Industria y Comercio

SUPERINTENDENCIA

OPOS

Organizacion CSA Ltda

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
INVENTARIO

NOU
EGI

SOLICITUD

PATENTE DE INVENCIÓN

21. EXPEDIENTE N° 07-137519

54. TITULO "GENES PARA DESHIDRINA Y PROMOTORES DEL CAFE"

51. CLASIFICACIÓN INTERNACIONAL C12N 5/04

71. SOLICITANTE NESTEC S.A. Y CORNELL RESEARCH FOUNDATION INC.

DOMICILIO VEVEY, SUIZA E ITHACA, ESTADO DE NUEVA YORK, ESTADOS UNIDOS DE AMERICA

74. REPRESENTANTE LEGAL ALICIA LLOREDA RICAURTE

T.P. 53.215

22. BOGOTÁ,D.C. 22 OCT 2007

(FORMA P 10)

(1)

PETITORIO

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

No. 07-137519- -00000-0000

Fecha: 2007-12-28 17:29:00 Dep. 2020 NUECREACIONI
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 578

Industria y Comercio
SUPERINTENDENCIA

FORMULARIO ÚNICO DE SOLICITUD DE PATENTE PCT ENTRADA EN FASE NACIONAL

06 Feb 2008

① SOLICITUD DE:

☒ Patente de Invención ☐ Patente de Modelo de Utilidad

☒ Capítulo I. ☐ Capítulo II.

② SOLICITANTE (71)	Nombre: **NESTEC S.A. Y CORNELL RESEARCH FOUNDATION INC.** Dirección: Avenue Nestlé 55, CH-1800 Vevey, Suiza y 20 Thornwood Drive, Suite 105 Ithaca, Nueva York 14850, Estados Unidos de América. Nacionalidad o Domicilio: Vevey, Suiza e Ithaca, Estado de Nueva York, Estados Unidos de América. Lugar de Constitución: Suiza y Estado de Nueva York, Estados Unidos de América. Teléfono: 3264270 Fax: 6069701 E-mail: jllcco@lloredacamacho.com	**IDENTIFICACION** C.C. ☐ NIT ☐ C.E. ☐ Otro ☐ Cual _____ Número _____
③ REPRESENTANTE O APODERADO	Nombre: **ALICIA LLOREDA RICAURTE** Dirección: Calle 72 No. 5-83 Piso 5º, Bogotá, Colombia Teléfono: 3264270 Fax: 6069701 E-mail: jllcco@lloredacamacho.com	C.C. ☒ NIT ☐ C.E. ☐ Otro ☐ Cual _____ Número 39.690.713 TP 53.215
④ INVENTOR(ES) (72)	Nombre: TANKSLEY, Steven, D.; LIN, Chenwei; BEN AMOR, Mohamed; MCCARTHY, James, Gerard y PETIARD, Vincent. Dirección: 17 Bridle Lane, Dryden, NY 13053; 600 Warren Rd, #4-3a, Ithaca, NY 14850; Avenue Nestlé 55, Tn-tebag-korba, XX 8070; 298 Vallee De Vautruchot, F-37210 Noizay XX y 28 Allee Cote Fleurie, F-37100 Tours XX. Nacionalidad o Domicilio: Estados Unidos de América, Tunisia y Francia.	

⑤ PCT (86)

Solicitud Internacional No. PCT/US2006/026234 Fecha: 05/07/06

Publicación Internacional No. WO 2007/005980 A3 Fecha: 11/01/07

⑥ Título (54) **"GENES PARA DESHIDRINA Y PROMOTORES DEL CAFÉ"**

⑦ Clasificación Internacional (51) C12N 5/04

⑧ Prioridad SI ☒ NO ☐

(33) País de Origen	(32) Fecha	(31) Número de Solicitud
US	06/07/05	60/696,890

⑨ Comprobante de Pago No. 07-98,720 **Fecha:** 4 de Diciembre de 2007

(10) **ANEXOS**

- [x] Comprobante de pago de la tasa de presentación de la solicitud.
- [x] Documento que acredita la existencia y representación legal cuando el solicitante sea persona jurídica.
- [x] Documento de Representación Legal.
- [ ] Documento de cesión del inventor al solicitante o a su causante.
- [ ] Declaraciones
- [ ] Copia de la solicitud internacional. (Resumen, descripción, reivindicaciones, dibujos)
- [x] Traducción de la solicitud internacional al castellano. (Resumen, descripción, reivindicaciones)
- [ ] Copia del examen preliminar internacional efectuado por la Autoridad Internacional de Examen Preliminar (IPEA).
- [ ] Traducción del examen preliminar internacional efectuado por la IPEA.
- [ ] De ser el caso, copia del contrato de acceso.
- [ ] De ser el caso, documento que acredite la licencia o autorización de uso de conocimientos tradicionales de las comunidades indígenas.
- [ ] Arte Final 12X12.
- [ ] De ser el caso, información sobre otras solicitudes de patente o títulos obtenidos en el extranjero por el mismo titular o su causante, relacionados parcial o totalmente con la invención de esta solicitud.
- [ ] De ser el caso, modificaciones efectuadas a la solicitud internacional.
- [x] Otros: reporte de búsqueda internacional con su traducción, documentos anexos al reporte de búsqueda, publicación con su traducción, opinión escrita de la autoridad de búsqueda internacional con su traducción, dibujos y listado de secuencias.

(11) **FIGURA CARACTERISTICA:**

(12) Solicito la concesión de la patente,

NOMBRE: ALICIA LLOREDA RICAURTE

FIRMA: _____

C.C. 39.690.713 Usaquén T.P. 53.215

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800176089-2

RECIBO OFICIAL DE CAJA : 07 - 98,720
FECHA : DICIEMBRE 4 DE 2007

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

No. 07-137519- -00000-0000

Fecha: 2007-12-28 17:29:00 Dep. 2020 NUECREACIONI
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 578

***** C O N S I G N A C I O N *****

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PGO	Vr. PAGO
LLOREDA Y CIA S. A	CONSIGNACION	BANCO POPULAR	050-00110-6	72074913	04/12/2007	9,522,900.00

***** C O N C E P T O *****

CANT.	RENTISTICO	CONCEPTO	TOTAL CONCEPTO
1	50005-01-01 SOLICITUDES	1 TRAMITES DE SOL. DE PATENTE DE IN	482,000.00
		TOTAL : $	482,000.00

SON: CUATROCIENTOS OCHENTA Y DOS MIL PESOS

RESPONSABLE : _____

RECIBO DE CAJA APLICADO AL EXPEDIENTE No._____

2

ANEXO

Los documentos que acreditan la calidad de REPRESENTANTES LEGALES en Bogotá de los doctores **ALICIA LLOREDA** y/o **GUSTAVO TAMAYO** y/o **IGNACIO SANTAMARIA** y/o **ENRIQUE ALVAREZ** para TODOS LOS ASUNTOS RELACIONADOS CON LA PROPIEDAD INDUSTRIAL de las sociedades peticionarias, que facultan a cada uno de dichos abogados para actuar individualmente, se están presentando con esta solicitud. En consecuencia, ruego a usted reconocer la personería de cada uno de dichos Representantes Legales para actuar dentro del presente expediente, a fin de que cualquiera de ellos pueda intervenir en cualquier tiempo respecto de la solicitud en él contenida.

ALICIA LLOREDA RICAURTE

T.P. No. 53.215

Nestec

TRADUCCIÓN OFICIAL No.283/04 DE UN DOCUMENTO ESCRITO EN FRANCÉS, EL CUAL PARA SU IDENTIFICACIÓN SE SELLA CON EL SELLO DEL TRADUCTOR. ESTA ES UNA TRADUCCIÓN OFICIAL EFECTUADA EL DIA 01 DE SEPTIEMBRE DE 2004 POR JEAN-JACQUES TURPIN, C.E.: 193849 DE BOGOTÁ, TRADUCTOR E INTÉRPRETE OFICIAL SEGÚN RESOLUCIÓN No. 2680/91 DEL 19 DE DICIEMBRE DE 1991, EXPEDIDA POR EL MINISTERIO DE JUSTICIA DE COLOMBIA.

Autenticaciones que se hallan a continuación de un documento de poder conferido en inglés y español por **NESTEC S.A.,** una sociedad organizada y existe de conformidad con las leyes de Suiza, domiciliada en Avenue Nestlé 55, CH-1800 Vevey, Suiza, y otorgado a los Doctores ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARÍA y/o ENRIQUE ÁLVAREZ, todos domiciliados en Bogotá, Colombia, Calle 72 No. 5 – 83 (Piso 6º).

Dado el 5 de agosto de 2004. Firma ilegible de la Señora Paula Nelson, Vice-Presidente de Nestec S.A.

CERTIFICADO NOTARIAL (SOCIEDAD)

Legalización por parte del suscrito Notario de la firma de la Señora **Paula NELSON** quien desempeña el cargo de **Vice-Presidente** de la compañía **NESTEC S.A.**, una sociedad legalmente constituida y existente de acuerdo con las leyes de **Suiza**, domiciliada en **Vevey (Suiza).** Además, dicho Notario certifica que el signatario, según las leyes de **Suiza**, tiene facultad para otorgar el presente documento. Los hechos anteriores le constan por haber examinado los documentos respectivos el 24 de marzo de 1997, y de todo ello da fe.

Firma de Jean-Daniel RUMPF.

Sello: Jean-Daniel RUMPF – Notario. Estampilla de 4 Frs.

283/04

Traducción Oficial No.
JEAN - JACQUES H. TURPIN
Traductor e Intérprete Oficial
FRANCES - ESPAÑOL
Res. Minjusticia No. 2680/91

COMO NOTARIO SEXTO (E) DEL CIRCULO DE BOGOTÁ, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA
09 NOV 2007
PABLO MENDEZ MIRANDA
NOTARIO (E)

4

Legalización No. 7.135. El Notario Jean-Daniel RUMPF, en Vevey, Distrito de Vevey, Cantón de Vaud, Suiza, certifica la firma puesta en la página anterior por la Señora Paula Nelson, Directora adjunta de Nestec S.A., en Vevey, la cual compromete en forma válida con su firma individual.

Dado en Vevey, el once de agosto de dos mil cuatro. Firma ilegible del Notario.

Sello: Jean-Daniel RUMPF – Notario. Estampilla de 4 Frs.

APOSTILLA

(Convención de La Haya del 5 de octubre de 1961)

1. - País: Suiza

   El presente documento público

2. - ha sido firmado por el Señor Jean-Daniel RUMPF

3. - obrando en calidad de Notario en Vevey

4. - con sello de Jean-Daniel RUMPF, Notario

CERTIFICADO

5. - En Lausana 6.- El 12 de agosto de 2004

7. - Por la Cancillería de Estado del Cantón de Vaud

8. - Bajo el número 6705

9. - Sello: Cancillería del Consejo de Estado – Libertad y Patria - Emolumento de 25 Fr.

10. - Por el Canciller de Estado, firma de P. Equey.

Es traducción fiel y completa de la cual se deja copia en los archivos de esta oficina para confrontación y a la cual me remito.

Traducción oficial No. 283/04.

Traductor: Jean-Jacques Turpin.

Bogotá, D.C., 1 de septiembre de 2004.

283/04

Traducción Oficial No.
JEAN - JACQUES H. TURPIN
Traductor e Intérprete Oficial
FRANCES - ESPAÑOL
Res. Minjusticia No. 2680/91

COMO NOTARIO SEXTO (E) DEL CIRCULO DE BOGOTA, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA

09 NOV 2007

PABLO MENDEZ BARAJAS
NOTARIO (E)

NO SE ASUME
RESPONSABILIDAD DEL TEXTO

2007 SEP -2 A 11: 09

JEFE DE LEGALIZACIONES
SANTAFE DE BOGOTA D.C.

MINISTERIO DE RELACIONES
EXTERIORES

061229

COTA FIRMA APARECE EN EL
DOCUMENTO DESEMPEÑA
LAS FUNCIONES INDICADAS

COMO NOTARIO SEXTO (E) DEL CIRCUL
DE BOGOTA, HAGO CONSTAR QUE LA
FOTOCOPIA CORRESPONDE CON EL ORIGINAL
QUE HE TENIDO A LA VISTA

09 NOV 2007

PABLO
NOTARIAS

Los abajo firmados, **NESTEC S.A.**, una sociedad organizada y existente de conformidad con las leyes de Suiza

con domicilio en **Avenue Nestlé 55, CH-1800 Vevey, Suiza**

nombramos por el presente a ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARIA y/o ENRIQUE ALVAREZ, domiciliados en Bogotá, Colombia, con oficinas en dicha ciudad, Calle 72 No. 5-83, piso 6o., nuestros representantes legales y verdaderos en Bogotá, con facultades para recibir notificaciones y nombrar apoderados judiciales o extrajudiciales, con el fin de dar cumplimiento al parágrafo 1o. del artículo 543 del decreto 410 de 1971.

Los doctores ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARIA y/o ENRIQUE ALVAREZ, o los apoderados que cualquiera de ellos designe, quedan especialmente facultados para representarnos ante las autoridades colombianas, administrativas, judiciales o de policía, en todos los trámites relativos a la propiedad industrial, a cuyo efecto les facultamos para dar ante dichas autoridades todos los pasos necesarios al objeto indicado; elevar solicitudes, declaraciones, reclamos, recursos y presentar desistimientos; formular descripciones, enmiendas, limitar el alcance de las solicitudes, presentar oposiciones y apelaciones; instaurar cancelaciones, renunciar total o parcialmente a los registros, abonar todos los impuestos y cuotas y efectuar todo otro pago determinado por la ley; recibir todos los documentos y valores, dando el descargo respectivo; llenar cualesquiera otros requisitos y tomar, en fin, todas las medidas que creyeren conducentes al resguardo de nuestros intereses; y en caso de producirse oposición, pasando los antecedentes a los tribunales, quedan facultados para tomar intervención como demandantes o demandados ante los jueces que sean competentes, pudiendo transar, someter a árbitros, desistir, percibir y apelar, con todas las demás facultades que resulten necesarias, y por el presente declaramos desde ahora válido y bueno cuanto dichos representantes y/o los apoderados que cualquiera de ellos nombre hicieren en nuestro beneficio.

Asimismo, facultamos a los doctores ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARIA y/o ENRIQUE ALVAREZ, para solicitar en la Oficina de Propiedad Industrial la protocolización de este documento y de cualquier poder que cualquiera de ellos otorgue a nombre nuestro.

Dado y firmado

The undersigned, **NESTEC S.A.**, a corporation organized and existing under the laws of Switzerland.

domiciled at **Avenue Nestlé 55, CH-1800 Vevey, Switzerland**

hereby appoint, ALICIA LLOREDA and/or GUSTAVO TAMAYO and/or IGNACIO SANTAMARIA and/or ENRIQUE ALVAREZ, of Bogota, Colombia, with offices in said town, Calle 72 No. 5-83, Sixth Floor, our true and legal representatives in Bogotá, with power to receive notifications and to appoint judicial or extrajudicial agents, so as to comply with paragraph 1o. of article 543 of decree 410 of 1971.

Doctors ALICIA LLOREDA and/or GUSTAVO TAMAYO and/or IGNACIO SANTAMARIA and/or ENRIQUE ALVAREZ, or the agents by any one of them appointed, are especially empowered to represent us before the administrative, judicial or police authorities of Colombia, in all proceedings pertaining to industrial property, for which purpose we authorize them to take before the said authorities all the necessary steps for the said end; to file petitions, declarations, claims, recourses and withdrawals, to draw up specifications; to file amendments, to limit the scope of applications, file oppositions; file cancellations total or partially waive registrations, to pay all installments and effect all other payments prescribed by law; to receive proper receipts, to comply with all other requirements and, in short, to take all those measures which they may deem fit for the protection of our interests; and in case of opposition, where the proceedings be passed over to the courts, we authorize said agents to act on our behalf, as complainants as well as defendants, before the competent judges, with full powers to make arrangements, to submit to arbitrators, to desist, to collect, to appeal and, in short, with all such further powers as may be required, and we hereby declare from now on as valid and good all that said representatives and/or the agents by any one of them appointed may do for our benefit.

We also authorize doctors ALICIA LLOREDA and/or GUSTAVO TAMAYO and/or IGNACIO SANTAMARIA and/or ENRIQUE ALVAREZ, to apply for protocolization of this document and of any power of attorney by any one of them conferred in our name, before the Office of Industrial Property.

Given and signed this 5th August 2004

Paula NELSON – Vice President

NOTARY SHOULD EXECUTE OTHER SIDE

DE BOGOTA, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA

09 NOV 2007

PABLO MEJIA DE LA CRUZ
NOTARIO (E)

CERTIFICADO NOTARIAL (SOCIEDAD)

Suiza
Vevey {ss

Hoy 11 de Agosto de 2004, compareció PAULA NELSON, a quien conozco y de quien me consta que firmó el anterior documento, y quien debidamente juramentado declaró y dijo que él es Vicepresidente de **NESTEC S.A.**, la compañía nombrada en el documento que antecede, una sociedad debidamente organizada y existente conforme a las leyes de **Suiza**, domiciliada en la ciudad de **Vevey**, **Suiza** legalmente constituida para un período de duración que aún no ha terminado, que ejerce su objeto social; conforme a las leyes de **Suiza**, que el declarante está autorizado por dicha Compañía para conferir este poder; que el sello estampado por él al pie de tal poder, en nombre y en representación de la expresada Compañía, es el sello oficial de ella; y que el objeto para el cual se confiere este instrumento está comprendido dentro del de la sociedad.
El suscrito Notario ha visto los Estatutos de incorporación firmados en **Vevey**, **Suiza**, con fecha 24 Mar 20, 1997, y certifica que de conformidad con la legislación de dicho País, la anterior declaración juramentada es prueba suficiente de que los hechos sobre que ella versa son ciertos o verdaderos, de lo cual da fe.

Notary Public (Notario Público).

NOTARIAL CERTIFICATE (CORPORATION)

Switzerland
Vevey {ss

On this 11th day of August, 2004, personally came *Paula NELSON*, to me known, and known to me to be the person who signed the foregoing instrument, who being by me duly sworn did depose and say that he/she is the *Vice President* of **NESTEC S.A.**, the company named within, a Corporation organized and existing under the laws of the **Switzerland**, domiciled at the City of **Vevey**, **Switzerland** legally constituted for a period of tenure not yet expired, which exercises its corporate purposes according to the laws of **Switzerland**, that the deponent is authorized by said Company to confer this power of attorney; that the seal affixed hereunto by the deponent in the name and on behalf of said Corporation is the corporate seal thereof; and that the purpose for which this instrument is granted is included within the corporate purposes of the Company.
The undersigned Notary Public has seen the articles of Incorporation of said Corporation executed at **Vevey**, **Switzerland**, under date of March 24 1997 and hereby certifies that under the laws of said County the foregoing acknowledgment is sufficient proof that the facts therein contained are true, which he attests.

J.-D. RUMPF NOTAIRE

Légalisation No 7'135.-
Le notaire JEAN-DANIEL RUMPF, à Vevey, District de Vevey, Canton de Vaud, Suisse, certifie authentique la signature apposée d'autre --- part par Mme Paula NELSON, directrice-adjointe de la société ----- NESTEC S.A., à Vevey, qu'elle engage valablement par sa signature - individuelle. ----------
VEVEY, LE ONZE AOÛT DEUX MILLE QUATRE. ----------

J.-D. RUMPF NOTAIRE

COMO NOTARIO SEXTO (E) DEL CIRCULO DE BOGOTA, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA
09 NOV 2007
PABLO MENDEZ BARAJAS
NOTARIO (E)

APOSTILLE
(Convention de La Haye du 5 octobre 1961)

1. Pays: SUISSE

Le présent acte public

2. a été signé par Jean-Daniel Rumpf

3. agissant en qualité de notaire à Vevey

4. est revêtu du sceau/timbre de J.-D. Rumpf notaire

Attesté

5. à Lausanne 6. le 12 août 2004

7. par la Chancellerie d'Etat du Canton de Vaud

8. sous No 6705

9. Sceau/timbre:

10. Signature
pr le Chancelier d'Etat:

P.Equey

Emolument: Fr. 25.-

COMO NOTARIO SEXTO (E) DEL [illegible]
DE BOGOTA, D.C. CONSTA QUE LA
FOTOCOPIA COINCIDE CON EL ORIGINAL
QUE HE TENIDO A LA VISTA

19 NOV 2007

[illegible] ARAJAS
NOTARIO (E)

TRADUCCION OFICIAL No.2004-11-08-**531** de un documento escrito en Inglés/ Español, la cual para efectos del presente lleva el sello personal de Martha Lucia Arciniegas M., Traductora e Intérprete Oficial en virtud de la Res 1523 /88 del Ministerio de Justicia.

PODER

Legalización de un Poder de Representación Legal conferido a ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARIA y/o ENRIQUE ALVAREZ por **CORNELL RESEARCH FOUNDATION INC.**

Documento escrito simultáneamente en inglés y en español

Dado y firmado el 9 de agosto del 2004

Por: Richard S. Cahoon

Vice Presidente

Procedimientos de Autenticación

Certificado Notarial (Corporación)

Legalización de la firma precedente

(fdo)

ROSEMARY REYNOLS

Notaria Pública, Estado de New York
Calificada en el Condado de Tompkins

No. 4895247

Comisión vence el 18 de Mayo del 2004

COMO NOTARIO SEXTO (E) DEL CIRCULO DE BOGOTA, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA

04 DIC 2007

PABLO MENDEZ CARAJAS
NOTARIO (E)

MARTHA LUCIA ARCINIEGAS M.
TRADUCTORA E INTERPRETE OFICIAL
RESOLUCION 1523/88 MINJUSTICIA

8

Estado de Nueva York)

) ss:

Oficina del Secretario del Condado de Tompkins)

Yo, **AURORA R. VALENTI,** Secretaria del mencionado Condado y de la Corte Suprema adjunta y para el mencionado Condado, siendo Tribunales Registro, que tiene por ley un sello, mediante el presente certifico que ROSEMARY REYNOLS

la funcionaria cuyo nombre está suscrito a la declaración extra- tribunal o certificado de evidencia adjunto, y ante quienes el mismo fue realizado, reside en el mencionado Condado; que en el momento de tomar dicha declaración, prueba o reconocimiento era un NOTARIO PUBLICO en y para el mencionado Condado, debidamente autorizado por las leyes del mencionado Estado, para tomar el mismo, y también para tomar prueba de reconocimiento de escrituras para territorios, escrituras de inmuebles, viviendas y propiedades para ser registradas en el Estado. Además certifico que reconozco que estoy familiarizada con la escritura de dicha funcionaria y verdaderamente creo que la firma de dicho certificado de prueba de reconocimiento es genuina y que dicho documento es ejecutado y reconocido de conformidad con las leyes del mencionado Estado

COMO NOTARIO SEXTO (E) DEL CIRCULO DE SANTA ELENA CERTIFICO QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA

04 DIC 2007

PABLO MENDEZ BARAJAS
NOTARIO (E)

En testimonio de lo cual

Firmo y coloco mi Sello oficial en Ithaca, New York, en el mencionado Condado este 13 de Agosto del 2004

(firma ilegible)_____Secretario _____Secretario Encargado

COMO NOTARIO SEXTO (E) DEL CIRCULO DE BOGOTA, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA

04 DIC 2007

PABLO MENDEZ BARAJAS
NOTARIO (E)

APOSTILLA

(*Convention de La Haye du Octobre 1961*)

1. País: **Estados Unidos de América**

Este documento público

2. ha sido firmado por **Aurora R. Valenti**

3. actuando como **Secretaria del Condado**

4. lleva el sello/estampilla del **Condado de Tompkins**

Certificado

5. En Nueva York, Nueva York 6. este día 18 de agosto del 2004

7. Por el Secretario Especial Encargado de Estado, Estado de Nueva York

8. No. NYC-10300839A

9. Sello/Estampilla

Aparece el Gran Sello del Estado de Nueva York.

10. Firma

(firma: ilegible)

James Bizzarri

Secretario Encargado Especial del Estado Nueva York

COMO NOTARIO SEXTO (E) DEL CIRCULO DE BOGOTA, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA

04 DIC 2007

PABLO MENDEZ ARANJAS
NOTARIO (E)

Según mi legal saber y entender, esta es una traducción fiel y completa de su original en inglés de la cual se deja copia para futuras referencias.

MARTHA LUCIA ARCINIEGAS M.
TRADUCTORA E INTERPRETE OFICIAL
RESOLUCION 1523/88 MINJUSTICIA

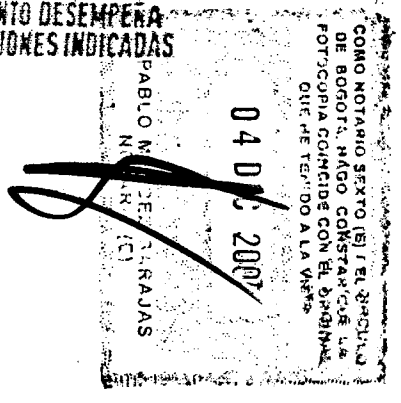
MINISTERIO DE
RELACIONES EXTERIORES

283166
Martha Arciniegas
CUYA FIRMA APARECE EN EL
DOCUMENTO DESEMPEÑA
LAS FUNCIONES INDICADAS

NO SE ASUME
RESPONSABILIDAD DEL TEXTO

2004 NOV 10 P 10:41

JEFE DE LEGALIZACIONES
BOGOTA D.C.



Cornell Research
JOSE ELORRIETA CAMACHO & CO.
98-016.61

Apostille

(Convention de La Haye du 5 Octobre 1961)

1. Country: **United States of America**

   This public document

2. has been signed by **Aurora R. Valenti**

3. acting in the capacity of **County Clerk**

4. bears the seal/stamp of the county of **Tompkins**

Certified

5. At New York, New York 6. the 18th day of August 2004

7. by Special Deputy Secretary of State, State of New York

8. No. NYC-10300839A

9. Seal/Stamp 10. Signature

COMO NOTARIO SEXTO (E) DEL CIRCULO DE BOGOTA, HAGO CONSTAR QUE LA FOTOCOPIA COINCIDE CON EL ORIGINAL QUE HE TENIDO A LA VISTA
04 DIC 2007
PA[illegible] MENDEZ BARAJAS
NOTARIO (E)

James Bizzarri
Special Deputy Secretary of State

04135595.RSL (REV: 2/6/96)

12

0030

Los abajo firmados, **CORNELL RESEARCH FOUNDATION INC.**, una sociedad organizada y existente de conformidad con las leyes del Estado de Nueva York, Estados Unidos de América.

Cornell Business & Technology Park

con domicilio en **20 Thornwood Drive, Suite 105**
Ithaca, Nueva York 14850,
Estados Unidos de América

nombramos por el presente a ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARIA y/o ENRIQUE ALVAREZ, domiciliados en Bogotá, Colombia, con oficinas en dicha ciudad, Calle 72 No. 5-83, piso 6o., nuestros representantes legales y verdaderos en Bogotá, con facultades para recibir notificaciones y nombrar apoderados judiciales o extrajudiciales, con el fin de dar cumplimiento al parágrafo 1o. del artículo 543 del decreto 410 de 1971.

Los doctores ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARIA y/o ENRIQUE ALVAREZ, o los apoderados que cualquiera de ellos designe, quedan especialmente facultados para representarnos ante las autoridades colombianas, administrativas, judiciales o de policía, en todos los trámites relativos a la propiedad industrial, a cuyo efecto les facultamos para dar ante dichas autoridades todos los pasos necesarios al objeto indicado; elevar solicitudes, declaraciones, reclamos, recursos y presentar desistimientos; formular descripciones, enmiendas, limitar el alcance de las solicitudes, presentar oposiciones y apelaciones; instaurar cancelaciones, renunciar total o parcialmente a los registros, abonar todos los impuestos y cuotas y efectuar todo otro pago determinado por la ley; recibir todos los documentos y valores, dando el descargo respectivo; llenar cualesquiera otros requisitos y tomar, en fin, todas las medidas que creyeren conducentes al resguardo de nuestros intereses; y en caso de producirse oposición, pasando los antecedentes a los tribunales, quedan facultados para tomar intervención como demandantes o demandados ante los jueces que sean competentes, pudiendo transar, someter a árbitros, desistir, percibir y apelar, con todas las demás facultades que resulten necesarias, y por el presente declaramos desde ahora válido y bueno cuanto dichos representantes y/o los apoderados que cualquiera de ellos nombre hicieren en nuestro beneficio.

Asimismo, facultamos a los doctores ALICIA LLOREDA y/o GUSTAVO TAMAYO y/o IGNACIO SANTAMARIA y/o ENRIQUE ALVAREZ, para solicitar en la Oficina de Propiedad Industrial la protocolización de este documento y de cualquier poder que cualquiera de ellos otorgue a nombre nuestro.

Dado y firmado

The undersigned, **CORNELL RESEARCH FOUNDATION INC.**, a corporation organized and existing under the laws of State of New York, United States of America.

Cornell Business & Technology Park

domiciled at **20 Thornwood Drive, Suite 105**
Ithaca, New York 14850,
United States of America

hereby appoint, ALICIA LLOREDA and/or GUSTAVO TAMAYO and/or IGNACIO SANTAMARIA and/or ENRIQUE ALVAREZ, of Bogota, Colombia, with offices in said town, Calle 72 No. 5-83, Sixth Floor, our true and legal representatives in Bogotá, with power to receive notifications and to appoint judicial or extrajudicial agents, so as to comply with paragraph 1o. of article 543 of decree 410 of 1971.

Doctors ALICIA LLOREDA and/or GUSTAVO TAMAYO and/or IGNACIO SANTAMARIA and/or ENRIQUE ALVAREZ, or the agents by any one of them appointed, are especially empowered to represent us before the administrative, judicial or police authorities of Colombia, in all proceedings pertaining to industrial property, for which purpose we authorize them to take before the said authorities all the necessary steps for the said end; to file petitions, declarations, claims, recourses and withdrawals, to draw up specifications; to file amendments, to limit the scope of applications, file oppositions; file cancellations total or partially waive registrations, to pay all installments and effect all other payments prescribed by law; to receive proper receipts, to comply with all other requirements and, in short, to take all those measures which they may deem fit for the protection of our interests; and in case of opposition, where the proceedings be passed over to the courts, we authorize said agents to act on our behalf, as complainants as well as defendants, before the competent judges, with full powers to make arrangements, to submit to arbitrators, to desist, to collect, to appeal and in short with all such further powers as may be required and we hereby declare from now on as valid and good all that said representatives and/or the agents by any one of them appointed may do for our benefit.

FOTOCOPIA COINCIDE CON EL ORIGINAL
PABLO ENRIQUEZ BARAJAS
NOTARIO (E)

We also authorize doctors ALICIA LLOREDA and/or GUSTAVO TAMAYO and/or IGNACIO SANTAMARIA and/or ENRIQUE ALVAREZ, to apply for protocolization of this document and of any power of attorney by any one of them conferred in our name, before the Office of Industrial Property.

Given and signed this 9th day of August 2004

Richard S. Cahoon
Sr. Vice President

13

CERTIFICADO NOTARIAL (SOCIEDAD)

Estados Unidos de América
Estado de **Nueva York** {ss
Condado de **Tompkins**

Hoy 9 de Agosto de 2004, compareció Richard S. Cahoon, a quien conozco y de quien me consta que firmó el anterior documento, y quien debidamente juramentado declaró y dijo que él es Vice Presidente de **CORNELL RESEARCH FOUNDATION INC.** la compañía nombrada en el documento que antecede, una sociedad debidamente organizada y existente conforme a las leyes del Estado de **Nueva York**, Estados Unidos de América, domiciliada en la ciudad de **Ithaca**, Condado de **Tompkins**, Estado de **Nueva York**, Estados Unidos de América, legalmente constituida para un período de duración que aún no ha terminado, que ejerce su objeto social; conforme a las leyes del Estado de **Nueva York**, Estados Unidos de América; que el declarante está autorizado por dicha Compañía para conferir este poder; que el sello estampado por él al pie de tal poder, en nombre y en representación de la expresada Compañía, es el sello oficial de ella; y que el objeto para el cual se confiere este instrumento está comprendido dentro del de la sociedad.

El suscrito Notario ha visto los Estatutos de incorporación firmados en Ithaca, Estado de Nueva York, con fecha Enero 11, 1932, y certifica que de conformidad con la legislación de dicho Estado la anterior declaración juramentada es prueba suficiente de que los hechos sobre que ella versa son ciertos o verdaderos, de lo cual da fe.

Notary Public (Notario Público).

NOTARIAL CERTIFICATE (CORPORATION)

United States of America
State of **New York** {ss
County of **Tompkins**

On this 9th day of August, 2004, personally came Richard S. Cahoon, to me known, and known to me to be the person who signed the foregoing instrument, who being by me duly sworn did depose and say that he is the Sr. Vice President of **CORNELL RESEARCH FOUNDATION INC.** the company named within, a Corporation organized and existing under the laws of the State of **New York**, United States of America, domiciled at the City of **Ithaca**, County of **Tompkins**, State of **New York**, United States of America, legally constituted for a period of tenure not yet expired, which exercises its corporate purposes according to the laws of the State of **New York**, United States of America; that the deponent is authorized by said Company to confer this power of attorney; that the seal affixed hereunto by the deponent in the name and on behalf of said Corporation is the corporate seal thereof; and that the purpose for which this instrument is granted is included within the corporate purposes of the Company.

The undersigned Notary Public has seen the articles of Incorporation of said Corporation executed at Ithaca, State of New York, under date of January 11, 1932, and hereby certifies that under the laws of said State the foregoing acknowledgment is sufficient proof that the facts therein contained are true, which he attests.

Rosemary Reynolds

ROSEMARY REYNOLDS
Notary Public, State of New York
No. 4895247
Qualified in Tompkins County
Term Expires May 18, 20__

ESTO (E) DEL CIRCULO
AGO CONSTAR QUE LA
EL ORIGINAL
A VISTA
04 DIC 2007
PABLO MENDEZ BARAJAS
NOTARIO (E)

State of New York
Tompkins County Clerk's Office } ss.

I, AURORA R. VALENTI, Clerk of said County, and of the Supreme and County Courts held in and for said County, being Courts of Record having a seal, do hereby certify that Rosemary Reynolds the officer whose name is subscribed to the deposition or certificate of proof or acknowledgment of the annexed instrument, and before whom the same was made, resides in said County: that at time of taking such deposition, proof or acknowledgment was a NOTARY PUBLIC in and for said County, duly authorized by the Laws of said State to take the same, and also to take proof of acknowledgment of deeds for lands, tenements and hereditaments to be recorded in the State. I further certify that I am well acquainted with the handwriting of such officer, and verily believe that the signature to such certificate of proof of acknowledgment is genuine, and that said instrument is executed and acknowledged in conformity with the laws of said State.

In testimony whereof I have hereunto set my hand and affixed my official Seal, at Ithaca, New York, in said County this 13 day of August, 2004

Aurora R Valenti, Clerk _____, Deputy Clerk

14

Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
DIVISIÓN DE NUEVAS CREACIONES
EXTRACTO PARA PUBLICACIÓN
SOLICITUD DE PATENTE DE INVENCIÓN PCT

(21) N° de solicitud:

(22) Fecha de solicitud: 28-12-07

(30) Prioridad (31) ***No. Prioridad*** (32) ***Fecha*** (33) ***País***
60/696,890 06-07-05 US

(51) Int Cl: C12N 5/04

(71) Solicitante(s): **NESTEC S.A. Y CORNELL RESEARCH FOUNDATION INC.**

(72) Inventor(es): TANKSLEY, Steven, D.; LIN, Chenwei; BEN AMOR, Mohamed; MCCARTHY, James, Gerard y PETIARD, Vincent.

(74) Apoderado: **ALICIA LLOREDA RICAURTE**

(85) Fecha límite inicio fase nacional: 06-02-08
(86) Datos relativos a la presentación de la solicitud PCT
Fecha de presentación de la solicitud: 05-07-06
No. de solicitud: ***PCT/US2006/026234***

(87) Publicación internacional
Fecha: ***11-01-07***
N° publicación: ***WO 2007/005980 A3***

(54) Titulo: **GENES PARA DESHIDRINA Y PROMOTORES DEL CAFÉ**

Resumen - Reivindicación(es):

Reivindicación número 1.-

Una molécula de ácido nucleico aislada de café (género *Coffea*) que tiene una secuencia codificante que codifica para una deshidrina o una proteína abundante embriogénica tardía (LEA).

REPORTE DE BÚSQUEDA INTERNACIONAL

Solicitud Internacional No.
PCT/US2006/026234

A. CLASIFICACION DE MATERIA OBJETO
INV. C07K14/415 C12N15/82 C12N5/04

Según la Clasificación de Patente Internacional (IPC) o según la clasificación nacional y la IPC

B. CAMPOS BUSCADOS

Documentación mínima buscada (sistema de clasificación seguido de símbolos de clasificación)
C07K C12N

Documentación buscada fuera de documentación mínima hasta donde dichos documentos se incluyan en los campos buscados

Base de datos electrónica consultada durante la búsqueda internacional (nombre de la base de datos y, donde sea práctico, buscar términos usados)
EPO-Internacional, BIOSIS, EMBASE, WPI Dato, Búsqueda de Secuencia

C. DOCUMENTOS CONSIDERADOS RELEVANTES

Categoría °	Citación de documento, con indicación, donde sea apropiado, de los textos relevantes	Relevante a la reivindicación No.
X	KASUGA MIE Y COL: "Una combinación del gen Arabidopsis DREB1A y el promotor rd29A inducible por estrés mejoran la tolerancia al estrés hídrico y de baja temperatura en tabaco por transferencia de gen" PLANT & CELL PHYSIOLOGY MARZO DE 2004, vol. 45, no. 3, Marzo de 2004 (2004-03), páginas 346-350, XP002421281 ISSN: 0032-0781	1-3, 5-39, 41-47
A	secuencias NTERD10B, AC AB049336	4

	-/--	

☒ En el recuadro C se enumeran más documentos.

☒ En el anexo se enumeran miembros de la familia de patentes.

° Categorías especiales de documentos citados:

'A' documento que define el estado general de la técnica el cual no se considera de relevancia particular

'E' documento anterior pero publicado en o después de la fecha de presentación internacional

'L' documento que puede arrojar dudas sobre reivindicaciones de prioridad o que se cita para establecer la fecha de publicación de otra citación u otra razón especial (según se especifica)

'O' documento que se refiere a una revelación oral, uso, exhibición u otros medios

'P' documento publicado antes de la fecha de presentación internacional pero después de la fecha de prioridad reivindicada

'T' documento posterior publicado después de la fecha de presentación internacional o fecha de prioridad y que no está en conflicto con la solicitud pero citado para entender el principio o teoría que sustenta la invención

'X' documento de relevancia particular; la invención reivindicada no puede considerarse nueva o no puede considerarse que comprende un paso inventivo cuando el documento se toma solo

'Y' documento de relevancia particular; la invención reivindicada no puede considerarse que comprende un paso inventivo cuando el documento se combina con uno o más documentos, siendo dicha combinación obvia para una persona experta en la técnica

'&' documento miembro de la misma familia de patentes

Fecha de la terminación real de la búsqueda internacional	Fecha de envío por correo del reporte de búsqueda internacional
6 de Junio de 2007	20/06/2007

Nombre y dirección de correo de la ISA/	Funcionario autorizado
Oficina de Patentes Europea, P. B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	LOUBRADOU-BOURGES, N

18

C. (Continuación) DOCUMENTOS CONSIDERADOS RELEVANTES

Categoría °	Citación de documento, con indicación, donde sea apropiado, de los textos relevantes	Relevante a la reivindicación No.
X	FERNANDEZ DIANA Y COL: "Genes de café (Coffea arabica L) tempranamente expresados durante la infección por el hongo de la roya (Hemileia vastatrix)" MOLECULAR PLANT PATHOLOGY, vol. 5, no. 6, Noviembre de 2004 (2004-11), páginas 527-536, XP002421282 ISSN: 1464-6722	1-3, 5-39, 41-47
A	página 532; tabla 2	4

X	WO 03/027249 A (SYNGENTA PARTICIPATIONS AG [CH]; SU WENPEI [US]; ANDON NANCY [US]; HAY) 3 de Abril de 2003 (2003-04-03)	1-3, 5-39, 41-47
A	reivindicación 1; secuencia 34	4

X	WO 01/77384 A (EPIGENOMICS AG [DE]; OLEK ALEXANDER [DE]; PIEPENBROCK CHRISTIAN [DE]); 18 de Octubre de 2001 (2001-10-18) secuencias 25203, 17545, 207467	10

X	KAYE CLAUDIA Y COL: "Caracterización de un gen de espinaca CAP160 y expresión de dos proteínas de aclimatación-fría de espinaca en tabaco" PLANT PHYSIOLOGY (ROCKVILLE), vol. 116, no. 4, Abril de 1998 (1998-04), páginas 1367-1377, XP002421536 ISSN: 0032-0889	1-3, 5-39, 41-47
A	el documento completo	4

A	BRETON G Y COL: "Aplicaciones Biotecnológicas de proteínas asociadas a congelamiento de plantas" BIOTECHNOLOGY ANNUAL REVIEW 2000, vol. 6, 2000, páginas 59-101, XP001062997 ISSN: 1387-2656 página 85	1-39, 41-47

A	ALLAGULOVA CH R Y COL: "Las dehidrinas de plantas: estructura y funciones" BIOCHEMISTRY. BIOKHIMII A SEPTIEMBRE DE 2003, vol. 68, no. 9, Septiembre de 2003 (2003-09), páginas 945-951, XP009079176 ISSN: 0006-2979 el documento completo	1-39, 41-47

A	WISE MICHAEL J Y COL: "Pregunta de POPP: Qué hacen las proteínas LEA?" TRENDS IN PLANT SCIENCE, vol. 9, no. 1, Enero de 2004 (2004-01), páginas 13-17, XP002421537 ISSN: 1360-1385 el documento completo	1-39, 41-47

	-/--	

C. (Continuación) DOCUMENTOS CONSIDERADOS RELEVANTES

Categoría °	Citación de documento, con indicación, donde sea apropiado, de los textos relevantes	Relevante a la reivindicación No.
P, X	WO 2005/122697 A (CARMEL HAIFA UNIVERSITY ECONOM [IL]; NEVO EVIATAR [IL]; SONG WEINING) 29 de Diciembre de 2005 (2005-12-29) reivindicaciones 1-13	1-3, 5-39, 41-47

P, X	CHENWEI LIN Y COL: "Café y tomate comparten repertorios comunes de genes como lo revela una secuenciación profunda de semillas y transcriptos de cereza" THEORETICAL AND APPLIED GENETICS; INTERNATIONAL JOURNAL OF PLANT BREEDING RESEARCH, SPRINGER-VERLAG, BE, vol. 112, no. 1, 1 de Diciembre de 2005 (2005-12-01), páginas 114-130, XP019322122 ISSN: 1432-2242 el documento completo	1-39, 41-47

P, X	HINNIGER CÉCILE Y COL: "Aislamiento y caracterización de cADN codificando tres dehidrinas expresadas durante el desarrollo del grano de Coffea canephora (Robusta)" ANNALS OF BOTANY MAYO DE 2006, vol. 97, no. 5, Mayo de 2006 (2006-05), páginas 755-765, XP009079141 ISSN: 0305-7364 el documento completo	1-39, 41-47

X	SATOS Y COL: "Análisis estructural del cromosoma 3 de Arabidopsis thaliana. I. Características secuenciales de las regiones de 4,504,864 pb cubiertas por los clones sesenta P1 y TAC" GENBANK, 28 de Abril 2000 (2000-04-28), XP002974277 secuencia Q9LT74	1-32

REPORTE DE BÚSQUEDA INTERNACIONAL

Información sobre miembros de familia de patentes

Solicitud Internacional No.
PCT/US2006/026234

Documento de patente citado en el reporte de búsqueda	Fecha de publicación	Miembro(s) de la familia de patentes	Fecha de publicación
WO 03027249 A	03-04-2003	EP 1576163 A2	21-09-2005
WO 0177384 A	18-10-2001	AU 5057201 A	23-10-2001
		EP 1268856 A2	02-01-2003
		US 2004241651 A1	02-12-2004
WO 2005122697 A	29-12-2005	NINGUNO	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2006/026234

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/415 C12N15/82 C12N5/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KASUGA MIE ET AL: "A combination of the Arabidopsis DREB1A gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer." PLANT & CELL PHYSIOLOGY MAR 2004, vol. 45, no. 3, March 2004 (2004-03), pages 346-350, XP002421281 ISSN: 0032-0781	1-3, 5-39, 41-47
A	sequences NTERD10B, AC AB049336	4
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 June 2007	20/06/2007

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	LOUBRADOU-BOURGES, N

Form PCT/ISA/210 (second sheet) (April 2005)

22

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2006/026234

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FERNANDEZ DIANA ET AL: "Coffee (Coffea arabica L.) genes early expressed during infection by the rust fungus (Hemileia vastatrix)" MOLECULAR PLANT PATHOLOGY, vol. 5, no. 6, November 2004 (2004-11), pages 527-536, XP002421282 ISSN: 1464-6722	1-3, 5-39, 41-47
A	page 532; table 2	4
X	WO 03/027249 A (SYNGENTA PARTICIPATIONS AG [CH]; SU WENPEI [US]; ANDON NANCY [US]; HAY) 3 April 2003 (2003-04-03)	1-3, 5-39, 41-47
A	claim 1; sequence 34	4
X	WO 01/77384 A (EPIGENOMICS AG [DE]; OLEK ALEXANDER [DE]; PIEPENBROCK CHRISTIAN [DE];) 18 October 2001 (2001-10-18) sequences 25203,17545,207467	10
X	KAYE CLAUDIA ET AL: "Characterization of a gene for spinach CAP160 and expression of two spinach cold-acclimation proteins in tobacco" PLANT PHYSIOLOGY (ROCKVILLE), vol. 116, no. 4, April 1998 (1998-04), pages 1367-1377, XP002421536 ISSN: 0032-0889	1-3, 5-39, 41-47
A	the whole document	4
A	BRETON G ET AL: "Biotechnological applications of plant freezing associated proteins." BIOTECHNOLOGY ANNUAL REVIEW 2000, vol. 6, 2000, pages 59-101, XP001062997 ISSN: 1387-2656 page 85	1-39, 41-47
A	ALLAGULOVA CH R ET AL: "The plant dehydrins: structure and putative functions." BIOCHEMISTRY. BIOKHIMII A SEP 2003, vol. 68, no. 9, September 2003 (2003-09), pages 945-951, XP009079176 ISSN: 0006-2979 the whole document	1-39, 41-47
A	WISE MICHAEL J ET AL: "POPP the question: What do LEA proteins do?" TRENDS IN PLANT SCIENCE, vol. 9, no. 1, January 2004 (2004-01), pages 13-17, XP002421537 ISSN: 1360-1385 the whole document	1-39, 41-47

-/--

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 2005/122697 A (CARMEL HAIFA UNIVERSITY ECONOM [IL]; NEVO EVIATAR [IL]; SONG WEINING []) 29 December 2005 (2005-12-29) claims 1-13	1-3, 5-39, 41-47
P,X	CHENWEI LIN ET AL: "Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts" THEORETICAL AND APPLIED GENETICS ; INTERNATIONAL JOURNAL OF PLANT BREEDING RESEARCH, SPRINGER-VERLAG, BE, vol. 112, no. 1, 1 December 2005 (2005-12-01), pages 114-130, XP019322122 ISSN: 1432-2242 the whole document	1-39, 41-47
P,X	HINNIGER CÉCILE ET AL: "Isolation and characterization of cDNA encoding three dehydrins expressed during Coffea canephora (Robusta) grain development." ANNALS OF BOTANY MAY 2006, vol. 97, no. 5, May 2006 (2006-05), pages 755-765, XP009079141 ISSN: 0305-7364 the whole document	1-39, 41-47
X	SATOS ET AL: "Structural analysis of arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones" GENBANK, 28 April 2000 (2000-04-28), XP002974277 sequence Q9LT74	1-32

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2006/026234

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 03027249 A	03-04-2003	EP 1576163 A2	21-09-2005
WO 0177384 A	18-10-2001	AU 5057201 A	23-10-2001
		EP 1268856 A2	02-01-2003
		US 2004241651 A1	02-12-2004
WO 2005122697 A	29-12-2005	NONE	

DOCUMENTOS ANEXOS AL REPORTE DE BÚSQUEDA INTERNACIONAL

1. XP002421281
2. XP002421282
3. WO 03/027249 A
4. WO 01/77384 A
5. XP002421536
6. XP001062997
7. XP009079176
8. XP002421537
9. WO 2005/122697 A
10. XP019322122
11. XP009079141
12. XP002974277

XP-002421281

Plant Cell Physiol. 45(3): 346–350 (2004)
JSPP © 2004

Short Communication

A Combination of the *Arabidopsis* DREB1A Gene and Stress-Inducible *rd29A* Promoter Improved Drought- and Low-Temperature Stress Tolerance in Tobacco by Gene Transfer

Mie Kasuga [1], Setsuko Miura [1], Kazuo Shinozaki [2,3] and Kazuko Yamaguchi-Shinozaki [1,3,4]

[1] *Biological Resources Division, Japan International Research Center for Agricultural Sciences (JIRCAS), 1-1 Ohwashi, Tsukuba, Ibaraki, 305-8686 Japan*

[2] *Laboratory of Plant Molecular Biology, RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, Ibaraki, 305-0074 Japan*

[3] *Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Japan*

The transcription factor DREB1A/CBF3 specifically interacts with the dehydration responsive element (DRE/CRT) and induces expression of genes involved in environmental stress tolerance in *Arabidopsis*. Overexpression of *DREB1A* improved drought- and low-temperature stress tolerance in tobacco. The stress-inducible *rd29A* promoter minimized the negative effects on the plant growth in tobacco. Furthermore, we detected overexpression of stress-inducible target genes of DREB1A in tobacco. These results indicate that a combination of the *rd29A* promoter and DREB1A is useful for improvement of various kinds of transgenic plants that are tolerant to environmental stress.

Keywords: Drought tolerance — Low-temperature stress tolerance — Stress-inducible promoter — Transgenic tobacco — Transcription factor.

Abbreviations: CaMV, cauliflower mosaic virus; DRE, dehydration responsive element; LEA, late embryogenesis abundant.

The nucleotide sequence reported in this paper has been submitted to DDBJ under accession numbers AB049335, AB049336, AB049337and AB049338.

Environmental stresses such as drought, high salt and low temperature affect plant growth and decrease crop productivity extremely. It is important to improve stress tolerance of the crop plant to increase crop yield under stress conditions. Plants respond and adapt to the environmental stresses with physiological and developmental changes. Moreover, expression of a variety of genes has been shown to be induced by these stresses in a variety of plants (Ingram and Bartels 1996, Thomashow 1999, Shinozaki and Yamaguchi-Shinozaki 1997, Shinozaki and Yamaguchi-Shinozaki 2000). The dehydration-responsive element (DRE) with the core sequence A/GCCGAC has been identified as a *cis*-acting promoter element in regulating gene expression in response to drought, high-salt, and cold stresses in *Arabidopsis* (Yamaguchi-Shinozaki and Shinozaki 1994). A similar motif has been identified as CRT (C-repeat) and LTRE (low-temperature-responsive element) in cold-inducible genes (Baker et al. 1994, Jiang et al. 1996).

The transgenic *Arabidopsis* plants overexpressing *CBF1* (*DREB1B*) cDNA under control of the cauliflower mosaic virus (CaMV) 35S promoter showed strong tolerance to freezing stress (Jaglo-Ottosen et al. 1998). Overexpression of the *DREB1A* (*CBF3*) cDNA under the control of the CaMV 35S promoter resulted in strong expression of target stress-inducible genes and the transgenic plants acquired higher tolerance to drought, high-salt, and freezing stresses (Liu et al. 1998, Kasuga et al. 1999, Gilmour et al. 2000). Six genes have been identified as the target stress-inducible genes of DREB1A using RNA gel blot analysis. We also identified further target stress-inducible genes of DREB1A using a 1,300 full-length *Arabidopsis* cDNA microarray (Seki et al. 2001). Twelve stress-inducible genes were identified as target genes of DREB1A. These DREB1A target genes contained the DRE or DRE-related core motifs in their promoter regions. Recently, Fowler and Thomashow (2002) reported 41 target genes of CBFs/DREBs using Affymetrix GeneChip arrays. Thus overexpression of DREBs/CBFs resulted in the induction of multiple stress tolerance genes under a non-stressed control condition in transgenic *Arabidopsis*, and therefore, the transgenics had improved tolerance to drought, high-salt and freezing stresses.

Recently, Jaglo et al. (2001) showed that constitutive overexpression of the *Arabidopsis* CBF genes in transgenic *Brassica napus* plants induced expression of orthologs of *Arabidopsis* CBF-targeted genes and increased the freezing tolerance of plants. Hsieh et al. (Hsieh et al. 2002a, Hsieh et al. 2002b) also reported that tomato plants ectopically expressing *Arabidopsis* CBF1/DREB1B showed enhanced resistance to water deficit, chilling and oxidative stresses. These results indicate that the CBF/DREB gene can be used to improve the multi-stress tolerance of agriculturally important crops by gene transfer. However, transgenic *Arabidopsis* overexpressing DREB1A caused severe growth retardation under normal

[4] Corresponding author: E-mail, kazukoys@jircas.affrc.go.jp; Fax, +81-29-838-6643.

BNSDOCID: <XP_____2421281A_I_>

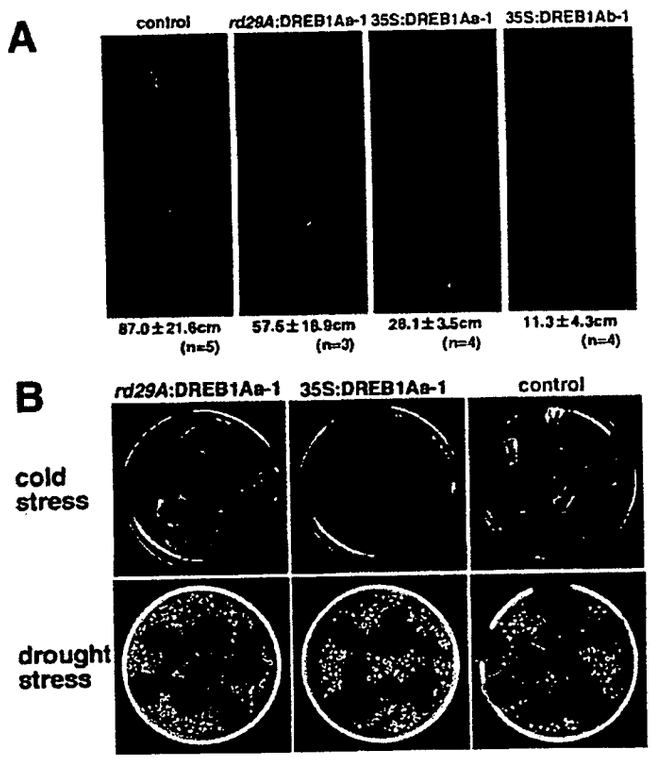


Fig. 1 (A) Phenotypes of the 35S:DREB1A and *rd29A*:DREB1A tobacco plants in relation to the control plant. (10 weeks) (B) Cold and drought tolerance of the 35S:DREB1A and *rd29A*:DREB1A tobacco plants. Cold stress: plants were exposed to a temperature of 1°C for 10 d and returned to 25°C for 2 d, drought stress: plants grown in pots were not watered for 2 weeks. To prepare transgenic tobacco (*Nicotiana tabacum*), we used the CaMV 35S constitutive promoter and the stress-inducible *Arabidopsis rd29A* promoter to overexpress the *Arabidopsis DREB1A* gene in tobacco (*Nicotiana tabacum*, cv. SR1). The 35S:DREB1A plasmid was constructed as described previously (Liu et al. 1998). The *rd29A*:DREB1A was constructed as described previously (Kasuga et al. 1999). The constructs were introduced into *Agrobacterium tumefaciens* 4404 as described previously (Yamaguchi-Shinozaki and Shinozaki 1994).

growth conditions. Transgenic tomato overexpressing CBF1 also showed the dwarf phenotype under unstressed normal growth conditions (Hsieh et al. 2002a, Hsieh et al. 2002b). Use of the stress-inducible *rd29A* promoter instead of the constitutive 35S CaMV promoter for the overexpression of DREB1A in transgenic *Arabidopsis* minimizes the negative effects on plant growth (Kasuga et al. 1999). In this study we analyzed the effect of overexpression of DREB1A on stress tolerance and growth retardation in transgenic tobacco plants using the constitutive CaMV35S promoter and the stress-inducible *rd29A* promoter.

First, we generated transgenic tobacco (*Nicotiana tabacum*) plants overexpressing the DREB1A cDNA driven by a modified 35S CaMV promoter (Mitsuhara et al. 1996) or the stress-inducible *rd29A* promoter (Kasuga et al. 1999). Twenty-nine antibiotic-resistant tobacco transformants carrying the 35S:DREB1A transgene (henceforth, 35S:DREB1A plants) and 42 antibiotic-resistant tobacco transformants carrying the *rd29A*:DREB1A transgene (henceforth, the *rd29A*:DREB1A plants) were generated using the leaf-disc method (Yamaguchi-Shinozaki and Shinozaki 1994). Expression levels of 14 and 30 lines of the 35S:DREB1A and *rd29A*:DREB1A plants, respectively, were examined by RNA gel blot analysis using the DREB1A cDNA as a probe. Most of the lines showed overexpression of the transgene (data not shown). We selected four and eight lines for the 35S:DREB1A and *rd29A*:DREB1A plants, which showed various levels of overexpression of each transgene, and used them for further analyses. Among four lines of 35S:DREB1A, two lines were T1 generation and the other two lines were T2 generation and all eight lines of *rd29A*:DREB1A were T1 generation. We also generated transgenic tobacco plants carrying pBI121 as control plants.

We observed the growth of the 35S:DREB1A and *rd29A*:DREB1A plants under normal growing conditions and compared them with the control plants at 10 weeks after sowing (Fig. 1A). All four 35S:DREB1A plants showed growth retardation and were classified into two groups including two plants for each group (Fig. 1; 35S:DREB1Aa and 35S:DREB1Ab). The 35S:DREB1Ab plants showed severer growth retardation than the 35S:DREB1Aa plants (Fig. 1A). By contrast, most of the *rd29A*:DREB1A plants exhibited slight growth retardation (Fig. 1A) while only two lines of *rd29A*:DREB1A showed growth retardation like 35S:DREB1Ab (data not shown).

To analyze the relationship between expression levels of the *DREB1A* gene and the growth of the transgenic plants, we carried out RNA gel blot hybridization. Expression of *DREB1A* is higher in the 35S:DREB1Ab plants than that in the 35S:DREB1Aa plants. Expression of the *DREB1A* gene in *rd29A*:

BNSDOCID: <XP_ .. ___2421281A__I_>

DREB1Aa tobacco is lower than that in both 35S:DREB1Aa and 35S:DREB1Ab under normal growth conditions (data not shown). The 35S:DREB1A plants revealed different levels of growth retardation under the same condition, which may be due to the different levels of expression of the *DREB1A* transgene as a result of the position effect.

We showed that overexpression of *DREB1A* driven by the 35S promoter also affected the growth of transgenic tobacco plants and their growth retardation was correlated to the level of the accumulated *DREB1A* mRNA in transgenic tobacco as well as in transgenic *Arabidopsis* under unstressed conditions (Fig. 1A). Such growth retardation was also reported using tomato overexpressing CBF1 (DREB1B) by Hsieh et al. (Hsieh et al. 2002a, Hsieh et al. 2002b). They also reported that these phenomena could be reversed by GA_3 treatment in the transgenic tomato plants and suggested that hyperaccumulation of the CBF1 protein in the transgenics may be interfering with GA biosynthesis. Therefore, we analyzed whether application of GA_3 could overcome the dwarf phenotype of the transgenic *Arabidopsis* and tobacco plants overexpressing *DREB1A* (data not shown). Leaves of transgenic *Arabidopsis* seemed to be enlarged by GA_3 treatment. However, this phenomenon was not specific to the transgenics. The control *Arabidopsis* plants were also enlarged by GA_3. On the other hand, transgenic tobacco did not show any reversal of growth retardation by GA_3 treatment. Thus, the dwarf phenotype or growth retardation of transgenic *Arabidopsis* and tobacco plants may be due not to the interference of GA biosynthesis but to some other mechanism, which is different from the mechanism reported in transgenic tomato plants (Hsieh et al. 2002a, Hsieh et al. 2002b).

The tolerance levels of the 35S:DREB1Aa and *rd29A*:DREB1Aa transgenic tobacco to freezing and drought stress were compared with those of the control plants grown in pots or on agar plates at 25°C for 3 weeks. When plants grown on agar were exposed to temperatures of 1°C for 10 d and then returned to 25°C for 2 d the control plants showed severe damage on their leaves. However, both 35S:DREB1Aa and *rd29A*:DREB1Aa transgenic plants revealed higher tolerance than the control plants (Fig. 1B). To test whether overexpression of the *DREB1A* gene enhances drought stress tolerance of the transgenic tobacco, we did not water the control and transgenic plants grown in pots for 2 weeks. Both 35S:DREB1Aa and *rd29A*:DREB1Aa tobacco plants showed higher stress tolerance to drought than the control plants (Fig. 1B; 35S:DREB1Aa-1 and *rd29A*:DREB1Aa-1).

To evaluate the effect of DREB1A overexpression on low-temperature stress tolerance, we conducted an electrolyte leakage test using the leaves of 35S:DREB1Aa, *rd29A*:DREB1Aa and control tobacco plants grown on agar plates at 1°C for 10 d. We collected the third and fourth leaves from the bottom for each transgenic plant and used them for the analysis. We observed the differences in growth between the control and transgenic plants but both plants were young enough and did not show bleaching of these leaves that occurs with aging.

Table 1 Drought and low-temperature stress tolerance of the transgenic tobacco plants

	Control	Cold	Drought
Ion leakage (%)			
35S:DREB1Aa	3.8 ±0.7	4.8 ±1.1*	5.6 ±2.2*
rd29A:DREB1Aa	5.2 ±2.1	9.1 ±1.7*	9.7 ±1.1*
Control plants	5.8 ±1.3	22.5 ±5.1	17.3 ±7.5
F_v/F_m			
35S:DREB1Aa	0.49±0.22	0.52±0.13*	0.61±0.10*
Control plants	0.35±0.27	0.20±0.19	0.05±0.02

Ion leakage and photosynthetic activity (F_v/F_m) were measured. The values are means ± SE of two to eight independent samples. Statistical significance compared with the value of the control plants was determined by *t*-test (*$p<0.01$). Membrane leakage and photosynthetic activity were measured according to the method described by Fan et al. (1997). Leaves of each control or transgenic plant were immersed in 10 ml of 0.4 M mannitol at room temperature with shaking at 200 rpm for 3 h, and the bathing solution was measured for conductivity with a conductivity meter (HORIBA DS-8F). The total conductivity was determined by boiling leaves for 10 min. The conductivity due to leakage was expressed as the percentage of the initial conductivity versus the total conductivity. Photosynthetic activities of leaves were assessed by use of the ratio of variable chlorophyll fluorescence (F_v) to maximum yield of chlorophyll fluorescence (F_m), expressed as F_v/F_m. The ratio F_v/F_m was monitored with a TEACHING PAM (HEINZ WALZ). Leaves were adapted to the dark for 10 min at room temperature before measurements were made.

Electrolyte leakage in both 35S:DREB1Aa and *rd29A*:DREB1Aa plants was reduced significantly compared with that in the control plants (Table 1). We also measured electrolyte leakage of the leaves of both control and transgenic plants grown in pots after drought stress treatments by stopping water for 2 weeks. Electrolyte leakage was clearly low in both 35S:DREB1Aa and *rd29A*:DREB1Aa plants as compared with that of the control plants (Table 1). Moreover, the 35S:DREB1Aa plants had higher photosynthetic activity than the control plants as shown by measuring chlorophyll fluorescence maximum photochemical efficiency of PSII in the dark-adapted state (Table 1: F_v/F_m). These results demonstrated that overexpression of DREB1A resulted in improved stress tolerance in tobacco plants.

Then, we analyzed the difference in *DREB1A* expression between the 35S:DREB1A and *rd29A*:DREB1A transgenic tobacco plants (Fig. 2A). The 35S:DREB1A, *rd29A*:DREB1A and control plants (pBI121) were grown on MS agar medium containing Kanamycin (200 mg liter^{-1}) for 3 weeks as described previously (Yamaguchi-Shinozaki and Shinozaki 1994) and exposed to drought- and low-temperature stress treatments. Isolation of total RNA and RNA gel blot hybridization were performed as described previously (Yamaguchi-Shinozaki and Shinozaki 1994). The *DREB1A* gene was slightly expressed under the control conditions in the *rd29A*:DREB1Aa tobacco plants, but after they were exposed to these

BNSDOCID: <XP_____2421281A_I_>

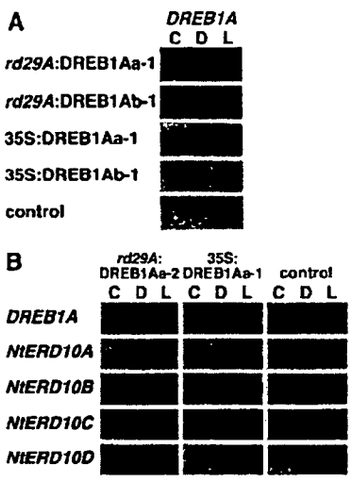


Fig. 2 RNA-gel blot analysis of the *DREB1A* and *NtERD10* genes. Each lane was loaded with 40 µg of total RNA prepared from the plants. C, 3-week-old plants growing under normal conditions; D, plants dehydrated in Petri dishes at room temperature for 5 h under dim light (100 lux); L, plants exposed to a temperature of 1°C for 10 d. (A) Expression of the *DREB1A* gene in the 35S:DREB1A and *rd29A*:DREB1A tobacco plants as compared with that in the control plants. (B) Expression of each *NtERD10* gene in the 35S:DREB1A and *rd29A*:DREB1A tobacco plants as compared with that in the control plants. The DNA fragments of the full-length *DREB1A*, *NtERD10A* and *NtERD10B* cDNAs and the 3'-terminal specific DNA fragment of the *NtERD10C* and *NtERD10D* cDNAs were used as probes.

stress conditions, the expression of the *DREB1A* gene was elevated to a much higher level. We detected stronger DREB1A expression in the *rd29A*:DREB1Ab plants under the control condition than in the *rd29A*:DREB1Aa tobacco plants, and the level of DREB1A expression was further increased by drought and low-temperature stresses in both transgenic plants. On the other hand, we detected almost constitutive expression of *DREB1A* in the 35S:DREB1Aa and 35S:DREB1Ab plants. However, we sometimes detected stronger expression of DREB1A in the 35S:DREB1A plants exposed to low temperature. A certain mechanism to stabilize the *DREB1A* mRNA may exist under low-temperature stress conditions in the transgenic tobacco plants (Fig. 2A).

Previously, we reported that DREB1A functions as a transcriptional activator in the expression of more than 40 stress-inducible genes in *Arabidopsis* (Kasuga et al. 1999, Seki et al. 2001). Some of the DREB1A target genes encode group 2 LEA proteins such as ERD10 and RD17. To identify the target genes of DREB1A in tobacco we constructed a cDNA library using mRNAs prepared from the 35S:DREB1A tobacco plants. To prepare cDNA libraries 20 g of whole transgenic tobacco plants grown on MS agar medium for 3 weeks were used. Total RNA, poly(A)+ RNA and cDNA were prepared as described previously (Yamaguchi-Shinozaki et al. 1992). We screened the library for cDNAs encoding group 2 LEA proteins from the library. The probes for the group 2 LEA-protein were amplified by PCR with the primers, 5'-GGATCCGGATCCAAACCAGGGGGTGATGAT-3' and 5'- GGATCCGGATCCTTAATCAGACACTTTTTC-3', using the *ERD10* cDNA as a template. We isolated four cDNAs for the group 2 LEA proteins of tobacco, and named them *NtERD10A*, *NtERD10B*, *NtERD10C* and *NtERD10D*, whose accession numbers are AB049335, AB049336, AB049337, AB049338, respectively.

We examined the expression of these four *NtERD10* genes in the control and transgenic tobacco plants (Fig. 2B). In the 35S:DREB1Aa tobacco plants, all the *NtERD10* genes except *NtERD10A* were expressed strongly even under normal growth conditions as compared with that in the control plants and the expression levels of these genes were significantly increased by drought and low-temperature stresses. In the *rd29A*:DREB1Aa tobacco plants, all the *NtERD10* genes were weakly expressed under the normal growth conditions like that in the control plants and strongly induced by drought and cold stresses. The expression level of these induced genes in the transgenics was higher than that in the control plants. These results indicate that these four *NtERD10* genes are target stress-inducible genes of DREB1A in tobacco and are probably responsible for the stress tolerance of the transgenic tobacco like *Arabidopsis* homologs. However, the degree of gene activation seemed to be limited in the 35S:DREB1A tobacco under the control condition (Fig. 2B). The *DREB1A* gene of *Arabidopsis* may not have enough activity to overexpress the target genes in tobacco plants or we could not detect the highest expression of the *NtERD10* genes because we used the transgenic line with a mild phenotype (35S:DREB1Aa: Fig. 1A).

Among the four *NtERD10* genes, the *NtERD10A* gene showed different expression from the other three genes (Fig. 2B). This gene is induced by drought but not by a low temperature while the other genes are induced by both drought and a low temperature like the DREB1A target genes in *Arabidopsis*. However, we detected overexpression of *NtERD10A* in both the *rd29A*:DREB1Ab-1 and 35S:DREB1Aa-1 plants under a low temperature, which indicated that overexpression of DREB1A affected the expression of *NtERD10A* (Fig. 2B). Therefore, we think that *NtERD10A* is one of the target genes of DREB1A in tobacco. However, this gene may not be the target gene of the endogenous DREB1A homolog in tobacco, or this stress-inducible gene may be suppressed specifically under a low temperature.

Use of the stress-inducible *rd29A* promoter instead of the constitutive 35S CaMV promoter for the overexpression of DREB1A minimized the negative effects on plant growth in transgenic *Arabidopsis* (Kasuga et al. 1999). We further showed that DRE in the *rd29A* promoter also functions in gene expression in response to stress in tobacco plants (Yamaguchi-Shinozaki and Shinozaki 1993), which suggests that the *rd29A* promoter regulates the expression of *DREB* in tobacco similarly does in *Arabidopsis*. These results indicate that the stress-

BNSDOCID: <XP_____2421281A_1_>

inducible *rd29A* promoter is quite useful to overexpress DREB1A for improving drought, salt and freezing stress tolerance not only in transgenic *Arabidopsis* but also in other kinds of transgenic plants such as tobacco.

Acknowledgments

The authors thank Ms. Kyoko Murai, Ms. Ekuko Ohgawara, Ms. Mie Yamamoto and Ms. Fumie Saito for their excellent technical assistance. This work was supported in part by the Program for Promotion of Basic Research Activities for Innovative Biosciences, which awarded a grant to K. Y.-S. This work was also supported in part by the Special Coordination Fund of the Science and Technology Agency, by the Human Frontier Science Program, and by a grant in aid from the Ministry of Education, Culture, Science, and Technology to K. S.

References

Baker, S.S., Wilhelm, K.S. and Thomashow, M.F. (1994) The 5'-region of *Arabidopsis thaliana cor15a* has *cis*-acting elements that confer cold-, drought- and ABA-regulated gene expression. *Plant Mol. Biol.* 24: 701–713.

Fan, L., Zheng, S. and Wang, X. (1997) Antisense suppression of phospholipase D retards abscisic acid- and ethylene-promoted senescence of postharvest Arabidopsis leaves. *Plant Cell* 9: 2183–2196.

Fowler, S. and Thomashow, M.F. (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* 14: 1675–1690.

Gilmour, S.J., Sebolt, A.M., Salazar, M.P., Everard, J.D. and Thomashow, M.F. (2000) Overexpression of the *Arabidopsis* CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant Physiol.* 124: 1854–1865.

Hsieh, T.H., Lee, J.T., Charng, Y.Y. and Chan, M.T. (2002a) Tomato plants ectopically expressing Arabidopsis CBF1 show enhanced resistance to water deficit stress. *Plant Physiol.* 130: 618–626.

Hsieh, T.H., Lee, J.T., Yang, P.T., Chiu, L.H., Charng, Y.Y., Wang, Y.C. and Chan, M.T. (2002b) Heterology expression of the Arabidopsis C-repeat/dehydration response element binding factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. *Plant Physiol.* 129: 1086–1094.

Ingram, J. and Bartels, D. (1996) The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47: 377–403.

Jaglo, K.R., Kleff, S., Amundsen, K.L., Zhang, X., Haake, V., Zhang, J.Z., Deits, T. and Thomashow, M.F. (2001) Components of the Arabidopsis C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiol.* 127: 910–917.

Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O. and Thomashow, M.F. (1998) *Arabidopsis* CBF1 overexpression induces cor genes and enhances freezing tolerance. *Science* 280: 104–106.

Jiang, C., Iu, B. and Singh, J. (1996) Requirement of a CCGAC *cis*-acting element for cold induction of the *BN115* gene from winter *Brassica napus*. *Plant Mol. Biol.* 30: 679–684.

Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1999) Improving plant drought, salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat. Biotechnol.* 17: 287–291.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain, separate two cellular signal transduction pathways in drought- and low temperature-responsive gene expression, respectively, in Arabidopsis. *Plant Cell* 10: 1391–1406.

Mitsuhara, I., Ugaki, M., Hirochika, H., Ohshima, M., Murakami, T., Gotoh, Y., Katayose, Y., Nakamura, S., Ryoso, H., Nishimiya, S., Ueno, K., Mochizuki, A., Tanimoto, H., Tsugawa, H., Otsuki, Y. and Ohashi, Y. (1996) Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. *Plant Cell Physiol.* 37: 49–59.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y. and Shinozaki, K. (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* 13: 61–72.

Shinozaki, K. and Yamaguchi-Shinozaki, K. (1997) Gene expression and signal transduction in water stress response. *Plant Physiol.* 115: 327–334.

Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Curr. Opin. Plant Biol.* 3: 217–223.

Thomashow, M.F. (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50: 571–599.

Yamaguchi-Shinozaki, K., Koizumi, M., Urao, S. and Shinozaki, K. (1992) Molecular cloning and characterization of 9 cDNAs for genes that are responsive to desiccation in *Arabidopsis thaliana*: Sequence analysis of one cDNA clone that encodes a putative transmembrane channel protein. *Plant Cell Physiol.* 33: 217–224.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (1993) The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of *rd22*, a gene responsive to dehydration stress in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 238: 17–25.

Yamaguchi-Shinozaki, K. and Shinozaki, K. (1994) A novel *cis*-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell* 6: 251–264.

(Received October 8, 2003; Accepted December 24, 2003)

BNSDOCID: <XP_____ 2421281A__I_>

XP-002431282

(2)

Control of Xylose Consumption by Xylose Transport in Recombinant *Saccharomyces cerevisiae*

Márk Gárdonyi,[1] Marie Jeppsson,[1] Gunnar Lidén,[2] Marie F. Gorwa-Grauslund,[1] Bärbel Hahn-Hägerdal[1]

[1]Department of Applied Microbiology, Lund University, P.O. Box 124, 22100 Lund, Sweden; telephone: + 46 46 222 8428; fax: + 46 46 222 4203; e-mail: Barbel.Hahn-Hagerdal@tmb.lth.se
[2]Department of Chemical Engineering II, Lund University, P.O. Box 124, 22100 Lund, Sweden

Received 18 June 2002; accepted 20 November 2002

DOI: 10.1002/bit.10631

Abstract: *Saccharomyces cerevisiae* TMB3001 has previously been engineered to utilize xylose by integrating the genes coding for xylose reductase (XR) and xylitol dehydrogenase (XDH) and overexpressing the native xylulokinase (XK) gene. The resulting strain is able to metabolize xylose, but its xylose utilization rate is low compared to that of natural xylose utilizing yeasts, like *Pichia stipitis* or *Candida shehatae*. One difference between *S. cerevisiae* and the latter species is that these possess specific xylose transporters, while *S. cerevisiae* takes up xylose via the high-affinity hexose transporters. For this reason, in part, it has been suggested that xylose transport in *S. cerevisiae* may limit the xylose utilization.

We investigated the control exercised by the transport over the specific xylose utilization rate in two recombinant *S. cerevisiae* strains, one with low XR activity, TMB3001, and one with high XR activity, TMB3260. The strains were grown in aerobic sugar-limited chemostat and the specific xylose uptake rate was modulated by changing the xylose concentration in the feed, which allowed determination of the flux response coefficients. Separate measurements of xylose transport kinetics allowed determination of the elasticity coefficients of transport with respect to extracellular xylose concentration. The flux control coefficient, C^J_{transp} for the xylose transport was calculated from the response and elasticity coefficients. The value of C^J_{transp} for both strains was found to be < 0.1 at extracellular xylose concentrations > 7.5 g L^{-1}. However, for strain TMB3260 the flux control coefficient was higher than 0.5 at xylose concentrations < 0.6 g L^{-1}, while C^J_{transp} stayed below 0.2 for strain TMB3001 irrespective of xylose concentration. © 2003 Wiley Periodicals, Inc. *Biotechnol Bioeng* 82: 818–824, 2003.

Keywords: xylose transport; xylose reductase; metabolic control analysis; *Saccharomyces cerevisiae*

Correspondence to: Bärbel Hahn-Hägerdal

Contract grant sponsor: The Swedish Energy Administration (Energimyndigheten); EU Commission
Contract grant number: QLK3-CT-1999-00080

© 2003 Wiley Periodicals, Inc.

INTRODUCTION

In industrial ethanol production *Saccharomyces cerevisiae* is the choice organism due to its high productivity and ethanol tolerance. It also tolerates fermentation inhibitors present in a lignocellulose hydrolysate better than other organisms (Lindén and Hahn-Hägerdal, 1989). However, *S. cerevisiae* is unable to utilize pentose sugars. Lignocellulose contains 5–35% xylan (Hayn et al., 1993; Hespell, 1998; Sjöström, 1993; Stenberg et al., 1998; Vinzant et al., 1994), which upon hydrolysis yields xylose. Hence, fuel ethanol production from pentose-rich sources, such as hardwood or agricultural residues will only be economically viable if xylose is fermented to ethanol (Hinman et al., 1989; Von Sivers and Zacchi, 1996). The genes necessary for xylose metabolism have been cloned and expressed in *S. cerevisiae* (Hallborn, et al., 1991; Ho and Chang, 1989; Kötter et al., 1990; Tantirungki, et al., 1993). The resulting strains were capable of utilizing xylose, but they compared poorly with natural xylose utilizing yeasts, like *Pichia stipitis*, in terms of growth rate and ethanol productivity (Kötter and Ciriacy, 1993; Wahlbom, 2002).

Saccharomyces cerevisiae takes up xylose from the media with very low affinity (K_M = 100–190 m*M*) via the high-affinity hexose transporters (Busturia and Lagunas, 1986; Hamacher et al., 2002; Heredia et al., 1968; Kotyk, 1967; Kötter and Ciriacy, 1993). It has been suggested that the transport does not limit the xylose utilization rate of recombinant xylose utilizing *S. cerevisiae* strains at xylose concentrations around 20 g L^{-1} (133 m*M*), since the initial uptake rate is 30-fold higher than the actual xylose utilization rate at this concentration (Kötter and Ciriacy, 1993). At lower concentrations, on the other hand, transport has been speculated to be a problem due to the fact that the enzymes in the xylose metabolic pathway have poor affinities to their substrates. Transport was suggested to control the xylose utilization rate in *S. cerevisiae* TMB3001, because the specific uptake rate increased with increasing residual xylose

BNSDOCID: <XP_____2421282A__I_>

concentration between 3 g L^{-1} and 13 g L^{-1} (20–85 m*M*) in anaerobic chemostat cultivations (Eliasson et al., 2000).

Transport is often described as either a "limiting" or "nonlimiting" step for utilization of xylose. However, there is no single rate-limiting step in biochemical systems, instead each step exercises a fraction of the total control over the metabolic rate (Fell, 1997; Kacser and Burns, 1973). Metabolic control analysis (MCA) has been used before to quantify the degree of control exercised by transport on the glycolytic flux in *S. cerevisiae*. Flux control coefficients between 0.6 and 1 have been obtained from kinetic models based on in vitro enzyme kinetics and intracellular metabolite concentrations (Cortassa and Aon, 1994; Galazzo and Bailey, 1990). Direct determination of the flux control coefficient in strains with different high-affinity glucose transporter expression yielded 0.9 (Ye, 1999). In the closely related *Saccharomyces bayanus* a flux control coefficient between 0.5 and 1.2 has been found by varying the extracellular glucose concentration (Diderich et al., 1999). Here we report for the first time the quantitative determination of the flux control exercised by transport on the specific xylose utilization rate in recombinant *S. cerevisiae*.

We determined the degree of control exercised by the xylose transport over the xylose utilization rate in aerobic chemostat cultures of a recombinant *S. cerevisiae* strain, TMB3001 (Eliasson et al., 2000). In addition, a strain with increased XR activity, TMB3260 (Johansson, 2001), was included in this study to determine whether XR overproduction improves aerobic xylose utilization and whether this influences the control coefficient of the transport step. We modulated the xylose transport rate by changing the xylose concentration in the feed. The flux control coefficients were calculated from the flux response to changes in extracellular xylose concentration and the elasticity of transport towards the extracellular xylose concentration. Maltose was used as cosubstrate, because TMB3001 grows poorly on xylose as sole carbon source.

MATERIALS AND METHODS

Strains and Cultivation Conditions

TMB3001 (CEN.PK 113-7A *his3*::YIpXR/XDH/XK) (Eliasson et al., 2000) and TMB3260 (TMB3001 *XYL1*$^+$) (Johansson, 2001) were cultivated in defined mineral medium (Verduyn et al., 1992) with xylose (1, 2, 5, 10, or 20 g L^{-1}) and maltose (2, 5 or 10 g L^{-1}) as carbon sources at 30°C. Dow Corning® Antifoam RD Emulsion (BDH Laboratory Supplies, Poole, UK), 0.05% (v/v) was added to prevent foaming. Sugar-limited chemostat cultivation was performed in Bioflo III fermentors (New Brunswick Scientific, Edison, NJ) with a working volume of 1.5 L and a dilution rate of 0.125 h^{-1}. Steady state was determined from on-line gas analysis. Sampling was done at least 5 residence times after the gas composition had stabilized. The fermentor was aerated with a flow rate of 2 L/min^{-1} and stirred at 500 rpm. A pH value of 5.5 was maintained by automatic addition of NaOH.

Extraction of Intracellular Sugars

Intracellular xylose and maltose were extracted using a slightly modified hot ethanol method (Gonzalez et al., 1997). The metabolism of 5 mL cells was quenched in 26 ml 70 m*M* Hepes in 60% methanol (v/v), pH 7.5, kept at −40°C in a dry ice/ethanol bath (de Koning and van Dam, 1992). The cell suspension was centrifuged at 5000*g* and −25°C for 5 min. The pellet was washed twice with 26 mL of the Hepes-methanol solution to remove extracellular xylose. Metabolites were extracted by dissolving the pellet in 5 mL of 75% (v/v) boiling ethanol with 0.25*M* Hepes, pH 7.5, followed by incubation for 3 min at 80°C, instantaneous freezing and storage at −80°C. The ethanol was removed in a DNA SpeedVac (Savant Instruments, Inc., Farmingdale, NY) during 1 h at medium drying rate (≈ 43°C). The samples were centrifuged at 14000 rpm for 1 min, and the supernatants were analyzed for xylose and maltose using column liquid chromatography (CLC). The intracellular concentrations were calculated on the assumption that cells contain 1.5 μL water (mg dry weight)$^{-1}$ (de Koning and van Dam, 1992).

Analysis

Xylose and maltose concentrations were determined using a Gilson CLC system (Gilson, Villiers-le-bel, France). An Aminex HPX-87H column (BioRad, Richmond, CA) and an R1D-10A refractive index detector (Shimadzu, Kyoto, Japan) were used. The column temperature was 45°C and 5 m*M* H_2SO_4 were used as mobile phase at a flow rate of 0.6 mL min^{-1}. The detection limit was 0.4 m*M* for xylose and 0.1 mM for maltose.

Zero-Trans-Influx Assay

Sugar uptake assays were performed at 30°C, pH 5.0 as previously described (Spencer-Martins and van Uden, 1985). D-[^{14}C]-xylose from Amersham International (Buckinghamshire, UK) was used at 50 cpm nmol^{-1}specific activity between 12.5–250 m*M* final concentrations. All measurements were done in duplicate. Kinetic constants were estimated by nonlinear regression analysis using commercial software (SigmaPlot, SPSS Inc., San Rafael, CA).

Calculations

Flux response coefficients can, in principle, be calculated by approximating the derivative

$$\frac{dJ}{dX} \cdot \frac{X}{J} \text{ by } \frac{\Delta J}{\Delta X} \cdot \frac{X}{J},$$

where J is the metabolic flux and X is the concentration of

BNSDOCID: <XP_____2421282A_I_>

the modulated feed component. However, it has been shown that this approximation gives large inaccuracies (Fell, 1997). Therefore, a hyperbolic equation (i.e., a Monod-type equation) was used to describe the relation between metabolic flux and extracellular xylose concentration:

$$J = q_X = \frac{q_{X,max} \cdot X_{ec}}{K_S + X_{ec}}, \quad (1)$$

where q_X is the specific steady-state xylose uptake rate, X_{ec} is the steady-state extracellular xylose concentration and K_S is the half-saturation constant. Flux response coefficients were then calculated from Equation (2):

$$R^J_{X_{ec}} = \frac{K_S}{K_S + X_{ec}} \quad (2)$$

The specific xylose transport rate was expressed according to the symmetrical four-state carrier model (Teusink et al., 1998):

$$v_{transp} = \frac{v_{transp,max}(X_{ec} - X_{ic})}{K_M + X_{ec} + X_{ic} + \alpha \frac{X_{ec}X_{ic}}{K_M}}, \quad (3)$$

where X_{ec} and X_{ic} are the extra- and intracellular xylose concentrations, K_M is the apparent Michaelis-Menten constant of xylose transport and α is a term accounting for the different mobilities of the free and loaded carriers. Model parameters values $v_{transp,max}$ and K_M were estimated by nonlinear parameter estimation from zero-trans-influx measurements. A value of 0.83 for α was calculated as described earlier (Teusink et al., 1998) using the previously determined 2.41-fold mobility change upon xylose binding (Kotyk, 1967).

Steady-state intracellular xylose concentrations were calculated using the steady-state requirement that the specific xylose transport rate must equal the metabolic flux ($v_{transp} = q_s$) and solving Equation (3) for X_{ic}. Elasticity coefficients were calculated from:

$$\varepsilon^{transp}_{X_{ec}} = \frac{\partial v_{transp}}{\partial X_{ec}} \cdot \frac{X_{ec}}{v_{transp}} \quad (4)$$

and flux control coefficients as:

$$C^J_{transp} = \frac{R^J_{X_{ec}}}{\varepsilon^{transp}_{X_{ec}}}. \quad (5)$$

Using the fitted parameter values it was possible to obtain analytical expressions for both the elasticity coefficient and the flux control coefficient.

Enzyme Activities

Two milliliters of chemostat culture were harvested by centrifugation (13.000 rpm, 1 min) and the pellets were stored at − 20°C. Proteins were extracted with the Y-PER reagent (Pierce, Rockford, IL) according to the manufacturer's recommendations. Protein concentrations were determined by the micro BCA method (Pierce, Rockford IL). Activities of XR and XDH were measured on COBAS-MIRA automatic analyzer (Roche, Mannheim, Germany) as described earlier (Eliasson et al., 2000).

RESULTS

Two recombinant xylose-utilizing *S. cerevisiae* strains, TMB3001 and TMB3260 with an extra copy of the *XYL1* gene behind the *PGK1* promoter, were investigated in the present study. TMB3260 had a 20-fold higher XR activity than TMB3001, 11.18 U mg^{-1} compared with 0.66 U mg^{-1}. These values are higher than those measured earlier under aerobic or anaerobic condition, 3.36 U mg^{-1} for TMB3260 (Johansson, 2001) and 0.42–0.52 U mg^{-1} for TMB3001 (Eliasson et al., 2000; Johansson, 2001). The XDH activity was, 5.12 U mg^{-1} for TMB3260 and 2.44 U mg^{-1} for TMB3001. For TMB3001 specific XDH activity between 2.74 and 3.37 U mg^{-1} has been reported previously (Eliasson et al., 2000).

Specific Xylose Utilization Rates

TMB3001 and TMB3260 were cultivated in carbon-limited chemostat cultures. Steady states were established in chemostat cultures with 1, 2, 5, 10, or 20 g L^{-1} (6.7, 13.3, 33.3, 66.7, or 133.3 m*M*) xylose media. Maltose, 5 g L^{-1} (13.9 m*M*) was included in the feed to support growth. Samples were taken at steady state for analysis of residual sugars and determination of biomass. TMB3260 consumed about half of the xylose in the feed, while TMB3001 only utilized 15%. The specific xylose consumption rate varied between 22.5 and 114 mg g^{-1} h^{-1} for TMB3260 and between 7.5 and 72.5 mg g^{-1} h^{-1} for TMB3001. The specific xylose utilization rate as a function of residual xylose concentration could be well described by Monod-type kinetics (Fig. 1A). The half saturation constant, K_S, and the maximal specific xylose utilization rate, $q_{X,max}$, were determined by nonlinear regression using Equation (1). Clearly, the xylose utilization pathway of TMB3260 was saturated at lower extracellular xylose concentrations than TMB3001, as indicated by the lower K_S value (Table I). The calculated maximum specific xylose utilization rate was almost 40% higher for TMB3260 than TMB3001 (Table I). Chemostat cultivation with lower (2 g L^{-1}; 5.6 m*M*) and higher (10 g L^{-1}; 27.8 m*M*) maltose concentrations and 2 g L^{-1} (13.3 mM) xylose were also investigated to determine whether maltose influenced the specific xylose utilization rate. Low and high maltose concentration in the media had no significant effect on the specific xylose consumption rate (not shown).

Kinetics of Xylose Transport

Cells were harvested from each steady state to determine zero-trans-influx kinetics. The parameter values did not vary significantly for different residual xylose concentrations. Therefore, average values for each strain were calcu-

BNSDOCID: <XP._____ _2421282A_I_>

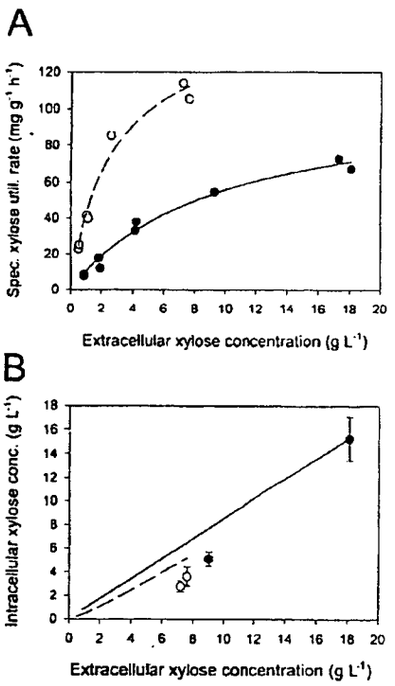


Figure 1. Specific xylose uptake rate (A) and intracellular xylose concentrations (B) at different extracellular xylose concentrations. Lines represent calculated values; circles are experimental results. Each circle corresponds to one steady-state culture and error bars show the standard deviation of two separate extractions. TMB3001 (continuous line, ●), TMB3260 (dashed line, ○).

lated (Table II). Furthermore, there was no significant difference between the kinetics of xylose transport in the two strains and the K_M values are in agreement with those published earlier (Busturia and Lagunas, 1986; Heredia et al., 1968; Kötter and Ciriacy, 1993; Kotyk, 1967) Maltose only weakly inhibited the initial xylose transport, K_i = 68.3 ± 29.0 m*M*, causing less than 1% reduction in xylose uptake rate at the maltose concentrations present in the fermentor, < 0.18 g L^{-1} (0.5 m*M*).

Intracellular Xylose Concentrations

We determined the intracellular xylose concentration, because it has been suggested that the intracellular glucose concentration may have a large impact on the rate of glucose transport in derepressed *S. cerevisiae* cells (Teusink et al., 1998). The intracellular xylose concentration was estimated from kinetic data and steady-state uptake rates and compared to measured values (Fig. 1B). Samples were taken from chemostat cultures with 10 or 20 g L^{-1} (66.7 or 133.3 m*M*) xylose and rapidly quenched at − 40°C. Intracellular metabolites were extracted and the xylose concentration was measured by column liquid chromatography. Measured intracellular xylose concentrations were in the same range as the values calculated from the kinetic expression although generally higher. Comparing extracts of washed and unwashed cells, we observed about 20% loss of intracellular metabolites other than xylose probably as a result of losing cells in the extensive washing procedure. This could possibly explain the deviation between measured and calculated values.

Table I. Kinetic constants for specific xylose uptake rate. K_x half saturation of the xylose uptake, q_{max} maximal specific xylose uptake rate.

Strain	K_X (m*M*)	q_{max} (mmol g^{-1} h^{-1})
TMB3001	62.6 ± 11.5	0.72 ± 0.06
TMB3260	17.2 ± 3.3	1.00 ± 0.07

Table II. Kinetic constants for initial xylose transport rate. K_M Michaelis constant for xylose transport, v_{max} maximal initial xylose transport rate.

Strain	K_M (m*M*)	v_{max} (mmol g^{-1} h^{-1})
TMB3001	143.4 ± 16.7	10.9 ± 3.0
TMB3260	129.7 ± 25.4	9.7 ± 2.2

MCA Coefficients

The flux control coefficient for transport (C^J_{transp}) was determined by dividing the flux response coefficient with the elasticity of the transport with respect to extracellular xylose. Flux response coefficients ($R^J_{X_e}$) were calculated according to Equation (2) for both strains at different xylose concentrations. The lower K_S of TMB3260 resulted in a faster declining flux response with increasing xylose levels in the fermentor (Fig. 2A). Elasticity coefficients ($\varepsilon^{transp}_{X_e}$) derived from the calculated intracellular xylose concentrations were lower for TMB3260 than for TMB3001 (Fig. 2B) as a result of the lower intracellular xylose levels. In strain TMB3260 the high response and low elasticity coefficients at low extracellular xylose concentration (<0.6 g L^{-1}: 4 m*M*) resulted in more than 50% control over the specific xylose uptake rate by the transport (Fig. 2C). However, with increasing extracellular xylose concentration the control diminished quickly due to the sharp decline of $R^J_{X_e}$ (Fig. 2A). In contrast, xylose transport in TMB3001 had little control over the specific uptake rate even at low xylose concentrations (Fig. 2C).

DISCUSSION

We investigated the control exercised by the transport step on the specific xylose utilization rate in two recombinant xylose utilising *S. cerevisiae* strains. We found that transport has very little control over the specific xylose utilization rate (C^J_{transp}< 0.15) in strain TMB3001. We also found

BNSDOCID: <XP_____2421282A__I_>

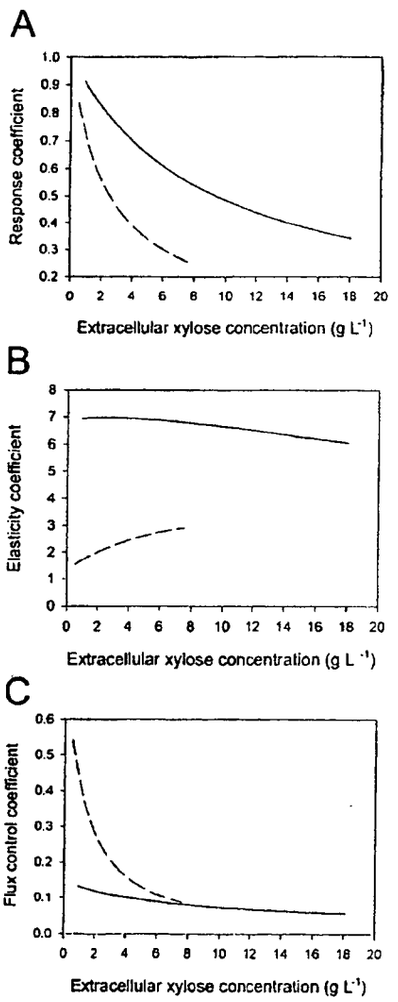


Figure 2. Calculated MCA coefficients shown as function of the extracellular xylose concentration. (A) Flux response coefficient. (B) Elasticity coefficient. (C) Flux control coefficient. TMB3001 (continuous line). TMB3260 (dashed line).

that in TMB3260, derived from TMB3001 by overproducing XR, transport has a flux control coefficient higher than 0.5 when the extracellular xylose concentration is below 0.6 g L^{-1} (4 m*M*). At higher residual xylose concentrations C^{J}_{transp} was reduced and it fell below 0.1 at 7.5 g L^{-1} (50 m*M*) xylose.

The specific xylose utilization rate of TMB3001 in aerobic sugar-limited chemostat was strongly dependent on the extracellular xylose concentration, as has been observed earlier in anaerobic sugar-limited chemostat cultures (Eliasson et al., 2000; Wahlbom et al., 2001). Based on the sensitivity of the xylose flux to changes in the extracellular xylose concentration, i.e., the large $R^{J}_{X_{ec}}$, it has been speculated that transport might have large control over the xylose utilization rate (Eliasson et al., 2000; Wahlbom, 2002). However, a large value of $R^{J}_{X_{ec}}$ does not necessarily mean a large flux control for the transport step, since the response coefficient is the product of elasticity and flux control coefficients. If the elasticity coefficient of the transport step is very large, then the flux control coefficient may have a low value. Indeed, we found that $\varepsilon^{transp}_{X_{ec}}$ was very high, and therefore xylose transport, in fact, had little control over the xylose utilization rate in TMB3001.

The high elasticity of the transport was the result of the high intracellular xylose concentration. It has been shown earlier that sugar uptake by facilitated diffusion is prone to inhibition by intracellular sugar (Teusink et al., 1998). This inhibition was found to be significant already at relatively low intracellular concentrations of glucose due to the requirement that the carrier be free when moving outwards. Steady-state intracellular xylose concentration and consequently $\varepsilon^{transp}_{X_{ec}}$ should be lower in a strain with an improved xylose metabolism pathway. TMB3260, with increased XR-activity, has been shown to utilize xylose 70% faster in oxygen-limited batch fermentation than TMB3001 (Johansson, 2001). In the aerobic chemostat cultures of the present study the specific xylose utilization rate was more than doubled in TMB3260 compared to TMB3001 at the same extracellular xylose concentration. The higher XR activity in TMB3260 also resulted in lower intracellular xylose concentration. Consequently, the elasticity was lower than in TMB3001 and the flux control coefficient was high at low residual xylose concentrations. The sharp decline of the flux control with increasing extracellular xylose concentration was primarily due to reduction of the flux response coefficient.

It has been suggested earlier for another recombinant *S. cerevisiae* strain that transport does not control the xylose utilization rate at 20 g L^{-1} (133.3 m*M*) extracellular xylose, because the zero-trans-influx of xylose was 30-fold higher than the steady-state xylose uptake rate (Kötter and Ciriacy, 1993). Although, the difference between initial xylose uptake rate and steady-state xylose uptake rate cannot be taken as a proof for low control by transport, our results agree with the previous suggestion. It has also been suggested that transport might be critical at lower extracellular xylose concentration, because enzymes in the xylose utilization pathway, XR and XDH, have poor affinity for their substrates (Kötter and Ciriacy, 1993). This hypothesis was partially confirmed by our results. In TMB3260, but not in TMB3001, transport had large control at extracellular xylose concentrations below 1 g L^{-1} (6.7 m*M*). It is conceivable that the elasticity of XR towards intracellular xylose would greatly increase when its concentration is low. According to the connectivity theorem

$$\frac{C^{J}_{transp}}{C^{J}_{XR}} = -\frac{\varepsilon^{XR}_{X_{ic}}}{\varepsilon^{transp}_{X_{ic}}}$$

BNSDOCID: <XP_____2421282A_I_>

(Kacser and Burns, 1973), consequently, an increase in $\varepsilon_{X_5}^{XR}$ or a decrease in $\varepsilon_{X_5}^{transp}$ results in a shift of control from XR to the transport step.

Our results show that xylose transport in *S. cerevisiae* strains has limited control over the specific xylose utilization rate, unless the xylose metabolism pathway is dramatically improved. Overexpression of *Hxt4*, *Hxt7*, or *Gal2* did not improve the xylose-utilization of TMB3001 and other recombinant *S. cerevisiae* strains (Hamacher et al., 2002). However, it would be beneficial to express an active xylose transporter, which is capable of taking up xylose in the cell against a concentration gradient, for utilization of xylose below 1 g L^{-1} (6.7 m*M*). Such transporters are common among the natural xylose utilizing yeasts (Alcorn and Griffin, 1978; Does and Bisson, 1989; Kilian and van Uden, 1988; Kilian et al., 1993; Lucas and van Uden, 1986; Nobre et al., 1999; Weierstall et al., 1999). In addition, it would be desirable to express a transporter, which is specific for xylose and not inhibited by glucose. This would enable cofermentation of xylose and glucose.

MFGG was supported by EU Marie Curie fellowship.

References

Alcorn ME, Griffin CC. 1978. A kinetic analysis of D-xylose transport in *Rhodotorula glutinis*. Biochim Biophys Acta 510:361–371.

Busturia A, Lagunas R. 1986. Catabolite inactivation of the glucose transport system in *Saccharomyces cerevisiae*. J Gen Microbiol 132:379–385.

Cortassa S, Aon MA. 1994. Metabolic control analysis of glycolysis and branching to ethanol production in chemostat cultures of *Saccharomyces cerevisiae* under carbon, nitrogen or phosphate limitations. Enzyme Microb Technol 16:761–770.

de Koning W, van Dam K. 1992. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Anal Biochem 204:118–123.

Diderich JA, Teusink B, Valkier J, Anjos J, Spencer-Martins I, van Dam K, Walsh MC. 1999. Strategies to determine the extent of control exerted by glucose transport on glycolytic flux in the yeast *Saccharomyces bayanus*. Microbiology 145:3447–3454.

Does AL, Bisson LF. 1989. Characterisation of xylose uptake in the yeasts *Pichia heedii* and *Pichia stipitis*. Appl Environ Microbiol 55: 159–164.

Eliasson A, Christensson C, Wahlbom CF, Hahn-Hägerdal B. 2000. Anaerobic xylose fermentation by recombinant *Saccharomyces cerevisiae* carrying *XYL1*, *XYL2*, and *XKS1* in mineral medium chemostat cultures. Appl Environ Microbiol 66:3381–3386.

Fell D. 1997. Understanding the control of metabolism. London, UK: Portland Press Ltd.

Galazzo JL, Bailey JE. 1990. Fermentation pathway kinetics and metabolic flux control in suspended and immobilized *Saccharomyces cerevisiae*. Enzyme Microb Technol 12:162–172.

Gonzalez B, Francois J, Renaud M. 1997. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. Yeast 13: 1347–1355.

Hallborn J, Walfridsson M, Airaksinen U, Ojamo H, Hahn-Hägerdal B, Penttilä M, Keränen S. 1991. Xylitol production by recombinant *Saccharomyces cerevisiae*. Biotechnology 9:1090–1095.

Hamacher T, Becker J, Gárdonyi M, Hahn-Hägerdal B, Boles E. 2002. Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. Microbiology 148:2783–2788.

Huyn M, Steiner W, Klinger R, Steinmueller H, Sinner M, Esterbauer H. 1993. In: Saddler JN, editor. Biotechnology in agriculture: Bioconversion of forest and agricultural plant residues. Wallingford, UK: CAB International. p 33–72.

Heredia CF, Sols A, DelaFuente G. 1968. Specificity of the constitutive hexose transport in yeast. Eur J Biochem 5:321–329.

Hespell RB. 1998. Extraction and characterization of hemicellulose from the corn fiber produced by corn wet-milling processes. J Agric Food Chem 46:2615–2619.

Hinman ND, Wright JD, Hoagland W, Wyman CE. 1989. Xylose fermentation. An economic analysis. Appl Biochem Biotechnol 20/21: 391–401.

Ho NWY, Chang S-F. 1989. Cloning of yeast xylulokinase gene by complementation of *E. coli* and yeast mutations. Enzyme Microb Technol 11:417–421.

Johansson B. 2001. Metabolic engineering of the pentose phosphate pathway of xylose fermenting *Saccharomyces cerevisiae*. PhD thesis. Lund University

Kacser H, Burns JA. 1973. The control of flux. Symp Soc Exp Biol 27: 65–104.

Kilian SG, van Uden N. 1988. Transport of xylose and glucose in the xylose-fermenting yeast *Pichia stipitis*. Appl Microbiol Biotechnol 27: 545–548.

Kilian SG, Prior BA, du Preez JC. 1993. The kinetics and regulation of D-xylose transport in *Candida utilis*. World J Microbiol Biotechnol 9: 357–360.

Kötter P, Ciriacy M. 1993. Xylose fermentation by *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 38:776–783.

Kötter P, Amore R, Hollenberg CP, Ciriacy M. 1990. Isolation and characterization of the *Pichia stipitis* xylitol dehydrogenase gene, *XYL2*, and construction of a xylose-utilizing *Saccharomyces cerevisiae* transformant. Curr Genet 18:493–500.

Kotyk A. 1967. Mobility of the free and of the loaded monosaccharide carrier in *Saccharomyces cerevisiae*. Biochim Biophys Acta 135: 112–119.

Lindén T, Hahn-Hägerdal B. 1989. Fermentation of lignocellulose hydrolysates with yeasts and xylose isomerase. Enzyme Microb Technol 11: 583–589.

Lucas C, van Uden N. 1986. Transport of hemicellulose monomers in the xylose-fermenting yeast *Candida shehatae*. Appl Microbiol Biotechnol 23:491–495.

Nobre A, Lucas C, Leão C. 1999. Transport and utilization of hexoses and pentoses in the halotolerant yeast *Debaryomyces hansenii*. Appl Environ Microbiol 65:3594–3598.

Sjöström E. 1993. Wood chemistry: fundamentals and applications. San Diego, CA: Academic Press.

Spencer-Martins I, van Uden N. 1985. Catabolite interconversion of glucose transport systems in the yeast *Candida wickerhamii*. Biochim Biophys Acta 812:168–172.

Stenberg K, Tengborg C, Galbe M, Zacchi G, Palmqvist E, Hahn-Hägerdal B. 1998. Recycling of process streams in ethanol production from softwoods based on enzymatic hydrolysis. Appl Biochem Biotechnol 70–72:697–708.

Tantirungkij M, Nakashima N, Seki T, Yoshida T. 1993. Construction of xylose-assimilating *Saccharomyces cerevisiae*. J Ferm Bioeng 75: 83–88.

Teusink B, Diderich JA, Westerhoff HV, van Dam K, Walsh MC. 1998. Intracellular glucose concentration in derepressed yeast cells consuming glucose is high enough to reduce the glucose transport rate by 50%. J Bacteriol 180:556–562.

Verduyn C, Postma E, Scheffers WA, van Dijken JP. 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8: 501–517.

Vinzant TB, Ponfick L, Nagle NJ, Ehrman CI, Reynolds JB, Himmel ME.

BNSDOCID: <XP_____2421282A__I_>

1994. SSF comparison of selected woods from southern sawmills. Appl Biochem Biotechnol 45–46:611–626.

Von Sivers M, Zacchi G. 1996. Ethanol from lignocellulosics: a review of the economy. Bioresour Technol 56:131–140.

Wahlbom CF. 2002. Metabolic engineering and random mutagenesis for improved xylose utilisation of *Saccharomyces cerevisiae*. PhD thesis. Lund University, Sweden.

Wahlbom CF, Eliasson A, Hahn-Hägerdal B. 2001. Intracellular fluxes in a recombinant xylose-utilizing *Saccharomyces cerevisiae* cultivated anaerobically at different dilution rates and feed concentrations. Biotechnol Bioeng 72:289–296.

Weierstall T, Hollenberg CP, Boles E. 1999. Cloning and characterization of three genes (*SUT1-3*) encoding glucose transporters of the yeast *Pichia stipitis*. Mol Microbiol 31:871–883.

Ye L. 1999. Glucose transport in *Saccharomyces cerevisiae*: effects on growth and metabolism. PhD thesis, University of Amsterdam, The Netherlands.

BNSDOCID: <XP_____2421282A__I_>

(3)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 April 2003 (03.04.2003)

PCT

(10) International Publication Number
WO 03/027249 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/30475

(22) International Filing Date:
26 September 2002 (26.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/325,277	26 September 2001 (26.09.2001)	US
60/370,526	4 April 2002 (04.04.2002)	US
60/370,620	4 April 2002 (04.04.2002)	US

(71) Applicant *(for all designated States except US)*: **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SU, Wenpei** [CN/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **ANDON, Nancy** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row CA 92007, San Diego, CA 92121 (US). **HAYNES, Paul** [GB/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **BRIGGS, Steven, P.** [US/US]; Torrey Mesa Research Institute, 3115 Merry field Row, San Diego, CA 92121 (US). **COOPER, Bret** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **GLAZEBROOK, Jane** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **GOFF, Stephen, A.** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **KATAGIRI, Fumiaki** [JP/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **KREPS, Joel** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **MOUGHAMER, Todd** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **PROVART, Nicholas** [CA/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **RICKE, Darrell** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US). **ZHU, Tong** [US/US]; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US).

(74) Agent: **TORCHIA, Timothy, E.**; Torrey Mesa Research Institute, 3115 Merryfield Row, San Diego, CA 92121 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/027249 A2

(54) Title: HIGH-PROTEIN-PHENOTYPE-ASSOCIATED PLANT GENES

(57) Abstract: Methods to identify genes associated with a high protein phenotype in plants are provided, as well as the gene identified thereby and their use.

HIGH-PROTEIN-PHENOTYPE-ASSOCIATED PLANT GENES

Cross-Reference to Related Applications

This application claims the benefit of U.S. Provisional Application No. 60/325,277, filed September 26, 2001, U.S. Provisional Application No. 60/370,526 filed April 4, 2002, and U.S. Provisional Application No. 60/370,620 filed April 4, 2002, each of which is incorporated herein by reference in its entirety.

Reference to Material Submitted on Compact Disc

The sequence listing accompanying this application is contained on compact disc. The material on the CD-ROM (filed herewith), on CD volumes labeled "COPY 1 – SEQUENCE LISTING", "COPY 2 – SEQUENCE LISTING", "COPY 3 – SEQUENCE LISTING" and "CRF", each containing a text file named "60011-PCT Seq List.txt" created September 26, 2002, having a size of 133,120 bytes, is hereby incorporated by reference in its entirety pursuant to 37 C.F.R. § 1.52(e)(5).

Field of the Invention

The present invention generally relates to the field of plant molecular biology, and more specifically to plant genes useful to alter the protein content or level in plants and to develop molecular markers for plant breeding..

Background of the Invention

Farmers grow conventional maize on an estimated 100 million hectares (200 million acres) throughout the developing world. Maize is the world's most widely grown cereal crop and an essential food source for millions of the world's poor. More than half of the world's malnourished children live in countries where maize is an important food. In 20 developing countries, primarily in Latin America and Africa, maize gruel is the main food mothers use to wean their babies, and maize is the single largest source of calories. But babies who subsist on maize can face a dangerous lack of protein during a critical stage of physical and mental development as diets high in maize lack two essential amino acids needed to prevent malnutrition.

In maize crops, the expression of storage protein genes directly affects the nutritional quality of the seed protein. The prolamine (zein) fraction of storage proteins comprises over

50% of the total protein in the mature seed, however, α-zein polypeptides which are especially abundant contain extremely low levels of the essential amino acids lysine and tryptophan. Thus, maize seed protein is deficient in these amino acids because such a large proportion of the total seed storage protein is contributed by the α-zeins (Mertz et al., 1964).

The development of breeding steps to improve maize based on the manipulation of zein profile is hampered by the complexity of the zein proteins. The term "zein" encompasses a family of some 100 related proteins. Zeins can be divided into four structurally distinct types: α-zeins include proteins with molecular weights of 19,000 and 22,000 daltons; β-zeins include proteins with a molecular weight of 14,000 daltons; (γzeins include proteins with molecular weights of 27,000 and 26,000 daltons; and δ-zeins include proteins having a molecular weight of 10,000 daltons. The α-zeins are the major zein proteins found in the endosperm of maize kernels. However, the complexity of zein proteins goes beyond these size classes. Protein sequence analyses indicates that there is microheterogenicity in zein amino acid sequences. This is in accord with isoelectric focusing analyses which show charge differences in zein proteins. Over 70 genes encoding the zein proteins have been identified (Rubenstein, 1982), and the zein genes appear to be located on at least three chromosomes. Thus, the zein proteins are encoded by a multigene family.

There are several mutations known to cause reductions in zein synthesis that lead to alterations in the amino acid content of the seed. For example, in the seeds of plants homozygous for the recessive mutation *opaque-2*, the zein content is reduced by approximately 50% (Tsai et al., 1978). The *opaque-2* mutation primarily affects synthesis of the 19 and 22 kD α-zein proteins, causing a significant decrease in the level of the 19 kD zein fraction and reducing the accumulation of the 22 kD zein fraction to barely detectable levels (Jones et al., 1977). In this mutant, there is a concomitant increase in the proportion of more nutritionally balanced proteins , e.g., albumins, globulins and glutelins, deposited in the seed. The net result of the altered storage protein patterns is an increase in the essential amino acids lysine and tryptophan in the mutant seed (Misra et al., 1972). However, *opaque-2* maize has low yields, chalky-looking grain, and susceptibility to pests and diseases.

Two other recessive mutations, *floury*-2 and *sugary*-1, result in increased levels of methionine in the seed. The increased methionine content in the seeds of *floury*-2 mutants is the result of a decrease in the zein/glutelin ratio, due to reductions in the levels of both the 19 and 22 kD ∀-zein fractions, and an apparent increase in the methionine content of the glutelin fraction (Hansel et al., 1973; Jones, 1978). In *sugary*-1 mutants, there is a decrease in zein

synthesis coupled with an increase in the methionine content of the zein and glutelin fractions (Paulis et al., 1978).

As demonstrated by the *opaque*-2, *floury*-2, and *sugary*-1 mutations, reductions in zein synthesis and/or changes in the relative proportions of the storage protein fractions can affect the overall amino acid composition of the seed. Unfortunately, poor agronomic characteristics (kernel softness, reduced yield, lowered resistance to disease) are associated with the *opaque* and *floury* mutations, preventing their ready application in commercial breeding.

Another way that genes can be down regulated in animals and plants involves the expression of antisense genes. A review of the use of antisense genes in manipulating gene expression in plants can be found in van der Krol et al. (1988a; 1988b). The inhibition of expression of several endogenous plant genes has been reported. For example, U.S. Patent No. 5,107,065 discloses down regulation of polygalacturonase activity by expression of an antisense gene. Other plant genes down regulated using antisense genes include the genes encoding chalcone synthase and the small subunit of ribulose-1,5-biphosphate carboxylase (van der Krol et al., 1988c; Rodermel et al., 1988).

Down regulation of gene expression in a plant may also occur through expression of a particular transgene. This type of down regulation is referred to as co-suppression and involves coordinate silencing of a transgene and a second transgene or a homologous endogenous gene (Matzke and Matzke, 1995). For example, cosuppression of a herbicide resistance gene in tobacco (Brandle et al., 1995), polygalacturonidase in tomato (Flavell, 1994) and chalcone synthase in petunia (U. S. Patent No. 5,034,323) have been demonstrated. Flavell (1994) suggested that multicopy genes, or gene families, must have evolved to avoid cosuppression in order for multiple copies of related genes to be expressed in a plant.

Recently, a new corn variety was prepared which contains nearly twice as much usable protein as other maize grown in the tropics and yields 10 percent more grain. The new maize variety, called "quality protein maize" (QPM), was developed through traditional plant breeding and looks and tastes like normal maize, but the nutritive value of its protein is nearly equivalent to cow's milk. In particular, the varieties produce 70-100 percent more of lysine and tryptophan. A bumper crop of the maize is expected in the coming months from more than one million hectares (2.5 million acres) currently under cultivation in 11 countries. Economists expect that by 2003, the number of hectares sown to QPM will triple to approximately 3.5 million hectares (8.75 million acres). Moreover, as incomes rise in Asia, researchers expect that the use of maize in animal feed will increase by more than three percent

42

each year between now and the year 2020. The high protein maize fattens pigs and poultry more efficiently, enabling poor farming families to increase their incomes. Pigs and poultry raised on this type of maize gain weight roughly twice as fast as animals fed on conventional maize. However, QPM is the result of more than three decades of scientific discovery.

Thus, there is a need for improved methods to alter the nutritional content of seeds and plants to produce kernels and plants with good agronomic characteristics, while maintaining the phenotype of the parent, e.g., kernel hardness, yield, and disease resistance.

Plants are increasingly used a "protein factories" for production of industrial or therapeutic polypeptides, such as antigens, antibodies (e.g., monoclonal), cytokines, vaccines. Methods for increased yield and/or quality or ease of downstream processing are needed.

Thus, there is a need for improved methods and compositons to alter the protein content of seeds and plants to produce kernels and plants with good characteristics for production of important polypeptides.

Summary of the Invention

Proteins and genes involved in a tropical high protein trait corn germplasm are disclosed, as well as their use to genetically modify cereals for higher protein yield and better protein quality. A total 11 genes (and thir orthologs) are identified for use in protein trait modification in cereals, particularly corn. These genes belong to two groups: one group of proteins is associated with seed protein storage and the other group is generally related to seed stress response or proteins that are unregulated during seed maturation. Possibly the stress response mechanism has co-evolved with the high protein trait. Higher protein yield in corn and other cereals can be achieved by manipulating the gene expression level of these genes and other regulatory genes regulating the stress mechanism.

Accordingly, the invention provides isolated nucleic acid molecules, e.g., DNA, comprising a plant nucleotide sequence encoding a polypeptide that is expressed in cells of a plant, e.g., embryos, mature embryos, endosperm, shoot, root, leaf and developing seed, from high protein varieties of plants, relative to cells of a plant from a corresponding lower protein variety. For example, the invention provides a nucleic acid molecule comprising a plant nucleotide sequence comprising an open reading frame encoding a polypeptide which is substantially similar to a polypeptide comprising any one of SEQ ID Nos. 1-36. To provide altered protein content to a

plant, this sequence may be overexpressed individually, in the sense or antisense orientation, or in combination with other sequences, to confer altered nutritional properties to the plant relative to a plant that does not comprise and/or express the sequence(s). Thus, in one embodiment, the protein content may be enhanced, while in another embodiment it may be reduced, e.g., low protein products such as rice for individuals that are intolerant or sensitive to certain proteins. Low protein content plants or seeds can be a superior form for production of heterologous industrially or therapeutically important proteins in plants, and plant seeds by, for example, reducing levels of abundant endogenous proteins. To avoid detrimental effects to the plants, such modulation can be controlled using inducible promoters. One system employs hybrid two component systems such as Gal4/C1, in which the controlled promoter(s) is on or off only in the hybrid, not the parental lines. The overexpression may be constitutive, or it may be preferable to express the sequence from an inducible promoter including a promoter which is responsive to external stimuli, such as chemical application, or environmental stimuli, so as to avoid possible deleterious effects on plant growth. High protein varieties of plants are those which have at least a 1%, preferably at least 5%, and more preferably at least 10%, increase in protein content or level relative to a corresponding control plant. For example, for maize, a high protein line or variety preferably may have a protein content in whole kernel that is at least 14.5%, more preferably at least 15.5%, in embyro at least 17%, more preferably 18.3%, and in endosperm at least 13.5%, more preferably at least 14.2%. High protein varieties of maize are well known to the art, see, for example, U.S. Patent Nos. 5,986,182, 5,936,143, 5,907,089, 5,900,528, 5,850,031, 5,824,855, 5,824,854, 5,763,756, and 5,675,065.

As described herein, protein expression profiles from embryos of normal and high protein varieties of maize were compared using two-dimensional SDS-PAGE analysis in order to identify differentially expressed genes. Application of proteomic technology to the high protein corn germplasm has revealed more than 120 genes that are differentially expressed in high protein lines. Such genes may encode structural or regulatory proteins, and hence are of potential use in manipulating protein content in maize (corn) and other cereals such as wheat and rice, e.g., for manipulating seed protein phenotype and for the development of molecular markers for plant breeding. Moreover, based on the proteomic approach, the results provide a novel function for

unknown or previously uncharacterized/ mischaracterized genes, and may lead to useful regulatory genes for particular traits, structural genes or molecular markers. Further, by using a segregating population, the results also provide the necessary means to identify genes specifically related to the high protein phenotype rather than those that are merely causally associated. Thus, the identified proteins (polypeptides) and their corresponding genes can be used to: 1) manipulate protein content or levels in corn and other cereal species, e.g., by using the genes as molecular markers in breeding or in transgenic plants; 2) isolate orthologs from other crop species such as rice and wheat; 3) generate antibodies and develop protein-based assays for breeding selection; and 4) identify common transcriptional regulatory elements and factors which bind those elements, i.e., the upstream regions of the genes associated with the high protein trait.

Non-protein based methods may also be employed to identify the nucleic acid molecules of the invention. For example, an array of nucleic acid samples, e.g., a plurality of oligonucleotides, each plurality corresponding to a different plant gene, on a solid substrate, e.g., a DNA chip, and probes corresponding to nucleic acid obtained from plant sources that express genes associated with protein content and probes to nucleic acid obtained from plant sources that do not express those genes or express the genes at a reduced level, can be used to systematically identify genes associated with increased protein levels.

Preferably, the nucleotide sequence in the nucleic acid molecule of the invention is from plant DNA, either a dicot or a monocot, which encodes a polypeptide that is substantially similar to a polypeptide comprising any one of SEQ ID NOs: 1-36. More preferably, the nucleotide sequence is from plant DNA that is substantially similar to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36. The term "substantially similar", when used herein with respect to a polypeptide means a polypeptide corresponding to a reference polypeptide, wherein the polypeptide has substantially the same structure and function as the reference polypeptide, e.g., where only changes in amino acid sequence are those which do not affect the polypeptide function. When used for a polypeptide or an amino acid sequence, the percentage of identity between the substantially similar and the reference polypeptide or amino acid sequence is at least 65%, 66%, 67%, 68%, 69%, 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, where the reference polypeptide is a polypeptide comprising any one of SEQ ID NOs: 1-36. One indication that two polypeptides are substantially similar to each other is that

an agent, e.g., an antibody, which specifically binds to one of the polypeptides, specifically binds to the other.

In its broadest sense, the term "substantially similar", when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence, wherein the corresponding sequence encodes a polypeptide having substantially the same structure and function as the polypeptide encoded by the reference nucleotide sequence. The term "substantially similar" is specifically intended to include nucleotide sequences wherein the sequence has been modified to optimize expression in particular cells. The percentage of identity between the substantially similar nucleotide sequence and the reference nucleotide sequence is at least 65%, 66%, 67%, 68%, 69%, 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, up to at least 99%, wherein the reference sequence is one which encodes a polypeptide comprising any one of SEQ ID NOs: 1-36, or the complement thereof. Sequence comparisons maybe carried out using a Smith-Waterman sequence alignment algorithm (see e.g. Waterman (1995) or http://www hto.usc.edu/software/seqaln/index.html). The localS program, version 1.16, is preferably used with following parameters: match: 1, mismatch penalty: 0.33, open-gap penalty: 2, extended-gap penalty: 2. Further, a nucleotide sequence that is "substantially similar" to a reference nucleotide sequence hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

Hence, the isolated nucleic acid molecules of the invention also include orthologs of the sequences encoding the polypeptides comprising the amino acid sequences disclosed herein, including, but not limited to, dicots and monocots, preferably cereal plants, e.g., wheat or rice. An ortholog is a gene from a different species that encodes a product having the same function as the product encoded by a gene from a reference organism. The encoded ortholog products likely have at least 70% sequence identity to each other. Hence, the invention

includes an isolated nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide having at least 70% identity to a polypeptide comprising one or more of the sequences disclosed herein. Databases such GenBank or one found at http://bioserver.myongjiac.kr/rjce.html (for rice) may be employed to identify sequences related to the disclosed sequences, e.g., orthologs in cereal crops such as rice. Alternatively, recombinant DNA techniques such as hybridization or PCR may be employed to identify sequences related to the disclosed sequences.

Preferably, the polypeptide has substantial identity, i.e., at least 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, and at least 99%, amino acid sequence identity to a polypeptide comprising any one of SEQ ID NOs: 1-36. The invention also provides anti-sense nucleic acid molecules corresponding to the genes identified herein. Also provided are expression cassettes, e.g., recombinant vectors, and host cells, comprising the nucleic acid molecule of the invention.

The nucleic acid molecules of the invention, their encoded polypeptides and compositions thereof, are useful to provide plants with enhanced protein content, identify common transcriptional regulatory factors which bind upstream of the coding region of genes associated with high protein content and as markers for breeding selection. The compositions of the invention include plant nucleic acid sequences and the amino acid sequences for the polypeptides or partial-length polypeptides encoded thereby which are useful to provide enhanced nutritional characteristics to a plant, preferably by enhancing protein content or levels. Methods of the invention involve stably transforming a plant with one or more of at least a portion of these nucleotide sequences operably linked to a promoter capable of driving expression of that nucleotide sequence in a plant cell. By "portion" or "fragment", as it relates to a nucleic acid molecule, sequence or segment of the invention, when it is linked to other sequences for expression, is meant a sequence having at least 80 nucleotides, more preferably at least 150 nucleotides, and still more preferably at least 400 nucleotides. If not employed for expressing, a "portion" or "fragment" means at least 9, preferably 12, more preferably 15, even more preferably at least 20, consecutive nucleotides, e.g., probes and primers (oligonucleotides), corresponding to the nucleotide sequence of the nucleic acid molecules of the invention. The method comprises introducing to a plant, plant cell, or plant tissue an expression cassette comprising at least one of nucleic acid molecules of the invention so as to yield a transformed differentiated plant, transformed cell or transformed tissue. Transformed

cells or tissue can be regenerated to provide a transformed differentiated plant. The transformed differentiated plant preferably expresses the nucleic acid molecule in an amount that yields a transformed plant having enhanced protein content, e.g., in seed, to a corresponding nontransformed plant. The present invention also provides a transformed plant prepared by the method, progeny and seed thereof.

A transformed (transgenic) plant of the invention includes plants, dicots or monocots, the genome of which is augmented by a nucleic acid molecule of the invention, or in which the corresponding gene has been disrupted, e.g., to result in a loss, a decrease or an alteration, in the function of the product encoded by the gene. The nucleic acid molecules of the invention are thus useful for targeted gene disruption, as well as for markers and probes.

The invention also includes recombinant nucleic acid molecules which have been modified so as to comprise codons other than those present in the unmodified sequence. The recombinant nucleic acid molecules of the invention include those in which the modified codons specify amino acids that are the same as those specified by the codons in the unmodified sequence, as well as those that specify different amino acids, i.e., they encode a variant polypeptide having one or more amino acid substitutions relative to the polypeptide encoded by the unmodified sequence.

The invention further includes a nucleotide sequence which is complementary to one (hereinafter "test" sequence) which hybridizes under stringent conditions with the nucleic acid molecules of the invention as well as RNA which is encoded by the nucleic acid molecule. When the hybridization is performed under stringent conditions, either the test or nucleic acid molecule of invention is preferably supported, e.g., on a membrane or DNA chip. Thus, either a denatured test or nucleic acid molecule of the invention is preferably first bound to a support and hybridization is effected for a specified period of time at a temperature of, e.g., between 55 and 70°C, in double strength citrate buffered saline (SC) containing 0.1% SDS followed by rinsing of the support at the same temperature but with a buffer having a reduced SC concentration. Depending upon the degree of stringency required such reduced concentration buffers are typically single strength SC containing 0.1% SDS, half strength SC containing 0.1% SDS and one-tenth strength SC containing 0.1% SDS.

The present invention also provides a method to identify a polypeptide which is associated with a high protein phenotype. The method comprises separating a plurality of polypeptides from a sample comprising polypeptides, wherein the sample is from a plant having a high protein content. Then the separated sample of polypeptides from a plant having

a high protein content is compared to a separated sample of polypeptides from a corresponding plant with lower protein content. Preferably, polypeptides are identified that are present in the sample from a plant having a high protein content that are not present in the sample from the plant with lower protein content.

Also provided is an isolated nucleic acid molecule comprising a nucleotide sequence that directs transcription, e.g., a promoter, of a linked nucleic acid fragment in a host cell, such as a plant cell. It is preferred that the nucleotide sequence is from plant genomic DNA which has at least 65%, 66%, 67%, 68%, 69%, 70%, e.g., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, and even 90% or more, e.g., 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, and 99%, nucleotide sequence identity to a sequence of a promoter from a plant gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36. The promoter sequence is preferably about 25 to 2000, e.g., 50 to 500 or 100 to 1400, nucleotides in length. In one embodiment of the invention, the isolated nucleic acid molecule comprises a plant nucleotide sequence which is the promoter region for a gene encoding any one of SEQ ID NOs: 1-36, or is structurally related to the promoter for a gene encoding SEQ ID NOs: 1-36, i.e., is an orthologous promoter, and is linked to a plant structural gene. Hence, the present invention further provides an expression cassette or a recombinant vector containing the nucleic acid molecule, and the vector may be a plasmid. Such cassettes or vectors, when present in a plant, plant cell or plant tissue result in transcription of the linked nucleic acid fragment in the plant, plant tissue or plant cell.

The expression cassettes or vectors of the invention may optionally include other regulatory sequences, e.g., transcription terminator sequences, introns and/or enhancers, and may be contained in a host cell. The expression cassette or vector may augment the genome of a transformed plant or may be maintained extrachromosomally. The expression cassette or vector may further have a Ti plasmid and be contained in an *Agrobacterium tumefaciens* cell; it may be carried on a microparticle, wherein the microparticle is suitable for ballistic transformation of a plant cell; or it may be contained in a plant cell protoplast. Further, the expression cassette can be contained in a plant, plant cell or plant tissue from a dicot or a monocot. In particular, the plant may be a cereal plant.

The present invention further provides a method of augmenting a plant genome by contacting plant cells with an expression cassette or vector of the invention, i.e., one having a nucleotide sequence that directs transcription of a linked nucleic acid fragment in a plant cell, wherein the nucleotide sequence is from plant genomic DNA that has at least 65%, and more

preferably at least 70%, identity to the sequence of a promoter from a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36 so as to yield transformed plant cells; and regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the linked fragment in the cells of the plant. The present invention also provides a plant prepared by the method, progeny and seed thereof.

Brief Description of the Figures

Figure 1 shows the protein content in various sources from high protein and control maize lines.

Figures 2A and 2B illustrate a two dimensional gel with proteins from a control (#530; panel A) or high protein (#465; panel B) maize line. Figures 2C and 2D illustrate another comparision of protein expression profile of high protein germplasm and normal corn line.

Figures 3A to 3H show the peptide and criteria (e.g., Xcro > 2 and Dcn > 0.01) employed to search databases for the corresponding full length protein for 18 of the proteins shown in the attached Sequence Listing which is incorporated herein.

Figures 4A and 4B are representative vectors for over- or under-expression of genes in seed.

The Sequence Listing shows the amino acid sequence of proteins, high protein phenotype genes and proteins, or the orthologs thereof, which are preferentially expressed in high protein maize lines relative to lines with lower protein content. Sequences 1 to 36 are the high protein involved proteins. Odd numbered SEQ ID Nos are are protein-encoding orfs and the even numbered SEQ ID NOs are amino acids sequences. Sequences 37 to 45 are representative peptides identified by the MS as described in the Examples.

Detailed Description of the Invention

Definitions

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, composed of monomers (nucleotides) containing a sugar, phosphate and a base which is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively

modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., 1991; Ohtsuka et al., 1985; Rossolini et al., 1994). A "nucleic acid fragment" is a fraction of a given nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. The terms "nucleic acid", "nucleotide sequence", "nucleic acid molecule", "nucleic acid fragment" or "nucleic acid sequence or segment" may also be used interchangeably with gene, cDNA, DNA and RNA encoded by a gene.

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. In the context of the present invention, an "isolated" or "purified" DNA molecule or an "isolated" or "purified" polypeptide is a DNA molecule or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule or polypeptide may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein or polypeptide having less than about 30%, 20%, 10%, 5%, (by dry weight) of contaminating protein. When the protein of the invention, or biologically active portion thereof, is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein-of- interest chemicals.

Fragments and variants of the disclosed nucleotide sequences and proteins or partial-length proteins encoded thereby are also encompassed by the present invention. By "fragment" or "portion" is meant a full length or less than full length of the nucleotide sequence encoding, or the amino acid sequence of, a polypeptide or protein. Alternatively, fragments or portions of a nucleotide sequence that are useful as hybridization probes generally do not encode fragment proteins retaining biological activity. Thus, fragments or portions of a nucleotide sequence may range from at least about 9 nucleotides, about 12 nucleotides, about 20 nucleotides, about 50 nucleotides, about 100 nucleotides or more.

The term "gene" is used broadly to refer to any segment of nucleic acid associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. For example, gene refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

"Naturally occurring" is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a protein or nucleotide sequence present in an organism (including a virus), which can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory, is naturally occurring.

A "marker gene" encodes a selectable or screenable trait.

"Selectable marker" is a gene whose expression in a cell gives the cell a selective advantage. The selective advantage possessed by the cells transformed with the selectable marker gene may be due to their ability to grow in the presence of a negative selective agent, such as an antibiotic or a herbicide, compared to the growth of non-transformed cells. The selective advantage possessed by the transformed cells, compared to non-transformed cells, may also be due to their enhanced or novel capacity to utilize an added compound as a nutrient, growth factor or energy source. Selectable marker gene also refers to a gene or a combination of genes whose expression in a cell gives the cell both a negative and/or a positive selective advantage.

The term "chimeric" refers to any gene or DNA that contains 1) DNA sequences, including regulatory and coding sequences, that are not found together in nature, or 2)

sequences encoding parts of proteins not naturally adjoined, or 3) parts of promoters that are not naturally adjoined. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or comprise regulatory sequences and coding sequences derived from the same source, but arranged in a manner different from that found in nature.

A "transgene" refers to a gene that has been introduced into the genome by transformation and is stably maintained. Transgenes may include, for example, DNA that is either heterologous or homologous to the DNA of a particular plant to be transformed. Additionally, transgenes may comprise native genes inserted into a non-native organism, or chimeric genes. The term "endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism but that is introduced by gene transfer.

The terms "protein," "peptide" and "polypeptide" may be used interchangeably herein.

By "variants" is intended substantially similar sequences. For nucleotide sequences, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical amino acid sequence of the native protein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis which encode the native protein, as well as those that encode a polypeptide having amino acid substitutions. Generally, nucleotide sequence variants of the invention will have at least 40, 50, 60, to 70%, e.g., preferably 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, generally at least 80%, e.g., 81%-84%, at least 85%, e.g., 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, to 98%, sequence identity to the native (endogenous) nucleotide sequence.

"DNA shuffling" is a method to introduce mutations or rearrangements, preferably randomly, in a DNA molecule or to generate exchanges of DNA sequences between two or more DNA molecules, preferably randomly. The DNA molecule resulting from DNA shuffling is a shuffled DNA molecule that is a non-naturally occurring DNA molecule derived from at least one template DNA molecule. The shuffled DNA preferably encodes a variant polypeptide modified with respect to the polypeptide encoded by the template DNA, and may have an altered biological activity with respect to the polypeptide encoded by the template DNA.

The nucleic acid molecules of the invention can be optimized for enhanced expression in plants of interest. See, for example, EPA035472; WO91/16432; Perlak et al., 1991; and Murray et al., 1989. In this manner, the genes or gene fragments can be synthesized utilizing plant-preferred codons. See, for example, Campbell and Gowri (1990) for a discussion of host-preferred codon usage. Thus, the nucleotide sequences can be optimized for expression in any plant. It is recognized that all or any part of the gene sequence may be optimized or synthetic. That is, synthetic or partially optimized sequences may also be used. Variant nucleotide sequences and proteins also encompass sequences and protein derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different coding sequences can be manipulated to create a new polypeptide possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994); Stemmer (1994); Crameri et al. (1997); Moore et al. (1997); Zhang et al. (1997); Crameri et al. (1998); and U.S. Patent Nos. 5,605,793 and 5,837,458.

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or essentially identical amino acid sequences, or where the nucleic acid sequence does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance the codons CGT, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations" which are one species of "conservatively modified variations." Every nucleic acid sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

"Recombinant DNA molecule" is a combination of DNA sequences that are joined together using recombinant DNA technology and procedures used to join together DNA sequences as described, for example, in Sambrook et al.

The terms "heterologous DNA sequence," "exogenous DNA segment" or "heterologous nucleic acid," each refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

A "homologous" DNA sequence is a DNA sequence that is naturally associated with a host cell into which it is introduced.

"Wild-type" refers to the normal gene, or organism found in nature without any known mutation.

"Genome" refers to the complete genetic material of an organism.

"Vector" is defined to include, inter alia, any plasmid, cosmid, phage or *Agrobacterium* binary vector in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable, and which can transform prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g., autonomous replicating plasmid with an origin of replication).

Specifically included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g., higher plant, mammalian, yeast or fungal cells).

"Cloning vectors" typically contain one or a small number of restriction endonuclease recognition sites at which foreign DNA sequences can be inserted in a determinable fashion without loss of essential biological function of the vector, as well as a marker gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Marker genes typically include genes that provide tetracycline resistance, hygromycin resistance or ampicillin resistance.

"Expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, the promoter can also be specific to a particular tissue or organ or stage of development.

Such expression cassettes will comprise the transcriptional initiation region of the invention linked to a nucleotide sequence of interest. Such an expression cassette is provided with a plurality of restriction sites for insertion of the gene of interest to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

The transcriptional cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region, a DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be native with the transcriptional initiation region, may be native with the DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also, Guerineau et al. (1991); Proudfoot (1991); Sanfacon et al. (1991); Mogen et al. (1990); Munroe et al. (1990); Ballas et al. (1989); Joshi et al. (1987).

An oligonucleotide corresponding to a nucleic acid molecule of the invention may be about 30 or fewer nucleotides in length (e.g., 9, 12, 15, 18, 20, 21 or 24, or any number between 9 and 30). Generally specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16-24 nucleotides in length may be

preferred. Those skilled in the art are well versed in the design of primers for use processes such as PCR. If required, probing can be done with entire restriction fragments of the gene disclosed herein which may be 100's or even 1000's of nucleotides in length.

"Coding sequence" refers to a DNA or RNA sequence that codes for a specific amino acid sequence and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", i.e., lacking an intron, such as in a cDNA or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a sequence of RNA which is contained in the primary transcript but which is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

The terms "open reading frame" and "ORF" refer to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides ('codon') in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

A "functional RNA" refers to an antisense RNA, ribozyme, or other RNA that is not translated.

The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

"Regulatory sequences" and "suitable regulatory sequences" each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. As is noted above, the term "suitable regulatory sequences" is not limited to promoters. However, some suitable regulatory sequences useful in the present invention will include, but are not limited to constitutive plant

promoters, plant tissue-specific promoters, plant development specific promoters, inducible plant promoters and viral promoters.

"5' non-coding sequence" refers to a nucleotide sequence located 5' (upstream) to the coding sequence. It is present in the fully processed mRNA upstream of the initiation codon and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency (Turner et al., 1995).

"3' non-coding sequence" refers to nucleotide sequences located 3' (downstream) to a coding sequence and include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., 1989.

The term "translation leader sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

The term "mature" protein refers to a post-translationally processed polypeptide without its signal peptide. "Precursor" protein refers to the primary product of translation of an mRNA. "Signal peptide" refers to the amino terminal extension of a polypeptide, which is translated in conjunction with the polypeptide forming a precursor peptide and which is required for its entrance into the secretory pathway. The term "signal sequence" refers to a nucleotide sequence that encodes the signal peptide.

The term "intracellular localization sequence" refers to a nucleotide sequence that encodes an intracellular targeting signal. An "intracellular targeting signal" is an amino acid sequence that is translated in conjunction with a protein and directs it to a particular sub-cellular compartment. "Endoplasmic reticulum (ER) stop transit signal" refers to a carboxy-terminal extension of a polypeptide, which is translated in conjunction with the polypeptide and causes a protein that enters the secretory pathway to be retained in the ER. "ER stop transit sequence" refers to a nucleotide sequence that encodes the ER targeting signal. Other intracellular targeting sequences encode targeting signals active in seeds and/or leaves and vacuolar targeting signals.

"Promoter" refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. "Promoter" includes a minimal promoter that is a short DNA sequence comprised of a TATA- box and other sequences that serve to specify the site of transcription initiation, to which regulatory elements are added for control of expression. "Promoter" also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements that is capable of controlling the expression of a coding sequence or functional RNA. This type of promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other upstream promoter elements bind sequence-specific DNA-binding proteins that mediate their effects. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also contain DNA sequences that are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions.

The "initiation site" is the position surrounding the first nucleotide that is part of the transcribed sequence, which is also defined as position +1. With respect to this site all other sequences of the gene and its controlling regions are numbered. Downstream sequences (i.e., further protein encoding sequences in the 3' direction) are denominated positive, while upstream sequences (mostly of the controlling regions in the 5' direction) are denominated negative.

Promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation are referred to as "minimal or core promoters." In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription. A "minimal or core promoter" thus consists only of all basal elements needed for transcription initiation, e.g., a TATA box and/or an initiator.

"Constitutive expression" refers to expression using a constitutive or regulated promoter. "Conditional" and "regulated expression" refer to expression controlled by a regulated promoter.

"Constitutive promoter" refers to a promoter that is able to express the gene that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant. Each of the transcription-activating elements do not exhibit an absolute tissue-specificity, but mediate transcriptional activation in most plant parts at a level of ≥1% of the level reached in the part of the plant in which transcription is most active.

"Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and include both tissue-specific and inducible promoters. It includes natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. New promoters of various types useful in plant cells are constantly being discovered, numerous examples may be found in the compilation by Okamuro et al. (1989). Since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity. Typical regulated promoters useful in plants include but are not limited to safener-inducible promoters, promoters derived from the tetracycline-inducible system, promoters derived from salicylate-inducible systems, promoters derived from alcohol-inducible systems, promoters derived from glucocorticoid-inducible system, promoters derived from pathogen-inducible systems, and promoters derived from ecdysome-inducible systems.

"Tissue-specific promoter" refers to regulated promoters that are not expressed in all plant cells but only in one or more cell types in specific organs (such as leaves or seeds), specific tissues (such as embryo or cotyledon), or specific cell types (such as leaf parenchyma or seed storage cells). These also include promoters that are temporally regulated, such as in early or late embryogenesis, during fruit ripening in developing seeds or fruit, in fully differentiated leaf, or at the onset of senescence.

"Inducible promoter" refers to those regulated promoters that can be turned on in one or more cell types by an external stimulus, such as a chemical, light, hormone, stress, or a pathogen.

"Operably-linked" refers to the association of nucleic acid sequences on single nucleic acid fragment so that the function of one is affected by the other. For example, a regulatory

DNA sequence is said to be "operably linked to" or "associated with" a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

"Expression" refers to the transcription and/or translation of an endogenous gene or a transgene in plants. For example, in the case of antisense constructs, expression may refer to the transcription of the antisense DNA only. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

"Altered levels" refers to the level of expression in transgenic cells or organisms that differs from that of normal or untransformed cells or organisms.

"Overexpression" refers to the level of expression in transgenic cells or organisms that exceeds levels of expression in normal or untransformed cells or organisms.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from an endogenous gene or a transgene.

"Co-suppression" and "transwitch" each refer to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar transgene or endogenous genes (U.S. Patent No. 5,231,020).

"Gene silencing" refers to homology-dependent suppression of viral genes, transgenes, or endogenous nuclear genes. Gene silencing may be transcriptional, when the suppression is due to decreased transcription of the affected genes, or post-transcriptional, when the suppression is due to increased turnover (degradation) of RNA species homologous to the affected genes. (English et al., 1996). Gene silencing includes virus-induced gene silencing (Ruiz et al., 1998).

"Silencing suppressor" gene refers to a gene whose expression leads to counteracting gene silencing and enhanced expression of silenced genes. Silencing suppressor genes may be of plant, non-plant, or viral origin. Examples include, but are not limited to HC-Pro, P1-HC-Pro, and 2b proteins. Other examples include one or more genes in TGMV-B genome.

"Transcription stop fragment" refers to nucleotide sequences that contain one or more regulatory signals, such as polyadenylation signal sequences, capable of terminating transcription. Examples include the 3' non-regulatory regions of genes encoding nopaline synthase and the small subunit of ribulose bisphosphate carboxylase.

"Translation stop fragment" refers to nucleotide sequences that contain one or more regulatory signals, such as one or more termination codons in all three frames, capable of terminating translation. Insertion of a translation stop fragment adjacent to or near the initiation codon at the 5' end of the coding sequence will result in no translation or improper translation. Excision of the translation stop fragment by site-specific recombination will leave a site-specific sequence in the coding sequence that does not interfere with proper translation using the initiation codon.

The terms "*cis*-acting sequence" and "*cis*-acting element" refer to DNA or RNA sequences whose functions require them to be on the same molecule. An example of a *cis*-acting sequence on the replicon is the viral replication origin.

The terms "*trans*-acting sequence" and "*trans*-acting element" refer to DNA or RNA sequences whose function does not require them to be on the same molecule.

"Chromosomally-integrated" refers to the integration of a foreign gene or DNA construct into the host DNA by covalent bonds. Where genes are not "chromosomally integrated" they may be "transiently expressed." Transient expression of a gene refers to the expression of a gene that is not integrated into the host chromosome but functions independently, either as part of an autonomously replicating plasmid or expression cassette, for example, or as part of another biological system such as a virus.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988); the local homology algorithm of Smith et al. (1981); the homology alignment algorithm of Needleman and Wunsch (1970); the search-for-similarity-method of Pearson and Lipman (1988); the algorithm of Karlin and Altschul (1990), modified as in Karlin and Altschul (1993).

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. (1988); Higgins et al. (1989); Corpet et al. (1988); Huang et al. (1992); and Pearson et al. (1994). The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul et al. (1990), are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. (1997). Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al., *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1989). See http://www.ncbi.nlm.nih.gov. Alignment may also be performed manually by inspection.

For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window, as measured by sequence comparison algorithms or by visual inspection. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions,

where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, more preferably at least 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions (see below). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, or even more preferably, 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch, 1970. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s)

substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, 1984; T_m 81.5°C + 16.6 (log M) +0.41 (%GC) - 0.61 (% form) - 500/L; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993). Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2X SSC wash at 65°C for 15 minutes (see, Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1X SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6X SSC at 40°C for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, more preferably about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long robes (e.g., >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2X (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, e.g., hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0. 1X SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

The following are examples of sets of hybridization/wash conditions that may be used to clone orthologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M

$NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

By "variant" polypeptide is intended a polypeptide derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Such variants may results form, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

Thus, the polypeptides of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the polypeptides can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985); Kunkel et al. (1987); U. S. Patent No. 4,873,192; Walker and Gaastra (1983), and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff et al. (1978). Conservative substitutions, such as exchanging one amino acid with another having similar properties, are preferred.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the polypeptides of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired activity. The deletions, insertions, and substitutions of the polypeptide sequence encompassed herein are not expected to produce radical changes in the characteristics of the polypeptide. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays.

Individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations," where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The

following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (I); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). See also, Creighton (1984). In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

"Production tissue" refers to mature, harvestable tissue consisting of non-dividing, terminally-differentiated cells. It excludes young, growing tissue consisting of germline, meristematic, and not-fully-differentiated cells.

"Germline cells" refer to cells that are destined to be gametes and whose genetic material is heritable.

The word "plant" refers to any plant, particularly to seed plant, and "plant cell" is a structural and physiological unit of the plant, e.g., a cell which comprises a cell wall or a protoplast. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, a plant tissue, or a plant organ.

"Plant tissue" includes differentiated and undifferentiated tissues or plants, including but not limited to roots, stems, shoots, leaves, pollen, seeds, tumor tissue and various forms of cells and culture such as single cells, protoplast, embryos, and callus tissue. The plant tissue may be in plants or in organ, tissue or cell culture.

The term "altered plant trait" means any phenotypic or genotypic change in a transgenic plant relative to the wild-type or non-transgenic plant host.

The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. Host cells containing the transformed nucleic acid fragments are referred to as "transgenic" cells, and organisms comprising transgenic cells are referred to as "transgenic organisms". Examples of methods of transformation of plants and plant cells include *Agrobacterium*-mediated transformation (De Blaere et al., 1987) and particle bombardment technology (Klein et al., 1987; U.S. Patent No. 4,945,050). Whole plants may be regenerated from transgenic cells by methods well known to the skilled artisan (see, for example, Fromm et al., 1990).

"Transformed," "transgenic," and "recombinant" refer to a host cell or organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced.

The nucleic acid molecule can be stably integrated into the genome generally known in the art and are disclosed in Sambrook et al. (1989). See also Innis et al. (1995); and Gelfand (1995); and Innis and Gelfand (1999). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially mismatched primers, and the like. For example, "transformed," "transformant," and "transgenic" plants or calli have been through the transformation process and contain a foreign gene integrated into their chromosome. The term "untransformed" refers to normal plants that have not been through the transformation process.

A "transgenic plant" is a plant having one or more plant cells that contain an expression vector.

"Transiently transformed" refers to cells in which transgenes and foreign DNA have been introduced (for example, by such methods as *Agrobacterium*-mediated transformation or biolistic bombardment), but not selected for stable maintenance.

"Stably transformed" refers to cells that have been selected and regenerated on a selection media following transformation.

"Transient expression" refers to transgene expression in cells, e.g., after transformation with recombinant virus or by such methods as *Agrobacterium*-mediated transformation, electroporation, or biolistic bombardment, but not selected for its stable maintenance.

"Genetically stable" and "heritable" refer to chromosomally-integrated genetic elements that are stably maintained in the plant and stably inherited by progeny through successive generations.

"Primary transformant" and "T0 generation" refer to transgenic plants that are of the same genetic generation as the tissue which was initially transformed (i.e., not having gone through meiosis and fertilization since transformation).

"Secondary transformants" and the "T1, T2, T3, etc. generations" refer to transgenic plants derived from primary transformants through one or more meiotic and fertilization cycles. They may be derived by self-fertilization of primary or secondary transformants or crosses of primary or secondary transformants with other transformed or untransformed plants.

"Significant increase" is an increase that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater.

"Significantly less" means that the decrease is larger than the margin of error inherent in the measurement technique, preferably a decrease by about 2-fold or greater.

I. The Nucleic Acid Molecules of the Invention and Polypeptide Encoded Thereby

This invention relates to isolated plant nucleic acid molecules, sequences and segments (fragments), the expression of which is increased in plants with increased protein content or levels, as well as the endogenous plant promoters for those expressed molecules, sequences or segments. Preferred sources for the nucleic acid molecules of the invention include, but are not limited to, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers; duckweed (*Lemna*, see WO 00/07210, which includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as follows: genus *Lemna (L. aequinoctialis, L. disperma, L. ecuadoriensis, L. gibba, L. japonica, L. minor, L. miniscula, L. obscura, L. perpusilla, L. tenera, L.* trisulca, L. turionifera, L. valdiviana); *genus* Spirodela (S. intermedia, S. polyrrhiza, S.punctata); *genus* Woffia *(Wa. angusta, Wa. arrhiza, Wa. australina, Wa. borealis, Wa. brasiliensis, Wa. columbiana, Wa. elongata, Wa. globosa, Wa. microscopica, Wa. neglecta)* and genus *Wofiella (W1. caudata, W1. denticulata, W1. gladiata, W1. hyalina, W1. lingulata, W1. repunda, W1. rotunda, and W1. neotropica*). Any other genera or species of *Lemnaceae*, if they exist, are also aspects of the present invention. *Lemna gibba, Lemna minor*, and *Lemna miniscula* are preferred, with *Lemna* minor and *Lemna miniscula* being most preferred. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of Lemnaceae - A Monograph Study. Geobatanischen Institut ETH, Stiftung Rubel, Zurich (1986)); vegetables including tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca*

sativa), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, cauliflower, Brassica, e.g., broccoli, cabbage, brussels sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, asparagus, and zucchini and ornamental plants include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia.

Other vegetable sources (and databases to identify orthologs of the invention) for the nucleic acid sequences of the invention include those are shown in Table 1.

Table 1

FAMILY	LATIN NAME	COMMON NAME	MAP REFERENCES RESOURCES	LINKS
Cucurbitaceae	*Cucumis sativus*	Cucumber		http://www.cucurbit.org/
	Cucumis melo	Melon		http://genome.cornell.edu/cgc/
	Citrullus lanatus	Watermelon		
	Cucurbita pepo	Squash – summer		
	Cucurbita maxima	Squash - winter		
	Cucurbita moschata	Pumpkin /butternut		
Total				http://www.nal.usda.gov/pgdic/Map_proj/

Solanaceae	*Lycopersicon esculentum*	Tomato	• 15x BAC on variety Heinz 1706 order from Clemson Genome center (www.genome.clemson.edu) • 11.6x BAC of L. cheesmanii (orginates from J. Giovannoni) available from Clemson genome center (www.genome.clemson.edu) • EST collection from TIGR (www.tigr.org/tdb/lgi/index.html) • EST collection from Clemsom Genome Center (www.genome.clemson.edu) • TAG 99:254-271, 1999 (esculentum x pennelli) • TAG 89:1007-1013, 1994 (peruvianum) • Plant Cell Reports 12:293-297, 1993	genome.cornell.edu/solgenes http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=solgenes http://genome.cornell.edu/tgc/ http://tgrc.ucdavis.edu/

	Capsicum annuum	Pepper		http://neptune.netimages.com/~chile/science.html
	Capsicum frutescens	Chile pepper		
	Solanum melongena	Eggplant		
	(Nicotiana tabacum)	(Tobacco)		
	(Solanum tuberosum)	(Potato)		
	(Petunia x hybrida hort. ex E. Vilm.)	(Petunia)	4x BAC of Petunia hybrida 7984 available from Clemson genome center (www.genome.clemson.edu)	
Total				http://www.nal.usda.gov/pgdic/Map_proj/
Brassicaceae	*Brassica oleracea* L. var. *italica*	Broccoli		http://res.agr.ca/ecorc/cwmt/crucifer/traits/index.htm http://geneous.cit.cornell.edu/cabbage/aboutcab.html

	Brassica oleracea L. var. *capitata*	Cabbage		
	Brassica rapa	Chinese Cabbage		
	Brassica oleracea L. var. *botrytis*	Cauliflower		
	Raphanus sativus var. *niger*	Daikon		
	(Brassica napus)	(Oilseed rape)		http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=brassicadb
		Arabidopsis	12x and 6x BACs on Columbia strain available from Clemson genome center (www.genome.clemson.edu)	http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=agr
Total				http://www.nal.usda.gov/pgdic/Map_proj/
Umbelliferae	*Daucus carota*	Carrot		
Compositae	*Lactuca sativa*	Lettuce		
	Helianthus annuus	(Sunflower)		
Total				

Chenopodiaceae	*Spinacia oleracea*	Spinach		
	(Beta vulgaris)	(Sugar Beet)		
Total				
Leguminosae	*Phaseolus vulgaris*	Bean	4.3x BAC available from Clemson genome center (www.genome.clemson.edu)	http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=beangenes
	Pisum sativum	Pea		
	(Glycine max)	(Soybean)	7.5x and 7.9x BACs available from Clemson genome center (www.genome.clemson.edu)	http://ars-genome.cornell.edu/cgi-bin/WebAce/webace?db=soybase
Total			http://www.nal.usda.gov/pgdic/Map_proj/	
Gramineae	*Zea mays*	Sweet Corn	Novartis BACs for Mo17 and B73 have been donated to Clemson Genome Center (www.genome.clemson.edu)	
	(Zea mays)	(Field Corn)		http://www.agron.missouri.edu/mnl/

Total			http://www.nal.usda.gov/pgdic/Map_proj/	
Liliaceae	*Allium cepa*	Onion		
		Leek		
		(Garlic)		
		(Asparagus)		
Total			http://www.nal.usda.gov/pgdic/Map_proj/	

Preferred forage and turf grass nucleic acid sources for use in the methods of the invention include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop. Preferably, the nucleic acid sources are crop plants and in particular cereals (for example, corn, alfalfa, sunflower, rice, *Brassica*, canola, soybean, barley, soybean, sugarbeet, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, etc.), and even more preferably corn and soybean.

According to one embodiment, the present invention is directed to a nucleic acid molecule comprising a nucleotide sequence isolated from any plant which encodes a polypeptide having at least 70% amino acid sequence identity to a polypeptide comprising SEQ ID NOs. 1-36 or a promoter for said nucleotide sequence. Thus, based on the nucleic acid sequences encoding the polypeptide of the present invention, orthologs of those sequences may be identified or isolated from the genome of any desired organism, preferably from another plant, according to well known techniques based on their sequence similarity to the coding sequences, e.g., hybridization, PCR or computer generated sequence comparisons. For example, all or a portion of a particular plant sequence is used as a probe that selectively hybridizes to other gene sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen source organism. Further, suitable genomic and cDNA libraries may be prepared from any cell or tissue of an organism. Such techniques include hybridization screening of plated DNA libraries (either plaques or colonies; see, e.g., Sambrook et al., 1989) and amplification by PCR using oligonucleotide primers preferably corresponding to sequence domains conserved among

related polypeptide or subsequences of the nucleotide sequences provided herein (see, e.g., Innis et al., 1990). These methods are particularly well suited to the isolation of gene sequences from organisms closely related to the organism from which the probe sequence is derived. The application of these methods using the coding sequences as probes is well suited for the isolation of gene sequences from any source organism, preferably other plant species. In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any plant of interest. Methods for designing PCR primers and PCR cloning are generally known in the art as discussed hereinabove.

In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen organism. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labeled with a detectable group such as ^{32}P, or any other detectable marker. Thus, for example, probes for hybridization can be made by labeling synthetic oligonucleotides based on the sequence of the invention. Methods for preparation of probes for hybridization and for construction of cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook et al. (1989). In general, sequences that hybridize to the sequences disclosed herein will have at least 40% to 50%, about 60% to 70% and even about 80% 85%, 90%, 95% to 98% or more identity with the disclosed sequences. That is, the sequence similarity of sequences may range, sharing at least about 40% to 50%, about 60% to 70%, and even about 80%, 85%, 90%, 95% to 98% sequence similarity.

The nucleic acid molecules of the invention can also be identified by, for example, a search of known databases for genes encoding polypeptides having a specified amino acid sequence identity. Methods of alignment of sequences for comparison are well known in the art and are described hereinabove.

Eleven proteins and their orthologs, of the invention, and their sequences are listed in the Sequence Listing, and are further described. Globulin-1 s allele precursor and Globulin-2 precursor are embryo storage protein. Reference describing these two gene family include (1) Biochem Genet 1989 Apr;27(3-4):239-51 Characterization of embryo globulins encoded by the maize Glb genes. Kriz AL. and (2) Characterization of the maize Globulin-2 gene and analysis of two null alleles. Kriz AL, Wallace NH Biochem Genet 1991 Jun;29(5-6):241-54,

which are incorporated by reference. Oleosin is a proteins associated with seed oil body. It is also an ABA inducible protein, further described in (1) Frandsen GI, Mundy J, Tzen JT. Oil bodies and their associated proteins, oleosin and caleosin. Physiol Plant. 2001 Jul;112(3):301-307 and (2) Crowe AJ, Abenes M, Plant A, Moloney MM. The seed-specific transactivator, ABI3, induces oleosin gene expression. Plant Sci. 2000 Feb 21;151(2):171-181. which are incorporated by reference. The 17.2 KD heat shock protein is a stress induced protein, and are further described in Heat shock proteins, Martin E Feder and Gretchen E. Hofmann 1999 Molecular cahperones, and the stress responses : Evolutionary and ecological physiology. Annu Rev Physiol 61:243-282, which is incorporated by reference. Glucose and ribitol dehydrogenase homolog is an embryo-specifc protein, up-regulated during seed maturation, and is further described in Alexander R, Alamillo JM, Salamini F, Bartels D Planta 1994;192(4):519-25 A novel embryo-specific barley cDNA clone encodes a protein with homologies to bacterial glucose and ribitol dehydrogenase, which is incorporated by reference. ZMPK1 precursor is a putative receptor protein kinase related to stress response, and further described in Zhang R, Walker JC (1993) Structure and expression of the *S* locus-related genes of maize. Plant Mol Biol 21: 1171-1174, whichis incorporated herein by reference. Glutathione S-transferase is an enzyme for transferring glutathione to many substrates, including cytotoxic substances, and is further described in Marrs K.A. The function and regulation of glutathione S- transferases in plants, Annu. Rev. Plant Physiol. Plant. Mol. Biol. 1996 , Vol. 47: 127-158, which is incorporated herein by reference. Thioredoxin dependent peroxidase is an enzyme involved in Antioxidative Defence System, further described in RB Van Huystee, Some Molecular Aspects Of Plant Peroxidase Biosynthetic Studies, Annu. Rev. Plant Physiol. Plant. Mol. Biol. 1987 , Vol. 38: 205-219, which is incorporated herein by reference. RAB28 protein is an ABA induced gene, in late embryogenesis in repsponse to water stress, further described in Pla M, Gomez J, Goday A, Pagès M Regulation of the abscisic acid-responsive gene rab28 in maize viviparous mutants, Mol Gen Genet. 1991 Dec;230(3):394-400, which is incoproated herein by reference. Dehydrin dHN1 belongs to a group of proteins that are stress induced and involved in stress tolerance, further described in Zeevaart JAD, Creelman RA 1988 Metabolisms and physiology of abscisic acids. Annu Rev Plant Physiol Mol Biol 39:439-473, which is incoproated herein by reference. Hydroxymethylglutaryl-CoA reductase which is a key enzyme involved in catalyzing an early reaction unique to isoprenoid biosynthesis., further described in (1) Kato-Emori S, Higashi K, Hosoya K, Kobayashi T, Ezura H. Cloning and characterization of the gene encoding 3-

hydroxy-3-methylglutaryl coenzyme A reductase in melon (Cucumis melo L. reticulatus). Mol Genet Genomics. 2001 Mar;265(1):135-42, and (2) Caelles C, Ferrer A, Balcells L, Hegardt FG, Boronat A. Isolation and structural characterization of a cDNA encoding Arabidopsis thaliana 3-hydroxy-3-methylglutaryl coenzyme A reductase. Plant Mol Biol. 1989 Dec;13(6):627-38, which are incorporated herein by reference.

As described in the Examples, a proteomics approach was used to identify genes that were differentially expressed in high protein corn lines. Over 150 differentially expressed protein spots were identified and analyzed as described in the experimental conditions. Provded herein are genes, their proteins, as shown in the Sequence Listing, and their orthologs, applicable to the methods and compositons of the present invention.

The nature of the candidate genes and their potential roles in contributing to the high protein phenotype is presented, however, the inventor is by no means to be limited by any one proposed mechanism. Among the proteins positively annotated, two groups of proteins are outstanding and are believed to be intimately related to the corn high protein phenotype: one group represents the seed storage proteins, including globulins and oleosin. These major seed protein storage components are believed to directly contribute to the high protein phenotype. A second group of proteins can be roughly characterized as stress related proteins, such as the heat shock protein, dehydrin and a regulatory gene rab28 involved in ABA related stress response.

These two groups of proteins or genes are part of the same mechanism that contributes to the high protein yield. For example, ABI3, which is a Arabidopsis gene that involved in seed storage protein biosynthesis, is reported as a key player in temperature stress. One hypothesis for this relationship is that plants that are more stress resistant, such as more heat tolerant, will grow better, therefore have more grain yield including grain protein. Based on this hypothesis we postulate that other global regulators that have a significant impact on stress related response can be used to manipulate seed protein content, such as the ABI3 from Arabidopsis and its homolog in rice, and in particular the genes of the present invention.

Additional data presented herein support these relationships and uses of these protein and genes as to modulate high protein trait. Details are provided in th Examples section. Seed storage proteins directly contribute to the high protein phenotype. Antibodies developed against the two embryo specific globulin proteins, glb1 and 2 (see Sequence Listing) were used to deterimine that their protein levels in the high protein inbred lines are significantly higher than in the control line (1.5-2 fold). Gene expression patterns of selected genes were studied

82

in rice gene expression profiling experiments. All three genes studied, the heat shock 17 gene, the glucose dehydrogenase gene and dehydrin gene (see Sequence Listing) were up-regulated in the time cource of rice seed development, coinciding with the development phase during which seed maturation occurs, indicating they play an important role during grain filling. (see Table 2) We also took a genetic approach to study the segregation of some of the genes with the high protein phenotype. A population of hybrid corn lines were used, derived from high protein line Wil500 as the high protein source. Table 3 demonstrates the correlation between the high protein trait and the HS 17 gene expression level.

Both seed protein content and protein quality can be changed by using these genes. In one embodiment a transgenic approach to over express or down regulate these genes in the seeds can be employed to increase grain protein content. Meanwhile, as is evident from their amino acid sequences (see Sequence Listing), some of these genes and proteins are biased to a special amino acid profile, over expressing of these proteins in seed can change the seed protein property. In particular, nutrionally enhanced seed, more complete or elevated in one or more amino acids, can be obtained. For example, poultry, like swine, have a specific amino acid that, if deficient, will reduce the animal's performance on the feed. For poultry, the limiting amino acid is methionine, while for swine the limiting amino acid is lysine. Accordingly, the present invention can provide feed that contains, or can be formulated to contain, an increase in the amount of methionine (or lysine for swine) and general protein to keep the desired protein gross energy ratio in the diet.

The genes disclosed herein are useful as genetic markers in marker assisted breeding programs to select high protein lines during plant breeding. For example, antibodies to the proteins of the invention (for use in ELISA for example) or DNA markers that are linked to these genes, or the genes themselves, find use to predict genomic profile and thus trait outcome of siblings in breeding practices.

Furthermore, the genes and proteins find use in production of effective protein production factories. For example, down regulation of one or more of the proteins can provide a seed that is reduced in these proteins thus allowing increased cellular resources for expression of industrially or therapeutically important polypeptides. This can best be done by inducible regulation of the one or more genes. In one embodiment reduction of storage protein content is achieved by anti-sense, for example RNAi methods. A two component Gal4/C1 system can be used to provide an inducible system. Two components in separate inbred lines are inactive, but create hybrids in which gene mosulation is activated to create plants with

lower protein yield. This methd also finds use to create low protein lines that are less allergenic.

Further details for use of these genes and proteins and their orthologs are presented herein.

II. Expression Cassettes of the Invention

The present invention also encompasses expression cassettes, preferably in the form of a recombinant vectors comprising the nucleic acid sequences of the invention. In such vectors, the expression cassette comprises regulatory elements for expression of the nucleotide sequences in a host cell capable of expressing the nucleotide sequences. Such regulatory elements usually comprise promoter and termination signals and preferably also comprise elements allowing efficient translation of polypeptides encoded by the nucleic acid sequences of the present invention. For efficient initiation of translation, sequences adjacent to the initiating methionine may require modification. For example, they can be modified by the inclusion of sequences known to be effective in plants. Joshi (1987) has suggested an appropriate consensus for plants and Clontech suggests a further consensus translation initiator (1993/1994 catalog, page 210). These consensuses are suitable for use with the nucleotide sequences of this invention. The sequences are incorporated into constructions comprising the nucleotide sequences, up to and including the ATG (whilst leaving the second amino acid unmodified), or alternatively up to and including the GTC subsequent to the ATG (with the possibility of modifying the second amino acid of the transgene).

Vectors comprising the nucleic acid sequences are usually capable of replication in particular host cells, e.g., as extrachromosomal molecules, and are therefore used to amplify the nucleic acid sequences of this invention in the host cells. In a preferred embodiment, host cells for such vectors are plant cells.

A. Promoters and Enhancers

Expression of the nucleotide sequences in transgenic plants is driven by promoters shown to be functional in plants. The choice of promoter will vary depending on the temporal and spatial requirements for expression, and also depending on the target species. In many cases, expression in multiple tissues is desirable. Although many promoters from dicotyledons have been shown to be operational in monocotyledons and *vice versa*, ideally dicotyledonous promoters are selected for expression in dicotyledons, and monocotyledonous promoters for expression in monocotyledons. However, there is no restriction to the provenance of selected

promoters; it is sufficient that they are operational in driving the expression of the nucleotide sequences in the desired cell.

These promoters include, but are not limited to, constitutive, inducible, temporally regulated, developmentally regulated, chemically regulated, stress-responsive, tissue-preferred and tissue-specific promoters. Promoter sequences are known to be strong or weak. A strong promoter provides for a high level of gene expression, whereas a weak promoter provides for a very low level of gene expression. An inducible promoter is a promoter that provides for the turning on and off of gene expression in response to an exogenously added agent, or to an environmental or developmental stimulus. A bacterial promoter such as the P_{tac} promoter can be induced to varying levels of gene expression depending on the level of isothiopropylgalactoside added to the transformed bacterial cells. An isolated promoter sequence that is a strong promoter for heterologous nucleic acid is advantageous because it provides for a sufficient level of gene expression to allow for easy detection and selection of transformed cells and provides for a high level of gene expression when desired.

Preferred promoters that are expressed constitutively include promoters from genes encoding actin or ubiquitin and the CaMV 35S and 19S promoters. The nucleotide sequences of this invention can also be expressed under the regulation of promoters that are chemically regulated. This enables the nucleic acid sequence or encoded polypeptide to be synthesized only when the crop plants are treated with the inducing chemicals. Preferred technology for chemical induction of gene expression is detailed in the published application EP 0 332 104 (to Ciba-Geigy) and U.S. Patent 5,614,395. A preferred promoter for chemical induction is the tobacco PR-1a promoter.

Tissue-specific or tissue-preferential promoters useful in the present invention. Also useful are promoters which confer seed-specific expression, such as those disclosed by Schernthaner et al. (1988); anther (tapetal) specific promoter B6 (Huffman et al.); and pistil-specific promoters such as a modified S13 promoter (Dzelkalns et al., 1993).

Preferred tissue specific expression patterns include green tissue-specific, root-specific, stem-specific, and flower-specific. Promoters suitable for expression in green tissue include many which regulate genes involved in photosynthesis and many of these have been cloned from both monocotyledons and dicotyledons. A preferred promoter is the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula, 1989). A preferred promoter for root-specific expression is that described by de Framond (1991; EP 0 452 269 to

Ciba-Geigy). A preferred stem specific promoter is that described in U.S. Patent No. 5,625,136 (to Ciba-Geigy) and which drives expression of the maize *trpA* gene.

Other promoters which direct specific or enhanced expression in certain plant tissues will be known to those of skill in the art in light of the present disclosure. These include, for example, the *rbcS* promoter, specific for green tissue; the *ocs, nos*, and *mas* promoters which have higher activity in roots or wounded leaf tissue; a truncated (-90 to +8) 35S promoter which directs enhanced expression in roots, an tubulin gene that directs expression in roots and promoters derived from zein storage protein genes which direct expression in endosperm. It is particularly contemplated that one may advantageously use the 16 bp *ocs* enhancer element from the octopine synthase (*ocs*) gene (Bonchez et al., 1989), especially when present in multiple copies, to achieve enhanced expression in roots.

Preferred plant promoters include, but are not limited to, a promoter such as the CaMV 35S promoter, an enhanced 35S promoter or others such as CaMV 19S, *nos*, *Adh1*, sucrose synthase, ∀-tubulin, ubiquitin, actin, *cab*, PEPCase or those associated with the R gene complex. Further suitable promoters may include the U2 and U5 snRNA promoters from maize, the promoter from alcohol dehydrogenase, the Z4 promoter from a gene encoding the Z4 22 kD zein protein, the Z10 promoter from a gene encoding a 10 kD zein protein, a Z27 promoter from a gene encoding a 27 kD zein protein, the A20 promoter from the gene encoding a 19 kD -zein protein, inducible promoters, such as the light inducible promoter derived from the pea rbcS gene and the actin promoter from rice; seed specific promoters, such as the phaseolin promoter from beans, may also be used. Other promoters useful in the practice of the invention are known to those of skill in the art.

Examples of tissue specific promoters which have been described include the lectin (Vodkin, 1983; Lindstrom et al., 1990,) corn alcohol dehydrogenase 1 (Vogel et al., 1992; Dennis et al., 1984), corn light harvesting complex (Simpson, 1985; Bansal et al., 1992), corn heat shock protein (Odell et al., 1985; Rochester et al., 1986), pea small subunit RuBP carboxylase (Poulsen et al., 1986; Cashmore et al., 1983), Ti plasmid mannopine synthase (Langridge et al., 1989), Ti plasmid nopaline synthase (Langridge et al., 1989), petunia chalcone isomerase (vanTunen et al., 1988), bean glycine rich protein 1 (Keller et al., 1989), truncated CaMV 35s (Odell et al., 1985), potato patatin (Wenzler et al., 1989), root cell (Yamamoto et al., 1990), maize zein (Reina et al., 1990; Kriz et al., 1987; Wandelt et al., 1989; Langridge et al., 1983; Reina et al., 1990), globulin-1 (Belanger et al., 1991), α-tubulin,

cab (Sullivan et al., 1989), PEPCase (Hudspeth & Grula, 1989), R gene complex-associated promoters (Chandler et al., 1989), and chalcone synthase promoters (Franken et al., 1991).

Inducible promoters that have been described include the ABA- and turgor-inducible promoters, the promoter of the auxin-binding protein gene (Schwob et al., 1993), the UDP glucose flavonoid glycosyl-transferase gene promoter (Ralston et al., 1988), the MPI proteinase inhibitor promoter (Cordero et al., 1994), and the glyceraldehyde-3-phosphate dehydrogenase gene promoter (Kohler et al., 1995; Quigley et al., 1989; Martinez et al., 1989).

Several tissue-specific regulated genes and/or promoters have been reported in plants. These include genes encoding the seed storage proteins (such as napin, cruciferin, beta-conglycinin, and phaseolin) zein or oil body proteins (such as oleosin), or genes involved in fatty acid biosynthesis (including acyl carrier protein, stearoyl-ACP desaturase. and fatty acid desaturases (fad 2-1)), and other genes expressed during embryo development (such as Bce4, see, for example. EP 255378 and Kridl et al., 1991). Particularly useful for seed-specific expression is the pea vicilin promoter (Czako et al., 1992). (See also U.S. Pat. No. 5,625,136, herein incorporated by reference.) Other useful promoters for expression in mature leaves are those that are switched on at the onset of senescence, such as the SAG promoter from Arabidopsis (Gan et al., 1995, 270 (5244), 1986-8).

A class of fruit-specific promoters expressed at or during antithesis through fruit development, at least until the beginning of ripening, is discussed in U.S. 4,943,674, the disclosure of which is hereby incorporated by reference. cDNA clones that are preferentially expressed in cotton fiber have been isolated (John et al., 1992). cDNA clones from tomato displaying differential expression during fruit development have been isolated and characterized (Mansson et al., 1985, Slater et al., 1985). The promoter for polygalacturonase gene is active in fruit ripening. The polygalacturonase gene is described in U.S. Patent No. 4,535,060, U.S. Patent No. 4,769,061, U.S. Patent No. 4,801,590, and U.S. Patent No. 5,107,065, which disclosures are incorporated herein by reference.

Other examples of tissue-specific promoters include those that direct expression in leaf cells following damage to the leaf (for example, from chewing insects), in tubers (for example, patatin gene promoter), and in fiber cells (an example of a developmentally-regulated fiber cell protein is E6 (John et al., 1992). The E6 gene is most active in fiber, although low levels of transcripts are found in leaf, ovule and flower.

The tissue-specificity of some "tissue-specific" promoters may not be absolute and may be tested by one skilled in the art using the diphtheria toxin sequence. One can also achieve

tissue-specific expression with "leaky" expression by a combination of different tissue-specific promoters (Beals et al., 1997). Other tissue-specific promoters can be isolated by one skilled in the art (see U.S. 5,589,379). Several inducible promoters ("gene switches") have been reported. Many are described in the review by Gatz (1996 and 1997). These include tetracycline repressor system, *Lac* repressor system, copper-inducible systems, salicylate-inducible systems (such as the PR1a system), glucocorticoid- (Aoyama, 1997) and ecdysome-inducible systems. Also included are the benzene sulphonamide- (U.S. Patent No. 5,364,780) and alcohol- (WO 97/06269 and WO 97/06268) inducible systems and glutathione S-transferase promoters. Other studies have focused on genes inducibly regulated in response to environmental stress or stimuli such as increased salinity, drought, pathogen and wounding. (Graham et al., 1985; Graham et al., 1985, Smith et al., 1986). Accumulation of metallocarboxypeptidase-inhibitor protein has been reported in leaves of wounded potato plants (Graham et al., 1981). Other plant genes have been reported to be induced methyl jasmonate, elicitors, heat-shock, anaerobic stress, or herbicide safeners.

Frequently it is desirable to have continuous or inducible expression of a DNA sequence throughout the cells of an organism in a tissue-independent manner. For example, increased resistance of a plant to infection by soil- and air borne pathogens might be accomplished by genetic manipulation of the plant's genome to comprise a continuous promoter operably linked to a heterologous or homologous pathogen-resistance gene such that pathogen-resistance proteins are continuously expressed throughout the plant's tissues.

Alternatively, it might be desirable to inhibit expression of a native DNA sequence within a plant's tissues to achieve a desired phenotype. In this case, such inhibition might be accomplished with transformation of the plant to comprise a constitutive, tissue-independent promoter operably linked to an antisense nucleotide sequence, such that constitutive expression of the antisense sequence produces an RNA transcript that interferes with translation of the mRNA of the native DNA sequence.

Other elements include those that can be regulated by endogenous or exogenous agents, e.g., by DNA binding proteins such as zinc finger proteins, including naturally occurring zinc finger proteins or chimeric zinc finger proteins (see, e.g., U.S. Patent No. 5,789,538, WO 99/48909; WO 99/45132; WO 98/53060; WO 98/53057; WO 98/53058; WO 00/23464; WO 95/19431; and WO 98/54311) or myb-like transcription factors. For example, a chimeric zinc finger protein may include amino acid sequences which bind to a specific DNA sequence (the

zinc finger) and amino acid sequences that activate (e.g., GAL 4 sequences) or repress the transcription of the sequences linked to the specific DNA sequence.

B. 5' and 3' Sequences

In addition to promoters, a variety of 3 transcriptional terminators are also available for use in the present invention. Transcriptional terminators are responsible for the termination of transcription and correct mRNA polyadenylation. The 3N nontranslated regulatory DNA sequence preferably includes from about 50 to about 1,000, more preferably about 100 to about 1,000, nucleotide base pairs and contains plant transcriptional and translational termination sequences. Appropriate transcriptional terminators and those which are known to function in plants include the CaMV 35S terminator, the tml terminator, the nopaline synthase terminator, the pea rbcS E9 terminator, the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*, and the 3N end of the protease inhibitor I or II genes from potato or tomato, although other 3N elements known to those of skill in the art can also be employed.

The 5N regulatory region of the expression cassette may also include other enhancing sequences. Numerous sequences have been found to enhance gene expression in transgenic plants. These include sequences which have been shown to enhance expression such as intron sequences (e.g., from *Adh1, bronze1* or the sucrose synthase intron) and viral leader sequences (e.g., from TMV, MCMV and AMV). For example, a number of non-translated leader sequences derived from viruses are known to enhance expression. Specifically, leader sequences from Tobacco Mosaic Virus (TMV), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (e.g., Gallie et al., 1987; Skuzeski et al., 1990). Other leaders known in the art include but are not limited to: Picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5 noncoding region) (Elroy-Stein et al., 1989); Potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Allison et al., 1986); MDMV leader (Maize Dwarf Mosaic Virus); Human immunoglobulin heavy-chain binding protein (BiP) leader, (Macejak et al., 1991); Untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling et al., 1987; Tobacco mosaic virus leader (TMV), (Gallie et al., 1989; and Maize Chlorotic Mottle Virus leader (MCMV) (Lommel et al., 1991. See also, Della-Cioppa et al., 1987.

C. Targeting Sequences

It may be preferable to target expression of the nucleotide sequences of the present invention to different cellular localizations in the plant. In some cases, localization in the cytosol may be desirable, whereas in other cases, localization in some subcellular organelle, e.g., the nucleus, may be preferred. Subcellular localization of transgene encoded enzymes is undertaken using techniques well known in the art. Typically, the DNA encoding the target peptide from a known organelle-targeted gene product is manipulated and fused upstream of the nucleotide sequence. Many such target sequences are known for the chloroplast and their functioning in heterologous constructions has been shown. The expression of the nucleotide sequences of the present invention is also targeted to the endoplasmic reticulum or to the vacuoles of the host cells. Techniques to achieve this are well-known in the art.

D. Marker Genes

In order to improve the ability to identify transformants, one may desire to employ a selectable or screenable marker gene as, or in addition to, the preselected nucleic acid sequence or segment. "Marker genes" are genes that impart a distinct phenotype to cells expressing the marker gene and thus allow such transformed cells to be distinguished from cells that do not have the marker. Such genes may encode either a selectable or screenable marker, depending on whether the marker confers a trait which one can 'select' for by chemical means, i.e., through the use of a selective agent (e.g., a herbicide, antibiotic, or the like), or whether it is simply a trait that one can identify through observation or testing, i.e., by 'screening' (e.g., the R-locus trait). Of course, many examples of suitable marker genes are known to the art and can be employed in the practice of the invention.

Included within the terms selectable or screenable marker genes are also genes which encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected by their catalytic activity. Secretable proteins fall into a number of classes, including small, diffusible proteins detectable, e.g., by ELISA; small active enzymes detectable in extracellular solution (e.g., ∀-amylase, ∃-lactamase, phosphinothricin acetyltransferase); and proteins that are inserted or trapped in the cell wall (e.g., proteins that include a leader sequence such as that found in the expression unit of extensin or tobacco PR-S).

With regard to selectable-secretable markers, the use of a gene that encodes a polypeptide that becomes sequestered in the cell wall, and which polypeptide includes a

unique epitope is considered to be particularly advantageous. Such a secreted antigen marker would ideally employ an epitope sequence that would provide low background in plant tissue, a promoter-leader sequence that would impart efficient expression and targeting across the plasma membrane, and would produce protein that is bound in the cell wall and yet accessible to antibodies. A normally secreted wall protein modified to include a unique epitope would satisfy all such requirements.

Numerous other possible selectable and/or screenable marker genes will be apparent to those of skill in the art in addition to the one set forth herein below. Therefore, it will be understood that the following discussion is exemplary rather than exhaustive. In light of the techniques disclosed herein and the general recombinant techniques which are known in the art, the present invention renders possible the introduction of any gene, including marker genes, into a recipient cell to generate a transformed plant cell, e.g., a monocot cell.

Possible selectable markers for use in connection with the present invention include, but are not limited to, a *neo* gene, which codes for kanamycin resistance and can be selected for using kanamycin, G418, a gene encoding resistance to bleomycin, and the like; a *bar* gene which codes for bialaphos resistance; a gene which encodes an altered EPSP synthase protein thus conferring glyphosate resistance; a nitrilase gene such as *bxn* from *Klebsiella ozaenae* which confers resistance to bromoxynil; a mutant acetolactate synthase gene (ALS) which confers resistance to imidazolinone, sulfonylurea or other ALS-inhibiting chemicals (European Patent Application 154,204, 1985); a methotrexate-resistant DHFR gene; a dalapon dehalogenase gene that confers resistance to the herbicide dalapon; or a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan. Where a mutant EPSP synthase gene is employed, additional benefit may be realized through the incorporation of a suitable chloroplast transit peptide, CTP (European Patent Application 0 218 571; 1987).

An illustrative embodiment of a selectable marker gene capable of being used in systems to select transformants is the genes that encode the enzyme phosphinothricin acetyltransferase, such as the *bar* gene from *Streptomyces hygroscopicus* or the *pat* gene from *Streptomyces viridochromogenes* (U.S. Patent No. 5,550,318). The enzyme phosphinothricin acetyltransferase (PAT) inactivates the active ingredient in the herbicide bialaphos, phosphinothricin (PPT). PPT inhibits glutamine synthetase, causing rapid accumulation of ammonia and cell death. The success in using this selective system in conjunction with monocots was particularly surprising because of the major difficulties which have been reported in transformation of cereals.

Screenable markers that may be employed include, but are not limited to, a ∃-glucuronidase or *uid*A gene (GUS) which encodes an enzyme for which various chromogenic substrates are known; an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues; a -lactamase gene, which encodes an enzyme for which various chromogenic substrates are known (e.g., PADAC, a chromogenic cephalosporin); a *xyl*E gene which encodes a catechol dioxygenase that can convert chromogenic catechols; an -amylase gene; a tyrosinase gene which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to form the easily detectable compound melanin; a -galactosidase gene, which encodes an enzyme for which there are chromogenic substrates; a luciferase (*lux*) gene, which allows for bioluminescence detection; or an aequorin gene, which may be employed in calcium-sensitive bioluminescence detection, or a green fluorescent protein.

Genes from the maize R gene complex are contemplated to be particularly useful as screenable markers. The R gene complex in maize encodes a protein that acts to regulate the production of anthocyanin pigments in most seed and plant tissue. Maize strains can have one, or as many as four, R alleles which combine to regulate pigmentation in a developmental and tissue specific manner. A gene from the R gene complex was applied to maize transformation, because the expression of this gene in transformed cells does not harm the cells. Thus, an R gene introduced into such cells will cause the expression of a red pigment and, if stably incorporated, can be visually scored as a red sector. If a maize line carries dominant alleles for genes encoding the enzymatic intermediates in the anthocyanin biosynthetic pathway (C2, A1, A2, Bz1 and Bz2), but carries a recessive allele at the R locus, transformation of any cell from that line with R will result in red pigment formation. Exemplary lines include Wisconsin 22 which contains the rg-Stadler allele and TR112, a K55 derivative which is r-g, b, Pl. Alternatively any genotype of maize can be utilized if the C1 and R alleles are introduced together.

A further screenable marker contemplated for use in the present invention is firefly luciferase, encoded by the *lux* gene. The presence of the *lux* gene in transformed cells may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry. It is also envisioned that this system may be developed for populational screening for bioluminescence, such as on tissue culture plates, or even for whole plant screening.

E. Other Sequences

A vector of the invention can also further comprise plasmid DNA. Plasmid vectors include additional DNA sequences that provide for easy selection, amplification, and transformation of the expression cassette in prokaryotic and eukaryotic cells, e.g., pUC-derived vectors such as pUC8, pUC9, pUC18, pUC19, pUC23, pUC119, and pUC120, pSK-derived vectors, pGEM-derived vectors, pSP-derived vectors, or pBS-derived vectors. The additional DNA sequences include origins of replication to provide for autonomous replication of the vector, additional selectable marker genes, preferably encoding antibiotic or herbicide resistance, unique multiple cloning sites providing for multiple sites to insert DNA sequences or genes encoded in the expression cassette, and sequences that enhance transformation of prokaryotic and eukaryotic cells.

Another vector that is useful for expression in both plant and prokaryotic cells is the binary Ti plasmid (as disclosed in Schilperoort et al., U.S. Patent No. 4,940,838) as exemplified by vector pGA582. This binary Ti plasmid vector has been previously characterized by An, cited *supra*. This binary Ti vector can be replicated in prokaryotic bacteria such as *E. coli* and *Agrobacterium*. The *Agrobacterium* plasmid vectors can be used to transfer the expression cassette to dicot plant cells, and under certain conditions to monocot cells, such as rice cells. The binary Ti vectors preferably include the nopaline T DNA right and left borders to provide for efficient plant cell transformation, a selectable marker gene, unique multiple cloning sites in the T border regions, the *col*E1 replication of origin and a wide host range replicon. The binary Ti vectors carrying an expression cassette of the invention can be used to transform both prokaryotic and eukaryotic cells, but is preferably used to transform dicot plant cells.

Virtually any DNA may be used for delivery to recipient cells to ultimately produce fertile transgenic plants in accordance with the present invention. For example, DNA segments in the form of vectors and plasmids, or linear DNA fragments, in some instance containing only the DNA element to be expressed in the plant, and the like, may be employed.

Vectors, plasmids, cosmids, YACs (yeast artificial chromosomes) and DNA segments for use in transforming such cells will, of course, generally comprise the cDNA, gene or genes which one desires to introduce into the cells. These DNA constructs can further include structures such as promoters, enhancers, polylinkers, or even regulatory genes as desired. The DNA segment or gene chosen for cellular introduction will often encode a protein which will be expressed in the resultant recombinant cells, such as will result in a screenable or selectable trait and/or which will impart an improved phenotype to the regenerated plant. However, this

may not always be the case, and the present invention also encompasses transgenic plants incorporating non-expressed transgenes.

III. Transformation

The expression cassettes of the present invention can be introduced into a host cell, e.g., a plant cell, in a number of art-recognized ways. Those skilled in the art will appreciate that the choice of method might depend on the type of cell, e.g., monocotyledonous or dicotyledonous, targeted for transformation. Vectors which may be used to transform plant tissue with the expression cassettes of the present invention include both *Agrobacterium* vectors and ballistic vectors, as well as vectors suitable for DNA-mediated transformation, e.g., direct uptake or via electroporation. However, cells other than plant cells may be transformed with the expression cassettes of the invention.

Suitable methods of transforming plant cells include, but are not limited to, microinjection (Crossway et al., 1986), direct DNA transfer to plant cells by PEG precipitation; liposomes; electroporation (Riggs et al., 1986, *Agrobacterium*-mediated transformation (Hinchee et al., 1988), direct gene transfer (Paszkowski et al., 1984), and ballistic particle acceleration using devices available from Agracetus, Inc., Madison, Wis. and BioRad, Hercules, Calif. (see, for example, Sanford et al., U.S. Pat. No. 4,945,050; and McCabe et al., 1988). Also see, Weissinger et al., 1988; Sanford et al., 1987 (onion); Christou et al., 1988 (soybean); McCabe et al., 1988 (soybean); Datta et al., 1990 (rice); Klein et al., 1988 (maize); Klein et al., 1988 (maize); Klein et al., 1988 (maize); Fromm et al., 1990 (maize); and Gordon-Kamm et al., 1990 (maize); Svab et al., 1990 (tobacco chloroplast); Koziel et al., 1993 (maize); Shimamoto et al., 1989 (rice); Christou et al., 1991 (rice); European Patent Application EP 0 332 581 (orchardgrass and other Pooideae); Vasil et al., 1993 (wheat); Weeks et al., 1993 (wheat).

In one embodiment, a nucleotide sequence of the present invention is directly transformed into the plastid genome. Plastid transformation technology is extensively described in U.S. Patent Nos. 5,451,513, 5,545,817, and 5,545,818, in PCT application no. WO 95/16783, and in McBride et al., 1994. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the gene of interest into a suitable target tissue, e.g., using biolistics or protoplast transformation (e.g., calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate orthologous recombination with the

plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab et al., 1990; Staub et al., 1992). This resulted in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub et al., 1993). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial *aadA* gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3N-adenyltransferase (Svab et al., 1993). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention. Typically, approximately 15-20 cell division cycles following transformation are required to reach a homoplastidic state. Plastid expression, in which genes are inserted by orthologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclear-expressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein. In a preferred embodiment, a nucleotide sequence of the present invention is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants homoplastic for plastid genomes containing a nucleotide sequence of the present invention are obtained, and are preferentially capable of high expression of the nucleotide sequence.

Agrobacterium tumefaciens cells containing a vector comprising an expression cassette of the present invention, wherein the vector comprises a Ti plasmid, are useful in methods of making transformed plants. Plant cells are infected with an *Agrobacterium tumefaciens* as described above to produce a transformed plant cell, and then a plant is regenerated from the transformed plant cell. Numerous *Agrobacterium* vector systems useful in carrying out the present invention are known. For example, U.S. Pat. No. 4,459,355 discloses a method for transforming susceptible plants, including dicots, with an *Agrobacterium* strain containing the Ti plasmid. The transformation of woody plants with an *Agrobacterium* vector is disclosed in U.S. Patent No. 4,795,855. Further, U.S. Patent No. 4,940,838 to Schilperoort et al. discloses a binary *Agrobacterium* vector (i.e., one in which the *Agrobacterium* contains one plasmid having the vir region of a Ti plasmid but no T region, and a second plasmid having a T region but no vir region) useful in carrying out the present invention.

It is particularly preferred to use the binary type vectors of Ti and Ri plasmids of *Agrobacterium spp.* Ti-derived vectors transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants, such as soybean, cotton, rape, tobacco, and rice (Pacciotti et al., 1985: Byrne et al., 1987; Sukhapinda et al., 1987; Lorz et al., 1985; Potrykus, 1985; Park et al., 1985: Hiei et al., 1994. The use of T-DNA to transform plant cells has received extensive study and is amply described (EP 120516; Hoekema, 1985; Knauf, et al., 1983; and An. et al., 1985. For introduction into plants, the nucleotide sequences of the invention can be inserted into binary vectors as described in the examples.

Transformation of plants can be undertaken with a single DNA molecule or multiple DNA molecules (i.e., co-transformation), and both these techniques are suitable for use with the expression cassettes of the present invention. Numerous transformation vectors are available for plant transformation, and the expression cassettes of this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the preferred transformation technique and the target species for transformation.

Preferred plant cells for transformation include, but are not limited to, cells from plant such as corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers; duckweed (*Lemna*, see WO 00/07210, which includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as follows: genus *Lemna (L. aequinoctialis, L. disperma, L. ecuadoriensis, L. gibba, L. japonica, L. minor, L. miniscula, L. obscura, L. perpusilla, L. tenera, L.* trisulca, L. turionifera, L. valdiviana); genus *Spirodela (S. intermedia, S. polyrrhiza,*

S.punctata); genus *Woffia (Wa. angusta, Wa. arrhiza, Wa. australina, Wa. borealis, Wa. brasiliensis, Wa. columbiana, Wa. elongata, Wa. globosa, Wa. microscopica, Wa. neglecta)* and genus *Wofiella (W1. caudata, W1. denticulata, W1. gladiata, W1. hyalina, W1. lingulata, W1. repunda, W1. rotunda, and W1. neotropica*). Any other genera or species of *Lemnaceae*, if they exist, are also aspects of the present invention. *Lemna gibba, Lemna minor*, and *Lemna miniscula* are preferred, with *Lemna* minor and *Lemna miniscula* being most preferred. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of Lemnaceae - A Monograph Study. Geobatanischen Institut ETH, Stiftung Rubel, Zurich (1986)); vegetables including tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, caluliflower, Brassica, e.g.,

97

broccoli, cabbage, brussels sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, asparagus, and zucchini and ornamental plants include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia. Other vegetables are in Table 1.

Preferred forage and turf grass for use in the methods of the invention include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop.

Preferably, plants or cells to be transformed are crop plants and in particular cereals (for example, corn, alfalfa, sunflower, rice, *Brassica*, canola, soybean, barley, soybean, sugarbeet, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, and the like), and even more preferably rice, corn and soybean.

In a preferred embodiment, the transformed host cells are monocot or dicot cells, including, but not limited to, wheat, corn (maize), rice, oat, barley, millet, rye, rape and alfalfa, as well as asparagus, tomato, egg plant, apple, pear, quince, cherry, apricot, pepper, melon, lettuce, cauliflower, *Brassica*, e.g., broccoli, cabbage, brussels sprout, sugar beet, sugar cane, sweetcorn, onion, carrot, leek, cucumber, tobacco, aubergine, beet, broad bean, carrot, celery, chicory, cotton, radish, pumpkin, hemp, buckwheat, orchardgrass, creeping bent top, redtop, ryegrass, tobacco, turfgrass, tall fescue, cow pea, endive, gourd, grape, raspberry, chenopodium, blueberry, pineapple, avocado, mango, banana, groundnut, nectarine, papaya, garlic, pea, peach, peanut, pepper, pineapple, plum, potato, safflower, snap bean, spinach, squashes, strawberry, sunflower, sorghum, sweet potato, turnip, watermelon, legumes such as *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, and the like; and ornamental crops including Impatiens, Begonia, Petunia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Ageratum, Amaranthus, Anthirrhinum, Aquilegia, Chrysanthemum, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossis, Zinnia, and the like. More preferably, the transformed host cells are monocot cells such as maize, rice, wheat, barley, oats, and sorghum, which can be regenerated into a transgenic plant.

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a vector of the present invention. The term "organogenesis," as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers; the term "embryogenesis," as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

The choice of plant tissue source for transformation will depend on the nature of the host plant and the transformation protocol. Useful tissue sources include callus, suspension culture cells, protoplasts, leaf segments, stem segments, tassels, pollen, embryos, hypocotyls, tuber segments, meristematic regions, and the like. The tissue source is selected and transformed so that it retains the ability to regenerate whole, fertile plants following transformation, i.e., contains totipotent cells. Type I or Type II embryonic maize callus and immature embryos are preferred *Zea mays* tissue sources. Selection of tissue sources for transformation of monocots is described in PCT publication WO 95/06128.

For certain plant species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the nptII gene which confers resistance to kanamycin and related antibiotics (Messing & Vierra, 1982); Bevan et al., 1983), the *bar* gene which confers resistance to the herbicide phosphinothricin (White et al., 1990, Spencer et al., 1990), the *hph* gene which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann), and the *dhfr* gene, which confers resistance to methotrexate (Bourouis et al., 1983).

Thus, the present invention also provides a transformed (transgenic) plant cell, *in planta* or *ex planta*, including, but not limited to, a transformed plant cell from plants such as corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat

(*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense, Gossypium hirsutum*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Cofea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers; duckweed (*Lemna*, see WO 00/07210, which includes members of the family *Lemnaceae*. There are known four genera and 34 species of duckweed as follows: genus *Lemna (L. aequinoctialis, L. disperma, L. ecuadoriensis, L. gibba, L. japonica, L. minor, L. miniscula, L. obscura, L. perpusilla, L. tenera, L.* trisulca, L. turionifera, L. valdiviana); genus *Spirodela (S. intermedia, S. polyrrhiza, S.punctata);* genus *Woffia (Wa. angusta, Wa. arrhiza, Wa. australina, Wa. borealis, Wa. brasiliensis, Wa. columbiana, Wa. elongata, Wa. globosa, Wa. microscopica, Wa. neglecta)* and genus *Wofiella (W1. caudata, W1. denticulata, W1. gladiata, W1. hyalina, W1. lingulata, W1. repunda, W1. rotunda, and W1. neotropica*). Any other genera or species of *Lemnaceae*, if they exist, are also aspects of the present invention. *Lemna gibba, Lemna minor*, and *Lemna miniscula* are preferred, with *Lemna* minor and *Lemna miniscula* being most preferred. *Lemna* species can be classified using the taxonomic scheme described by Landolt, Biosystematic Investigation on the Family of Duckweeds: The family of Lemnaceae - A Monograph Study. Geobatanischen Institut ETH, Stiftung Rubel, Zurich (1986)); vegetables including tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as

100

silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g., lentil, and false indigo, *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radiata pine, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, chenopodium, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, caluliflower, Brassica, e.g., broccoli, cabbage, brussels sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, asparagus, and zucchini and ornamental plants include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrhinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossos, and Zinnia, as well as from vegetables including those described in Table 1.

In a preferred embodiment, the transformed plants, include, but are not limited to, transformed wheat, corn (maize), rice, oat, barley, millet, rye, rape and alfalfa, as well as asparagus, tomato, egg plant, apple, pear, quince, cherry, apricot, pepper, melon, lettuce, cauliflower, *Brassica*, e.g., broccoli, cabbage, brussels sprout, sugar beet, sugar cane, sweetcorn, onion, carrot, leek, cucumber, tobacco, aubergine, beet, broad bean, carrot, celery, chicory, cotton, radish, pumpkin, hemp, buckwheat, orchardgrass, creeping bent top, redtop, ryegrass, tobacco, turfgrass, tall fescue, cow pea, endive, gourd, grape, raspberry, chenopodium, blueberry, pineapple, avocado, mango, banana, groundnut, nectarine, papaya, garlic, pea, peach, peanut, pepper, pineapple, plum, potato, safflower, snap bean, spinach, squashes, strawberry, sunflower, sorghum, sweet potato, turnip, watermelon, legumes such as *Arachis*, e.g., peanuts, *Vicia,* e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, trifolium, *Phaseolus*, e.g., common bean and lima bean, *Pisum,* e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, Lotus, e.g., trefoil, lens, e.g.,

lentil, and false indigo, and the like; and ornamental crops including Impatiens, Begonia, Petunia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Ageratum, Amaranthus, Anthirrhinum, Aquilegia, Chrysanthemum, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossis, Zinnia, and the like. Preferably, the transformed plants are transformed monocot such as maize, rice, wheat, barley, oats, and sorghum.

IV. Identification of Transgenic Plants

To confirm the presence of the preselected nucleic acid segment(s) or "transgene(s)" in the regenerating plants, a variety of assays may be performed. Such assays include, for example, "molecular biological" assays well known to those of skill in the art, such as Southern and Northern blotting, *in situ* hybridization and nucleic acid-based amplification methods such as PCR or RT-PCR; "biochemical" assays, such as detecting the presence of a protein product, e.g., by immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as leaf or root assays; and also, by analyzing the phenotype of the whole regenerated plant, e.g., for disease or pest resistance.

DNA may be isolated from cell lines or any plant parts to determine the presence of the preselected nucleic acid segment through the use of techniques well known to those skilled in the art. Note that intact sequences will not always be present, presumably due to rearrangement or deletion of sequences in the cell.

The presence of nucleic acid elements introduced through the methods of this invention may be determined by polymerase chain reaction (PCR). Using this technique discreet fragments of nucleic acid are amplified and detected by gel electrophoresis. This type of analysis permits one to determine whether a preselected nucleic acid segment is present in a stable transformant, but does not prove integration of the introduced preselected nucleic acid segment into the host cell genome. In addition, it is not possible using PCR techniques to determine whether transformants have exogenous genes introduced into different sites in the genome, i.e., whether transformants are of independent origin. It is contemplated that using PCR techniques it would be possible to clone fragments of the host genomic DNA adjacent to an introduced preselected DNA segment.

Positive proof of DNA integration into the host genome and the independent identities of transformants may be determined using the technique of Southern hybridization. Using this technique specific DNA sequences that were introduced into the host genome and flanking

host DNA sequences can be identified. Hence the Southern hybridization pattern of a given transformant serves as an identifying characteristic of that transformant. In addition it is possible through Southern hybridization to demonstrate the presence of introduced preselected DNA segments in high molecular weight DNA, i.e., confirm that the introduced preselected DNA segment has been integrated into the host cell genome. The technique of Southern hybridization provides information that is obtained using PCR, e.g., the presence of a preselected DNA segment, but also demonstrates integration into the genome and characterizes each individual transformant.

It is contemplated that using the techniques of dot or slot blot hybridization which are modifications of Southern hybridization techniques one could obtain the same information that is derived from PCR, e.g., the presence of a preselected DNA segment.

Both PCR and Southern hybridization techniques can be used to demonstrate transmission of a preselected DNA segment to progeny. In most instances the characteristic Southern hybridization pattern for a given transformant will segregate in progeny as one or more Mendelian genes (Spencer et al., 1992); Laursen et al., 1994) indicating stable inheritance of the gene. The nonchimeric nature of the callus and the parental transformants (R_0) was suggested by germline transmission and the identical Southern blot hybridization patterns and intensities of the transforming DNA in callus, R_0 plants and R_1 progeny that segregated for the transformed gene.

Whereas DNA analysis techniques may be conducted using DNA isolated from any part of a plant, RNA may only be expressed in particular cells or tissue types and hence it will be necessary to prepare RNA for analysis from these tissues. PCR techniques may also be used for detection and quantitation of RNA produced from introduced preselected DNA segments. In this application of PCR it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique will demonstrate the presence of an RNA species and give information about the integrity of that RNA. The presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and will only demonstrate the presence or absence of an RNA species.

While Southern blotting and PCR may be used to detect the preselected DNA segment in question, they do not provide information as to whether the preselected DNA segment is being expressed. Expression may be evaluated by specifically identifying the protein products of the introduced preselected DNA segments or evaluating the phenotypic changes brought about by their expression.

Assays for the production and identification of specific proteins may make use of physical-chemical, structural, functional, or other properties of the proteins. Unique physical-chemical or structural properties allow the proteins to be separated and identified by electrophoretic procedures, such as native or denaturing gel electrophoresis or isoelectric focussing, or by chromatographic techniques such as ion exchange or gel exclusion chromatography. The unique structures of individual proteins offer opportunities for use of specific antibodies to detect their presence in formats such as an ELISA assay. Combinations of approaches may be employed with even greater specificity such as Western blotting in which antibodies are used to locate individual gene products that have been separated by electrophoretic techniques. Additional techniques may be employed to absolutely confirm the identity of the product of interest such as evaluation by amino acid sequencing following purification. Although these are among the most commonly employed, other procedures may be additionally used.

Assay procedures may also be used to identify the expression of proteins by their functionality, especially the ability of enzymes to catalyze specific chemical reactions involving specific substrates and products. These reactions may be followed by providing and quantifying the loss of substrates or the generation of products of the reactions by physical or chemical procedures. Examples are as varied as the enzyme to be analyzed.

Very frequently the expression of a gene product is determined by evaluating the phenotypic results of its expression. These assays also may take many forms including but not limited to analyzing changes in the chemical composition, morphology, or physiological properties of the plant. Morphological changes may include greater stature or thicker stalks. Most often changes in response of plants or plant parts to imposed treatments are evaluated under carefully controlled conditions termed bioassays.

V. Utility

Once an expression cassette of the invention has been transformed into a particular plant species, it may be propagated in that species or moved into other varieties of the same species,

particularly including commercial varieties, using traditional breeding techniques. Particularly preferred plants of the invention include the agronomically important crops listed above. The genetic properties engineered into the transgenic seeds and plants described above are passed on by sexual reproduction and can thus be maintained and propagated in progeny plants. The present invention also relates to a transgenic plant cell, tissue, organ, seed or plant part obtained from the transgenic plant. Also included within the invention are transgenic descendants of the plant as well as transgenic plant cells, tissues, organs, seeds and plant parts obtained from the descendants.

Preferably, the expression cassette in the transgenic plant is sexually transmitted. In one preferred embodiment, the coding sequence is sexually transmitted through a complete normal sexual cycle of the R0 plant to the R1 generation. Additionally preferred, the expression cassette is expressed in the cells, tissues, seeds or plant of a transgenic plant in an amount that is different than the amount in the cells, tissues, seeds or plant of a plant which only differs in that the expression cassette is absent.

The transgenic plants produced herein are thus expected to be useful for a variety of commercial and research purposes. Transgenic plants can be created for use in traditional agriculture to possess traits beneficial to the grower (e.g., agronomic traits such as resistance to water deficit, pest resistance, herbicide resistance or increased yield), beneficial to the consumer of the grain harvested from the plant (e.g., improved nutritive content in human food or animal feed), or beneficial to the food processor (e.g., improved processing traits). In such uses, the plants are generally grown for the use of their grain in human or animal foods. However, other parts of the plants, including stalks, husks, vegetative parts, and the like, may also have utility, including use as part of animal silage or for ornamental purposes. Often, chemical constituents (e.g., oils or starches) of maize and other crops are extracted for foods or industrial use and transgenic plants may be created which have enhanced or modified levels of such components.

Transgenic plants may also find use in the commercial manufacture of proteins or other molecules, where the molecule of interest is extracted or purified from plant parts, seeds, and the like. Cells or tissue from the plants may also be cultured, grown *in vitro*, or fermented to manufacture such molecules.

The transgenic plants may also be used in commercial breeding programs, or may be crossed or bred to plants of related crop species. Improvements encoded by the expression

cassette may be transferred, e.g., from maize cells to cells of other species, e.g., by protoplast fusion.

The transgenic plants may have many uses in research or breeding, including creation of new mutant plants through insertional mutagenesis, in order to identify beneficial mutants that might later be created by traditional mutation and selection. An example would be the introduction of a recombinant DNA sequence encoding a transposable element that may be used for generating genetic variation. The methods of the invention may also be used to create plants having unique "signature sequences" or other marker sequences which can be used to identify proprietary lines or varieties.

Thus, the transgenic plants and seeds according to the invention can be used in plant breeding which aims at the development of plants with improved properties conferred by the expression cassette, such as tolerance of viruses or other pests, or other stresses. The various breeding steps are characterized by well-defined human intervention such as selecting the lines to be crossed, directing pollination of the parental lines, or selecting appropriate descendant plants. Depending on the desired properties different breeding measures are taken. The relevant techniques are well known in the art and include but are not limited to hybridization, inbreeding, backcross breeding, multiline breeding, variety blend, interspecific hybridization, aneuploid techniques, etc. Hybridization techniques also include the sterilization of plants to yield male or female sterile plants by mechanical, chemical or biochemical means. Cross pollination of a male sterile plant with pollen of a different line assures that the genome of the male sterile but female fertile plant will uniformly obtain properties of both parental lines. Thus, the transgenic seeds and plants according to the invention can be used for the breeding of improved plant lines which for example increase the effectiveness of conventional methods such as herbicide or pesticide treatment or allow to dispense with said methods due to their modified genetic properties. Alternatively new crops with improved stress tolerance can be obtained which, due to their optimized genetic "equipment", yield harvested product of better quality than products which were not able to tolerate comparable adverse developmental conditions.

The invention will be further described by the following examples which is not intended to limit the scope of the invention.

Example 1

To identify novel genes that are associated with the high protein phenotype in selected lines of *Zea mays*, a differential analysis of four high protein lines and two control lines as well as a segregating population derived from a high protein line and a normal line was conducted. High protein corn lines (Wil500, Wil578, WIO465), control lines (WICY530 and LH59) and a segregating population derived from cross LH59XWIL578 (total of 53 lines) were obtained from Wilson Genetics. High protein maize refers to germplasm having elevated levels of protein in the seed, typically above 14.5 % in the whole kernel, above 17% in the embryo and above 13.5% in the endosperm (see Figure 1).

The following proteomic approaches were used: 1) extraction of proteins from tissue including total kernels, mature/developing embyros, root and leaf from 2 week old seedlings, optionally exposed to fertilizer; 2) two-dimensional (2-D) separation of proteins by size and charge using gels using isoelectric focusing (IEF) and SDS-PAGE at three pH ranges (e.g., pH 3-10, pH 4-7 and pH 7-10); 3) image analysis of silver stained gels to identify differentially expressed proteins (by visual inspection and PDQUEST software); 4) gel excision and trypsin digestion of selected protein spots; 5) analysis of resulting tryptic peptides using MALDI-TOF mass spectrometry; and 6) database searching using protein sequence information for protein identification using SEQUEST.

Materials and Methods

Sample preparation and gel electrophoresis. Embryos from mature corn kernel from both high protein lines and normal corn lines were cut out of the seeds and directly homogenized in a solution containing 7 uM urea, 2 uM thiourea, 0.5% Triton X-100 and 60 mM DTT. The first dimension for isoelectric focusing was carried out on a BioRad IPG system essentially as described by the manufacturer
using three pH gradient strips, pH 3-10, pH 4-7 and pH 5-8 for 45kvhr. Subsequent to loading the IEF strips on the second dimension, the IEF strips were re-equilibrated with a solution (2% SDS, 50 mM Tris, pH 6.9, 10% glycerol and 7 mM urea), and directly applied to a BioRad 8-16% gradient SDS-PAGE gel for electrophoresis. The resultant gels were stained with silver using a BioRad silver staining kit according to the manufacturer's recommendations. 2D PAGE profiles were laser scanned and comparative analyses were performed using PDQuest

software package (BioRad) . Only spots that were present/completely absent between normal and high protein lines were selected for further analysis. Protein spots were cut out of the gel either manually or using the BioRad spot cutter.

Trypsin digestion. Gel pieces were transferred to an eppendorf tube or a polypropylene 96 well plate. 100 ul acetonitrile was added to dehydrate the gel. After removing the acetonitrile by speed vacuum, the gels were contacted with 50 mM NH_4HCO_3 and trypsin at 10 ng/ul and digested overnight at 37 degrees C. Peptides were extracted by 3 washes with 5% formic acid in 50% acetonitrile. The combined supernatants were dried down in a Speedvac and the peptides were redissolved in 6 ul of 0.1% formic acid for MS analysis.

MS/MS analysis and data analysis. All analysis were performed on a Finnigan LCQ ion trap mass spectrometer that was run and operated as described in Link et al. (1997). The peptide sequence raw data was searched against a cereal database by SEQUEST software. To determine the function of the genes identified as being differentially expressed, a number of criteria were considered: the statistical score from SEQUEST, xcorr and deltCN, the peptide length and terminal sequence, the quality of the spectrum from the peptides, the number of peptides from the same protein spots that were identified in the same search, and the molecular weight and pI of the protein.

Results

Approximately 100 2-D gels under three different conditions were analyzed with samples from 2 normal lines, 3 high protein lines and 10 selected lines from the segregating population. Using data from mature embryos from Wil500, WIO465, LH59 and WICY530, approximately 120 differentially expressed proteins in were ideneified and isolated. Figure 2 is an example of two gels, one with proteins from control maize embryos (pH 5-8, spots 13-18, panel A) and another with proteins from a high protein line (pH 5-8, spots 1-12, panel B). Figures 2C and 2D are further examples, in which the arrow points to a readily identifiable difference area that contains the various forms of globulins proteins in embryo as described in the invention.

Thirty-eight of the differentially expressed maize proteins or their orthologs are listed in the Sequence Listing. For example, the following proteins were found to be differentially expressed: globulin-1 s allele precursor, globulin 2 precursor, glucose and ribitol dehydrogenase, glutathione S-transferase, rab28 protein (maize), heat shock protein 17.2, oleosin 16 kD protein, and putative receptor protein kinase zmpk1 precursor. In general, there was more globulin in all the high protein lines tested and there appears to be a very different

mature product of the glb1 and glb2 genes in high protein lines. These differences may occur due to regulatory processes, allelic variation, at the mRNA and/or protein level or post-translationally. Figures 3A to 3H show a subset of the 38 proteins which were identified using different criteria (Xcorr and dCN).

This genetic information is useful for marker development for breeding purposes in maize and for seed protein content manipulation in cereals in general.

Example 2

Gene expression levels of three of the genes of the invention, with sequences in the Sequence Listing, were examined during seed development in rice using microarray technology. Relative gene expression levels were determined and are presented in Table 2. All three genes were up regulated in the time course of rice seed development. The gene expression levels were determined by hybridizing the rice mRNA isolated at various developmental stages to an Affymetrix gene chip containing rice gene sequences. The rice genechip covered about 20,000 rice genes. A similar pattern of gene experession during corn seed maturation si expected.

Table 2

Developmetal stages	17 kd Heat shock	Dehydrin	GRD
seed development f_Seed day 0 anthesis	50.08	35.49	117.72
seed development g_Seed day 2 post anthesis	131.87	36.5	132.49
seed development h_Seed day 4 post anthesis	193.08	45.24	149.95
seed development I_Seed day 7 post anthesis	383.84	151.97	325.3
seed development j_Seed day 9 post anthesis	465.35	194.87	426.05
seed development k_11-day_post anthesis	737.56	356.98	784.33
seed development l_14-day_post anthesis	632.7	457.08	840.55
seed development m_17-day_post anthesis	1,150.06	438.75	1,486.17
seed development n_19-day_post anthesis	1,268.10	433.28	1,702.69

Example 4

Co-segregation of 17kD heat shock protein gene expression with high protein phenotype was observed. Tagman analysis (real time PCR) of 17 kD heat shock protein gene expression was compared with unbiqutin gene expression level as a reference. 30DAP embryos were used in these experiments. PCR primers were designed according the sequences shown in the Sequence Listing. This result demonstrates the cosegregation of HS17 gene expression with the high protein phenotype in the hybrids tested. These are the same corn lines as showed in Figure 1.

Table 3

Corn lines	**HS 17 gene**	UBQ gene
73/93 High Protein	1.59	1
73/98 High Protein	0.69	1
73/88 High Protein	5.05	1
73/92 Low Protein	0.07	1
73/76 Low Protein	0.01	1
73/84 Low Protein	0.07	1
LH59 Normal Protein	0.29	1
WIL500 high protein	5.33	1

Taqman analysis for key candidate gene expression was performed as follows. For one step RT-PCR amplification, total RNA was used in a 50 ml reaction using the master mixture of a Taq-Man One-Step RT-PCR Mix Reagnets (cat # 4309169, lot# 0006014) (PE Biosystems, Foster City, CA), following the manufacturer's protocol. The one step RT-PCR was conducted with an ABI Prism* 7900 HT Sequence Detection System (AB Applied Biosystems, Foster City, CA). The reactions were incubated for 30 min at 48° C for reverse transcription, and for 40 cycles of 15 s at 95° C, 60 s at 60° C for amplification. The ramp rate was set at 100%

between two different temperature set points. 50 ml Reaction was composed of 6.25 ml of 2 mM forward primer, AtTRX3-F (gtgtggaaatgacacagattgtga), 6.25 ml of 2 mM reverse primer, AtTRX3-R (agacgggtgcaatgaaacg), 6.25 ml of 2 mM TaqMan probe (6FAM-agacttcactgcaacatggtgcccac-TAMRA), AtTRX3_TaqMan, 1.25 ml of 40x MultiScribe & Rnase inhibitor Mix, 5 ml of template RNA (50 ng total RNA), and 25 ml of Master Mix w/o UNG (Taq-Man One-Step RT-PCR Mix Reagnets: cat # 4309169, lot# 0006014) (PE Biosystems, Foster City, CA). Data collection was processed between two temperature set points of 95° C and 60° C during amplification. The fold change in TRX3 transcript was determined following the ABI Prism 7900HT Sequence Detection System User Guide (Applied Biosystems): Fold change = 2-DCt, where DCt = -(Ct TRX - CtTRX-STD1ng) threshold 0.36507.

Example 5

Modulation of high protein trait by genes of the invention is readily determined using plant transformation sytems as described herein and as known in the art. In one embodiment, the Gateway cloning system was used to introduce genes of the invention into agrotransformation vectors for cereals, with seed specific promoters. See Figures 4A and 4B. The embryo specific promoter is a globulin promoter, and the ADPGPP gene promoter is used as the endosperm specific promoter. Use of these promoter constructs allows ease of cloning various genes under the control of these promoter to overexpress and/or downregulate the expression of these genes.

Gateway System cloning of pOPT003 & pOPT004 was as follows. Two oligos (NJ001 for & NJ002rev) were designed to amplify the Gateway Cassette A. These oligos contain restriction enzymes (Bcl I and Spe I) to clone into Xba I and BamH I sites of the pNOV4000 and pNOV4002 vector (note that Xba I is compatible with Spe I site and Bcl I is compatible with BamH I site). The sGFP-M5 gene of the pNOV4000 and pNOV4002 plasmid is replaced with the Gateway cassette A in which we generated pOPT001 and pOPT002 vector. pOPT001, pOPT002 and pNOV2117 (agro) vector were digested and ligated with Kpn I and Hind III sites. The final products were transformed into DB3.1 E.Coli cells, and the pOPT003 and pOPT004 vectors were generated, as shown in the figures.

Protein determination was done as follows. For green house generated materials, seed protein were determined by elemental analysis, nitrogen to caculate the total protein yield, using conversion factor 6.25. A N/protein analyzer, FLASH EA 1112 Series , from CE

Instruments were used in our experiment. Protein content for field generated seed materials were determined by NIR (Near Infrared) analysis.

Example 6

Vector construction for overexpression and gene "knockout" experiments.

Overexpression

Vectors used for expression of full-length genes of the embodiments of the invention of interest in plants (overexpression) are designed to overexpress the protein of interest and are of two general types, biolistic and binary, depending on the plant transformation method to be used.

For biolistic transformation (biolistic vectors), the requirements are as follows:

1. a backbone with a bacterial selectable marker (typically, an antibiotic resistance gene) and origin of replication functional in *Escherichia coli* (*E. coli*; eg. ColE1), and
2. a plant-specific portion consisting of:
    a. a gene expression cassette consisting of a promoter (eg. ZmUBIint MOD), the gene of interest (typically, a full-length cDNA) and a transcriptional terminator (eg. *Agrobacterium tumefaciens nos* terminator);
    b. a plant selectable marker cassette, consisting of a promoter (eg. rice Act1D-BV MOD), selectable marker gene (eg. phosphomannose isomerase, PMI) and transcriptional terminator (eg. CaMV terminator).

Vectors designed for transformation by *Agrobacterium tumefaciens* (*A. tumefaciens*; binary vectors) consist of:

1. a backbone with a bacterial selectable marker functional in both *E. coli* and *A. tumefaciens* (eg. spectinomycin resistance mediated by the *aadA* gene) and two origins of replication, functional in each of aforementioned bacterial hosts, plus the *A. tumefaciens virG* gene;
2. a plant-specific portion as described for biolistic vectors above, except in this instance this portion is flanked by *A. tumefaciens* right and left border sequences which mediate transfer of the DNA flanked by these two sequences to the plant.

Knock out vectors

Vectors designed for reducing or abolishing expression of a single gene or of a family or related genes (knockout vectors) are also of two general types corresponding to the

methodology used to downregulate gene expression: antisense or double-stranded RNA interference (dsRNAi).

Anti-sense

For antisense vectors, a full-length or partial gene fragment (typically, a portion of the cDNA) can be used in the same vectors described for full-length expression, as part of the gene expression cassette. For antisense-mediated down-regulation of gene expression, the coding region of the gene or gene fragment will be in the opposite orientation relative to the promoter; thus, mRNA will be made from the non-coding (antisense) strand *in planta*.

dsRNAi

For dsRNAi vectors, a partial gene fragment (typically, 300 to 500 basepairs long) is used in the gene expression cassette, and is expressed in both the sense and antisense orientations, separated by a spacer region (typically, a plant intron, eg. the OsSH1 intron 1, or a selectable marker, eg. conferring kanamycin resistance). Vectors of this type are designed to form a double-stranded mRNA stem, resulting from the basepairing of the two complementary gene fragments *in planta*.

Biolistic or binary vectors designed for overexpression or knockout can vary in a number of different ways, including eg. the selectable markers used in plant and bacteria, the transcriptional terminators used in the gene expression and plant selectable marker cassettes, and the methodologies used for cloning in gene or gene fragments of interest (typically, conventional restriction enzyme-mediated or Gateway™ recombinase-based cloning). An important variant is the nature of the gene expression cassette promoter driving expression of the gene or gene fragment of interest in most tissues of the plants (constitutive, eg. ZmUBIint MOD), in specific plant tissues (eg. maize ADP-gpp for endosperm-specific expression), or in an inducible fashion (eg. GAL4bsBz1 for estradiol-inducible expression in lines constitutively expressing the cognate transcriptional activator for this promoter).

Insertion of a gene of the embodiments of the invention into Expression Vector

A validated rice cDNA clone such as the *OsPT11* cDNA prepared in Example 14 above, in pCR2.1-TOPO is subcloned using conventional restriction enzyme-based cloning into a vector, downstream of the maize ubiquitin promoter and intron, and upstream of the *Agrobacterium tumefaciens* nos 3' end transcriptional terminator. The resultant gene expression cassette (promoter, gene of the embodiments of the invention and terminator) is

further subcloned, using conventional restriction enzyme-based cloning, into the pNOV2117 binary vector, generating pNOVCAND.

The pNOVCAND binary vector is designed for transformation and over-expression of the gene of the embodiments of the invention in monocots. It consists of a binary backbone containing the sequences necessary for selection and growth in *Escherichia coli* DH-5α (Invitrogen) and *Agrobacterium tumefaciens* LBA4404, including the bacterial spectinomycin antibiotic resistance *aadA* gene from *E. coli* transposon Tn7, origins of replication for *E. coli* (ColE1) and *A. tumefaciens* (VS1), and the *A. tumefaciens virG* gene. In addition to the binary backbone, pNOV2117 contains the T-DNA portion flanked by the right and left border sequences, and including the Positech™ (Syngenta) plant selectable marker and the gene of the embodiments of the invention gene expression cassette. The Positech™ plant selectable marker confers resistance to mannose and in this instance consists of the maize ubiquitin promoter driving expression of the PMI (phosphomannose isomerase) gene, followed by the cauliflower mosaic virus transcriptional terminator.

This is exemplified in Rice Transformation as follows. pNOVCAND is transformed into a rice cultivar (Kaybonnet) using Agrobacterium-mediated transformation, and mannose-resistant calli are selected and regenerated.

Agrobacterium is grown on YPC solid plates for 2-3 days prior to experiment initiation. Agrobacterial colonies are suspended in liquid MS media to an OD of 0.2 at λ600nm. Acetosyringone is added to the agrobacterial suspension to a concentration of 200μM and agro is induced for 30min.

Three-week-old calli which are induced from the scutellum of mature seeds in the N6 medium (Chu, C.C. et al., Sci, Sin., 18, 659-668(1975)) are incubated in the agrobacterium solution in a 100 x 25 petri plate for 30 minutes with occasional shaking. The solution is then removed with a pipet and the callus transfered to a MSAs medium which is overlayed with sterile filter paper.

Co-Cultivation is continued for 2 days in the dark at 22°C.

Calli are then placed on MS-Timetin plates for 1 week. After that they are tranfered to PAA + mannose selection media for 3 weeks.

Growing calli (putative events) are picked and transfered to PAA+ mannose media and cultivated for 2 weeks in light.

Colonies are tranfered to MS20SorbKinTim regeneration media in plates for 2 weeks in light. Small plantlets are transferred to MS20SorbKinTim regeneration media in GA7 containers. When they reach the lid, they are transfered to soil in the greenhouse.

Expression of the gene of the embodiments of the invention in transgenic T_0 plants is analyzed. Additional rice cultivars, such as but not limited to, Nipponbare, Taipei 309 and Fuzisaka 2 are also transformed and assayed for expression of the gene product of the embodiments of the invention and enhanced protein expression.

References

Altschul et al., J. Mol. Biol. 215:403 (1990).

Altschul et al., Nucleic Acids Res. 25:3389 (1997).

An et al., EMBO J. 4:277 (1985).

Aoyama T. et al., N-H Plant Journal, 11:605 (1997).

Ballas et al., Nucleic Acids Res. 17:7891 (1989).

Bansal et al., Proc. Natl. Acad. Sci. USA, 89:3654 (1992).

Batzer et al., Nucleic Acid Res., 19:5081 (1991).

Beals et al., Plant Cell, 9:1527 (1997).

Belanger et al., Genetics, 129:863 (1991).

Bevan et al., Nature, 304:184 (1983).

Blochinger & Diggelmann, Mol Cell Biol, 4:2929.

Bourouis et al., EMBO J., 2:1099 (1983).

Brandle et al., Biotech. 13:994 (1995).

Byrne et al. Plant Cell Tissue and Organ Culture, 8:3 (1987).

Campbell and Gowri, Plant Physiol., 92:1 (1990).

Chandler et al., Plant Cell, 1:1175 (1989).

Christou et al., Biotechnology, 9:957 (1991).

Christou et al., Plant Physiol., 87:671 (1988).

Cordero et al., Plant J., 6:141 (1994).

Corpet et al., Nucleic Acids Res. 16:10881 (1988).

Crameri et al., Nature Biotech., 15:436 (1997).

Crameri et al., Nature, 391:288 (1998).

Crossway et al., BioTechniques, 4:320 (1986).

Czako et al., Mol. Gen. Genet. 23 5 (1), 33-40 (1992).

Datta et al., Bio/Technology, 8, 736 (1990).

Dayhoff et al., Atlas of Protein Sequence and Structure, Natl. Biomed. Res. Found., Washington, C.D. (1978).

De Blaere et al., Meth. Enzymol. 143:277 (1987).

de Framond, FEBS, 290:103 (1991).

Della-Cioppa et al., Plant Physiology, 84:965 (1987).

Dennis et al., Nucleic Acids Res., 12:3983 (1984).

Dzelkalns et al., Plant Cell, 5:855 (1993).

Elroy-Stein et al., PNAS USA, 86:6126 (1989).

English, et al., Plant Cell 8:179 (1996).

Eulgem et al., Trends in Plant Sci. 5:199 (2000).

Flavell, Proc. Natl. Acad. Sci USA 91:3490 (1994).

Franken et al., EMBO J., 10:2605 (1991).

Fromm et al., Bio/Technology 8:833 (1990).

Gallie et al., Molecular Biology of RNA, 237 (1989).

Gallie et al., Nucl. Acids Res., 15:8693 (1987).

Gan et al., Science (1995), 270 (5244), 1986-8).

Gatz Current Opinion in Biotechnology, 7:168 (1996).

Gatz, C., Annu. Rev. Plant Physiol. Plant Mol. Biol., 48:89 (1997).

Gelfand, eds., PCR Strategies (Academic Press, New York (1995)).

Gordon-Kamm et al., Plant Cell, 2, 603 (1990).

Graham et al., Biochem. Biophys. Res. Comm., 101:1164 (1981).

Graham et al., J. Biol. Chem., 260:6555 (1985).

Graham et al., J. Biol. Chem., 260:6561 (1985).

Guerineau et al., Mol. Gen. Genet. 262:141 (1991).

Hansel et al., Cereal Chem. 50:383 (1973).

Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA, 89:10915 (1989).

Hiei et al., Plant J. 6:271 (1994).

Higgins et al., Gene 73:237 (1988).

Higgins et al., CABIOS 5:151 (1989).

Hinchee et al., Biotechnology, 6:915 (1988).

Hoekema, In: The Binary Plant Vector System. Offset-drukkerij Kanters B.V.

Huang et al., CABIOS 8:155 (1992).

Hudspeth & Grula, Plant Molec. Biol., 12:579 (1989).

Huffman et al., J. Cell. Biochem., 17B: Abstract.

Ingelbrecht et al., Plant Cell, 1:671 (1989).

Innis et al., eds., PCR Protocols: A Guide to Methods and Applications (Academic Press, New York (1995).

Innis and Gelfand, eds., PCR Methods Manual (Academic Press, New York) (1999).

Jobling et al., Nature, 325:622 (1987).

John et al., Proc. Natl. Acad. Sci. USA 89(13):5769 (1992).

Jones, Biochem. Genet. 16:27 (1978).

Jones et al., Adv Bot Res 24:89 (1997).

Joshi et al., Nucleic Acid Res. 15:9627 (1987).

Joshi, NAR, 15:6643 (1987).

Karlin and Altschul, Proc. Natl. Acad Sci. USA 872264 (1990).

Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90:5873 (1993).

Keller et al., Genes Dev., 3:1639 (1989).

Klein et al., Bio/Technology, 6:559 (1988).

Klein et al., Nature (London) 327:70 (1987).

Klein et al., Plant Physiol., 91:440 (1988).

Klein et al., Proc. Natl. Acad. Sci. USA, 85:4305 (1988).

Kohler et al., Plant Mol. Biol., 29:1293 (1995).

Knauf, et al., Genetic Analysis of Host Range Expression by *Agrobacterium* In: Molecular Genetics of the Bacteria-Plant Interaction, Puhler, A. ed., Springer-Verlag, New York, 245 (1983).

Koziel et al., Biotechnology, 11:194 (1993).

Kridl et al., Seed Science Research, 1:209 (1991).

Kriz et al., Mol. Gen. Genet., 207:90 (1987).

Kunkel, Proc. Natl. Acad. Sci. USA, 82:488 (1985).

Kunkel et al., Methods in Enzymol., 154:367 (1987).

Langridge et al., Cell, 34:1015 (1983).

Lindstrom et al., Der. Genet., 11:160 (1990).

Link et al., Electrophoresis, 18:1314 (1997).

Lommel et al., Virology, 81:382 (1991).

Lorz et al. Mol. Gen. Genet. 199:178 (1985).

Macejak et al., Nature, 353:90 (1991).

Mansson et al., Gen. Genet., 200:356 (1985).

Martinez et al., J. Mol. Biol., 208:551 (1989).

Matzke and Matzke, Plant Physiol. 107:679 (1995).

McBride et al., Proc. Natl. Acad. Sci. USA, 91:7301 (1994).

McCabe et al., Bio/Technology, 6:923 (1988).

McDowell et al., Plant Cell 10:1861 (1998).

Meinkoth and Wahl, Anal. Biochem., 138:267 (1984).

Mertz et al., Science 145:279 (1964).

Messing & Vierra, Gene, 19:259 (1982).

Misra et al., Science 176:1425 (1972).

Mogen et al., Plant Cell 2:1261 (1990).

Moore et al., J. Mol. Biol., 272:336 (1997).

Munroe et al., Gene 91:151 (1990).

Murray et al., Nucleic Acids Res., 17:477 (1989).

Myers and Miller, CABIOS 4:11 (1988).

Needleman and Wunsch, J. Mol. Biol. 48:443 (1970).

Odell et al., Nature, 313:810 (1985).

Odell et al., Nature, 313:810 (1985).

Ohtsuka et al., J. Biol. Chem., 260:2605 (1985).

Okamuro et al., Biochemistry of Plants, 15:1 (1989).

Paszkowski et al., EMBO J., 3:2717 (1984).

Pacciotti et al. Bio/Technology 3:241 (1985).

Park et al., J. Plant Biol. 38(4):365 (1985).

Paulis et al., Cereal Chem. 55:705 (1978).

Pearson et al., Meth. Mol. Biol. 24:307 (1994).

Pearson and Lipman, Proc. Natl. Acad. Sci. 85:2444 (1988).

Penninck et al., Plant Cell 8:2309 (1996).

Perlak et al., Proc. Natl. Acad. Sci. USA, 88:3324 (1991).

Potrykus Mol. Gen. Genet. 199:183 (1985).

Proudfoot, Cell 64:671 (1991).

Quigley et al., J. Mol. Evol., 29:412 (1989).

Ralston et al., Genetics, 119:185 (1988).

Reina et al., Nucleic Acids Res., 18:6425 (1990).

Reina et al., Nucleic Acids Res., 18:7449 (1990).

Riggs et al., Proc. Natl. Acad. Sci. USA, 83:5602 (1986).

Rochester et al., (1986).

Rodermel et al., Cell 55:673 (1988).

Rossolini et al., Mol. Cell. Probes, 8:91 (1994).

Rubenstein, In: Maize for Biological Research (Sheridan ed., U of N. Dakota Press, pp 189-196 (1982)).

Ruiz et al., Plant Cell 10:937 (1998).

Sambrook et al., Molecular Cloning: A Laboratory Manual (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York) (1989).

Sanfacon et al., Genes Dev. 5:141 (1991).

Sanford et al., Particulate Science and Technology, 5:27 (1987).

Schernthaner et al., EMBO J., 7:1249 (1988).

Schwob et al., Plant J., 4:423 (1993).

Shimamoto et al., Nature, 338:274 (1989).

Simpson, Plant Mol. Biol., 19:699 (1985).

Skuzeski et al., Plant Molec. Biol., 15:65 (1990).

Slater et al., Plant Mol. Biol., 5:137 (1985).

Smith et al., Adv. Appl. Math. 2:482 (1981).

Smith et al., Planta 168:94 (1986).

Spencer et al., Theor Appl Genet, 79:625 (1990).

Staub et al., EMBO J., 12:601 (1993).

Staub et al., Plant Cell, 4:39 (1992).

Stemmer, Nature, 370:389 (1994).

Stemmer, Proc. Natl. Acad. Sci. USA, 91:10747 (1994).

Sukhapinda et al. Plant Mol. Biol. 8:209 (1987).

Sullivan et al., Mol. Gen. Genet., 215:431 (1989).

Svab et al., Proc. Natl. Acad. Sci. USA, 87:8526 (1990).

Svab et al., Proc. Natl. Acad. Sci. USA, 90:913 (1993).

Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York (1993).

Tsai et al., Biochem. Genet. 16:883 (1978).

Turner et al., Molecular Biotechnology, 3:225 (1995).

van der Krol et al., 1988a; 1988b, 1988c.

VanTunen et al., EMBO J., 7:1257 (1988).

Vasil et al., Biotechnology, 11:1553 (1993).

Vodkin, Prog. Clin. Biol. Res., 138:87 (1983).

Vogel et al., EMBO J., 11:157 (1992).

Walker and Gaastra, eds., Techniques in Molecular Biology, MacMillan Publishing Company, New York (1983).

Wandelt et al., Nucleic Acids Res., 17:2354 (1989).

Waterman, M.S. Introduction to Computational Biology: Maps, sequences and genomes. Chapman & Hall. London (1995).

Weeks et al., Plant Physiol., 102:1077 (1993).

Weissinger et al., Annual Rev. Genet., 22:421 (1988).

Wenzler et al., Plant Mol. Biol., 13:347 (1989).

White et al., Nucl Acids Res, 18:1062 (1990).

Yamamoto et al., Nucleic Acids Res., 18:7449 (1990).

Yu et al., Proc Natl Acad Sci USA 95:7819 (1998).

Zhang et al., Proc. Natl. Acad. Sci. USA, 94:4504 (1997).

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

We claim:

1. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which hybridizes under low stringency conditions to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

2. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which hybridizes under high stringency conditions to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

3. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which hybridizes under moderate stringency conditions to a nucleic acid segment encoding a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

4. An isolated nucleic acid molecule comprising a plant nucleotide sequence or its complement which encodes a polypeptide that is substantially similar to a polypeptide comprising any one of SEQ ID NOs: 1-36, wherein the nucleotide sequence does not encode any one of SEQ ID NOs: 1-36.

5. The isolated nucleic acid molecule of claim 1, 2, 3 or 4 which is DNA.

6. The isolated nucleic acid molecule of claim 1, 2, 3 or 4 which is RNA.

7. The isolated nucleic acid molecule of claim 4 wherein the nucleotide sequence encodes a polypeptide having at least 90% amino acid sequence identity to the polypeptide comprising any one of SEQ ID NOs: 1-36.

8. The isolated nucleic acid molecule of claim 4 wherein the nucleotide sequence encodes a polypeptide having at least 80% amino acid sequence identity to the polypeptide comprising any one of SEQ ID NOs: 1-36.

9. The isolated nucleic acid molecule of claim 4 wherein the nucleotide sequence encodes a polypeptide having at least 70% amino acid sequence identity to the polypeptide comprising any one of SEQ ID NOs: 1-36.

10. A polypeptide encoded by the nucleic acid molecule of claim 1, 2, 3 or 4.

11. An expression cassette comprising the nucleic acid molecule of claim 1, 2, 3 or 4 operably linked to suitable regulatory sequences.

12. The expression cassette of claim 11 which is linked to a promoter for expression in a plant.

13. A recombinant vector comprising the nucleic acid molecule of claim 1, 2, 3 or 4.

14. A host cell comprising the expression cassette of claim 11.

15. A host cell comprising the isolated nucleic acid molecule of claim 1, 2, 3 or 4.

16. The host cell of claim 15 which is selected from the group consisting of yeast, bacteria and plant.

17. A transformed plant, or seed thereof, the genome of which is augmented with the nucleic acid molecule of claim 1, 2, 3 or 4 which is expressed in an amount which confers increased protein content to the plant.

18. A transformed plant, or seed thereof, the genome of which is genetically altered so as to inhibit the expression of a gene corresponding to the nucleic acid molecule of claim 1, 2, 3 or 4.

19. The plan, or seed thereof, of claim 18 which is altered by T-DNA insertion, transposon insertion, or targeted DNA insertion.

20. The plant, or seed thereof, of claim 18 in which expression is inhibited by transcription or post-transcriptional mechanisms.

21. The plant, or seed thereof, of claim 17 or 18 which is a monocot.

22. The plant, or sed thereof, of claim 17 or 18 which is a dicot.

23. A method of expressing a nucleic acid molecule in a cell, comprising:
introducing the nucleic acid molecule of claim 1, 2, 3 or 4 into a cell so as to express the nucleic acid molecule.

24. The method of claim 23 wherein the cell is a plant cell.

25. The method of claim 23 wherein the cell is a monocot cell.

26. The method of claim 23 wherein the cell is a dicot cell.

27. A composition comprising the nucleic acid molecule of claim 1, 2, 3 or 4.

28. A composition comprising the polypeptide of claim 10.

29. A method to confer altered nutritional qualities to a plant, comprising:
a) contacting plant cells with an expression cassette comprising the nucleic acid molecule of claim 1, 2, 3 or 4 so as to yield transformed plant cells; and
b) regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the nucleic acid molecule in the cells of the plant in an amount effective to alter the protein content of the transformed plant relative to a corresponding plant which does not comprise the expression cassette.

30. A method to confer altered nutritional qualities to a plant, comprising:
a) contacting plant cells with an expression cassette comprising a nucleotide sequence encoding a polypeptide comprising any one of SEQ ID NOs: 1-36 so as to yield transformed plant cells; and

b) regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the nucleotide sequence in the cells of the plant in an amount effective to alter the protein content of the transformed plant relative to a corresponding plant which does not comprise the expression cassette.

31. A transformed plant prepared by the method of claim 29 or 30.

32. A seed of the plant of claim 31.

33. A progeny plant of the plant of claim 31.

34. An isolated nucleic acid molecule comprising a plant nucleotide sequence that directs transcription of an operatively linked nucleic acid fragment in a host cell, which nucleotide sequence corresponds to plant genomic DNA which is substantially similar to a nucleic acid segment which directs the transcription of a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36.

35. The nucleic acid molecule of claim 34 wherein the nucleotide sequence has at least 90% identity to the nucleic acid segment.

36. A recombinant vector comprising the nucleic acid molecule of claim 34.

37. The vector of claim 36 which is a plasmid.

38. An expression cassette comprising the nucleic acid molecule of claim 34 operatively linked to an open reading frame.

39. The expression cassette of claim 38 operably linked to other suitable regulatory sequences.

40. A host cell comprising the expression cassette of claim 38.

41. A transformed plant, the genome of which is augmented with the expression cassette of claim 38.

42. A plant cell containing the expression cassette of claim 38.

43. A transformed plant comprising transformed plant cells, the transformed plant cells containing the expression cassette of claim 38.

44. The transformed plant of claim 43 wherein the plant is a dicot.

45. The cell of claim 42 which is a dicot cell.

46. The transformed plant of claim 43 wherein the plant is a monocot.

47. The cell of claim 42 which is a monocot cell.

48. The transformed plant of claim 43 which is a cereal plant.

49. A method of augmenting a plant genome, comprising:

a) contacting plant cells with the expression cassette of claim 38 so as to yield transformed plant cells; and

b) regenerating the transformed plant cells to provide a differentiated transformed plant, wherein the differentiated transformed plant expresses the nucleic molecule in the cells of the plant.

50. A transformed plant prepared by the method of claim 49.

51. A seed of the plant of claim 50.

52. A progeny plant of the plant of claim 50.

53. A method of using a plant promoter, comprising: introducing the expression cassette of claim 38 to a plant cell and detecting the expression of the product of the open reading frame.

54. An isolated nucleic acid molecule comprising a plant nucleotide sequence that directs transcription of an operatively linked nucleic acid fragment in a plant cell, which nucleotide sequence corresponds to plant genomic DNA which hybridizes under low stringency conditions a nucleic acid segment that directs transcription of a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36.

55. An isolated nucleic acid molecule comprising a plant nucleotide sequence that directs transcription of an operatively linked nucleic acid fragment in a plant cell, which nucleotide sequence corresponds to plant genomic DNA which hybridizes under high stringency conditions a nucleic acid segment that directs transcription of a gene encoding a polypeptide comprising any one of SEQ ID NOs: 1-36.

56. A recombinant vector comprising the expression cassette of claim 38.

57. A plant cell comprising the vector of claim 56.

58. A transformed plant, the cells of which comprise the vector of claim 56.

59. The nucleic acid molecule of claim 34, 54 or 55 wherein the nucleotide sequence is 25 to 2000 nucleotides in length.

60. The expression cassette of claim 38 wherein the open reading frame is in an antisense orientation.

61. The expression cassette of claim 38 wherein the open reading frame is in a sense orientation.

62. The expression cassette of claim 12 wherein the nucleic acid molecule is in antisense orientation.

63. The expression cassette of claim 12 wherein the nucleic acid molecule is in sense orientation.

64. An antibody that binds to the polypeptide of claim 10.

65. A method for marker-assisted selection of plants having a desired property, comprising:

a) contacting a probe comprising at least a portion of a nucleic acid sequence comprising an open reading frame encoding a polypeptide comprising any one of SEQ ID NOs: 1-36 with a nucleic acid sample from a plant in an amount sufficient to form complexes; and

b) detecting or determining the amount of complex formation.

66. A method for marker-assisted selection of plants having a desired property, comprising:

a) contacting a probe comprising at least a portion of the nucleic acid molecule of claim 1, 2, 3 or 4 with a nucleic acid sample from a plant in an amount sufficient to form complexes; and

b) detecting or determining the amount of complex formation

67. A method for marker-assisted selection of plants having a desired property, comprising:

a) contacting a sample comprising plant proteins with the antibody of claim 64 in an amount sufficient to form complexes; and

b) detecting or determining the amount of complex formation.

68. A method to identify transcription factors for genes associated with high protein content in plants, comprising:

a) contacting the nucleic acid molecule of claim 34, 54 or 55 with a sample comprising transcription factor polypeptides so as to form a complex between the nucleic acid molecule and at least one transcription factor; and

b) detecting or determining complex formation.

69. The method of claim 68 further comprising identifying the transcription factor in the complex.

70. A method of feeding livestock, which comprises feeding livestock a plant of any of the claims 17-22, 31-33, 41-48, 50-52 or 57-58, or a plant part thereof.

71. The method of claim 70 where the plant part is grain or seed.

72. A manufacturing process, which comprises milling grain produced on a plant of any of the claims 17-22, 31-33, 41-48, 50-52 or 57-58, or a plant part thereof.

73. A product which comprises milled grain produced on a plant of any of the claims 17-22, 31-33, 41-48, 50-52 or 57-58, or a plant part therof.

74. The product of claim 73 that is a human or animal food product.

75. A method of producing an industrially or therapeutically important protein in a plant or part thereof, such as a seed, comprising modulaton of high protein phenotype by over or under expressing one or more high protein phenotype genes in the host plant.

76. The method of claim 78 wherein under expressing or down regulation of one or more of the genes provides a plant or part thereof, with an increased ability to produce the industrially or therapeutically important polypeptide.

Figure 1

Seed protein content in the high protein lines and in selected hybrids

Maize materials	Protein %
1 Inbreds	
LH59	13.5
WI500	15.6
2. Hybrids	
73/93	11.9
73/88	11.9
73/98	11.9
73/92	8.2
73/76	8.4
73/84	8.5

The high protein materials used in this study were from Wilson Genetics, originally derived from a tropical germplasm. WIL500 is the key high protein source and was the inbred used in most of this study for high protein source. The normal protein control used in this study is LH59. Seeds of LH59 and WIL 500 were generated from green house generated seed materials

The hybrids listed in this study were derived from recombinant inbred lines crossed with a tester line called JHAx412B. The inbred lines used to generate the hybrids were recombinant inbred lines (F4 generation) generated from a cross between WI500 and a normal protein line, BIIJ208.

FIGURE 2A

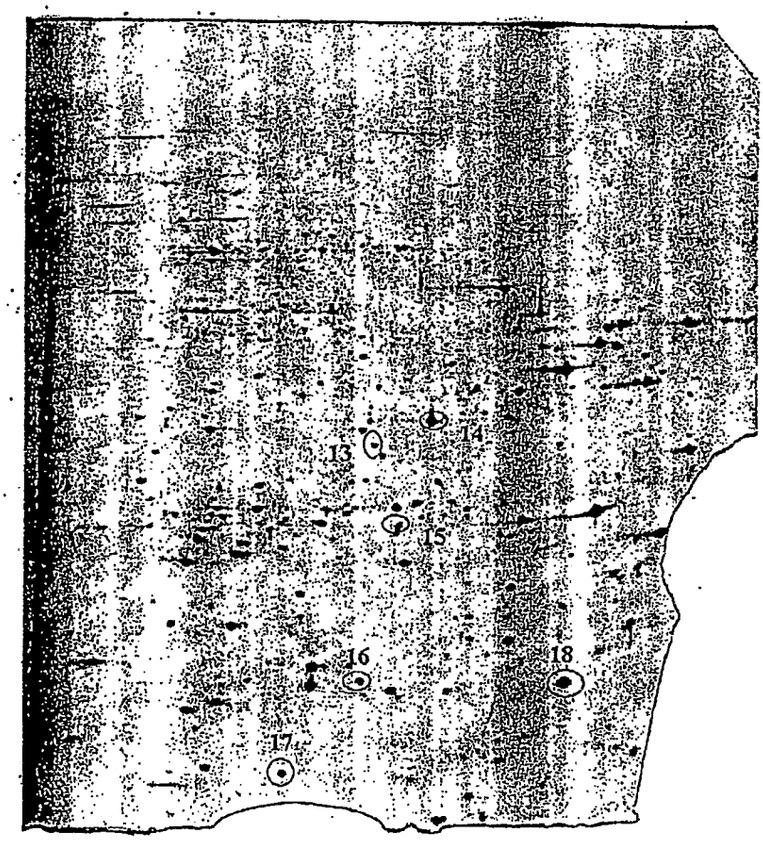


FIGURE 2B

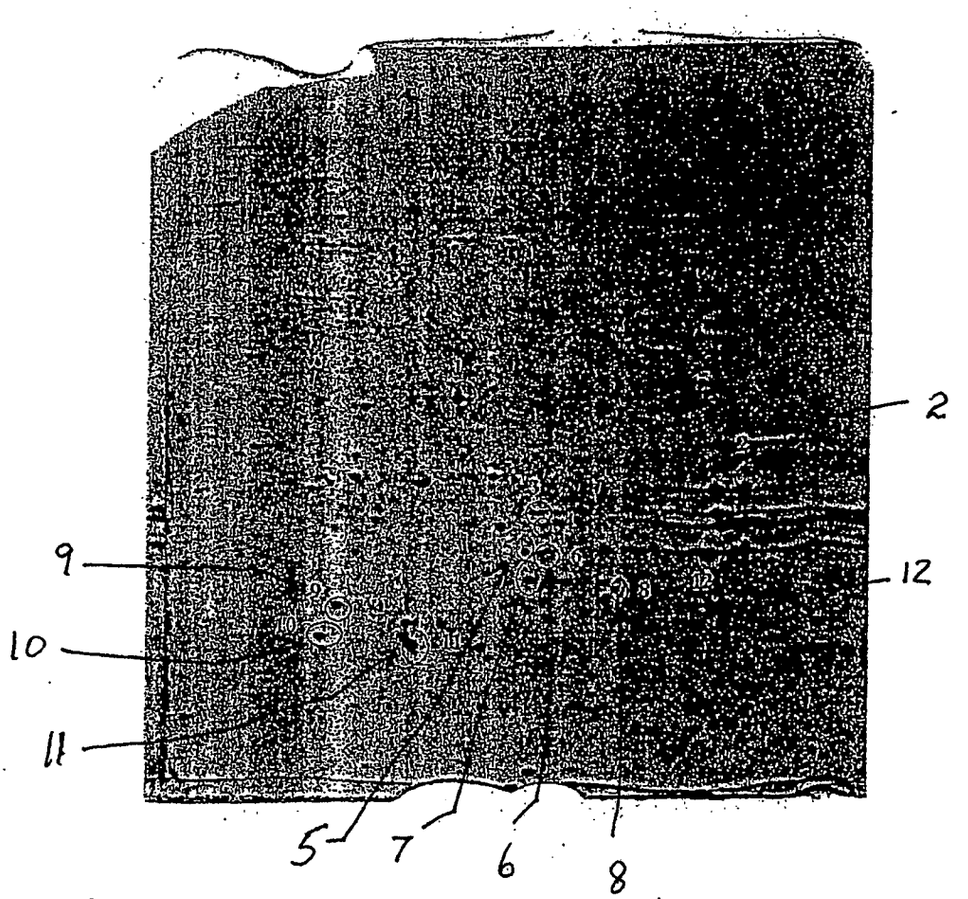
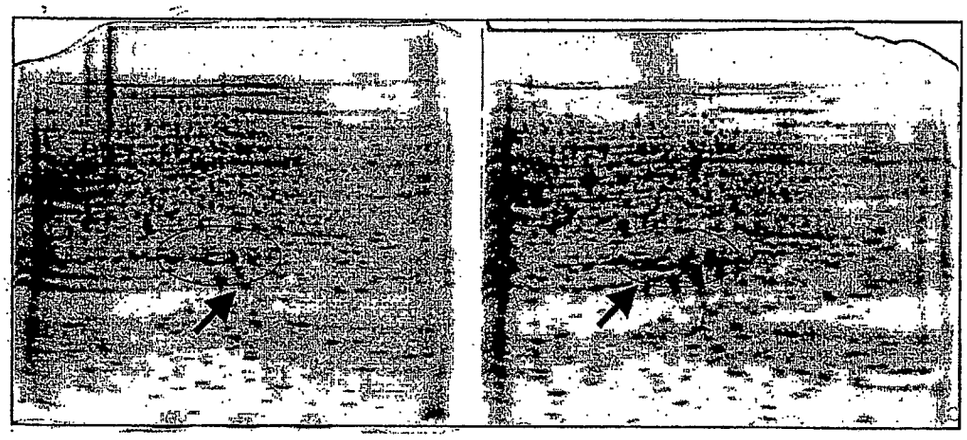


Figure 2C and 2D

Comparision of protein expression profile of high protein germplasm and normal corn line

This example demonstrated the proteomic approach used to identify proteins that are related to the high protein phenotype. Seed proteins were separated by two dimensional electrophoresis and the differentially expressed spots were identified by mass spec as described in the methods. 30DAP embryos were used. The arrow points to a clear difference area that contains the various forms of globulins proteins in embryo.

2C 2D



LH 59 WI500

FIGURE 3A

Spot wp0824a_1 **Corn seed embryo**

Apparent mass: 50.0, apparent pI: 6.5

XCorr	dCn	Sp	RSp	Ions	Ref	Sequence1664.8 (-0.5)
4.4007	0.425	2421.4	1	21/26	PIR2:S15675	R.FTHELLEDAVGNYR.V

PIR2:S15675 globulin-2 precursor - maize

MKVPVLLLLV SLCFSLALAW QTDTESGSGR PYHYGEESFR HWTRSRQGRF
RVLERFTHEL LEDAVGNYRV AELEAAPRAF LQPSHYDADE VMFVKEGEGV
IVLLRGGKRE SFCVREGDVM VIPAGAVVYS ANTHQSEWFR VVMLLSPVVS
TSGRFEEFFP IGGESPESFL SVFSDDVIQA SFNTRREEWE KVFEKQSKGE
ITTASEEQIR ELSRSCSRGG RSSRSEGGDS GSSSSKWEIK PSSLTDKKPT
HSNSHGRHYE ITGDECPHLR LLDMDVGLAN IARGSMMAPS YNTRANKIAI
VLKGQGYFEM ACPHVSGGRS SPRRERGHGR EEEEEREEEQ GGGGGQKSRS
YRQVKSRIRE GSVIVIPAGH PTALVAGEDK NLAVLCFEVN ASFDDKVFLA
GTNSALQKMD RPAKLLAFGA DEEQQVDRVI GAQKDAVFLR GPQSHRVSSV

Position	MH+	Sequence (link:NCBI Blast)
56-69	1664.8149	FTHELLEDAVGNYR

Spot wp0824a_4 **corn seed embryo**

Apparent mass: 17.9 apparent pI: 5.8
3.4360 0.365 1424.1 2 19/26 PIR2:S72545 +7 N.AGLENGVLTVTVPK.A

PIR2:S72545 heat shock protein 16.9 - pearl millet

MSLVRRGNVF DPFSMDLWDP FDNMFRSIVP SSSSSDTAAF ANARIDWKET
PEVHVFKADL PGVKKEEVKV EVEDGNVLVI SGQRSKEKED KNDRWHRVER
SSGQFVRRFR LPEDAKTDQV NAGLENGVLT VTVPKAEGKK PEVKAIEISG

Position	MH+	Sequence (link:NCBI Blast)
122-13	1398.6426	AGLENGVLTVTVPK

FIGURE 3B

Spot wp0824a_6 corn seed embryo

Apparent mass: 21.5 apparent pI: 6.2
3.6105 0.302 1575.8 1 18/26 PIR2:T04358 K.GLAYEYLEQDLGNK.S

PIR2:T04358 glutathione transferase (EC 2.5.1.18)

MAEEKKQGLQ LLDFWVSPFG QRCRIAMDEK **GLAYEYLEQD** LGNKSELLLR
ANPVHKKIPV LLHDGRPVCE SLVIVQYLDE AFPAAAPALL PADPYARAQA
RFWADYVDKK LYDCGTRLWK LKGDGQAQAR AEMVEILRTL EGALGDGPFF
GGDALGFVDV ALVPFTSWFL AYDRFGGVSV EKECPRLAAW AKRCAERPSV
AKNLYPPEKV YDFVCGMKKR LGIE

Position	MH+	Sequence (link:NCBI Blast)
31-44	1613.7642	GLAYEYLEQDLGNK

Spot wp0810w_11 corn seed embryo

Apparent mass: 26.0 apparent pI: 8.4

3.7625 0.367 1493.8 1 19/30 X14312.1_0 +1 K.IDLQTAQQLQNQDDNR.G

X14312.1_0 Arabidopsis CRA1 gene for 12S seed storage protein. //start:stop=196:1951 //PID=; Arabidopsis
thaliana; 12S seed storage protein
MARVSSLLSF CLTLLILFHG YAAQQGQQGQ QFPNECQLDQ LNALEPSHVL
KSEAGRIEVW DHHAPQLRCS GVSFARYIIE SKGLYLPSFF NTAKLSFVAK
GRGLMGKVIP GCAETFQDSS EFQPRFEGQG QSQRFRDMHQ KVEHIRSGDT
IATTPGVAQW FYNDGQQPLV IVSVFDLASH QNQLDRNPRP FYLAGNNPQG
QVWLQGREQQ PQKNIFNGFG PEVIAQALKI **DLQTAQQLQN QDDNRGNIVR**
VQGPFGVIRP PLRGQRPQEE EEEEGRHGRH GNGLEETICS ARCTDNLDDP
SRADVYKPQL GYISTLNSYD LPILRFIRLS ALRGSIRQNA MVLPQWNANA
NAILYETDGE AQIQIVNDNG NRVFDGQVSQ GQLIAVPQGF SVVKRATSNR
FQWVEFKTNA NAQINTLAGR TSVLRGLPLE VITNGFQISP EEARRVKFNT
LETTLTHSSG PASYGRPRVA AA

Position	MH+	Sequence (link:NCBI Blast)
230-245	1900.9998	IDLQTAQQLQNQDDNR

FIGURE 3C

Spot wp0801w_21 corn seed embryo

Apparent mass: 75.0 apparent pI: 5.5

3.1818 0.377 1116.3 1 20/ 42 AF121355.1_0 +1
K.VTVANVESGGEFTVSSADDILK.A

AF121355.1_0 Arabidopsis thaliana peroxiredoxin TPx1 mRNA, complete cds. //start:stop=54:542 //PID=; Arabidopsis thaliana; thioredoxin-dependent peroxidase; peroxiredoxin TPx

MAPIAVGDVV PDGTISFFDE NDQLQTASVH SLAAGKKVIL FGVPGAFTPT CSMKHVPGFI EKAEELKSKG VDEIICFSVN DPFVMKAWGK TYPENKHVKF VADGSGEYTH LLGLELDLKD KGLGVRSRRF ALLLDDLKVT VANVESGGEF TVSSADDILK AL

Position	MH+	Sequence (link:NCBI Blast)
139-160	2239.4414	VTVANVESGGEFTVSSADDILK

FIGURE 3D

Spot nla0808_11 corn seed embryo

Apparent mass: 65.0 apparent pI: 6.6
:
3.4352 0.173 2003.0 1 18/ 20 SW:GLB1_MAIZE +3 K.AEEVDEVLGSR.R

SW:GLB1_MAIZE P15590 zea mays (maize). globulin-1 s allele precursor (glb1-s) (7s-like). 7/1999
MVSARIVVLL AVLLCAAAAV ASSWEDDNHH HHGGHKSGRC VRRCEDRPWH
QRPRCLEQCR EEEREKRQER SRHEADDRSG EGSSEDERER EQEKEEKQKD
RRPYVFDRRS FRRVVRSEQG SLRVLRPFDE VSRLLRGIRD YRVAVLEANP
RSFVVPSHTD AHCIGYVAEG EGVVTTIENG ERRSYTIKQG HVFVAPAGAV
TYLANTDGRK KLVITKILHT ISVPGEFQFF FGPGGRNPES FLSSFSKSIQ
RAAYKTSSDR LERLFGRHGQ DKGIIVRATE EQTRELRRHA SEGGHGPHWP
LPPFGESRGP YSLLDQRPSI ANQHGQLYEA DARSFHDLAE HDVSVSFANI
TAGSMSAPLY NTRSFKIAYV PNGKGYAEIV CPHRQSQGGE SERERGKGRR
SEEEEESSEE QEEVGQGYHT IRARLSPGTA FVVPAGHPFV AVASRDSNLQ
IVCFEVHADR NEKVFLAGAD NVLQKLDRVA KALSFASKAE EVDEVLGSRR
EKGFLPGPKE SGGHEEREQE EEEREERHGG RGERERHGRE EREKEEEERE
GRHGRGRREE VAETLLRMVT ARM

Position	MH+	Sequence (link:NCBI Blast)
489-499	1204.2793	AEEVDEVLGSR

FIGURE 3E

Spot nla0808_4 corn seed embryo

Apparent mass: 65.0 apparent pI:. 6.8

4.3004 0.354 2163.0 1 36/ 72 PIR2:S18545
R.NDGTARPGGVAASMAAAAR.L

PIR2:S18545 rab28 protein - maize

MSQEQPRRPS GHEETSGGGE QGAVRYGDVF PAVSGGLAEK PVARRTATMQ
SAENLVFGQT LKGGPAAAMQ SAATTNERMG AVGHDQATDA TAVQGVTVSE
TRVPGGGRIV TEFVAGQAVG QYLARDDDGG GGIAGPGAGA GVAGKDITKV
TIGEALEATA LAAGDAPVER SDAARIQAAE ARATGLDANV PGGLARQAQS
AAAANSWAWG DEDKATLGDV LANATARLVA DKPVESADAL GVAGAENRNR
NDGTARPGGV AASMAAAARL NRNEAVWE

Position	MH+	Sequence (link:NCBI Blast)
251-269	1744.9233	NDGTARPGGVAASMAAAAR

FIGURE 3F

Spot nla0808_7. corn seed embryo

Apparent mass: 55.0 apparent pI: 6.8

4.4214 0.477 1987.8 1 24/36 SW:ZEAD_MAIZE +1
R.AQQLQQLVLANLAAYSQQH.Q
SW:ZEAD_MAIZE P24450 zea mays (maize). zein-alpha precursor (19 kd) (pms2).

MAAKIFCFLM LLGLSASVAT ATIFPQCSQA PIASLLPPYL SPAVSSMCEN
PIVQPYRIQQ AIATGILPLS PLFLQQPSAL LQQLPLVHLV AQNIRAQQLQ
QLVLANLAAY SQQHQFLPFN QLAALNSAAY LQQQLPFSQL VAAYPRQFLP
FNQLAALNSA AYLQQQQLLP FSQLADVSPA AFLTQQQLLP FYLHAMPNAG
TLLQLQQLLP FNQLALTNST VFYQQPIIGG ALF

3.2085 0.226 2050.3 1 17/20 SW:ZEAA_MAIZE R.LQQAIAASILR.S

SW:ZEAA_MAIZE P06678 zea mays (maize). zein-alpha precursor (19 kd) (clone 19d1). 6/1994
MAAKIFALLA LLALSANVAT ATIIPQCSQQ YLSPVTAARF EYPTIQSYRL
QQAIAASILR SLALTVQQPY ALLQQPSLVN LYLQRIVAQQ LQQQLLPTIN
QVVAANLDAY LQQQQFLPFN QLAGVNPAAY LQAQQLLPFN QLVRSPAAFL
LQQQLLPFHL QVVANIAAFL QQQQLLPFYP QVVGNINAFL QQQQLLPFYP
QDVANNVAFL QQQQLLPFSQ LALTNPTTLL QQPTIGGAIF

Position	MH+	Sequence (link:NCBI Blast)
96-114	2125.3921	AQQLQQLVLANLAAYSQQH
50-60	1184.4243	LQQAIAASILR

Spot nla0808_8 corn seed embryo

Apparent mass: 21.0 apparent pI: 6.9

2.7962 0.324 1236.8 1 17/26 SW:OLE1_MAIZE R.GATGGGGGYGDLQR.G

SW:OLE1_MAIZE P13436 zea mays (maize). oleosin zm-i (oleosin 16 kd) (lipid body-associated major protein) (lipid
body-associated protein 13). 7/1998

MADHHRGATG GGGGYGDLQR GGGMHGEAQQ QQKQGAMMTA
LKAATAATFG GSMLVLSGLI LAGTVIALTV ATPVLVIFSP VLVPAAIALA
LMAAGFVTSG
GLGVAALSVF SWMYKYLTGK HPPAADQLDH AKARLASKAR DVKDAAQHRI
DQAQGS

Position	MH+	Sequence (link:NCBI Blast)
7-20	1266.3126	GATGGGGGYGDLQR

FIGURE 3G

Spot nla0907_1 corn seed embryo

Apparent mass: 60.0 apparent pI: 6.0

2.9706 0.127* 954.7 1 17/ 24 PIR2:T06212 K.VALVTGGDSGIGR.A

PIR2:T06212 glucose and ribitol dehydrogenase homolog - barley
MASQKFPPQQ QDCQPGKEHA MDPRPEAIIK NYKSANKLQG KVALVTGGDS
GIGRAVCLCL ALEGATVNFT YVKGHEDKDA EETLQALRDI KSRTGAGEPK
ALSGDLGYEE NCRRVVEEVA NAHGGRVDIL VNNAAEQYVR PCITEITEQD
LERVFRTNIF SYFLMTKFAV KHMGPGSSII NTTSVNAYKG NATLLDYTAT
KGAIVAFTRA LSMQLAEKGI RVNGVAPGPI WTPLIPASFP EEKVKQFGSE
VPMKRAGQPS EVAPSFVFLA SEQDSSYISG QILHPNGGTI VNS

Position	MH+	Sequence (link:NCBI Blast)
42-54	1202.353	VALVTGGDSGIGR

Spot nla0907_12 corn seed embryo

Apparent mass: 16.5 apparent pI: 7.0

3.7833 0.110 1211.9 1 27/ 48 SW:HS11_MEDSA +8 R.IDWKETPEAHVFK.A

SW:HS11_MEDSA P27879 medicago sativa (alfalfa). 18.1 kd class i heat shock protein (fragment). 4/1993
DPFSLDVWDP FKDFPFTNSA LSASSFPQEN SAFVSTRIDW KETPEAHVFK
ADLPGLKKEE VKVEIEDDRV LQISGERNVE KEDKNDQWHR VERSSGKFMR
RFRLPENAKM DQVKAAMENG VLTVTVPKEE IKKPEVKSIE ISS

Position	MH+	Sequence (link:NCBI Blast)
102-109	975.1355	FRLPENAK
38-50	1600.8143	IDWKETPEAHVFK

FIGURE 3H

Spot nla0907_16 corn seed embryo

Apparent mass: 21.5 apparent pI: 6.2

3.2768 0.167 1376.0 1 17/26 PIR2:T04358 K.GLAYEYLEQDLGNK.S

PIR2:T04358 glutathione transferase (EC 2.5.1.18) - maize
MAEEKKQGLQ LLDFWVSPFG QRCRIAMDEK **GLAYEYLEQD LGNKSELLLR**
ANPVHKKIPV LLHDGRPVCE SLVIVQYLDE AFPAAAPALL PADPYARAQA
RFWADYVDKK LYDCGTRLWK LKGDGQAQAR AEMVEILRTL EGALGDGPFF
GGDALGFVDV ALVPFTSWFL AYDRFGGVSV EKECPRLAAW AKRCAERPSV
AKNLYPPEKV YDFVCGMKKR LGIE

Position	MH+	Sequence (link:NCBI Blast)
31-44	1613.7642	GLAYEYLEQDLGNK

Spot nla0907_7 corn seed embryo

Apparent mass: 25.0 apparent pI: 6.6

5.4645 0.423 2528.5 1 39/100 PIR2:S05545
H.GHGATGHVDQYGNPVGGVEHGTGGMR.H

PIR2:S05545 dehydrin 3 - maize
MEYGQQGQHG **HGATGHVDQY GNPVGGVEHG TGGMRHGTGT**
GGMGQLGEHG GAGMGGGQFQ PAEEHKTGG ILHRSGSSSS SSSEDDGMGG
RRKKGIKEKIKEKLPGGHKD DQHATATTGG AYGQQGHTGS AYGQQGHTGG
AYATGTEGTG EKKGIMDKIK EKLPGQH

Position	MH+	Sequence (link:NCBI Blast)
10-35	2548.7119	GHGATGHVDQYGNPVGGVEHGTGGMR

Figure 4A

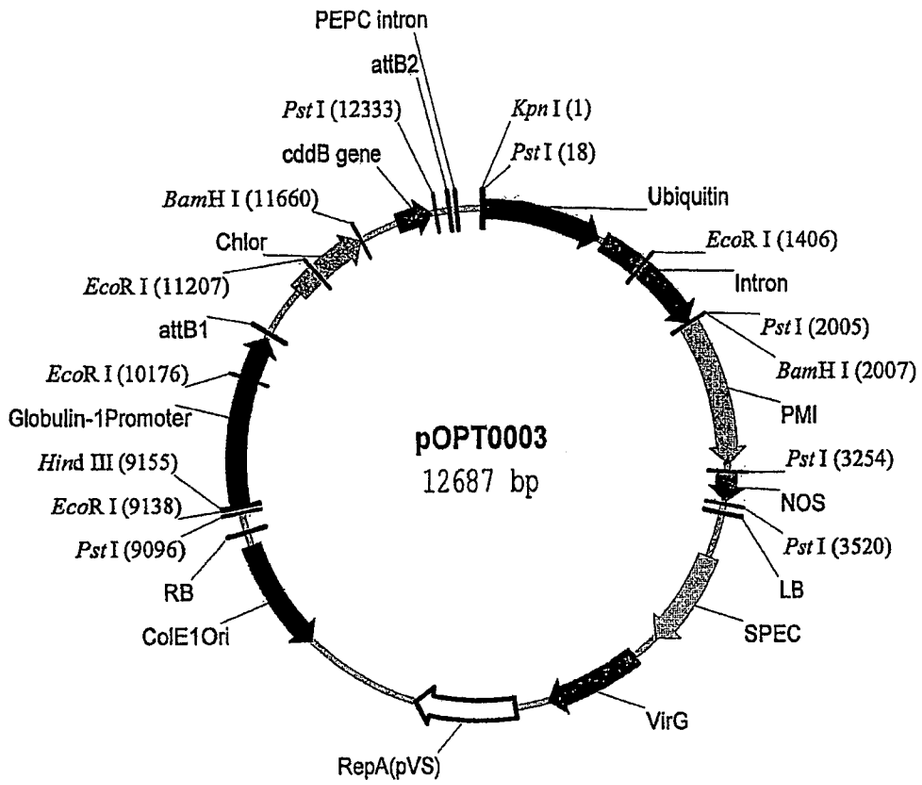

PEPC intron
attB2
Pst I (12333)
Kpn I (1)
cddB gene
Pst I (18)
BamH I (11660)
Ubiquitin
Chlor
EcoR I (1406)
EcoR I (11207)
Intron
attB1
Pst I (2005)
EcoR I (10176)
BamH I (2007)
Globulin-1Promoter
PMI
pOPT0003
12687 bp
Hind III (9155)
Pst I (3254)
EcoR I (9138)
NOS
Pst I (9096)
Pst I (3520)
RB
LB
ColE1Ori
SPEC
VirG
RepA(pVS)

Figure 4B

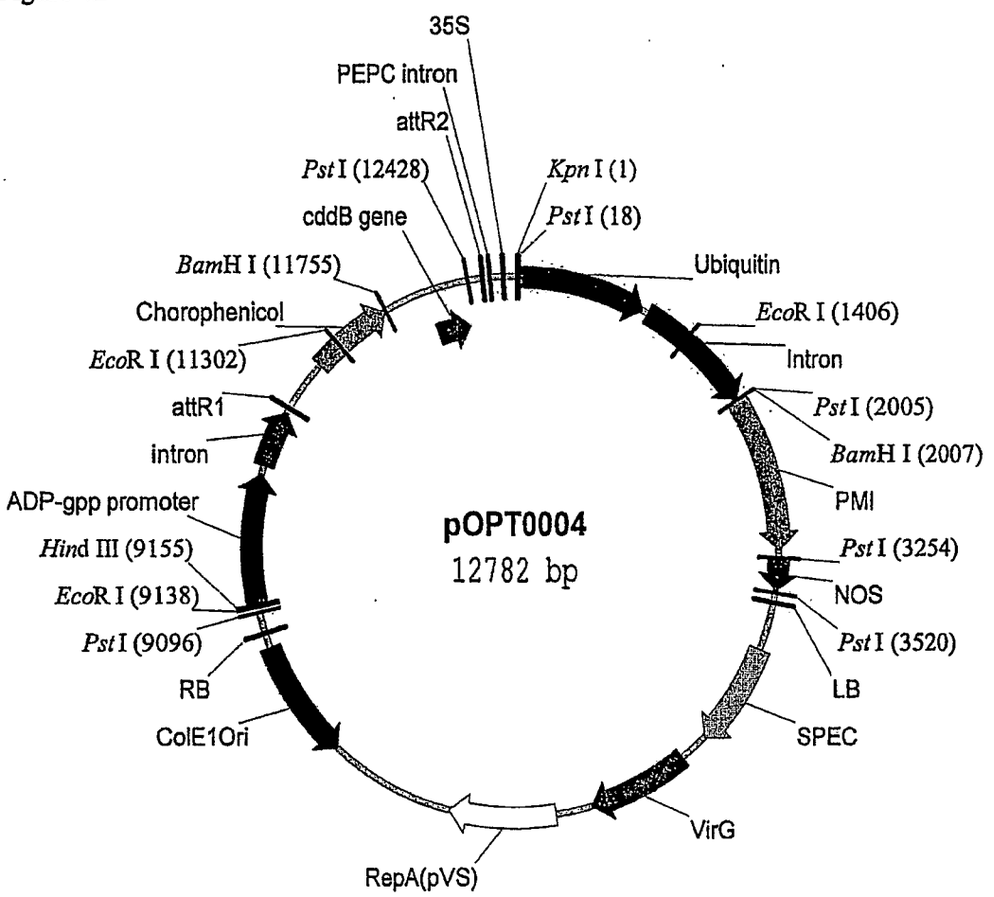
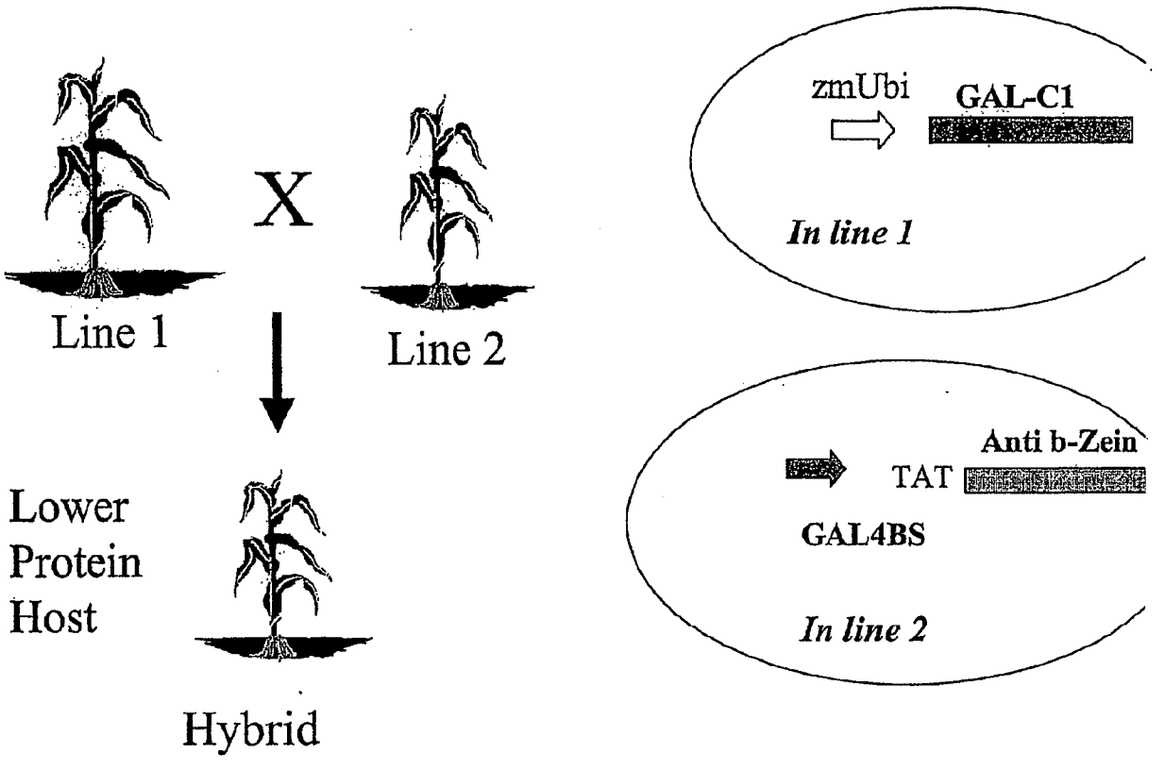

35S
PEPC intron
attR2
Pst I (12428)
Kpn I (1)
cddB gene
Pst I (18)
*Bam*H I (11755)
Ubiquitin
Chorophenicol
*Eco*R I (1406)
*Eco*R I (11302)
Intron
attR1
Pst I (2005)
intron
*Bam*H I (2007)
ADP-gpp promoter
PMI
pOPT0004
12782 bp
Hind III (9155)
Pst I (3254)
*Eco*R I (9138)
NOS
Pst I (9096)
Pst I (3520)
RB
LB
ColE1Ori
SPEC
VirG
RepA(pVS)

Figure 5
Producing lower storage protein maize grain



Sequence Listing

<110> Su, Wenpei
Andon, Nancy
Haynes, Paul
Briggs, Steven P.
Cooper, Bret
Glazebrook, Jane
Goff, Stephen A.
Katagiri, Fumiaki
Kreps, Joel
Moughamer, Todd
Provart, Nicholas
Ricke, Darrell
Zhu, Tong

<120> High-protein-phenotype-associated plant genes

<130> 60011-PCT

<140>

<141> 2002-09-26

<150> US 60/325,277
<151> 2001-09-26

<150> US 60/370,526
<151> 2002-04-04

<150> US 60/370,620
<151> 2002-04-04

<160> 45

<170> PatentList.pl version 3.0.4 (C) 2001 Syngenta

<210> 1
<211> 1722
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)..(1722)
<223> GLOBULIN-1 S ALLELE PRECURSOR
gene:GLB1-S
sp|P15590|GLB1_MAIZE

<400> 1
```
atg gtg agc gcc aga atc gtt gtc ctc ctc gcc gtc ctc cta tgc gct       48
Met Val Ser Ala Arg Ile Val Val Leu Leu Ala Val Leu Leu Cys Ala
1               5                   10                  15

gcc gcc gca gtc gcg tcg tcc tgg gag gac gac aac cac cac cac cac       96
Ala Ala Ala Val Ala Ser Ser Trp Glu Asp Asp Asn His His His His
            20                  25                  30

ggg ggc cac aag tcc ggg cga tgc gtg cgg cgg tgc gag gac cgg ccc      144
Gly Gly His Lys Ser Gly Arg Cys Val Arg Arg Cys Glu Asp Arg Pro
        35                  40                  45
```

```
tgg cac cag cgc ccc cgg tgc ctg gag cag tgc agg gag gag gag cgg      192
Trp His Gln Arg Pro Arg Cys Leu Glu Gln Cys Arg Glu Glu Glu Arg
    50                  55                  60

gag aag cgg caa gag cgg agc agg cac gag gcc gac gac cgc agc ggc      240
Glu Lys Arg Gln Glu Arg Ser Arg His Glu Ala Asp Asp Arg Ser Gly
65                  70                  75                  80

gag ggc tcg tcg gag gat gag cgc gag cgc gag cag gag aag gag gag      288
Glu Gly Ser Ser Glu Asp Glu Arg Glu Arg Glu Gln Glu Lys Glu Glu
                85                  90                  95

aag cag aag gac cgg cgg ccg tac gtg ttc gac cgg cgc agc ttt cgt      336
Lys Gln Lys Asp Arg Arg Pro Tyr Val Phe Asp Arg Arg Ser Phe Arg
            100                 105                 110

cgc gtg gtc cgg agc gag cag ggg tcc ctg agg gtg ctc cgg ccg ttc      384
Arg Val Val Arg Ser Glu Gln Gly Ser Leu Arg Val Leu Arg Pro Phe
        115                 120                 125

gac gag gtg tcc agg ctc ctc cgc ggc atc cgg gac tac cgc gtg gcg      432
Asp Glu Val Ser Arg Leu Leu Arg Gly Ile Arg Asp Tyr Arg Val Ala
    130                 135                 140

gtc ctg gag gcg aac ccg cgc tcg ttc gtg gtg ccc agc cac acc gac      480
Val Leu Glu Ala Asn Pro Arg Ser Phe Val Val Pro Ser His Thr Asp
145                 150                 155                 160

gcg cac tgc atc ggc tac gtg gcg gaa ggc gag gga gtg gtg acg acg      528
Ala His Cys Ile Gly Tyr Val Ala Glu Gly Glu Gly Val Val Thr Thr
                165                 170                 175

atc gag aac ggc gag agg cgg tcg tac acc atc aag caa ggc cac gtc      576
Ile Glu Asn Gly Glu Arg Arg Ser Tyr Thr Ile Lys Gln Gly His Val
            180                 185                 190

ttc gtg gcg ccg gcc ggg gcg gtc acc tac ctg gcc aac acc gac ggc      624
Phe Val Ala Pro Ala Gly Ala Val Thr Tyr Leu Ala Asn Thr Asp Gly
        195                 200                 205

cgg aag aaa ctg gtc atc acc aag atc ctc cat acc atc tcc gtg cct      672
Arg Lys Lys Leu Val Ile Thr Lys Ile Leu His Thr Ile Ser Val Pro
    210                 215                 220

ggc gag ttc cag ttc ttc ttc ggc ccc ggc ggg agg aac ccg gaa tcg      720
Gly Glu Phe Gln Phe Phe Phe Gly Pro Gly Gly Arg Asn Pro Glu Ser
225                 230                 235                 240

ttc ctg tcg agc ttc agc aag agc atc cag aga gct gcg tac aag acc      768
Phe Leu Ser Ser Phe Ser Lys Ser Ile Gln Arg Ala Ala Tyr Lys Thr
                245                 250                 255

tcg agc gac cgg ctg gag agg ctg ttc ggg agg cat ggg cag gac aag      816
Ser Ser Asp Arg Leu Glu Arg Leu Phe Gly Arg His Gly Gln Asp Lys
            260                 265                 270

ggg atc atc gtg cgt gcc acg gag gag cag acc cgc gag ctg cgg cgc      864
Gly Ile Ile Val Arg Ala Thr Glu Glu Gln Thr Arg Glu Leu Arg Arg
        275                 280                 285
```

```
cac gcc tcg gag ggc ggc cac ggc ccg cac tgg ccc ctg ccg ccg ttc      912
His Ala Ser Glu Gly Gly His Gly Pro His Trp Pro Leu Pro Pro Phe
    290                 295                 300

ggc gag tcg cgc ggc ccc tac agc ctc ctg gac cag cgg ccc agc atc      960
Gly Glu Ser Arg Gly Pro Tyr Ser Leu Leu Asp Gln Arg Pro Ser Ile
305                 310                 315                 320

gcc aac cag cac ggg cag ctc tac gag gcc gac gcg cgc agc ttc cac     1008
Ala Asn Gln His Gly Gln Leu Tyr Glu Ala Asp Ala Arg Ser Phe His
                325                 330                 335

gac ctc gcc gag cac gac gtc agc gtc tcc ttc gcc aac atc acc gcg     1056
Asp Leu Ala Glu His Asp Val Ser Val Ser Phe Ala Asn Ile Thr Ala
            340                 345                 350

ggg tcc atg agc gcg cca ttg tac aac acc cgt tcg ttc aag atc gcc     1104
Gly Ser Met Ser Ala Pro Leu Tyr Asn Thr Arg Ser Phe Lys Ile Ala
        355                 360                 365

tac gtg ccg aac ggc aag ggc tac gcc gag atc gtg tgc ccg cac cgc     1152
Tyr Val Pro Asn Gly Lys Gly Tyr Ala Glu Ile Val Cys Pro His Arg
    370                 375                 380

cag tcg cag ggc ggc gag agc gag cgc gag cgc ggc aag ggc agg agg     1200
Gln Ser Gln Gly Gly Glu Ser Glu Arg Glu Arg Gly Lys Gly Arg Arg
385                 390                 395                 400

agc gaa gaa gaa gaa gaa tcg tct gag gag cag gag gaa gtc ggg cag     1248
Ser Glu Glu Glu Glu Glu Ser Ser Glu Glu Gln Glu Glu Val Gly Gln
                405                 410                 415

ggg tac cac acc atc cgg gcg cgg ctg tca ccg ggc acg gcg ttc gtg     1296
Gly Tyr His Thr Ile Arg Ala Arg Leu Ser Pro Gly Thr Ala Phe Val
            420                 425                 430

gtg ccc gcg ggc cac ccg ttc gtc gcg gtg gcg tcc cgg gac agc aac     1344
Val Pro Ala Gly His Pro Phe Val Ala Val Ala Ser Arg Asp Ser Asn
        435                 440                 445

ctc cag atc gtg tgc ttc gag gtc cac gcc gac agg aac gag aag gtg     1392
Leu Gln Ile Val Cys Phe Glu Val His Ala Asp Arg Asn Glu Lys Val
    450                 455                 460

ttc ctg gcc ggc gcc gac aac gtg ctg cag aag ctc gac cgg gtc gcc     1440
Phe Leu Ala Gly Ala Asp Asn Val Leu Gln Lys Leu Asp Arg Val Ala
465                 470                 475                 480

aag gcg ctg tca ttc gcc tcc aag gcg gag gag gtg gac gag gtg ctc     1488
Lys Ala Leu Ser Phe Ala Ser Lys Ala Glu Glu Val Asp Glu Val Leu
                485                 490                 495

ggc tcg cgg cgc gag aag ggg ttc ctt cct ggc ccc aag gag agc ggc     1536
Gly Ser Arg Arg Glu Lys Gly Phe Leu Pro Gly Pro Lys Glu Ser Gly
            500                 505                 510

ggc cac gag gag cgg gag cag gag gag gag gaa cgc gaa gaa cgc cac     1584
Gly His Glu Glu Arg Glu Gln Glu Glu Glu Glu Arg Glu Glu Arg His
        515                 520                 525

ggc ggg cgt ggg gag agg gaa cgc cac gga cgt gag gag cgg gag aaa     1632
```

```
Gly Gly Arg Gly Glu Arg Glu Arg His Gly Arg Glu Glu Arg Glu Lys
    530                 535                 540

gag gag gag gaa cgc gaa gga cgc cac ggc cgc ggg cgc cgc gag gaa     1680
Glu Glu Glu Glu Arg Glu Gly Arg His Gly Arg Gly Arg Arg Glu Glu
545                 550                 555                 560

gtg gcg gag acg ctc ctg agg atg gtg acc gcc agg atg tga             1722
Val Ala Glu Thr Leu Leu Arg Met Val Thr Ala Arg Met
                565                 570
```

<210> 2
<211> 573
<212> PRT
<213> Zea mays

<400> 2

```
Met Val Ser Ala Arg Ile Val Val Leu Leu Ala Val Leu Leu Cys Ala
1               5                   10                  15

Ala Ala Ala Val Ala Ser Ser Trp Glu Asp Asp Asn His His His His
            20                  25                  30

Gly Gly His Lys Ser Gly Arg Cys Val Arg Arg Cys Glu Asp Arg Pro
        35                  40                  45

Trp His Gln Arg Pro Arg Cys Leu Glu Gln Cys Arg Glu Glu Glu Arg
    50                  55                  60

Glu Lys Arg Gln Glu Arg Ser Arg His Glu Ala Asp Asp Arg Ser Gly
65                  70                  75                  80

Glu Gly Ser Ser Glu Asp Glu Arg Glu Arg Glu Gln Glu Lys Glu Glu
                85                  90                  95

Lys Gln Lys Asp Arg Arg Pro Tyr Val Phe Asp Arg Arg Ser Phe Arg
            100                 105                 110

Arg Val Val Arg Ser Glu Gln Gly Ser Leu Arg Val Leu Arg Pro Phe
        115                 120                 125

Asp Glu Val Ser Arg Leu Leu Arg Gly Ile Arg Asp Tyr Arg Val Ala
    130                 135                 140

Val Leu Glu Ala Asn Pro Arg Ser Phe Val Val Pro Ser His Thr Asp
145                 150                 155                 160

Ala His Cys Ile Gly Tyr Val Ala Glu Gly Glu Gly Val Val Thr Thr
                165                 170                 175
```

Ile Glu Asn Gly Glu Arg Arg Ser Tyr Thr Ile Lys Gln Gly His Val
180 185 190

Phe Val Ala Pro Ala Gly Ala Val Thr Tyr Leu Ala Asn Thr Asp Gly
195 200 205

Arg Lys Lys Leu Val Ile Thr Lys Ile Leu His Thr Ile Ser Val Pro
210 215 220

Gly Glu Phe Gln Phe Phe Phe Gly Pro Gly Gly Arg Asn Pro Glu Ser
225 230 235 240

Phe Leu Ser Ser Phe Ser Lys Ser Ile Gln Arg Ala Ala Tyr Lys Thr
245 250 255

Ser Ser Asp Arg Leu Glu Arg Leu Phe Gly Arg His Gly Gln Asp Lys
260 265 270

Gly Ile Ile Val Arg Ala Thr Glu Glu Gln Thr Arg Glu Leu Arg Arg
275 280 285

His Ala Ser Glu Gly Gly His Gly Pro His Trp Pro Leu Pro Pro Phe
290 295 300

Gly Glu Ser Arg Gly Pro Tyr Ser Leu Leu Asp Gln Arg Pro Ser Ile
305 310 315 320

Ala Asn Gln His Gly Gln Leu Tyr Glu Ala Asp Ala Arg Ser Phe His
325 330 335

Asp Leu Ala Glu His Asp Val Ser Val Ser Phe Ala Asn Ile Thr Ala
340 345 350

Gly Ser Met Ser Ala Pro Leu Tyr Asn Thr Arg Ser Phe Lys Ile Ala
355 360 365

Tyr Val Pro Asn Gly Lys Gly Tyr Ala Glu Ile Val Cys Pro His Arg
370 375 380

Gln Ser Gln Gly Gly Glu Ser Glu Arg Glu Arg Gly Lys Gly Arg Arg
385 390 395 400

Ser Glu Glu Glu Glu Glu Ser Ser Glu Glu Gln Glu Glu Val Gly Gln
405 410 415

Gly Tyr His Thr Ile Arg Ala Arg Leu Ser Pro Gly Thr Ala Phe Val
420 425 430

Val Pro Ala Gly His Pro Phe Val Ala Val Ala Ser Arg Asp Ser Asn
435 440 445

Leu Gln Ile Val Cys Phe Glu Val His Ala Asp Arg Asn Glu Lys Val
450 455 460

Phe Leu Ala Gly Ala Asp Asn Val Leu Gln Lys Leu Asp Arg Val Ala
465 470 475 480

Lys Ala Leu Ser Phe Ala Ser Lys Ala Glu Glu Val Asp Glu Val Leu
485 490 495

Gly Ser Arg Arg Glu Lys Gly Phe Leu Pro Gly Pro Lys Glu Ser Gly
500 505 510

Gly His Glu Glu Arg Glu Gln Glu Glu Glu Glu Arg Glu Glu Arg His
515 520 525

Gly Gly Arg Gly Glu Arg Glu Arg His Gly Arg Glu Glu Arg Glu Lys
530 535 540

Glu Glu Glu Glu Arg Glu Gly Arg His Gly Arg Gly Arg Arg Glu Glu
545 550 555 560

Val Ala Glu Thr Leu Leu Arg Met Val Thr Ala Arg Met
565 570

<210> 3
<211> 2003
<212> DNA
<213> Zea mays

<220>
<221> misc_feature
<223> Globulin-2 precursor

<400> 3
cgcacacacc cgagcatatc acagtgacac tacacgatgg tgagcgccag aatcgttgtc 60

ctcctcgccg tcctcctatg cgctgccgcc gcagtcgcgt cgtcctggga ggacgacaac 120

caccaccacc acgggggcca caagtccggg cgatgcgtgc ggcggtgcga ggaccggccc 180

tggcaccagc gcccccggtg cctggagcag tgcagggagg aggagcggga gaagcggcaa 240

gagcggagca ggcacgaggc cgacgaccgc agcggcgagg gctcgtcgga ggatgagcgc 300

gagcgcgagc aggagaagga ggagaagcag aaggaccggc ggccgtacgt gttcgaccgg 360
cgcagctttc gtcgcgtggt ccggagcgag caggggtccc tgagggtgct ccggccgttc 420
gacgaggtgt ccaggctcct ccgcggcatc cgggactacc gcgtggcggt cctggaggcg 480
aacccgcgct cgttcgtggt gcccagccac accgacgcgc actgcatcgg ctacgtggcg 540
gaaggcgagg gagtggtgac gacgatcgag aacggcgaga ggcggtcgta caccatcaag 600
caaggccacg tcttcgtggc gccggccggg gcggtcacct acctggccaa caccgacggc 660
cggaagaaac tggtcatcac caagatcctc cataccatct ccgtgcctgg cgagttccag 720
ttcttcttcg gccccggcgg gaggaacccg gaatcgttcc tgtcgagctt cagcaagagc 780
atccagagag ctgcgtacaa gacctcgagc gaccggctgg agaggctgtt cgggaggcat 840
gggcaggaca aggggatcat cgtgcgtgcc acggaggagc agacccgcga gctgcggcgc 900
cacgcctcgg agggcggcca cggcccgcac tggcccctgc cgccgttcgg cgagtcgcgc 960
ggcccctaca gcctcctgga ccagcggccc agcatcgcca accagcacgg gcagctctac 1020
gaggccgacg cgcgcagctt ccacgacctc gccgagcacg acgtcagcgt ctccttcgcc 1080
aacatcaccg cggggtccat gagcgcgcca ttgtacaaca cccgttcgtt caagatcgcc 1140
tacgtgccga acggcaaggg ctacgccgag atcgtgtgcc cgcaccgcca gtcgcagggc 1200
ggcgagagcg agcgcgagcg cggcaagggc aggaggagcg aagaagaaga agaatcgtct 1260
gaggagcagg aggaagtcgg gcaggggtac cacaccatcc gggcgcggct gtcaccgggc 1320
acggcgttcg tggtgcccgc gggccacccg ttcgtcgcgg tggcgtcccg ggacagcaac 1380
ctccagatcg tgtgcttcga ggtccacgcc gacaggaacg agaaggtgtt cctggccggc 1440
gccgacaacg tgctgcagaa gctcgaccgg gtcgccaagg cgctgtcatt cgcctccaag 1500
gcggaggagg tggacgaggt gctcggctcg cggcgcgaga aggggttcct tcctggcccc 1560
aaggagagcg gcggccacga ggagcgggag caggaggagg aggaacgcga agaacgccac 1620
ggcgggcgtg gggagaggga acgccacgga cgtgaggagc gggagaaaga ggaggaggaa 1680
cgcgaaggac gccacggccg cgggcgccgc gaggaagtgg cggagacgct cctgaggatg 1740
gtgaccgcca ggatgtgagg ccggccgtgc tcgccaaaac gagcaggaag caacgagagg 1800
gtggcgcgcg accgacgtgc gtacgtagca tgagcctgag tggagacgtt ggacgtgtat 1860
gtatatacct ctctgcgtgt taactatgta cgtaagcggc aggcagtgca ataagtgtgg 1920
ctctgtagta tgtacgtgcg ggtacgatgc tgtaagctac tgaggcaagt ccataaataa 1980
ataatgacac gtgcgtgttc tat 2003

<210> 4

<211> 450
<212> PRT
<213> Zea mays

<220>
<221> misc_feature
<223> Globulin-2 precursor

<400> 4

Met Lys Val Pro Val Leu Leu Leu Leu Val Ser Leu Cys Phe Ser Leu
1 5 10 15

Ala Leu Ala Trp Gln Thr Asp Thr Glu Ser Gly Ser Gly Arg Pro Tyr
20 25 30

His Tyr Gly Glu Glu Ser Phe Arg His Trp Thr Arg Ser Arg Gln Gly
35 40 45

Arg Phe Arg Val Leu Glu Arg Phe Thr His Glu Leu Leu Glu Asp Ala
50 55 60

Val Gly Asn Tyr Arg Val Ala Glu Leu Glu Ala Ala Pro Arg Ala Phe
65 70 75 80

Leu Gln Pro Ser His Tyr Asp Ala Asp Glu Val Met Phe Val Lys Glu
85 90 95

Gly Glu Gly Val Ile Val Leu Leu Arg Gly Gly Lys Arg Glu Ser Phe
100 105 110

Cys Val Arg Glu Gly Asp Val Met Val Ile Pro Ala Gly Ala Val Val
115 120 125

Tyr Ser Ala Asn Thr His Gln Ser Glu Trp Phe Arg Val Val Met Leu
130 135 140

Leu Ser Pro Val Val Ser Thr Ser Gly Arg Phe Glu Glu Phe Phe Pro
145 150 155 160

Ile Gly Gly Glu Ser Pro Glu Ser Phe Leu Ser Val Phe Ser Asp Asp
165 170 175

Val Ile Gln Ala Ser Phe Asn Thr Arg Arg Glu Glu Trp Glu Lys Val
180 185 190

Phe Glu Lys Gln Ser Lys Gly Glu Ile Thr Thr Ala Ser Glu Glu Gln
195 200 205

Ile Arg Glu Leu Ser Arg Ser Cys Ser Arg Gly Gly Arg Ser Ser Arg
210 215 220

Ser Glu Gly Gly Asp Ser Gly Ser Ser Ser Ser Lys Trp Glu Ile Lys
225 230 235 240

Pro Ser Ser Leu Thr Asp Lys Lys Pro Thr His Ser Asn Ser His Gly
245 250 255

Arg His Tyr Glu Ile Thr Gly Asp Glu Cys Pro His Leu Arg Leu Leu
260 265 270

Asp Met Asp Val Gly Leu Ala Asn Ile Ala Arg Gly Ser Met Met Ala
275 280 285

Pro Ser Tyr Asn Thr Arg Ala Asn Lys Ile Ala Ile Val Leu Lys Gly
290 295 300

Gln Gly Tyr Phe Glu Met Ala Cys Pro His Val Ser Gly Gly Arg Ser
305 310 315 320

Ser Pro Arg Arg Glu Arg Gly His Gly Arg Glu Glu Glu Glu Glu Arg
325 330 335

Glu Glu Glu Gln Gly Gly Gly Gly Gly Gln Lys Ser Arg Ser Tyr Arg
340 345 350

Gln Val Lys Ser Arg Ile Arg Glu Gly Ser Val Ile Val Ile Pro Ala
355 360 365

Gly His Pro Thr Ala Leu Val Ala Gly Glu Asp Lys Asn Leu Ala Val
370 375 380

Leu Cys Phe Glu Val Asn Ala Ser Phe Asp Asp Lys Val Phe Leu Ala
385 390 395 400

Gly Thr Asn Ser Ala Leu Gln Lys Met Asp Arg Pro Ala Lys Leu Leu
405 410 415

Ala Phe Gly Ala Asp Glu Glu Gln Gln Val Asp Arg Val Ile Gly Ala
420 425 430

Gln Lys Asp Ala Val Phe Leu Arg Gly Pro Gln Ser His Arg Val Ser
435 440 445

Ser Val
450

<210> 5
<211> 723
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)..(723)
<223> Oleosin ZM-I

<400> 5

```
atg gca gcc aag att ttt gcc ctc ctt gcc ctc ctt gct ctt tca gca       48
Met Ala Ala Lys Ile Phe Ala Leu Leu Ala Leu Leu Ala Leu Ser Ala
1               5                  10                  15

aac gtt gct acc gcg act att att cca caa tgc tca caa caa tac ctc       96
Asn Val Ala Thr Ala Thr Ile Ile Pro Gln Cys Ser Gln Gln Tyr Leu
            20                  25                  30

tct ccg gtg aca gcc gcg aga ttt gaa tac cca act ata caa tcc tac      144
Ser Pro Val Thr Ala Ala Arg Phe Glu Tyr Pro Thr Ile Gln Ser Tyr
        35                  40                  45

agg cta caa cag gcc atc gca gca agc atc tta cgg tcg tta gca ttg      192
Arg Leu Gln Gln Ala Ile Ala Ala Ser Ile Leu Arg Ser Leu Ala Leu
    50                  55                  60

act gtc caa caa cca tat gcc cta ttg caa caa cca tcc tta gtg aat      240
Thr Val Gln Gln Pro Tyr Ala Leu Leu Gln Gln Pro Ser Leu Val Asn
65                  70                  75                  80

cta tat ctc caa aga atc gta gca caa caa cta caa caa caa ttg ctt      288
Leu Tyr Leu Gln Arg Ile Val Ala Gln Gln Leu Gln Gln Gln Leu Leu
                85                  90                  95

cca aca atc aat caa gta gtt gca gcg aac ctt gat gct tac ctc cag      336
Pro Thr Ile Asn Gln Val Val Ala Ala Asn Leu Asp Ala Tyr Leu Gln
            100                 105                 110

caa caa caa ttt ctt cca ttc aat caa cta gct ggg gtg aac cct gct      384
Gln Gln Gln Phe Leu Pro Phe Asn Gln Leu Ala Gly Val Asn Pro Ala
        115                 120                 125

gct tac ttg cag gca caa cag cta cta cca ttc aac caa ctt gtc agg      432
Ala Tyr Leu Gln Ala Gln Gln Leu Leu Pro Phe Asn Gln Leu Val Arg
    130                 135                 140

agc cct gct gcc ttc tta ctg cag caa cag ttg ttg cca ttc cat cta      480
Ser Pro Ala Ala Phe Leu Leu Gln Gln Gln Leu Leu Pro Phe His Leu
145                 150                 155                 160

caa gtt gtg gca aac att gct gct ttc ttg caa caa caa caa ttg ctg      528
Gln Val Val Ala Asn Ile Ala Ala Phe Leu Gln Gln Gln Gln Leu Leu
                165                 170                 175

cca ttt tac cca cag gtt gtg gga aac att aac gcc ttc ttg caa cag      576
```

```
Pro Phe Tyr Pro Gln Val Val Gly Asn Ile Asn Ala Phe Leu Gln Gln
            180                 185                 190

caa cag ttg ctg cca ttc tac cca cag gat gtg gca aac aat gtc gcc      624
Gln Gln Leu Leu Pro Phe Tyr Pro Gln Asp Val Ala Asn Asn Val Ala
        195                 200                 205

ttc tta caa caa caa caa ttg ctg cca ttt agc caa ctt gct ttg acg      672
Phe Leu Gln Gln Gln Gln Leu Leu Pro Phe Ser Gln Leu Ala Leu Thr
    210                 215                 220

aat cct acc acc tta ttg cag cag ccc acc att ggt ggt gcc atc ttc      720
Asn Pro Thr Thr Leu Leu Gln Gln Pro Thr Ile Gly Gly Ala Ile Phe
225                 230                 235                 240

tag                                                                  723
```

<210> 6
<211> 240
<212> PRT
<213> Zea mays

<400> 6

```
Met Ala Ala Lys Ile Phe Ala Leu Leu Ala Leu Leu Ala Leu Ser Ala
1               5                   10                  15

Asn Val Ala Thr Ala Thr Ile Ile Pro Gln Cys Ser Gln Gln Tyr Leu
            20                  25                  30

Ser Pro Val Thr Ala Ala Arg Phe Glu Tyr Pro Thr Ile Gln Ser Tyr
        35                  40                  45

Arg Leu Gln Gln Ala Ile Ala Ala Ser Ile Leu Arg Ser Leu Ala Leu
    50                  55                  60

Thr Val Gln Gln Pro Tyr Ala Leu Leu Gln Gln Pro Ser Leu Val Asn
65                  70                  75                  80

Leu Tyr Leu Gln Arg Ile Val Ala Gln Gln Leu Gln Gln Gln Leu Leu
                85                  90                  95

Pro Thr Ile Asn Gln Val Val Ala Ala Asn Leu Asp Ala Tyr Leu Gln
            100                 105                 110

Gln Gln Gln Phe Leu Pro Phe Asn Gln Leu Ala Gly Val Asn Pro Ala
        115                 120                 125

Ala Tyr Leu Gln Ala Gln Gln Leu Leu Pro Phe Asn Gln Leu Val Arg
    130                 135                 140
```

```
Ser Pro Ala Ala Phe Leu Leu Gln Gln Gln Leu Leu Pro Phe His Leu
145                 150                 155                 160

Gln Val Val Ala Asn Ile Ala Ala Phe Leu Gln Gln Gln Gln Leu Leu
                165                 170                 175

Pro Phe Tyr Pro Gln Val Val Gly Asn Ile Asn Ala Phe Leu Gln Gln
            180                 185                 190

Gln Gln Leu Leu Pro Phe Tyr Pro Gln Asp Val Ala Asn Asn Val Ala
        195                 200                 205

Phe Leu Gln Gln Gln Gln Leu Leu Pro Phe Ser Gln Leu Ala Leu Thr
    210                 215                 220

Asn Pro Thr Thr Leu Leu Gln Gln Pro Thr Ile Gly Gly Ala Ile Phe
225                 230                 235                 240
```

<210> 7
<211> 740
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (79)..(534)
<223> 17.2 kD Heat Shock Protein

<400> 7

```
aacacgagcc cgaagcactc ttgcaatcca ctgagttctg tttgttgaga cgcatagagc      60

tagctgctag cgtcgaca atg tcg ctc gtg agg cgc agc aac gtg ttc gac      111
                       Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp
                       1               5                   10

ccc ttc tcg atg gac ctc tgg gat ccc ttc gac acc atg ttc cgc tcc      159
Pro Phe Ser Met Asp Leu Trp Asp Pro Phe Asp Thr Met Phe Arg Ser
            15                  20                  25

atc gtc ccg tcg gcg acc tcc acc aac tcc gag act gcc gcc ttc gcc      207
Ile Val Pro Ser Ala Thr Ser Thr Asn Ser Glu Thr Ala Ala Phe Ala
        30                  35                  40

agc gcc cgc atc gac tgg aag gag acg ccc gag gcg cac gtc ttc aag      255
Ser Ala Arg Ile Asp Trp Lys Glu Thr Pro Glu Ala His Val Phe Lys
    45                  50                  55

gcc gac ctc ccc ggc gtc aag aag gag gag gtc aag gtt gag gtc gaa      303
Ala Asp Leu Pro Gly Val Lys Lys Glu Glu Val Lys Val Glu Val Glu
60                  65                  70                  75

gac ggc aac gtg ctg gtc atc agc ggc cag cgc agc agg gag aag gag      351
Asp Gly Asn Val Leu Val Ile Ser Gly Gln Arg Ser Arg Glu Lys Glu
```

```
                80                  85                  90

gac aag gac gac aag tgg cac cgt gtc gag cgc agc agt ggc cag ttc      399
Asp Lys Asp Asp Lys Trp His Arg Val Glu Arg Ser Ser Gly Gln Phe
            95                  100                 105

atc agg cgc ttc cgc ctg ccg gat gac gcc aag gtg gat cag gtc aag      447
Ile Arg Arg Phe Arg Leu Pro Asp Asp Ala Lys Val Asp Gln Val Lys
        110                 115                 120

gct ggc ctc gag aac ggc gtg ctc acg gtc acc gtg cct aag gcg gaa      495
Ala Gly Leu Glu Asn Gly Val Leu Thr Val Thr Val Pro Lys Ala Glu
    125                 130                 135

gag aag aag cct gag gtg aag gct att gag atc tct ggt tgagcatcca      544
Glu Lys Lys Pro Glu Val Lys Ala Ile Glu Ile Ser Gly
140                 145                 150

atccaatatg gacgtggatg aaggtgtact gctgctggtc cgtggctgtc gctgtcctgt    604

gtggatgttt cctgtatctt ctacagtata taatgtactt ccgtctgttt cgtttgtatg    664

tacaatctca atcttgcggg tatcgttcat gtatcccttt gaataataac aaataaaatc    724

gggtttgtca cggtaa                                                   740
```

<210> 8
<211> 152
<212> PRT
<213> Zea mays

<400> 8

```
Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp Pro Phe Ser Met Asp
1               5                   10                  15

Leu Trp Asp Pro Phe Asp Thr Met Phe Arg Ser Ile Val Pro Ser Ala
            20                  25                  30

Thr Ser Thr Asn Ser Glu Thr Ala Ala Phe Ala Ser Ala Arg Ile Asp
        35                  40                  45

Trp Lys Glu Thr Pro Glu Ala His Val Phe Lys Ala Asp Leu Pro Gly
    50                  55                  60

Val Lys Lys Glu Glu Val Lys Val Glu Val Glu Asp Gly Asn Val Leu
65                  70                  75                  80

Val Ile Ser Gly Gln Arg Ser Arg Glu Lys Glu Asp Lys Asp Asp Lys
                85                  90                  95

Trp His Arg Val Glu Arg Ser Ser Gly Gln Phe Ile Arg Arg Phe Arg
            100                 105                 110
```

PS6

```
Leu Pro Asp Asp Ala Lys Val Asp Gln Val Lys Ala Gly Leu Glu Asn
        115                 120                 125

Gly Val Leu Thr Val Thr Val Pro Lys Ala Glu Glu Lys Lys Pro Glu
    130                 135                 140

Val Lys Ala Ile Glu Ile Ser Gly
145                 150
```

<210> 9

<211> 469

<212> DNA

<213> Oryza sativa

<220>

<221> CDS

<222> (1)..(450)

<223> 17.2 kD Heat Shock Protein

<400> 9

```
atg tcg ctg gtg agg cgc agc aac gtg ttc gac cca ttc tcc ctc gac      48
Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp Pro Phe Ser Leu Asp
1               5                   10                  15

ctc tgg gac ccc ttc gac agc gtg ttc cgc tcc gtc gtc ccg gcc acc      96
Leu Trp Asp Pro Phe Asp Ser Val Phe Arg Ser Val Val Pro Ala Thr
            20                  25                  30

tcc gac aac gac acc gcc gcc ttc gcc aac gcc cgc atc gac tgg aag     144
Ser Asp Asn Asp Thr Ala Ala Phe Ala Asn Ala Arg Ile Asp Trp Lys
        35                  40                  45

gag acg ccg gag tcg cac gtc ttc aag gcc gac ctc ccc ggc gtc aag     192
Glu Thr Pro Glu Ser His Val Phe Lys Ala Asp Leu Pro Gly Val Lys
    50                  55                  60

aag gag gag gtg aag gtg gag gtg gag gaa ggc aac gtg ctg gtg atc     240
Lys Glu Glu Val Lys Val Glu Val Glu Glu Gly Asn Val Leu Val Ile
65                  70                  75                  80

agc ggg cag cgc agc aag gag aag gag gac aag aac gac aag tgg cac     288
Ser Gly Gln Arg Ser Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp His
                85                  90                  95

cgc gtg gag cgc agc agc ggg cag ttc atg cgg cgg ttc cgg ctg ccg     336
Arg Val Glu Arg Ser Ser Gly Gln Phe Met Arg Arg Phe Arg Leu Pro
            100                 105                 110
```

```
gag aac gcc aag gtg gac cag gtg aag gcc ggc atg gag aac ggc gtg      384
Glu Asn Ala Lys Val Asp Gln Val Lys Ala Gly Met Glu Asn Gly Val
        115                 120                 125

ctc acc gtc acc gtg ccc aag gcc gag gtc aag aag ccc gag gtg aag      432
Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val Lys
    130                 135                 140

gcc att gag atc tct ggc taaaatggtg aaaacggga                        469
Ala Ile Glu Ile Ser Gly
145                 150
```

<210> 10

<211> 150

<212> PRT

<213> Oryza sativa

<400> 10

```
Met Ser Leu Val Arg Arg Ser Asn Val Phe Asp Pro Phe Ser Leu Asp
1               5                   10                  15

Leu Trp Asp Pro Phe Asp Ser Val Phe Arg Ser Val Val Pro Ala Thr
            20                  25                  30

Ser Asp Asn Asp Thr Ala Ala Phe Ala Asn Ala Arg Ile Asp Trp Lys
        35                  40                  45

Glu Thr Pro Glu Ser His Val Phe Lys Ala Asp Leu Pro Gly Val Lys
    50                  55                  60

Lys Glu Glu Val Lys Val Glu Val Glu Glu Gly Asn Val Leu Val Ile
65                  70                  75                  80

Ser Gly Gln Arg Ser Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp His
                85                  90                  95

Arg Val Glu Arg Ser Ser Gly Gln Phe Met Arg Arg Phe Arg Leu Pro
            100                 105                 110

Glu Asn Ala Lys Val Asp Gln Val Lys Ala Gly Met Glu Asn Gly Val
        115                 120                 125

Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val Lys
    130                 135                 140
```

Ala Ile Glu Ile Ser Gly
145 150

<210> 11

<211> 460

<212> DNA

<213> wheat

<220>

<221> CDS

<222> (1)..(453)

<223>

<400> 11

```
atg tcg atc gtg agg cgg acg aac gtg ttc gac ccc ttc gcc gac ctc       48
Met Ser Ile Val Arg Arg Thr Asn Val Phe Asp Pro Phe Ala Asp Leu
1               5                  10                  15

tgg gcg gac ccc ttc gac acc ttc cgc tcc atc gtc ccg gcg atc tca       96
Trp Ala Asp Pro Phe Asp Thr Phe Arg Ser Ile Val Pro Ala Ile Ser
            20                  25                  30

ggc ggc ggc agc gag acg gct gcg ttc gcc aac gcc cgg atg gac tgg      144
Gly Gly Gly Ser Glu Thr Ala Ala Phe Ala Asn Ala Arg Met Asp Trp
        35                  40                  45

aag gag acc ccc gaa gcg cac gtc ttc aag gcc gac ctc ccc ggc gtg      192
Lys Glu Thr Pro Glu Ala His Val Phe Lys Ala Asp Leu Pro Gly Val
    50                  55                  60

aag aag gag gag gtc aag gtg gag gtg gag gac ggc aac gtg ctc gtc      240
Lys Lys Glu Glu Val Lys Val Glu Val Glu Asp Gly Asn Val Leu Val
65                  70                  75                  80

gtc agc ggc gag cgt aca aag gag aag gag gac aag aac gac aag tgg      288
Val Ser Gly Glu Arg Thr Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp
                85                  90                  95

cac cgc gtg gag cgc agc agc ggc aag ttc gtg cgg cgc ttc cgg ctg      336
His Arg Val Glu Arg Ser Ser Gly Lys Phe Val Arg Arg Phe Arg Leu
            100                 105                 110

ctg gag gac gcc aag gtg gag gag gtg aag gcc ggg ctg gag aac ggg      384
Leu Glu Asp Ala Lys Val Glu Glu Val Lys Ala Gly Leu Glu Asn Gly
        115                 120                 125

gtg ctc acc gtc acc gtg ccc aag gcc gag gtc aag aag ccc gag gtg      432
```

Val Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val
130 135 140

aag gcc atc cag atc tcc ggc tgagtat 460
Lys Ala Ile Gln Ile Ser Gly
145 150

<210> 12

<211> 151

<212> PRT

<213> wheat

<400> 12

Met Ser Ile Val Arg Arg Thr Asn Val Phe Asp Pro Phe Ala Asp Leu
1 5 10 15

Trp Ala Asp Pro Phe Asp Thr Phe Arg Ser Ile Val Pro Ala Ile Ser
20 25 30

Gly Gly Gly Ser Glu Thr Ala Ala Phe Ala Asn Ala Arg Met Asp Trp
35 40 45

Lys Glu Thr Pro Glu Ala His Val Phe Lys Ala Asp Leu Pro Gly Val
50 55 60

Lys Lys Glu Glu Val Lys Val Glu Val Glu Asp Gly Asn Val Leu Val
65 70 75 80

Val Ser Gly Glu Arg Thr Lys Glu Lys Glu Asp Lys Asn Asp Lys Trp
85 90 95

His Arg Val Glu Arg Ser Ser Gly Lys Phe Val Arg Arg Phe Arg Leu
100 105 110

Leu Glu Asp Ala Lys Val Glu Glu Val Lys Ala Gly Leu Glu Asn Gly
115 120 125

Val Leu Thr Val Thr Val Pro Lys Ala Glu Val Lys Lys Pro Glu Val
130 135 140

Lys Ala Ile Gln Ile Ser Gly
145 150

<210> 13

<211> 876
<212> DNA
<213> Hordeum vulgare

<220>
<221> CDS
<222> (1)..(876)
<223> PIR2 Accesion T06212
Glucose and Ribitol Dehydrogenoase Homolog

<400> 13

```
atg gcg tcg cag aag ttc ccg ccg cag cag cag gac tgc cag ccc ggc        48
Met Ala Ser Gln Lys Phe Pro Pro Gln Gln Gln Asp Cys Gln Pro Gly
1               5                   10                  15

aag gag cac gcc atg gac ccc cgc ccc gag gcc atc atc aag aac tac        96
Lys Glu His Ala Met Asp Pro Arg Pro Glu Ala Ile Ile Lys Asn Tyr
            20                  25                  30

aag tcg ggc caa caa gct cca ggg caa ggt ggc gct ggt gac cgg cgg       144
Lys Ser Gly Gln Gln Ala Pro Gly Gln Gly Gly Ala Gly Asp Arg Arg
        35                  40                  45

cga ctc ggg cat cgg gcg cgc ggt gtg cct gtg cct cgc gct gga ggg       192
Arg Leu Gly His Arg Ala Arg Gly Val Pro Val Pro Arg Ala Gly Gly
    50                  55                  60

cgc gac ggt gaa ctt cac gta cgt gaa ggg gca cga gga caa gga cgc       240
Arg Asp Gly Glu Leu His Val Arg Glu Gly Ala Arg Gly Gln Gly Arg
65                  70                  75                  80

gga gga gac cct gca ggc gct ccg cga cat caa gtc ccg cac cgg cgc       288
Gly Gly Asp Pro Ala Gly Ala Pro Arg His Gln Val Pro His Arg Arg
                85                  90                  95

cgg cga gcc caa ggc gct ctc ggg cga cct cgg gta cga gga gaa ctg       336
Arg Arg Ala Gln Gly Ala Leu Gly Arg Pro Arg Val Arg Gly Glu Leu
            100                 105                 110

ccg cag ggt ggt gga gga ggt ggc caa cgc gca cgg cgg ccg cgt gga       384
Pro Gln Gly Gly Gly Gly Gly Gly Gln Arg Ala Arg Arg Pro Arg Gly
        115                 120                 125

cat cct cgt gaa caa cgc ggc cga gca gta cgt ccg ccc ctg cat cac       432
His Pro Arg Glu Gln Arg Gly Arg Ala Val Arg Pro Pro Leu His His
    130                 135                 140

cga gat cac cga gca gga cct gga gcg cgt gtt ccg cac caa cat ctt       480
Arg Asp His Arg Ala Gly Pro Gly Ala Arg Val Pro His Gln His Leu
145                 150                 155                 160

ctc cta ctt cct cat gac caa gtt cgc cgt gaa gca cat ggg gcc cgg       528
Leu Leu Leu Pro His Asp Gln Val Arg Arg Glu Ala His Gly Ala Arg
                165                 170                 175

gtc cag cat cat caa cac cac ctc cgt gaa cgc gta caa ggg caa cgc       576
Val Gln His His Gln His His Leu Arg Glu Arg Val Gln Gly Gln Arg
            180                 185                 190

gac gct gct gga cta cac ggc cac caa ggg cgc cat cgt ggc ctt cac       624
Asp Ala Ala Gly Leu His Gly His Gln Gly Arg His Arg Gly Leu His
```

```
        195                 200                 205

ccg cgc gct gtc gat gca gct ggc gga gaa ggg gat ccg cgt caa cgg      672
Pro Arg Ala Val Asp Ala Ala Gly Gly Glu Gly Asp Pro Arg Gln Arg
    210                 215                 220

cgt ggc gcc ggg gcc cat ctg gac gcc cct cat ccc ggc ctc ctt ccc      720
Arg Gly Ala Gly Ala His Leu Asp Ala Pro His Pro Gly Leu Leu Pro
225                 230                 235                 240

gga gga gaa ggt gaa gca gtt cgg gtc cga ggt gcc cat gaa gcg cgc      768
Gly Gly Glu Gly Glu Ala Val Arg Val Arg Gly Ala His Glu Ala Arg
                245                 250                 255

cat gca gcc cag cga ggt cgc gcc cag ctt cgt ctt cct tgc cag cga      816
His Ala Ala Gln Arg Gly Arg Ala Gln Leu Arg Leu Pro Cys Gln Arg
            260                 265                 270

gca gga ctc ctc cta cat ctc cgg cca gat cct cca ccc caa cgg tgg      864
Ala Gly Leu Leu Leu His Leu Arg Pro Asp Pro Pro Pro Gln Arg Trp
        275                 280                 285

tac cat cgt caa                                                     876
Tyr His Arg Gln
    290
```

<210> 14
<211> 292
<212> PRT
<213> Hordeum vulgare

<400> 14

```
Met Ala Ser Gln Lys Phe Pro Pro Gln Gln Gln Asp Cys Gln Pro Gly
1               5                   10                  15

Lys Glu His Ala Met Asp Pro Arg Pro Glu Ala Ile Ile Lys Asn Tyr
            20                  25                  30

Lys Ser Gly Gln Gln Ala Pro Gly Gln Gly Gly Ala Gly Asp Arg Arg
        35                  40                  45

Arg Leu Gly His Arg Ala Arg Gly Val Pro Val Pro Arg Ala Gly Gly
    50                  55                  60

Arg Asp Gly Glu Leu His Val Arg Glu Gly Ala Arg Gly Gln Gly Arg
65                  70                  75                  80

Gly Gly Asp Pro Ala Gly Ala Pro Arg His Gln Val Pro His Arg Arg
                85                  90                  95

Arg Arg Ala Gln Gly Ala Leu Gly Arg Pro Arg Val Arg Gly Glu Leu
            100                 105                 110
```

```
Pro Gln Gly Gly Gly Gly Gly Gly Gln Arg Ala Arg Arg Pro Arg Gly
        115                 120                 125

His Pro Arg Glu Gln Arg Gly Arg Ala Val Arg Pro Pro Leu His His
    130                 135                 140

Arg Asp His Arg Ala Gly Pro Gly Ala Arg Val Pro His Gln His Leu
145                 150                 155                 160

Leu Leu Leu Pro His Asp Gln Val Arg Arg Glu Ala His Gly Ala Arg
                165                 170                 175

Val Gln His His Gln His His Leu Arg Glu Arg Val Gln Gly Gln Arg
            180                 185                 190

Asp Ala Ala Gly Leu His Gly His Gln Gly Arg His Arg Gly Leu His
        195                 200                 205

Pro Arg Ala Val Asp Ala Ala Gly Gly Glu Gly Asp Pro Arg Gln Arg
    210                 215                 220

Arg Gly Ala Gly Ala His Leu Asp Ala Pro His Pro Gly Leu Leu Pro
225                 230                 235                 240

Gly Gly Glu Gly Glu Ala Val Arg Val Arg Gly Ala His Glu Ala Arg
                245                 250                 255

His Ala Ala Gln Arg Gly Arg Ala Gln Leu Arg Leu Pro Cys Gln Arg
            260                 265                 270

Ala Gly Leu Leu Leu His Leu Arg Pro Asp Pro Pro Pro Gln Arg Trp
        275                 280                 285

Tyr His Arg Gln
    290
```

<210> 15
<211> 1187

<212> DNA

<213> Soybean

<220>

<221> CDS

<222> (38)..(916)

<223> Glucose and Ribitol Dehydrogenoase Homolog

<400> 15

```
gatctccata tctgttccca aaagcttgtt tgtaaga atg gct tcc ggt gaa cag      55
                                        Met Ala Ser Gly Glu Gln
                                        1               5

aaa ttc cct cct caa caa caa caa aca cag cct ggg aag gag cat gct      103
Lys Phe Pro Pro Gln Gln Gln Gln Thr Gln Pro Gly Lys Glu His Ala
            10                  15                  20

atg act cca gta ccc caa ttc act agc cct gac tac aag cct tca aat      151
Met Thr Pro Val Pro Gln Phe Thr Ser Pro Asp Tyr Lys Pro Ser Asn
        25                  30                  35

aaa ctt caa ggg aag att gca tta gtc act ggg ggt gat tct ggg att      199
Lys Leu Gln Gly Lys Ile Ala Leu Val Thr Gly Gly Asp Ser Gly Ile
    40                  45                  50

gga cga gcg gtg tgt aac ttg ttt gcc tta gaa ggt gct acc gtg gcc      247
Gly Arg Ala Val Cys Asn Leu Phe Ala Leu Glu Gly Ala Thr Val Ala
55                  60                  65                  70

ttc acg tat gtg aag ggg cat gag gac aag gac gcg agg gac aca ttg      295
Phe Thr Tyr Val Lys Gly His Glu Asp Lys Asp Ala Arg Asp Thr Leu
                75                  80                  85

gaa atg atc aag aga gca aag act tcg gat gcc aag gat cca atg gca      343
Glu Met Ile Lys Arg Ala Lys Thr Ser Asp Ala Lys Asp Pro Met Ala
            90                  95                  100

ata gca tct gat ttg ggt tac gat gag aac tgc aag agg gtg gtt gat      391
Ile Ala Ser Asp Leu Gly Tyr Asp Glu Asn Cys Lys Arg Val Val Asp
        105                 110                 115

gag gtc gtg agt gct tat ggt tgt att gac att ctg gtc aac aat gca      439
Glu Val Val Ser Ala Tyr Gly Cys Ile Asp Ile Leu Val Asn Asn Ala
    120                 125                 130

gct gag cag tac gag tgt gga acc gtg gag gac ata gac gag cct agg      487
Ala Glu Gln Tyr Glu Cys Gly Thr Val Glu Asp Ile Asp Glu Pro Arg
135                 140                 145                 150

ctt gag agg gtc ttt cgt aca aat atc ttc tcc tat ttc ttc atg gcg      535
Leu Glu Arg Val Phe Arg Thr Asn Ile Phe Ser Tyr Phe Phe Met Ala
                155                 160                 165

agg cat gcc ttg aag cac atg aag gaa gga agc agc att atc aac acg      583
Arg His Ala Leu Lys His Met Lys Glu Gly Ser Ser Ile Ile Asn Thr
            170                 175                 180
```

```
aca tca gtg aat gca tac aag gga cat gcg aaa cta ttg gac tac acg      631
Thr Ser Val Asn Ala Tyr Lys Gly His Ala Lys Leu Leu Asp Tyr Thr
        185                 190                 195

tcc acc aag ggg gca att gtg gcc tat aca agg ggt ctt gcc ctt cag      679
Ser Thr Lys Gly Ala Ile Val Ala Tyr Thr Arg Gly Leu Ala Leu Gln
    200                 205                 210

ctg gtg agt aag gga att cgg gtt aat ggg gtg gct cca ggg ccc att      727
Leu Val Ser Lys Gly Ile Arg Val Asn Gly Val Ala Pro Gly Pro Ile
215                 220                 225                 230

tgg acc cct ttg ata cca gcc tct ttc aag gag gaa gaa acg gcc caa      775
Trp Thr Pro Leu Ile Pro Ala Ser Phe Lys Glu Glu Glu Thr Ala Gln
                235                 240                 245

ttt gga gcg cag gtc cca atg aag aga gct ggt caa cct att gag gtt      823
Phe Gly Ala Gln Val Pro Met Lys Arg Ala Gly Gln Pro Ile Glu Val
            250                 255                 260

gct cct tcc tat gtt ttt ctt gct tcc aac caa tgc tcc tct tac ata      871
Ala Pro Ser Tyr Val Phe Leu Ala Ser Asn Gln Cys Ser Ser Tyr Ile
        265                 270                 275

act gga caa gtc ctt cac ccc aat ggt gga acc gtt gtg aat ggt          916
Thr Gly Gln Val Leu His Pro Asn Gly Gly Thr Val Val Asn Gly
    280                 285                 290

taaaccgttg gtgatgatga tattcgggat gaatatatgt ggcgagagta gtaggccagt    976

gttacgtttt gtgtgaatgt tttacgatgt gttttaatgc atggctaact cactcaggtc   1036

ctctctgcac tgttagaggt ggggcttgga ggattatcca cttttgaatg tacgagttat   1096

tagcctaaga aaatgtgtct tttgtagcca attatatgta aacaagtaaa agtatataat   1156

aaagatcggt atgtataagg tttaaacttt a                                  1187
```

<210> 16

<211> 293

<212> PRT

<213> Soybean

<400> 16

```
Met Ala Ser Gly Glu Gln Lys Phe Pro Pro Gln Gln Gln Gln Thr Gln
1               5                   10                  15

Pro Gly Lys Glu His Ala Met Thr Pro Val Pro Gln Phe Thr Ser Pro
            20                  25                  30

Asp Tyr Lys Pro Ser Asn Lys Leu Gln Gly Lys Ile Ala Leu Val Thr
```

```
        35                  40                  45

Gly Gly Asp Ser Gly Ile Gly Arg Ala Val Cys Asn Leu Phe Ala Leu
    50                  55                  60

Glu Gly Ala Thr Val Ala Phe Thr Tyr Val Lys Gly His Glu Asp Lys
65                  70                  75                  80

Asp Ala Arg Asp Thr Leu Glu Met Ile Lys Arg Ala Lys Thr Ser Asp
                85                  90                  95

Ala Lys Asp Pro Met Ala Ile Ala Ser Asp Leu Gly Tyr Asp Glu Asn
            100                 105                 110

Cys Lys Arg Val Val Asp Glu Val Val Ser Ala Tyr Gly Cys Ile Asp
        115                 120                 125

Ile Leu Val Asn Asn Ala Ala Glu Gln Tyr Glu Cys Gly Thr Val Glu
    130                 135                 140

Asp Ile Asp Glu Pro Arg Leu Glu Arg Val Phe Arg Thr Asn Ile Phe
145                 150                 155                 160

Ser Tyr Phe Phe Met Ala Arg His Ala Leu Lys His Met Lys Glu Gly
                165                 170                 175

Ser Ser Ile Ile Asn Thr Thr Ser Val Asn Ala Tyr Lys Gly His Ala
            180                 185                 190

Lys Leu Leu Asp Tyr Thr Ser Thr Lys Gly Ala Ile Val Ala Tyr Thr
        195                 200                 205

Arg Gly Leu Ala Leu Gln Leu Val Ser Lys Gly Ile Arg Val Asn Gly
    210                 215                 220

Val Ala Pro Gly Pro Ile Trp Thr Pro Leu Ile Pro Ala Ser Phe Lys
225                 230                 235                 240

Glu Glu Glu Thr Ala Gln Phe Gly Ala Gln Val Pro Met Lys Arg Ala
                245                 250                 255

Gly Gln Pro Ile Glu Val Ala Pro Ser Tyr Val Phe Leu Ala Ser Asn
            260                 265                 270

Gln Cys Ser Ser Tyr Ile Thr Gly Gln Val Leu His Pro Asn Gly Gly
        275                 280                 285
```

Thr Val Val Asn Gly
290

<210> 17
<211> 1194

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (119)..(982)

<223> PIR2 Accesion T06212 Glucose and Ribitol Dehydrogenase Homolog

<400> 17
cacttcgtca agagtcttat catcaaacgt tacgttctcg ttttctcaaa ttccaaacaa 60

taccaagaaa cctctacttg aaaagagaag attatcgccg cgtgtgtgcc taagagcg 118

atg gcg tct gag aaa caa aaa caa cat gca caa cct ggc aaa gaa cat 166
Met Ala Ser Glu Lys Gln Lys Gln His Ala Gln Pro Gly Lys Glu His
1 5 10 15

gtc atg gaa tca agc cca caa ttc tct agc tca gat tac caa cct tcc 214
Val Met Glu Ser Ser Pro Gln Phe Ser Ser Ser Asp Tyr Gln Pro Ser
20 25 30

aac aag ctt cgt ggt aag gtg gcg ttg ata act ggt gga gac tct ggg 262
Asn Lys Leu Arg Gly Lys Val Ala Leu Ile Thr Gly Gly Asp Ser Gly
35 40 45

att ggt cga gcc gtg gga tac tgt ttt gca tcc gaa gga gct act gtg 310
Ile Gly Arg Ala Val Gly Tyr Cys Phe Ala Ser Glu Gly Ala Thr Val
50 55 60

gct ttc act tac gtg aag ggt caa gaa gaa aaa gat gca caa gag acc 358
Ala Phe Thr Tyr Val Lys Gly Gln Glu Glu Lys Asp Ala Gln Glu Thr
65 70 75 80

cta caa atg ttg aag gag gtc aaa acc tcg gac tcc aag gaa cct atc 406
Leu Gln Met Leu Lys Glu Val Lys Thr Ser Asp Ser Lys Glu Pro Ile
85 90 95

gcc att cca acg gat tta gga ttt gac gaa aac tgc aaa agg gtc gtt 454
Ala Ile Pro Thr Asp Leu Gly Phe Asp Glu Asn Cys Lys Arg Val Val
100 105 110

gat gag gtc gtt aat gct ttt ggc cgc atc gat gtt ttg atc aat aac 502

```
Asp Glu Val Val Asn Ala Phe Gly Arg Ile Asp Val Leu Ile Asn Asn
        115                 120                 125

gca gca gag cag tac gag agc agc aca atc gaa gag att gat gag cct      550
Ala Ala Glu Gln Tyr Glu Ser Ser Thr Ile Glu Glu Ile Asp Glu Pro
    130                 135                 140

agg ctt gag cga gtc ttc cgt aca aac atc ttt tct tac ttc ttt ctc      598
Arg Leu Glu Arg Val Phe Arg Thr Asn Ile Phe Ser Tyr Phe Phe Leu
145                 150                 155                 160

aca agg cat gcg ttg aag cat atg aag gaa gga agc agc att atc aac      646
Thr Arg His Ala Leu Lys His Met Lys Glu Gly Ser Ser Ile Ile Asn
                165                 170                 175

acc act tcg gtg aat gcc tac aag gga aac gct tca ctt ctc gac tac      694
Thr Thr Ser Val Asn Ala Tyr Lys Gly Asn Ala Ser Leu Leu Asp Tyr
            180                 185                 190

acc gct aca aaa gga gcg att gtg gcg ttt act cga gga ctt gca ctt      742
Thr Ala Thr Lys Gly Ala Ile Val Ala Phe Thr Arg Gly Leu Ala Leu
        195                 200                 205

cag cta gct gag aaa gga atc cgt gtc aat ggt gtg gct cct ggt cca      790
Gln Leu Ala Glu Lys Gly Ile Arg Val Asn Gly Val Ala Pro Gly Pro
    210                 215                 220

ata tgg aca ccc ctt atc cca gca tca ttc aat gag gag aag att aag      838
Ile Trp Thr Pro Leu Ile Pro Ala Ser Phe Asn Glu Glu Lys Ile Lys
225                 230                 235                 240

aat ttt ggg tct gag gtt ccg atg aaa aga gcg ggt cag cca att gaa      886
Asn Phe Gly Ser Glu Val Pro Met Lys Arg Ala Gly Gln Pro Ile Glu
                245                 250                 255

gtg gca cca tcc tat gtt ttc ttg gcg tgt aac cac tgc tct tct tac      934
Val Ala Pro Ser Tyr Val Phe Leu Ala Cys Asn His Cys Ser Ser Tyr
            260                 265                 270

ttc act ggt caa gtt ctt cac cct aat gga gga gct gtg gta aat gcg      982
Phe Thr Gly Gln Val Leu His Pro Asn Gly Gly Ala Val Val Asn Ala
        275                 280                 285

taagcgtgga gtggacaaga ccggtcttaa cgtcttagac cattaaatat gatatgatga   1042

tgttgtttga gtttaggggc tttttgttat gttggtaatg tgttacgtcc gtatatgttg   1102

gtaatgtgtt gcgtccgtac cttctgtagc aaaagtatgt gtttaataaa gactttacct   1162

cgacaatagt cggggcctgt tcccttaaaa gt                                 1194
```

<210> 18

<211> 288

<212> PRT

<213> Arabidopsis thaliana

<400> 18

```
Met Ala Ser Glu Lys Gln Lys Gln His Ala Gln Pro Gly Lys Glu His
1               5                  10                  15

Val Met Glu Ser Ser Pro Gln Phe Ser Ser Ser Asp Tyr Gln Pro Ser
            20                  25                  30

Asn Lys Leu Arg Gly Lys Val Ala Leu Ile Thr Gly Gly Asp Ser Gly
        35                  40                  45

Ile Gly Arg Ala Val Gly Tyr Cys Phe Ala Ser Glu Gly Ala Thr Val
    50                  55                  60

Ala Phe Thr Tyr Val Lys Gly Gln Glu Glu Lys Asp Ala Gln Glu Thr
65                  70                  75                  80

Leu Gln Met Leu Lys Glu Val Lys Thr Ser Asp Ser Lys Glu Pro Ile
                85                  90                  95

Ala Ile Pro Thr Asp Leu Gly Phe Asp Glu Asn Cys Lys Arg Val Val
            100                 105                 110

Asp Glu Val Val Asn Ala Phe Gly Arg Ile Asp Val Leu Ile Asn Asn
        115                 120                 125

Ala Ala Glu Gln Tyr Glu Ser Ser Thr Ile Glu Glu Ile Asp Glu Pro
    130                 135                 140

Arg Leu Glu Arg Val Phe Arg Thr Asn Ile Phe Ser Tyr Phe Phe Leu
145                 150                 155                 160

Thr Arg His Ala Leu Lys His Met Lys Glu Gly Ser Ser Ile Ile Asn
                165                 170                 175

Thr Thr Ser Val Asn Ala Tyr Lys Gly Asn Ala Ser Leu Leu Asp Tyr
            180                 185                 190

Thr Ala Thr Lys Gly Ala Ile Val Ala Phe Thr Arg Gly Leu Ala Leu
        195                 200                 205

Gln Leu Ala Glu Lys Gly Ile Arg Val Asn Gly Val Ala Pro Gly Pro
    210                 215                 220

Ile Trp Thr Pro Leu Ile Pro Ala Ser Phe Asn Glu Glu Lys Ile Lys
225                 230                 235                 240
```

```
Asn Phe Gly Ser Glu Val Pro Met Lys Arg Ala Gly Gln Pro Ile Glu
                245                 250                 255

Val Ala Pro Ser Tyr Val Phe Leu Ala Cys Asn His Cys Ser Ser Tyr
            260                 265                 270

Phe Thr Gly Gln Val Leu His Pro Asn Gly Gly Ala Val Val Asn Ala
```

<210> 19
<211> 2451
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)..(2451)
<223> RECEPTOR PROTEIN KINASE ZMPK1 PRECURSOR (EC 2.7.1.-) -

<400> 19

```
atg cct cgt cct ctt gca gct ctc ctc tct acc gct tgc atc ctc tcc      48
Met Pro Arg Pro Leu Ala Ala Leu Leu Ser Thr Ala Cys Ile Leu Ser
1               5                   10                  15

ttc ttc atc gct ctg ttt cca aga gct gca tca tcc cga gac atc cta      96
Phe Phe Ile Ala Leu Phe Pro Arg Ala Ala Ser Ser Arg Asp Ile Leu
            20                  25                  30

cca ctg ggt tcc tct ctc gta gtc gag tcc tac gaa tcc agc acc tta     144
Pro Leu Gly Ser Ser Leu Val Val Glu Ser Tyr Glu Ser Ser Thr Leu
        35                  40                  45

caa tca tca gac ggg aca ttc tcc tct ggc ttc tac gaa gtc tac acc     192
Gln Ser Ser Asp Gly Thr Phe Ser Ser Gly Phe Tyr Glu Val Tyr Thr
    50                  55                  60

cat gcc ttc aca ttc tca gta tgg tac tca aag acg gag gcg gcg gcg     240
His Ala Phe Thr Phe Ser Val Trp Tyr Ser Lys Thr Glu Ala Ala Ala
65                  70                  75                  80

gcc aac aac aag acc atc gtg tgg agc gca aac cct gac cgc cct gtc     288
Ala Asn Asn Lys Thr Ile Val Trp Ser Ala Asn Pro Asp Arg Pro Val
                85                  90                  95

cat gcc agg agg tcg gct cta acc ctg caa aag gac ggc aac atg gtg     336
His Ala Arg Arg Ser Ala Leu Thr Leu Gln Lys Asp Gly Asn Met Val
            100                 105                 110

ctc acc gac tac gac ggc gca gcc gtg tgg cga gct gat ggc aac aac     384
Leu Thr Asp Tyr Asp Gly Ala Ala Val Trp Arg Ala Asp Gly Asn Asn
        115                 120                 125

ttc acc ggc gtc cag cgt gct cgg ctc ctg gac acc ggg aac ctc gtc     432
Phe Thr Gly Val Gln Arg Ala Arg Leu Leu Asp Thr Gly Asn Leu Val
    130                 135                 140
```

```
atc gag gac tca gga ggt aac act gta tgg cag agt ttc gat tcc cca      480
Ile Glu Asp Ser Gly Gly Asn Thr Val Trp Gln Ser Phe Asp Ser Pro
145                 150                 155                 160

acg gac act ttc ctg ccg acg cag ctc atc act gct gcg acc aga tta      528
Thr Asp Thr Phe Leu Pro Thr Gln Leu Ile Thr Ala Ala Thr Arg Leu
                165                 170                 175

gtc ccc aca acc caa tcg cgt agt cct ggt aac tac atc ttc cgc ttc      576
Val Pro Thr Thr Gln Ser Arg Ser Pro Gly Asn Tyr Ile Phe Arg Phe
            180                 185                 190

agc gac ctc tca gtg ctg tcg ctt ata tac cac gtg cct caa gtc tca      624
Ser Asp Leu Ser Val Leu Ser Leu Ile Tyr His Val Pro Gln Val Ser
        195                 200                 205

gac ata tac tgg cca gac cct gac cag aac ctc tac cag gat ggc cgg      672
Asp Ile Tyr Trp Pro Asp Pro Asp Gln Asn Leu Tyr Gln Asp Gly Arg
    210                 215                 220

aac cag tat aac agt acg agg tta gga atg ctt act gat agc ggg gtg      720
Asn Gln Tyr Asn Ser Thr Arg Leu Gly Met Leu Thr Asp Ser Gly Val
225                 230                 235                 240

ctt gcc tcg agc gac ttc gct gat ggt cag gcg ctt gtg gcc tcc gac      768
Leu Ala Ser Ser Asp Phe Ala Asp Gly Gln Ala Leu Val Ala Ser Asp
                245                 250                 255

gta ggg ccg ggc gtc aag aga agg cta act ctt gac cct gat ggc aat      816
Val Gly Pro Gly Val Lys Arg Arg Leu Thr Leu Asp Pro Asp Gly Asn
            260                 265                 270

ctc cgt ctg tac agc atg aac gat tca gat ggg tca tgg tca gtt tca      864
Leu Arg Leu Tyr Ser Met Asn Asp Ser Asp Gly Ser Trp Ser Val Ser
        275                 280                 285

atg gta gca atg acc cag cct tgc aat att cac ggt ttg tgt ggt cct      912
Met Val Ala Met Thr Gln Pro Cys Asn Ile His Gly Leu Cys Gly Pro
    290                 295                 300

aat ggc atc tgc cac tac tca ccc aca cct aca tgt tcg tgc cca cca      960
Asn Gly Ile Cys His Tyr Ser Pro Thr Pro Thr Cys Ser Cys Pro Pro
305                 310                 315                 320

ggt tat gcg acg agg aac ccg ggt aac tgg act gaa ggc tgt atg gct     1008
Gly Tyr Ala Thr Arg Asn Pro Gly Asn Trp Thr Glu Gly Cys Met Ala
                325                 330                 335

att gtc aac aca acc tgt gac cgc tat gac aag agg tct atg aga ttt     1056
Ile Val Asn Thr Thr Cys Asp Arg Tyr Asp Lys Arg Ser Met Arg Phe
            340                 345                 350

gtg cga ctt ccc aat acg gat ttt tgg ggg tcg gat cag caa cat ctt     1104
Val Arg Leu Pro Asn Thr Asp Phe Trp Gly Ser Asp Gln Gln His Leu
        355                 360                 365

ctg tcg gtt tct ctt cga act tgt agg gat atc tgc atc agt gac tgc     1152
Leu Ser Val Ser Leu Arg Thr Cys Arg Asp Ile Cys Ile Ser Asp Cys
    370                 375                 380

acc tgt aaa ggc ttt cag tat cag gaa ggc aca gga tca tgc tat cca     1200
```

```
Thr Cys Lys Gly Phe Gln Tyr Gln Glu Gly Thr Gly Ser Cys Tyr Pro
385                 390                 395                 400

aaa gct tat ctt ttc agt gga aga acc tac cca aca tct gac gtg cga    1248
Lys Ala Tyr Leu Phe Ser Gly Arg Thr Tyr Pro Thr Ser Asp Val Arg
                405                 410                 415

acg ata tat ctc aag ctt cca aca ggg gtc agt gtt tca aat gcc ctt    1296
Thr Ile Tyr Leu Lys Leu Pro Thr Gly Val Ser Val Ser Asn Ala Leu
            420                 425                 430

att cca cgt tcc gac gtg ttc gat tcc gtg ccc cgt cgt ctc gac tgc    1344
Ile Pro Arg Ser Asp Val Phe Asp Ser Val Pro Arg Arg Leu Asp Cys
        435                 440                 445

gat cgg atg aac aaa agc atc aga gaa ccg ttt cca gat gtg cac aag    1392
Asp Arg Met Asn Lys Ser Ile Arg Glu Pro Phe Pro Asp Val His Lys
    450                 455                 460

acc ggc gga gga gaa tcg aaa tgg ttt tac ttc tat ggg ttc ata gct    1440
Thr Gly Gly Gly Glu Ser Lys Trp Phe Tyr Phe Tyr Gly Phe Ile Ala
465                 470                 475                 480

gca ttt ttt gtc gtt gaa gtt tcc ttc att tcg ttt gcg tgg ttc ttt    1488
Ala Phe Phe Val Val Glu Val Ser Phe Ile Ser Phe Ala Trp Phe Phe
                485                 490                 495

gtt ttg aag aga gaa ctc agg cca tct gaa cta tgg gcg tct gag aaa    1536
Val Leu Lys Arg Glu Leu Arg Pro Ser Glu Leu Trp Ala Ser Glu Lys
            500                 505                 510

ggt tac aaa gca atg act agt aat ttt aga agg tac agc tac agg gaa    1584
Gly Tyr Lys Ala Met Thr Ser Asn Phe Arg Arg Tyr Ser Tyr Arg Glu
        515                 520                 525

ctt gtg aag gcg acc aga aaa ttc aag gtt gag cta ggg agg gga gaa    1632
Leu Val Lys Ala Thr Arg Lys Phe Lys Val Glu Leu Gly Arg Gly Glu
    530                 535                 540

tca ggc act gtg tac aaa ggt gtc cta gaa gat gat agg cat gtg gct    1680
Ser Gly Thr Val Tyr Lys Gly Val Leu Glu Asp Asp Arg His Val Ala
545                 550                 555                 560

gtg aag aag ctg gag aat gta agg caa ggc aag gaa gtg ttt cag gct    1728
Val Lys Lys Leu Glu Asn Val Arg Gln Gly Lys Glu Val Phe Gln Ala
                565                 570                 575

gag cta agt gta att ggg agg atc aac cac atg aac ctt gtg agg ata    1776
Glu Leu Ser Val Ile Gly Arg Ile Asn His Met Asn Leu Val Arg Ile
            580                 585                 590

tgg ggc ttc tgt tca gag gga tct cat agg ttg ttg gtc tcc gaa tat    1824
Trp Gly Phe Cys Ser Glu Gly Ser His Arg Leu Leu Val Ser Glu Tyr
        595                 600                 605

gtt gag aat gga tca ctg gct aac att ttg ttc agt gaa gga ggc aac    1872
Val Glu Asn Gly Ser Leu Ala Asn Ile Leu Phe Ser Glu Gly Gly Asn
    610                 615                 620

atc tta ttg gac tgg gag gga agg ttc aac att gcg tta ggt gtt gca    1920
Ile Leu Leu Asp Trp Glu Gly Arg Phe Asn Ile Ala Leu Gly Val Ala
```

```
625                 630                 635                 640

aaa ggg tta gcc tat ctc cac cat gag tgc tta gag tgg gtc atc cac     1968
Lys Gly Leu Ala Tyr Leu His His Glu Cys Leu Glu Trp Val Ile His
                645                 650                 655

tgt gat gtg aaa cct gag aac ata ctg tta gac caa gct ttt gag ccc     2016
Cys Asp Val Lys Pro Glu Asn Ile Leu Leu Asp Gln Ala Phe Glu Pro
            660                 665                 670

aag atc act gac ttt ggg ttg gtg aag ttg ctg aac aga gga ggg tcc     2064
Lys Ile Thr Asp Phe Gly Leu Val Lys Leu Leu Asn Arg Gly Gly Ser
        675                 680                 685

acc cag aac gta tcc cat gtc aga gga acg cta ggt tac att gca cct     2112
Thr Gln Asn Val Ser His Val Arg Gly Thr Leu Gly Tyr Ile Ala Pro
    690                 695                 700

gag tgg gtt tcc agc ctc ccg atc aca gca aaa gtc gat gta tac agt     2160
Glu Trp Val Ser Ser Leu Pro Ile Thr Ala Lys Val Asp Val Tyr Ser
705                 710                 715                 720

tat ggg gtt gtg cta ctg gag cta ttg aca ggc acc aga gtt tca gag     2208
Tyr Gly Val Val Leu Leu Glu Leu Leu Thr Gly Thr Arg Val Ser Glu
                725                 730                 735

ttg gtg gga ggc aca gat gag gtg cat agt atg ctt aga aag ctt gtc     2256
Leu Val Gly Gly Thr Asp Glu Val His Ser Met Leu Arg Lys Leu Val
            740                 745                 750

agg atg ctt tct gcc aaa ctt gaa ggg gag gaa caa tcg tgg att gat     2304
Arg Met Leu Ser Ala Lys Leu Glu Gly Glu Glu Gln Ser Trp Ile Asp
        755                 760                 765

ggg tat ctg gat tca aaa ctg aat cgt cca gtc aac tat gtg caa gca     2352
Gly Tyr Leu Asp Ser Lys Leu Asn Arg Pro Val Asn Tyr Val Gln Ala
    770                 775                 780

aga aca ctg atc aaa ttg gcg gtc tcc tgc ttg gag gaa gac aga agc     2400
Arg Thr Leu Ile Lys Leu Ala Val Ser Cys Leu Glu Glu Asp Arg Ser
785                 790                 795                 800

aaa aga ccg act atg gaa cat gca gtc cag acc ctc ctg tca gct gat     2448
Lys Arg Pro Thr Met Glu His Ala Val Gln Thr Leu Leu Ser Ala Asp
                805                 810                 815

gac                                                                 2451
Asp
```

<210> 20
<211> 817
<212> PRT
<213> Zea mays

<400> 20

```
Met Pro Arg Pro Leu Ala Ala Leu Leu Ser Thr Ala Cys Ile Leu Ser
1               5                   10                  15
```

```
Phe Phe Ile Ala Leu Phe Pro Arg Ala Ala Ser Ser Arg Asp Ile Leu
            20                  25                  30

Pro Leu Gly Ser Ser Leu Val Val Glu Ser Tyr Glu Ser Ser Thr Leu
        35                  40                  45

Gln Ser Ser Asp Gly Thr Phe Ser Ser Gly Phe Tyr Glu Val Tyr Thr
    50                  55                  60

His Ala Phe Thr Phe Ser Val Trp Tyr Ser Lys Thr Glu Ala Ala Ala
65                  70                  75                  80

Ala Asn Asn Lys Thr Ile Val Trp Ser Ala Asn Pro Asp Arg Pro Val
                85                  90                  95

His Ala Arg Arg Ser Ala Leu Thr Leu Gln Lys Asp Gly Asn Met Val
            100                 105                 110

Leu Thr Asp Tyr Asp Gly Ala Ala Val Trp Arg Ala Asp Gly Asn Asn
        115                 120                 125

Phe Thr Gly Val Gln Arg Ala Arg Leu Leu Asp Thr Gly Asn Leu Val
    130                 135                 140

Ile Glu Asp Ser Gly Gly Asn Thr Val Trp Gln Ser Phe Asp Ser Pro
145                 150                 155                 160

Thr Asp Thr Phe Leu Pro Thr Gln Leu Ile Thr Ala Ala Thr Arg Leu
                165                 170                 175

Val Pro Thr Thr Gln Ser Arg Ser Pro Gly Asn Tyr Ile Phe Arg Phe
            180                 185                 190

Ser Asp Leu Ser Val Leu Ser Leu Ile Tyr His Val Pro Gln Val Ser
        195                 200                 205

Asp Ile Tyr Trp Pro Asp Pro Asp Gln Asn Leu Tyr Gln Asp Gly Arg
    210                 215                 220

Asn Gln Tyr Asn Ser Thr Arg Leu Gly Met Leu Thr Asp Ser Gly Val
225                 230                 235                 240

Leu Ala Ser Ser Asp Phe Ala Asp Gly Gln Ala Leu Val Ala Ser Asp
                245                 250                 255
```

```
Val Gly Pro Gly Val Lys Arg Arg Leu Thr Leu Asp Pro Asp Gly Asn
            260                 265                 270

Leu Arg Leu Tyr Ser Met Asn Asp Ser Asp Gly Ser Trp Ser Val Ser
        275                 280                 285

Met Val Ala Met Thr Gln Pro Cys Asn Ile His Gly Leu Cys Gly Pro
    290                 295                 300

Asn Gly Ile Cys His Tyr Ser Pro Thr Pro Thr Cys Ser Cys Pro Pro
305                 310                 315                 320

Gly Tyr Ala Thr Arg Asn Pro Gly Asn Trp Thr Glu Gly Cys Met Ala
                325                 330                 335

Ile Val Asn Thr Thr Cys Asp Arg Tyr Asp Lys Arg Ser Met Arg Phe
            340                 345                 350

Val Arg Leu Pro Asn Thr Asp Phe Trp Gly Ser Asp Gln Gln His Leu
        355                 360                 365

Leu Ser Val Ser Leu Arg Thr Cys Arg Asp Ile Cys Ile Ser Asp Cys
    370                 375                 380

Thr Cys Lys Gly Phe Gln Tyr Gln Glu Gly Thr Gly Ser Cys Tyr Pro
385                 390                 395                 400

Lys Ala Tyr Leu Phe Ser Gly Arg Thr Tyr Pro Thr Ser Asp Val Arg
                405                 410                 415

Thr Ile Tyr Leu Lys Leu Pro Thr Gly Val Ser Val Ser Asn Ala Leu
            420                 425                 430

Ile Pro Arg Ser Asp Val Phe Asp Ser Val Pro Arg Arg Leu Asp Cys
        435                 440                 445

Asp Arg Met Asn Lys Ser Ile Arg Glu Pro Phe Pro Asp Val His Lys
    450                 455                 460

Thr Gly Gly Gly Glu Ser Lys Trp Phe Tyr Phe Tyr Gly Phe Ile Ala
465                 470                 475                 480

Ala Phe Phe Val Val Glu Val Ser Phe Ile Ser Phe Ala Trp Phe Phe
                485                 490                 495

Val Leu Lys Arg Glu Leu Arg Pro Ser Glu Leu Trp Ala Ser Glu Lys
```

500 505 510

Gly Tyr Lys Ala Met Thr Ser Asn Phe Arg Arg Tyr Ser Tyr Arg Glu
515 520 525

Leu Val Lys Ala Thr Arg Lys Phe Lys Val Glu Leu Gly Arg Gly Glu
530 535 540

Ser Gly Thr Val Tyr Lys Gly Val Leu Glu Asp Asp Arg His Val Ala
545 550 555 560

Val Lys Lys Leu Glu Asn Val Arg Gln Gly Lys Glu Val Phe Gln Ala
565 570 575

Glu Leu Ser Val Ile Gly Arg Ile Asn His Met Asn Leu Val Arg Ile
580 585 590

Trp Gly Phe Cys Ser Glu Gly Ser His Arg Leu Leu Val Ser Glu Tyr
595 600 605

Val Glu Asn Gly Ser Leu Ala Asn Ile Leu Phe Ser Glu Gly Gly Asn
610 615 620

Ile Leu Leu Asp Trp Glu Gly Arg Phe Asn Ile Ala Leu Gly Val Ala
625 630 635 640

Lys Gly Leu Ala Tyr Leu His His Glu Cys Leu Glu Trp Val Ile His
645 650 655

Cys Asp Val Lys Pro Glu Asn Ile Leu Leu Asp Gln Ala Phe Glu Pro
660 665 670

Lys Ile Thr Asp Phe Gly Leu Val Lys Leu Leu Asn Arg Gly Gly Ser
675 680 685

Thr Gln Asn Val Ser His Val Arg Gly Thr Leu Gly Tyr Ile Ala Pro
690 695 700

Glu Trp Val Ser Ser Leu Pro Ile Thr Ala Lys Val Asp Val Tyr Ser
705 710 715 720

Tyr Gly Val Val Leu Leu Glu Leu Leu Thr Gly Thr Arg Val Ser Glu
725 730 735

Leu Val Gly Gly Thr Asp Glu Val His Ser Met Leu Arg Lys Leu Val
740 745 750

Arg Met Leu Ser Ala Lys Leu Glu Gly Glu Glu Gln Ser Trp Ile Asp
755 760 765

Gly Tyr Leu Asp Ser Lys Leu Asn Arg Pro Val Asn Tyr Val Gln Ala
770 775 780

Arg Thr Leu Ile Lys Leu Ala Val Ser Cys Leu Glu Glu Asp Arg Ser
785 790 795 800

Lys Arg Pro Thr Met Glu His Ala Val Gln Thr Leu Leu Ser Ala Asp
805 810 815

Asp

<210> 21

<211> 1434

<212> DNA

<213> Zea mays

<220>

<221> CDS

<222> (1)..(1434)

<223> RECEPTOR PROTEIN KINASE ZMPK1 PRECURSOR maize

<400> 21

atg ggg atc agc aag gga ggc agc ggc aag gag gcg aag aag ccg ccg 48
Met Gly Ile Ser Lys Gly Gly Ser Gly Lys Glu Ala Lys Lys Pro Pro
1 5 10 15

ctg ctg ctg ggg cga ttc gag gtc ggg aag ctg ctg ggg cag ggc aac 96
Leu Leu Leu Gly Arg Phe Glu Val Gly Lys Leu Leu Gly Gln Gly Asn
20 25 30

ttc gcc aag gtg tac cac gcg cgc aac gtg gcc acc ggc gag gag gtg 144
Phe Ala Lys Val Tyr His Ala Arg Asn Val Ala Thr Gly Glu Glu Val
35 40 45

gcg atc aag gtg atg gag aag gag aag atc ttc aag tcg ggg ctc acg 192
Ala Ile Lys Val Met Glu Lys Glu Lys Ile Phe Lys Ser Gly Leu Thr
50 55 60

```
gcg cac atc aag cgg gag atc gcc gtg ctc cgg cgc gtc cgc cac ccg      240
Ala His Ile Lys Arg Glu Ile Ala Val Leu Arg Arg Val Arg His Pro
65                  70                  75                  80

cac atc gtg cag ctg tac gag gtg atg gcc acc aag ctc cgg atc tac      288
His Ile Val Gln Leu Tyr Glu Val Met Ala Thr Lys Leu Arg Ile Tyr
                85                  90                  95

ttc gtc atg gag tac gtc cgc ggc ggc gag ctg ttc gcg cgc gtg gcg      336
Phe Val Met Glu Tyr Val Arg Gly Gly Glu Leu Phe Ala Arg Val Ala
            100                 105                 110

cgg ggg cgg ctg ccc gag gcc gac gcg cgg cgc tac ttc cag cag ctg      384
Arg Gly Arg Leu Pro Glu Ala Asp Ala Arg Arg Tyr Phe Gln Gln Leu
        115                 120                 125

gtg tcc gcc gtc gcg ttc tgc cac gcg cgc ggg gtg ttc cac cgc gac      432
Val Ser Ala Val Ala Phe Cys His Ala Arg Gly Val Phe His Arg Asp
    130                 135                 140

atc aag ccg gag aac ctc ctc gtc gac gac gcc ggc gac ctc aag gtg      480
Ile Lys Pro Glu Asn Leu Leu Val Asp Asp Ala Gly Asp Leu Lys Val
145                 150                 155                 160

tcc gac ttc ggg ctc tcc gcg gtg gcg gac ggg atg cgg cgc gac ggg      528
Ser Asp Phe Gly Leu Ser Ala Val Ala Asp Gly Met Arg Arg Asp Gly
                165                 170                 175

ctg ttc cac acg ttc tgc ggc acg ccg gcg tac gtc gcg ccg gag gtg      576
Leu Phe His Thr Phe Cys Gly Thr Pro Ala Tyr Val Ala Pro Glu Val
            180                 185                 190

ctg tcg cgc cgc ggg tac gac gcc gcc ggg gcc gac ctc tgg tcc tgc      624
Leu Ser Arg Arg Gly Tyr Asp Ala Ala Gly Ala Asp Leu Trp Ser Cys
        195                 200                 205

ggc gtc gtg ctc ttc gtc ctc atg gcc ggc tac ctc ccc ttc cag gac      672
Gly Val Val Leu Phe Val Leu Met Ala Gly Tyr Leu Pro Phe Gln Asp
    210                 215                 220

cgc aac ctc gcc ggc atg tac cgc aag atc cac aag ggc gac ttc cgc      720
Arg Asn Leu Ala Gly Met Tyr Arg Lys Ile His Lys Gly Asp Phe Arg
225                 230                 235                 240

tgc ccc aag tgg ttc tcg ccg gag ctc atc cgc ctc ctc cgc ggc gtc      768
Cys Pro Lys Trp Phe Ser Pro Glu Leu Ile Arg Leu Leu Arg Gly Val
                245                 250                 255

ctc gtc acc aac ccg cag cgc cgc gcc acc gcc gag ggg atc atg gag      816
Leu Val Thr Asn Pro Gln Arg Arg Ala Thr Ala Glu Gly Ile Met Glu
            260                 265                 270

aac gag tgg ttc aag atc ggc ttc cgc cgc ttc tcc ttc cgc gtc gag      864
Asn Glu Trp Phe Lys Ile Gly Phe Arg Arg Phe Ser Phe Arg Val Glu
        275                 280                 285

gac gac cgc acc ttc acc tgc ttc gaa ctt gac gac gac gcc gcc gtc      912
Asp Asp Arg Thr Phe Thr Cys Phe Glu Leu Asp Asp Asp Ala Ala Val
    290                 295                 300

gac gcg ccc acc tcg ccg ccg gac acg ccg cgg aca gtg gac agc ggc      960
```

```
Asp Ala Pro Thr Ser Pro Pro Asp Thr Pro Arg Thr Val Asp Ser Gly
305                 310                 315                 320

gac gtc ggc gct gct ccg acg cga cca aga aaa gcc ggg agc ctg acg      1008
Asp Val Gly Ala Ala Pro Thr Arg Pro Arg Lys Ala Gly Ser Leu Thr
                325                 330                 335

tcg tgc gac tcg gcg ccc ctg aac gcg ttc gac atc atc tcc ttc tcc      1056
Ser Cys Asp Ser Ala Pro Leu Asn Ala Phe Asp Ile Ile Ser Phe Ser
            340                 345                 350

ccg ggg ttc gac ctc tca gga ctc atc ccg gag cag cag aaa cac acg      1104
Pro Gly Phe Asp Leu Ser Gly Leu Ile Pro Glu Gln Gln Lys His Thr
        355                 360                 365

gcg agg ttc gtg tcg gcg gcg ccg gtg gag gtg atc gtg gcg acg ctg      1152
Ala Arg Phe Val Ser Ala Ala Pro Val Glu Val Ile Val Ala Thr Leu
    370                 375                 380

gag gcg gcc gcg gcg gcg gcg ggc atg gcg gtg cgg gag agg gag gac      1200
Glu Ala Ala Ala Ala Ala Ala Gly Met Ala Val Arg Glu Arg Glu Asp
385                 390                 395                 400

ggg tcg atc agc atg gag ggg aca cgc gag ggc gag cac ggc gcg ctg      1248
Gly Ser Ile Ser Met Glu Gly Thr Arg Glu Gly Glu His Gly Ala Leu
                405                 410                 415

gcg gtg gcc gcg gag atc tac gag ctc acg ccg gag ctg gtg gtg gtg      1296
Ala Val Ala Ala Glu Ile Tyr Glu Leu Thr Pro Glu Leu Val Val Val
            420                 425                 430

gag gtg cgg cgg aag gcc ggc ggc gcc gcc gag tac gag gag ttc ttc      1344
Glu Val Arg Arg Lys Ala Gly Gly Ala Ala Glu Tyr Glu Glu Phe Phe
        435                 440                 445

cgg gcg cgg ctc aag cca agc ctc cgc gag ctc gtc tgc gac gac cgg      1392
Arg Ala Arg Leu Lys Pro Ser Leu Arg Glu Leu Val Cys Asp Asp Arg
    450                 455                 460

cca tgc ccg gag gac tcc ggc gag ctc tcc cgg agc ctt tga              1434
Pro Cys Pro Glu Asp Ser Gly Glu Leu Ser Arg Ser Leu
465                 470                 475
```

<210> 22

<211> 477

<212> PRT

<213> Zea mays

<400> 22

```
Met Gly Ile Ser Lys Gly Gly Ser Gly Lys Glu Ala Lys Lys Pro Pro
1               5                   10                  15

Leu Leu Leu Gly Arg Phe Glu Val Gly Lys Leu Leu Gly Gln Gly Asn
```

20 25 30

Phe Ala Lys Val Tyr His Ala Arg Asn Val Ala Thr Gly Glu Glu Val
35 40 45

Ala Ile Lys Val Met Glu Lys Glu Lys Ile Phe Lys Ser Gly Leu Thr
50 55 60

Ala His Ile Lys Arg Glu Ile Ala Val Leu Arg Arg Val Arg His Pro
65 70 75 80

His Ile Val Gln Leu Tyr Glu Val Met Ala Thr Lys Leu Arg Ile Tyr
85 90 95

Phe Val Met Glu Tyr Val Arg Gly Gly Glu Leu Phe Ala Arg Val Ala
100 105 110

Arg Gly Arg Leu Pro Glu Ala Asp Ala Arg Arg Tyr Phe Gln Gln Leu
115 120 125

Val Ser Ala Val Ala Phe Cys His Ala Arg Gly Val Phe His Arg Asp
130 135 140

Ile Lys Pro Glu Asn Leu Leu Val Asp Asp Ala Gly Asp Leu Lys Val
145 150 155 160

Ser Asp Phe Gly Leu Ser Ala Val Ala Asp Gly Met Arg Arg Asp Gly
165 170 175

Leu Phe His Thr Phe Cys Gly Thr Pro Ala Tyr Val Ala Pro Glu Val
180 185 190

Leu Ser Arg Arg Gly Tyr Asp Ala Ala Gly Ala Asp Leu Trp Ser Cys
195 200 205

Gly Val Val Leu Phe Val Leu Met Ala Gly Tyr Leu Pro Phe Gln Asp
210 215 220

Arg Asn Leu Ala Gly Met Tyr Arg Lys Ile His Lys Gly Asp Phe Arg
225 230 235 240

Cys Pro Lys Trp Phe Ser Pro Glu Leu Ile Arg Leu Leu Arg Gly Val
245 250 255

Leu Val Thr Asn Pro Gln Arg Arg Ala Thr Ala Glu Gly Ile Met Glu
260 265 270

```
Asn Glu Trp Phe Lys Ile Gly Phe Arg Arg Phe Ser Phe Arg Val Glu
        275                 280                 285

Asp Asp Arg Thr Phe Thr Cys Phe Glu Leu Asp Asp Asp Ala Ala Val
    290                 295                 300

Asp Ala Pro Thr Ser Pro Pro Asp Thr Pro Arg Thr Val Asp Ser Gly
305                 310                 315                 320

Asp Val Gly Ala Ala Pro Thr Arg Pro Arg Lys Ala Gly Ser Leu Thr
                325                 330                 335

Ser Cys Asp Ser Ala Pro Leu Asn Ala Phe Asp Ile Ile Ser Phe Ser
            340                 345                 350

Pro Gly Phe Asp Leu Ser Gly Leu Ile Pro Glu Gln Gln Lys His Thr
        355                 360                 365

Ala Arg Phe Val Ser Ala Ala Pro Val Glu Val Ile Val Ala Thr Leu
    370                 375                 380

Glu Ala Ala Ala Ala Ala Ala Gly Met Ala Val Arg Glu Arg Glu Asp
385                 390                 395                 400

Gly Ser Ile Ser Met Glu Gly Thr Arg Glu Gly Glu His Gly Ala Leu
                405                 410                 415

Ala Val Ala Ala Glu Ile Tyr Glu Leu Thr Pro Glu Leu Val Val Val
            420                 425                 430

Glu Val Arg Arg Lys Ala Gly Gly Ala Ala Glu Tyr Glu Glu Phe Phe
        435                 440                 445

Arg Ala Arg Leu Lys Pro Ser Leu Arg Glu Leu Val Cys Asp Asp Arg
    450                 455                 460

Pro Cys Pro Glu Asp Ser Gly Glu Leu Ser Arg Ser Leu
465                 470                 475
```

<210> 23
<211> 672
<212> DNA
<213> Zea mays

<220>
<221> CDS

<222> (1)..(672)
<223> PIR2:T04358 Glutathione S-Transferase 1
glutathione transferase (EC 2.5.1.18)

<400> 23

```
atg gcc gag gag aag aag cag ggc ctg cag ctg ctg gac ttc tgg gtg      48
Met Ala Glu Glu Lys Lys Gln Gly Leu Gln Leu Leu Asp Phe Trp Val
1               5                  10                  15

agc cca ttc ggg cag cgc tgc cgc atc gcg atg gac gag aag ggc ctg      96
Ser Pro Phe Gly Gln Arg Cys Arg Ile Ala Met Asp Glu Lys Gly Leu
            20                  25                  30

gcc tac gag tac ctg gag cag gac ctg ggg aac aag agc gag ctg ctc     144
Ala Tyr Glu Tyr Leu Glu Gln Asp Leu Gly Asn Lys Ser Glu Leu Leu
        35                  40                  45

ctc cgc gcc aac ccg gtg cat aag aag atc ccc gtg ctg ctg cac gac     192
Leu Arg Ala Asn Pro Val His Lys Lys Ile Pro Val Leu Leu His Asp
    50                  55                  60

ggc cgc ccc gtc tgc gag tcc ctc gtc atc gtg cag tac ctc gac gag     240
Gly Arg Pro Val Cys Glu Ser Leu Val Ile Val Gln Tyr Leu Asp Glu
65                  70                  75                  80

gcg ttc ccg gcg gcg gcg ccg gcg ctg ctc ccc gcc gac ccc tac gcg     288
Ala Phe Pro Ala Ala Ala Pro Ala Leu Leu Pro Ala Asp Pro Tyr Ala
                85                  90                  95

cgc gcg cag gcc cgc ttc tgg gcg gac tac gtc gac aag aag ctc tac     336
Arg Ala Gln Ala Arg Phe Trp Ala Asp Tyr Val Asp Lys Lys Leu Tyr
            100                 105                 110

gac tgc ggc acc cgg ctg tgg aag ctc aag ggg gac ggc cag gcg cag     384
Asp Cys Gly Thr Arg Leu Trp Lys Leu Lys Gly Asp Gly Gln Ala Gln
        115                 120                 125

gcg cgc gcc gag atg gtc gag atc ctc cgc acg ctg gag ggc gcg ctc     432
Ala Arg Ala Glu Met Val Glu Ile Leu Arg Thr Leu Glu Gly Ala Leu
    130                 135                 140

ggc gac ggg ccc ttc ttc ggc ggc gac gcc ctc ggc ttc gtc gac gtc     480
Gly Asp Gly Pro Phe Phe Gly Gly Asp Ala Leu Gly Phe Val Asp Val
145                 150                 155                 160

gcg ctc gtg ccc ttc acg tcc tgg ttc ctc gcc tac gac cgc ttc ggc     528
Ala Leu Val Pro Phe Thr Ser Trp Phe Leu Ala Tyr Asp Arg Phe Gly
                165                 170                 175

ggc gtc agc gtg gag aag gag tgc ccg agg ctg gcc gcc tgg gcc aag     576
Gly Val Ser Val Glu Lys Glu Cys Pro Arg Leu Ala Ala Trp Ala Lys
            180                 185                 190

cgc tgc gcc gag cgc ccc agc gtc gcc aag aac ctc tac ccg ccc gag     624
Arg Cys Ala Glu Arg Pro Ser Val Ala Lys Asn Leu Tyr Pro Pro Glu
        195                 200                 205

aag gtc tac gac ttc gtc tgc ggg atg aag aag agg ctg ggc atc gag     672
Lys Val Tyr Asp Phe Val Cys Gly Met Lys Lys Arg Leu Gly Ile Glu
    210                 215                 220
```

<210> 24
<211> 224
<212> PRT
<213> Zea mays

<400> 24

Met Ala Glu Glu Lys Lys Gln Gly Leu Gln Leu Leu Asp Phe Trp Val
1 5 10 15

Ser Pro Phe Gly Gln Arg Cys Arg Ile Ala Met Asp Glu Lys Gly Leu
20 25 30

Ala Tyr Glu Tyr Leu Glu Gln Asp Leu Gly Asn Lys Ser Glu Leu Leu
35 40 45

Leu Arg Ala Asn Pro Val His Lys Lys Ile Pro Val Leu Leu His Asp
50 55 60

Gly Arg Pro Val Cys Glu Ser Leu Val Ile Val Gln Tyr Leu Asp Glu
65 70 75 80

Ala Phe Pro Ala Ala Ala Pro Ala Leu Leu Pro Ala Asp Pro Tyr Ala
85 90 95

Arg Ala Gln Ala Arg Phe Trp Ala Asp Tyr Val Asp Lys Lys Leu Tyr
100 105 110

Asp Cys Gly Thr Arg Leu Trp Lys Leu Lys Gly Asp Gly Gln Ala Gln
115 120 125

Ala Arg Ala Glu Met Val Glu Ile Leu Arg Thr Leu Glu Gly Ala Leu
130 135 140

Gly Asp Gly Pro Phe Phe Gly Gly Asp Ala Leu Gly Phe Val Asp Val
145 150 155 160

Ala Leu Val Pro Phe Thr Ser Trp Phe Leu Ala Tyr Asp Arg Phe Gly
165 170 175

Gly Val Ser Val Glu Lys Glu Cys Pro Arg Leu Ala Ala Trp Ala Lys
180 185 190

Arg Cys Ala Glu Arg Pro Ser Val Ala Lys Asn Leu Tyr Pro Pro Glu
195 200 205

```
Lys Val Tyr Asp Phe Val Cys Gly Met Lys Lys Arg Leu Gly Ile Glu
    210                 215                 220
```

<210> 25

<211> 945

<212> DNA

<213> Oryza sativa

<220>

<221> CDS

<222> (6)..(698)

<223> Glutathione S-Transferase 1

<400> 25

```
tagcc atg gcg gag gag aag gag ctg gtg ctg ctc gat ttc tgg gtg agc      50
      Met Ala Glu Glu Lys Glu Leu Val Leu Leu Asp Phe Trp Val Ser
      1               5                  10                  15

ccg ttc ggg cag cgg tgc cgg atc gcc atg gcg gag aag ggg ctg gag      98
Pro Phe Gly Gln Arg Cys Arg Ile Ala Met Ala Glu Lys Gly Leu Glu
                20                  25                  30

ttc gag tac cgc gag gag gac ctc ggc aac aag agc gac ctc ctc ctc     146
Phe Glu Tyr Arg Glu Glu Asp Leu Gly Asn Lys Ser Asp Leu Leu Leu
            35                  40                  45

cgc tcc aac ccc gtc cac agg aag atc ccc gtc ctc ctc cac gcc ggc     194
Arg Ser Asn Pro Val His Arg Lys Ile Pro Val Leu Leu His Ala Gly
        50                  55                  60

cgc ccc gtc tcc gag tcc ctc gtc atc ctc cag tac ctc gac gac gcg     242
Arg Pro Val Ser Glu Ser Leu Val Ile Leu Gln Tyr Leu Asp Asp Ala
    65                  70                  75

ttc ccc ggc acc ccc cac ctc ctc cct ccg ggg aac tcc ggc gac gcc     290
Phe Pro Gly Thr Pro His Leu Leu Pro Pro Gly Asn Ser Gly Asp Ala
80                  85                  90                  95

gac gcc gcg ttc gcg cgc gcc acg gcg agg ttc tgg gcg gac tac gtc     338
Asp Ala Ala Phe Ala Arg Ala Thr Ala Arg Phe Trp Ala Asp Tyr Val
                100                 105                 110

gac agg aag ctc tac gac tgc ggg tcc agg ctg tgg agg ctc aag ggt     386
Asp Arg Lys Leu Tyr Asp Cys Gly Ser Arg Leu Trp Arg Leu Lys Gly
            115                 120                 125

gag ccg cat gcg gcg gcg ggg cgc gag atg gcg gag atc ctc cgc acg     434
Glu Pro His Ala Ala Ala Gly Arg Glu Met Ala Glu Ile Leu Arg Thr
        130                 135                 140
```

```
ctg gag gcg gag ctc ggc gac cgg gag ttc ttc ggc ggc ggc ggc ggc      482
Leu Glu Ala Glu Leu Gly Asp Arg Glu Phe Phe Gly Gly Gly Gly Gly
    145                 150                 155

ggc agg ctc ggg ttc gtc gac gtc gcg ctc gtg ccg ttc acg gcg tgt      530
Gly Arg Leu Gly Phe Val Asp Val Ala Leu Val Pro Phe Thr Ala Cys
160                 165                 170                 175

tcc aca gct act gag agg tgc ggc ggg ttc agc gtg gag gag gtg gcg      578
Ser Thr Ala Thr Glu Arg Cys Gly Gly Phe Ser Val Glu Glu Val Ala
                180                 185                 190

ccg agg ctg gcg gcg tgg gcg cgg cgg cgc ggc cgg atc gac tcc gtc      626
Pro Arg Leu Ala Ala Trp Ala Arg Arg Arg Gly Arg Ile Asp Ser Val
            195                 200                 205

gtc aag cac ctc ccc tcg ccg gag aag gtc tac gac ttc gtc ggc gtc      674
Val Lys His Leu Pro Ser Pro Glu Lys Val Tyr Asp Phe Val Gly Val
        210                 215                 220

ctc aag aag aag tac ggc gtc gag tagatcggtg gatgcgaagt tgcagggatc      728
Leu Lys Lys Lys Tyr Gly Val Glu
    225                 230

gattggcggt tgcgttcgca acgtgaacga ttcgtccgtt gtttcagtgg ccaagtgtgt      788

gtgagtttgt tgttaccgtt gagtgcttgt gtgtgggatg gttggtggca gcagagagtt      848

gcctccgatt ctctgagata gtcactaaat aaagtttgtc ctttgaaact aaaaaaagtt      908

ggctttggtt aaaaaaaaaa aaaaaaaaaa aaaaaaa                              945
```

<210> 26

<211> 231

<212> PRT

<213> Oryza sativa

<400> 26

```
Met Ala Glu Glu Lys Glu Leu Val Leu Leu Asp Phe Trp Val Ser Pro
1               5                   10                  15

Phe Gly Gln Arg Cys Arg Ile Ala Met Ala Glu Lys Gly Leu Glu Phe
            20                  25                  30

Glu Tyr Arg Glu Glu Asp Leu Gly Asn Lys Ser Asp Leu Leu Leu Arg
        35                  40                  45

Ser Asn Pro Val His Arg Lys Ile Pro Val Leu Leu His Ala Gly Arg
    50                  55                  60
```

Pro Val Ser Glu Ser Leu Val Ile Leu Gln Tyr Leu Asp Asp Ala Phe
65 70 75 80

Pro Gly Thr Pro His Leu Leu Pro Pro Gly Asn Ser Gly Asp Ala Asp
85 90 95

Ala Ala Phe Ala Arg Ala Thr Ala Arg Phe Trp Ala Asp Tyr Val Asp
100 105 110

Arg Lys Leu Tyr Asp Cys Gly Ser Arg Leu Trp Arg Leu Lys Gly Glu
115 120 125

Pro His Ala Ala Ala Gly Arg Glu Met Ala Glu Ile Leu Arg Thr Leu
130 135 140

Glu Ala Glu Leu Gly Asp Arg Glu Phe Phe Gly Gly Gly Gly Gly Gly
145 150 155 160

Arg Leu Gly Phe Val Asp Val Ala Leu Val Pro Phe Thr Ala Cys Ser
165 170 175

Thr Ala Thr Glu Arg Cys Gly Gly Phe Ser Val Glu Glu Val Ala Pro
180 185 190

Arg Leu Ala Ala Trp Ala Arg Arg Arg Gly Arg Ile Asp Ser Val Val
195 200 205

Lys His Leu Pro Ser Pro Glu Lys Val Tyr Asp Phe Val Gly Val Leu
210 215 220

Lys Lys Lys Tyr Gly Val Glu
225 230

<210> 27
<211> 486
<212> DNA
<213> Oryza sativa

<220>
<221> CDS
<222> (1)..(486)
<223> LOCUS AF203879 peroxiredoxin ACCESSION AF203879
Thioredoxin-dependent peroxidase

<400> 27
atg gcc ccg gtt gcc gtg ggc gac acc ctc ccc gac ggc cag ctg ggg 48
Met Ala Pro Val Ala Val Gly Asp Thr Leu Pro Asp Gly Gln Leu Gly

```
1               5                   10                  15

tgg ttc gac ggg gag gac aag ctg cag cag gtc tcc gtc cac ggc ctc      96
Trp Phe Asp Gly Glu Asp Lys Leu Gln Gln Val Ser Val His Gly Leu
            20                  25                  30

gcc gcc ggc aag aag gtc gtc ctc ttc ggc gtc ccc ggt gcc ttc acc     144
Ala Ala Gly Lys Lys Val Val Leu Phe Gly Val Pro Gly Ala Phe Thr
        35                  40                  45

ccg acc tgc agc aat cag cat gtg cca gga ttc ata aat cag gct gag     192
Pro Thr Cys Ser Asn Gln His Val Pro Gly Phe Ile Asn Gln Ala Glu
    50                  55                  60

cag ctc aaa gcc aag ggt gta gac gac atc ttg ctt gtc agt gtt aac     240
Gln Leu Lys Ala Lys Gly Val Asp Asp Ile Leu Leu Val Ser Val Asn
65                  70                  75                  80

gac ccc ttt gtc atg aag gcg tgg gca aag tca tac cct gag aat aag     288
Asp Pro Phe Val Met Lys Ala Trp Ala Lys Ser Tyr Pro Glu Asn Lys
                85                  90                  95

cat gtg aaa ttc ctt gcc gat ggt ttg gga aca tac acc aag gca ctt     336
His Val Lys Phe Leu Ala Asp Gly Leu Gly Thr Tyr Thr Lys Ala Leu
            100                 105                 110

ggt ctt gag ctt gac ctt tcg gag aaa ggg ctt ggt att cgt tcg aga     384
Gly Leu Glu Leu Asp Leu Ser Glu Lys Gly Leu Gly Ile Arg Ser Arg
        115                 120                 125

cgg ttt gct ctc ctt gct gac aac ctc aag gtt act gtt gca aac att     432
Arg Phe Ala Leu Leu Ala Asp Asn Leu Lys Val Thr Val Ala Asn Ile
    130                 135                 140

gag gaa ggt ggc caa ttc aca atc tct ggt gct gag gag atc ctc aag     480
Glu Glu Gly Gly Gln Phe Thr Ile Ser Gly Ala Glu Glu Ile Leu Lys
145                 150                 155                 160

gca ctg                                                            486
Ala Leu
```

<210> 28
<211> 162
<212> PRT
<213> Oryza sativa

<400> 28

```
Met Ala Pro Val Ala Val Gly Asp Thr Leu Pro Asp Gly Gln Leu Gly
1               5                   10                  15

Trp Phe Asp Gly Glu Asp Lys Leu Gln Gln Val Ser Val His Gly Leu
            20                  25                  30

Ala Ala Gly Lys Lys Val Val Leu Phe Gly Val Pro Gly Ala Phe Thr
        35                  40                  45
```

Pro Thr Cys Ser Asn Gln His Val Pro Gly Phe Ile Asn Gln Ala Glu
50 55 60

Gln Leu Lys Ala Lys Gly Val Asp Asp Ile Leu Leu Val Ser Val Asn
65 70 75 80

Asp Pro Phe Val Met Lys Ala Trp Ala Lys Ser Tyr Pro Glu Asn Lys
85 90 95

His Val Lys Phe Leu Ala Asp Gly Leu Gly Thr Tyr Thr Lys Ala Leu
100 105 110

Gly Leu Glu Leu Asp Leu Ser Glu Lys Gly Leu Gly Ile Arg Ser Arg
115 120 125

Arg Phe Ala Leu Leu Ala Asp Asn Leu Lys Val Thr Val Ala Asn Ile
130 135 140

Glu Glu Gly Gly Gln Phe Thr Ile Ser Gly Ala Glu Glu Ile Leu Lys
145 150 155 160

Ala Leu

<210> 29

<211> 647

<212> DNA

<213> Arabidopsis thaliana

<220>

<221> CDS

<222> (60)..(548)

<223> Thioredoxin-dependent peroxidase

<400> 29
ccacgcgtcc gcaaaactct tctattttcc tctgtcttca aaaccacaga gatctcttc 59

atg gct cca att act gtc ggc gat gtt gta cca gac gga act atc tct 107
Met Ala Pro Ile Thr Val Gly Asp Val Val Pro Asp Gly Thr Ile Ser
1 5 10 15

ttc ttc gat gaa aat gat cag ctt cag acc gtc tcc gtt cac tct atc 155

```
Phe Phe Asp Glu Asn Asp Gln Leu Gln Thr Val Ser Val His Ser Ile
            20                  25                  30

gcc gcc ggt aaa aaa gtc att ctc ttt ggt gtt cct ggt gct ttc act      203
Ala Ala Gly Lys Lys Val Ile Leu Phe Gly Val Pro Gly Ala Phe Thr
        35                  40                  45

ccc aca tgc agt atg agc cat gtg cct gga ttc att ggg aaa gca gag      251
Pro Thr Cys Ser Met Ser His Val Pro Gly Phe Ile Gly Lys Ala Glu
    50                  55                  60

gag ctg aag tca aag ggt att gat gag atc att tgc ttt agt gtg aat      299
Glu Leu Lys Ser Lys Gly Ile Asp Glu Ile Ile Cys Phe Ser Val Asn
65                  70                  75                  80

gat cca ttt gtg atg aag gca tgg gga aaa aca tat cca gag aac aag      347
Asp Pro Phe Val Met Lys Ala Trp Gly Lys Thr Tyr Pro Glu Asn Lys
                85                  90                  95

cat gtg aag ttt gta gca gat ggg tct gga gaa tac acg cat ctt ctt      395
His Val Lys Phe Val Ala Asp Gly Ser Gly Glu Tyr Thr His Leu Leu
            100                 105                 110

gga ctt gag ctt gac ctt aag gac aag ggt tct ggt att agt tca ggg      443
Gly Leu Glu Leu Asp Leu Lys Asp Lys Gly Ser Gly Ile Ser Ser Gly
        115                 120                 125

aga ttc gct ttg ttg ctt gat aac ctt aag gtg act gta gcc aat gtt      491
Arg Phe Ala Leu Leu Leu Asp Asn Leu Lys Val Thr Val Ala Asn Val
    130                 135                 140

gaa tct ggt ggc gag ttc acg gtt tcc agc gca gag gat att ctc aag      539
Glu Ser Gly Gly Glu Phe Thr Val Ser Ser Ala Glu Asp Ile Leu Lys
145                 150                 155                 160

gct ctt taa gaaactttat cgtttcgctt gttgtattgt gaatctaaac              588
Ala Leu

tgctgtatgt gaagaagaga tttctatagc ttgatttcaa tcaaaaaaaa aaaaaaaaa     647
```

<210> 30

<211> 162

<212> PRT

<213> Arabidopsis thaliana

<400> 30

```
Met Ala Pro Ile Thr Val Gly Asp Val Val Pro Asp Gly Thr Ile Ser
1               5                   10                  15

Phe Phe Asp Glu Asn Asp Gln Leu Gln Thr Val Ser Val His Ser Ile
            20                  25                  30
```

Ala Ala Gly Lys Lys Val Ile Leu Phe Gly Val Pro Gly Ala Phe Thr
35 40 45

Pro Thr Cys Ser Met Ser His Val Pro Gly Phe Ile Gly Lys Ala Glu
50 55 60

Glu Leu Lys Ser Lys Gly Ile Asp Glu Ile Ile Cys Phe Ser Val Asn
65 70 75 80

Asp Pro Phe Val Met Lys Ala Trp Gly Lys Thr Tyr Pro Glu Asn Lys
85 90 95

His Val Lys Phe Val Ala Asp Gly Ser Gly Glu Tyr Thr His Leu Leu
100 105 110

Gly Leu Glu Leu Asp Leu Lys Asp Lys Gly Ser Gly Ile Ser Ser Gly
115 120 125

Arg Phe Ala Leu Leu Leu Asp Asn Leu Lys Val Thr Val Ala Asn Val
130 135 140

Glu Ser Gly Gly Glu Phe Thr Val Ser Ser Ala Glu Asp Ile Leu Lys
145 150 155 160

Ala Leu

<210> 31
<211> 1846
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (494)..(1159)
<223> Rab28 protein; abscisic acid inducible; rab28 gene
EMBL no. X59138;
PIR2 no. S18545

<220>
<221> CDS
<222> (1289)..(1456)
<223>

<400> 31
gtcgacggtg gaaactggaa acgggcctaa tggatgcggt actaagccag gtccagtaaa 60

tttgtcgtgc agagatgccc cgtggacttt agaggagccc acgataggac tctggcgcag 120

```
ttcggcctgc aagtaaaggc ccatgactcg ccgccgccgt ctggaccgct gacacgcatg    180

ccctgatgct ccccttccg ggagcttctt catccagctt gcagccggac agcccttgcg    240

ctcgcgccac gtgggcatgc cgccgcgacg cgcctcctct tgctcgtctc cacgtctctc    300

gcctctgcaa cacatgacat atgtccctcc tcgagctccc cgcacgccgc atataatcgc    360

aacgatccaa tccacactca ccatcaagtc tcaagaggca gctaaattaa accaacaagc    420

cgtgatctgt actgtagtag caacttggtt cttggtaggt gagatagtga tacagcaggt    480

agctcgagcg aga atg agc cag gag cag ccg agg agg ccg tcc ggc cat        529
                Met Ser Gln Glu Gln Pro Arg Arg Pro Ser Gly His
                1               5                   10

gag gag acg agc ggc ggc gga gag cag ggc gcc gtc cgc tac ggc gac      577
Glu Glu Thr Ser Gly Gly Gly Glu Gln Gly Ala Val Arg Tyr Gly Asp
        15                  20                  25

gtg ttc ccg gcg gtg agc ggg ggc ctc gcg gag aag ccc gtg gcg cgc      625
Val Phe Pro Ala Val Ser Gly Gly Leu Ala Glu Lys Pro Val Ala Arg
    30                  35                  40

agg acc gcc acg atg cag tcg gcg gag aac ctg gtg ttc ggc cag acg      673
Arg Thr Ala Thr Met Gln Ser Ala Glu Asn Leu Val Phe Gly Gln Thr
45                  50                  55                  60

ctc aag ggc ggc ccg gcg gcg gcc atg cag tcc gcg gcc acc acc aac      721
Leu Lys Gly Gly Pro Ala Ala Ala Met Gln Ser Ala Ala Thr Thr Asn
                65                  70                  75

gag cgc atg ggc gcc gtc ggg cac gac cag gcc acg gac gcc acc gcc      769
Glu Arg Met Gly Ala Val Gly His Asp Gln Ala Thr Asp Ala Thr Ala
            80                  85                  90

gtg cag ggc gtc acc gtc tcc gag acc cgc gtc cct ggc ggc ggc cgc      817
Val Gln Gly Val Thr Val Ser Glu Thr Arg Val Pro Gly Gly Gly Arg
        95                  100                 105

atc gtc acc gag ttc gtc gcc ggg cag gct gtc ggc cag tac ctc gcg      865
Ile Val Thr Glu Phe Val Ala Gly Gln Ala Val Gly Gln Tyr Leu Ala
    110                 115                 120

cgg gac gac gat ggc ggc ggc ggc atc gcc ggc ccc ggc gcc gga gcg      913
Arg Asp Asp Asp Gly Gly Gly Gly Ile Ala Gly Pro Gly Ala Gly Ala
125                 130                 135                 140

gga gtt gca ggt aag gat atc aca aag gtg acc atc ggc gag gcg ctc      961
Gly Val Ala Gly Lys Asp Ile Thr Lys Val Thr Ile Gly Glu Ala Leu
                145                 150                 155

gag gcg acg gcg ctc gcg gcg ggt gac gcg ccg gtg gag cgc agc gac     1009
Glu Ala Thr Ala Leu Ala Ala Gly Asp Ala Pro Val Glu Arg Ser Asp
            160                 165                 170

gcg gcc cgc atc cag gcg gcg gag gcg cgc gcc acg ggg ctg gac gcg     1057
Ala Ala Arg Ile Gln Ala Ala Glu Ala Arg Ala Thr Gly Leu Asp Ala
        175                 180                 185

aac gtg ccc ggc ggc ctg gcc cgg cag gcg cag tcg gcc gcg gcg gcc     1105
```

```
Asn Val Pro Gly Gly Leu Ala Arg Gln Ala Gln Ser Ala Ala Ala Ala
    190                 195                 200

aac tcg tgg gcg tgg gga gac gag gac aag gcc acg ctc ggc gac gtc     1153
Asn Ser Trp Ala Trp Gly Asp Glu Asp Lys Ala Thr Leu Gly Asp Val
205                 210                 215                 220

ctg gcg gtacgagtca cgaacacgac gtgccatcgt tttcgtttcg tgccgctgct     1209
Leu Ala


gctatatatc tgacagtgcg tgttggtggt gcaacagagc agagatcttt tgactatttg  1269

ttctttgtcg tacgtgcag aac gcg acg gcg agg ttg gtg gcg gac aag ccg    1321
                    Asn Ala Thr Ala Arg Leu Val Ala Asp Lys Pro
                            225                 230

gtg gag agc gcc gat gcg ttg ggg gtg gct ggc gcg gag aac cgc aac    1369
Val Glu Ser Ala Asp Ala Leu Gly Val Ala Gly Ala Glu Asn Arg Asn
    235                 240                 245

agg aac gac ggg acg gcg agg ccc gga ggc gtg gcg gcg tcc atg gct    1417
Arg Asn Asp Gly Thr Ala Arg Pro Gly Gly Val Ala Ala Ser Met Ala
250                 255                 260                 265

gcg gcc gca cgg ctc aac cgt aac gag gcg gtc tgg gag tgaagcagct     1466
Ala Ala Ala Arg Leu Asn Arg Asn Glu Ala Val Trp Glu
                270                 275

gcctggagag gagacacgtg cgtgtcctgg actctgaagt cctcgtcttt ttttttgttc  1526

gctagctagc tctgtacctc agcgcacgct ttacctacgt ccattcaggc gatcgagctg  1586

tgtaaatatg tagtatgtga cggctcagaa cgtgtcagtg tgtgtaactc gacatcaggc  1646

gatcgagctg tgtaaatatg tagtgttgta ccttcgtgca atataataaa gtaagatacg  1706

cgcgcgtcaa aagcgtgacc ggtgtaagat atactccgta tgcacataat taaggtgcat  1766

gctgtatatg gtgaagatgt gttaccccag tcgtacaaac caatatattt agccgggcgt  1826

gcggcgtgct atagtcgggt                                              1846


<210>  32
<211>  278
<212>  PRT
<213>  Zea mays

<400>  32

Met Ser Gln Glu Gln Pro Arg Arg Pro Ser Gly His Glu Glu Thr Ser
1               5                   10                  15

Gly Gly Gly Glu Gln Gly Ala Val Arg Tyr Gly Asp Val Phe Pro Ala
            20                  25                  30

Val Ser Gly Gly Leu Ala Glu Lys Pro Val Ala Arg Arg Thr Ala Thr
        35                  40                  45
```

Met Gln Ser Ala Glu Asn Leu Val Phe Gly Gln Thr Leu Lys Gly Gly
50 55 60

Pro Ala Ala Ala Met Gln Ser Ala Ala Thr Thr Asn Glu Arg Met Gly
65 70 75 80

Ala Val Gly His Asp Gln Ala Thr Asp Ala Thr Ala Val Gln Gly Val
85 90 95

Thr Val Ser Glu Thr Arg Val Pro Gly Gly Gly Arg Ile Val Thr Glu
100 105 110

Phe Val Ala Gly Gln Ala Val Gly Gln Tyr Leu Ala Arg Asp Asp Asp
115 120 125

Gly Gly Gly Gly Ile Ala Gly Pro Gly Ala Gly Ala Gly Val Ala Gly
130 135 140

Lys Asp Ile Thr Lys Val Thr Ile Gly Glu Ala Leu Glu Ala Thr Ala
145 150 155 160

Leu Ala Ala Gly Asp Ala Pro Val Glu Arg Ser Asp Ala Ala Arg Ile
165 170 175

Gln Ala Ala Glu Ala Arg Ala Thr Gly Leu Asp Ala Asn Val Pro Gly
180 185 190

Gly Leu Ala Arg Gln Ala Gln Ser Ala Ala Ala Ala Asn Ser Trp Ala
195 200 205

Trp Gly Asp Glu Asp Lys Ala Thr Leu Gly Asp Val Leu Ala Asn Ala
210 215 220

Thr Ala Arg Leu Val Ala Asp Lys Pro Val Glu Ser Ala Asp Ala Leu
225 230 235 240

Gly Val Ala Gly Ala Glu Asn Arg Asn Arg Asn Asp Gly Thr Ala Arg
245 250 255

Pro Gly Gly Val Ala Ala Ser Met Ala Ala Ala Ala Arg Leu Asn Arg
260 265 270

Asn Glu Ala Val Trp Glu
275

<210> 33
<211> 501
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)..(501)
<223> DEHYDRIN DHN1 (RAB-17 PROTEIN)

<400> 33

```
atg gag tac ggt cag cag ggg cag cac ggc cac ggc gcc acg ggc cat       48
Met Glu Tyr Gly Gln Gln Gly Gln His Gly His Gly Ala Thr Gly His
1               5                  10                  15

gtc gac cag tac ggc aac cca gtc ggc ggc gtc gag cac ggc acc ggc       96
Val Asp Gln Tyr Gly Asn Pro Val Gly Gly Val Glu His Gly Thr Gly
            20                  25                  30

ggc atg agg cac ggc acg gga acc ggc ggc atg ggc cag ctg ggt gag      144
Gly Met Arg His Gly Thr Gly Thr Gly Gly Met Gly Gln Leu Gly Glu
        35                  40                  45

cac ggc ggc gct ggc atg ggt ggc ggg cag ttc cag cct gcg agg gag      192
His Gly Gly Ala Gly Met Gly Gly Gly Gln Phe Gln Pro Ala Arg Glu
    50                  55                  60

gag cac aag acc ggc ggc atc ctg cat cgc tcc ggc agc tcc agc tcc      240
Glu His Lys Thr Gly Gly Ile Leu His Arg Ser Gly Ser Ser Ser Ser
65                  70                  75                  80

agc tcg tcg gag gac gac ggc atg ggc gga agg agg aag aag gga atc      288
Ser Ser Ser Glu Asp Asp Gly Met Gly Gly Arg Arg Lys Lys Gly Ile
                85                  90                  95

aag gag aag atc aag gag aag ctg ccc gga ggc cac aag gac gac cag      336
Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly His Lys Asp Asp Gln
            100                 105                 110

cac gcc acg gcg acg acc ggc ggc gcc tac ggg cag cag gga cac acc      384
His Ala Thr Ala Thr Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr
        115                 120                 125

ggc agc gcc tac ggg cag cag gga cac acc ggc ggc gcc tac gcc acc      432
Gly Ser Ala Tyr Gly Gln Gln Gly His Thr Gly Gly Ala Tyr Ala Thr
    130                 135                 140

ggc acc gag ggc acc ggc gag aag aaa ggc att atg gac aag atc aaa      480
Gly Thr Glu Gly Thr Gly Glu Lys Lys Gly Ile Met Asp Lys Ile Lys
145                 150                 155                 160

gag aag ctg ccc gga cag cac                                         501
Glu Lys Leu Pro Gly Gln His
                165
```

<210> 34
<211> 167
<212> PRT
<213> Zea mays

<400> 34

Met Glu Tyr Gly Gln Gln Gly Gln His Gly His Gly Ala Thr Gly His
1 5 10 15

Val Asp Gln Tyr Gly Asn Pro Val Gly Gly Val Glu His Gly Thr Gly
20 25 30

Gly Met Arg His Gly Thr Gly Thr Gly Gly Met Gly Gln Leu Gly Glu
35 40 45

His Gly Gly Ala Gly Met Gly Gly Gly Gln Phe Gln Pro Ala Arg Glu
50 55 60

Glu His Lys Thr Gly Gly Ile Leu His Arg Ser Gly Ser Ser Ser Ser
65 70 75 80

Ser Ser Ser Glu Asp Asp Gly Met Gly Gly Arg Arg Lys Lys Gly Ile
85 90 95

Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly His Lys Asp Asp Gln
100 105 110

His Ala Thr Ala Thr Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr
115 120 125

Gly Ser Ala Tyr Gly Gln Gln Gly His Thr Gly Gly Ala Tyr Ala Thr
130 135 140

Gly Thr Glu Gly Thr Gly Glu Lys Lys Gly Ile Met Asp Lys Ile Lys
145 150 155 160

Glu Lys Leu Pro Gly Gln His
165

<210> 35

<211> 1243

<212> DNA

<213> wheat

<220>

<221> CDS

<222> (51)..(1226)

<223> Dehydrin

<400> 35

```
gtaaacacat cagcactagt agatttcacg agtcagaagc tcagcgcaag atg gag        56
                                                        Met Glu
                                                        1

aac cag gca cac atc gcc ggc gag aag aag ggc atc atg gag aag atc      104
Asn Gln Ala His Ile Ala Gly Glu Lys Lys Gly Ile Met Glu Lys Ile
        5                   10                  15

aag gag aag ctc ccc ggc ggc cac ggc gac cac aag gag acc gct ggt      152
Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Lys Glu Thr Ala Gly
    20                  25                  30

acc cac ggg cac gcc gcc acg gcg acg cat ggt gcc ccg gcc acc ggt      200
Thr His Gly His Ala Ala Thr Ala Thr His Gly Ala Pro Ala Thr Gly
35                  40                  45                  50

ggt gcc tac ggg cag cag ggt cac gct gga acc acc ggc acg ggg ttg      248
Gly Ala Tyr Gly Gln Gln Gly His Ala Gly Thr Thr Gly Thr Gly Leu
                55                  60                  65

cat ggc gcc cac gcc ggc gag aag aag ggc gtg atg gag aac atc aag      296
His Gly Ala His Ala Gly Glu Lys Lys Gly Val Met Glu Asn Ile Lys
            70                  75                  80

gac aag ctc cct ggt ggc cac gag gac cac cag cag acc ggt ggc cac      344
Asp Lys Leu Pro Gly Gly His Glu Asp His Gln Gln Thr Gly Gly His
        85                  90                  95

tac ggg cag cag gga cac gcc ggc acg gcg acg cat ggc acc ccg gct      392
Tyr Gly Gln Gln Gly His Ala Gly Thr Ala Thr His Gly Thr Pro Ala
    100                 105                 110

acc gct ggc acc tat ggg caa cag ggg cat acc ggc acg gcg acg cat      440
Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala Thr His
115                 120                 125                 130

ggc acc cca gcg acc ggt ggc acc tat ggg gag cag gga cac acc gga      488
Gly Thr Pro Ala Thr Gly Gly Thr Tyr Gly Glu Gln Gly His Thr Gly
                135                 140                 145

gtg acc ggc acg ggg acg cac ggc acc ggc gag aag aag ggc ctc atg      536
Val Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly Leu Met
            150                 155                 160

gag aac atc aag gag aag ctc cct ggt ggc cat ggt gac cac cag cag      584
Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Gln Gln
        165                 170                 175

acc gct ggc acc tac ggg cag cag gga cac gtc ggc acg ggg aca cat      632
Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Val Gly Thr Gly Thr His
    180                 185                 190

ggc gcc ccg gct acc ggc ggg gcc tac ggg cag cat gaa cac gcc gga      680
Gly Ala Pro Ala Thr Gly Gly Ala Tyr Gly Gln His Glu His Ala Gly
```

```
195                 200                 205                 210

gtg gcc ggc gcg gga aca tac ggc acc ggc gag aag aag ggc gtc atg      728
Val Ala Gly Ala Gly Thr Tyr Gly Thr Gly Glu Lys Lys Gly Val Met
                215                 220                 225

gag aac atc aag gac aag ctc cct ggc ggc cac ggc gac cac cag cag      776
Glu Asn Ile Lys Asp Lys Leu Pro Gly Gly His Gly Asp His Gln Gln
            230                 235                 240

acc ggt ggc acc tac ggg cag cag gga cac acc ggc acg gcg acg cat      824
Thr Gly Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala Thr His
        245                 250                 255

ggc acc ccg gcc ggc ggc ggc acc tat gag cag cac gga cac acc ggg      872
Gly Thr Pro Ala Gly Gly Gly Thr Tyr Glu Gln His Gly His Thr Gly
    260                 265                 270

atg acc ggc acg ggg aca cac ggc acc ggc gag aag aag ggc gtc atg      920
Met Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly Val Met
275                 280                 285                 290

gag aac atc aag gag aag ctc ccc ggt ggc cac ggc gac cac cag cag      968
Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Gln Gln
                295                 300                 305

acc ggt gga gcc tac ggg cag cag gga cac acc ggc acg gcg acg cat     1016
Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr Gly Thr Ala Thr His
            310                 315                 320

ggc act ccg gct ggc ggc ggc acc tac ggg cag cat gca cac act gga     1064
Gly Thr Pro Ala Gly Gly Gly Thr Tyr Gly Gln His Ala His Thr Gly
        325                 330                 335

atg acc ggc acg gag acg cac ggc acc acg gcc acc ggc ggc acc cat     1112
Met Thr Gly Thr Glu Thr His Gly Thr Thr Ala Thr Gly Gly Thr His
    340                 345                 350

ggg cag cac gga cac gcc gga acg act ggc act ggg aca cac ggc acc     1160
Gly Gln His Gly His Ala Gly Thr Thr Gly Thr Gly Thr His Gly Thr
355                 360                 365                 370

gac ggg gtg ggc gag aag aag agc ctc atg gac aag atc aag gac aag     1208
Asp Gly Val Gly Glu Lys Lys Ser Leu Met Asp Lys Ile Lys Asp Lys
                375                 380                 385

ctg cct gga cag cac tga gcccggtgtg ccgacgg                          1243
Leu Pro Gly Gln His
            390
```

<210> 36

<211> 391

<212> PRT

<213> wheat

<400> 36

```
Met Glu Asn Gln Ala His Ile Ala Gly Glu Lys Lys Gly Ile Met Glu
1               5                  10                  15

Lys Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His Lys Glu Thr
            20                  25                  30

Ala Gly Thr His Gly His Ala Ala Thr Ala Thr His Gly Ala Pro Ala
        35                  40                  45

Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Ala Gly Thr Thr Gly Thr
    50                  55                  60

Gly Leu His Gly Ala His Ala Gly Glu Lys Lys Gly Val Met Glu Asn
65                  70                  75                  80

Ile Lys Asp Lys Leu Pro Gly Gly His Glu Asp His Gln Gln Thr Gly
                85                  90                  95

Gly His Tyr Gly Gln Gln Gly His Ala Gly Thr Ala Thr His Gly Thr
            100                 105                 110

Pro Ala Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala
        115                 120                 125

Thr His Gly Thr Pro Ala Thr Gly Gly Thr Tyr Gly Glu Gln Gly His
    130                 135                 140

Thr Gly Val Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly
145                 150                 155                 160

Leu Met Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His
                165                 170                 175

Gln Gln Thr Ala Gly Thr Tyr Gly Gln Gln Gly His Val Gly Thr Gly
            180                 185                 190

Thr His Gly Ala Pro Ala Thr Gly Gly Ala Tyr Gly Gln His Glu His
        195                 200                 205

Ala Gly Val Ala Gly Ala Gly Thr Tyr Gly Thr Gly Glu Lys Lys Gly
    210                 215                 220

Val Met Glu Asn Ile Lys Asp Lys Leu Pro Gly Gly His Gly Asp His
225                 230                 235                 240
```

```
Gln Gln Thr Gly Gly Thr Tyr Gly Gln Gln Gly His Thr Gly Thr Ala
                245                 250                 255

Thr His Gly Thr Pro Ala Gly Gly Gly Thr Tyr Glu Gln His Gly His
            260                 265                 270

Thr Gly Met Thr Gly Thr Gly Thr His Gly Thr Gly Glu Lys Lys Gly
        275                 280                 285

Val Met Glu Asn Ile Lys Glu Lys Leu Pro Gly Gly His Gly Asp His
    290                 295                 300

Gln Gln Thr Gly Gly Ala Tyr Gly Gln Gln Gly His Thr Gly Thr Ala
305                 310                 315                 320

Thr His Gly Thr Pro Ala Gly Gly Gly Thr Tyr Gly Gln His Ala His
                325                 330                 335

Thr Gly Met Thr Gly Thr Glu Thr His Gly Thr Thr Ala Thr Gly Gly
            340                 345                 350

Thr His Gly Gln His Gly His Ala Gly Thr Thr Gly Thr Gly Thr His
        355                 360                 365

Gly Thr Asp Gly Val Gly Glu Lys Lys Ser Leu Met Asp Lys Ile Lys
    370                 375                 380

Asp Lys Leu Pro Gly Gln His
385                 390
```

<210> 37
<211> 5
<212> PRT
<213> Zea mays

<400> 37

```
Ile Ser Tyr Glu Leu
1               5
```

<210> 38
<211> 15
<212> PRT
<213> Zea mays

<400> 38

Phe Asn Asn Lys Val Phe Cys Leu Met Phe Val Ala Ser Gln Lys
1 5 10 15

<210> 39
<211> 6
<212> PRT
<213> Zea mays

<400> 39

Glu Ser Thr Phe Leu Asp
1 5

<210> 40
<211>
<212> PRT
<213> Zea mays

<400> 40

<400> 3

Pro Gln Cys Ser Gln Gln Tyr Leu Ser Pro Val Thr Ala Ala Arg
1 5 10 15

<210> 41
<211>
<212> PRT
<213> Zea mays

<400> 41

Pro Ser Ala Thr Ser Thr Asn Ser Glu Thr Ala Ala Phe Ala Ser Ala
1 5 10 15

Arg

<210> 42
<211>
<212> PRT
<213> barley

<400> 42

Lys Val Ala Leu Val Thr Gly Gly Asp Ser Gly Ile Gly Arg
1 5 10

<210> 43
<211>
<212> PRT
<213> Zea mays

<400> 43

Lys Gly Leu Ala Tyr Glu Tyr Leu Glu Gln Asp Leu Gly Asn Lys
1 5 10 15

<210> 44
<211>
<212> PRT
<213> Zea mays

<400> 44

Arg Pro Gly Gly Val Ala Ala Ser Met Ala Ala Ala Ala Arg
1 5 10

<210> 45
<211>
<212> PRT
<213> Zea mays

<400> 45

Thr Gly Gly Met Arg His Gly Thr Gly Thr Thr Gly Gly Met Gly Gln
1 5 10 15

Leu Gly Glu His Gly Gly Ala Gly Met Gly Gly Gly Gln Phe Gln Pro
20 25 30

Ala Arg

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
18. Oktober 2001 (18.10.2001)

PCT

(10) Internationale Veröffentlichungsnummer
WO 01/77384 A2

(51) Internationale Patentklassifikation[7]: C12Q 1/68, G01N 33/50

(21) Internationales Aktenzeichen: PCT/IB01/00713

(22) Internationales Anmeldedatum:
6. April 2001 (06.04.2001)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
100 19 173.8 7. April 2000 (07.04.2000) DE

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: **EPIGENOMICS AG** [DE/DE]; Kastanienallee 24, 10435 Berlin (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: **OLEK, Alexander** [DE/DE]; Schröderstrasse 13, 10115 Berlin (DE). **PIEPENBROCK, Christian** [DE/DE]; Schwartzkopffstrasse 7 B, 10115 Berlin (DE). **BERLIN, Kurt** [DE/DE]; Marienkäferweg 4, 14532 Stahnsdorf (DE).

(74) Anwalt: **SCHUBERT, Klemens**; Joachimstrasse 9, 10119 Berlin (DE).

(81) Bestimmungsstaaten *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Bestimmungsstaaten *(regional)*: ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:
— *ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts*
— *mit dem Sequenzprotokollteil der Beschreibung in elektronischer Form getrennt veröffentlicht; auf Antrag vom Internationalen Büro erhältlich*

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

WO 01/77384 A2

(54) **Title:** DETECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNP'S) AND CYTOSINE-METHYLATIONS

(54) **Bezeichnung:** DETEKTION VON SNPs UND CYTOSIN-METHYLIERUNGEN

(57) **Abstract:** The invention relates to a set of oligonucleotides or peptide nucleic acid (PNA) oligomers and to a method suited for detecting cytosine methylations and single nucleotide polymorphisms (SNP's) in genomic DNA samples. Said method is used for diagnosing and/or prognosing detrimental occurrences for patients or individuals such as medical disorders.

(57) **Zusammenfassung:** Beschrieben wird ein Satz von Oligonukleotiden oder PNA-Oligomeren und ein Verfahren, das sich zur Detektion von Cytosin-Methylierungen und SNPs in genomischen DNA-Proben eignet. Dieses Verfahren dient der Diagnose und/oder Prognose nachteiliger Ereignisse für Patienten oder Individuen, wie Erkrankungen.

Detektion von SNPs und Cytosin-Methylierungen

Die vorliegende Erfindung beschreibt einen repräsentativen Satz von Oligonukleotiden oder PNA (Peptide Nucleic Acid) Oligomeren, welche sich zum gleichzeitigen Detektieren von SNPs (Single Nucleotide Polymorphisms) und Cytosin-Methylierungen in genomischen DNA-Proben zur Unterscheidung von Zelltypen besonders eignet, sowie ein verwendetes Verfahren.

Die nach den methodischen Entwicklungen der letzten Jahre in der Molekularbiologie gut studierten Beobachtungsebenen sind die Gene selbst, die Übersetzung dieser Gene in RNA und die daraus entstehenden Proteine. Wann im Laufe der Entwicklung eines Individuums welches Gen angeschaltet wird und wie Aktivieren und Inhibieren bestimmter Gene in bestimmten Zellen und Geweben gesteuert wird, ist mit hoher Wahrscheinlichkeit mit Ausmaß und Charakter der Methylierung der Gene bzw. des Genoms korrelierbar. Insofern äußern sich pathogene Zustände in einem veränderten Methylierungsmuster einzelner Gene oder des Genoms.

Stand der Technik

Das Humangenomprojekt, die Erstsequenzierung des menschlichen Genoms, wird in den nächsten Jahren abgeschlossen werden. Durch dieses Projekt wird es möglich werden, alle etwa 100.000 Gene zu identifizieren. Die Sequenzinformation öffnet ungeahnte Möglichkeiten für die Aufklärung der Genfunktionen. Dies wiederum eröffnet die Möglichkeit Pharmakogenetik und Pharmakogenomik zu betreiben. Die Pharmakogenetik und Pharmakogenomik zielt auf den Einsatz von Medikamenten in Abhängigkeit eines Genotypen. Dadurch soll die Effektivität von Medikamenten gesteigert werden. Der notwendige Zwischenschritt ist die Bestimmung der Po-

lymorphismen und Genotypen, die mit einem bestimmten Ansprechen assoziiert sind. Verlangt werden deshalb immer effizientere Genotypisierungsmethoden.

Derzeit gibt es zwei Kategorien von polymorphen Markern, die zur Genotypisierung eingesetzt werden, Mikrosatelliten und Single Nucleotide Polymorphisms (SNPs). Mikrosatelliten sind hoch polymorph, d.h. sie haben ein Vielzahl von Allelen. Sie sind dadurch charakterisiert, dass ein repetitives Sequenzelement, mit einer unterschiedlichen Anzahl Wiederholungen für unterschiedliche Allele, von konservierten Sequenzen frankiert ist. Durchschnittlich gibt es einen Mikrosatellitenmarker pro 1 Million Basen. Eine Karte von 5.000 positionierten Mikrosatellitenmarkern wurde von CEPH publiziert. Mikrosatelliten werden durch die Größenbestimmung von Produkten einer PCR mit Primern der konservierten, flankierenden Sequenz genotypisiert. Die Fluoreszenz markierten PCR Produkte werden auf Gelen aufgetrennt.

Es gibt vergleichsweise wenige beschriebene SNP Marker. Eine Karte mit 300.000 SNP Markern wird derzeit vom SNP Konsortium entwickelt und wird öffentlich zugänglich sein. Sind die SNP Marker identifiziert, können sie genomischen Positionen zugeordnet werden. Es ist angestrebt, 150.000 SNP Marker bis zum Jahr 2001 zu kartieren (Mashall, E. (1999); Science, 284, 406-407). Es gibt eine Handvoll Genotypisierungsmethoden für SNPS. Einige basieren auf der Auftrennung von Produkten auf Gelen, wie der oligonucleotide ligase assay (OLA). Er eignet sich daher eher für den mittleren Durchsatz. Andere vertrauen auf reine Hybridisierung, die jedoch nicht die gleiche Stringenz hat. DNA Arrays (DNA chip) eignen sich für die Analyse einer großen Anzahl SNPs in einer beschränkten Anzahl von Individuen. Bis jetzt sind Beispiele gezeigt worden, in denen 1.500 SNPs auf einem DNA Chip genotypi-

siert wurden. Die wirkliche Stärke von DNA Chips liegt in Ansätzen, wie der Resequenzierung und der Expressionsanalyse. Ansätze weiche Primerverlängerung anwenden sind gezeigt worden (Head, S. R. et al., (1999); Mol Cell Probes, 13(2), 81-87). Diese haben den Vorteil, wenn mit fluoreszenzmarkierten Terminatorbasen gearbeitet wird, dass die Resultate mit einem einfachen ELISA Lesegerät gesammelt werden können.

Es gibt einige SNP Genotypisierungsmethoden, die Massenspektrometrie zur Analyse verwenden. Diese haben den wesentlichen Vorteil, dass die allelspezifischen Produkte eine physische Darstellung der Produkte sind und nicht ein z. B. fluoreszierendes Signal, dass indirekt dem Produkt zugeordnet wird.

Die Matrix-assisted laser desorption/ionization time-of-flight Massenspektrometrie (MALDI) hat die Analytik von Biomolekülen revolutioniert (Karas, M. & Hillenkamp, F. Anal. Chem. 60, 2299-2301 (1988)). MALDI ist in verschiedenen Varianten zur Analyse von DNA eingesetzt worden. Die Varianten reichen von primer extension bis Sequenzierung (Liu, Y.-H., et al. Rapid Commun. Mass Spectrom. 9, 735-743 (1 995); Ch'ang, L.-Y., et al. Rapid Commun. Mass Spectrom. 9, 772-774 (1995); Little, D.P., et al. J. Mol. Med. 75, 745-750 (1997); Haff, L. & Smirnov, I.P. Genome Res. 7, 378-388 (1997), Fei, Z., Ono, T. & Smith, L.M. Nucieic Acids Res. 26, 2827-2828 (1998); Ross, P., Hall, L., Smirnov, 1. & Haff, L. Nature Biotech. 16, 1347-1351 (1998); Ross, P.L., Lee, K. & Belgrader, P. Anal. Chem. 69, 4197-4202 (1997); Griffin, T.J., Tang, W. & Smith, L.M. Nature Biotech. 15, 1368-1372 (1997)). Der größte Nachteil all dieser Methoden ist, dass alle eine gründliche Aufreinigung der Produkte vor der MALDI Analyse bedingen. Spin column Aufreinigung oder der Einsatz von

magnetic bead technology oder reversed-phase Aufreinigung sind notwendig.

Die Analyse von DNA im MALDI ist stark abhängig vom Ladungszustand des Produktes. Eine 100-fache Verbesserung der Empfindlichkeit in der MALDI Analyse kann dadurch erzielt werden, dass der Ladungszustand auf dem zu analysierenden Produkt so kontrolliert wird, dass nur eine einzige positive oder negative Überschussladung vorhanden ist. So modifizierte Produkte sind auch wesentlich weniger anfällig auf die Ausbildung von Addukten (z.B. mit Na und K, Gut, I.G. and Beck, S. (1 995) Nucleic Acids Res., 23, 1367-1373; Gut, I.G., Jeffery, W.A., Pappin, D.J.C. and Beck, S. Rapid Commun. Mass Spectrom., 11, 43-50 (1997)). Ein SNP Genotypisierungsverfahren, welches von diesen Bedingungen Gebrauch macht, mit dem Namen "GOOD Assay" wurde kürzlich vorgestellt (Sauer, S. et al., Nucleic Acids Research, Methods online, 2000, 28, el 3).

Die am häufigsten in der DNA eukaryotischer Zellen kovalent modifizierte Base ist 5-Methylcytosin. Sie spielt beispielsweise eine Rolle in der Regulation der Transkription, beim genetischen Imprinting und in der Tumorgenese. Die Identifizierung von 5-Methylcytosin als Bestandteil genetischer Information ist daher von erheblichem Interesse. 5-Methylcytosin-Positionen können jedoch nicht durch Sequenzierung identifiziert werden, da 5-Methylcytosin das gleiche Basenpaarungsverhalten aufweist wie Cytosin. Darüber hinaus geht bei einer PCR-Amplifikation die epigenetische Information, welche die 5-Methylcytosine tragen, vollständig verloren.

Eine relativ neue und die mittlerweile am häufigsten angewandte Methode zur 4 Untersuchung von DNA auf 5-Methylcytosin beruht auf der spezifischen Reaktion von Bisulfit mit Cytosin, das nach anschließender alkalischer

Hydrolyse in Uracil umgewandelt wird, welches in seinem Basenpaarungsverhalten dem Thymidin entspricht. 5-Methylcytosin wird dagegen unter diesen Bedingungen nicht modifiziert. Damit wird die ursprüngliche DNA so umgewandelt, dass Methylcytosin, welches ursprünglich durch sein Hybridisierungsverhalten vom Cytosin nicht unterschieden werden kann, jetzt durch "normale" molekularbiologische Techniken als einzig verbliebenes Cytosin beispielsweise durch Amplifikation und Hybridisierung oder Sequenzierung nachgewiesen werden kann. Alle diese Techniken beruhen auf Basenpaarung, welche jetzt voll ausgenutzt wird. Der Stand der Technik, was die Empfindlichkeit betrifft, wird durch ein Verfahren definiert, welches die zu untersuchende DNA in einer Agarose-Matrix einschließt, dadurch die Diffusion und Renaturierung der DNA (Bisulfit reagiert nur an einzelsträngiger DNA) verhindert und alle Fällungs- und Reinigungsschritte durch schnelle Dialyse ersetzt (Olek, A. et al., Nucl. Acids. Res. 1996, 24, 5064-5066). Mit dieser Methode können einzelne Zellen untersucht werden, was das Potential der Methode veranschaulicht. Allerdings werden bisher nur einzelne Regionen bis etwa 3000 Basenpaare Länge untersucht, eine globale Untersuchung von Zellen auf Tausenden von möglichen Methylierungsanalysen ist nicht möglich. Allerdings kann auch dieses Verfahren keine sehr kleinen Fragmente aus geringen Probenmengen zuverlässig analysieren. Diese gehen trotz Diffusionsschutz durch die Matrix verloren.

Eine Übersicht über die weiteren bekannten Möglichkeiten, 5-Methylcytosine nachzuweisen, kann aus dem folgenden Übersichtsartikel entnommen werden: Rein, T., DePamphilis, M. L., Zorbas, H., Nucleic Acids Res. 1998, 26, 2255. Die Bisulfit-Technik wird bisher bis auf wenige Ausnahmen (z. B. Zechnigk, M. et al., Eur. J. Hum. Gen. 1997, 5, 94-98) nur in der Forschung angewendet. Immer aber werden kurze, spezifische Stücke eines bekannten Gens nach einer

Bisulfit-Behandlung amplifiziert und entweder komplett sequenziert (Olek, A. und Walter, J., Nat. Genet. 1997, 17, 275-276) oder einzelne Cytosin-Positionen durch eine "PrimerExtension-Reaktion" (Gonzalgo, M. L. und Jones, P. A., Nucl. Acids Res. 1997, 25, 2529-2531, WO-A 95/00669) oder einen Enzymschnitt (Xiong, Z. und Laird, P. W., Nucl. Acids. Res. 1997, 25, 2532-2534) nachgewiesen. Zudem ist auch der Nachweis durch Hybridisierung beschrieben worden (Olek et al., WO-A 99/28498).

Weitere Publikationen die sich mit der Anwendung der Bisulfit-Technik zum Methylierungsnachweis bei einzelnen Genen befassen, sind: Xiong, Z. und Laird, P. W. (1997), Nucl. Acids Res. 25, 2532; Gonzalgo, M. L. und Jones, P. A. (1997), Nucl. Acids Res. 25, 2529; Grigg, S. und Clark, S. (1994), Bioassays 16, 431; Zeschnik, M. et al. (1997), Human Molecular Genetics 6, 387; Teil, R. et al. (1994), Nucl. Acids Res. 22, 695; Martin, V. et al. (1 995), Gene 157, 261 WO-A 97/46705 und WO-A 95/15373.

Eine Übersicht über den Stand der Technik in der Oligomer Array Herstellung lässt sich aus einer im Januar 1999 erschienen Sonderausgabe von Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999) und der dort zitierten Literatur entnehmen.

Für die Abtastung eines immobilisierten DNA-Arrays sind vielfach fluoreszent markierte Sonden verwendet worden. Besonders geeignet für Fluoreszenzmarkierungen ist das einfache Anbringen von Cy3 und Cy5 Farbstoffen am 5'-OH der jeweiligen Sonde. Die Detektion der Fluoreszenz der hybridisierten Sonden erfolgt beispielsweise über ein Konfokalmikroskop. Die Farbstoffe Cy3 und Cy5 sind, neben vielen anderen, kommerziell erhältlich.

Aufgabenstellung

Die vorliegende Erfindung soll einen Satz von Oligonukleotiden oder PNA-Oligomeren und ein Verfahren bereitstellen, welche sich zum gleichzeitigen Detektieren von SNPs (single nucleotide polymorphisms) und Cytosin-Methylierungen in genomischen DNA-Proben besonders eignen.

Beschreibung

Die Aufgabe wird also durch einen Satz von Oligonukleotiden oder PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und zur Detektion des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA gelöst, wobei die Basensequenzen ausgewählt sind aus SEQ-ID: 1 bis SEQ-ID: 382046.

Es ist ferner erfindungsgemäß, dass der erfindungsgemäße Satz sowohl die Basensequenzen mit der SEQ-ID: 1 bis SEQ-ID: 382046 selbst und/oder die durch Verlängerung, Verkürzung oder Veränderung der genannten Sequenzen mit der SEQ-ID: 1 bis SEQ-ID: 382046 enthält. Der erfindungsgemäße Satz kann sich erfindungsgemäß also aus unveränderten Sequenzen und/oder in erfindungsgemäßer Weise veränderten Sequenzen zusammensetzen.

Die vorliegende Erfindung beschreibt einen Satz von Oligomersonden (Oligonukleotiden und/oder PNA-Oligomeren) zur Detektion von Single Nucleotide Polymorphismen und/oder des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA, der besonders bevorzugt mindestens 10 der aufgeführten Oligonukleotid oder PNA Sequenzen ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-

ID: 382046 umfasst, oder aber mindestens 10 PNA-Oligomer oder Oligonukleotidsequenzen, die wiederum die dort aufgelisteten Sequenzen umfassen, nämlich die Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046.

In einer weiteren Variante des Verfahrens umfasst der Satz von Oligomersonden (Oligonukleotiden und/oder PNA-Oligomeren) zur Detektion von Single Nucleotid Polymorphismen und/oder des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA mindestens 100 Oligonukleotid oder PNA Sequenzen, ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, oder aber mindestens 100 PNA-Oligomere oder Oligonukleotidsequenzen, die wiederum die dort aufgelisteten Sequenzen umfassen, nämlich die Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046.

Besonders bevorzugt ist der Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen und des Cytosin-Methylierungszustandes in der chemisch vorbehandelten genomischen DNA dadurch gekennzeichnet, dass die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um eine weitere Base verlängert vorliegen, wobei die Basen A, T oder C sein können.

Besonders bevorzugt ist der Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen und des Cytosin-Methylierungszustandes in der chemisch vorbehandelten genomischen DNA wiederum dadurch gekennzeichnet, dass die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um eine weitere Base verlängert vorliegen, wobei die Basen A, T oder G sein können.

Bevorzugt ist der Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen und des Cytosin-Methylierungszustandes in der chemisch vorbehandelten genomischen DNA dadurch gekennzeichnet, dass die Basense-

quenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei weitere Base verlängert vorliegen, wobei die Basen A, T oder C sein können.

Bevorzugt ist der Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen und des Cytosin-Methylierungszustandes in der chemisch vorbehandelten genomischen DNA dadurch gekennzeichnet, dass die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei weitere Base verlängert vorliegen, wobei die Basen A, T oder G sein können.

Der Satz von PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen und des Cytosin-Methylierungszustandes in der chemisch vorbehandelten genomischen DNA ist besonders bevorzugt dadurch gekennzeichnet, dass jeweils am 5'-Ende und/oder am 3'-Ende des Oligomers eine Nukleobase weggelassen wird.

Der Satz von PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen und des Cytosin-Methylierungszustandes in der chemisch vorbehandelten genomischen DNA ist vorzugsweise dadurch gekennzeichnet, dass jeweils am 5'-Ende und/oder am 3'-Ende des Oligomers mindestens zwei Nukleobasen weggelassen werden.

Ein repräsentativer Satz von Oligonukleotiden und/oder PNA-Oligomeren, umfassend Oligomere und/oder Oligonukleotide gemäß den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, soll zur Detektion von Cytosin-Methylierungen und Einzelnukleotidpolymorphismen in genomischer DNA zur Unterscheidung von Zelltypen oder Geweben oder zur Untersuchung der Zelldifferenzierung verwendet werden. Dazu werden die folgenden Verfahrensschritte nacheinander ausgeführt:

Im ersten Verfahrensschritt wird eine genomische DNA Probe derart chemisch behandelt, dass an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

Die zu analysierende genomische DNA wird bevorzugt aus üblichen Quellen für DNA erhalten, wie z. B. Zelllinien, Blut, Sputum, Stuhl, Urin, Gehirn-Rückenmarks-Flüssigkeit, in Paraffin eingebettetes Gewebe, histologische Objektträger und alle möglichen Kombinationen hiervon.

Bevorzugt wird dazu die oben beschriebene Behandlung genomischer DNA mit Bisulfit (Hydrogensulfit, Disulfit) und anschließender alkalischer Hydrolyse verwendet, die zu einer Umwandlung nicht methylierter Cytosin-Nukleobasen in Uracil führt.

In einer bevorzugten Variante des Verfahrens führt man die Amplifikation mittels der Polymerasekettenreaktion (PCR) durch, wobei eine thermostabile DNA-Polymerase verwendet wird.

In einem zweiten Verfahrensschritt werden aus der chemisch vorbehandelten genomischen DNA mehr als zehn unterschiedliche Fragmente amplifiziert, die jeweils weniger als 2000 Basenpaare lang sind unter Verwendung synthetischer Oligonukleotide als Primer.

In einer besonders bevorzugten Variante des Verfahrens sind die Oligonukleotide oder PNA-Oligomere an definierten Stellen an eine Festphase gebunden.

In einer wiederum bevorzugten Variante des Verfahrens sind unterschiedliche Oligonukleotid und/oder PNA-

Oligomersequenzen auf einer ebenen Festphase in Form eines rechtwinkligen oder hexagonalen Gitters angeordnet.

Die Festphasenoberfläche besteht bevorzugt aus Silizium, Glas, Polystyrol, Aluminium, Stahl, Eisen, Kupfer, Nickel, Silber oder Gold.

Im dritten Verfahrensschritt hybridisiert man die Amplifikate an einen Satz von Oligonukleotiden oder PNA-Oligomeren, die mindestens 10 der oben genannten Sequenzen umfassen, nämlich die Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046.

In einer bevorzugten Variante des Verfahrens führt man die Amplifikation von mehreren DNA-Abschnitten in einem Reaktionsgefäß durch.

In einer bevorzugten Variante des Verfahrens liegen die Basensequenzen des erfindungsgemäßen Satzes von Oligonukleotiden, nämlich die Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um eine weitere Base verlängert vor, wobei die Basen entweder A, T oder G sein können.

In einer bevorzugten Variante des Verfahrens liegen die Basensequenzen des erfindungsgemäßen Satzes von Oligonukleotiden mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei weitere Base verlängert vor, wobei die Basen entweder A, T oder C sein können.

In einer bevorzugten Variante des Verfahrens liegen die Basensequenzen des erfindungsgemäßen Satzes von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen und des Cytosin-Methylierungszustandes mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei

weitere Basen verlängert vor, wobei die Basen entweder A, T oder G sein können.

In einer bevorzugten Variante des Verfahrens wird ein Satz von PNA-Oligomeren aus den oben genannten Basensequenzen verwendet, wobei jeweils am 5'-Ende und/oder am 3'-Ende des Oligomers eine Nukleobase weggelassen wird.

In einer bevorzugten Variante des Verfahrens wird ein Satz von PNA-Oligomeren aus den oben genannten Basensequenzen verwendet, wobei jeweils am 5'-Ende und/oder am 3'-Ende des Oligomers mindestens zwei Nukleobasen weggelassen werden.

In einer weiteren bevorzugten Variante des Verfahrens werden mindestens 10 der oben genannten Oligonukleotid oder PNA Sequenzen, nämlich ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, zur Detektion des Cytosin-Methylierungszustandes verwendet, oder aber mindestens 10 PNA-Oligomer oder Oligonukleotidsequenzen, die wiederum die oben genannten Sequenzen umfassen, nämlich die Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046.

In einer weiteren bevorzugten Variante des Verfahrens werden mindestens 100 der oben genannten Oligonukleotid oder PNA Sequenzen, nämlich ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, zur Detektion des Cytosin-Methylierungszustandes verwendet, oder aber mindestens 100 PNA-Oligomer oder Oligonukleotidsequenzen, die wiederum die oben genannten Sequenzen umfassen, nämlich die Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046.

In einer wiederum bevorzugten Variante des Verfahrens ist mindestens ein Primer an eine Festphase gebunden.

In einer erneut bevorzugten Variante des Verfahrens sind unterschiedliche Amplifikate auf der Festphase in Form eines rechtwinkligen oder hexagonalen Gitters angeordnet.

Die Festphasenoberfläche besteht bevorzugt aus Silizium, Glas, Polystyrol Aluminium, Stahl, Eisen, Kupfer, Nickel, Silber oder Gold.

Im letzten Schritt des Verfahrens detektiert man die hybridisierten Amplifikate. Die an den Amplifikaten angebrachten Markierungen sind an jeder Position der Festphase identifizierbar, an der sich eine Oligonukleotidsequenz befindet.

In einer bevorzugten Variante des Verfahrens sind die Markierungen der Amplifikate Fluoreszenzmarkierungen.

In einer bevorzugten Variante des Verfahrens sind die Markierungen der Amplifikate Radionuklide.

In einer weiteren bevorzugten Variante des Verfahrens tragen die Amplifikate ablösbare Massenmarkierungen, die in einem Massenspektrometer nachgewiesen werden.

In einer wiederum bevorzugten Variante des Verfahrens werden die Amplifikate, Fragmente der Amplifikate oder zu den Amplifikaten komplementäre Sonden im Massenspektrometer nachgewiesen.

In einer erneut bevorzugten Variante des Verfahrens weisen die erzeugten Fragmente im Massenspektrometer zur besseren Detektierbarkeit eine einzelne positive oder negative Nettoladung auf.

Der erfindungsgemäße Satz von Oligonukleotiden und/oder PNA-Oligomeren dient bevorzugt zur Diagnose und/oder

Prognose nachteiliger Ereignisse für Patienten oder Individuen.

Vorzugsweise wird der erfindungsgemäße Satz von Oligonukleotiden und/oder PNA-Oligomeren zur Diagnose und/oder Prognose nachteiliger Ereignisse für Patienten oder Individuen verwendet, wobei diese nachteiligen Ereignisse mindestens einer der folgenden Kategorien angehören: unerwünschte Arzneimittelwirkungen; Krebserkrankungen; CNS-Fehlfunktionen, Schäden oder Krankheit; aggressive Symptome oder Verhaltensstörungen; klinische, psychologische und soziale Konsequenzen von Gehirnverletzungen; psychotische Störungen und Persönlichkeitsstörungen; Demenz und/oder assoziierte Syndrome; kardiovaskuläre Krankheit, Fehlfunktion und Schädigung; Fehlfunktion, Schädigung oder Krankheit des gastrointestinalen Traktes; Fehlfunktion, Schädigung oder Krankheit des Atmungssystems; Verletzung, Entzündung, Infektion, Immunität und/oder Rekonvaleszenz; Fehlfunktion, Schädigung oder Krankheit des Körpers als Abweichung im Entwicklungsprozess; Fehlfunktion, Schädigung oder Krankheit der Haut, der Muskeln, des Bindegewebes oder der Knochen; endokrine und metabolische Fehlfunktion, Schädigung oder Krankheit; Kopfschmerzen oder sexuelle Fehlfunktion.

Vorzugsweise wird der erfindungsgemäße Satz von Oligonukleotiden und/oder PNA-Oligomeren zur Unterscheidung von Zelltypen oder Geweben oder zur Untersuchung der Zelldifferenzierung verwendet.

Weiterer Gegenstand der Erfindung ist ein Kit, das mindestens 10 Oligonukleotide oder PNA-Oligomere und Primer zur Herstellung der Amplifikate sowie eine Anleitung zur Durchführung des Verfahrens enthält.

Erfindungsgemäß bevorzugt ist ein Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA , wo die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um eine weitere Base verlängert vorliegen, wobei die Basen entweder A, T oder C sein können.

Erfindungsgemäß bevorzugt ist ein Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA, wo die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um eine weitere Base verlängert vorliegen, wobei die Basen entweder A, T oder G sein können.

Erfindungsgemäß bevorzugt ist ein Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA, wo die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei weitere Basen verlängert vorliegen, wobei die Basen entweder A, T oder C sein können.

Erfindungsgemäß bevorzugt ist ein Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA , wo die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei weitere Basen verlängert vorliegen, wobei die Basen entweder A, T oder G sein können.

Erfindungsgemäß bevorzugt ist ein Satz von PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA , wo jeweils am 5'-Ende und/oder am 3'-Ende des Oligomers eine Nukleobase weggelassen wird.

Erfindungsgemäß bevorzugt ist ein Satz von PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA, wo jeweils am 5'- und/oder am 3'-Ende des Oligomers mindestens zwei Nukleobasen weggelassen werden.

Erfindungsgemäß bevorzugt ist ein Satz von Oligomersonden (Oligonukleotide und/oder PNA-Oligomere) zur Detektion des Cytosin-Methylierungszustandes und/oder von Single Nucleotide Polymorphismen in chemisch vorbehandelter genomischer DNA, umfassend mindestens 10 der oben genannten Oligonukleotid oder PNA Sequenzen.

Erfindungsgemäß bevorzugt ist ein Satz von Oligomersonden (Oligonukleotide und/oder PNA-Oligomere) zur Detektion des Cytosin-Methylierungszustandes und/oder von Single Nucleotide Polymorphismen in chemisch vorbehandelter genomischer DNA, umfassend mindestens 100 der oben nachstehend genannten Oligonukleotid oder PNA Sequenzen.

Gegenstand der vorliegenden Erfindung ist zudem ein Verfahren zur Analyse eines repräsentativen Satzes von Cytosin-Methylierungen und Single Nucleotide Polymorphismen in genomischen DNA-Proben zur Unterscheidung von Zelltypen.

Im ersten Schritt des Verfahrens wandelt man in einer genomischen DNA Probe durch chemische Behandlung an der 5-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um.

Im zweiten Schritt des Verfahrens amplifiziert man aus dieser chemisch behandelten genomischen DNA mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind unter Verwendung synthetischer Oligonukleotide als Primer.

Im dritten Schritt des Verfahrens hybridisiert man die Amplifikate an einen Satz von Oligonukleotiden oder PNA-Oligomeren, umfassend mindestens 10 der oben genannten Sequenzen ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046 oder aber Sequenzen, welche in der oben beschriebenen Weise verlängert, verkürzt oder verändert wurden.

Im vierten Verfahrensschritt entfernt man die nicht hybridisierten Amplifikate.

Im letzten Verfahrensschritt detektiert man die hybridisierten Amplifikate.

Erfindungsgemäß bevorzugt ist es, dass man die chemische Behandlung mittels einer Lösung eines Bisulfits, Hydrogensulfits oder Disulfits durchfuhrt.

Erfindungsgemäß bevorzugt ist es, dass die Amplifikation mittels der Polymerasekettenreaktion (PCR) durchgeführt wird.

Erfindungsgemäß bevorzugt ist es, dass die Oligonukleotide oder PNA-Oligomere an definierten Stellen an eine Festphase gebunden sind.

Erfindungsgemäß bevorzugt ist es, dass unterschiedliche Oligonukleotid und/oder PNA-Oligomersequenzen auf einer ebenen Festphase in Form eines rechtwinkligen oder hexagonalen Gitters angeordnet sind.

Erfindungsgemäß bevorzugt ist es, dass an den Amplifikaten angebrachte Markierungen an jeder Position der Festphase, an der sich eine Oligonukleotidsequenz befindet, identifizierbar sind.

Erfindungsgemäß bevorzugt ist es, dass bei der Amplifikation mindestens ein Primer an eine Festphase gebunden ist.

Erfindungsgemäß bevorzugt ist es, dass unterschiedliche Amplifikate auf der Festphase in Form eines rechtwinkligen oder hexagonalen Gitters angeordnet sind.

Erfindungsgemäß bevorzugt ist es, dass die Markierungen der Amplifikate Fluoreszenzmarkierungen sind.

Erfindungsgemäß bevorzugt ist es, dass die Markierungen der Amplifikate Radionuklide sind.

Erfindungsgemäß bevorzugt ist es, dass die Amplifikate ablösbare Massenmarkierungen tragen, die in einem Massenspektrometer nachgewiesen werden.

Erfindungsgemäß bevorzugt ist es, dass die Amplifikate, Fragmente der Amplifikate oder zu den Amplifikaten komplementäre Sonden im Massenspektrometer nachgewiesen werden.

Erfindungsgemäß bevorzugt ist es, dass zur besseren Detektierbarkeit im Massenspektrometer die erzeugten Fragmente eine einzelne positive oder negative Nettoladung aufweisen.

Erfindungsgemäß bevorzugt ist es, dass man die Detektion mittels Matrix assistierter Laser Desorptions/Ionisations Massenspektrometrie (MALDI) oder mittels Elektrospray Massenspektrometrie (ESI) durchführt und visualisiert.

Erfindungsgemäß bevorzugt ist es, dass die Polymerasen hitzebeständige DNA-Polymerasen sind.

Erfindungsgemäß bevorzugt ist es, dass die Amplifikation von mehreren DNA-Abschnitten in einem Reaktionsgefäß durchfuhrt wird.

Erfindungsgemäß bevorzugt ist es, dass die Festphasenoberfläche aus Silizium, Glas, Polystyrol, Aluminium, Stahl, Eisen, Kupfer, Nickel, Silber oder Gold besteht.

Erfindungsgemäß bevorzugt ist es, dass die genomische DNA aus einer DNA-Probe erhalten wurde, wobei Quellen für DNA z. B. Zelllinien, Blut, Sputum, Stuhl, Urin, Gehirn-Rückenmarks-Flüssigkeit, in Paraffin eingebettetes Gewebe, histologische Objektträger und alle möglichen Kombinationen hiervon umfasst.

Gegenstand der Erfindung ist zudem die Verwendung eines Satzes von mindestens 10 der oben genannten Oligonukleotide und/oder PNA-Oligomere, ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, oder aber von mindestens 10 Oligomeren oder Oligonukleotiden, welche die oben genannten Sequenzen umfassen, zur Diagnose und/oder Prognose nachteiliger Ereignisse für Patienten oder Individuen.

Erfindungsgemäß bevorzugt ist die Verwendung eines Satzes von mindestens 10 der oben genannten Oligonukleotide und/oder PNA-Oligomere, ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, oder aber von mindestens 10 Oligomeren oder Oligonukleotiden, welche die oben genannten Sequenzen umfassen, zur Diagnose und/oder Prognose nachteiliger Ereignisse für Patienten oder Individuen, wobei diese nachteiligen Ereignisse mindestens einer folgenden Kategorien angehören: unerwünschte Arzneimittelwirkungen', Krebserkrankungen; CNS-Fehlfunktionen, Schäden oder Krankheit; aggressive Symptome oder Verhaltensstörungen; klinische, psychologische und soziale Konsequenzen von Gehirnverletzungen, psychotische Störungen und Persönlichkeitsstörungen; Demenz und/oder assoziierte Syndrome; kardiovaskuläre Krankheit, Fehlfunktion und Schädigung; Fehlfunktion, Schädigung oder Krankheit des gastrointestinalen Traktes, Fehlfunktion, Schädigung oder Krankheit des Atmungssystems; Verletzung, Entzündung, Infektion, Immunität und/oder Rekonvaleszenz; Fehlfunktion, Schädigung oder Krankheit des Körpers als Abweichung im Entwicklungsprozess; Fehlfunktion, Schädigung oder Krankheit der Haut, der Muskeln, des Bindegewebes oder der Knochen; endokrine und metabolische Fehlfunktion, Schädigung oder Krankheit; Kopfschmerzen oder sexuelle Fehlfunktion.

Erfindungsgemäß bevorzugt ist zudem die Verwendung eines Satzes von mindestens 10 der oben genannten Oligonukleotide und/oder PNA-Oligomere, ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, oder aber von mindestens 10 Oligomeren oder Oligonukleotiden, welche die oben genannten Sequenzen umfassen, zur Unterscheidung von Zelltypen oder Geweben oder zur Untersuchung der Zelldifferenzierung.

Gegenstand der vorliegenden Erfindung ist zudem ein Kit, enthaltend mindestens 10 der oben genannten Oligonukleotide und/oder PNA-Oligomere, ausgewählt aus den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046, oder aber von mindestens 10 Oligomeren oder Oligonukleotiden, welche die oben genannten Sequenzen umfassen, und Primer zur Herstellung der Amplifikate sowie eine Anleitung zur Durchführung des erfindungsgemäßen Verfahrens.

Das Sequenzprotokoll mit den Sequenzen SEQ-ID: 1 bis SEQ-ID: 382046 liegt der Internationalen Anmeldung in elektronisch lesbarer Form bei und ist Bestandteil dieser Anmeldung.

Patentansprüche

1. Satz von Oligonukleotiden oder PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und zur Detektion des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA, ausgewählt aus den Basensequenzen SEQ-ID: 1 bis SEQ-ID: 382046.

2. Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA nach Anspruch 1, dadurch gekennzeichnet, dass die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um eine weitere Base verlängert vorliegen, wobei die Basen entweder A, T oder C sein können.

3. Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA nach Anspruch 1, dadurch gekennzeichnet, dass die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um eine weitere Base verlängert vorliegen, wobei die Basen entweder A, T oder G sein können.

4. Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA nach Anspruch 1, dadurch gekennzeichnet, dass die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei weitere Basen verlängert

vorliegen, wobei die Basen entweder A, T oder C sein können.

5. Satz von Oligonukleotiden zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA nach Anspruch 1, dadurch gekennzeichnet, dass die Basensequenzen mehrheitlich am 5'-Ende und/oder am 3'-Ende jeweils um mindestens zwei weitere Basen verlängert vorliegen, wobei die Basen entweder A, T oder G sein können.

6. Satz von PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA nach Anspruch 1, dadurch gekennzeichnet, dass jeweils am 5'- und/oder am 3'-Ende des Oligomers eine Nukleobase weggelassen wird.

7. Satz von PNA (Peptide Nucleic Acid) Oligomeren zur Detektion von Einzelnukleotidpolymorphismen (SNPs, single nucleotide polymorphisms) und des Cytosin-Methylierungszustandes in chemisch vorbehandelter genomischer DNA nach Anspruch 1, dadurch gekennzeichnet, dass jeweils am 5'- und/oder am 3'-Ende des Oligomers mindestens zwei Nukleobasen weggelassen werden.

8. Satz von Oligomersonden (Oligonukleotide und/oder PNA-Oligomere) zur Detektion des Cytosin-Methylierungszustandes und/oder von Single Nucleotide Polymorphismen in chemisch vorbehandelter genomischer DNA, umfassend mindestens 10 der Oligonukleotid oder PNA Sequenzen der Ansprüche 1 bis 7.

9. Satz von Oligomersonden (Oligonukleotide und/oder PNA-Oligomere) zur Detektion des Cytosin-Methylierungszustandes und/oder von Single Nucleotide Polymorphismen in chemisch vorbehandelter genomischer DNA, umfassend mindestens 100 der Oligonukleotid oder PNA Sequenzen der Ansprüche 1 bis 7.

10. Verfahren zur Analyse eines repräsentativen Satzes von Cytosin-Methylierungen und Single Nucleotide Polymorphismen in genomischen DNA-Proben zur Unterscheidung von Zelltypen, dadurch gekennzeichnet, dass man folgende Schritte ausführt:

a) in einer genomischen DNA Probe wandelt man durch chemische Behandlung an der 5-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um;

b) aus dieser chemisch behandelten genomischen DNA amplifiziert man mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind unter Verwendung synthetischer Oligonukleotide als Primer;

c) man hybridisiert die Amplifikate an einen Satz von Oligonukleotiden oder PNA-Oligomeren, umfassend mindestens 10 Sequenzen der Ansprüche 1 bis 9;

d) man entfernt die nicht hybridisierten Amplifikate;

e) man detektiert die hybridisierten Amplifikate.

11. Verfahren nach Anspruch 10, dadurch gekennzeichnet, dass man die chemische Behandlung mittels einer Lösung eines Bisulfits, Hydrogensulfits oder Disulfits durchfuhrt.

12. Verfahren nach Anspruch 10 oder 11 dadurch gekennzeichnet, dass die Amplifikation mittels der Polymerasekettenreaktion (PCR) durchgeführt wird.

13. Verfahren nach einem der Ansprüche 10 bis 12 dadurch gekennzeichnet, dass die Oligonukleotide oder PNA-Oligomere an definierten Stellen an eine Festphase gebunden sind.

14. Verfahren nach Anspruch 13, dadurch gekennzeichnet, dass unterschiedliche Oligonukleotid und/oder PNA-Oligomersequenzen auf einer ebenen Festphase in Form eines rechtwinkligen oder hexagonalen Gitters angeordnet sind.

15. Verfahren nach Anspruch 13 oder 14, dadurch gekennzeichnet, dass an den Amplifikaten angebrachte Markierungen an jeder Position der Festphase, an der sich eine Oligonukleotidsequenz befindet, identifizierbar sind.

16. Verfahren nach einem der Ansprüche 10 bis 12, dadurch gekennzeichnet, dass bei der Amplifikation mindestens ein Primer an eine Festphase gebunden ist.

17. Verfahren nach Anspruch 16, dadurch gekennzeichnet, dass unterschiedliche Amplifikate auf der Festphase in Form eines rechtwinkligen oder hexagonalen Gitters angeordnet sind.

18. Verfahren nach einem der Ansprüche 10 bis 12 oder 15, dadurch gekennzeichnet, dass die Markierungen der Amplifikate Fluoreszenzmarkierungen sind.

19. Verfahren nach einem der Ansprüche 10 bis 12 oder 15, dadurch gekennzeichnet, dass die Markierungen der Amplifikate Radionuklide sind.

20. Verfahren nach einem der Ansprüche 10 bis 17, dadurch gekennzeichnet, dass die Amplifikate ablösbare Massenmarkierungen tragen, die in einem Massenspektrometer nachgewiesen werden.

21. Verfahren nach einem der Ansprüche 10 bis 17 dadurch gekennzeichnet, dass die Amplifikate, Fragmente der Amplifikate oder zu den Amplifikaten komplementäre Sonden im Massenspektrometer nachgewiesen werden.

22. Verfahren nach einem der Ansprüche 20 bis 21 dadurch gekennzeichnet, dass zur besseren Detektierbarkeit im Massenspektrometer die erzeugten Fragmente eine einzelne positive oder negative Nettoladung aufweisen.

23. Verfahren gemäß einem der Ansprüche 20 bis 22 dadurch gekennzeichnet, dass man die Detektion mittels Matrix assistierter Laser Desorptions/Ionisations Massenspektrometrie (MALDI) oder mittels Elektrospray Massenspektrometrie (ESI) durchfuhrt und visualisiert.

24. Verfahren nach einem der voranstehenden Ansprüche, wobei die Polymerasen hitzebeständige DNA-Polymerasen sind.

25. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, dass die Amplifikation von meh-

reren DNA-Abschnitten in einem Reaktionsgefäß durchführt wird.

26. Verfahren nach einem der Ansprüche 14 bis 17 dadurch gekennzeichnet, dass die Festphasenoberfläche aus Silizium, Glas, Polystyrol, Aluminium, Stahl, Eisen, Kupfer, Nickel, Silber oder Gold besteht.

27. Verfahren nach einem der Ansprüche 10 bis 26, wobei die genomische DNA aus einer DNA-Probe erhalten wurde, wobei Quellen für DNA z. B. Zelllinien, Blut, Sputum, Stuhl, Urin, Gehirn-Rückenmarks-Flüssigkeit, in Paraffin eingebettetes Gewebe, histologische Objektträger und alle möglichen Kombinationen hiervon umfasst.

28. Verwendung eines Satzes von Oligonukleotiden und/oder PNA-Oligomeren gemäss einem der Ansprüche 1 bis 9 zur Diagnose und/oder Prognose nachteiliger Ereignisse für Patienten oder Individuen.

29. Verwendung eines Satzes von Oligonukleotiden und/oder PNA-Oligomeren gemäss Anspruch 28 zur Diagnose und/oder Prognose nachteiliger Ereignisse für Patienten oder Individuen, wobei diese nachteiligen Ereignisse mindestens einer folgenden Kategorien angehören: unerwünschte Arzneimittelwirkungen; Krebserkrankungen; CNS-Fehlfunktionen, Schäden oder Krankheit; aggressive Symptome oder Verhaltensstörungen; klinische, psychologische und soziale Konsequenzen von Gehirnverletzungen; psychotische Störungen und Persönlichkeitsstörungen; Demenz und/oder assoziierte Syndrome; kardiovaskuläre Krankheit, Fehlfunktion und Schädigung; Fehlfunktion, Schädigung oder Krankheit des gastrointestinalen Traktes; Fehlfunktion, Schädigung oder Krankheit des Atmungssystems; Verletzung,

Entzündung, Infektion, Immunität und/oder Rekonvaleszenz; Fehlfunktion, Schädigung oder Krankheit des Körpers als Abweichung im Entwicklungsprozess; Fehlfunktion, Schädigung oder Krankheit der Haut, der Muskeln, des Bindegewebes oder der Knochen; endokrine und metabolische Fehlfunktion, Schädigung oder Krankheit; Kopfschmerzen oder sexuelle Fehlfunktion.

30. Verwendung eines Satzes von Oligonukleotiden und/oder PNA-Oligomeren nach einem der Ansprüche 1 bis 9 zur Unterscheidung von Zelltypen oder Geweben oder zur Untersuchung der Zelldifferenzierung.

31. Kit, enthaltend mindestens 10 Oligonukleotide oder PNA-Oligomere gemäss Anspruch 1 und Primer zur Herstellung der Amplifikate sowie eine Anleitung zur Durchführung des Verfahrens nach einem der Ansprüche 10 bis 27.

XP-002421536

Plant Physiol. (1998) 116: 1367–1377

Characterization of a Gene for Spinach CAP160 and Expression of Two Spinach Cold-Acclimation Proteins in Tobacco[1]

Claudia Kaye[2], Lisa Neven[3], Andrea Hofig[4], Qin-Bao Li, Dale Haskell, and Charles Guy*

Plant Molecular and Cellular Biology Program, Department of Environmental Horticulture, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, Florida 32611–0670

The cDNA sequence for CAP160, an acidic protein previously linked with cold acclimation in spinach (*Spinacia oleracea* L.), was characterized and found to encode a novel acidic protein of 780 amino acids having very limited homology to a pair of *Arabidopsis thaliana* stress-regulated proteins, rd29A and rd29B. The lack of similarity in the structural organization of the spinach and Arabidopsis genes highlights the absence of a high degree of conservation of this cold-stress gene across taxonomic boundaries. The protein has several unique motifs that may relate to its function during cold stress. Expression of the CAP160 mRNA was increased by low-temperature exposure and water stress in a manner consistent with a probable function during stresses that involve dehydration. The coding sequences for CAP160 and CAP85, another spinach cold-stress protein, were introduced into tobacco (*Nicotiana tabacum*) under the control of the 35S promoter using *Agrobacterium tumefaciens*-based transformation. Tobacco plants expressing the proteins individually or coexpressing both proteins were evaluated for relative freezing-stress tolerance. The killing temperature for 50% of the cells of the transgenic plants was not different from that of the wild-type plants. As determined by a more sensitive time/temperature kinetic study, plants expressing the spinach proteins had slightly lower levels of electrolyte leakage than wild-type plants, indicative of a small reduction of freezing-stress injury. Clearly, the heterologous expression of two cold-stress proteins had no profound influence on stress tolerance, a result that is consistent with the quantitative nature of cold-stress-tolerance traits.

Although the concept that cold acclimation in plants involves altered gene expression during the induction of freezing tolerance is more than 20 years old, it has only been in recent years that most of the information concerning the influence of low temperature on gene expression has become available (Thomashow, 1994; Hughes and Dunn, 1996). Similar findings with cold-water fishes (Lin and Gross, 1981), bacteria (Goldstein et al., 1990), and yeast (Kondo and Inouye, 1991) parallel the growing body of information on low-temperature-regulated genes in plants. In plants a significant portion of the genes demonstrated to be regulated in response to low temperature are, as yet, without a known function. Some, based on sequence homologies (Hahn and Walbot, 1989; Gilmour et al., 1992; Guo et al., 1992; Houde et al., 1992; Sutton et al., 1992; Neven et al., 1993), belong to a superfamily of developmentally and environmentally regulated proteins linked with reduced water activities (Dure, 1993; Close, 1996). Others have properties in common with this superfamily, such as boiling solubility, regulation by ABA, or responsiveness to water stress (Cattivelli and Bartels, 1990; Kurkela and Franck, 1990; Lin and Thomashow, 1992; Orr et al., 1992; Nordin et al., 1993; Yamaguchi-Shinozaki and Shinozaki, 1993a, 1993b).

Freezing stress is a desiccative physical stress (Pearce, 1988), and during slow equilibrium freezing in plants, unfrozen water is withdrawn from the inside of the cell by the lower vapor pressure of extracellular ice (Mazur, 1963). This causes progressive dehydration, which leads to a drastic loss of cellular volume and an increase in the concentration of intracellular solutes. The extreme dehydration that can result from freezing has major effects on cellular components such as membrane structure (Steponkus, 1984; Steponkus et al., 1988), which in turn reduces the ability of the cell to survive a freeze/thaw cycle. Consequently, during the course of evolution plants have acquired adaptive mechanisms to help ensure their survival in situations in which freezing occurs (Sakai and Larcher, 1987; Guy et al., 1992). The idea that many of these proteins might function not only during water stress but also during freezing stress to afford some manner of protection is appealing but largely untested. Previous work with spinach (*Spinacia oleracea* L.) in our laboratory has supported a relationship between drought stress and cold-acclimation responses (Guy et al., 1992; Neven et al., 1993).

[1] This research was supported in part by the U.S. Department of Agriculture/National Research Initiative Competitive Grants Program (grant no. 93-37100-8906). This is Florida Agricultural Experiment Station journal series no. R-06081.

[2] Present address: Centre de Coopération Internationale en Recherche Agronomique pour le Développement, 2477 ave du Val de Montferrand, 34032 Montpellier, cedex 1, France.

[3] Present address: U.S. Department of Agriculture-Agricultural Research Service, Yakima Agricultural Research Laboratory, 5230 Konnowac Pass Road, Wapato, WA 98951.

[4] Present address: Department of Oral Biology, University of Florida, Gainesville, FL 32610.

* Corresponding author; e-mail clg@gnv.ifas.ufl.edu; fax 1-352-392-3870.

Abbreviations: CAP, cold-acclimation protein; IPTG, isopropyl-1-thio-β-galactoside; LEA, late embryogenesis abundant; LT_{50}, killing temperature for 50% of the cells; MS, Murashige and Skoog.

BNSDOCID: <XP_____2421536A_I_>

The major question yet unanswered is this: By what mechanisms do plants become freezing tolerant? Undoubtedly, the process is one of great complexity (Steponkus, 1984; Sakai and Larcher, 1987). The finding that many cold-stress proteins are related to proteins linked to water stress or anhydrous conditions is pivotal because it points to a specific facet of freezing stress that can be more critically examined. In spinach two cold-stress proteins, CAP85 and CAP160, have been shown to share features in common with proteins associated with water stress (Guy et al., 1992; Neven et al., 1993). Both of these proteins exist in soluble form in spinach cells, and are actively synthesized and accumulated in leaf tissue exposed to 5°C (Guy and Haskell, 1987, 1988) or subjected to water stress at 25°C (Guy et al., 1992).

One approach to determining the role of a stress protein of known or unknown function is to modify its natural expression (Artus et al., 1996) or to insert its gene into another organism in which it does not naturally exist (Georges et al., 1990; Hightower et al., 1991; McKown and Warren, 1991; Kenward et al., 1993). Here we report the cloning and characterization of cDNAs and genomic sequences for one of these proteins, CAP160, and the influence of expression of both CAP160 and CAP85 proteins on freezing-stress injury in tobacco (*Nicotiana tabacum*).

MATERIALS AND METHODS

Isolation of CAP160 cDNAs and Genomic Clones

Using a monoclonal antibody (2H8) raised against purified CAP160, cDNAs were isolated by immunoscreening a Uni-Zap (Stratagene) library prepared with poly(A^+) RNA from spinach (*Spinacia oleracea* L.) leaf tissue that had been exposed to 5°C for 2 d (Neven et al., 1993). Sixteen clones were selected by the immunoscreening. Four different clones were partially sequenced and all appeared identical. Clone IIa contained the largest insert and was sequenced in its entirety using the *Taq* DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems) on an automated sequencer (Applied Biosystems). Subclones were prepared by cutting at internal restriction sites, followed by cloning into pBluescript (Stratagene). All regions of the cDNA were sequenced several times and, where necessary, synthetic oligonucleotide primers were used to complete sequencing. Spinach DNA was partially restricted with *Eco*RI and cloned into EMBL4. The library was screened with CAP160 cDNA clone IIa. Overlapping clones representing the 5′ and 3′ regions were used to construct the complete gene sequence. Database searches for homology to CAP160 DNA and protein sequences were performed with BLAST (Altschul et al., 1990).

Nucleic Acid Isolation and Analyses

RNA was isolated using a phenol-LiCl procedure (Ausubel et al., 1991), and portions were separated onto a 1.2% agarose gel containing formaldehyde. Gels were stained with ethidium bromide to visualize rRNA bands for verification of equal loading, photographed, and then blotted onto a nylon membrane (Hybond-N, Amersham) and fixed with UV light (Stratalinker 1800, Stratagene). Blots were hybridized with the random-primed ^{32}P-labeled cDNA clone IIa, then washed in 2× SSC, 0.2% SDS at room temperature, 0.5× SSC, 0.2% SDS at 68°C, and 0.1× SSC, 0.2% SDS at 68°C. Autoradiography was generally performed for 24 h at −80°C. DNA was isolated by the method of Gordon-Kamm et al. (1990), cut with restriction enzymes as indicated, and analyzed by DNA-blot hybridization.

CAP Expression in *Escherichia coli*

Fusion protein expression in *E. coli* XL1-Blue cells harboring pBluescript containing a cDNA insert was induced with IPTG. XL1-Blue cells containing pBluescript without an insert were included in all of the experiments as a control. Induced cells were pelleted by centrifugation at 1,600*g* for 15 min in a JA20 rotor (Beckman), resuspended in lysis buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 1 mM PMSF, 10% Suc, 1 mg/mL lysozyme), and incubated for 10 min on ice. Triton X-100 was added to 0.1% final concentration and incubated for 10 min on ice. The solution was centrifuged at 48,000*g* in a JA20 rotor for 1 h, and the supernatant was removed and mixed with 2× SDS buffer. Protein sequestered in inclusion bodies was extracted from the pellet with SDS buffer. Approximately 5 μg of total protein was fractionated by SDS-PAGE in 10% acrylamide gels, blotted to PVDF membranes, and probed with anti-CAP160 or anti-CAP85 antibodies (Guy et al., 1992).

Plant Transformation and Growth Conditions

CAP160 and CAP85 (Neven et al., 1993) cDNA sequences were put under the control of the cauliflower mosaic virus 35S promoter by subcloning a 2.7-kb *Bam*HI-*Xho*I fragment of CAP160 and a 1.8-kb *Bam*HI-*Xho*I fragment of CAP85 into the broad-host-range vector pVSTIII (Malik and Wa-

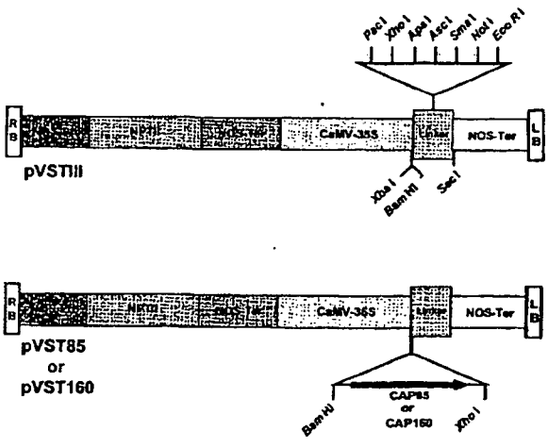


Figure 1. Constructs used in the tobacco transformation experiments (Malik and Wahab, 1993). The right (RB) and left (LB) borders are indicated by the open boxes at the ends of the constructs.

BNSDOCID: <XP_____2421536A_1_>

hab, 1993), creating plasmids pVST-160 and pVST-85, respectively (Fig. 1). These two plasmids were transferred separately into *Agrobacterium tumefaciens* strain LBA4404 by direct transformation (Hofgen and Willmitzer, 1988).

Leaf discs from surface-sterilized young leaves of greenhouse-grown tobacco (*Nicotiana tabacum* cv Xanthi) plants were infected with *A. tumefaciens* harboring pVST-160, pVST85, or pVSTIII. Successful transformants were selected on regeneration medium containing 4.3 g/L MS salts (Murashige and Skoog, 1962), 1 mL/L 1000× B5 vitamin mixture, 0.3 mg/L IAA, 10 mg/L dimethylallylaminopurine, 30 g/L Suc, 100 mg/L kanamycin, and 250 mg/L cefotaxime or carbenicillin. Young shoots were transferred to this medium with no hormones and no antibiotic for rooting. Rooted plants were transferred to soil and maintained either in a growth chamber at 25°C or in the greenhouse. Seeds were collected from all primary transformants and sown on solid medium containing MS salts and 200 mg/L kanamycin. Plantlets that remained green were transferred to soil and grown under greenhouse conditions.

Protein Analysis

Leaf material was frozen in liquid nitrogen and then powdered using a mortar and pestle. The frozen powder was transferred directly to SDS buffer (Neven et al., 1993) and boiled for 5 min. Insoluble proteins were removed by spinning the samples for 10 min in a microcentrifuge. The supernatant was transferred to fresh tubes and stored at −20°C. Total protein content was determined by the Bradford method (1976). The presence of CAP160 and CAP85 were determined by protein-blot analysis of crude protein extracts using monoclonal antibodies 2H8 and 5A10, respectively. Equal loading of protein-blot lanes was verified by Coomassie brilliant blue staining of a duplicate gel.

Partial sequencing of CAP160 purified from spinach tissue was accomplished after cleavage with cyanogen bromide as previously described (Neven et al., 1993).

Tissue prints were made from flower, stem, and pedicel tissue from cold-acclimated and nonacclimated transformed tobacco and control, mock-transformed tobacco (Cassab and Varner, 1987). The resulting membranes were probed with CAP85 monoclonal antibody 5A10 (Neven et al., 1993), CAP160 monoclonal antibody 2H8 (Haskell et al., 1993), or dehydrin antibody (Close et al., 1993).

Freeze Tests

Leaves were harvested from either greenhouse- or growth-chamber-grown plants at the temperatures indicated. Leaf discs (three per tube) were frozen at a cooling rate of 0.5 or 1°C/h, as indicated. All samples were equilibrated for 1 h at 0°C, and ice nucleation was induced at −1°C by placing a small ice chip in contact with the tissue. Frozen samples were thawed at 4°C in a refrigerator overnight. The leaf discs were placed in 5 mL of double-distilled water and incubated for 2 h. Electrolyte loss was measured with a conductivity meter. After the initial conductivity measurement, the relative electrolyte loss was determined by refreezing the tissue at −80°C, thawing, and measuring the conductivity a second time.

Drought Tolerance

Young plants at the three-leaf stage were transferred to vermiculite and watered every other day with 5% PEG in one-tenth MS salts. Control plants received one-tenth MS salts alone. After 2 weeks, the concentration of PEG was raised to 10%. Six weeks later the plants were evaluated visually and fresh weights were determined.

RESULTS

The CAP160 protein is known to be actively synthesized and/or accumulated in leaf tissue exposed to 5°C (Guy and

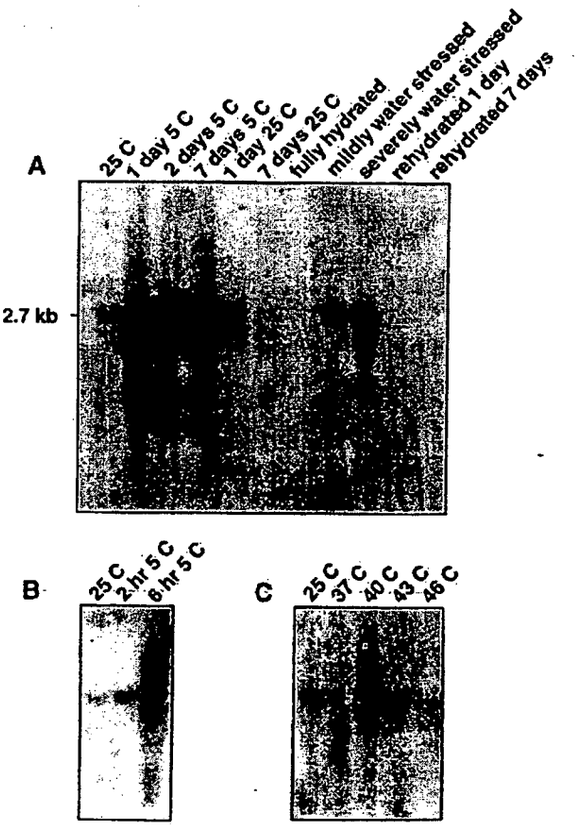


Figure 2. RNA-blot analyses of the influence of cold acclimation, deacclimation, water stress, and heat shock on the CAP160 steady-state mRNA abundance. A, Acclimation, deacclimation, and water-stress responses. B, Short-term response to low temperature. C, Response to heat shock. The cDNA hybridizes to a major 2.7-kb mRNA. Each lane was loaded with 5 μg of total RNA, and equal loading was verified by ethidium-bromide staining before blotting. Similar results were observed in several hybridization experiments. Details regarding the physiologic parameters for the water-stressed samples have been published elsewhere (Guy et al., 1992; Neven et al., 1993).

BNSDOCID: <XP_____2421536A_I_>

Haskell, 1987, 1988) or during water stress at 25°C (Guy et al., 1992). Therefore, any cDNAs for this protein selected by the antibody screening should hybridize with mRNA that was up-regulated during low-temperature exposure or water stress. Several of the clones hybridized strongly to a 2.7-kb mRNA that was up-regulated during exposure to 5°C and was down-regulated upon return to 25°C (Fig. 2A). Similarly, the cDNAs hybridized to a 2.7-kb mRNA that was abundant in water-stressed leaf tissue, but was low or absent in nonstressed tissue (Fig. 2A). The mRNA for CAP160 reached a maximum abundance in leaf tissue within 24 h of 5°C exposure and remained high for up to 7 d of low-temperature exposure. The transcript abundance was slightly increased at 2 h of 5°C and by 6 h had accumulated to a high level (Fig. 2B). Expression of CAP160 was moderately stimulated by a heat-shock treatment of 40°C for 2 h, but not by higher or lower heat-shock temperatures (Fig. 2C).

Sequence analysis of several cDNAs indicated an open reading frame encoding an acidic protein of 780 amino acids (Fig. 3). The position of the first ATG in the sequence has a similarity to a translation initiation site consensus (Joshi, 1987), and the open reading frame is broken by an in-frame stop codon 20 bp upstream from the first ATG, suggesting that the cDNA contains the entire coding sequence. The deduced protein has a molecular mass of 83.9 kD and a pI of 4.6. Whereas the calculated pI is in good agreement with the pI of 4.5 estimated from two-dimensional gels for spinach CAP160, the molecular mass of the deduced protein is little more than one-half of that estimated by SDS-PAGE (Guy and Haskell, 1987).

Protein-blot analyses of *E. coli* possessing pBluescript without an insert showed no immunoreactivity with anti-CAP160 antibodies, whereas cells containing the CAP160 cDNA expressed immunoreactive proteins with or without induction with IPTG (Fig. 4A). Expression was higher with IPTG, and the level of CAP160 accumulation in *E. coli* appeared to be similar to that found in cold-acclimated spinach leaves (Guy et al., 1992). Two major bands (120 and 140 kD) and a number of minor bands were present, each of which migrated faster than leaf CAP160 (Fig. 4A, upper band in lanes 7 and 16). The lower doublet bands in the leaf extracts of 115 and 110 kD are thought to be proteolytic breakdown products of CAP160 because they generally increase while the 160-kD band decreases with time after extraction. Induced bacterial cells contained a minor band at 160 kD (Fig. 4A, lanes 2, 4, 12, and 14) that comigrated with leaf CAP160 (Fig. 4A, lanes 7 and 16). Because an in-frame stop codon is present before the first ATG, it is probable that the immunoreactive recombinant protein of 160 kD in induced cells is not a β-galactosidase fusion protein, but an authentic-size CAP160 protein resulting from the reinitiation of translation at the first ATG site. The proteins of < 160 kD may be proteolytic products or may result from alternative reinitiation of translation at ATG sites farther downstream. Nevertheless, a small percentage

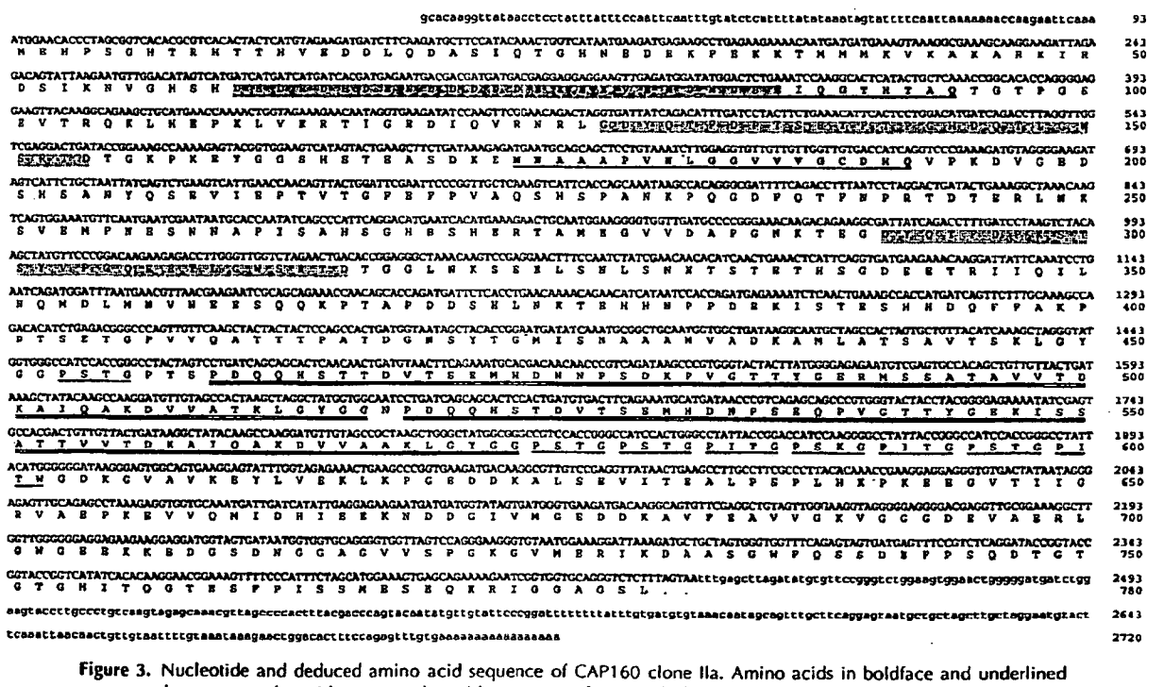


Figure 3. Nucleotide and deduced amino acid sequence of CAP160 clone IIa. Amino acids in boldface and underlined correspond to cyanogen bromide-generated peptide sequences from purified CAP160. The dark-shaded box denotes a motif rich in negatively charged residues, and the lighter-shaded box denotes a repeat sequence. An imperfectly tandem repeated sequence is double underlined, and a repeating four-residue motif of P(I/S)(T/K)(C/W) is underlined.

BNSDOCID: <XP_____2421536A_1_>

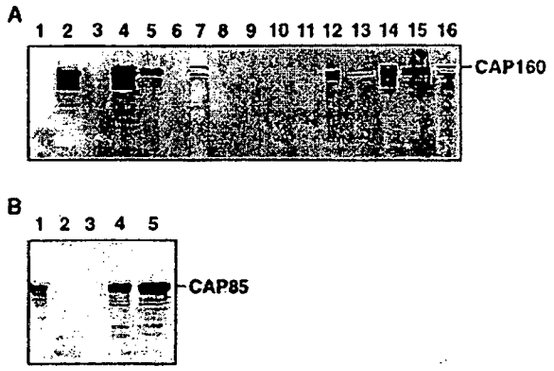


Figure 4. CAP160 and CAP85 protein expression in *E. coli* XL1-Blue cells. A, Lane 1, pBluescript without an insert; lane 2, CAP160 clone IIa; lane 3, clone IIb; lane 4, clone VIII; lane 5, clone VII; lane 6, clone IX; lane 7, spinach leaf CAP160; lane 8, soluble fraction of pBluescript without an insert minus IPTG; lane 9, insoluble fraction of pBluescript without an insert minus IPTG; lane 10, soluble fraction of pBluescript without an insert plus IPTG; lane 11, insoluble fraction of pBluescript without an insert plus IPTG; lane 12, soluble fraction of CAP160 IIa minus IPTG; lane 13, insoluble fraction of CAP160 IIa minus IPTG; lane 14, soluble fraction of CAP160 IIa plus IPTG; lane 15, insoluble fraction of CAP160 IIa plus IPTG; and lane 16, spinach leaf CAP160. B, Lane 1, Spinach leaf CAP85; lanes 2 and 3, CAP85 out-of-frame clone S42a plus IPTG; and lanes 4 and 5, CAP85 in-frame clone S42aE plus IPTG.

of the protein produced appeared to be full length. This was not unexpected because eukaryotic proteins are often found to be unstable when expressed in *E. coli* (Ausubel et al., 1991).

Whether the truncated products retained biological activity remains unknown. CAP160 protein products were present in both soluble and insoluble forms, and it appeared that a significant portion of the expressed protein was not sequestered in inclusion bodies. The 160-kD form was observed only in the soluble fraction (Fig. 4A, lanes 12 and 14). The reason for the low mobility of CAP160 in SDS-PAGE is unknown, but might result from the high content of Ser, Ala, and negatively charged residues (Kondo and Inouye, 1991). The CAP85 cDNA, in which the deduced protein sequence is smaller than the apparent size estimated by SDS-PAGE (Neven et al., 1993), also produced a protein that comigrated with CAP85 from leaf tissue (Fig. 4B).

Additional evidence linking the cDNA of clone IIa with CAP160 is provided by protein sequence information. Two amino acid sequences from peptide fragments prepared by cyanogen bromide cleavage of purified CAP160 closely matched the deduced sequences at residues 83 to 94 and 174 to 192 (Fig. 3).

CAP160 has several unusual motifs that may have a functional significance. It has an acidic region toward the N terminus in which 18 of 26 residues are either aspartic acid or Glu, including a central core where five aspartic acids are followed by four Glu residues (Fig. 3). Two imperfect repeats of 29 residues each are found in the N-terminal half of the protein. They are separated from each other by 133 residues. These two repeats are rich in Thr (24%), Gly (14%), and Asp (14%). Another two large, imperfect repeats of 58 and 56 amino acids, separated by 1 residue, are found beginning at residue 460. Thr (approximately 13%) is the single most abundant residue in these repeats, followed by Val (10.5%). Immediately following the second large repeat is a series of seven 4-residue repeats with the consensus of P(S/I)TG. Preceding the first large repeat is one copy of the 4-residue repeat PSTG. Given the presence of the Pro and Gly in this repeating sequence, this region might be a loop or a solvent-exposed region of the protein. The Ser and Thr residues both bearing hydroxyls might be potential phosphorylation sites on the protein. All of these characteristic features of CAP160 were individually used to do BLAST searches of the databases for matches that could provide insight into its function. In addition, the CAP160 sequence was searched for known protein motifs found in the Prosite library (release 12.2, 3/95) of the Genetics Computer Group (Madison, WI) sequence analysis software package, but no convincing matches were found that would give a clue to its function or identity.

A search of the GenBank database revealed no compelling sequence matches for the entire protein. Although it was suspected that CAP160 was likely the homolog of a pair of *Arabidopsis thaliana* proteins with several names (*cor78, lti140, rd29A, rd29B,* and *lti65*), the total protein sequence similarity was less than 20%. Nevertheless, four discrete regions within the CAP160 sequence did have high homology with the Arabidopsis proteins (Fig. 5), leading us to believe that CAP160 is a probable homolog to these proteins. The two Arabidopsis proteins also have an acidic region (EEDDDDDELEPE or DVEDDDDEYDEQDPE) with similarities to the acidic region of CAP160, which could have some importance to a common function.

Replicated blots of genomic DNA using cDNA probes showed some blot-to-blot variation, but overall produced the predicted banding pattern based on the DNA sequence from genomic clones for CAP160 (Fig. 6). The banding

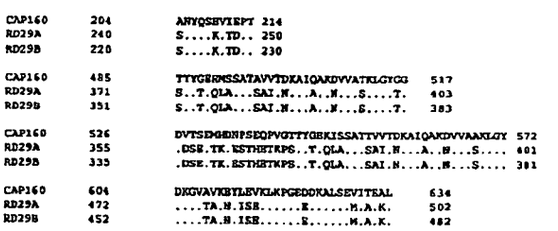


Figure 5. Alignment of CAP160 protein sequence with the Arabidopsis rd29A and rd29B proteins. Identical residues are indicated by dots.

BNSDOCID: <XP_____2421536A_I_>

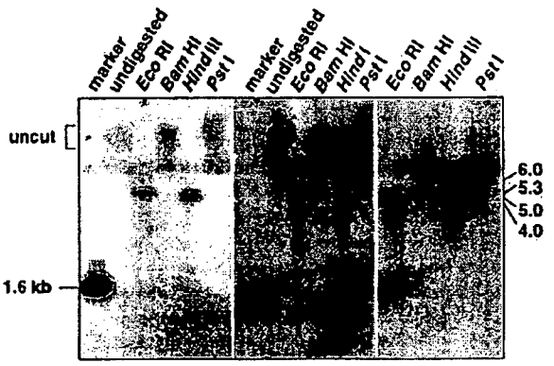


Figure 6. Genomic DNA blot for CAP160. Results of three different experiments are shown. Clone IIa was used to make the hybridization probe.

pattern was relatively simple even when stringency was reduced, indicating that CAP160 is not a member of a large, high-homology gene family.

The isolation of genomic clones for CAP160 permitted the characterization of the organization of the gene. The CAP160 protein is encoded on seven exons extending over 8 kb. The exons range from 71 to 1254 bp in length, whereas the introns tend to be larger than the exons, ranging from a minimum of 440 bp to a maximum of 1437 bp (Fig. 7). Characterization of genomic clones with differences in restriction fragment length revealed the existence of a 352-bp deletion in intron 1 (Fig. 7). Aside from this deletion, the nucleotide sequences in both intron and exon regions of all isolated clones were virtually identical, suggesting the possibility that if it does not represent a second gene sequence, this clone represents an allelic sequence.

Subcellular-localization studies were conducted with leaf and hypocotyl tissue to determine where in the cell CAP160 is found. These studies were performed at the same time as studies of another spinach CAP, the results of which have been previously reported (Neven et al., 1993). This previous report includes the marker protein analysis for relative fraction contamination. The major organellar fractions were prepared, and these extracts were probed with 2H8, a monoclonal antibody specific for CAP160, using protein-blot methods (Fig. 8). A major portion of CAP160 was present in a freely soluble fraction, which would be consistent with a protein present in the cytosol and not tightly associated with any macromolecular structure that would be pelleted at relatively low g forces. However, an equally large proportion was found in the mitochondrial fraction. Whether the protein was inside or outside of the mitochondria was not determined. Protein blots using antibodies for marker proteins for all of the fractions shown in Figure 8 indicated that the presence of the strong CAP160 reactivity in the mitochondrial fraction was well in excess of the documented cross-contamination (Neven et al., 1993), and thus was likely to reflect an actual cofractionation.

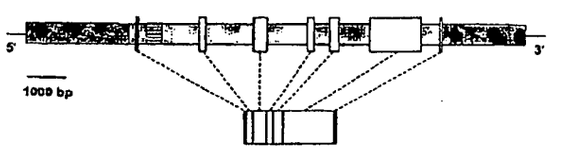


Figure 7. Structure of the CAP160 gene. Flanking sequences are indicated by dark shading, noncoding and intron sequences are lightly shaded, and exons are shown as unshaded boxes. The hatched area in intron 1 shows the position of a large deletion present in some sequences.

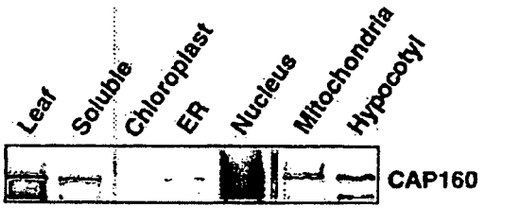


Figure 8. Subcellular localization of CAP160. The presence of CAP160 was detected by protein blot using a monoclonal antibody, 2H8, which is specific for CAP160. For marker protein analysis and relative contamination for each subcellular fraction, see Neven et al. (1993).

Expression of CAP160 and CAP85 in Transgenic Tobacco

Constructs using the cDNAs for CAP160 and CAP85 (Neven et al., 1993) were placed under the control of the 35S promoter and used in *A. tumefaciens*-based transformation experiments. Putative transformants were assayed for expression of CAP160 and CAP85 by protein-blot analyses using highly specific monoclonal antibodies. In these experiments protein was isolated from plants grown under either growth-chamber conditions (25°C) or greenhouse conditions. Some of the leaves (as indicated) were taken from plants that were grown under sterile conditions in sealed containers with a very high humidity level. Several transformants were found to express the mRNA for each introduced gene (Fig. 9). No cross-reacting protein was

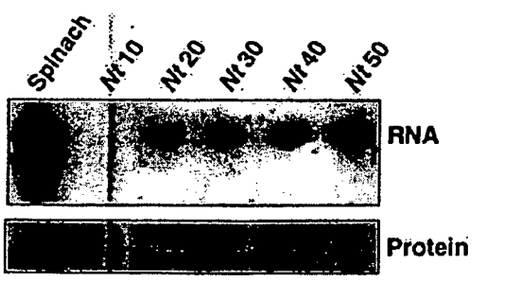


Figure 9. Analysis of spinach CAP160 RNA and protein expression in transgenic lines of tobacco. Spinach RNA and protein are shown for reference. *Nt*10 was transformed with the vector only. Hybridizations were done with radiolabeled cDNA clone IIa, and protein detection was with monoclonal antibody 2H8. RNA blots were loaded with 5 μg of total RNA per lane. Protein blots were loaded with 20 μg of total buffer-soluble protein per lane.

BNSDOCID: <XP_____2421536A__I_>

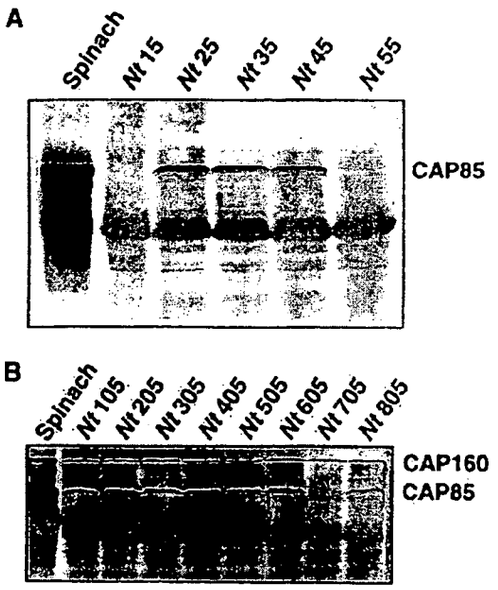


Figure 10. Analysis of spinach CAP85 expression in tobacco (A) and coexpression of CAP85 and CAP160 in progeny from a cross of lines expressing individually either CAP85 or CAP160 (B). Cross-reacting proteins were detected with a mixture of monoclonal antibody 2H8 for CAP160 and monoclonal antibody 5A10 for CAP85.

found in either wild-type tobacco (not shown) or in tobacco transformed with the vector alone. CAP85 was resolved as a doublet consisting of an intense 85-kD band and a smaller and less-intense 80- to 83-kD band in spinach (Neven et al., 1993), and its expression also resulted in a doublet in transgenic tobacco. R_0 CAP85 plants (*Nt*25 and *Nt*35) were crossed with R_0 CAP160 plants (*Nt*40), and the progeny were evaluated for expression of the two proteins. A large number of the plants expressed both proteins, a few expressed only CAP160, but no plants were found that expressed only CAP85 or neither of the two proteins (Fig. 10).

Tissue prints were done on nitrocellulose with cold-acclimated or nonacclimated stem, pedicel, and flower tissue taken from transformed tobacco plants to assess expression patterns at the tissue level. The resulting membranes were exposed to CAP85, CAP160, or both antibodies. Both proteins were expressed in all tissues examined, and no difference was seen between cold-acclimated and nonacclimated tissue (not shown). When the same tissues were analyzed by tissue printing and probed with a dehydrin antibody (Close et al., 1993), greater staining was seen in the cold-acclimated tissue than in the nonacclimated tissue. Tissue prints from control, mock-transformed tobacco did not react with either of the two monoclonal antibodies for the spinach proteins, but did react with the anti-dehydrin antibody.

Seeds were collected from all of the R_0 plants, regardless of whether they expressed the introduced CAP genes. Seeds were sown on kanamycin and the number of green plants (kanamycin resistant) versus white plants was evaluated. The number of white plants versus green plants was somewhat lower than the expected Mendelian ratio of 1:3 (1:6.6 for CAP160, 1:5 for CAP85), indicating the possibility that more than one, but probably only a few, copies of the transferred genes integrated into the genome. Selected green plants were evaluated by protein-blot analysis, and all expressed either the expected CAP160 or CAP85. Green, 1-month-old plantlets were transferred to soil, and at the three-leaf stage they were transferred into either cold (3°C) or warm (20°C) conditions. After 14 d, protein was isolated and analyzed by protein blot. In all cases the quantities of CAP85 and CAP160 protein were lower in the plants grown in the cold (data not shown).

Freezing Tests

The first test for increased hardiness was an electrolyte-leakage test performed on leaf discs of transformed and mock-transformed plants. Initial tests were conducted on R_0 plants grown at 25°C. Only one transformant, CAP160 *Nt*40, exhibited any evidence of a decrease in electrolyte leakage, but even that was not statistically significant (Table I). A sample from each plant was taken each time the electrolyte-leakage test was performed to assay for CAP160 or CAP85 expression.

The second set of experiments was performed on the R_1 generation. In this case, two sets of experiments were run. The first was a determination of LT_{50} as described above (Fig. 11). A slight improvement was seen in the hardiness of plants expressing the spinach protein only at −1°C, but overall the LT_{50} values were not different. The second test, a "kinetics-of-freezing" analysis, was designed to reveal more subtle differences between the plants. In this test, samples were placed at the approximated LT_{50} of −2°C and removed at various times. At all times, the plants expressing the spinach proteins exhibited less electrolyte leakage than the untransformed control plants. Because of the large variability in the mock-transformed plants, no significant difference was found between it and either the untransformed plants or those expressing the spinach proteins (Fig. 12).

Whole plants expressing CAP160 or CAP85, mock-transformed (pVSTIII, vector only) plants, or control cv Xanthi plants (nontransformed) were transferred at ap-

Table I. *Freezing tolerance of an R_0 transgenic tobacco line expressing the spinach CAP160 protein*

The data are presented as the arcsin-transformed percentage of the relative electrolyte leakage. Two different mock-transformed lines were used and the data were combined. The experiment was repeated five times.

Temperature	Electrolyte Leakage	
	Transformed controls	CAP160 (*Nt*40)
°C		
0.0	26 ± 1	30 ± 7
−1.0	30 ± 2	31 ± 3
−1.5	31 ± 3	34 ± 4
−2.0	50 ± 6	43 ± 11
−2.5	60 ± 6	52 ± 12

BNSDOCID: <XP_ _____2421536A_I_>

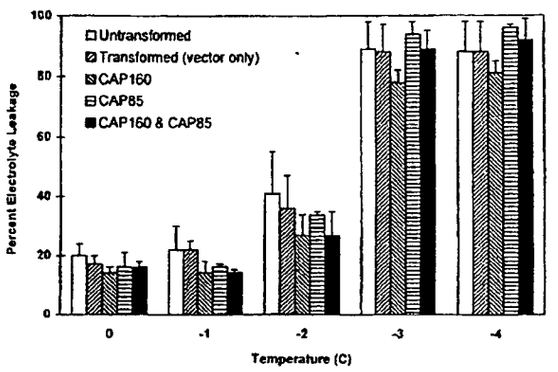


Figure 11. Influence of CAP85 or CAP160 expression or coexpression of both proteins on the severity of injury after freezing. Error bars represent the SD of two to four experiments.

proximately the three-leaf stage into a cold room. These plants were incubated for 1 month at 5°C. Visual inspection revealed no obvious difference in either the amount of leaf damage or plant size. The temperature was reduced to approximately 0°C, and although all of the plants suffered minor leaf damage, all survived (data not shown). Protein analyses indicated no differences in the levels of either spinach protein over that of similar plants kept at 20°C.

DISCUSSION

The function of CAP160 in spinach and its role in plant cold stress remains obscure. Homology searches indicate that CAP160, in its entirety, is a novel protein, the biological function of which cannot be identified by shared homology with proteins of known function. CAP160 does have limited similarity to the Arabidopsis *lti78*, or COR78 protein, a low-temperature-induced (Horvath et al., 1993; Nordin et al., 1993) and desiccation-responsive protein, also known as rd29A and rd29B (Yamaguchi-Shinozaki, 1993a, 1993b). The shared homology is enough for us to consider it a spinach homolog to these proteins. The primary structure of CAP160 shows no similarity to antifreeze or ice-nucleation proteins, the function of which is to alter the freezing of water (Green and Warren, 1985; Davies and Hew, 1990). However, several noteworthy features of the primary structure of CAP160 (Fig. 3) could provide helpful clues to its functional role in low-temperature-stress response mechanisms. These include a highly negatively charged motif at residues 61 to 86, which includes a run of nine consecutive aspartates and glutamates and several repeat motifs, a large tandem imperfect repeat, and a four-residue imperfect-repeating motif present in eight copies. All of these features were used to search the databases for matching sequences that might provide a clue to CAP160's function. Only the acidic residue tract 61 to 86 yielded a noteworthy similarity.

A variety of proteins possess extensive tracts of negatively charged residues, including nucleolins (Lapeyre et al., 1987), some transcription factors (Hisatake et al., 1991), Na/Ca,K exchanger (Reilander et al., 1992), spermine-binding protein (Chang et al., 1987), calsequestrin (Yazaki et al., 1990), and ubiquinol-Cyt *c* reductase protein (Van Loon et al., 1984). Of these, the relationship to that of nucleolin seems most notable. Nucleolin is a major nucleolar phosphoprotein in eukaryotic cells (Fang and Yeh, 1993) that has sequence-specific RNA-binding activity that regulates, or participates in, rRNA processing (Hanakahi et al., 1997). CAP160 lacks many of the structural features of plant nucleolin, such as nuclear-localization sequences (Bogre et al., 1996), and we do not consider it to be a nucleolin. However, it has some unusual similarities, including the fact that it is a phosphoprotein (Guy and Haskell, 1989) and is extremely sensitive to proteolytic cleavage even in purified form (D. Haskell and C. Guy, unpublished data). Nucleolin has a self-cleaving activity that resides in the C-terminal domain and that apparently recognizes repeated motifs within the molecule (Fang and Yeh, 1993). CAP160 has repeats, although not as many as are found in nucleolins.

Because one of the major effects of low-temperature exposure is a significant reduction in translation rate, perhaps CAP160's function is to participate as an accessory protein in the assembly of ribosomes or to help accelerate translation in some way. Evidence exists to suggest that the capacity of the translation apparatus is enhanced during cold acclimation (Guy et al., 1985) during a time when CAP160 is accumulating. CAP160's subcellular localization appears to be predominately cytosolic, which would be consistent with such a function. However, given its apparent cofractionation with mitochondria, alternative functions are also a possibility.

Analysis of CAP160 gene structure revealed further differences from that of the Arabidopsis cold- and drought-responsive *rd29* genes. CAP160 has more and much larger introns than either of the two Arabidopsis genes, six versus three each. In addition, the two Arabidopsis genes are within 2 kb of each other, but we could find no evidence from genomic clones that such a close tandem arrangement exists for CAP160 in spinach.

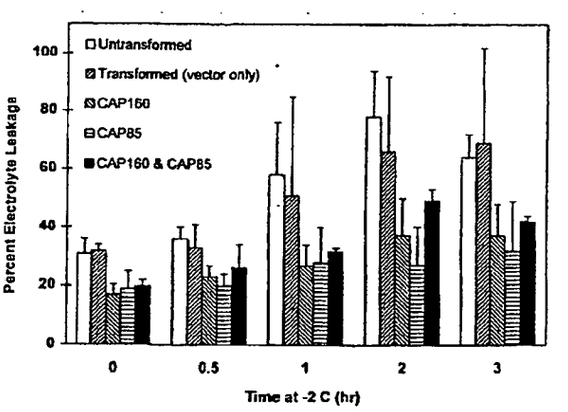


Figure 12. Differential sensitivity of tobacco expressing spinach cold-stress proteins in response to a time-course freeze/thaw regime. Error bars represent the SD of five experiments.

BNSDOCID: <XP..... .2421536A_1_>

The coding sequences from cDNA clones for two genes that are up-regulated during cold acclimation in spinach were placed under the control of the nearly constitutive cauliflower mosaic virus promoter and transformed separately into tobacco, a freezing-sensitive plant. These two genes were expressed in the transformed tobacco as determined by protein-blot analysis, whereas no cross-reacting proteins to the monoclonal antibodies were found in nontransformed tobacco. However, both of the proteins were expressed in tobacco at an apparently lower level than under low-temperature-induction conditions in spinach. Tests designed to determine whether their expression conferred increased tolerance to low-temperature stress indicated little overall benefit to tobacco. Freezing tolerance is considered to be a quantitative trait (Hayes et al., 1993); therefore, it is not surprising that the heterologous expression of one or two genes induced during cold acclimation would significantly alter the LT_{50} of freezing-sensitive tobacco. This result is similar to that when a cold-regulated, osmotin-like protein was overexpressed in potato and had no effect on freezing tolerance (Zhu et al., 1996), but it is in contrast to the constitutive overexpression of the COR15a protein in Arabidopsis (Artus et al., 1996).

In the work with Arabidopsis, overall cell survival was only subtly enhanced after a freeze/thaw cycle for overexpressing cells, but a more dramatic, positive benefit of overexpression of COR15a was the preservation of chloroplast electron-transport function by this chloroplast stromal protein. Other examples of this approach include several attempts to enhance yeast and plant freezing tolerance through the introduction of cold-water fish antifreeze proteins (Georges et al., 1990; Hightower et al., 1991; McKown and Warren, 1991; Kenward et al., 1993). Expression of fish antifreeze proteins has not significantly altered freezing of plant tissues (Hightower et al., 1991). The results of the expression of spinach cold-stress proteins in tobacco do not rule out the possibility that major enhancements in the tolerance of freezing-sensitive plants might be achieved through the expression of one or a few stress proteins, but instead further support the view of the quantitative nature of plant freezing-stress tolerance (Hayes et al., 1993).

CAP85 and CAP160 are continuously and highly expressed during cold acclimation and under continuing cold treatment, making them unlikely candidates for regulatory genes, which in all likelihood would be rare messages expressed very early during cold acclimation. A large number of genes are induced in the process of cold acclimation (Kaye and Guy, 1995). Although many of these genes fall into one of a few multigene families, particularly the LEA/dehydrin family, their function remains unknown. Most are probably involved in the metabolic adjustment of the plant during cold acclimation. The two clones used in this study, CAP85 and CAP160, are both members of the LEA/dehydrin superfamily. LEA and dehydrin proteins are known to accumulate during the later stages of seed development or during water stress and are thought to play a role during seed dry-down (Close et al., 1993). Recent evidence suggests that at least one of the spinach proteins may have cryoprotective properties (Kazuoka and Oeda, 1992). Purified CAP85 was shown to protect lactate dehydrogenase in vitro from a freeze/thaw inactivation (Kazuoka and Oeda, 1992). This opens the door to the possibility that many low-temperature-regulated genes may have cryoprotective properties (Artus et al., 1996). More likely, CAP160's role, given its subcellular localization and its induction during drought stress, will be stabilization of supramolecular structures such as membranes, ribosomes, or cytoskeletal elements; acting as that of a proteinaceous compatible solute that reduces the toxic effects of cellular solutes at a high concentration during dehydration; or that of an enzyme that synthesizes a product with compatible solute properties. Regardless of its mode of action, CAP160 represents a new class of intracellular, low-temperature-stress proteins with a function correlated with enhanced freezing tolerance only in spinach.

ACKNOWLEDGMENTS

We are grateful to J. Anderson, C. Zhang, D. Sung, and an anonymous reviewer for their constructive comments during the preparation and review of the manuscript, and T. Close for the gift of the anti-dehydrin antibody. Facilities and assistance for protein and DNA sequencing, antibody production, and computer services were provided by the core laboratories of the Interdisciplinary Center for Biotechnology Research. We also thank the Florida High Tech Council for support.

Received August 21, 1997; accepted December 9, 1997.
Copyright Clearance Center: 0032–0889/98/116/1367/11.
The accession numbers for the sequences described in this article are AF035535 (cDNA) and AF035536 (genomic clone).

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Artus NN, Uemura M, Steponkus P, Gilmour SJ, Lin C, Thomashow MF (1996) Constitutive expression of the cold-regulated *Arabidopsis thaliana COR15a* gene affects both chloroplast and protoplast freezing tolerance. Proc Natl Acad Sci USA 93: 13404–13409

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1991) Current Protocols in Molecular Biology, Vol 2. Wiley, New York

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–254

Bogre L, Jonak C, Mink M, Meskiene I, Traas J, Ha DT, Swoboda I, Plank C, Wagner E, Heberle-Bors E, and others (1996) Developmental and cell cycle regulation of alfalfa nucMs1, a plant homolog of the yeast Nsr1 and mammalian nucleolin. Plant Cell 8: 417–428

Cassab GI, Varner JE (1987) Immunocytolocalization of extensin in developing soybean seed coats by immunogold-silver staining and by tissue printing on nitrocellulose paper. J Cell Biol 105: 2581–2588

Cattivelli L, Bartels D (1990) Molecular cloning and characterization of cold-regulated genes in barley. Plant Physiol 93: 1504–1510

Chang CS, Saltzman AG, Hiipakka RA, Huang IY, Liao SS (1987) Prostatic spermine-binding protein: cloning and nucleotide sequence of cDNA, amino acid sequence, and androgenic control of mRNA level. J Biol Chem 262: 2826–2831

Close TJ (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol Plant 97: 795–803

BNSDOCID: <XP._____2421536A__I.>

Close TJ, Fenton RD, Moonan F (1993) A view of plant dehydrins using antibodies specific to the carboxy terminal peptide. Plant Mol Biol 23: 279–286

Davies PL, Hew CL (1990) Biochemistry of fish antifreeze proteins. FASEB J 4: 2460–2468

Dure L III (1993) A repeating 11-mer amino acid motif and plant desiccation. Plant J 3: 363–369

Fang SH, Yeh NH (1993) The self-cleaving activity of nucleolin determines its molecular dynamics in relation to cell proliferation. Exp Cell Res 208: 48–53

Georges F, Saleem M, Cutler AJ (1990) Design and cloning of a synthetic gene for the flounder antifreeze protein and its expression in plant cells. Gene 91: 159–165

Gilmour SJ, Artus NN, Thomashow MF (1992) cDNA sequence analysis and expression of two cold-regulated genes of *Arabidopsis thaliana*. Plant Mol Biol 18: 13–21

Goldstein J, Pollitt NS, Inouye M (1990) Major cold shock protein of *Escherichia coli*. Proc Natl Acad Sci USA 87: 283–287

Gordon-Kamm WJ, Spencer TM, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams WR Jr, Willets NG (1990) Transformation of maize cells and regeneration of fertile transgenic plants. Plant Cell 2: 603–618

Green RL, Warren GJ (1985) Physical and functional repetition in a bacterial ice nucleation gene. Nature 317: 645–648

Guo W, Ward RW, Thomashow MF (1992) Characterization of a cold-regulated wheat gene related to *Arabidopsis thaliana cor47*. Plant Physiol 100: 915–922

Guy C, Haskell D, Neven L, Klein P, Smelser C (1992) Hydration-state-responsive proteins link cold and drought stress in spinach. Planta 188: 265–270

Guy CL, Haskell D (1987) Induction of freezing tolerance in spinach is associated with the synthesis of cold acclimation induced proteins. Plant Physiol 84: 872–878

Guy CL, Haskell D (1988) Detection of polypeptides associated with the cold acclimation process in spinach. Electrophoresis 9: 787–796

Guy CL, Haskell D (1989) Preliminary characterization of high molecular mass proteins associated with cold acclimation in spinach. Plant Physiol Biochem 27: 777–784

Guy CL, Niemi KJ, Brambl R (1985) Altered gene expression during cold acclimation of spinach. Proc Natl Acad Sci USA 82: 3673–3677

Hahn M, Walbot V (1989) Effects of cold-treatment on protein synthesis and mRNA levels in rice leaves. Plant Physiol 91: 930–938

Hanakahi LA, Dempsey LA, Li MJ, Maizels N (1997) Nucleolin is one component of the B cell-specific transcription factor and switch region binding protein, LR1. Proc Natl Acad Sci USA 94: 3605–3610

Haskell DW, Anderson JV, Guy CL (1993) Antigen binding of a mouse monoclonal IgG_1 is inactivated by heating but not by freeze/thaw cycling. Cryobiology 30: 532–535

Hayes PM, Blake T, Chen THH, Tragoonrung S, Chen F, Pan A, Liu B (1993) Quantitative trait loci on barley (*Hordeum vulgare* L.) chromosome 7 associated with components of winter hardiness. Genome 36: 66–71

Hightower R, Baden C, Penzes E, Lund P, Dunsmuir P (1991) Expression of antifreeze proteins in transgenic plants. Plant Mol Biol 17: 1013–1021

Hisatake K, Nishimura T, Maeda Y, Hanada K, Song CZ, Muramatsu M (1991) Cloning and structural analysis of cDNA and the gene for mouse transcription factor UBF. Nucleic Acids Res 19: 4631–4637

Hofgen R, Willmitzer L (1988) Storage of competent cells for *Agrobacterium* transformation. Nucleic Acids Res 16: 9877

Horvath DP, McLarney BK, Thomashow MF (1993) Regulation of *Arabidopsis thaliana* L. (Heynh) cor78 in response to low temperature. Plant Physiol 103: 1047–1053

Houde M, Danyluk J, Laliberté JF, Rassart E, Dhindsa RS, Sarhan F (1992) Cloning, characterization, and expression of a cDNA encoding a 50-kilodalton protein specifically induced by cold acclimation in wheat. Plant Physiol 99: 1381–1387

Hughes MA, Dunn MA (1996) The molecular biology of plant acclimation to low temperature. J Exp Bot 296: 291–305

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15: 6643–6653

Kaye C, Guy CL (1995) Perspectives of plant cold tolerance: physiology and molecular responses. Sci Prog 78: 271–299

Kazuoka T, Oeda K (1992) Heat-stable COR (cold-regulated) proteins associated with freezing tolerance in spinach. Plant Cell Physiol 33: 1107–1114

Kenward KD, Altschuler M, Hildebrand D, Davies PL (1993) Accumulation of type I fish antifreeze protein in transgenic tobacco is cold-specific. Plant Mol Biol 23: 377–385

Kondo K, Inouye M (1991) TIP 1, a cold shock-inducible gene of *Saccharomyces cerevisiae*. J Biol Chem 266: 17537–17544

Kurkela S, Franck M (1990) Cloning and characterization of a cold- and ABA-inducible Arabidopsis gene. Plant Mol Biol 15: 137–144

Lapeyre B, Bourbon H, Amalric F (1987) Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. Proc Natl Acad Sci USA 84: 1472–1476

Lin CL, Thomashow MF (1992) DNA sequence analysis of a complementary DNA for cold-regulated *Arabidopsis* gene *cor15* and characterization of the COR15 polypeptide. Plant Physiol 99: 519–525

Lin Y, Gross JK (1981) Molecular cloning and characterization of winter flounder antifreeze cDNA. Proc Natl Acad Sci USA 78: 2825–2829

Malik VS, Wahab SZ (1993) Versatile vectors for expressing genes in plants. J Plant Biochem Biotechnol 2: 69–70

Mazur PJ (1963) Kinetics of water loss from cells at subzero temperatures and the likelihood of intracellular freezing. Gen Physiol 47: 347–369

McKown RL, Warren GJ (1991) Enhanced survival of yeast expressing an antifreeze gene analogue after freezing. Cryobiology 28: 474–482

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15: 473–497

Neven LG, Haskell DW, Hofig A, Li QB, Guy CL (1993) Characterization of a spinach gene responsive to low temperature and water stress. Plant Mol Biol 21: 291–305

Nordin K, Vahala T, Palva ET (1993) Differential expression of two related, low-temperature-induced genes in *Arabidopsis thaliana* (L.) Heynh. Plant Mol Biol 21: 641–653

Orr W, Lu B, White TC, Robert LS, Singh J (1992) Complementary DNA sequence of a low temperature-induced *Brassica napus* gene with homology to the *Arabidopsis thaliana kin1* gene. Plant Physiol 98: 1532–1534

Pearce RS (1988) Extracellular ice and cell shape in frost-stressed cereal leaves: a low temperature scanning-electron microscopy study. Planta 175: 313–324

Reilander H, Achilles A, Friedel U, Maul G, Lottspeich F, Cook NJ (1992) Primary structure and functional expression of the Na/Ca,K-exchanger from bovine rod photoreceptors. EMBO J 11: 1689–1695

Sakai A, Larcher W (1987) Frost Survival of Plants: Responses and Adaption to Freezing Stress. Springer-Verlag, Berlin, pp 138–153

Steponkus PL (1984) Role of the plasma membrane in freezing injury and cold acclimation. Annu Rev Plant Physiol 35: 543–584

Steponkus PL, Uemura M, Balsamo RA, Arvinte T, Lynch DV (1988) Transformation of the cryobehavior of rye protoplasts by modification of the plasma membrane lipid composition. Proc Natl Acad Sci USA 85: 9026–9030

Sutton F, Ding X, Kenefick DG (1992) Group 3 LEA gene HVA1 regulation by cold-acclimation and deacclimation in two barley cultivars with varying freezing resistance. Plant Physiol 99: 338–340

Thomashow MF (1994) *Arabidopsis thaliana* as a model for studying mechanisms of plant cold tolerance. *In* EM Meyerowitz, CR Somerville, eds, Arabidopsis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 807–834

3NSDOCID: <XP_____2421536A_I_>

Van Loon AP, De Groot RJ, De Haan M, Dekker A, Grivell LA (1984) The DNA sequence of the nuclear gene coding for the 17-kd subunit VI of the yeast ubiquinol-cytochrome *c* reductase: a protein with an extremely high content of acidic amino acids. EMBO J **3:** 1039–1043

Yamaguchi-Shinozaki K, Shinozaki K (1993a) Arabidopsis DNA encoding two desiccation-responsive *rd29* genes. Plant Physiol **101:** 1119–1120

Yamaguchi-Shinozaki K, Shinozaki K (1993b) Characterization of the expression of a desiccation-responsive *rd29* gene of *Arabidopsis thaliana* and analysis of its promoter in transgenic plants. Mol Gen Genet **236:** 331–340

Yazaki PJ, Salvatori S, Dahms AS (1990) Amino acid sequence of chicken calsequestrin deduced from cDNA: comparison of calsequestrin and aspartactin. Biochem Biophys Res Commun **170:** 1089–1095

Zhu B, Chen TH, Li PH (1996) Analysis of late-blight disease resistance and freezing tolerance in transgenic potato plants expressing sense and antisense genes for an osmotin-like protein. Planta **198:** 70–77

BNSDOCID: <XP_____2421536A_I_>

(6)

© 2000 Elsevier Science B.V. All rights reserved.
Biotechnology Annual Review.
Volume 6.
M.R. El-Gewely, editor.

XP-001062997

Biotechnological applications of plant freezing associated proteins

P.D. 2000
p. 59-101 = (43)

Ghislain Breton[1], Jean Danyluk[1], Fran ois Ouellet [2] and Fathey Sarhan[1]
[1]D partement des Sciences biologiques, Universit du Qu bec à Montr al, Montr al, Qu bec, Canada; [2]Plant Gene Expression Center, USDA/ARS, Albany, California, USA

Abstract. Plants use a wide array of proteins to protect themselves against low temperature and freezing conditions. The identification of these freezing tolerance associated proteins and the elucidation of their cryoprotective functions will have important applications in several fields. Genes encoding structural proteins, osmolyte producing enzymes, oxidative stress scavenging enzymes, lipid desaturases and gene regulators have been used to produce transgenic plants. These studies have revealed the potential capacity of different genes to protect against temperature related stresses. In some cases, transgenic plants with significant cold tolerance have been produced. Furthermore, the biochemical characterization of the cold induced antifreeze proteins and dehydrins reveals many applications in the food and the medical industries. These proteins are being considered as food additives to improve the quality and shelf-life of frozen foods, as cryoprotective agents for organ and cell cryopreservation, and as chemical adjuvant in cancer cryosurgery.

Keywords: antifreeze proteins, cold acclimation, COR proteins, cryoprotection, cryosurgery, dehydrins, freezing tolerance, frozen food, genetic engineering, organ preservation, osmoprotectants, osmotic stress, oxidative stress, plants, regulators, transgenic plants.

Introduction

Life is virtually impossible without the simple water molecule. Its high polarity, lesser density of its solid state compared to its liquid state and the relative stability of its hydrogen bonds make it a unique molecule. The water molecules in their liquid state form a large hydrogen-bonded network that ensures the structural stability of biological molecules such as proteins, nucleic acids and the fluid bilayer structure of membranes. This property plays a key role in preserving the functional integrity of living cells. However, as liquid water changes to its solid state at the freezing point of a cell, ice crystals grow through the cells, destroying them and releasing their contents. Like all isothermic organisms, humans can generate enough metabolic heat to enjoy the luxury of the high temperature associated with the liquid state of water. This phenomenon seems so natural to us that we often associate heat with life and cold with death. However, many organisms can adapt and survive temperature extremes. Some of these organisms have the ability to tolerate freezing temperatures as low as -50 C [1].

Address for correspondence: Dr. Fathey Sarhan, Universit du Qu bec Montr al, D partement des Sciences biologiques, C.P. 8888, Succ. Centre Ville, Montr al, Qu bec, Canada, H3C 3P8. Tel : 514-987-3000 3363#. Fax : 514-987-4647. E-mail: sarhan.fathey@uqam.ca

BNSDOCID: <XP_____1062997A__I_>

Non-isothermic animals like fishes, insects and frogs have developed many ways to face winter. Some hide and avoid freezing in deep water or soil while those that live at the surface adopt different mechanisms. Some try to keep their internal fluid liquid at all cost and others prepare themselves for winter by regulating gradually the ice formation until complete freezing of their internal fluids [2]. Surface animals synthesize antifreeze proteins (AFPs) and/or accumulate osmoprotectants such as the sugar alcohol glycerol. Unicellular organisms which are almost immobile like bacteria and algae have also developed efficient strategies to live in cold winter conditions. Cold tolerant prokaryotes are usually classified as mesophilic and psychrophilic. The adaptation of mesophilic bacteria to cold occurs mainly through physiological means whereas psychrobacteria have adapted to low temperature (LT) environments through evolution [3–5].

Plants are multicellular organisms that, unlike animals, cannot escape when confronted by stressful conditions. They either adapt to the new environment or simply die. To ensure survival, several plants have evolved efficient strategies that help them to tolerate extreme winter conditions. These mechanisms are genetically controlled and induced upon exposure to LT [6,7]. This response, termed cold acclimation, is associated with the development of freezing tolerance (FT). Plants avoid freezing damage by using two strategies [1]. The first is freezing avoidance and involves supercooling of some cells and tissues to temperatures as low as -40°C. Supercooling occurs because of a lack of nucleating substances necessary for ice initiation and by the use of barriers for ice growth in some tissues. Although whole plants do not supercool to this extent, the few degrees of protection afforded by supercooling is a significant avoidance mechanism. The most prevalent strategy in overwintering plants is the development of tolerance to freezing [1]. At high freezing temperatures, ice nucleation is initiated in extracellular spaces because of a lower solute concentration [6]. This causes water migration from the cell to the extracellular ice crystal resulting in the loss of vital intracellular water. The degree of cellular dehydration increases with decreasing temperature to a point where dehydration and/or ice crystal formation is sufficient to kill the cells. There exists a great diversity in the capacities of plants to acclimate and survive freezing temperatures. For example, spring cereals, peas and potatoes withstand a few degrees of frost, while winter cereals and cabbage tolerate temperatures to about -33°C, and certain woody plants can survive temperatures lower than -50°C [1,7]. Understanding how these plants survive such freezing temperatures is not only a scientific challenge, but also has important economic applications. Different aspects of plant responses to LT have been reviewed in detail elsewhere and will not be covered here [1,6,8]. The goal of this review is to discuss the different biotechnological applications developed from the basic research in the area of LT tolerance in plants.

BNSDOCID: <XP_ ____1062997A__I_>

Molecular genetics and improvements of low temperature tolerance in plants

Winter survival and crop productivity are influenced by many different winter stresses [9–11] such as absolute freezing temperature, length of freezing period, ice encasement, flooding and oxidative stress caused by LT induced photoinhibition. Exposure of plants to LT produces morphological, biochemical, and physiological changes that are often highly correlated with plant FT and winter survival. This suggests that the phenotypic expression of FT is controlled by a large number of genes with complex interactions. The complexity of the LT response has made it difficult to separate genes responsible for LT acclimation and cold hardiness from those associated with metabolic adjustments to LT. Our inability to effectively characterize LT responses emphasizes the need for a better understanding of winter stresses, the LT responsive genes responsible for adaptation to winter stresses and their interaction with the environment [9]. Because of this complexity, progress in improving cold tolerance by traditional breeding of important crop species has been slow. To overcome this problem, considerable attention has been given to the potential use of genetic engineering to improve this important trait. Identification of genes contributing to LT tolerance and the understanding of their regulation would provide a previously unavailable opportunity for coupling traditional plant breeding methods with biotechnology for crop improvement. Analysis of the literature reveals a growing list of cold-regulated (COR) genes, with at least 50 now identified in different species [8].

The identification of a large number of COR genes makes it difficult to select the right gene for genetic engineering of LT tolerance in crop species that are hard to transform. Thus, transgenic studies with model species are an attractive approach to evaluate the potential of individual genes before proceeding with the transformation of crop species. Although these studies have had a limited success in improving FT, they provide valuable information on the mechanisms underlying this important trait. These studies have been classified into three groups based on the presumed function of the gene that was evaluated. The first represents genes encoding structural proteins that might be involved in protecting the cell during LT stress (Table 1). The second represents genes encoding enzymes involved in the biosynthesis of different osmoprotectants, in the desaturation of lipids, and in the antioxidative response (Tables 2 to 5). The third represents regulatory genes that control the overall mechanism of LT responses. These studies also revealed the potential and limitations of introducing single genes to modify LT tolerance in plants. In addition, they improved our knowledge of the different metabolic pathways associated with LT tolerance in plants. Many important factors need to be considered for a successful transformation [12]. These include the transformation techniques, choice of plant species, inducibility and tissue specificity of the promoter used, organellar targeting signal, post-translational modifications, availability of prosthetic group and precursors, microenvironment, and possible side-reactions of the new product.

BNSDOCID: <XP_____1062997A__!_>

Transgenic plants with structural proteins

Antifreeze proteins

AFPs were first identified in fishes as the causative agents of serum freezing point depression [13]. These proteins are unique in that they cause a freezing point depression far greater than would be predicted from colligative properties alone. Later studies demonstrated that AFPs lower the freezing point by interacting directly with the ice surface and inhibiting the binding of additional water molecules to the ice crystal lattice. This property is known as thermal hysteresis [14,15]. AFPs are also known to inhibit the recrystallization of ice, which is the growth of larger ice crystals at the expense of smaller ice crystals [16]. Larger ice crystals increase the possibility of physical damage within frozen tissues. The inhibition of the recrystallization of ice occurs at very low AFP concentrations (nM) and may be the function of AFPs in freezing tolerant organisms [17]. In addition, they may be involved in protecting cell membranes from cold-induced damage [18]. AFPs have been isolated and characterized from several species. The proteins for which sequence or structural information has been obtained include the five types of fish AFPs, two types from insects, three AFPs from rye plants and one from carrot [19–21]. These AFPs are unrelated in composition and structure. The levels of thermal hysteresis range from 0.2–0.6 C for plant AFPs [19], 1–2 C for fish AFPs [20], 5–10 C for insect AFPs [22]. Because the properties of these proteins suggested that they may be useful in protecting plants against cold temperatures, many transgenic studies were initiated to evaluate their potential (Table 1).

It was demonstrated in 1989 that vacuum infiltration of fish AFP into plant leaves produced a depression of the freezing temperature of the tissue [23]. This result led Hightower et al. [24] to transform tomato and tobacco plants with a synthetic AFP based on type I AFP from winter flounder and containing three, alanine-rich, 11 amino acid repeats. Transgenic plants accumulated the corresponding mRNAs, but did not show any antifreeze activity as measured by ice recrystallization inhibition. However, when fusion proteins containing five repeats and targeted to the cytoplasm or the apoplast were expressed in tomato, the transgenic lines were shown to accumulate the chimeric proteins to 0.1% of total soluble leaf proteins and to possess antifreeze activity. Unfortunately, the effects of transgene expression on cold tolerance of tissues were not examined. A synthetic fusion protein containing 5 repeats was also expressed in yeast [25]. The expressed protein was shown to inhibit the recrystallization of ice *in vitro*. Yeast cells expressing this AFP showed a two fold increase in survival after rapid freezing compared to non-transformed cells. In another study, tobacco plants transformed with the type I AFP from winter flounder expressed the AFP mRNA but not the protein under normal growth conditions [26]. However, AFP accumulated to detectable levels only after exposure of the plants to cold, possibly due to a higher stability of the α-helical protein at LT. Whether or not this amount of protein had any antifreeze activity was

BNSDOCID: <XP_____1062997A__I_>

Table 1. Transgenic plants expressing different structural proteins.

Class	Source and Protein	Target organism and Location	Effects on cold tolerance	Refs.
AFP	Fish AFP I *afa 3, afa5*	Tomato (cytoplasm and apoplast) Tobacco	Not tested	24
	Fish AFP I *afa5*	Yeast	Yes	25
	Fish AFP I	Tobacco	Not tested	26
	Fish AFP I (with potato codon usage)	Potato (apoplast)	Yes	27
	Fish AFP II	Tobacco (cytoplasm and apoplast)	No	28
	Carrot AFP	Arabidopsis	Not tested	29
	Carrot AFP	Tobacco	Not tested	21
Ice nucleator	*S. commersonii* INAZ	Tobacco Potato	Not tested	37,38
COR/LEA	*C. plantagineum* LEA group II	Tobacco	No	48
	Spinach CAP85 LEA group II	Tobacco	Yes	49
	Spinach CAP160 novel	Tobacco	Yes	49
	Barley HVA1 LEA group III	Rice	Not tested	50
	Wheat Em LEA group I	Yeast	No	51
	Tomato LE25 LEA group IV	Yeast	Yes	52
	Arabidopsis COR15a	Arabidopsis (chloroplast)	Yes	53
	Barley COR14b	Arabidopsis	Not tested	55

not determined. Recently, Wallis et al. [27] used codon usage frequencies observed in higher plants to construct a synthetic AFP gene based on type I AFP, and added a signal peptide directing it to the extracellular space. A transgenic potato line that expressed this protein to 0.6% of total leaf proteins showed a significant decrease of electrolyte leakage after freezing at -2 C compared to control plants. These results provided the first evidence of the capacity of AFPs to prevent freezing injuries in plant tissues. More recently, a chimeric construct containing the type II AFP gene from fish fused to the plant signal sequence directing secretion to the apoplast was introduced into tobacco [28]. The AFPs that accumulated to 2% of total apoplast proteins were shown to be active in both thermal hysteresis and the inhibition of ice recrystallization tests. However, visual evaluation of plants in the field following light frosts (-0.1 to -3 C) did not reveal any difference between transgenic and wild type plants. This study also reported that the production of plants that accumulate the protein in the cytosol was not successful. The transgenic

BNSDOCID: <XP _____1062997A__I_>

plants were sterile and only expressed the transcript for type II AFP. It has been proposed that the extensively disulfide-bonded protein is toxic when targeted to this compartment [28]. Transgenic tobacco and Arabidopsis plants expressing an active carrot AFP have been produced [21,29]. However, the cold tolerance of the transgenic plants was not assessed.

To date, the efforts to improve plant cold tolerance with AFPs have not been successful. Factors such as the sensitivity of the plant to freezing, the low levels of expression and subcellular localization can be responsible, in part, for such results. Future engineering studies with AFPs will also have to examine other considerations to design realistic transformation strategies. The possibility that AFPs may interact and stabilize cell membranes at temperatures above freezing [18] may be used to protect membranes in chilling sensitive plants. Further characterization is needed to define AFPs that possess this property and which of these AFPs can be expressed in plants to high levels without harmful effects. In addition, the identification of insect AFPs with very high levels of thermal hysteresis offers a way to promote supercooling of tissues in freezing sensitive plants. If these proteins can be made to accumulate to high levels in different compartments of the cell, a freezing point depression of a couple degrees could have an important role in frost protection of plants that cannot survive freeze-induced dehydration of the protoplasm. Finally, the identification of plant AFPs that possess enzymatic activities corresponding to three classes of pathogenesis-related proteins suggest that these proteins could be used to engineer freezing tolerant plants with both increased antifreeze activity in the apoplast and disease resistance. These proteins were shown to inhibit the recrystallization of ice which is known to take place most rapidly at temperatures just below freezing and when environmental temperatures fluctuate within the subzero range [30]. Because this activity is mediated by low concentrations of AFPs, targeting high levels of accumulation in the apoplast is realistic and can confer some protection to freezing tolerant plants. Because of the many potential applications of this group of proteins, the molecular and biochemical characterization of these proteins and their evaluation in transgenic studies need to be pursued.

Ice nucleators

Frost-hardy plants may increase their FT by initiating intercellular ice formation at high subfreezing temperatures [31]. Propagation of intercellular ice at subzero temperatures close to 0 C promotes gradual dehydration of cells, depressing both the cellular freezing and ice nucleation point. Siminovitch and Scarth [32] reported that lethal intracellular freezing occurs in acclimated plants that are allowed to supercool, and that initiation of ice formation at high subfreezing temperatures favoured cell survival. However, many hardy and non-hardy plant species can supercool extensively because they have no internal ice-nucleating agents active at temperatures higher than -5 C [33]. In these plants, supercooling allows a degree of protection during short-term exposure to high freezing temperatures but

BNSDOCID: <XP_____ _1062997A_ I_>

does not eliminate the risk of spontaneous intracellular freezing that causes cell death. On the other hand, ice nucleation active bacteria can catalyze ice formation at temperatures as warm as -2 C [34]. Genes encoding ice nucleation proteins from bacteria have been cloned and structurally characterized [35]. A similar high temperature ice nucleation activity has been detected within freezing tolerant insects and animals [36]. To determine if bacterial ice nucleation proteins can function in plant cells, the ice nucleation gene (*inaZ*) from *Pseudomonas syringae* was introduced into tobacco, a freezing sensitive species, and *Solanum commersonii*, a freezing tolerant species [37] (Table 1). Transformants of both species showed increased ice nucleation activity over untransformed controls. The temperature of the warmest freezing initiation in intact leaves was increased from approximately -7 C in the untransformed controls to -4 C in the transformants. It was also shown that 48 hours of LT are required for the maximum ice nucleation activity [38]. Similarly, LT exposure is required for maximum ice nucleation activity in bacteria [39]. The warmest ice nucleation temperature detected in transgenic plants was lower than that conferred by this gene in *P. syringae* (-2 C). This difference is probably due to a lower amount of INAZ protein or an inappropriate localization. In bacteria, these proteins are located in the outer membrane where they form large aggregates that serve as templates for high temperature ice nucleation [35]. Therefore, it remains to be determined if the INAZ protein can be targeted and assembled correctly in a compartment such as the apoplast where it would be most beneficial. These studies demonstrate the feasibility of expressing the bacterial gene *inaZ* in higher plants and open the door to engineering plants capable of initiating ice formation at higher subzero temperatures to improve survival. An alternative strategy to the use of bacterial ice nucleators may one day be the use of plant endogenous ice nucleators. Although they are less active, studies with rye have shown that ice-nucleating activity in cold acclimated plants may be mediated to some extent by a protein component [40]. The identification of this gene and the characterization of the encoded product will ultimately reveal how it functions and consequently its potential in high temperature ice nucleation.

COR/LEA proteins

During exposure of plants to freezing temperatures, the extracellular water freezes before the intracellular water because of its higher freezing point. The resulting lower water potential in the apoplast will draw water out of the cell, leading to dehydration of the protoplasm. It is well established that freeze-induced membrane damage results primarily from the severe dehydration associated with freezing [41,42]. Furthermore, the dehydration stress that occurs during freezing is likely to be similar to the one occurring during drought and salt stress, and so is expected to trigger some common genetic and physiological responses. The identification and characterization of COR genes in many plant species has revealed that many of them encode proteins with unknown functions [8]. A partial list of these genes is found in Tables 2 and 3 of Thomashow's recent review [8]. Analysis of these tables

BNSDOCID: <XP_____1062997A__I_>

reveals that at least two novel genes encode hydrophobic polypeptides, 9 novel genes encode hydrophilic polypeptides, while other genes encode the hydrophilic homologs of groups II and III of Late Embryogenesis Abundant (LEA) proteins. LEA proteins were initially found to be induced during late embryogenesis, prior to seed desiccation, and in plants during water stress. They were divided into different groups based on conserved structural features [43,44].

The hydrophilic COR proteins of unknown functions are referred to here as COR/LEA proteins. They are unusually hydrophilic, boiling soluble and many are composed largely of repeated amino acid sequence motifs. Sequence analysis reveals that many contain regions capable of forming amphipathic α-helices. Based on these properties, it has been suggested that these proteins may increase tolerance to freezing and dehydration by stabilizing proteins and membranes during dehydrative conditions [8,45–47]. Although many COR/LEA genes have been identified, only a handful have so far been evaluated in transgenic studies (Table 1).

Of all the genes induced during cold acclimation, the most widely surveyed encode proteins belonging to the group II LEA family [8]. A cDNA encoding a member of this group from the resurrection plant *Craterostigma plantagineum* was introduced into tobacco [48]. Transgenic plants that expressed this protein did not show any changes in the tolerance to osmotic stress as measured by the ion-leakage test. In another study, two genes, *cap85* (LEA group II) and *cap160* (novel gene) from spinach, were constitutively expressed in tobacco [49]. Plants expressing the proteins, individually or together, were indistinguishable from wild type plants as assessed by the FT of detached leaves. However, when a kinetic study of freezing at -2 C was carried out, transgenic plants expressing the spinach proteins had slightly lower rates of electrolyte leakage indicating a reduction of freezing stress injury. The potential of overexpressing a group III LEA gene from barley was assessed in rice [50]. Transgenic rice plants showed an increased tolerance to water deficit and salinity as reflected by their higher growth rates, delayed development of symptoms and improved recovery. However, the effect of cold stress on the performance of these transgenic plants was not measured. Although other groups of LEA proteins were not yet shown to be regulated by cold temperatures, they have been evaluated in transgenic studies. For example, the group I LEA protein from wheat was overexpressed in yeast cells and shown to attenuate the growth inhibition normally observed in high osmolarity media [51]. However, no correlation was established between the expression of this protein and increased survival after overnight freezing. Overexpression of a class IV LEA protein from tomato in yeast conferred a small but significant increase in FT [52]. After 24h at -20 C, 8% of the transformed yeast strain survived compared to 2% for the control strain.

Genes encoding novel proteins have also been expressed in plants. The COR15a gene from Arabidopsis, which encodes a protein that is targeted to the stromal compartment of chloroplasts, was constitutively expressed in transgenic *Arabidopsis* plants [53]. Non-acclimated transgenic plants showed an increase of 1 to 2 C in the FT of both chloroplasts frozen *in situ* and isolated leaf protoplasts

BNSDOCID: <XP_ ____1062997A__I_>

frozen *in vitro* over the temperature range of -4 to -8 C. Subsequent investigations have indicated that COR15a may increase FT by decreasing the propensity of membranes to form hexagonal II phase lipids, a major cause of membrane damage in non-acclimated plants over the temperature range of about -4 to -8 C [54]. These results led this group to propose that proteins that contain possible amphipathic α-helices such as COR15a may alter the intrinsic curvature of membrane monolayers thus leading to a shift in the lamellar-to-hexagonal II phase transition temperature [8,54]. Another novel gene, COR14b from barley, was introduced into *Arabidopsis* plants [55]. This gene also encodes a protein that is targeted to the stromal compartment of chloroplasts. However, the effect of its expression on the FT of chloroplasts or whole plants was not measured.

Because FT is considered to be a multigenic trait, it is not surprising that in the studies described above, the expression of one or two COR/LEA proteins does not significantly alter the tolerance to freezing. However, the demonstration that an amphipathic α-helix containing protein defers the occurrence of the freeze-induced formation of the hexagonal II phase to lower temperatures is significant. Many COR/LEA proteins contain such regions and have been localized to cellular compartments such as the nucleus [56], the cytoplasm [56], the chloroplast [57] and the paramural space [47]. If these proteins are found to function in a manner similar to COR15a, they may be used to cryoprotect different membranes. A modest increase of FT by 1 or 2 C could have a tremendous impact on freezing tolerant crops in early fall and late spring or in freezing sensitive species that suffer membrane injuries due to the formation of the hexagonal II phase [54]. However, newer strategies to improve FT will have to consider that alteration of membrane lipid composition [42] and the accumulation of sugars and compatible solutes [58–59] also contribute to decreasing membrane injuries caused by freezing.

Metabolic engineering to improve cold tolerance

When plants are exposed to abiotic stresses such as low and freezing temperatures, membrane fluidity and membrane bound processes are perturbed [60]. In addition, when extracellular ice forms, there is an efflux of water from the cells causing both a dehydration of the protoplast and an increase in the concentration of intracellular solutes, thus putting a strain on membranes and macromolecules [6]. Under conditions in which chloroplasts are exposed to excess excitation energy generated by the interaction of light and LT, there is an increase in photoreduction of oxygen and concomitant production of reactive oxygen intermediates, such as superoxides and peroxides, which can damage membranes and enzymes [12]. Plants react to these stresses by inducing various responses. They synthesize compatible solutes and sugars needed for osmotic adjustment and osmoprotection [6]. They modify the lipid species and their degree of unsaturation to preserve membrane fluidity and prevent freeze induced membrane damage [60,61]. In addition, they activate antioxidant enzymes such as peroxidase and superoxide dismutase, that scavenge reactive oxygen intermediates [62,63]. Knowledge

BNSDOCID: <XP_____1062997A_..._I_>

generated from these studies was used to design strategies to improve stress tolerance [58,59,64,65]. Among these approaches are: 1) the introduction of genes encoding enzymes that catalyze the conversion of naturally occurring substrates into products with osmoprotective properties, 2) genes encoding lipid-modifying enzymes, and 3) genes encoding enzymes involved in the antioxidative response.

Amino acid derived osmoprotectants

Osmoprotectants (also termed compatible solutes) occur in all organisms from archaebacteria to higher plants and animals [64,66]. They are highly soluble compounds that carry no net charge at physiological pH and are non-toxic at high concentrations. Osmoprotectants serve to raise osmotic pressure in the cytoplasm and can also stabilize proteins and membranes when salt levels or temperatures are unfavourable [59,64]. Osmoprotectants therefore play important roles in the adaptation of cells to various adverse environmental conditions [67]. There are many amino acid derived osmoprotectants that have been identified in plants under stressful conditions like proline [68–73], 3-dimethylsulfoniopropionate [64], polyamines [74,75], trigonelline [76] and betaines [64,68,77,78].

Although there are many examples of amino acid derived osmoprotectants, metabolic engineering studies to date have concentrated on increasing proline and betaine accumulation [12,64] (Table 2). Proline biosynthesis from the amino acid glutamate can be achieved by using the *p5cs* gene from mothbean [*Vigna aconitifolia*] encoding a bifunctional enzyme containing the catalytic activities of γ-glutamyl kinase and glutamic-γ-semialdehyde dehydrogenase [12]. Two studies have used this gene to produce transgenic plants with increased proline production [79,80] but tolerance to LT was not measured. However, transgenic tobacco plants with increased proline content [79] were shown to be more osmotolerant than control plants. In the second study, this gene was placed under the control of an ABA inducible promoter thus leading to overproduction of proline only under stress conditions [80]. Transgenic rice plants that accumulated proline to high levels during salt and water stresses showed an enhanced biomass production compared to non-transformed control plants.

Glycine betaine is synthesized in plants via a two-step oxidation of choline which is derived from serine [64,81]. In spinach, the oxidation steps are carried out by the two chloroplastic enzymes choline mono-oxygenase (CMO) and betaine aldehyde dehydrogenase (BADH). The first reaction is ferredoxin-dependent and oxidizes choline to betaine aldehyde. The second reaction is NAD-dependent and converts betaine aldehyde to betaine [64]. This pathway is not the only one in plants because studies with transgenic tobacco plants expressing the barley BADH showed that this enzyme was targeted to peroxisomes. This suggests that in monocots BADH may be mostly localized in peroxisomes [82]. Since there is no reduced ferredoxin in the peroxisomes [64], it is unlikely that the choline-oxidizing enzyme in monocots is CMO. In the bacterium *E. coli*, the conversion of choline to betaine occurs via the two choline dehydrogenases (BET A and BET B) and in the soil

BNSDOCID: <XP_____1062997A_I_>

Table 2. Transgenic plants accumulating amino acid related osmoprotectants.

Class	Source and Protein	Target organism and Location	Effects on cold tolerance	Refs.
Proline	Mothbean P5CS	Tobacco	Not tested	79
	Mothbean P5CS	Rice	Not tested	80
Betaine	*A.globiformis* COD A	*Synechococcus*	Yes	83
	A.globiformis COD A	Arabidopsis (chloroplast)	Yes	84-88
	A.globiformis COD A	Rice cytosol and chloroplast	Yes	89
	Barley BADH	Tobacco	Not tested	82
	Spinach BADH	Tobacco	Not tested	90
	Spinach CMO	Tobacco	Not tested	91
	E. coli BADH (BET B)	Tobacco	Not tested	92
	E. coli (BET A)	Tobacco	Not tested	93
	E. coli (BET A and BET B)	*Synechococcus*	Not tested	94

bacterium *Arthrobacter globiformis* by the one-step choline oxidase (COD A) [64,77].

To date, only Murata et al. have studied the effects of betaine accumulation on LT tolerance. They transferred the *codA* gene into several species such as the cyanobacterium *Synechococcus sp.* strain PCC 7942 [83], *Arabidopsis thaliana* [84–88] and rice [89]. Compared to control cells, transformed cyanobacteria cells expressing *codA* showed an enhanced ability to grow at the usually non permissive temperature of 20 C in a medium supplemented with 1mM choline chloride [83]. Furthermore, transformed cells were more resistant to LT-induced photoinhibition. These effects were attributed to betaine accumulation mediated by the *codA* gene [83]. *Arabidopsis* plants expressing codA showed an enhanced growth, germination and survival in the presence of NaCl [85]. When plants were exposed to 5 C under continuous light followed by a recovery period at 22 C, leaves of wild type plants exhibited symptoms of chlorosis whereas those of transformed plants did not. The activity of photosystem II in transgenic plants remained at a higher level than in wild type plants under LT, high light and salt stresses [85,88]. Furthermore, transgenic *Arabidopsis* plants had a greater tolerance to LT, heat stress and high salt during germination and early growth [84,86,87]. Wild-type seedlings germinating at LT were morphologically distorted while transgenic plants were not [86]. Cold sensitive rice plants engineered to express the codA protein, in the chloroplast or cytosol, showed a faster recovery rate compared to wild type plants after exposure to salt stress [89]. This study also showed that accumulation of betaine in chloroplasts confers a higher degree of tolerance to salt- and LT-induced photoinhibition [89].

In spite of the encouraging results regarding the effect of betaine accumulation on

BNSDOCID: <XP_____1062997A_I_>

LT tolerance in transgenic plants, poor accumulation of betaine has been observed due to a limited choline supply [91]. These low levels of accumulation are not enough for betaine to function as an osmotic adjustment agent. Therefore, the observed effects in these studies are probably due to the stabilization of protein structure and function [89]. Research must be continued to further control the flux of choline in transgenic plants and may require additional engineering steps [91]. Studies aimed at improving betaine production in plants are worthwhile because it was recently shown that exogenous application of betaine could improve FT by more than 5 C in both non-acclimated and cold acclimated wheat plants [95]. This increase in tolerance is the highest reported for plants treated with exogenous chemicals or for stably transformed plants. In addition, *in vitro* cryoprotection assays have shown that betaine can protect liposomes, bacteria and plant tissue during freezing [96–99].

Sugar related osmoprotectants

Accumulation of sugar related compounds such as sucrose, raffinose, sorbitol and fructans during cold acclimation of plants has been observed repeatedly [100–109]. Many of these compounds are believed to stabilize membranes and proteins during the dehydrative conditions induced by freezing or drought [6]. According to Crowe et al. [110], sugars can replace water in the maintenance of dehydrated proteins or membranes in a physical state similar to that seen in the presence of water. In addition, an alternative function has been proposed for the disaccharides sucrose and trehalose which can form a glass phase in the dry state [111]. A glass is a liquid of such high viscosity that it is capable of slowing chemical reactions. It was proposed that glass formation (vitrification) could lead to long-term stability of a living system [112]. Transgenic plants expressing different sugar biosynthetic enzymes have been produced with the primary objective to study carbohydrate metabolism and not to evaluate their potential effects on tolerance to LT or other stresses [113–116]. However, some recent advances have furthered our understanding of carbohydrate metabolism and are noteworthy to mention because they could be used in molecular engineering strategies (Table 3).

Trehalose accumulation has long been detected during the onset of desiccation tolerance of organisms such as yeast and resurrection plant [117,118] but its presence was never observed in tissues of other plants, probably due to the presence of high levels of trehalase activity. A most unexpected finding in recent research was the discovery that the trehalose biosynthetic pathway is ubiquitous in plants [117]. Transgenic tobacco plants overexpressing the yeast trehalose phosphate synthase gene (*tps1*) showed a significant tolerance to drought [119,120]. However, these trehalose-accumulating plants exhibited reduced growth under non-stressful conditions [119], and multiple phenotypic alterations [120]. Transgenic tobacco and potato plants were also produced which express the *E. coli* trehalose-6-phosphate synthase and trehalase-6-phosphatase genes [*tps1* and *tpp*] [121,122]. These plants had an improved growth performance during drought but also showed

BNSDOCID: <XP_____1062997A__1_>

Table 3. Transgenic plants expressing enzymes involved in the biosynthesis of sugar related osmoprotectants.

Class	Source and Protein	Target organism and Location	Effects on cold tolerance	Refs.
Trehalose	Yeast TPS1	Tobacco	Not tested	119
	Yeast TPS1	Tobacco	Not tested	120
	E. coli TPS1 AND TPP	Tobacco Potato	Not tested	121,122
	Yeast TPS1	Yeast	Yes	124
Fructans	*B. subtilis* Fructosyl transferase	Tobacco	Not tested	128,129
Mannitol	*E. coli* MTLDH	Tobacco (cytoplast and chloroplast)	Not tested	130,131
	E. coli MTLDH	*Arabidopsis*	Not tested	132
	E. coli MTLDH	Tobacco	Not tested	133
	E. coli MTLDH	Tobacco	Not tested	134
D-ononitol	Ice plant IMT1	Tobacco	Not tested	136
Inositol	*S. polyrrhiza* INS3PS	*Arabidopsis*	Not tested	137
Sorbitol	Apple S6PDH	Tobacco	Not tested	138
	Apple S6PDH	Tobacco	Not tested	139

some morphological alterations. Because these alterations may be caused by some as yet unidentified function of trehalose in plants [117], the negative effects on plant development may in the future be alleviated by using a promoter sequence that drives expression in specific cell types or under stress conditions [121]. A general observation from these studies is that trehalose accumulates to low levels in transgenic plants. It was shown that this problem could be circumvented by inhibiting endogenous trehalase, an enzyme involved in trehalose turnover [121]. Furthermore, it was recently shown that expressing trehalose phosphate synthase or neutralizing trehalase could improve FT in yeast [123,124]. Based on these results, attempts to engineer trehalose accumulation will undoubtedly be pursued.

Several flowering plant species accumulate, as their carbohydrate storage pool, the linear and branched polymers of fructose called fructans [115]. The degree of polymerisation of plant fructans varies greatly but generally the fructosyl chain is about 30 to 50 fructosyl residues long. The global distribution of fructan-accumulating plants shows that they are especially abundant in temperate climate zones with seasonal drought or frost, but that they are almost absent in tropical

BNSDOCID: <XP_____1062997A__I_>

regions [125]. The physiological role of fructans is not fully understood. They are thought to lower the cellular sucrose concentration to prevent sugar-induced feedback inhibition of photosynthesis [126] and also be involved in drought and cold tolerance [127]. Unlike starch which is insoluble and accumulates in amyloplasts, fructans are soluble molecules that accumulate preferentially in vacuoles. A bacterial fructan biosynthetic enzyme was used to transform tobacco plants [128,129]. The fructan-producing transgenic plants were found to be more resistant to PEG-induced drought stress, as determined by their growth properties and biomass accumulation. In this study, the use of a bacterial enzyme resulted in the production of fructans with a degree of polymerisation of more than 100000 fructosyl residues [129]. Because such molecules are not normally present in plants, they may themselves induce a stress response that affects the drought tolerance of these plants [115]. The availability of plant fructan biosynthetic genes will allow the detailed study, using transgenic plants, of the effects of fructan accumulation on stress tolerance [115].

Polyols such as mannitol [130–134], D-ononitol [135,136], inositol [137] and sorbitol [138–139] form another class of compounds with osmoprotective activities that have been the targets of molecular engineering studies. These sugar alcohols are closely related to sugars, as they represent the chemically reduced form of an aldose or ketose sugar. It has been suggested that sugar alcohols, because of their water-like hydroxyl groups, may mimic the structure of water and maintain an artificial sphere of hydration around macromolecules [58,65,114]. While transgenic *Arabidopsis* plants accumulating inositol did not show any significant difference in stress tolerance, the transgenic plants accumulating mannitol, D-ononitol and sorbitol did show some enhanced tolerance to drought or salt stress. These studies are particularly instructive because they show that engineering a new pathway for sorbitol accumulation in tobacco resulted in growth defects and necrosis [139]. This was attributed, in part, to interference with inositol biosynthesis and alterations in carbohydrate allocation and transport. Furthermore, engineering a new pathway for D-ononitol accumulation in tobacco makes D-ononitol production stress inducible because its precursor inositol is also stress inducible in this plant [136]. These studies highlight the importance of metabolic analyses of different transgenic and stressed plants to estimate *in vivo* pathway fluxes and intermediate pool sizes [77]. Once available, such data could be incorporated into metabolic models and used to predict the impact of inserting a transgene on pathway flux.

Lipid modifications

In cold sensitive plants, temperatures ranging from 0 C to about 12 C restrict growth and may even result in the plant's death [140]. A widely held view was that the high level of disaturated species of phosphatidylglycerol (PG) in the chloroplast membranes was at the base of the chilling sensitivity of these plants [60]. Many enzymes capable of lipid desaturation in plants and other organisms have been identified [141,142], and some were used to produce transgenic plants to investigate

BNSDOCID: <XP_____1062997A_I_>

Table 4. Transgenic plants expressing lipid modifying enzymes.

Class	Source and Protein	Target organism and Location	Effects on cold tolerance	Refs.
Acyltransferase	*Arabidopsis* G3PAT	Tobacco	Yes	144
	Squash G3PAT		No increase	
	E. coli G3PAT	*Arabidopsis* (chloroplast)	No increase	146
	Arabidopsis G3PAT	Rice	Yes	145
Desaturase	*Arabidopsis fad7*	Tobacco	Yes	148,149
	Arabidopsis fad3	Tobacco	No	150-152
	A.nidulans Δ9-desaturase	Tobacco (chloroplast)	Yes	153

the effects of changing lipid unsaturation on chilling sensitivity [60,143] (Table 4). The first direct evidence of a link between the level of saturated lipids and chilling sensitivity was provided by Murata and colleagues [144]. They cloned the cDNAs for the plastid glycerol-3-phosphate acyltransferase from squash, a chilling-sensitive species, and from *Arabidopsis*, a chilling-resistant species. The enzyme from squash preferentially transfers saturated lipid species to the *sn-1* position of glycerol-3-phosphate while the *Arabidopsis* enzyme selectively transfers unsaturated lipid species. Transgenic tobacco plants expressing the squash cDNA contained elevated levels of disaturated PG (76%), whereas plants expressing the *Arabidopsis* cDNA contained slightly reduced levels of disaturated PG (28%) compared with wild type plants (36%). Plants expressing the squash enzyme showed severe signs of chilling sensitivity as measured by LT-induced photoinhibition in leaf disks, and development of chlorosis in LT-exposed plants. This observation has now been extended to rice plants that have been transformed with the cDNA from *Arabidopsis* [145]. Transgenic rice plants expressing this gene had higher levels of unsaturated species of PG and were found to have higher net photosynthetic rates when exposed to 17 C compared to control plants. Other groups also analysed the effects of increased levels of saturated molecular species of PG by targeting the product of the *plsB* gene of *E. coli* to the chloroplasts of *Arabidopsis* [146] or studying the *fab 1* mutant of *Arabidopsis* [147]. Transgenic plants expressing the *plsB* gene displayed a chilling-sensitive phenotype while *fab 1* mutants did not. However, *fab 1* mutants did show growth retardation after prolonged exposure to LT, and eventually died if exposed for more than 4 weeks. Experiments with tobacco and *Synechocystis* [60] have suggested that higher levels of unsaturated membrane lipids accelerate the recovery process from the LT-induced photoinhibition and may be the determining factor in the chilling-insensitive phenotype.

An alternative approach to modify the level of unsaturation in membrane lipids is to express fatty acid desaturase genes in plants. Transgenic tobacco plants have been produced that overexpress the chloroplastic fad7 ω-3 desaturase from *Arabidopsis* responsible for the formation of trienoic fatty acids (16:3, 18:3) in leaf tissues

BNSDOCID: <XP_____1062997A__I_>

[148,149], the microsomal fad 3 ω-3 desaturase from *Arabidopsis* responsible for the formation of 18:2, 18:3 fatty acids [150–152], and the Δ9-desaturase from *Anacystis nidulans* targeted to the chloroplast which introduces a cis-double bond at the Δ9 position of both 16 and 18 carbon saturated fatty acids [153]. All transgenic plants contained higher levels of unsaturated fatty acids compared to control plants, but only the fad 7 and Δ9 desaturase expressing plants exhibited a significant increase in chilling tolerance. Failure to detect an increase in the LT tolerance of transgenic plants expressing the microsomal *fad 3* gene may indicate that unsaturation of chloroplast lipids may play a more important role in chilling tolerance [142,152]. Results from these transgenic studies indicate that there may exist a realistic potential in alleviating the chilling sensitivity of some species by combining several engineering steps such as decreasing disaturated PG and increasing unsaturation.

Compared to chilling sensitive plants, freezing tolerant plants can grow at low positive temperatures and induce adaptive changes that enable them to withstand subsequent exposures to freezing. In these plants, the primary cause of freezing injury is the destabilization of the plasma membrane during freeze-induced dehydration [41]. Several studies have now documented the subtle changes induced by cold acclimation in the plasma membrane lipid composition of rye [154], *Solanum* species [155], oat [156], and *Arabidopsis* [156]. Changes that have been proposed to increase the cryostability of the plasma membrane during freeze-induced dehydration are an increase in the proportion of di-unsaturated species of phosphatidylcholine (PC) and phosphatidylethanolamine and a decrease in the proportion of cerebrosides and sterol containing lipids [156]. Although membrane-engineering studies [157,158] have shown that increasing the proportion of di-unsaturated species of PC in the plasma membrane decreases membrane injury, no such genetic studies have yet been reported. Manipulating the proportion of certain lipid species in the model plant *Arabidopsis*, containing endogenous mechanisms for frost tolerance, could be a valuable tool in evaluating the potential benefits or limitations of these changes on the cryostability of the plasma membrane during freeze-induced dehydration. However, studies with the unicellular organism *Acholeplasma laislawii* [159] indicate that the proportion of different lipid species may be controlled by the activity of the enzymes that appear to be responsive to the physical state of the membrane in which they operate. If such a feedback mechanism is in operation in freezing tolerant plants during cold acclimation, it may prove difficult to genetically manipulate the relative proportions of lipid species to enhance the cryostability of the plasma membrane. A greater understanding of the mechanisms that bring about changes in lipid species in these freezing tolerant plants will indicate if this genetic engineering approach has potential.

Antioxidative response

Free radicals and other active derivatives of oxygen are inevitable by-products of biological redox reactions. During normal growth conditions, plants are well

BNSDOCID: <XP_____1062997A__I_>

adapted for minimizing the damage that occurs from misuse of photosynthetic energy transfer. However, perturbation of electron transport under conditions of photoinhibition can lead to the transfer of an excess of electrons to oxygen and create an oxidative stress. Such conditions are encountered by crops that are exposed to periods of cold or drought during a growing season, and this affects plant productivity [62,63]. During environmental stresses, plants increase the synthesis of non-enzymatic antioxidants such as the tripeptide glutathione (GSH) and vitamins C (ascorbate) and E (α-tocopherol), as well as antioxidant enzymes such as superoxide dismutase (SOD), glutathione peroxidase (GPX), glutathione reductase (GR), ascorbate peroxidase (APX) and catalases [160]. However, the degree of enhancement of the antioxidative response in plants is limited by the relatively poor capacity for induction of the enzymes of the defense system [161]. Therefore, to evaluate if tolerance to oxidative stress could be increased, many studies have characterized transgenic plants expressing genes encoding different oxygen radical scavenging enzymes (Table 5). These studies have been extensively reviewed elsewhere [62,161–166].

SODs are enzymes that react with superoxide radicals at almost diffusion-limited rates to produce hydrogen peroxide [167]. Although there has been some inconsistencies in the phenotypes of transgenic plants expressing SOD, the overall observation is that elevated levels of SOD in a variety of cellular compartments can lead to a protection from oxidative stresses caused by the herbicide methyl viologen (MV), high light intensity and low and freezing temperatures [163–165]. The conflicting results in these studies have been attributed to the sensitivity of Cu/Zn SOD to peroxide compared to Mn SOD, the suborganellar association of Cu/Zn SOD with PSI, and possible induction of an antioxidative response by the peroxide produced [161,163–165]. Preliminary studies with transgenic plants that overexpress APX or GR show that they had reduced damage following either MV or photooxidative treatment [164,165]. Furthermore, the expression of a combination of antioxidant enzymes, GR and SOD, in transgenic tobacco plants was shown to have synergistic effects on stress tolerance [168]. Further characterization of transgenic plants overexpressing GR [169] and different forms of SOD [170–173] have been reported. The justification of such studies was recently demonstrated in yeast by showing the importance of the cytoplasmic SOD for resistance to freeze-thaw stress [174]. The overexpression of bacterial GR in the chloroplasts of *L. esculentum* did not lower its chilling sensitivity [169] while transgenic maize plants that overproduced Fe SOD or MnSOD in chloroplasts displayed better growth rates during chilling and enhanced tolerance to MV [171,172,175]. Field trials of two elite alfalfa plants transformed with either a Mn SOD targeted to the mitochondrion or the chloroplast show that, after one winter, most transgenic plants had higher survival rates and herbage yield than control plants [173]. This extends the previous observation that overexpression of Mn SOD in transgenic alfalfa plants often improves the winter survival of this crop [176,177].

Recent studies have also evaluated the potential of increasing resistance to

BNSDOCID: <XP_____1062997A__I_>

Table 5. Examples of transgenic plants expressing enzymes involved in the antioxidative response.

Class	Source and Protein	Target organism and Location	Effects on cold tolerance	Refs.
SOD	Petunia	Tobacco (chloroplast)	Not tested	191,192
	Cu/Zn SOD (chloroplastic)	Tomato (chloroplast)	Yes	
	Pea	Tobacco	Not tested	193
	Cu/Zn SOD (chloroplastic)	(cytoplasm and chloroplast)		
	Tomato	Potato	Not tested	194
	Cu/Zn SOD	(cytoplasm and chloroplast)		
	Pea	Tobacco	Yes	195,196
	Cu/Zn SOD	(chloroplast)		
	Tobacco	Alfalfa	Yes	173,176
	Mn SOD	(mitochondria and chloroplast)		177
	Tobacco	Tobacco	Not tested	197-199
	Mn SOD	(mitochondria and chloroplast)		
	Tobacco	Maize	Yes	171,175
	Mn SOD	(chloroplast)		
	Arabidopsis	Tobacco	Yes	200
	Fe SOD	(chloroplast)		
	Arabidopsis	Poplar	No	170
	Fe SOD	(chloroplast)		
	Arabidopsis	Maize	Not tested	172
	Fe SOD	(chloroplast)		
GR	*E. coli*	Tobacco	Not tested	168,
	GR	(cytoplasm and chloroplast)		204,205
	E. coli	Poplar	Not tested	201-203,208
	GR and GS	(chloroplast)		
	Pea	Tobacco	Not tested	206,207
	GR	(cytoplasm and chloroplast)		
γ-ECS	*E. coli*	Poplar	Not tested	209-211
	γ-ECS	(cytoplasm and chloroplast)		
	E. coli	Tobacco	Not tested	190
	γ-ECS	(chloroplast)		
OTHERS	Pea	Tobacco	Not tested	212
	APX	(chloroplast)		
	E. coli	Tobacco	Not tested	182
	Catalase	(chloroplast)		
	E. coli	Tobacco	Yes	181
	GST/GPX			
	Tobacco	Tobacco	Not tested	185
	AOX			
	B. juncea	Tobacco	Not tested	187
	Glyoxalase I			

oxidative stress by using novel approaches. Mannitol, a compatible solute, was previously shown to act as a hydroxyl radical scavenger *in vitro* [178]. The targeting of mannitol biosynthesis to chloroplasts in transgenic tobacco plants resulted in an increased resistance to MV-induced oxidative stress [133]. Another promising approach is to use enzymes with both glutathione S-transferase and glutathione peroxidase activities (GST/GPX) that are believed to play an important role in

BNSDOCID: <XP__ ____1062997A__I_>

the scavenging of the toxic by-products of lipid peroxidation in plants [179,180]. The expression of a cDNA encoding a GST/GPX enzyme conferred, in young transgenic tobacco seedlings a growth advantage during chilling and salt stress [181]. However, this advantage was transient because after 2 to 3 weeks no differences were discernible. Another group has characterized the overexpression of a bacterial catalase in chloroplasts of tobacco [182]. Photosynthesis of transgenic plants was more tolerant to high irradiance under drought conditions compared to wild-type plants. However, the severity of the stress resulted in the complete inactivation of endogenous ascorbate peroxidase in transgenic and wild-type plants. These results suggest that under such conditions qualitative modifications of the antioxidative machinery may be more important than quantitative modifications. Another enzyme that has been proposed to function in plants by limiting formation of reactive oxygen species is alternative oxidase (AOX) localized in mitochondria [183,184]. Transgenic cultured tobacco cells with increased levels of AOX had lower levels of formation of reactive oxygen species in mitochondria [185]. Limiting oxidative stress in mitochondria may find applications because recent work with maize [186] has shown that chilling involves other subcellular organelles in the production of peroxide such as the mitochondria. Another group has studied the potential of overexpressing an enzyme of the glyoxalase system in transgenic tobacco plants [187]. The glyoxalase system is involved in the protection against methylglyoxal toxicity, generated at low levels spontaneously or enzymatically from glyceraldehyde 3-phosphate and dihydroxyacetone phosphate [188,189]. Glyoxalase I, the enzyme used in their study [187], catalyses the formation of S-D-lactoylglutathione from the hemithioacetal formed non-enzymatically from methylglyoxal and reduced GSH [188]. Glyoxalase II catalyses the hydrolysis of the product to D-lactic acid and regenerates the reduced GSH. Transgenic plants overexpressing glyoxalase I showed a significant improvement in tolerance to methylglyoxal and high salt, as tested in a detached leaf disc senescence assay [187]. These results suggest that a potential advantage of overexpressing enzymes of the glyoxalase system would be in reducing the level of methylglyoxal under stress conditions and enabling regeneration of GSH. Another approach that was attempted is to engineer an increased GSH biosynthesis capacity [190], a major non-enzymatic antioxidant that is perceived to be particularly important in protecting chloroplasts from oxidative damage. Increased GSH biosynthetic capacity in the chloroplasts of transgenic tobacco plants resulted in greatly enhanced oxidative stress, which was manifested as light intensity-dependent chlorosis or necrosis. Because foliar pools of both GSH and γ-glutamylcysteine (the immediate precursor to GSH) were altered, it was proposed that transgenic plants suffered continuous oxidative damage caused by a failure of the redox-sensing process in the chloroplast [190].

Studies in this field have demonstrated that some components of the antioxidative response show a potential in protecting plants against oxidative stress. However, further knowledge is needed on the level of oxidative stress that cellular compartments experience under different abiotic stresses. Such information would

BNSDOCID: <XP_____1062997A__I_>

allow the development of strategies to compensate for the different responses observed in various species or under different abiotic stresses. Moreover, the increased expression of a single antioxidative response component may not have a drastic effect on the tolerance to a particular abiotic stress but may still have an impact on biomass production during a growing season. A case that illustrates this point is the study of transgenic alfalfa plants expressing MnSOD in the chloroplast or mitochondria [173]. This study also showed that in some independent transgenic plants, winter survival and yield actually decreased and that the transformation event had a greater influence on survival and yield than the subcellular site of SOD targeting. These observations stress the importance that a reliable evaluation of a transgene's potential in alleviating antioxidative or other stresses may require the analysis of many independent transgenic plants to detect a tendency. Furthermore, this study did not reveal any differences in cellular viability during freezing between transgenic and control plants. Therefore, field survival of transgenic plants may be mediated by the beneficial effect of increased SOD activity under other field stress conditions such as ice encasement, flooding, or long-term freezing [173]. The use of various transgenic plants is indispensable to evaluate the potential of components of the antioxidant response system for increased field survival and productivity during abiotic stresses.

Global regulators (Master switch)

Increasing evidence provided by genetic and molecular analyses have revealed that the cold-induced increase in FT is regulated by a multisignal transduction network [213,214]. The current working model proposes that cold is first sensed by either the plasma membrane or cytoplasmic and chloroplastic molecular complexes which transduce the signal via a cascade of kinases and phosphatases, leading to the activation of specific transcription factors which in turn activate the transcription of the plethora of COR genes. This differential gene expression is responsible for both the metabolic adjustment of growth at LT and the development of FT. Understanding the regulation of these molecular events is the goal of many laboratories. This knowledge may provide means to produce, by genetic transformation, transgenic plants in which all the genes needed for FT are expressed. Groundbreaking results have been obtained thus far by identifying and overexpressing two signal transduction molecular targets, a transcription factor and a kinase.

The recent cloning of the transcription factors CBF1 and DREB1A by the groups led by Michael Thomashow and Kazuo Shinozaki marked an important milestone in the field of molecular genetics of cold acclimation [215,216]. These proteins and related members bind to the *cis*-acting element CCGAC present in the promoter region of COR genes [215–217]. The genes encoding these proteins were expressed in transgenic *Arabidopsis* plants. The results were striking: not only were COR genes expressed at higher levels, but most importantly, the plants showed significant increases in FT [216,218,219]. These results undoubtedly represent the most significant advance in the field of FT improvement, as was outlined in some recent

BNSDOCID: <XP_____1062997A_I_>

reviews [220,221]. It remains to be determined, however, if CBF/DREB homologs exist in other species particularly in wheat and rye, the two cultivated crops showing the best FT. Although the molecular components involved in regulating LT gene expression in these species have yet to be identified, recent studies with cereals have revealed the chromosome that has a major positive impact on cold tolerance and gene expression [46]. Studies with the Cheyenne/Chinese Spring chromosome substitution series have shown that chromosome 5A positively regulates the expression of at least four COR gene families correlated with tolerance. This chromosome was also the one most frequently found to have the greatest effect on tolerance in wheat and barley. These findings suggest that chromosome 5A contains gene(s) that constitute the 'master switch' governing the expression of COR genes in cereals (Fig. 1). While the exact nature and number of genes that compose the master switch on chromosome 5 are not known, these results suggest that more realistic transformation strategies could also be used in cereals for the production of more tolerant crops.

In wheat, as many as 15 out of 21 chromosomes have been shown to influence cold tolerance [222]. Thus, it is unlikely that single regulators like CBF/DREB in *Arabidopsis* or the master switch in cereals would be the sole determinants of COR gene expression, especially since so many genes show an increased expression at LT [8,223]. In addition, several genes possibly involved in FT (*sfr*) in *Arabidopsis* remain to be identified or cloned [224], and it is likely that some of them will be cold inducible but that their promoter region will lack the CCGAC element bound by the CBF/DREB proteins. An example is the *WCS19* gene from wheat, which is cold inducible but shows no CCGAC element in its promoter region [225, Ouellet and Sarhan, unpublished results]. It is therefore likely that other transcription factors involved in the activation of COR genes exist and must be identified so as to increase the tools at our disposal for elaborating strategies for engineering cold tolerance.

Another promising approach for the improvement of stress tolerance in plants was recently shown by the transformation of yeast and tobacco cells with *At-DBF2*, a serine/threonine protein kinase [226]. This gene was isolated from Arabidopsis and is the homologue of the yeast *Dbf2* gene, which encodes a protein kinase present in the multisubunit CCR4 general transcriptional complex. *At-DBF2* was isolated by screening for osmotolerance in yeast, after transformation with an *Arabidopsis* cDNA library prepared from siliques. The rationale behind the screen was that siliques contain the maturing seeds where is triggered a developmental program that will lead to the acquisition of desiccation (hence osmotic) tolerance. When At-DBF2 or DBF2 were overexpressed in yeast, the transformants showed increased tolerance to osmotic, salt, cold and heat stresses. When the tobacco BY2 cells were transformed with At-DBF2, they showed increased tolerance to salt, osmotic, cold and heat stresses, but not to oxidative stress. In addition, the authors showed that the expression of the At-DBF2 gene is upregulated by the same stresses for which the protein confers tolerance. These results provide evidence that At-DBF2 is involved positively in a general osmotic stress response pathway. Thus,

BNSDOCID: <XP_____1062997A__I_>

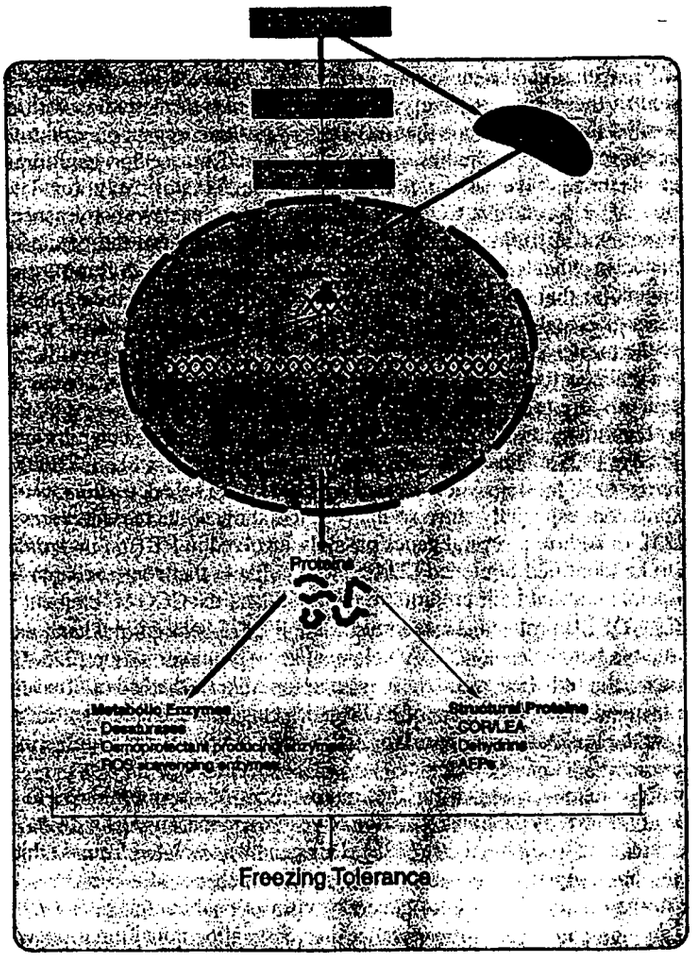


Fig. 1. Master switch hypothesis: Hypothetical model describing how low temperature in association with light regulates the expression of cold responsive genes. Low temperature is first perceived by putative membrane associated receptor(s) and/or chloroplast redox sensing mechanisms, and then transmitted to the nucleus. This triggers the global regulator (master switch). These transcription factors activate the promoters (red helix) of cold-regulated genes (blue helix). The transcribed mRNAs are translated to different proteins that contribute to the development of freezing tolerance. Illustrated by Allan Sachan.

BNSDOCID: <XP_____1062997A__I_>

transformation of crops with At-DBF2 could provide the benefit of improving the tolerance to many stresses, including the major causes of crop losses like drought, cold and salinity.

An important factor that must be considered when planning the genetic engineering of plants is the selection of an appropriate promoter to drive the expression of the transgene. Very often, plants modified to overexpress genes of interest have been transformed with constructs bearing the gene under the control of a constitutive promoter. The most widely used promoters are those of the maize ubiquitin gene and of the CaMV35S gene, for monocots and dicots respectively. However, the overexpression of a gene resulting from the use of a constitutive promoter is not necessarily an objective that is desirable in all cases. Indeed, few studies have focused on the physiological consequences related to the constitutive expression of genes that are normally inducible, and thus expressed only when needed. In some cases, chimeric promoter regions have been constructed to combine specific characteristics for expression. Zhu et al. [80] fused an ABA-responsive element from the barley *Hva22* gene, a rice actin 1 minimal promoter, and an *Hva22* intron to obtain a stress-inducible promoter. This element was used to increase the level of Δ^1-pyrroline-5-carboxylate synthetase in transgenic rice plants, which led to a drought and salt-inducible increase of the proline content and a concomitant increase in tolerance to those stresses. The activity of the promoter at LT was not reported. Another example of a chimeric promoter is discussed below.

Detrimental effects of constitutive expression of normally tightly regulated genes can occur. Plants transformed to express the DREB1A gene constitutively (from the CaMV35S promoter) showed better FT, but also severe growth retardation [219]. This detrimental effect was not reported for the homologous CBF1 overexpression [218]. When the expression of DREB1A was regulated by the cold-inducible *rd29A* promoter, plants achieved an even higher FT but most importantly, minimal effects on growth were noted [219]. These results provide convincing evidence that inducible expression may represent the best strategy to be used for crop improvement. In this respect, the promoter of the cold-inducible *wcs120* gene from wheat should also prove to be a useful tool for regulation of transgene expression [227].

Another factor to consider is the species specificity of the promoter used. Towards this end, some studies have focused on the identification of promoters that are more efficient in different plants. For example, several constructs bearing, in different combinations, fragments of the CaMV35S promoter, an intron of the bean phaseolin gene, the Ω sequence of TMV, and termination sequences from the CaMV 35S or nopaline synthase genes have been tested in rice (monocot) and tobacco (dicot) [228]. It was shown that the most efficient constructs for rice were not the same as for tobacco, suggesting differences in the specificity of gene expression between monocot and dicot plants. Schenk et al. [229] reported on the activity of a promoter from the sugarcane badnavirus in transient and transgenic expression. They showed that the promoter was active in several monocot and dicot species, to a level similar to that of the maize ubiquitin and CaMV 35S promoters. The viral promoter

BNSDOCID: <XP_____1062997A_1_>

conferred constitutive expression of the transgene. It could thus be a useful tool for the high level expression of foreign genes in both monocot and dicot transgenic plants. To our knowledge, only one study reported on the activity of a cold-responsive promoter in different species [227]. The *wcs120* promoter was found to be active and, most of all, cold inducible in both monocot and dicot species in *in vitro* transient assays, whether they are cold tolerant or not. This characteristic makes it a very important tool to exploit for crop improvement of cold tolerance. The use of a broad range, cold-inducible promoter such as that of the *wcs120* gene would allow the expression of genes only when the plant is under cold stress conditions. It would also alleviate the need to develop different constructs for different species. However, the effectiveness of this promoter in different transgenic plants remains to be evaluated.

Molecular markers

Another important application of the knowledge gained from basic studies of COR genes is the development of cold tolerance markers. The expression of several genes such as *wcs120*, *wcor 410* and *wcs 719* was found to be highly correlated with the capacity of plants and tissues to develop FT [47,230,231]. Using a number of wheat and rye cultivars differing in their acclimation capacity, it was found that the accumulation of these proteins correlates well with the capacity of each cultivar to develop FT. In addition, a near perfect correlation was also observed between the accumulation of the WCS120 protein and the increase in LT_{50} of both winter and spring wheat. These data support the concept that the anti-WCOR410, WCS120 and WCS719 antibodies may represent useful tools for breeders to select for FT in cereals. In a similar study, Lim et al. [232] reported that a 25-kDa member of LEA group II proteins could serve as a genetic marker to distinguish between cold-hardy and less hardy genotypes of the woody plant *Rhododendron* [232]. The relative level of this protein within a genotype can also serve as a physiological indicator of FT status under a range of phenological (acclimation) or developmental (age) conditions. The use of molecular markers is easy, rapid, inexpensive and non-destructive compared to invasive, time consuming methods currently used to evaluate the LT_{50}. These antibodies are currently used to assess the segregation of FT in different crosses and to study the inheritance of their respected genes. The use of these molecular markers was instrumental in mapping and regulation studies. Using a combination of Southern and western analyses on the ditelocentric series of Chinese Spring wheat, in which one homologous pair of chromosome arms is missing in each line, several COR genes were mapped to specific chromosome arms [233]. In addition, the chromosome that regulates cold tolerance and gene expression was also identified using these molecular markers. These studies provided the basis of the master switch hypothesis discussed in the previous section.

BNSDOCID: <XP_____1062997A__I_>

Potential applications to food products

Freezing is a widely used technique for food preservation during storage and transport. It reduces spoilage by inhibiting microbial growth and helps in extending shelf life of a wide variety of foods. However, there are many examples of food stuffs that cannot be frozen without unacceptable changes in quality, particularly the high water containing tissues, such as fruits, vegetables, fish and meats [24]. It is known that extracellular and intracellular ice formation results in undesirable changes in taste and texture due to cellular dehydration, concentration of solutes, and mechanical factors that can collectively lead to destruction of the plasma membrane [234,235]. Ice formation in frozen foods is influenced by both rate of freezing and storage temperature. In foods that are frozen slowly, ice crystals are large and randomly distributed and ice fills the extracellular space causing dehydration of the cell and forcing cells apart. When plant tissues are frozen rapidly, water does not translocate across the plasma membrane and small uniformly distributed ice crystals are formed within the cell. Although foods are usually flash frozen to produce small ice crystals, these ice crystals may grow larger over time through a process known as recrystallization. Recrystallization occurs in frozen foods because temperature gradients form within the products during freezing or thawing, or because temperatures fluctuate during storage or when products are in transit. For example, home frost-free freezer temperatures may rise to near 0 C daily during defrost cycles, resulting in extensive recrystallisation and limiting the length of time that foods can be stored without a dramatic loss of quality. These freeze/thaw cycles cause membrane damage and produce a loss in the liquid-retaining properties of cells, thereby causing meat, vegetable or fruit tissues to lack turgor upon thawing. Furthermore, cellulose dehydration in fruits and vegetables during freezing may account for changes in textural qualities [234]. These changes are usually associated with the degradation of chlorophylls and the oxidation of other pigments that reduce the quality and the nutritional value of thawed fruits and vegetables. Studies of frozen beef have shown that recrystallization increases with higher storage temperature and damages cellular membranes, causing greater protein denaturation and exudation from the tissue during thawing [236]. The recrystallization of ice is also a problem in foods that are eaten while frozen. For example, ice cream stored for long periods develops a gritty texture due to the growth of individual ice crystals. Low-fat frozen dessert formulations often contain more water, thus increasing problems associated with ice recrystallization.

Presently, cryoprotectants such as glycerol, sugars and dimethyl sulfoxide (DMSO) are used to control ice crystal formation in biological tissues. However, these compounds function colligatively and must be present at relatively high concentrations (usually >0.5M) to be effective. At these concentrations, cryoprotectants have adverse effects on the taste, texture and even toxicity of frozen foods. For example, the addition of sugars may reduce the direct injury to fruit caused by ice formation, but these sugars may also prevent reabsorption of water during thawing and actually increase the loss of water and nutrients from the thawed

BNSDOCID: <XP_____1062997A__I_>

fruit. It is also possible to decrease recrystallization using high molecular weight polymers that act as stabilizers [237–239]. For example, gelatin at concentrations up to 0.5 mg mL^{-1} reduces the recrystallization of ice during freezing and thawing [240]. On the other hand, polymers such as xanthan and guar gums may alter crystal growth indirectly by increasing solution viscosity and lowering the nucleation temperature [239]. However, these polymers are also used at relatively high concentrations (0.5% w/w), which may affect the consistency of the final product and prohibit their use as additives in all types of frozen foods.

Antifreeze proteins

To find an alternative to the currently available cryoprotectants, scientists have long considered using substances made by organisms that survive freezing conditions. The first substance, with properties to protect an organism from freezing, was discovered in fish of the southern polar ocean [241]. The characterization of these substances led to the identification of AFPs. Recently, proteins with similar functions were found in cold tolerant plants and insects [19,22]. AFPs are distinct from other modifiers of ice crystal growth known to date because they inhibit the growth of ice noncolligatively, by adsorbing onto the crystal surface and inhibiting the further addition of water molecules [15]. AFPs are effective in modifying crystal growth at concentrations 200 to 500 times lower than solutes that act colligatively [242]. AFPs also inhibit the recrystallization of ice at concentrations as low as 25 µg of protein per liter [243]. At these very low concentrations, AFPs are unlikely to directly affect the taste or consistency of frozen foods. Thus, AFPs are attracting the attention of food technologists interested in controlling the way ice crystals grow in frozen foods. They are considered as prime candidates for food additives and even for incorporation into the genomes of the raw-commodity [244,245].

The use of AFPs in frozen foods has been studied previously using type I AFPs and AFGPs (antifreeze glycoproteins) isolated from fish caught in polar oceans. In studies of frozen meat, Payne et al [246] demonstrated that beef and lamb soaked in solutions containing 0.1 mg AFP or AFGP ml^{-1} for 4.5 to 6 hr, then frozen to -20 C for 2 to 3 days, had smaller ice crystals than controls when the meat was examined by light and scanning electron microscopy. Moreover, meat treated with AFGPs before freezing and then stored frozen for 8 weeks also exhibited less drip with no detrimental effect on flavour, texture, tenderness, juiciness, and overall quality [247]. Likewise, the flesh of fishes naturally containing AFP exhibits significantly less drip and is juicier in sensory trials compared with flesh of fish that do not produce AFPs [248]. In studies involving frozen desserts that are eaten while frozen, synthetic AFPs similar to type I AFPs from fish have been shown to inhibit ice recrystallization in popsicles and ice cream [249]. The presence of AFPs in these products presumably inhibits ice recrystallization and preserves the smooth, creamy texture of a high quality product.

Although preliminary studies provided good evidence that fish AFPs may be useful additives to frozen foods, the widespread use of these proteins in frozen

3NSDOCID: <XP_____1062997A__1_>

foods has been limited by both ethical and marketing considerations. First of all, the production of transgenic plants harbouring genes encoding fish AFPs was considered to be unethical because vegetarians may not wish to eat plant products containing fish proteins. Secondly, the addition of fish AFPs to frozen foods such as ice cream was shown to be unacceptable to consumers, who think the food may taste "fishy". Rather than abandoning the idea of using AFPs in frozen foods, food producers have sought an alternate source of the proteins. The recent discovery of AFPs in plants provided this alternative, and renewed the interest of using AFPs to reduce freezing damages and to improve the overall quality of frozen plant products. The most studied AFPs from plants are those from rye [19]. These AFPs accumulate only during cold acclimation of cereals and were found to be similar to pathogenesis related (PR) proteins such as endoglucanases, endochitinases and thaumatin. The dual activities of these proteins could serve two functions as additives in frozen foods by inhibiting the recrystallization of ice and by reducing microbial activity within the food products [30]. This will undoubtedly have a tremendous impact on the quality and shelf life of both frozen foods and fresh produce that needs to be stored at low positive temperatures. The use of plant AFPs as food additives is very attractive due to its potentially easier acceptance by consumers, its relatively low production cost and its presence in renewable resources. There is a strong possibility that we will see plant AFPs as a new class of frozen food additives in the near future.

Dehydrins

The other group of plant proteins that may become the target of new biotechnological applications in the food and cosmetic industries is the dehydrin group. This protein family is synonymous with LEA group II proteins or the D-11 family. Dehydrins are the most conspicuous of soluble proteins induced by dehydrative stresses. A literature survey revealed that they have been detected in well over 100 independent studies of drought stress, cold acclimation, salinity stress, and embryo development [45,250]. They have a broad size range, show no compelling similarity in amino acid sequence to any known enzyme or protein of known function, and accumulate to levels in excess of 1% of total soluble proteins. Dehydrins can be classified by their content in different amino acid segments named K, S and Y, which are usually repeated several times. The K segments are predicted to form amphipathic α-helices. Dehydrins are also peculiar by their high hydrophilicity and boiling solubility which renders their purification very easy. Several lines of evidence are consistent with a role of dehydrins in membrane interactions including immunolocalization, plasma membrane association and adoption of an α-helical structure by several cereal dehydrins in the presence of SDS [47,251]. It was also shown that dehydrins could protect enzymes from freezing damage *in vitro*. For example, the WCS120 protein at 0.2 μM has the same efficiency as 250 mM sucrose in *in vitro* cryoprotection assays of lactate dehydrogenase, meaning that WCS120 is over a million-fold more efficient than sucrose on a molar basis

BNSDOCID: <XP_____1062997A__I_>

[56]. Similar results have been obtained with other dehydrins [252,253]. It was proposed that dehydrins, by acting synergistically with compatible solutes, may replace water for the "solvation" of membranes and proteins, thus alleviating the dehydration-induced membrane destabilization and protein coagulation [45]. Second, these proteins may reduce the incidence of events leading to the formation of non-bilayer structures between membranes. This can be achieved by the proteins abundant association with different lipid species of membranes, thus causing a steric hindrance to the interaction between membrane bilayers or preventing lipid demixing that occurs during dehydration. Finally, dehydrins may also function in counteracting the irreversibly damaging effects of the increasing ionic concentration during dehydration. High ionic concentration decreases interbilayer repulsion due to charge screening and leads to the interaction between bilayers and the induction of non-bilayer structures.

As plants are the only multicellular organisms able to tolerate freezing and dehydration stresses (plant seeds lose 80% of their water content during maturation and remain viable) and because these processes are highly correlated with dehydrin accumulation, these specific plant proteins may find very innovative applications in the future. For example, as a complement to AFPs, dehydrins may further protect frozen foods against storage shrink and textural degradation. It can also protect membranes against water loss and ion leakage. As another example, dehydrins may be used in ointment against frostbite because ointments have to date been shown to give at best only a negligible protection against frostbite of the face and ears [254,255]. Frostbite involves exactly the same phenomenon as plant cell freezing: ice formation, increasing osmolality of the extracellular fluid and intracellular dehydration [256,257].

Applications in medicine

Freeze-associated proteins have also been considered for potential applications in medicine. So far, AFPs have been used in studies on cryodestruction of undesirable cells such as cancer cells [258] and in the protection of vital molecules, tissues and organs [259].

Cryodestruction of tumour cells

In the case of cryosurgery, cooled probes inserted into tumorogenic tissue with intra-operative monitoring produce local freezing to destroy the unwanted tissues without affecting the surrounding healthy ones. However, recent ultrasound studies have shown that the volume of tissue which later became necrotic is only 70 to 80% of the size of the frozen volume [260,261]. This can lead to complications such as recurrence of cancer in these peripherical regions. The use of concentrated AFPs was found to enhance the destruction of cells or tissues frozen *in vitro* under these fast and highly hypothermic conditions [262,263]. These encouraging results led Pham et al. [264] to test AFPs *in vivo*. In a study with nude mice containing

BNSDOCID: <XP_____1062997A__I_>

subcutaneous metastatic prostate tumours, they showed that the preoperative injection of a 10 mg/ml AFP solution into the tumour prior to freezing enhanced the destruction of cancerous cells under thermal conditions which normally yield cell survival. At concentrations higher than about 5 mg/ml of AFP, independently of the thermal conditions, the ice crystals that form develop a micronsize highly destructive spicular shape. These findings suggest that the adjunctive use of AFPs in cryosurgery could reduce cancerous cell survival in the periphery of the cryogenic lesions and thus prevent or reduce the complications from undesirable tissues that survive freezing. It is important, however, to conduct clinical trials to evaluate the potential and efficiency of this technique in humans.

Cryoprotection

The ability to cryopreserve human organs and tissues would have an important impact in medicine, particularly in organ transplantation and reproduction [259]. Up till now, chemical substances like dimethyl sulfoxide (DMSO), glycerol and sucrose are used as cryoprotective agents. Because these substances are used at high concentrations that may cause toxicity and side effects, AFPs, especially from plants, could be an interesting alternative. These proteins are natural products and have proven effects in organisms that survive freezing. Prospects for cryopreservation are numerous, ranging from oocytes and platelets to whole organs and bodies.

Assisted reproduction relies on embryo cryopreservation which involves many ethical considerations. Oocyte cryopreservation may provide a good alternative and allow the banking of supernumerary oocytes following superovulation techniques [265,266]. Although promising results have been obtained with cryopreserved oocytes [267,268], variability has precluded routine application of the technology in clinical practice [269]. Exposure to cryoprotective agents and LT has been reported to disrupt the oocyte cytoskeleton, and harden zona pellucida, rendering the oocyte impenetrable to sperm. To circumvent this problem, the potential of using AFPs to protect oocytes during the freezing process has been evaluated. Fish AFGPs have been shown to protect immature pig oocytes during hypothermia [270], and cryopreservation [271–274]. Further studies will indicate if they can be used routinely for the cryopreservation of oocytes.

Another potential application of AFPs is the cryopreservation of human platelets. Stored platelets can be kept only for five days at room temperature before microbial contamination and progressive aging induced activation [275]. Cooling below 5 C increases ion leakage and platelet activation by calcium entry. A preliminary study has shown that AFGP can prevent ion leakage and thereby improve the survival of platelets at temperatures as low as 4 C [275]. Certainly, more research is required to determine the most appropriate AFP that is needed to increase the membrane stability of the platelets and hence storage time. The recent identification of many plant stress associated proteins may provide additional tools that can be used to tackle this problem. The development of an efficient technique to prolong the

BNSDOCID: <XP_____1062997A_ I >

storage of human platelets would have enormous health bënefits. Human platelets are essential for blood clotting, and transfusions of platelets are routinely given to burn victims, chemotherapy patients and patients afflicted with thrombocytopenia, a side effect of many invasive surgeries such as cardiopulmonary bypass or bone marrow transplantation [275].

Cryopreservation of complex tissues and organs is much more difficult than single cell preservation. These consist of different cell types, whose membranes are in physical contact. Since the cooling rate and the lowest permissive temperature seem to be specific for each cell type, the challenge to cryopreserve whole tissues and organs is far from being resolved [276,277]. It may seem obvious that the solution is to avoid or prevent ice formation and may be achieved by vitrification [111]. However, the highly concentrated vitrification solution that is needed has undesirable toxic effects. To circumvent this problem, Rubinsky et al. [278] added AFPs to the vitrification solution used to preserve mammalian livers. After six hours at -3 C, the liver was still able of bile production in a culture dish when thawed to 37 C. Microscopic analysis of tissue slices also showed that the damage to the structural integrity of the liver was relatively small and that the hepatocytes appeared mostly intact. These results led the authors to conclude that AFPs have a protective effect on the whole liver during freezing.

Although these results are encouraging for the future of cryobiology, further research must be pursued to study also the cryopreservation of other organs. For example, a heart stays viable for only four to eight hours after it is removed from its donor [279]. Longer preservation times may save thousands of lives each week. Studies with fish AFP have shown that they can sometimes exacerbate the damage during freezing of rat cardiac explants [280] and cardiomyocytes [281]. If the negative effect of these AFPs is of structural nature, the problem may be circumvented by using plant AFPs or other freezing associated proteins like dehydrins which are completely different in structure and may have a complementary function [19]. In a near future, the characterization of cold and freezing associated proteins from organisms that survive freezing temperatures, particularly plants that survive temperatures as low as -50 C, may lead to novel applications. These include near freezing point long-term surgeries, frostbite protection, organ and platelet preservation and whole body cryostorage for deep space travel.

Acknowledgments

Research in our laboratory has been supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Fonds pour la Formation de Chercheurs et l'Aide la Recherche of the Ministère de l' ducation du Qu bec.

BNSDOCID: <XP_____1062997A__I_>

References

1. Burke MJ, Gusta LV, Quamme HA, Weiser CJ, Li PH. Freezing and injury in plants. Annu Rev Plant Physiol 1976;27:507–528.
2. Storey KB, Storey JM. Natural freeze tolerance in ectothermic vertebrates. Annu Rev Physiol 1992;54:619–637.
3. Panoff J-M, Thammavongs B, Gu guen M, Boutibonnes P. Cold stress responses in mesophilic bacteria. Cryobiology 1998;36:75–83.
4. Kempf B, Bremer E. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Arch Microbiol 1998;170:319–330.
5. Russell NJ. Molecular adaptations in psychrophilic bacteria: potential for biotechnological applications. Adv Biochem Eng Biotechnol 1998;61:1–21.
6. Guy CL. Cold acclimation and freezing stress tolerance: role of protein metabolism. Annu Rev Plant Physiol Plant Mol Biol 1990;41:187–223.
7. Thomashow MF. Molecular genetics of cold acclimation in higher plants. Adv Genet 1990;28:99–131.
8. Thomashow MF. Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 1999;50:571–599.
9. Fowler DB, Limin AE, Ritchie JT. Low-temperature tolerance in cereals: models and genetic interpretation. Crop Sci 1999;39:626–633.
10. Huner NPA, quist G, Sarhan F. Energy balance and acclimation to light and cold. Trends Plant Sci 1998;3:224–230.
11. Varapetian BB, Jackson MB. Plant adaptations to anaerobic stress. Ann Bot 1997;79:3–20.
12. Holmberg N, B low L. Improving stress tolerance in plants by gene transfer. Trends Plant Sci 1998;3:61–66.
13. DeVries AL, Komatsu SK, Feeney RE. Chemical and physical properties of freezing point-depressing glycoproteins from antarctic fishes. J Biol Chem 1970;245:2901–2908.
14. DeVries AL. Role of glycopeptides and peptides in inhibition of crystallization of water in polar fishes. Philos T Roy Soc B 1984;304:575–588.
15. DeVries AL. Antifreeze glycopeptides and peptides : interactions with ice and water. Methods Enzymol 1986;127:293–303.
16. Knight CA, DeVries AL, Oolman LD. Fish antifreeze protein and the freezing and recrystallization of ice. Nature 1984;308:295–296.
17. Knight CA, Duman JG. Inhibition of recrystallization of ice by insect thermal hysteresis proteins: A possible cryoprotective role. Cryobiology 1986;23:256–262.
18. Hays LM, Feency RE, Crowe LM, Crowe JH, Oliver AE. Antifreeze glycoproteins inhibit leakage from liposomes during thermotropic phase transitions. Proc Natl Acad Sci USA 1996;93:6835–6840.
19. Griffith M, Antikainen M, Hon W-C, Pihakaski-Maunsbach K, Yu X-M, Chun JU, Yang DSC. Antifreeze proteins in winter rye. Physiol Plant 1997;100:327–332.
20. Ewart KV, Lin Q, Hew CL. Structure, function and evolution of antifreeze proteins. Cell Mol Life Sci 1999;55:271–283.
21. Worrall D, Elias L, Ashford D, Smallwood M, Sidebottom C, Lillford P, Telford J, Holt C, Bowles D. A carrot leucine-rich-repeat protein that inhibits ice recrystallization. Science 1998;282:115–117.
22. Tyshenko MG, Doucet D, Davies PL, Walker VK. The antifreeze potential of the spruce budworm thermal hysteresis protein. Nat Biotechnol 1997;15:887–890.
23. Cutler AJ, Saleem M, Kendall E, Gusta LV, Georges F, Fletcher GL. Winter flounder antifreeze protein improves the cold hardiness of plant tissues. J Plant Physiol 1989;135:351–354.
24. Hightower R, Baden C, Penzes E, Lund P, Dunsmuir P. Expression of antifreeze proteins in transgenic plants. Plant Mol Biol 1991;17:1013–1021.
25. McKown RL, Warren GJ. Enhanced survival of yeast expressing an antifreeze gene analogue

BNSDOCID: <XP_____1062997A_I_>

after freezing. Cryobiology 1991;28:474–482.
26. Kenward KD, Altschuler M, Hildebrand D, Davies PL. Accumulation of type I fish antifreeze protein in transgenic tobacco is cold-specific. Plant Mol Biol 1993;23:377–385.
27. Wallis JG, Wang H, Guerra DJ. Expression of a synthetic antifreeze protein in potato reduces electrolyte release at freezing temperatures. Plant Mol Biol 1997;35:323–330.
28. Kenward KD, Brandle J, McPherson J, Davies PL. Type II fish antifreeze protein accumulation in transgenic tobacco does not confer frost resistance. Transgen Res 1999;8:105–117.
29. Meyer K, Keil M, Naldrett MJ. A leucine-rich repeat protein of carrot that exhibits antifreeze activity. FEBS Lett 1999;447:171–178.
30. Griffith M, Ewart KV. Antifreeze proteins and their potential use in frozen foods. Biotech Adv 1995;13:375–402.
31. Gusta LV, Fowler DB. Factors affecting the cold survival of winter cereals. Can J Plant Sci 1977;57:213–219.
32. Siminovitch D, Scarth GW. A study of the mechanism of frost injury to plants. Can J Res 1938;C16:467–481.
33. Levitt J. Responses of plants to environmental stresses. Vol 1: Chilling, Freezing, and High temperature stresses. New York: Academic Press, 1980;79–96.
34. Lindow SE. Competitive exclusion of epiphytic bacteria by ice *Pseudomonas syringae* mutants. Appl Environ Microbiol 1987;53:2520–2527.
35. Warren GJ. Identification and analysis of *ina* genes and proteins. Lee Jr. RE, Warren GJ, Gusta LV (ed) Biological ice nucleation and its applications. St Paul, Minn: APS Press, 1995;85–99.
36. Lee RE Jr, Costanzo JP. Biological ice nucleation and ice distribution in cold-hardy ectothermic animals. Annu Rev Physiol 1998;60:55–72.
37. Baertlein DA, Lindow SE, Panopoulos NJ, Lee SP, Mindrinos MN, Chen THH. Expression of a bacterial ice nucleation gene in plants. Plant Physiol 1992;100:1730–1736.
38. Van Zee K, Baertlein DA, Lindow SE, Panopoulos N, Chen THH. Cold requirement for maximal activity of the bacterial ice nucleation protein INAZ in transgenic plants. Plant Mol Biol 1996;30:207–211.
39. Rogers JS, Stall RE, Burke MJ. Low-temperature conditioning of the ice nucleation active bacterium, *Erwinia herbicola*. Cryobiology 1987;24:270–279.
40. Brush RA, Griffith M, Mlynarz A. Characterization and quantification of intrinsic ice nucleators in winter rye [*Secale cereale*] leaves. Plant Physiol 1994;104:725–735.
41. Steponkus PL. Role of the plasma membrane in freezing injury and cold acclimation. Annu Rev Plant Physiol 1984;35:543–584.
42. Steponkus PL, Uemura M, Webb MS. Membrane destabilization during freeze-induced dehydration. In: Close TL, Bray EA. (eds) Plant responses to cellular dehydration during environmental stress. Rockville, MD: American Society of Plant Physiologists, 1993;37–47.
43. Dure III L, Crouch M, Harada J, Ho T-HD, Mundy J, Quatrano RS, Thomas T, Sung ZR. Common amino acid sequence domains among the LEA proteins of higher plants. Plant Mol Biol 1989;12:475–486.
44. Dure III L. Structural motifs in LEA proteins of higher plants. In: Close TJ, Bray EA (eds) Responses of plants to cellular dehydration during environmental stress. Rockville, MD: American Society of Plant Physiologists, 1993;91–103.
45. Close TJ. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol Plant 1996;97:795–803.
46. Sarhan F, Ouellet F, Vazquez-Tello A. The wheat *wcs 120* gene family. A useful model to understand the molecular genetics of freezing tolerance in cereals. Physiol Plant 1997;101:439–445.
47. Danyluk J, Perron A, Houde M, Limin A, Fowler B, Benhamou N, Sarhan F. Accumulation of an acidic dehydrin in the vicinity of the plasma membrane during cold acclimation of wheat. Plant Cell 1998;10:623–638.
48. Iturriaga G, Schneider K, Salamini F, Bartels D. Expression of desiccation-related proteins from the resurrection plant *Craterostigma plantagineum* in transgenic tobacco. Plant Mol Biol 1992;20:555–558.

BNSDOCID: <XP._____1062997A__I_>

49. Kaye C, Neven L, Hofig A, Li Q-B, Haskell D, Guy C. Characterization of a gene for spinach CAP160 and expression of two spinach cold-acclimation proteins in tobacco. Plant Physiol 1998;116:1367–1377.
50. Xu D, Duan X, Wang B, Hong B, Ho T-HD, Wu R. Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. Plant Physiol 1996;110:249–257.
51. Swire-Clark GA, Marcotte Jr WR. The wheat LEA protein Em functions as an osmoprotective molecule in *Saccharomyces cerevisiae*. Plant Mol Biol 1999;39:117–128.
52. Imai R, Chang L, Ohta A, Bray EA, Takagi M. A lea-class gene of tomato confers salt and freezing tolerance when expressed in *Saccharomyces cerevisiae*. Gene 1996;170:243–248.
53. Artus NN, Uemura M, Steponkus PL, Gilmour SJ, Lin C, Thomashow MF. Constitutive expression of the cold-regulated *Arabidopsis thaliana COR15a* gene affects both chloroplast and protoplast freezing tolerance. Proc Natl Acad Sci USA 1996;93:13404–13409.
54. Steponkus PL, Uemura M, Joseph RA, Gilmour SJ, Thomashow MF. Mode of action of the *COR15a* gene on the freezing tolerance of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 1998;95:14570–14575.
55. Crosatti C, Polverino de Laureto P, Bassi R, Cattivelli L. The interaction between cold and light controls the expression of the cold-regulated barley gene *cor14b* and the accumulation of the corresponding protein. Plant Physiol 1999;119:671–680.
56. Houde M, Daniel C, Lachapelle M, Allard F, Laliberti S, Sarhan F. Immunolocalization of freezing-tolerance-associated proteins in the cytoplasm and nucleoplasm of wheat crown tissues. Plant J 1995;8:583–593.
57. Lin C, Thomashow MF. DNA sequence analysis of a complementary DNA for cold-regulated *Arabidopsis* gene *cor15* and characterization of the COR15 polypeptide. Plant Physiol 1992;99:519–525.
58. Bartels D, Nelson D. Approaches to improve stress tolerance using molecular genetics. Plant Cell Environ 1994;17:659–667.
59. Bohnert HJ, Shen B. Transformation and compatible solutes. Sci Hort 1999;78:237–260.
60. Nishida I, Murata N. Chilling sensitivity in plants and cyanobacteria: the crucial contribution of membrane lipids. Annu Rev Plant Physiol Plant Mol Biol 1996;47:541–568.
61. Uemura M, Steponkus PL. A contrast of the plasma membrane lipid composition of oat and rye leaves in relation to freezing tolerance. Plant Physiol 1994;104:479–496.
62. Foyer CH, Descourvières P, Kunert KJ. Protection against oxygen radicals: an important defence mechanism studied in transgenic plants. Plant Cell Environ 1994;17:507–523.
63. Leegood RC. Effects of temperature on photosynthesis and photorespiration. In: Smirnoff N (ed) Environment and Plant Metabolism: flexibility and acclimation. Oxford: BIOS Scientific Publishing, 1995:45–62.
64. McNeil SD, Nuccio ML, Hanson AD. Betaines and related osmoprotectants. Targets for metabolic engineering of stress resistance. Plant Physiol 1999;120:945–949.
65. Bohnert HJ, Jensen RG. Strategies for engineering water-stress tolerance in plants. Trends Biotechnol 1996;14:89–97.
66. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. Living with water stress: evolution of osmolyte systems. Science 1982;217:1214–1222.
67. Yancey PH. Compatible and counteracting solutes. In: Strange K (ed) Cellular and Molecular Physiology of Cell Volume Regulation. Boca Raton, FL: CRC Press, 1994:81–109.
68. Naidu BP, Paleg LG, Aspinall D, Jennings AC, Jones GP. Amino acid and glycine betaine accumulation in cold-stressed wheat seedlings. Phytochem 1991;30:407–409.
69. Murelli C, Rizza F, Marinone A, Dulio A, Terzi V, Cattivelli L. Metabolic changes associated with cold-acclimation in contrasting cultivars of barley. Physiol Plant 1995;94:87–93.
70. Bertrand A., Paquin R. Influence de la temp rature d'endurcissement sur la tol rance au gel de la luzerne et sa teneur en sucres, amidon et proline. Can J Plant Sci 1991;71: 737–747.
71. Wanner LA, Junttila O. Cold-induced freezing tolerance in Arabidopsis. Plant Physiol 1999;120:391–400.

BNSDOCID: <XP_____1062997A_I_>

72. Delauney AJ, Verma DPS. Proline biosynthesis and osmoregulation in plants. Plant J 1993;4:215–223.
73. Hare, PD, Cress WA, Van Staden J. Proline synthesis and degradation: a model system for elucidating stress-related signal transduction. J Exp Bot 1999;50:413–434.
74. Kumar A, Minocha SC. Transgenic manipulation of polyamine metabolism. In: Lindsey K (ed) Transgenic Plant Research. Amsterdam: Harwood, 1998;187–199.
75. Bouchereau A, Aziz A, Larher F, Martin-Tanguy J. Polyamines and environmental challenges: recent development. Plant Sci 1999;140:103–125.
76. Tramontano WA, Jouve D. Trigonelline accumulation in salt-stressed legumes and the role of other osmoregulators as cell cycle control agents. Phytochem 1997;44:1037–1040.
77. Nuccio ML, Rhodes D, McNeil SD, Hanson AD. Metabolic engineering of plants for osmotic stress resistance. Curr Opin Plant Biol 1999;2:128–134.
78. Nomura M, Muramoto Y, Yasuda S, Takabe T, Kishitani S. The accumulation of glycinebetaine during cold acclimation in early and late cultivars of barley. Euphytica 1995;83:247–250.
79. Kavi Kishor PB, Hong Z, Miao G-H, Hu C-AA, Verma DPS. Overexpression of Δ^1-pyrroline-5-carboxylate synthetase increases proline production and confers osmotolerance in transgenic plants. Plant Physiol 1995;108:1387–1394.
80. Zhu B, Su J, Chang M, Verma DPS, Fan Y-L, Wu R. Overexpression of a Δ^1-pyrroline-5-carboxylate synthetase gene and analysis of tolerance to water- and salt-stress in transgenic rice. Plant Sc 1998;139:41–48.
81. Rhodes D, Hanson AD. Quaternary ammonium and tertiary sulfonium compounds in higher plants. Annu Rev Plant Physiol Plant Mol Biol 1993;44:357–384.
82. Nakamura T, Yokota S, Muramoto Y, Tsutsui K, Oguri Y, Fukui K, Takabe T. Expression of a betaine aldehyde dehydrogenase gene in rice, a glycine betaine nonaccumulator, and possible localization of its protein in peroxisomes. Plant J 1997;11:1115–1120.
83. Deshnium P, Gombos Z, Nishiyama Y, Murata N. The action *In vivo* of glycine betaine in enhancement of tolerance of *Synechococcus* sp. Strain PCC 7942 to low temperature. J Bacteriol 1997;179:339–344.
84. Hayashi H, Alia, Sakamoto A, Nonaka H, Chen THH, Murata N. Enhanced germination under high-salt conditions of seeds of transgenic *Arabidopsis* with a bacterial gene (*codA*) for choline oxidase. J Plant Res 1998;111:357–362.
85. Hayashi H, Alia, Mustardy L, Deshnium P, Ida M, Murata N. Transformation of *Arabidopsis thaliana* with the *codA* gene for choline oxidase: accumulation of glycinebetaine and enhanced tolerance to salt and cold stress. Plant J 1997;12:133–142.
86. Alia, Hayashi H, Chen THH, Murata N. Transformation with a gene for choline oxidase enhances the cold tolerance of *Arabidopsis* during germination and early growth. Plant Cell Environ 1998;21:232–239.
87. Alia, Hayashi H, Sakamoto A, Murata N. Enhancement of the tolerance of *Arabidopsis* to high temperatures by genetic engineering of the synthesis of glycinebetaine. Plant J 1998;16:155–161.
88. Alia, Kondo Y, Sakamoto A, Nonaka H, Hayashi H, Saradhi PP, Chen TH, Murata N. Enhanced tolerance to light stress of transgenic *Arabidopsis* plants that express the *codA* gene for a bacterial choline oxidase. Plant Mol Biol 1999;40:279–288.
89. Sakamoto A, Alia, Murata N. Metabolic engineering of rice leading to biosynthesis of glycinebetaine and tolerance to salt and cold. Plant Mol Biol 1998;38:1011–1019.
90. Rathinasabapathi B, McCue KF, Gage DA, Hanson AD. Metabolic engineering of glycine betaine synthesis: plant betaine aldehyde dehydrogenase lacking typical transit peptides are targeted to tobacco chloroplasts where they confer betaine aldehyde resistance. Planta 1994;193:155–162.
91. Nuccio ML, Russell BL, Nolte KD, Rathinasabapathi B, Gage DA, Hanson AD. The endogenous choline supply limits glycine betaine synthesis in transgenic tobacco expressing choline monooxygenase. Plant J 1998;16:487–496.
92. Lilius G, Holmberg N, Bulow L. Enhanced NaCl stress tolerance in transgenic tobacco express-

BNSDOCID: <XP_____1062997A__I_>

ing bacterial choline dehydrogenase. Biotechnology 1996;14:177–180.
93. Holmstr m K-O, Welin B, Mandal A, Kristiansdottir I, Teeri TH, Lamark T, Strøm AR, Palva ET. Production of the *Escherichia coli* betaine-aldehyde dehydrogenase, an enzyme required for the synthesis of the osmoprotectant glycine betaine, in transgenic plants. Plant J 1994;6:749–758.
94. Nomura M, Ishitani M, Takabe T, Rai AK, Takabe T. *Synechococcus* sc. PCC 7942 transformed with *Escherichia coli bet* genes produces glycine betaine from choline and acquires resistance to salt stress. Plant Physiol 1995;107:703–708.
95. Allard F, Houde M, Kr I M, Ivanov A, Huner NPA, Sarhan F. Betaine improves freezing tolerance in wheat. Plant Cell Physiol 1998;39:1194–1202.
96. Zhao Y, Aspinall D, Paleg LG. Protection of membrane integrity in *Medicago sativa* L. by glycinebetaine against the effects of freezing. J Plant Physiol 1992;140:541–543.
97. Ko R, Tombras Smith L, Smith GM. Glycine betaine confers enhanced osmotolerance and cryotolerance on *Listeria monocytogenes*. J Bacteriol 1994;176:426–431.
98. Higgins J, Hodges NA, Olliff CJ, Phillips AJ. A comparative investigation of glycinebetaine and dimethylsulphoxide as liposome cryoprotectants. J Pharm Pharmacol 1987;39:577–582.
99. Lloyd AW, Ollif CJ, Rutt KJ. A study of modified betaines as cryoprotective additives. J Pharm Pharmacol 1994;46:704–707.
100. Bachmann M, Matile P, Keller F. Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L. Plant Physiol 1994;105:1335–1345.
101. Mitchell DE, Madore MA. Patterns of assimilate production and translocation in muskmelon [*Cucumis melo* L.]. Plant Physiol 1992;99:966–971.
102. Olien CR, Clark JL. Changes in soluble carbohydrate composition of barley, wheat, and rye during winter. Agron J 1993;85:21–29.
103. Antikainen M, Pihakaski S. Early developments in RNA, protein, and sugar levels during cold stress in winter rye [*Secale cereale*] leaves. Ann Bot 1994;74:335–341.
104. Turner LB, Pollock CJ. Changes in stolon carbohydrates during the winter in four varieties of white clover [*Trifolium repens* L.] with contrasting hardiness. Ann Bot 1998;81:97–107.
105. Castonguay Y., Nadeau P., Lechasseur P, Chouinard L. Differential accumulation of carbohydrates in alfalfa cultivars of contrasting winterhardiness. Crop Sci 1995;35:509–516.
106. Pollock CJ, Eagles CF, Sims IM. Effect of photoperiod and irradiance changes upon development of freezing tolerance and accumulation of soluble carbohydrate in seedlings of *Lolium perenne* grown at 2 C. Ann Bot 1988;62:95–100.
107. Imanishi HT, Suzuki T, Masuda K, Harada T. Accumulation of raffinose and stachyose in shoot apices of *Lonicera caerulea* L. during cold acclimation. Sci Hortic 1998;72:255–263.
108. V g jfalvi A, Kerepesi I, Galiba G, Tischner T, Sutka J. Frost hardiness depending on carbohydrate changes during cold acclimation in wheat. Plant Sci 1999;144:85–92.
109. Yoshida M, Abe J, Moriyama M, Kuwabara T. Carbohydrate levels among winter wheat cultivars varying in freezing tolerance and snow mold resistance during autumn and winter. Physiol Plant 1998;103:8–16.
110. Crowe JH, Hoekstra FA, Crowe LM. Anhydrobiosis. Annu Rev Physiol 1992;54:579–599.
111. Crowe JH, Carpenter JF, Crowe LM. The role of vitrification in anhydrobiosis. Annu Rev Physiol 1998;60:73–103.
112. Bruni F, Leopold AC. Cytoplasmic glass formation in maize embryos. Seed Sci Res 1992;2:251–253.
113. Stitt M, Sonnewald U. Regulation of metabolism in transgenic plants. Annu Rev Plant Physiol Plant Mol Biol 1995;46:341–368.
114. Stoop JMH, Williamson JD, Pharr DM. Mannitol metabolism in plants: a method for coping with stress. Trends Plant Sci 1996;1:139–144.
115. Vijn I, Smeekens S. Fructan: more than a reserve carbohydrate? Plant Physiol 1999;120:351–359.
116. Popp M, Smirnoff N. Polyol accumulation and metabolism during water deficit. In: Smirnoff N. (ed) Environment and Plant Metabolism: flexibility and acclimation. Oxford: BIOS Scientific

BNSDOCID: <XP ___ ___1062997A_I_>

Publishing, 1995;199–215.
117. Goddijn O, Smeekens S. Sensing trehalose biosynthesis in plants. Plant J 1998;14:143–146.
118. Singer MA, Lindquist S. Thermotolerance in *Saccharomyces cerevisiae*: the Yin and Yang of trehalose. Trends Biotechnol 1998;16:460–468.
119. Holmstr m K-O, M ntyl E, Welin B, Mandal A, Palva ET, Tunnela OE, Londesborough J. Drought tolerance in tobacco. Nature 1996;379:683–684.
120. Romero C, Bell s JM, Vay JL, Serrano R, Culi ez-Maci FA. Expression of the yeast *trehalose-6-phosphate synthase* gene in transgenic tobacco plants: pleiotropic phenotypes include drought tolerance. Planta 1997;201:293–297.
121. Goddijn OJM, Verwoerd TC, Voogd E, Krutwagen RWHH, de Graaf PTHM, Poels J, Van Dun K, Ponstein AS, Damm B, Pen J. Inhibition of trehalase activity enhances trehalose accumulation in transgenic plants. Plant Physiol 1997;113:181–190.
122. Pilon-Smits EAH, Terry N, Sears T, Kim H, Zayed A, Hwang S, van Dun K, Voogd E, Verwoerd TC, Krutwagen RWHH, Goddijn OJM. Trehalose-producing transgenic tobacco plants show improved growth performance under drought stress. J Plant Physiol 1998;152:525–532.
123. Shima J, Hino A, Yamada-Iyo C, Suzuki Y, Nakajima R, Watanabe H, Mori K, Takano H. Stress tolerance in doughs of *Saccharomyces cerevisiae* trehalase mutants derived from commercial Baker's yeast. Appl Environ Microbiol 1999;65:2841–2846.
124. Soto T, Fernandez J, Vicente-Soler J, Cansado J, Gacto M. Accumulation of trehalose by overexpression of *tps1*, coding for trehalose-6-phosphate synthase, causes increased resistance to multiple stresses in the fission yeast *Schizosaccharomyces pombe*. Appl Environ Microbiol 1999;65:2020–2024.
125. Hendry GAF, Wallace RK. The origin, distribution, and evolutionary significance of fructans. In: Suzuki M, Chatterton NJ (eds) Science and Technology of fructans. Boca Raton, FL: CRC Press, 1993;119–139.
126. Pollock CJ. Fructans and the metabolism of sucrose in vascular plants. New Phytol 1986;104:1–24.
127. Wiemken A, Sprenger N., Boller T. Fructan: an extension of sucrose by sucrose. In: Pontis HG, Salerno GL, Echeverria EJ (eds) Sucrose metabolism, Biochemistry, physiology and molecular biology. Rockville, MD: American Society of Plant Physiologists, 1995;178–189.
128. Ebskamp MJ, van der Meer IM, Spronk BA, Weisbeek PJ, Smeekens SC. Accumulation of fructose polymers in transgenic tobacco. Biotechnology 1994;12:272–275.
129. Pilon-Smits EAH, Ebskamp MJM, Paul MJ, Jeuken MJW, Weisbeek PJ, Smeekens SCM. Improved performance of transgenic fructan-accumulating tobacco under drought stress. Plant Physiol 1995;107:125–130.
130. Tarczynski MC, Jensen RG, Bohnert HJ. Expression of a bacterial *mtlD* gene in transgenic tobacco leads to production and accumulation of mannitol. Proc Natl Acad Sci USA 1992;89:2600–2604.
131. Tarczynski MC, Jensen RG, Bohnert HJ. Stress protection of transgenic tobacco by production of osmolyte mannitol. Science 1993;259:508–510.
132. Thomas JC, Sepahi M, Arendall B, Bohnert HJ. Enhancement of seed germination in high salinity by engineering mannitol expression in *Arabidopsis thaliana*. Plant Cell Environ 1995;18:801–806.
133. Shen B, Jensen RG, Bohnert HJ. Increased resistance to oxidative stress in transgenic plants by targeting mannitol biosynthesis to chloroplasts. Plant Physiol 1997;113:1177–1183.
134. Karakas B, Ozias-Akins P, Stushnoff C, Suefferheld M, Rieger M. Salinity and drought tolerance of mannitol-accumulating transgenic plants. Plant Cell Environ 1997;20:609–616.
135. Vernon DM, Tarczynski MC, Jensen RG, Bohnert HJ. Cyclitol production in transgenic tobacco. Plant J 1993;4:199–205.
136. Sheveleva E, Chmara W, Bohnert HJ, Jensen RG. Increased salt and drought tolerance by D-ononitol production in transgenic *Nicotiana tabacum* L. Plant Physiol 1997;115:1211-1219.
137. Smart CC, Flores S. Overexpression of D-*myo*-inositol-3-phosphate synthase leads to elevated

BNSDOCID: <XP_____1062997A__I_>

levels of inositol in *Arabidopsis*. Plant Mol Biol 1997;33:811–820.
138. Tao R, Uratsu SL, Dandekar AM. Sorbitol synthesis in transgenic tobacco with apple cDNA encoding NADP-dependent sorbitol-6-phosphate dehydrogenase. Plant Cell Physiol 1995;366:525–532.
139. Sheveleva EV, Marquez S, Chmara W, Zegeer A, Jensen RG, Bohnert HJ. Sorbitol-6-phosphate dehydrogenase expression in transgenic tobacco. High amounts of sorbitol lead to necrotic lesions. Plant Physiol 1998;117:831–839.
140. Lyons JM. Chilling injury in plants. Annu Rev Plant Physiol 1973;24:445–466.
141. Shanklin J, Cahoon EB. Desaturation and related modifications of fatty acids. Annu Rev Plant Physiol Plant Mol Biol 1998;49:611–641.
142. Napier JA, Michaelson LV, Stobart AK. Plant desaturases: harvesting the fat of the land. Curr Opin Plant Biol 1999;2:123–127.
143. Somerville C. Direct tests of the role of membrane lipid composition in low-temperature-induced photoinhibition and chilling sensitivity in plants and cyanobacteria. Proc Natl Acad Sci USA 1995;92:6215–6218.
144. Murata N, Ishizaki-Nishizawa O, Higashi S, Hayashi H, Tasaka Y, Nishida I. Genetically engineered alteration in the chilling sensitivity of plants. Nature 1992;356:710–713.
145. Yokoi S, Higashi SI, Kishitani S, Murata N, Toriyama K. Introduction of the cDNA for Arabidopsis glycerol-3-phosphate acyltransferase (GPAT) confers unsaturation of fatty acids and chilling tolerance of photosynthesis of rice. Mol Breeding 1998;4:269–275.
146. Wolter FP, Schmidt R, Heinz E. Chilling sensitivity of *Arabidopsis thaliana* with genetically engineered membrane lipids. EMBO J 1992;11:4685–4692.
147. Wu J, Browse J. Elevated levels of high-melting-point phosphatidylglycerols do not induce chilling sensitivity in an *Arabidopsis* mutant. Plant Cell 1995;7:17–27.
148. Kodoma H, Hamada T, Horiguchi G, Nishimura M, Iba K. Genetic enhancement of cold tolerance by expression of a gene for chloroplast ω-3 fatty acid desaturase in transgenic tobacco. Plant Physiol 1994;105:601–605.
149. Kodoma H, Horiguchi G, Nishiuchi T, Nishimura M, Iba K. Fatty acid desaturation during chilling acclimation is one of the factors involved in confering low-temperature tolerance to young tobacco leaves. Plant Physiol 1995;107:1177–1185.
150. Kodoma H, Akagi H, Kusumi K, Fujimura T, Iba K. Structure, chromosomal location and expression of a rice gene encoding the microsome ω-3 fatty acid desaturase. Plant Mol Biol 1997;33:493–502.
151. Hamada T, Kodoma H, Hishimura M, Iba K. Modification of fatty acid composition by over- and antisense-expression of a microsomal ω-3 fatty acid desaturase gene in transgenic tobacco. Transgenic Res 1996;5:115–121.
152. Hamada T, Kodoma H, Takeshita K, Utsumi H, Iba K. Characterization of transgenic tobacco with an increased α-linolenic acid level. Plant Physiol 1998;118:591–598.
153. Ishizaki-Nishizawa O, Fujii T, Azuma M, Sekiguchi K, Murata N, Ohtani T, Toguri T. Low-temperature resistance of higher plants is significantly enhanced by a nonspecific cyanobacterial desaturase. Nat Biotechnol 1996;14:1003–1006.
154. Lynch DV, Steponkus PL. Plasma membrane lipid alterations associated with cold acclimation of winter rye seedlings [*Secale cereale* L. cv Puma]. Plant Physiol 1987;83:761–767.
155. Palta JP, Whitaker BD, Weiss LS. Plasma membrane lipids associated with genetic variability in freezing tolerance and cold acclimation of *Solanum* species. Plant Physiol 1993;103:793–803.
156. Uemura M, Joseph RA and Steponkus PL. Cold acclimation of *Arabidopsis thaliana*: Effect on plasma membrane lipid composition and freeze-induced lesions. Plant Physiol 1995;109:15–30.
157. Steponkus PL, Uemura M, Balsamo RA, Arvinte T, Lynch DV. Transformation of the cryobehavior of rye protoplasts by modification of the plasma membrane lipid composition. Proc Natl Acad Sci USA 1988;85:9026–9030.
158. Sugarawa Y, Steponkus PL. Effect of cold acclimation and modification of the plasma membrane

BNSDOCID: <XP___1062997A__I_>

lipid composition on lamellar-to-hexagonal II phase transitions in rye protoplasts. Cryobiology 1990;27:667.
159. Quinn PJ. Engineering frost resistance in plants by genetic manipulation. In: Harwood JL (ed) Plant lipid biosynthesis: fundamentals and agricultural applications. New York: Cambridge Univ Press, 1998; 156–183.
160. Noctor G, Foyer CH. Ascorbate and glutathione: keeping active oxygen under control. Annu Rev Plant Physiol Plant Mol Biol 1998;49:249–279.
161. Foyer CH, Lopez-Delgado H, Dat JF, Scott IM. Hydrogen peroxide- and glutathione-associated mechanisms of acclimatory stress tolerance and signalling. Physiol Plant 1997;100:241–254.
162. H rouart D, Bowler C, Willekens H, Van Camp W, Slooten L, Van Montagu M, Inz D. Genetic engineering of oxidative stress resistance in higher plants. Philos T Roy Soc B 1993;342:235–240.
163. Allen RD. Dissection of oxidative stress tolerance using transgenic plants. Plant Physiol 1995;107:1049–1054.
164. Allen RD, Webb RP, Schake SA. Use of transgenic plants to study antioxidant defenses. Free Radical Bio Med 1997;23:473–479.
165. Slooten L, Van Montagu M, Inz D. Manipulation of oxidative stress tolerance in transgenic plants. In: Lindsey K. Transgenic Plant Research. Amsterdam: Harwood, 1998;241–262.
166. Van Breusegem F, Van Montagu M, Inz D. Engineering stress tolerance in maize. Outlook Agric 1998;27:115–124.
167. Bowler C, Van Montagu M, Inz D. Superoxide dismutase and stress tolerance. Annu Rev Plant Physiol Plant Mol Biol 1992;43:83–116.
168. Aono M, Saji H, Sakamoto A, Tanaka K, Kondo N, Tanaka K. Paraquat tolerance of transgenic *Nicotiana tabacum* with enhanced activities of glutathione reductase and superoxide dismutase. Plant Cell Physiol 1995;36:1687–1691.
169. Br ggemann W, Beyel V, Brodka M, Poth H, Weil M, Stockhaus J. Antioxidants and antioxidative enzymes in wild-type and transgenic *Lycopersicon* genotypes of different chilling tolerance. Plant Sci 1999;140:145–154.
170. Arisi A-CM, Cornic G, Jouanin L, Foyer, CH. Overexpression of iron superoxide dismutase in transformed poplar modifies the regulation of photosynthesis at low CO_2 partial pressures or following exposure to the prooxidant herbicide methyl viologen. Plant Physiol 1998; 117:565–574.
171. Van Breusegem F, Slooten L, Stassart J-M, Botterman J, Moens T, Van Montagu M, Inz D. Effects of overproduction of tobacco MnSOD in maize chloroplasts on foliar tolerance to cold and oxidative stress. J Exp Bot 1999;50:71–78.
172. Van Breusegem F, Slooten L, Stassart J-M, Moens T, Botterman J, Van Montagu M, Inz D. Overproduction of *Arabidopsis thaliana* FeSOD confers oxidative stress tolerance to transgenic maize. Plant Cell Physiol 1999;40:515–523.
173. McKersie BD, Bowley SR, Jones KS. Winter survival of transgenic alfalfa overexpressing superoxide dismutase. Plant Physiol 1999;119:839–847.
174. Park J-I, Grant CM, Davies MJ, Dawes IW. The cytoplasmic Cu,Zn superoxide dismutase of *Saccharomyces cerevisiae* is required for resistance to freeze-thaw stress. J Biol Chem 1998;273:22921–22928.
175. Van Breusegem F, Kushnir S, Slooten L, Bauw G, Botterman J, Van Montagu M, Inz D. Processing of a chimeric protein in chloroplasts is different in transgenic maize and tobacco plants. Plant Mol Biol. 1998;38:491–496.
176. McKersie BD, Bowley SR, Harjanto E, Leprince O. Water-deficit tolerance and field performance of transgenic alfalfa overexpressing superoxide dismutase. Plant Physiol 1996;111:1177–1181.
177. McKersie BD, Chen Y, de Beus M, Bowley SR, Bowler C, Inz D, D'Halluin K, Botterman J. Superoxide dismutase enhances tolerance of freezing stress in transgenic alfalfa (*Medicago sativa* L.). Plant Physiol 1993;103:1155–1163.
178. Smirnoff N, Cumbes QJ. Hydroxyl radical scavenging activity of compatible solutes. Phyto-

BNSDOCID: <XP_____1062997A__I_>

chemistry. 1989;28:1057–1060.
179. Criqui MC, Jamet E, Parmentier Y, Marbach J, Durr A, Fleck J. Isolation and characterization of a plant cDNA showing homology to animal glutathione peroxidase. Plant Mol Biol 1992;18:623–627.
180. Beeor-Tzahar T, Ben-Hayyim G, Holland D, Faltin Z, Eshdat Y. A stress associated citrus protein is a distinct plant phospholipid hydroperoxide glutathione peroxidase. FEBS Lett 1995;366:151–155.
181. Roxas VP, Smith RK Jr, Allen ER, Allen RD. Overexpression of glutathione-S-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. Nat Biotechnol 1997;15:988–991.
182. Shikanai T, Takeda T, Yamauchi H, Sano S, Tomizawa K-I, Yokota A, Shigeoka S. Inhibition of ascorbate peroxidase under oxidative stress in tobacco having bacterial catalase in chloroplasts. FEBS Lett 1998;428:47–51.
183. Purvis AC. Role of the alternative oxidase in limiting superoxide production by plant mitochondria. Physiol Plant 1997;100:165–170.
184. Wagner AM, Moore AL. Structure and function of the plant alternative oxidase: its putative role in the oxygen defence mechanism. Biosci Rep 1997;17:319–333.
185. Maxwell DP, Wang Y, McIntosh L. The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells. Proc Natl Acad Sci USA 1999;96:8271–8276.
186. Kingston-Smith AH, Harbinson J, Foyer CH. Acclimation of photosysnthesis, H_2O_2 content and antioxidants in maize (zea mays) grown at sub-optimal temperatures. Plant Cell Environ 1999; 22:1071–1083.
187. Veena, Reddy VS, Sopory SK. Glyoxalase I from *Brassica juncea*: molecular cloning, regulation and its over-expression confer tolerance in transgenic tobacco under stress. Plant J 1999;17:385–395.
188. Thornalley PJ. Modification of the glyoxalase system in disease processes and prospects for therapeutic strategies. Biochem Soc Trans 1993;21:531–534.
189. Thornalley PJ. The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. Biochem J 1990;269:1–11.
190. Creissen G, Firmin J, Fryer M, Kular B, Leyland N, Reynold H, Pastori G, Wellburn F, Baker N, Wellburn A, Mullineaux P. Elevated glutathione biosynthetic capacity in the chloroplasts of transgenic tobacco plants paradoxically causes increased oxidative stress. Plant Cell 1999;11:1277–1291.
191. Pitcher LH, Brennan E, Hurley A, Dunsmuir P, Tepperman JM, Zilinskas BA. Overproduction of petunia chloroplastic Copper/Zinc superoxide dismutase does not confer ozone tolerance in transgenic tobacco. Plant Physiol 1991;97:452–455.
192. Tepperman JM, Dunsmuir P. Transformed plants with elevated levels of chloroplastic SOD are not more resistant to superoxide toxicity. Plant Mol Biol 1990;14:501–511.
193. Pitcher LH, Zilinskas BA. Overexpression of Cu/Zn SOD in the cytosol of transgenic tobacco confers partial resistance to ozone-induced foliar necrosis. Plant Physiol 1996;110:583–588.
194. Perl A, Perl-Treves R, Galili S, Aviv D, Shalgi E, Malkin S, Galun E. Enhanced oxidative-stress defense in transgenic potato expressing tomato Cu,Zn superoxide dismutases. Theor Appl Genet 1993;85:568–576.
195. Sen Gupta A, Webb RP, Holaday AS, Allen RD. Overexpression of superoxide dismutase protects plants from oxidative stress. Plant Physiol 1993;103:1067–1073.
196. Sen Gupta A, Heinen JL, Holaday AS, Burke JJ, Allen RD. Increased resistance to oxidative stress in transgenic plants that overexpress chloroplastic Cu/Zn superoxide dismutase. Proc Natl Acad Sci USA 1993;90:1629–1633.
197. Bowler C, Slooten L, Vandenbranden S, De Rycke R, Botterman J, Sybesma C, Van Montagu M, Inz D. Manganese superoxide dismutase can reduce cellular damage mediated by oxygen radicals in transgenic plants. EMBO J 1991;10:1723–1732.
198. Slooten L, Capiau K, Van Camp W, Van Montagu M, Sybesma C, Inz D. Factors affecting

BNSDOCID: <XP_____1062997A__I_>

the enhancement of oxidative stress tolerance in transgenic tobacco overexpressing manganese superoxide dismutase in the chloroplasts. Plant Physiol 1995;107:737–750.
199. Van Camp W, Willekens H, Bowler C, Van Montagu M, Inz D, Reupold-Popp P, Sandermann Jr H, Langebartels C. Elevated levels of superoxide dismutase protect transgenic plants against ozone damage. Bio/Technology 1994;12:165–168.
200. Van Camp W, Capiau K, Van Montagu M, Inz D, Slooten L. Enhancement of oxidative stress tolerance in transgenic tobacco plants overproducing Fe-superoxide dismutase in chloroplasts. Plant Physiol 1996;112:1703–1714.
201. Foyer C, Lelandais M, Galap C, Kunert KJ. Effects of elevated cytosolic glutathione reductase activity on the cellular glutathione pool and photosynthesis in leaves under normal and stress conditions. Plant Physiol 1991;97:863–872.
202. Foyer CH, Lelandais M, Jouanin L, Kunert KJ. Over-expression of enzymes of glutathione metabolism in poplar (*Populus tremula* X P. alba). Bull Soc Lux Clin Biol 1993;sp issue:229–240.
203. Foyer CH, Souriau N, Perret S, Lelandais M, Kunert K-J, Pruvost C, Jouanin L. Overexpression of glutathione reductase but not glutathione synthetase leads to increases in antioxydant capacity and resistance to photoinhibition in poplar trees. Plant Physiol 1995;109:1047–1057.
204. Aono M, Kubo A, Saji H, Natori T, Tanaka K, Kondo N. Resistance to active oxygen toxicity of transgenic *Nicotiana tabacum* that express the gene for glutathione reductase from *Escherichia coli*. Plant Cell Physiol 1991;32:691–697.
205. Aono M, Kubo A, Saji H, Tanaka K, Kondo N. Enhanced tolerance to photo-oxidative stress in transgenic *Nicotiana tabacum* with high chloroplastic glutathione reductase activity. Plant Cell Physiol 1993;34:129–135.
206. Broadbent P, Creissen GP, Kular B, Wellburn AR, Mullineaux PM. Oxidative stress responses in transgenic tobacco containing altered levels of glutathione reductase activity. Plant J 1995;8:247–255.
207. Creissen G, Reynolds H, Xue Y, Mullineaux P. Simultaneous targeting of pea glutathione reductase and of a bacterial fusion protein to chloroplasts and mitochondria in transgenic tobacco. Plant J 1995;8:167–175.
208. Strohm M, Jouanin L, Kunert KJ, Pruvost C, Polle A, Foyer CH, Rennenberg H. Regulation of glutathione synthesis in leaves of transgenic poplar (*Populus tremula* × *P. alba*) overexpressing glutathione synthetase. Plant J 1995;7:141–145.
209. Noctor G, Strohm M, Jouanin L, Kunert KJ, Foyer CH, Rennenberg H. Synthesis of glutathione in leaves of transgenic poplar (*Populus tremula* × *P. alba*) overexpressing γ-glutamylcysteine synthetase. Plant Physiol 1996;112:1071–1078.
210. Arisi A-CM, Noctor G, Foyer CH, Jouanin L. Modification of thiol contents in poplars (*Populus tremula* X *P. alba*) overexpressing enzymes involved in glutathione synthesis. Planta 1997;203:362–372.
211. Noctor G, Arisi A-CM, Jouanin L, Foyer CH. Manipulation of glutathione and amino acid biosynthesis in the chloroplast. Plant Physiol 1998;118:471–482.
212. Torsethaugen G, Pitcher LH, Zilinska BA, Pell EJ. Overproduction of ascorbate peroxidase in tobacco chloroplast does not provide protection against ozone. Plant Physiol 1997;114:529–537.
213. Ishitani M, Xiong L, Stevenson B, Zhu JK. Genetic analysis of osmotic and cold stress signal transduction in *Arabidopsis*: interactions and convergence of abscisic acid-dependent and abscisic acid-independent pathways. Plant Cell 1997;9:1935–1949.
214. Shinozaki K, Yamaguchi-Shinozaki K. Molecular responses to drought and cold stress. Curr Opin Biotech 1996;7:161–167.
215. Stockinger EJ, Gilmour SJ, Thomashow MF. *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci USA 1997;94:1035–1040.
216. Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. Two tran-

3NSDOCID: <XP_____1062997A__I_>

scriptional factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low- temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell 1998;10:1391–1406.
217. Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF. Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. Plant J 1998;16:433–442.
218. Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF. *Arabidopsis CBF1* overexpression induces *COR* genes and enhances freezing tolerance. Science 1998;280:104–106.
219. Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotechnol 1999;17:287–291.
220. Sarhan F, Danyluk J. Engineering cold-tolerant crops-throwing the master switch. Trends Plant Sci 1998;3:289–290.
221. Warren GJ. Cold stress: manipulating freezing tolerance in plants. Curr Biol 1998;8:R514–R516.
222. Stushnoff C, Fowler DB, Brule-Babel A. Breeding and selection for resistance to low temperature. In: Vose PB (ed) Plant Breeding- A contemporary basis. Oxford, UK: Pergamon Press, 1984;115–136.
223. Hughes MA, Dunn MA. The molecular biology of plant acclimation to low temperature. J Exp Bot 1996;47:291–305.
224. Thorlby G, Veale E, Butcher K, Warren G. Map position of *SFR* genes in relation to other freezing-related genes of *Arabidopsis thaliana*. Plant J 1999;17:445–452.
225. Chauvin LP, Houde M, Sarhan F. A leaf-specific gene stimulated by light during wheat acclimation to low temperature. Plant Mol Biol 1993;23:255–265.
226. Lee JH, van Montagu M, Verbruggen N. A highly conserved kinase is an essential component for stress tolerance in yeast and plant cells. Proc Natl Acad Sci USA 1999;96:5873–5877.
227. Ouellet F, Vazquez-Tello A, Sarhan F. The wheat *wcs120* promoter is cold-inducible in both monocotyledonous and dicotyledonous species. FEBS Lett 1998;423:324–328.
228. Mitsuhara I, Ugaki M, Hirochika H, Ohshima M, Murakami T, Gotoh Y, Katayose Y, Nakamura S, Honkura R, Nishimiya S, Ueno K, Mochizuki A, Tanimoto H, Tsugawa H, Otsuki Y, Ohashi Y. Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. Plant Cell Physiol 1996;37:49–59.
229. Schenk PM, Sagi L, Remans T, Dietzgen RG, Bernard MJ, Graham MW, Manners JM. A promoter from sugarcane bacilliform badnavirus drives transgene expression in banana and other monocot and dicot plants. Plant Mol Biol 1999;39:1221–1230.
230. Houde M, Dhindsa RS, Sarhan F. A molecular marker to select for freezing tolerance in Gramineae. Mol Gen Genet 1992;234:43–48.
231. Carpentier E. Accumulation et r gulation d'une prot ine homologue des facteurs d polym risant l'actine lors de l'acclimatation au froid chez le bl . M.Sc. thesis, Universit du Qu bec Montr al, 1998.
232. Lim CC, Krebs SL, Arora R. A 25-kDa dehydrin associated with genotype-and age-dependent leaf freezing-tolerance in *Rhododendron*: a genetic marker for cold hardiness? Theor Appl Genet 1999;99:910–918.
233. Limin AE, Danyluk J, Chauvin LP, Fowler DB, Sarhan F. Chromosome mapping in low-temperature induced wcs120 family genes and regulation of cold-tolerance expression in wheat. Mol Gen Genet 1997;253:720–727.
234. Reeve RM, Brown MS. Some structural and histochemical changes related to frozen fruits and vegetables. Cryobiology 1966;3:214–223.
235. Mohr WP. Freeze-thaw damage to protoplasmic structures in high moisture, edible plant tissues. J Texture Stud 1971;2:316–327.
236. Martino MN, Zaritzky NE. Ice crystal size modifications during frozen beef storage. J Food Sci

BNSDOCID: <XP._____1062997A__I_>

1988;53:1631–1637.
237. Holt C. The effect of polymers on ice crystal growth. In: Balt WB (ed) Food freezing: Today and tomorrow. London: Springer-Verlag, 1991;81–86.
238. Caldwell KB, Goff HD, Stanley DW. A low-temperature scanning electron microscopy study of ice cream. II. Influence of selected ingredients and processes. Food Struct 1992;11:11–23.
239. Goff HD, Sahagian ME. Glass transitions in aqueous carbohydrate solutions and their relevance to frozen food stability. Thermochim Acta 1996;280:449–464.
240. Smith CE, Schwartzberg HG. Ice crystal size changes during ripening in freeze concentration. Biotech Prog 1985;1:111–120.
241. Scholander PF, Flag W, Walters V, Irving. Climatic adaptation in arctic and tropical poikilotherms. Physiol Zool 1953;26:67–92.
242. Fletcher GL, Kao MH, Fourney RM. Antifreeze peptides confer freezing resistance to fish. Can J Zool 1986;64:1897–1901.
243. Griffith M, Antikainen M. Extracellular ice formation in freezing-tolerant plants. Adv Low-Temp 1997;3:107–139.
244. Feeney RE, Yeh Y. Antifreeze proteins: current status and possible food uses. Trends Food Sci Tech 1998;9:102–106.
245. Fletcher GL, Goddard SV, Wu Y. Antifreeze proteins and their genes: from basic research to business opportunity. Chemtech 1999;29:17–28.
246. Payne SR, Sandford D, Harris A, Young OA. The effects of antifreeze proteins on chilled and frozen meat. Meat Sci 1994;37:429–438.
247. Payne SR, Young OA. Effects of pre-slaughter administration of antifreeze proteins on frozen meat quality. Meat Sci 1995;41:147–155.
248. Payne SR, Wilson PW. Comparison of the freeze–thaw characteristics of Antartic Cod (*Dissostichus mawsoni*) and Black Cod (*Paranotothenia augustata*)-possible effects of antifreeze glycoproteins. J Musc Foods 1994;5:233–244.
249. Warren GJ, Mueller GM, McKown RL. Ice crystal growth suppression polypeptides and methods of preparation. US patent 5118792. 1992.
250. Choi DW, Zhu B, Close TJ. The barley (*Hordeum vulgare* L.) dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 *Dhn* genes of cv Dicktoo. Theor Appl Genet 1999;98:1234–1247.
251. Ismail AM, Hall AE, Close TJ. Purification and Partial Characterization of a Dehydrin Involved in Chilling Tolerance during Seedling Emergence of Cowpea. Plant Physiol 1999;120:237–244.
252. Kazuoka T, Oeda K. Purification and characterization of COR85-oligomeric complex from cold-acclimated spinach. Plant Cell Physiol 1994;35:601–611.
253. Wisniewski M, Webb R, Balsamo R, Close TJ, Yu X-M, Griffith M. Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: a dehydrin from peach (*Prunus persica*). Physiol Plant 1999;105:600–608.
254. Lehmuskallio E. Cold protecting ointments and frostbite. A questionnaire study of 830 conscripts in Finland. Acta Derm Venereol 1999;79:67–70.
255. Lehmuskallio E, Anttonen H. Thermophysical effects of ointments in cold: an experimental study with a skin model. Acta Derm Venereol 1999;79:33–36.
256. Bracker MD. Environmental and thermal injury. Clin Sports Med 1992; 11:419–436.
257. Berg A, Aas P, Lund T. Frostbite injuries. Tidsskr Nor Laegeforen 1999;119:382–385.
258. Gage AA, Baust J Mechanisms of tissue injury in cryosurgery. Cryobiology 1998;37:171–186.
259. Karlsson JO, Toner M. Long-term storage of tissues by cryopreservation: critical issues. Biomaterials 1996;17:243–256.
260. Lam CM, Shimi SM, Cuschieri A. Ultrasonic characterization of hepatic cryolesions. Arch Surg 1995;130:1068–1072.
261. Saliken JC, Cohen J, Miller R, Rothert M. Laboratory evaluation of ice formation around a 3-mm accuprobe. Cryobiology 1995;32:285–295.
262. Koushafar H, Rubinsky B. Effect of antifreeze proteins on frozen primary prostatic adenocarci-

BNSDOCID: <XP_____1062997A__I_>

noma cells. Urology 1997;49:421–425.
263. Koushafar H, Pham L, Lee C, Rubinski B. Chemical adjuvant cryosurgery with antifreeze proteins. J Surg Oncol 1997;66:114–121.
264. Pham L, Dahiya R, Rubinsky B. An *in vivo* study of antifreeze protein adjuvant cryosurgery. Cryobiology 1999;38:169–175.
265. Bernard A, Fuller BJ. Cryopreservation of human oocytes: A review of current problems and perspectives. Hum Reprod Update 1996;2:193–207.
266. Porcu E. Freezing of oocytes. Curr Opin Obstet Gynecol 1999;11:297–300.
267. Porcu E, Fabbri R, Ciotti PM, Petracchi S, Seracchioli R, Flamigni C. Ongoing pregnancy after intracytoplasmic sperm injection of epididymal spermatozoa into cryopreserved human oocytes. J Assist Reprod Genet 1999;16:283–285.
268. Tucker MJ, Morton PC, Wright G, Sweitzer CL, Massey JB. Clinical application of human egg cryopreservation. Hum Reprod 1998;13:3156–3159.
269. O'Neil L, Paynter SJ, Fuller BJ, Shaw RW, DeVries AL. Vitrification of mature mouse oocytes in a 6M Me_2SO solution supplemented with antifreeze glycoproteins: the effect of temperature. Cryobiology 1998;37:59–66.
270. Rubinsky B, Arav A, Mattioli M, DeVries AL. The effect of antifreeze glycopeptides on membrane potential changes at hypothermic temperatures. Biochem Biophys Res Comm 1990;173:1369–1374.
271. Rubinsky B, Arav A, DeVries AL. The cryoprotective effect of antifreeze glycopeptides from antartic fishes. Cryobiology 1992;29:69–79.
272. Rubinsky B, Arav A, Fletcher GL. Hypothermic protection – a fundamental property of "antifreeze" proteins. Biochem Biophys Res Comm 1991;180:566–571.
273. Rubinsky B, Arav A, DeVries AL. Cryopreservation of oocytes using directional cooling and antifreeze glycoproteins. Cryo-Lett 1991;12:93–106.
274. Arav A, Rubinsky B, Fletcher G, Seren E. Cryogenic protection of oocytes with antifreeze proteins. Mol Reprod Dev 1993;36:488–493.
275. Tablin F, Oliver AE, Walker NJ, Crowe LM, Crowe JH. Membrane phase transition of intact human platelets: correlation with cold-induced activation. J Cell Physiol 1996;168:305–313.
276. Yoshida K, Matsui Y, Wei T, Kaibori M, Kwon AH, Yamane A, Kamiyama Y. A novel conception for liver preservation at a temperature just above freezing point. J Surg Res 1999;81:216–223.
277. Vannereau H, Novakoviteh G, Carin M. Cryobiology of complex tissues. Contracept Fertil Sex 1998;26:573–577.
278. Rubinsky B, Arav A, Hong J-S, Lee CY. Freezing of mammalian livers with glycerol and antifreeze proteins. Biochem Biophys Res Comm 1994;200:732–741
279. Storey KB, Storey JM. Lifestyles of the cold and frozen. The Sciences. 1999;39:33–37.
280. Wang T, Zhu Q, Yang X, Layne JR DeVries AL Antifreeze glycoproteins from antarctic notothenoid fishes fail to protect the rat cardiac explant during hypothermic and freezing preservation. Cryobiology 1994;31:185–192.
281. Mugnano JA, Wang T, Layne JR Jr, DeVries AL, Lee RE Jr. Antifreeze glycoproteins promote intracellular freezing of rat cardiomyocytes at high subzero temperatures. Am J Physiol 1995;269:R474–479.

BNSDOCID: <XP_____1062997A__1_>

Biochemistry (Moscow), Vol. 68, No. 9, 2003, pp. 945-951. Translated from Biokhimiya, Vol. 68, No. 9, 2003, pp. 1157-1165.
Original Russian Text Copyright © 2003 by Allagulova, Gimalov, Shakirova, Vakhitov.

XP009079176

(7)

REVIEW

The Plant Dehydrins: Structure and Putative Functions

Ch. R. Allagulova, F. R. Gimalov, F. M. Shakirova*, and V. A. Vakhitov

Institute of Biochemistry and Genetics, Ufa Scientific Center, Russian Academy of Sciences, pr. Oktyabrya 69, Ufa 450054, Russia; fax: (3472) 35-6100; E-mail: shakirova@anrb.ru

Received January 15, 2003

Abstract—This review deals with recent data on the structure and biochemical properties of dehydrins, proteins that are normally synthesized in maturating seeds during their desiccation, and also in vegetative tissues of plants treated with abscisic acid or exposed to environmental stress factors that result in cellular dehydration. The dehydrins are considered as stress proteins involved in formation of plant protective reactions against dehydration. The generally accepted classification of dehydrins is based on their structural features, such as the presence of conserved sequences, designated as Y-, S-, and K-segments. The K-segment representing a highly conserved 15 amino acid motif (EKKGIMDKIKEKLPG) forming amphiphilic α-helix has been found in all dehydrins. The pathways of regulation of dehydrin gene expression, putative functions of dehydrins, and molecular mechanisms of their actions are discussed.

Key words: dehydrins, desiccation, stress factors, abscisic acid

The evolution of plants has resulted in the elaboration of protective mechanisms responsible for adaptation to constantly changing environmental conditions of growth. Stress-induced changes in expression of certain genes represent the genetic basis for the development of the protective reactions. It is well known that intensity of growth and development of plants depends on their water supply. However, dehydration is a normal process during seed maturation, which is realized by programmed expression of specific genes and the phytohormone abscisic acid (ABA) plays a key role in the regulation of this process. Several genes specifically expressed in this ontogenetic period have been recognized. They include genes encoding so-called LEA (Late Embryogenesis Abundant) proteins. LEA proteins have been found in ABA treated vegetating plants and also under the stress conditions that result in cellular dehydration such as drought, salinity, and low temperatures. That is why LEA proteins can be considered as stress proteins. Dehydrins represent one of the groups of the LEA protein family, which are involved in reactions of plants to drought, salinity, and dehydration. In the present review we have summarized current knowledge on structure, properties, and functional importance of dehydrins.

A previously unknown protein population found in embryos at late stages of embryogenesis was originally described by Dure et al. [1-3]. These proteins were called LEA (late embryogenesis abundant) proteins. Usually these proteins are not seen in premature embryos and young seedlings. It is interesting that appearance of the LEA proteins in embryonal tissue depends on content of endogenous ABA, whose level is usually increased during seed formation and especially before their maturation That is why LEA proteins have been called ABA-inducible proteins. For example, in mature cotton-plant seeds 18 LEA mRNA have been identified, and 13 of these were induced by ABA treatment in premature embryos [3]. Later the existence of LEA proteins was also found in other plant species representing various taxonomic groups [4-10]. This implies wide distribution of LEA proteins in the plant kingdom.

The fact of intensive synthesis and accumulation of LEA proteins in seeds at the late stages of their maturation, during dehydration, suggests the involvement of these proteins in protective reactions promoting maintenance of embryo structures under conditions of water deficit [11]. The existence of genes encoding LEA-proteins in the genetic programs controlling normal seed maturation is very "useful" for the vegetating plants because impairments of water balance (e.g., under drought and low temperature conditions) are always accompanied by induction of LEA proteins; (the latter is also observed during ABA treatment) [12]. This suggests involvement of LEA proteins in protection of plant cells

* To whom correspondence should be addressed.

0006-2979/03/6809-0945$25.00 ©2003 MAIK "Nauka/Interperiodica"

BNSDOCID: <XP_____9079176A_I_>

against damage induced by dehydration, but particular functions of these proteins are still unknown and represent a field for various speculations [13-17].

STRUCTURE AND PROPERTIES OF DEHYDRINS

LEA proteins of group II [10, 17], also known as the D-11 family of LEA proteins [11], are the most studied among drought-induced water-soluble proteins. These proteins have been called "dehydrins" [18]. Dehydrins have been isolated and purified from maize seeds [19, 20], vigna seeds [21], from barley and maize seedlings [18], and from bark tissue of peach [22]. Arabidopsis dehydrins RAB18, LTI29, LTI30, and COR47 [23] and wheat dehydrin WCOR410 [24] have been isolated from recombinant *E. coli* cells. Isolation and purification of native dehydrins allowed the investigation of their biochemical properties *in vitro*. These proteins are characterized by a wide range of molecular masses from 9 to 200 kD [10, 25]. Dehydrins are thermostable and maintain their integrity in aqueous solutions at up to 100°C. Generally, dehydrins are enriched with glycine and lysine residues, but they lack cysteine and tryptophan [22]. Analysis of the amino acid composition of a 35 kD dehydrin isolated from seedlings of cowpea (*Vigna unguiculata*) revealed that (apart from absence of cysteine and tryptophan), it is characterized by a high content of glycine, threonine, asparagine, glutamine, serine, and aspartic and glutamic acids; this explains the highly hydrophilic nature of this protein [21]. In arabidopsis LTI30 dehydrin with molecular mass of 21 kD, glycine, histidine, lysine, and threonine represent 56% of the total amino acids, whereas cysteine, tryptophan, arginine, and valine were not found [23].

The existence of conservative sequences denominated as K-, S-, and Y-segments [10] are structural features of dehydrins (and the presence of a K-segment is obligate for all dehydrins). The K-segment is a highly conservative Lys-rich motif that consists of 15 amino acid residues (EKKGIMDKIKEKLPG). However, some single amino acid substitutions and structural modifications may occur within the segment. For example, sequencing of cloned cDNA fragments and analysis of corresponding amino acid sequences revealed that the K-segment from *Pseudotsuga menziesii* dehydrins consists of the following amino acids (Q/E)K(P/A)G(M/L)LDKIK(A/Q)(K/M)(I/L)PG [15, 26].

Computer-aided analysis of amino acid sequences of K-segments demonstrated the possibility of formation of amphipathic α-helix [14]. Circular dichroism (CD) data also support formation of secondary α-helical structure by the K-segments. For example, the CD spectrum of vigna dehydrin suggests that in the presence of 10% SDS up to 15% of amino acids are involved in formation of secondary α-helical structure [21]. Similar results have been obtained for maize dehydrin G50 [19]. Interestingly, this protein lacking any hydrophobic sites may be involved in hydrophobic interactions. This suggests that the potential ability for hydrophobic interactions may be attributed to formation by K-segment of secondary structure as amphipathic α-helix, which combines hydrophilic and hydrophobic features [19]. In dehydrins the number of K-segment repeats may vary from 1 to 12 [10, 15, 23]. Thus, in spite of initial designation of dehydrins as dehydration-induced proteins, only K-segment proteins are now referred to this group.

Many dehydrins contain an S-segment that consists of serine residues; this site may be phosphorylated. The latter was demonstrated for maize RAB17 and tomato TAS14 dehydrins [20, 27, 28]. Phosphorylation of S-segment is suggested to promote dehydrin interaction with specific signal peptides followed by their translocation into the nucleus [10, 15, 29]. The N-terminal region of many dehydrins contains another conservative sequence, the Y-segment. This sequence, (V/T)DEYGNP, shares significant homology with the nucleotide-binding site of plant and bacterial chaperones; however, binding of dehydrin Y-segments with nucleotides remains to be well documented [10, 15].

Besides these three conservative segments, dehydrins also contain less conservative regions enriched with glycine and polar amino acid residues (especially threonine). These regions were named ϕ-segments [15, 21]. For example, the N-terminal region of maize RAB17 dehydrin contains paired ϕ-segments positioned between K-segments [15]. In some dehydrins ϕ-segments are localized as single copies between K-segments; they are enriched with proline and alanine residues (e.g., *Arabidopsis thaliana* COR47 protein [15]). Thus, the structure and size of dehydrin molecules, which can vary from 82 to 575 amino acid residues, is determined by combination and number of K-, S-, Y-, and ϕ-segments [10, 15].

CLASSIFICATION OF DEHYDRINS

The classification of dehydrins proposed by T. J. Close [10] is based on the presence of highly conservative segments. According to their combination, dehydrins have been subdivided into five classes: Y_nSK_2, K_n, K_nS, SK_n, and Y_2K_n.

Y_nSK_2 dehydrins. This is the most abundant class of dehydrins that contain from one to three Y-segments, one S-segment, and two K-segments. These are basic or neutral proteins. They are induced by drought or ABA (but not by cold temperatures). For example, barley dehydrins DHN1, DHN2, DHN3, DHN4, DHN6, and DHN9 share the common structure YSK_2. In seedlings they are accumulated in response to ABA treatment and dehydration, but their content remains unchanged during

BNSDOCID: <XP_ _____9079176A _I_>

287

hypothermia [30]. The barley dehydrin DHN12 also belongs to this class. Its gene is specifically expressed during the embryonal stage of development (i.e., it is not expressed in vegetative tissues even under stress conditions [31]).

K_n dehydrins. Dehydrins of this class contain from two to nine K-segments. These acidic or neutral proteins are induced by cold temperatures, dehydration, and ABA [10]. For example, blackberry K_5-dehydrin of molecular mass of 60-65 kD accumulates in flower buds during low temperature acclimatization [32]. Expression of barley *Dhn5B* gene encoding dehydrin K_9 is induced by low temperature treatment and to a lesser extent by drought and ABA [30, 33]. Soft wheat contains proteins of the WCS120 family (Wheat Cold Specific proteins), which are homologous to barley DHN5 dehydrin. Synthesis of proteins is specifically induced under low temperature and the level of their accumulation positively correlates with chilling tolerance of wheat and other cereals. WCS120 proteins differ from other dehydrins by the presence of a repeating glycine-rich motif: TGGTYQQGHTGTT [25].

SK_n dehydrins. Proteins of this class contain one S-segment and two or three K-segments [10]. They are acidic dehydrins, which are preferentially induced by low temperatures. For example, at 4°C the level of accumulation of WCOR410 protein (Wheat Cold Regulated Protein), which is an SK_n dehydrin, is significantly higher in winter wheat than in spring wheat plants [24]. The gene encoding barley acidic dehydrin DHN8 (a homolog of WCOR410 wheat protein) is actively expressed during low temperature treatment and to a lesser extent (than Y_nSK_2 proteins) during drought and ABA treatment [30]. Canadian fir dehydrin encoded by the *PgDhn1* gene is also an SK_n dehydrin. Increased expression of this dehydrin was observed in vegetative tissues during low temperature treatment, drought, salinity, wound stress, and treatment with jasmonate and methyl-jasmonate [34].

K_nS dehydrins. In contrast to dehydrins of other classes, the K-segment of K_nS dehydrins begins with the motif (H/Q)KEG (in others a similar site contains EKKG). This is the characteristic feature of these proteins. Expression of rice Wsi724 and medic Cas15a and 15b dehydrins that belong to this class is induced by dehydration and cold temperatures [10, 15].

Y_2K_n dehydrins. Acidic dehydrins of this class usually contain two Y-segments and one or two K-segments [10, 15]. These include vigna DHN1 dehydrin, which is associated with seed chilling tolerance during germination [35].

Some proteins cannot be assigned to a certain class of dehydrins. For example, chickweed leaves contain H26 protein, which has the SK_3S segment composition; it probably represents an intermediate form of proteins of SK_n and K_nS proteins [10]. One spinach protein has the following segment structure: YK_{11}. It shares some features of K_n-, Y_2K_n-, and K_nS-classes [10, 29].

Subdivision of dehydrins into several classes depending on their structural organization and biochemical properties was quite reasonable because of involvement of these proteins in plant response to various growing conditions. Synthesis of basic Y_nSK_2 dehydrins is induced by drought and ABA but not by low temperature treatments, whereas acidic dehydrins of K_nS, SK_n, and Y_2K_n classes are preferentially accumulated by plants in response to cold temperatures.

DISTRIBUTION OF DEHYDRINS

Dehydrins were initially found in flower plants [6, 18]. Immunological studies and screening of cDNA and genome libraries revealed the presence of dehydrins in other taxonomic groups [9, 26, 36-40]. Detection of dehydrins in various plant species can employ antibodies against synthetic polypeptide homolog of K-segment, the obligate component of dehydrins [8-10, 15, 29, 37]. In higher plants dehydrins are accumulated in seeds at later stages of embryogenesis and in vegetative tissues treated with ABA or subjected to stress factors (drought, salinity, and cold temperature). In lower plants dehydrin content is also increased under unfavorable environmental conditions. A certain correlation exists between the level of dehydrins and plant tolerance to dehydration. For example, cells of the drought-tolerant moss *Tortula ruralis* are characterized by increased constitutive expression of dehydrins [36]. In plants of the fern *Polypodium virginianum* and the moss *Exomotheca holstii*, dehydrin synthesis is induced by dehydration and ABA treatment [41, 42]. Immune analysis of the blue-green algae *Anabaena* and *Calothrix* cells revealed accumulation of 40 kD dehydrin in response to increase of salt concentration and polyethylene glycol addition to the cultivation medium [8]. Combination of constitutive and stress-induced dehydrin expression was found in the brown algae *Fucus spiralis*, *F. vesiculosus*, and *F. evanescens* [37] and the lichen *Selaginella lepidophylla* [15]. Thus, dehydrins are widely distributed in the plant kingdom. They have been found in the fungus *Pneumocystis carinii* (Accession No. NCBI CAC43457). Dehydrin homolog sequences were also found in *Escherichia coli* cells (Accession No. NCBI AAB18249), *Chlamidia trachomatis* [43], and even in *Drosophila melanogaster* cells [44].

REGULATION OF DEHYDRIN EXPRESSION

Dehydrin encoding genes are expressed in response to various stress-factors inducing moisture deficit. This implies the existence of many pathways for regulation of this process. Since stress-factors usually induce accumulation of ABA, treatment with this phytohormone also causes dehydrin gene expression in various plants [18, 30,

BNSDOCID: <XP_____9079176A__I_>

288

39, 41, 45-49]. So, dehydrins are referred to as RAB (Responsive to Abscisic acid)-proteins [6, 10, 15, 18, 46, 50, 51].

However, certain evidence exists that drought- and cold temperature-induced dehydrin genes were not expressed in plants treated with ABA [52-54]. Use of ABA-deficient sunflower mutants *nd-1* and *w-1* revealed that they accumulated transcripts of the *HaDhn1a* gene during seed embryogenesis and in seedlings subjected to water stress, and this accumulation was comparable to that found in wild type plants [55]. Thus, regulation of dehydrin gene expression occurs via ABA-dependent and ABA-independent pathways [30, 55-57].

Genes are sensitive to ABA due to the presence of certain *cis*-elements, known as ABA Responsive Elements (ABRE) in their promoters. These elements contain the ACGT sequence, denominated as the G-box; it should be noted that the presence of a single copy is not enough for ABA-dependent gene expression [30, 46, 50, 55, 58, 59]. For example, the promoter site of barley ABA-inducible *dhn1* gene contains a fragment, TACGTCCcgccccagacgaGGCCGCG, which consists of a G-box and GC-motif; its removal was accompanied by loss of promoter sensitivity to ABA [60]. Promoter sensitivity to ABA also depends on the presence of the MYC-element, containing the CACCTG sequence, and the MYB-element, which contains the TAACTG-motif [46, 58]. Their presence was demonstrated during sequencing of promoter regions of barley dehydrin genes [30]. Responsiveness to ABA signals also requires ABRE coupling elements (CE) [46, 61]. It was shown that ABRE-elements interact with bZIP transcription factors [50, 59, 62]. ABA-induced gene expression probably involves MYC- and MYB-transcription factors interacting with corresponding *cis*-elements [46, 59] and also transcription factors VP/ABI3 [50, 59, 63, 64].

ABA-independent pathways of regulation of dehydrin gene expression are realized by specific sites, such as DRE (Drought-Responsive Elements) [65] also known as CRT (C-repeat)-elements [66, 67]. These elements were found in promoters of genes expressed in response to dehydration, salinity, and cold temperatures. In the barley dehydrin gene promoters the number of repeats of DRE-elements may vary from 1 to 3 [67]. CBF (CRT-Binding Factor)/DREB (DRE-Binding Factor) transcription factors interact with DRE/CRT-elements [67-70]. These factors regulate transcription activity of dehydrin genes [67, 71-73].

Promoters of wheat *wcs120* dehydrin genes induced by cold temperatures contain the CANNTG motif; these genes may have up to eight repeats of this motif [25], which is the binding site for bHLH class transcription factors. Some other factors, such as $CPRF_3$ (Common Plant Regulatory Factor) and GBF_3 (G-Box Binding Factor), which are related to bZIP proteins, also specifically interact with this motif [25].

MOLECULAR MECHANISMS RESPONSIBLE FOR PROTECTIVE EFFECTS OF DEHYDRINS

Although functions of dehydrins remain unknown, their induction in vegetative tissues during cell dehydration suggests their involvement in protective reactions. In fact, during water deficit transcription activity of sunflower dehydrin gene *HaDhn2* is significantly higher in drought-tolerant than in drought-sensitive lines [74]. The dependence between drought-tolerance of plants and dehydrin accumulation was also found in sorghum plants [75]. In salinity-tolerant rice lines, the level of ABA-induced dehydrin expression was significantly higher than in sensitive plants [76]. Many plants demonstrate a direct interrelationship between the level of dehydrins and cold-tolerance [16, 22, 25, 32, 52, 77-82].

Cold-tolerance of plants depends on the level of starch accumulation. Starch is the main source of di- and monosaccharides that protect cells against dehydration-induced damages. So, maintenance of active α-amylase is a very important factor for starch degradation. Interestingly, in wooden plants starch degradation occurs even at temperatures below 0°C [83]. Cold acclimatization of birch seedlings is accompanied by accumulation of a dehydrin of 24 kD in amyloplasts. *In vitro* experiments revealed that the presence of this protein promoted maintenance of catalytic activity of α-amylase and appropriate level of starch hydrolysis during temperature decrease [81]. This suggests that dehydrins maintain adequate local water concentration near α-amylase, which is an important precondition for enzyme–substrate complex formation and starch degradation [81]. This water-retaining capacity of dehydrins is also supported by studies of the effect of peach PCA60 dehydrin on the shape of ice crystals, decrease in temperature, and the rate of their formation [22]. So, it is reasonable to draw a certain analogy between dehydrins and arctic fish antifreeze proteins. α-Helical sites of these proteins determine their affinity to the surface of ice crystal lattice, and this prevents subsequent growth of the ice crystals [22]. Cryoprotective activity of dehydrins was also demonstrated for cold-induced spinach CAP85, wheat WCS66, and peach PCA60 dehydrins, which were able to prevent inactivation of rabbit lactate dehydrogenase induced by a freeze–thaw cycle [84, 85, 22]. Administration of cold-inducible dehydrin gene *cor15a* into arabidopsis plants increased tolerance of isolated protoplasts and chloroplasts to freezing [86, 87].

Dehydrins are involved in osmotic regulation of plant cells. For example, study of various tissues of arabidopsis revealed that in stoma closing cells RAB18 dehydrin prevents water loss in nuclei during increase of osmotic pressure in cytoplasm at opened stomas; LTI29 and ERD14 dehydrins accumulating in root tips increase water flow to dividing meristematic cells [39]. Dehydrin involvement in regulation of osmotic potential of plant

BNSDOCID: <XP__..__9079176A__I_>

289

cells is possibly related to their synthesis during water stress [39].

Specific functions of dehydrins may be attributed to their particular localization in the cell. Dehydrins are localized in cytoplasm and nucleus [10, 15]. This was clearly demonstrated for wheat WCS120, maize RAB17, and tomato TAS14 dehydrins [25, 27, 28], peach PCA60 dehydrin [22], and birch 24 kD dehydrin [81]. Immunomicroscopic studies revealed that treatment of maize seeds with ABA for 15 min resulted in dehydrin accumulation in cytoplasm; during prolonged treatment they were detected in nucleus, and the maximal increase in dehydrin level in the nucleus and cytoplasm was observed after 48 h incubation with ABA [88]. In the nucleus dehydrins were associated with heterochromatin, euchromatin, nucleoplasm and nucleoli; in cytoplasm they were found mainly in cytosol, some quantity of dehydrins was also detected in organelles [88].

Nuclear localization of dehydrins in plant cells is associated with their protective role in stabilization of transcription machinery to unfavorable conditions. For example, pea embryo p16 dehydrin with molecular mass of 16 kD was isolated from the fraction of histone H3; this suggests involvement of this dehydrin in mechanisms responsible for structural integrity of chromatin during cell desiccation in the process of seed formation [89, 90]. Mechanisms underlying dehydrin translocation into the nucleus are not perfectly understood, but certain evidence exists that S-segment is involved in this process. For example, the stretch from amino acid residues 66 to 96, which contains S-segment with adjacent protein kinase CK2 recognition site and a sequence sharing similarity with NLS (Nuclear Localization Signal) viral signal peptides was identified in maize RAB17 dehydrin [91]. Since mutations of two amino acid residues in the putative NLS-fragment attenuated RAB17 dehydrin translocation into nucleus and mutations of three amino acid residues in protein kinase CK2 binding site completely blocked dehydrin phosphorylation and its transportation into the nucleus, it was concluded that maize RAB17 dehydrin translocation into the nucleus requires phosphorylation of serine residues in the S-segment; (the involvement of NLS-like sequences in this process is also possible) [91]. Dehydrins of wheat WCS120 family lack S-segment and the corresponding signal sequences and so it is proposed that the wheat dehydrins are translocated into the nucleus via protein–protein interactions [25].

It should be noted that not all dehydrins have been found in the nucleus. For example, cold-induced wheat WCOR410 dehydrin is located on the outer surface of the plasma membrane and in intercellular space (but not inside cells); it is a peripheral rather than an integral protein. In this connection, a reasonable question arises: how are WCOR410 dehydrins delivered to their correct destination [24]? Sequencing of this dehydrin revealed the presence of significant amounts of acidic and hydroxyl-containing amino acids. This may determine the ability of the protein for interaction with membrane phospholipids and steroids [24]. Dehydrins can be associated not only with plasmalemma, but also with intracellular membrane structures (as shown for parenchymal cells of maize scutellum [88] and onion epidermal cells [15]). Membrane localization of dehydrins may suggest their involvement in stabilization of cell membrane structures; this is especially important under unfavorable (environmental) conditions [10, 15]. Maintenance of spatial organization of biological membranes involves structured water molecules [92]. The dehydration process may damage integrity of hydrate cover of macromolecules; this, in turn, may result in shortening of the distance between membrane bilayers, mixing their components and finally to irreversible changes in the structure of biomembranes [93]. Interaction between dehydrins and membrane macromolecules with damaged hydrate cover may exert a protective effect on the membrane structures [10, 15, 24].

Structural features of dehydrins suggest their involvement in hydrophobic and hydrophilic interactions. Due to the presence of K-segments forming amphiphilic α-helix, dehydrins may interact with lipid components of biomembranes and (like chaperones) with hydrophobic sites of partially denatured proteins [10, 15, 94]. Dehydrins may also interact with water-soluble low molecular weight components of cytoplasm and polar groups of proteins [10, 15]. Maize G50 dehydrin is an example of a dehydrin demonstrating hydrophobic interactions [19]. The presence of ϕ-segments may also increase hydrophilicity of dehydrin structure. Since ϕ-segments are characterized by increased proportion of polar amino acid residues, they may be involved in hydrophilic interactions between dehydrins and polar groups of macromolecules and also between dehydrins and such low molecular weight components of nucleoplasm and cytosol as sugars, amino acids, and water molecules. Due to the presence of K- and ϕ-segments, dehydrins may prevent further denaturation and aggregation of partially denatured macromolecules by forming a protecting shell; this stabilizes cell structures at all levels of their organization [10, 15].

Putative functions of dehydrins also include involvement in detoxification of salt ion excess observed during cell dehydration. It was shown that RAB18, LTI29, LTI30, and COR47 dehydrins from *Arabidopsis thaliana* can interact with Cu^{2+} and Ni^{2+}. The dehydrin LTI30 can also bind Co^{2+} and Zn^{2+}; however, it should be noted that these dehydrins lack any specific metal-binding motifs. Metal ion binding capacity of *Arabidopsis thaliana* dehydrins may be attributed to the presence of a high number of histidine residues [22]. Wheat WCOR410 dehydrin contains up to 43% of charged amino acid residues, which may be involved in ion binding; this reduces the probability of sedimentation and crystallization of cell sap salts during dehydration [24]. These data may serve as an argument for the chelating functions of dehydrins.

BNSDOCID: <XP_____9079176A_I_>

 XP009079176

Thus, dehydrins are characteristic proteins of late ontogenesis. They play an important role in seed desiccation during the maturation period and so they have been pooled into an independent group of proteins. Accumulation of dehydrins in vegetating plants in response to cold temperatures, salinity, and moisture deficit attracts much interest in relation to possible protective functions of dehydrins under conditions of cell dehydration. Dehydrin gene expression is regulated by ABA, and so these proteins are considered as components of the ABA-controlled anti-stress subprograms. Detailed study of structural organization of dehydrins revealed the existence of repeated conservative domains providing specific spatial configuration that underlines dehydrin involvement in the reactions preventing loss of water. Dehydrins are widely distributed in plants. They significantly vary in amino acid composition (from 100 to 600 amino acid residues) and by molecular mass. The amino acid composition of these proteins is characterized by high content of charged and polar residues, and this determines their biochemical properties including thermostability. This may promote their specific protective functions under conditions of cell dehydration: dehydrins may prevent coagulation of macromolecules and maintain integrity of crucial cell structures. Many researchers consider dehydrins together with other proteins accumulating in plant cells during water cells. These include some emulsifiers and chaperones stabilizing macromolecule surface in the cells. However, the precise physiological functions of dehydrins remain to be clarified.

This work was supported by the Program for Support of Leading Scientific Schools (00-15-97899, 00-15-97810, NSh - 1785.2003.4, NSh - 2217.2003.4).

REFERENCES

1. Dure, L., III, Greenway, S. C., and Galau, G. A. (1981) *Biochemistry*, 20, 4162-4168.
2. Galau, G. A., and Dure, L., III (1981) *Biochemistry*, **20**, 4169-4178.
3. Galau, G. A., Hughes, D. W., and Dure, L., III (1986) *Plant Mol. Biol.*, 7, 155-170.
4. Choi, J. H., Liu, L., Borkird, C., and Sung, Z. R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1906-1910.
5. Hong, B., Uknes, S. J., and Ho, T.-H. D. (1988) *Plant Mol. Biol.*, **11**, 495-506.
6. Mundy, J., and Chua, N.-H. (1988) *EMBO J.*, 7, 2279-2286.
7. Lits, J. C., Colwell, G. W., Chakerian, R. L., and Quatrano, R. S. (1987) *Nucleic Acids Res.*, 15, 3607-3618.
8. Close, T. J., and Lammers, P. J. (1993) *Plant Physiol.*, **101**, 773-779.
9. Close, T. J., Fenton, R. D., and Moonan, F. (1993) *Plant Mol. Biol.*, 23, 279-286.
10. Close, T. J. (1996) *Physiol. Plant.*, **96**, 795-803.
11. Dure, L., III, Crouch, M., Harada, J., Ho, T.-H. D., Mundy, J., Quatrano, R., Thomas, T., and Sung, Z. R. (1989) *Plant Mol. Biol.*, **12**, 475-486.
12. Bray, E. A. (1993) *Plant Physiol.*, **103**, 1035-1040.
13. Dure, L., III (1993) in *Response of Plants to Cellular Dehydration during Environmental Stress* (Close, T. J., and Bray, E., eds.) American Society for Plant Physiology, Rockville, pp. 91-103.
14. Baker, J., Steele, C., and Dure, L., III (1988) *Plant Mol. Biol.*, **11**, 277-291.
15. Close, T. J. (1997) *Physiol. Plant.*, **100**, 291-296.
16. Zhu, B., Choi, D.-W., Fenton, R., and Close, T. J. (2000) *Mol Gen. Genet.*, **264**, 145-153.
17. Ramanjulu, S., and Bartels, D. (2002) *Plant Cell Environment*, **25**, 141-151.
18. Close, T. J., Kortt, A. A., and Chandler, P. M. (1989) *Plant Mol. Biol.*, **13**, 95-108.
19. Ceccardi, T. L., Meyer, N. C., and Close, T. J. (1994) *Prot. Exp. Purific.*, **5**, 266-269.
20. Campbell, S. A., Crone, D. E., Ceccardi, T. L., and Close, T. J. (1998) *Plant Mol. Biol.*, **38**, 417-423.
21. Ismail, A. M., Hall, A. E., and Close, T. J. (1999) *Pla Physiol.*, **120**, 237-244.
22. Wisniewski, M., Webb, R., Balsamo, R., Close, T. J., Yu, X.-M., and Grifith, M. (1999) *Physiol. Plant.*, **105**, 600-608.
23. Svensson, J., Palva, E. T., and Welin, B. (2000) *Prot. Exp. Purific.*, **20**, 169-178.
24. Danyluk, J., Perron, A., Houde, M., Limin, A., Fowler, B., Benhamou, N., and Sarhan, F. (1998) *Plant Cell*, **10**, 623-638.
25. Sarhan, F., Ouellet, F., and Vazquez-Tello, A. (1997) *Physiol. Plant.*, **101**, 439-445.
26. Jarvis, S. B., Taylor, M. A., MacLeod, M. R., and Davies, H. V. J. (1996) *Plant Physiol.*, **147**, 559-566.
27. Goday, A., Jensen, A. B., Culianez-Macia, F. A., Alba, M. M., Figueras, M., Serratosa, J., Torrent, M., and Pages, M. (1994) *Plant Cell*, **6**, 360-361.
28. Godoy, J., Lunar, R., Torres-Schumann, S., Moreno, J., Rodrigo, R. M., and Pintor-Toro, J. A. (1994) *Plant. Mol. Biol.*, **26**, 1921-1934.
29. Campbell, S. A., and Close, T. J. (1997) *New Phytol.*, **137**, 61-74.
30. Choi, D.-W., Zhu, B., and Close, T. J. (1999) *Theor. App' Genet.*, **98**, 1234-1247.
31. Choi, D.-W., and Close, T. J. (2000) *Theor. Appl. Genet.*, **100**, 1274-1278.
32. Levi, A., Panta, G. R., Parmentier, C. M., Muthalif, M. M., Arora, R., Shanker, S., and Rowland, L. (1999) *Physiol. Plant.*, **107**, 98-109.
33. Close, T. J., Meyer, N. C., and Radick, J. (1995) *Plant Physiol.*, **107**, 289-290.
34. Richard, S., Morency, M. J., Drevet, C., Jouanin, L., and Seguin, A. (2000) *Plant Mol. Biol.*, **1**, 1-10.
35. Ismail, A. M., Hall, A. E., and Close, T. J. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 13566-13570.
36. Bewley, J. D., Reynolds, T. L., and Oliver, M. J. (1993) in *Response of Plants to Cellular Dehydration during Environmental Stress* (Close, T. J., and Bray, E., eds.) American Society for Plant Physiology, Rockville, pp. 193-201.
37. Li, R., Brawley, S. H., and Close, T. J. (1998) *J. Phycol.*, 34, 642-650.
38. Kontunen-Soppela, S., Taulavuori, K., Taulavuori, E., Lahdesmaki, P., and Laine, K. (2000) *Physiol. Plant.*, **109**, 404-409.

BNSDOCID: <XP_____9079176A__I_>

39. Nylander, M., Svensson, J., Palva, E. T., and Welin, B. V. (2001) *Plant Mol. Biol.*, **45**, 263-279.
40. Bomal, C., Le, V. Q., and Tremblay, F. M. (2002) *Physiol. Plant.*, **115**, 523-530.
41. Reynolds, T. L., and Bewley, J. D. (1993) *J. Exp. Bot.*, **269**, 1771-1779.
42. Hellwege, E. M., Dietz, K.-J., Volk, O. H., and Hartung, W. (1994) *Planta*, **194**, 525-531.
43. Stephens, R. S., Kalman, S., Lammel, C. J., Fan, J., Marathe, R., Aravind, L., Mitchell, W. P., Olinger, L., Tatusov, H. L., Zhao, Q., Koonin, E. V., and Davis, R. W. (1998) *Science*, 282, 754-759.
44. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., and George, R. A. (2000) *Science*, **287**, 2185-2195.
45. Robertson, M., and Chundler, P. M. (1992) *Plant Mol. Biol.*, **19**, 1031-1044.
46. Busk, P. K., and Pages, M. (1998) *Plant Mol. Biol.*, **37**, 425-435.
47. Lang, V., Robertson, M., and Chandler, P. M. (1998) *Theor. Appl. Genet.*, **96**, 1193-1199.
48. Jiang, Y., and Huang, B. (2002) *Crop Science*, **42**, 202-207.
49. Welling, A., Moritz, T., Palva, E. T., and Juntilla, O. (2002) *Plant Physiol.*, **129**, 1633-1641.
50. Leung, J., and Giraudat, J. (1998) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 199-222.
51. Tamminen, I., Makele, P., Heino, P., and Palva, E. T. (2001) *Plant J.*, **25**, 1-8.
52. Houde, M., Danyluk, J., Laliberte, J.-F., Rassart, E., Dhindsa, R. S., and Sarhan, F. (1992) *Plant Physiol.*, **99**, 1381-1387.
53. Kiyosue, T., Yamaguchi-Shinozaki, K., Shinozaki, K., Kamada, H., and Harada, H. (1993) *Plant Mol. Biol.*, **21**, 1053-1068.
54. Whitsitt, M. S., Collins, R. G., and Mullet, J. E. (1997) *Plant Physiol.*, **114**, 917-925.
55. Giordany, T., Natali, L., D'Ercole, A., Pugliesi, C., Fambrini, M., Vernieri, P., Vitagliano, C., and Cavallini, A. (1999) *Plant. Mol. Biol.*, **39**, 739-748.
56. Shinozaki, K., and Yamaguchi-Shinozaki, K. (1997) *Plant Physiol.*, **115**, 327-334.
57. Zhu, J.-K. (2002) *Annu. Rev. Plant Biol.*, 53, 247-273.
58. Rock, C. D. (2000) *New Phytol.*, **148**, 357-396.
59. Finkelstein, R. R., Gampala, S. S. L., and Rock, C. D. (2002) *Plant Cell*, **14**, 15-45.
60. Robertson, M., Cuming, A. C., and Chundler, P. M. (1995) *Physiol. Plant.*, **94**, 470-478.
61. Shen, Q., Zhang, P., and Ho, T.-H. D. (1996) *Plant Cell*, **8**, 1107-1119.
62. Gago, G. M., Almoguera, C., Jordano, J., Gonzalez, D. H., and Chan, R. L. (2002) *Plant Cell Environment*, 25, 633-640.
63. Parcy, F., and Giraudat, J. (1997) *Plant J.*, 11, 693-702.
64. McCarty, D. R., Hattori, T., Carson, C. B., Vasil, V., Lazar, M., and Vasil, I. K. (1991) *Cell*, **66**, 895-905.
65. Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994) *Plant Cell*, **6**, 251-264.
66. Stockinger, E. J., Gilmour, S. J., and Thomashow, M. F. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 1035-1040.
67. Thomashow, M. F., Gilmour, S., Stockinger, E., Jaglo-Ottosen, K. R., and Zarka, D. G. (2001) *Physiol. Plant.*, **112**, 171-175.
68. Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998) *Plant Cell*, **10**, 1391-1406.
69. Shinwari, Z. K., Nakashima, K., Miura, S., Kasuga, M., Seki, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998) *Biochem. Biophys. Res. Commun.*, **250**, 161-170.
70. Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schanbenberger, O., and Thomashow, M. F. (1998) *Science*, **280**, 104-106.
71. Thomashow, M. F. (1998) *Plant Physiol.*, **118**, 1-7.
72. Thomashow, M. F. (1999) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **50**, 571-599.
73. Choi, D.-W., Rodriguez, E. M., and Close, T. J. (2002) *Plant Physiol.*, **129**, 1781-1787.
74. Cellier, F., Conejero, G., Breitler, J. C., and Casse, F. (1998) *Plant Physiol.*, **116**, 319-328.
75. Wood, A. J., and Goldsbrough, P. B. (1997) *Physiol. Plant.*, **99**, 144-152.
76. Moons, A., Bauw, G., Prinsen, E., Montagu, M. V., and van der Straeten, D. (1995) *Plant Physiol.*, **107**, 177-186.
77. Robertson, A. J., Weninger, A., Wilen, R. W., Fu, P., and Gusta, L. V. (1994) *Plant Physiol.*, **106**, 1213-1216.
78. Arora, R., and Wisniewski, M. E. (1994) *Plant Physiol.*, **105**, 95-101.
79. Arora, R., Rowland, L. J., and Panta, G. R. (1997) *Physiol. Plant.*, **101**, 8-16.
80. Artlip, T., Callahan, A., Bassett, C., and Wisniewski, M. E. (1997) *Plant Mol. Biol.*, **33**, 61-70.
81. Rinne, P., Kaikuranta, P., van der Plas, L., and van der Schoot, C. (1999) *Planta*, **209**, 377-388.
82. Lim, C. C., Krebs, S. L., and Arora, R. (1999) *Theor. Appl. Genet.*, **99**, 912-920.
83. Rinne, P., Tuominen, H., and Juntilla, O. (1994) *Tree Physiol.*, **14**, 549-561.
84. Kazuoka, T., and Odeda, K. (1994) *Plant Cell Physiol.*, **35**, 601-611.
85. Houde, M., Daniel, C., Lachapelle, M., Allard, F., Laliberte, S., and Sarhan, F. (1995) *Plant J.*, **8**, 583-593.
86. Artus, N. N., Uemura, M., Steponkus, P. L., Gilmour, S., Lin, C., and Thomashow, M. F. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 13404-13409.
87. Thomashow, M. F., Artus, N., Gilmour, S., Stockinger, E., Wilhelm, K., Zarka, D., Joseph, R. A., Uemura, M., and Steponkus, P. J. (1996) *Plant Physiol.*, **111**, 29.
88. Egerton-Warburton, L. M., Balsamo, R. A., and Close, T. J. (1997) *Physiol. Plant.*, **101**, 545-555.
89. Castillo, J., Rodrigo, M. I., Marquez, J. A., Zunniga, A., and Franco, L. (2000) *Eur. J. Biochem.*, **267**, 2156-2165.
90. Castillo, J., Zunniga, A., Franco, L., and Rodrigo, M. I. (2002) *Eur. J. Biochem.*, **269**, 4641-4648.
91. Jensen, A. B., Goday, A., Figueras, M., Jessop, A. C., and Pages, M. (1998) *Plant J.*, **13**, 691-697.
92. Israelachvili, J., and Wennerstron, H. (1996) *Nature*, **379**, 219-225.
93. Steponkus, P. L., Uemure, M., and Webb, M. S. (1993) in *Response of Plants to Cellular Dehydration during Environmental Stress* (Close, T. J., and Bray, E., eds.) American Society for Plant Physiology, Rockville, pp. 37-47.
94. Mayhew, M., and Hartl, F. U. (1996) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F., ed.) American Society for Microbiology, Washington, DC, pp. 922-937.

BNSDOCID: <XP_____9079176A_I_>

XP-002421537

POPP the question: what *do* LEA proteins do?

Michael J. Wise[1] and Alan Tunnacliffe[2]

[1]Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK CB2 3EH
[2]Institute of Biotechnology, University of Cambridge, Tennis Court Road, Cambridge, UK CB2 1QT

Late embryogenesis abundant (LEA) proteins are produced in maturing seeds and anhydrobiotic plants, animals and microorganisms, in which their expression correlates with desiccation tolerance. However, their function has remained obscure for 20 years. We argue that novel computational tools devised for non-globular proteins might now overcome this problem. Predictions arising from bioinformatics fit well with recent data on Group 3 proteins, which potentially form cytoskeletal filaments, and suggest experimentally testable functions for these and other LEA protein groups.

Late embryogenesis abundant (LEA) proteins were first identified 20 years ago in cotton and wheat [1–3] but their function remains a puzzle. LEA genes have been cloned from many plant species and at least six different groups of LEA proteins have been defined on the basis of expression pattern and sequence [4–6] (Table 1). As their name suggests, they are produced in abundance during late embryo development, comprising up to 4% of cellular protein [7], with maximum expression post-abscission during desiccation [8]. Their expression is linked to the acquisition of desiccation tolerance in orthodox seeds, pollen and anhydrobiotic plants, but many LEA proteins are induced by cold, osmotic stress or exogenous abscisic acid, or can even be expressed constitutively [e.g. *dhnX* from *Arabidopsis* (DHX1_ARATH)] [9]. They have been variously proposed to protect cellular structures from the effects of water loss by acting as a hydration buffer, by sequestering ions, by direct protection of other proteins or membranes, or by renaturing unfolded proteins [4,5]. These suggestions for function have been made on a largely intuitive basis and are supported by relatively little evidence. However, some effect on stress tolerance seems apparent because tomato, wheat and barley LEA proteins have been shown to confer increased resistance to osmotic or freeze stresses when introduced into yeast [10–13], and a barley LEA protein improved tolerance to water deficit in transgenic rice [14] and wheat [15]. Furthermore, *in vitro*, an algal LEA protein diminished freeze damage of the enzyme lactate dehydrogenase [16].

Few clues about function have been derived from protein structure because all three major groups of LEA proteins seem to be natively unfolded, at least in the hydrated state [17], even though secondary structure prediction programs suggest a high degree of folding for Group 3 and Group 6 LEA proteins, chiefly into amphipathic α-helices. By contrast, many Group 1 and Group 2 LEA proteins are predicted to have significant unstructured ('loop') regions and, in almost every case, the proportion reported as loop in four-state predictions (α-helix, β-sheet, loop or no prediction) by the PredictProtein server

Table 1. LEA protein groups mapping to POPP-derived superfamilies

LEA group	Sequence motif	LEA superfamily	Consensus POPP[a]
1a	GGQTRREQLGEEGYSQMGRK	4	+E, +G, +EG, +GE, +GG, +KG, +QE, +RK, +GGE, +KGG
1b		6	+E, +G, +DE, +EG, +ES, +GG, +GQ, +RE, +RK, +ARE, +DES, +REG
2a	DEYGNP (Y domain)	1	+G, −L, +EK, +GG, +GT, +EKL, +IKE, +KEK, +KIK, +KKG, +KLP, +LPG
2a	EEKK (K domain)	10	−F, +G, −I, +AG, +EK, +GG, +GQ, +KE, +SS, +EKL, +GAG, +IKE, +KEK, +KLP, +LPG, +SSS
2b	S_n (S segment)	3	−F, −I, −L, −R, −W, +DK, +EK, +KK, +KL, +LP, +TH, +EKK, +KEK, +KLP, +LPG
3a	TAQAAKEKAGE	2	+A, −C, +E, −F, −I, +K, −L, −P, +AE, +AK, +EK, +ET, +GE, +GK, +KE, +AAE, +AKD, +EKA
3b		5	+A, −I, +K, −L, −P, +Q, +T, −V, +AA, +AQ, +EK, +KE, +KT, +QA, +QQ, +QS, +QT, +TQ, +AAK, +AQA, +EKT, +QAA, +TQQ
6	?	7	+A, −F, −L, +AA, +AE, +MQ, +QS, +VA, +AAA, +GVA, +QSA, +SAA

[a]A consensus POPP (Protein or Oligonucleotide Probability Profile) is shown for each superfamily, where '+' indicates significant overexpression of a peptide, and '−' indicates significant underexpression. Some minor superfamilies are not shown. The sequences of proteins generally considered to be Group 4 are known [30] to redistribute consistently to Group 2 (PM1_SOYBN, O24442 and LE13_GOSHI) and Group 3 (LE11_HELAN and LE25_LYCES). The two Group 5 proteins, LE29_GOSHI and Q93Y63 are found among the Group 3 LEA proteins. Notice that there is considerable overlap in the consensus POPPs across subgroups, so further work is needed to confirm the biological significance of this refinement.

Corresponding author: Alan Tunnacliffe (at10004@biotech.cam.ac.uk).

http://plants.trends.com 1360-1385/$ - see front matter © 2003 Elsevier Ltd. All rights reserved. doi:10.1016/j.tplants.2003.10.012

BNSDOCID: <XP_ _2421537A_I_>

(http://cubic.bioc.columbia.edu/predictprotein/predictprotein.html) [18] exceeds the proportion reported as structured. The mystery surrounding the role of LEA proteins was deepened further by the discovery of homologues in organisms other than plants, including the nematodes *Caenorhabditis elegans*, *Steinernema feltiae* and *Aphelenchus avenae*, and the prokaryotes *Deinococcus radiodurans*, *Bacillus subtilis* and *Haemophilus influenzae* [19–22]. Some evidence from these organisms supports a role for LEA proteins in desiccation tolerance [23] but, again, precise function is unknown.

Novel bioinformatics for low-complexity proteins

A mechanism is therefore needed to form reasoned hypotheses relating to LEA protein function that can then be tested experimentally. Bioinformatics might represent a way forward, although conventional tools (such as the BLAST algorithm [24]) demonstrate only limited relatedness between LEA proteins and other polypeptides. One reason for this is that LEA proteins contain regions of low sequence complexity (peptides whose sequences display significant regularities) [25]. For example, using the low-complexity-peptide search tool 0j.py [26], a cross-section of LEA proteins spanning all groups attracted a median score of 11 (a high score indicates low complexity). For comparison, 116 745 sequences at least 50 amino acids long taken from SwissProt attract a median score of 3. This is partly because Group 3 LEA proteins have up to 13 repetitions of the consensus 11-mer TAQAAKEKAGE, Group 2 LEA proteins have repeats of either or both of DEYGNP and EEKK, and most have a poly-serine stutter. Low-complexity proteins and protein domains are routinely masked out during sequence similarity searches because their inclusion can severely skew search statistics attuned to globular proteins [27].

To address this and similar problems, a new computational method called POPP (Protein or Oligonucleotide Probability Profile) analysis has been developed [28] that allows proteins to be compared based on similarities in their peptide compositions rather than similarities in their sequences. The rationale is simple: proteins or protein domains with similar function are likely to have the same short peptides (mono-, di- and tri-peptides) over- or under-represented. Furthermore, because local-alignment sequence comparison tools such as BLAST typically score matches over longer oligopeptides and also introduce penalties for non-contiguous matches, they might miss the similarities that POPP analysis is designed to detect.

At the simplest level, POPPs show up any over- or under-representation of particular amino acids in a protein sequence. Thus, we find that the description of the three major Groups of LEA proteins as hydrophilic is borne out because charged and/or polar residues feature strongly. Lysine is strongly represented in Groups 2 and 3, and moderately strongly in Group 1. Group 1 also shows highly significant over-representation of arginine. Similarly, glutamate is strongly over-represented in Groups 1 and 3, and moderately over-represented in Group 2. The finding that cysteine is lacking in these groups is also confirmed but, in addition, we can now add that phenylalanine, tryptophan, isoleucine, leucine and asparagine are also significantly under-represented. By contrast, glycine is strongly represented in Group 1 and Group 2 but is found only marginally more than expected by chance in Group 3. This means that analyses of the type performed by Adriana Garay-Arroyo *et al.* [29], which characterized LEA proteins as hydrophilic with a high glycine content and used these parameters to search databases for similar proteins, named 'hydrophilins', are somewhat limited in scope.

Beyond amino acid representation, POPPs, together with other machine-learning (automated classification) techniques, have first been used to examine relatedness of LEA protein sequences [30]. Those proteins with overlapping POPPs can be clustered, and this process can be taken further by identifying similarity between clusters, leading to a grouping of proteins into families and superfamilies (SFs). Even though SF numbers are generated automatically and therefore need have no connection with known LEA groups, POPP SF designations align mainly with group assignments, at least for the major LEA protein groups (1, 2 and 3) (Table 1). The SF designations introduce further refinement of the categorization of LEA proteins, which might provide insight into their function. For example, Group 2 LEA proteins appear to split into two subgroups: Group 2a, which includes all the Group 2 proteins not associated with cold tolerance and all the Group 2 proteins that are literally late in embryogenesis (SF1 and a splinter superfamily, SF10); and Group 2b LEA proteins, which are broadly associated with cold tolerance, together with the few that are explicitly not produced during late embryogenesis (SF3). Subgroups of Group 3 proteins, corresponding to SF2 (Group 3a) and SF5 (Group 3b), can also be identified, and similarly, subgroups of Group 1 proteins can also be identified. More radically, Groups 4 and 5 disappear altogether under the POPP and machine-learning spotlights, with former members of these groups consistently redistributing into LEA Groups 2 and 3 [30] (Table 1). This can be rationalized by inspecting the sequences, showing (for example) that Group 5 proteins have a similar repeat motif to that found in Group 3 members.

POPP prediction of protein function

The power of the POPP approach lies in the ability to interrogate a database of POPPs for proteins of known function with similar peptide profiles to query proteins of unknown function, in this case the LEA proteins. When this is done against a database of POPPs representing SwissProt and the keywords associated with the hits are then clustered, there are several interesting matches. For example, Group 1a (SF4) shares POPP vectors with histones, especially histone H4, and with other chromosomal and nuclear proteins, including proteins known to bind DNA. Group 1b (SF6), by contrast, matches a group of proteins with somewhat broader functions, such as RNA or ATP binding and gyrase and chaperone activities (Table 2). Each superfamily is identified with its own functional theme and, although there is overlap between these themes, each SF has a characteristic pattern of functional matches. So, Group 3b (SF5) proteins also match with nuclear, DNA binding proteins and with histones but,

BNSDOCID: <XP____.___2421537A__1_>

Table 2. Key words and phrases associated with each group and superfamily[a]

LEA group	LEA superfamily	Key words and phrases	Possible function(s)
1a	4	Histone H4, chromosomal protein, nuclear protein, methanofuran, DNA binding	DNA binding nuclear protein
1b	6	dsRNA binding, DNA gyrase, (DNA) breakage, CLP, ATP binding	Nucleic acid unwinding or nucleic acid repair Molecular chaperone
2a	1	(DNA) break, ATP binding, DNA topoisomerase, protein biosynthesis, topoisomerase, repair	DNA unwinding or repair
2a	10	Nuclear protein, DNA binding, transcription regulation, intermediate filament, keratin, chaperone, homeobox, coiled coil, HMG box domain, cytoskeletal	DNA-binding nuclear protein; regulation of transcription? Cytoskeleton
2b	3	Coiled coil, nuclear protein, caldesmon, histone H1, chaperone, tropomyosin, filament, (DNA) break, DNA topoisomerase	DNA unwinding or repair Cytoskeleton Ca^{2+} binding Molecular chaperone
3a	2	Chaperone, coiled coil, tropomyosin, stress, filament, phosphorylation, caldesmon, elongation factor, neurofilament, actin binding, cytoskeleton, rotamase	Molecular chaperone Cytoskeleton Ca^{2+}-binding
3b	5	Coiled coil, histone H1, filament, nuclear protein, neurofilament, antigenic, flagella, HAMP domain, synuclein, peptidoglycan anchor, DNA binding, hsp70	Chromatin-associated nuclear protein Filament Kinase or phosphatase?
6	7	GroEL protein, nuclear protein, histone H1, chaperonin, DNA binding, HAMP domain, synuclein, transcription regulation	Molecular chaperone Chromatin-associated nuclear protein; transcription factor?

Abbreviations: dsRNA, double-stranded RNA; HAMP, the group of histidine kinases, adenylyl cyclases, methyl binding proteins and phosphatases (see InterPro entry IPR003660); HMG, high mobility group (a DNA binding domain); LEA, late embryogenesis abundant proteins.

[a]The list of protein identifiers was generated by a search of a POPP (Protein or Oligonucleotide Probability Profile) version of SwissProt, with each consensus POPP as the query, and where then submitted to the Protein Annotators' Assistant (after deleting any LEA proteins). The most significant key words and phrases are listed here, together with some of the functional themes that emerge. CLP is an ATP-dependent plant chaperone.

more specifically, with H1 histones. And although Group 2b (SF3) has DNA recognition as one putative function, there is also relatedness to proteins that bind other polymers, such as intermediate filaments (IFs). These functions are largely hypothetical, but there is tremendous scope here for experimental exploration of potential functions for the various LEA proteins. Many relevant experiments are immediately obvious; for example, DNA binding suggests nuclear localization, which is relatively straightforward to demonstrate, as are ATP and GTP hydrolase activities.

One obvious question that arises from the above analysis is why any of this is relevant if LEA proteins are unstructured. It seems clear that LEA proteins, at least the three major Groups, form part of the growing family of 'natively unfolded' or 'intrinsically disordered' proteins [31,32]. A large proportion of the proteome of many species is predicted to comprise partially or wholly unfolded proteins. For example, in *Drosophila*, up to 17% of proteins could be wholly disordered. However, lack of structure does not mean lack of function because folding of these proteins can occur on binding to biological target molecules [33–35]. Examples include the haem-induced folding of apocytochrome c and the increase in structure seen in the CFP-10 major T-cell antigen of *Mycobacterium tuberculosis* when it binds its partner protein ESAT-6 [36,37]. There are also many examples of local folding of proteins being coupled to nucleic acid binding, such as when transcription factors lymphoid enhancer factor 1 and alcohol dehydrogenase regulatory gene 1 protein recognize their respective DNA recognition elements, or bacteriophage λ antitermination protein N complexes with its cognate RNA [33].

Almost all natively unfolded proteins have defined partner molecules, which can be small molecules such as nucleotides or cations as well as macromolecules. Do LEA proteins also have structure-inducing binding partners? Possibly, but we would expect to have identified these by now, at least for macromolecular partners, which would probably be present at the same high concentrations, and with the same expression profiles, as the LEA proteins themselves. Alternatively, LEA proteins might exert their function in the unstructured state. This is not altogether far-fetched if LEA proteins function as a hydration buffer, ion sink or water replacement molecule. Indeed, if LEA proteins function as water replacement molecules, a random coil arrangement might be preferable, allowing the LEA protein to adjust its conformation to facilitate hydrogen bonding with essentially any macromolecular or membrane surface. An additional hypothesis is that LEA proteins might act as a 'molecular shield', simply decreasing the frequency of collisions between partially denatured proteins, thereby reducing the possibility of exposed hydrophobic regions seeding damaging aggregate formation.

Desiccation-induced folding of LEA proteins

Another possibility is that abiotic stress can also induce folding and, indeed, recent experimental data show a conformational shift on dehydration of the Group 3a (SF2) LEA protein from the nematode *Aphelenchus avenae*, AavLEA1 [17]. Although AavLEA1 is unstructured in solution, Fourier transform infrared (FT-IR) spectroscopic analysis shows that the protein becomes more folded, developing a significant α-helical component, when dried. Furthermore, spectral components were present that were consistent with the formation of superhelical – possibly

BNSDOCID: <XP _____2421537A__I_ >

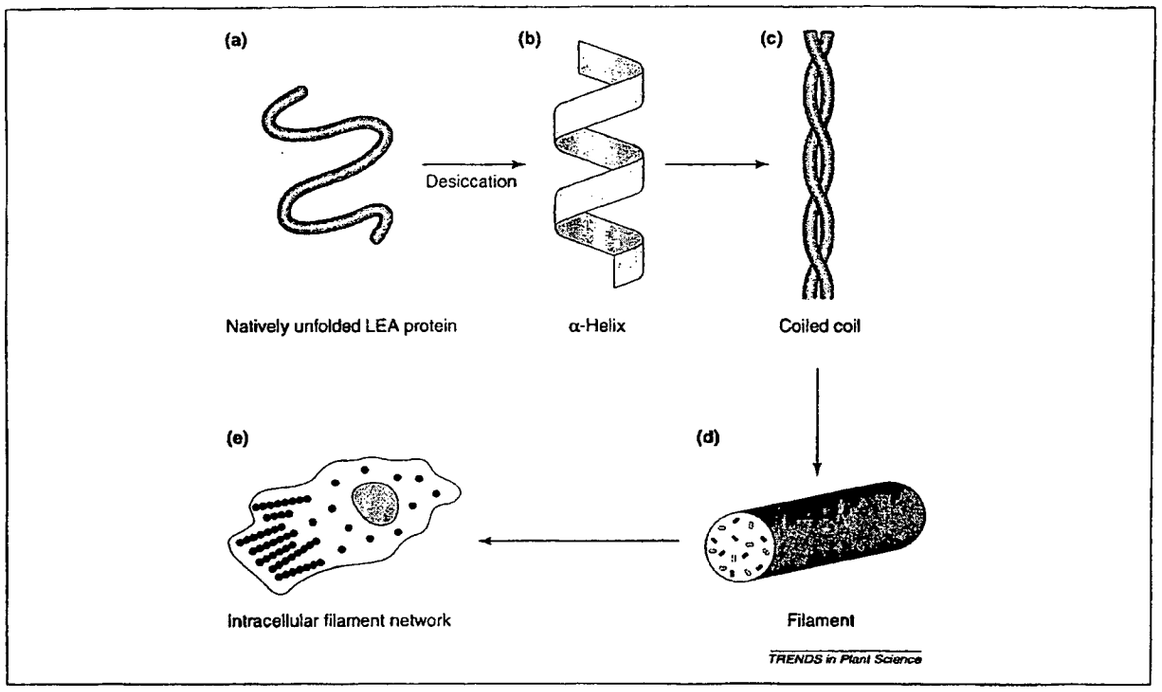


Figure 1. Hypothetical formation of intracellular filaments by Group 3a (SF2) LEA proteins, triggered by desiccation. A Group 3a (SF2) LEA protein from an anhydrobiotic nematode has been shown to shift from (a) an unfolded to (b) a more folded state on drying [17]. When dry, it exhibits a strong α-helical component and seems to form superhelical structures. The bioinformatics for this superfamily indicates similarity with proteins that form (c) coiled coils and (d) filaments associated with the cytoskeleton, including neurofilaments, which are a type of intermediate filament (Table 2). Therefore, one possibility is that Group 3a (SF2) LEA proteins also (e) generate filamentous structures within the desiccating cell, which might increase its mechanical strength in a similar fashion to intermediate filaments.

coiled-coil-like – structures. This is an extremely unusual observation because protein dehydration is more often associated with a loss of structure and aggregation [38] than with an increase in structure, folding or subunit assembly. It offers the exciting possibility that structural shifts in Group 3a (SF2) LEA proteins, and possibly other categories of LEA proteins, depend on the availability of water. In experiments using circular dichroism spectroscopy, other unstructured LEA (or similar) proteins have been shown to exhibit some increase in folding when water activity is decreased using, for example, trifluoroethanol [39–41] or a high salt concentration [42]. A partially characterized protein from *Typha latifolia* (bullrush) pollen, which is probably a Group 3 LEA protein, was also shown by FT-IR spectroscopy to form α-helical structures on drying [43].

Interestingly, these findings fit well with predictions arising from the novel bioinformatics, which suggest that Group 3a (SF2) LEA proteins are similar to other proteins that form part of the cytoskeleton (Figure 1). Through coiled-coil and superhelical filament formation, this suggests a cytoskeletal role for these proteins analogous to that of IFs. Although not found in plants, conventional IFs increase the mechanical strength of some metazoan cells. We thus suggest that Group 3a (SF2) LEA proteins might behave similarly in plant cells experiencing water loss, counteracting the physical stresses imposed during desiccation (which can lead to severe plasmolysis and cell deformation) [44]. If these proteins do indeed form filaments on desiccation, it should be possible to observe these by immunocytochemistry or other forms of microscopy.

Perspectives

POPP predictions of function for all the major LEA protein groups can be followed up in a similar fashion and offer a basis for improved experimental understanding of this intriguing family of proteins. The application of these novel bioinformatics tools to other protein families such as the set of glycine-rich plant proteins (originally thought to be cell-wall proteins but now known to be more broadly distributed [45]) is expected to contribute to the functional genomics programmes now under way, particularly when applied at the level of protein domains, as opposed to entire protein sequences. In general, the analysis of 'unusual' (generally, non-globular) proteins demands the deployment of novel software tools. We believe these tools offer opportunities for reasoned hypothesis formation relating to protein function, instead of the largely intuitive approach adopted previously.

Acknowledgements

M.W. is the Bristol-Myers Squibb Senior Research Fellow and A.T. is the AWG Senior Research Fellow, Pembroke College, Cambridge.

References

1 Dure, L. III *et al.* (1981) Developmental biochemistry of cottonseed embryogenesis and germination – changing messenger ribonucleic acid populations as shown by *in vitro* and *in vivo* protein synthesis. *Biochemistry* 20, 4162–4168

BNSDOCID: <XP_____2421537A__I_>

2 Grzelczak, Z.F. *et al.* (1982) Synthesis and turnover of proteins and messenger RNA in germinating wheat embryos. *Can. J. Biochem.* 60, 389–397
3 Galau, G.A. *et al.* (1986) Abscisic acid induction of cloned cotton late embryogenesis abundant (LEA) messenger RNAs. *Plant Mol. Biol.* 7, 155–170
4 Bray, E.A. (1993) Molecular responses to water deficit. *Plant Physiol.* 103, 1035–1040
5 Cuming, A.C. (1999) LEA proteins. In *Seed Proteins* (Shewry, P.R. and Casey, R., eds), pp. 753–780, Kluwer Academic
6 Bray, E.A. *et al.* (2000) Responses to abiotic stress. In *Biochemistry and Molecular Biology of Plants* (Buchanan, B.B. *et al.*, eds), pp. 1158–1203, American Society of Plant Physiologists
7 Roberts, J.K. *et al.* (1993) Cellular concentrations and uniformity of cell-type accumulation of two LEA proteins in cotton embryos. *Plant Cell* 5, 769–780
8 Hughes, D.W. and Galau, G.A. (1989) Temporally modular gene expression during cotyledon development. *Genes Dev.* 3, 358–369
9 Welin, B.V. *et al.* (1994) Characterization and differential expression of *dhn/lea/rab*-like genes during cold acclimation and drought stress in *Arabidopsis thaliana*. *Plant Mol. Biol.* 26, 131–144
10 Imai, R. *et al.* (1996) A LEA class gene of tomato confers salt and freezing tolerance when expressed in *Saccharomyces cerevisiae*. *Gene* 170, 243–248
11 Honjoh, K-I. *et al.* (1999) Introduction of the *hiC6* gene, which encodes a homologue of a late embryogenesis abundant (LEA) protein, enhances freezing tolerance of yeast. *J. Plant Physiol.* 155, 509–512
12 Swire-Clark, G.A. and Marcotte, W.R. Jr (1999) The wheat LEA protein Em functions as an osmoprotective molecule in *Saccharomyces cerevisiae*. *Plant Mol. Biol.* 39, 117–128
13 Zhang, L. *et al.* (2000) Expression of plant group 2 and group 3 LEA genes in *Saccharomyces cerevisiae* revealed functional divergence among LEA proteins. *J. Biochem.* 127, 611–616
14 Xu, D. *et al.* (1996) Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.* 110, 249–257
15 Sivamani, E. *et al.* (2000) Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. *Plant Sci.* 155, 1–9
16 Honjoh, K. *et al.* (2000) Cryoprotective activities of group 3 late embryogenesis abundant proteins from *Chlorella vulgaris* C-27. *Biosci. Biotechnol. Biochem.* 64, 1656–1663
17 Goyal, K. *et al.* (2003) Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *J. Biol. Chem.* 278, 12977–12984
18 Rost, B. (1996) PHD: predicting 1D protein structure by profile based neural networks. *Methods Enzymol.* 266, 525–539
19 Solomon, A. *et al.* (2000) Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat-stable protein. *Parasitology* 121, 409–416
20 Dure, L. III (2001) Occurrence of a repeating 11-mer amino acid sequence motif in diverse organisms. *Protein Pept. Lett.* 8, 115–122
21 Browne, J. *et al.* (2002) Plant desiccation gene found in a nematode. *Nature* 416, 38
22 Stacy, R.A.P. and Aalen, R.B. (1998) Identification of sequence homology between the internal hydrophilic repeated motifs of Group 1 late-embryogenesis-abundant proteins in plants and hydrophilic repeats of the general stress protein GsiB of *Bacillus subtilis*. *Planta* 206, 476–478
23 Battista, J.R. *et al.* (2001) Inactivation of two homologues of proteins presumed to be involved in the desiccation tolerance of plants sensitizes *Deinococcus radiodurans* R1 to desiccation. *Cryobiology* 43, 133–139
24 Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
25 Wootton, J.C. (1994) Sequences with 'unusual' amino acid compositions. *Curr. Opin. Struct. Biol.* 1144, 413–421
26 Wise, M.J. (2001) 0j.py: a software tool for low complexity proteins and protein domains. *Bioinformatics* 17, S288–S295
27 Altschul, S.F. *et al.* (1994) Issues in searching molecular sequence databases. *Nat. Genet.* 66, 119–129
28 Wise, M.J. (2002) The POPPs: clustering and searching using peptide probability profiles. *Bioinformatics* 18, S38–S45
29 Garay-Arroyo, A. *et al.* (2000) Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* 275, 5668–5674
30 Wise, M.J. (2003) LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles. *BMC Bioinformatics* 4, 52
31 Uversky, V.N. *et al.* (2000) Why are natively unfolded proteins unstructured under physiologic conditions? *Proteins* 41, 415–427
32 Dunker, A.K. *et al.* (2001) Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59
33 Dyson, H.J. and Wright, P.E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12, 54–60
34 Dunker, A.K. *et al.* (2002) Intrinsic disorder and protein function. *Biochemistry* 41, 6573–6582
35 Uversky, V.N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem.* 269, 2–12
36 Stellwagen, E. *et al.* (1972) The conformation of horse heart apocytochrome c. *J. Biol. Chem.* 247, 8074–8077
37 Renshaw, P.S. *et al.* (2002) Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6–CFP-10 complex. Implications for pathogenesis and virulence. *J. Biol. Chem.* 277, 21598–21603
38 Dong, A.C. *et al.* (1995) Infrared spectroscopic studies of lyophilization-induced and temperature-induced protein aggregation. *J. Pharmacol. Sci.* 84, 415–424
39 Soulages, J.L. *et al.* (2002) Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis abundant protein from soybean. *Plant Physiol.* 128, 822–832
40 Lisse, T. *et al.* (1996) The recombinant dehydrin-like desiccation stress protein from the resurrection plant *Craterostigma plantagineum* displays no defined three-dimensional structure in its native state. *Biol. Chem.* 377, 555–561
41 Boothe, J.G. *et al.* (1997) Purification, characterization, and structural analysis of a plant low-temperature-induced protein. *Plant Physiol.* 113, 367–376
42 Russouw, P.S. *et al.* (1995) Isolation and characterization of a heat-soluble protein from pea (*Pisum sativum*) embryos. *Seed Sci. Res.* 5, 137–144
43 Wolkers, W.F. *et al.* (2001) Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses *in vitro*. *Biochim. Biophys. Acta* 1544, 196–206
44 Wolfe, J. *et al.* (1986) Hydration, dehydration, and the stresses and strains in membranes. In *Membranes, Metabolism and Dry Organisms* (Leopold, A.C., ed.), pp. 286–305, Cornell University Press
45 Sachetto-Martins, G. *et al.* (2000) Plant glycine-rich proteins: a family or just proteins with a common motif? *Biochim. Biophys. Acta* 1492, 1–14

Letters to *Trends in Plant Science*

If you wish to comment on an article recently published in *Trends in Plant Science* or would like to discuss issues of general current interest to plant scientists, please write a Letter to the Editor. Letters should be no more than 750 words long with a maximum of 12 references and one small figure. Letters should be e-mailed to plants@current-trends.com.

The decision to publish rests with the Editor, and the author(s) of any *Trends in Plant Science* article criticized in a Letter will normally be invited to reply.

BNSDOCID: <XP_____2421537A_I_>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 December 2005 (29.12.2005)

PCT

(10) International Publication Number
WO 2005/122697 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/IL2005/000656

(22) International Filing Date: 20 June 2005 (20.06.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/580,726 21 June 2004 (21.06.2004) US

(71) Applicant *(for all designated States except US)*: **CARMEL-HAIFA UNIVERSITY ECONOMIC CORP. LTD** [IL/IL]; IBM Building, Haifa University, Mount Carmel, 31905 Haifa (IL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NEVO, Eviatar** [IL/IL]; Institute of Evolution, Haifa University, Mount Carmel, 31905 Haifa (IL). **SONG, Weining** [AU/IL]; Institute of Evolution, Haifa University, Mount Carmel, 31905 Haifa (IL). **DU, Xianghong** [AU/IL]; Institute of Evolution, Haifa University, Mount Carmel, 31905 Haifa (IL).

(74) Agent: **BEN-AMI, Paulina**; Ben-Ami & Associates, P.O. BOX 94, 76100 Rehovot (IL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/122697 A2

(54) **Title:** TRANSGENIC PLANTS CONTAINING A DEHYDRIN GENE

(57) **Abstract:** The present invention provides a new dehydrin gene isolated from wild barley of the Negev desert capable of improving tolerance to abiotic stresses such as, drought, cold/freezing or high salinity stress in plants, and transgenic plants transformed thereby.

TRANSGENIC PLANTS CONTAINING A DEHYDRIN GENE

FIELD OF THE INVENTION

The present invention relates to a new dehydrin gene, designated *Dhn1*, isolated from wild barley of the Negev desert, an extremely arid region, encoding a dehydrin 1 protein, designated DHN1, to vectors comprising it, and to transgenic plants containing said gene.

BACKGROUND OF THE INVENTION

Worldwide, millions of people are chronically undernourished while the projected growth of human population will add almost three billions more by the year of 2050 as estimated by the United Nations. Environmental stresses have been recognized as the most important limiting factor in world crop production. A wide range of physiological and molecular responses takes place when plants are subject to abiotic stresses such as drought, salinity and low temperature.

Dehydrins (DHNs) are water-soluble lipid-associating proteins that accumulate during low-temperature or water-deficit conditions, and are thought to play a role in freezing- and drought-tolerance in plants. DHNs are among the most abundant and most frequently observed proteins induced by abiotic stresses in plants, as well as exogenous abscisic acid (Close, 1996). DHNs are also known as Group II late embryogenesis abundant (LEA) proteins which appear to be evolutionarily conserved among photosynthetic organisms in addition to yeast (Borovskii et al., 2002; Campbell and Close, 1997; Close and Lammers, 1993; Li et al., 1998; Mtwisha et al., 1998). This family of proteins shares virtually no similarity with any enzymes or proteins of known function. Structurally, they are characterized by combinations of typical domains, with the K-segment, a lysine-rich 15-amino acid consensus sequence (EKKGIMDKIKEKPLG), as the most conserved in all plants whereas the S-segment is a phosphorylatable serine-rich

tract, the Y-segment an N-terminal conserved sequence and the Φ-segment repeated units rich in glycine and polar amino acids (Close, 1996, 1997). DHNs are supposed to form putative amphiphilic α-helices which are considered to bind to intracellular molecules, in particular membranes and proteins, primarily by hydrophobic interactions (Close, 1997). Detection of dehydrin can be made through antibodies prepared against the conserved sequences (Close et al., 1993). It has been shown that DHNs are associated with endomembrane systems and plasma membrane (Asghar et al., 1994; Egerton-Warburton et al., 1997; Danyluk et al. 1998) and maize DHN1 can bind to lipid vesicles that contain acidic phospholipids (Koag et al., 2003).

Although great progress has been achieved in the studies on DHNs in the last 20 years, their function has remained unclear, a situation similar to that of other LEA proteins (Wise and Tunnacliffe, 2004). Some effect of tomato le4 on improved stress tolerance was reported when introduced into yeast (Zhang et al., 2000). However, as a whole, overexpressing DHNs in plants has not improved stress tolerance significantly (Kaye et al., 1998; Welin, 1994). In contrast, engineering plant stress tolerance through regulons has fared much better compared with that through target genes like dehydrins (Jaglo-Ottosen et al., 1998; Liu et al., 1998; Kasuga et al., 1999).

Dhn genes exist as multi-gene families in plants. In cultivated barley (*Hordeum vulgare*), there are more than 12 *Dhn* genes identified and mapped in chromosomes 3H, 4H, 5H, and 6H in barley cv Dicktoo (Choï et al., 1999 and 2000). After a cluster of *Dhn* loci (*Dhn*1/2 and *Dhn*4a) was found to overlap the major quantitative trait locus (QTL) for winterhardiness on chromosome 7(5H) (Pan et al. 1994), several additional studies indicate that *Dhn* genes are associated with environmental stress tolerance (Campbell and Close, 1997). The studies of gene expression of *Dhn* in barley, which have been more comprehensive than other systems, reveal that the expression profiles of *Dhn* genes are highly variable from one *Dhn* gene to another, the majority are strongly drought-induced while others are cold-induced (Choi et al., 1999; Choi et al., 2000).

The loss of genetic diversity of some of the world's crops has accelerated alarmingly in recent decades with many crops becoming increasingly susceptible to diseases, pests and environmental stresses. Wild barley (*H. vulgare* ssp. *spontaneum*) is the progenitor of cultivated barley (*H. vulgare*) and is still widely distributed over the eastern Mediterranean rim and western Asia (Nevo, 1992). It occupies both primary and secondary man-made habitats and its distribution center lies in the Fertile Crescent (Badr et al., 2000). Its highest genetic diversity is displayed in Israel and Jordan despite the small area of these countries. In northern Israel and particularly the Eastern Galilee and Golan Heights, it comprises massive and continuously distributed central populations fading out to small, semi-isolated, and isolated populations in the xeric southern steppes and deserts (Nevo, 1992; Owuor et al., 1999; Turpeinen et al., 2001; Baek et al., 2003). Wild barley has, particularly with accessions from the desert regions, unique resistance to water stress (Nevo, 1992). As it can survive in severely drought-stressed environments, wild barley is a potential source of genetic variation to improve drought tolerance in elite barley cultivars. With the genomic sequencing advancing rapidly among crop species, gene transfer based on native or homologous genes from related lines and species are a more attractive option for crop improvement. Wild and cultivated barleys can provide a model system to investigate the molecular basis of drought tolerance as well as natural selection for adaptation to stressed environments.

Two groups pioneered manipulation of plants to increase their tolerance to environmental stresses such as cold, freezing, drought or high salinity, but both groups did it through engineering of regulatory elements, e.g. engineering of transcription factors, and not through structural genes coding for a protein.

The results of the first group (Thomashow et al., Michigan State University) are disclosed in several patents. US 5,891,859 and US 5,892,009 describe a gene, designated as CBF1, encoding a protein, CBF1, which binds to a region regulating expression of genes which promote cold temperature and dehydration tolerance in plants, and can be used to transform microorganisms and plants. US 6,417,428 discloses a plant comprising a recombinant molecule comprising a polynucleotide

that encodes a polypeptide that binds to a cold or dehydration transcription regulating region comprising the sequence CCG. US 6,706,866 discloses a non-naturally occurring binding protein comprising an amino acid sequence capable of binding to a CCG regulatory sequence and an amino acid sequence which forms a transcription activation region, and plants that overexpress said proteins and exhibit increased tolerance to environmental stresses such as cold, freezing, drought or high salinity.

The results of the second group (Shinozaki et al., Japan International Research Center) are also disclosed in several patents. US 6,495,742 and US 6,670,528 describe a transcription factor gene encoding a protein that regulates the transcription of genes located downstream of a stress responsive element, wherein said stress is dehydration stress, low temperature stress or salt stress, and transgenic plants transformed therewith, and transgenic plants transformed with a DNA coding for the *Arabidopsis Thaliana* transcription factor DERB1A, that binds to a stress responsive element.

SUMMARY OF THE INVENTION

In one aspect, the present invention relates to a dehydrin gene isolated from wild barley (accession 20-05), herein designated *Dhn 1*. SEQ ID NO:1 comprises the coding region of said *Dhn1* gene. This gene codes for the new dehydrin1 protein DHN1 of SEQ ID NO:2.

The present invention further relates to plants transformed with said *Dhn1* gene and overexpressing DHN1, said plants exhibiting tolerance to abiotic stresses, such as drought, cold/freezing and high salinity stresses.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 depicts the construct pNEdhn of the invention. 35S, CaMV35S promoter. Dhn1, *Dhn1* coding region. NOS, Nos terminator.

Fig. 2A shows a comparison of the amino acid sequences of different DHN1 proteins: the DHN1 of the invention (SEQ ID NO: 2) from wild barley (Genebank

of Institute of Evolution at Haifa University, accession number 20-05) and the DNH1s from the barley cultivars Dicktoo (4105103) (SEQ ID NO: 3), Himalaya137 (296198) (SEQ ID NO: 4), Himalaya139 (118483) (SEQ ID NO: 5), and Georgie139 (CAA66970) (SEQ ID NO: 6). Dashes indicate where a sequence has been expanded to allow optimal alignment. The arrow indicates the intron position. The Y-, S-, K- and Φ segments are in boldface. **Fig. 2B** depicts the cladogram of DHN1 sequences derived from Fig. 2A. Alignments of the amino acid sequences were carried out with the Clustal-W and Boxshade programs.

Figs. 3A-3C show the effects of drought stress on control and transgenic plants overexpressing DHN1. Plants were grown for 3 weeks with normal watering, withheld water for 5 weeks. Control and transgenic plants are shown under the same conditions. **Fig. 3A** shows wild-type plant without drought stress; **Fig. 3B** shows transgenic line N26 drought-stressed; **Fig. 3C** shows control plant drought-stressed.

Fig. 4 shows effects of severe drought stress and re-watering on transgenic plants overexpressing DHN1. Eight-week old plants with normal watering were transplanted to dry soils, re-watered 24 hours after transplanting, photograph was taken 36 hours after re-watering. Right, transgenic line N9. Left, control plant.

DETAILED DESCRIPTION OF THE INVENTION

In one aspect, the present invention relates to an isolated DNA having a nucleotide sequence coding for the dehydrin 1 (DHN1) protein of SEQ ID NO:2.

In preferred embodiments, the isolated DNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, isolated DNA having a nucleotide sequence differing from SEQ. ID NO: 1 in codon sequence due to the degeneracy of the genetic code, and cDNA consisting of the nucleotide sequence coding for the dehydrin 1 (DHN1) protein of SEQ ID NO:2.

The DNA of SEQ ID NO: 1 was obtained from genomic DNA and includes the coding region of the dehydrin 1 gene, herein designated *dhn1*, of the wild-type barley (Genebank of Institute of Evolution at Haifa University, Israel, accession

number 20-05) found in the northern Negev region of Israel. This DNA codes for the protein dehydrin 1, herein designated DHN1, of the amino acid sequence as shown in SEQ. ID NO: 2.

The present invention further relates to a chimeric DNA construct capable of expression in plant cells, comprising: (a) a DNA sequence coding for the DHN1 protein of SEQ ID NO: 2 such as cDNA coding for the DHN1 protein of SEQ ID NO: 2, and particularly SEQ ID NO: 1 and (b) DNA sequences enabling expression of the DHN1 protein in plant cells.

The invention still further relates to a recombinant vector comprising the DNA having a nucleotide sequence encoding for the dehydrin 1 (DHN1) protein of SEQ ID NO: 2 such as the DNA sequence in SEQ ID NO: 1 or cDNA coding for the DHN1 protein of SEQ ID NO: 2, in particular a recombinant expression vector comprising a chimeric gene construct in which the DNA of SEQ ID NO: 1 coding for the DHN1 protein of SEQ ID NO: 2, is operably linked to DNA sequences enabling expression of the DHN1 protein in plant cells.

The vector used to introduce the nucleic acid into the plant cell may be a plasmid, in which the DNA encoding the DHN1 protein is inserted into a unique restriction endonuclease cleavage site. The DNA is inserted into the vector using standard cloning procedures readily known in the art. This generally involves the use of restriction enzymes and DNA ligases, as described, for example, by Sambrook et al. (1989). The resulting plasmid which includes nucleic acid encoding the DHN1 protein can then be used to transform a plant cell (See generally, Gelvin and Schilperoort, 1994).

For plant transformation, the plasmid preferably also includes a selectable marker for plant transformation. Commonly used plant selectable markers include the kanamycin-resistance gene, kanamycin phosphotransferase II (nptII) gene, the hygromycin-resistance gene, hygromycin phosphotransferase (hpt) gene, the phosphinothricin acetyl transferase gene (bar), the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene, neomycin 3'-O-phosphotransferase (npt II) gene, or acetolactate synthase (ALS) gene. In preferred embodiments of the invention, the

selectable marker is the hpt or nptII gene, that allow selection of the transformants with hygromycin or kanamycin, respectively.

The plasmid may also include a reporter gene that, upon expression, provides a clear indication that genetic transformation did take place. Commonly used reporter genes are beta-glucuronidase (GUS), luciferase and green fluorescent protein (GFP). Reporter genes are often placed downstream of the promoter region and in the proximity of the gene of interest to ensure that they are expressed together and not separated by crossover events.

The plasmid preferably also includes suitable promoters for expression of the nucleic acid encoding the DHN1 protein and for expression of the selectable marker gene. The cauliflower mosaic virus 35S (35S CaMV) promoter is commonly used for plant transformation, as well as the rice actin 1 (Act1), the Ubiquitin 1 (Ubi1), the alpha-amylase gene promoter, and promoters of genes induced by stress. In one preferred embodiment, the promoter used in the present invention is the 35S CaMV promoter. Also the dhn1 promoter of the whole *dhn1* gene, of the sequence of SEQ ID NO:7, can be used in the present invention. In the plasmid, the nucleic acid encoding the DHN1 protein may be under the control of one promoter and the marker gene may be under control of the same or of a different promoter.

For plant transformation, the plasmid also preferably includes a nucleic acid molecule encoding a 3' terminator such as that from the 3' non-coding region of genes encoding a proteinase inhibitor, actin, or nopaline synthase (nos). In a preferred embodiment, the plasmid of the present invention includes the nopaline synthase (NOS) terminator.

In one preferred embodiment of the invention, the recombinant vector is represented by the construct of Fig. 1.

The plasmid is preferably a binary vector in which the genes of interest are inserted within the T-DNA borders. Examples of such vectors that can be used in the present invention are vectors obtainable from commercial sources such as the pCambia 1301, the pBI121, which contains a low-copy RK2 origin of replication, the neomycin phosphotransferase (nptII) marker gene with a nopaline synthase

(NOS) promoter and a NOS 3' polyadenylation signal, the pBI101 and functionally similar vectors described by Becker et al. (1992), and the pPZPY112 vector.

For the transformation of the plants, any method suitable for transformation of cereals can be used such as Agrobacterium-mediated transformation or particle bombardment (also known as biolistic transformation).

In the Agrobacterium-mediated transformation, plant cells are contacted with an inoculum of the bacteria transformed with the plasmid comprising the gene that encodes for the DHN1 protein of the invention, for example by inoculating the plant cells with a suspension of the transformed bacteria. Suitable bacteria from the genus Agrobacterium that can be utilized to transform plant cells include the species *Agrobacterium rhizogenes* and *Agrobacterium tumefaciens*, preferably *A. tumefaciens* strains LBA4404 or EHA105. *Agrobacterium* spp. are transformed with the plasmid by conventional methods well-known in the art.

Using the floral dip transformation protocol, the *A. tumefaciens* strain carrying the *Dhn1* gene in the binary vector is grown in a growth medium in the presence of antibiotic, e.g. kanamycin, to select for bacterial cells harboring the binary plasmid. The bacteria are spun down and resuspended in a 5% sucrose solution, followed by the addition of a surfactant, e.g. Silwet L-77. After mixing, the plants (after flowering) are dipped into the bacteria solution and kept under a dome or cover for 16-24 hours. The seeds are recovered and putative transformants are selected by plating the sterilized seeds on an antibiotic, e.g. hygromycin, and transplanting the putative transformants to soil. For higher rates of transformation, plants may be dipped two or three times at seven day intervals.

In cereals, the most commonly used tissue for transformation of wheat and barley has been the immature embryo from the developing grain (Weeks et al., 1993; Nehra et al., 1994; Becker et al., 1994; Wan and Lemaux, 1994; Barcelo and Lazzeri, 1995).

For the transformation of cereal plants, direct insertion of the DNA of interest may be carried out using the particle bombardment (biolistics) technique. According to this approach, rapidly propelled tungsten or gold microprojectiles

(which are smaller than the plant cells) coated with the DNA of interest are blasted into cells. After the target tissue is bombarded with the DNA-coated particles under vacuum, the DNA of interest disperses from the particles within the cells and then integrates into their genome (Harwood et al., 2000). In an alternative, the particle bombardment technique can be combined with the Agrobacterium technique to facilitate transformation (Tingay et al., 1997; Cheng et al., 1997)

The invention further relates to a transgenic plant transformed with a DNA coding for the DHN1 protein of SEQ ID NO: 2 such as SEQ ID NO: 1 or cDNA coding for the DHN1 protein of SEQ ID NO: 2, said DNA being operably linked to DNA sequences enabling expression of the DHN1 protein in plant cells and subsequent improvement of tolerance of the plant to abiotic stresses, such as drought, high salinity and cold/freezing stresses.

Thus, the invention includes a transgenic plant which contains in its cells a chimeric gene construct capable of expression in plant cells, comprising: (a) a DNA sequence of SEQ ID NO: 1 coding for the DHN1 protein of SEQ ID NO: 2, and (b) DNA sequences enabling expression of the DHN1 protein in plant cells.

The transgenic plant according to the invention include, without being limited to, cereals such as wheat, barley, corn, rice, oat, and forage and turf grasses, or any other food crop species such as beans, including soybeans, peas, tomatoes, and oilseed rape.

The invention will now be illustrated by the following non-limiting examples.

EXAMPLES

Materials and Methods

(i) Plant material. The wild barley accession 20-05 from Sde Boqer (Negev Highlands, Israel) was originally collected by Prof. Eviatar Nevo and maintained at the Institute of Evolution, University of Haifa (Israel). *Arabidopsis thaliana* ecotype Columbia was used for genetic transformation.

(ii) Isolation of genomic DNA. The Genomic DNA of wild barley was isolated using a method based on Weining and Langridge (1991) with modifications

(Song and Henry, 1995). DNA was isolated on a small scale from ~200 mg of young leaves. The leaves were grounded to a powder in microfuge tubes under liquid nitrogen. The powder was then mixed with 0.6 ml of extraction buffer (2% sarkosyl (sodium lauryl sarcosinate), 0.1 M Tris-HCl, 10 mM EDTA, pH8.0) and subsequently, with 0.6 ml of phenol/chloroform (1:1). The whole mixture was shaken for 20-30 s and the aqueous phase recovered after centrifugation. The phenol/chloroform extraction was repeated and the DNA precipitated with ethanol.

(iii) Plant Transformation. The putative open reading frame of the *Dhn1* gene (SEQ ID NO:1) in wild barley accession 20-05 was amplified by PCR using primers 5'-atcgagatctgttcgtacttcgtagcacc-3' (SEQ ID NO:8) and 5'-gctggtaaccacgacatatatcggaga-3' (SEQ ID NO:9). PCR conditions were as follows: 94°C for 3 min, followed by 37 cycles of 94°C for 30 s, 64°C for 30 s, 72°C for 1 min, and 72°C for 5 min. PCR products were cloned into an pGEM®-T Easy Vectors (Promega) and confirmed by sequence analysis. The *Dhn1* gene was excised using *Nco*I and *Bst* EII and the insert was ligated into the *Bam*HI-*Sac*I restricted pCambia 1301 vector between the CaMV 35S promoter and the nopaline synthase (NOS) terminator sequence of *A. tumefaciens*. The chimeric construct 35SCaMV-*Dhn1*-NOS **(Fig. 1)** was introduced into *A. tumefaciens* strain EHA105.

Arabidopsis thaliana plants were transformed according to the floral dip transformation protocol of Clough and Bent (1998) with *A. tumefaciens* carrying the construct. *A. tumefaciens* carrying construct pNEdhn **(Fig. 1,** 35SCaMV-Dhn1-NOS) was grown at 28°C in LB with kanamycin to OD_{600} = 0.8 and harvested in 5% sucrose solution with 0.05% (500 µl/L) Silwet L-77 before *Arabidopsis* plants were dipped. Selection of putative transformants was performed as described by Clough and Bent (1998). To select for transformants, sterilized seeds were suspended in 0.1% (w/v) sterile agarose and plated on 20 µg mL^{-1} hygromycin. Transgenic lines resistant to hygromycin were transferred to pots and moved into a greenhouse. A small leaf sample from these transgenic plants was used to isolate genomic DNA for monitoring the presence of wild barley *Dhn1* sequence by PCR (primer sequences: 5'-gacgagggatggccacaagactga-3' (SEQ ID NO:10); 5'-agtaacgcatggctgcggatgcta-

3'(SEQ ID NO:11). These lines were advanced to the T3 population and were found to be phenotypically uniform for kanamycin resistance and positive for PCR screening.

(iii) Southern and Northern Analysis: Southern- and northern-hybridizations were carried out with approximately 10 μg of genomic DNA and 30 μg of total RNA per lane. Genomic DNA was isolated from transgenic lines and wild-type *Arabidopsis* as described by Murray and Thompson (1980) and digested with *Hin*dIII, then transferred to a Hybond-N^+ nylon membrane (Amersham, Buckinghamshire, UK). Southern hybridizations (65°C) were carried out as described by Sambrook et al. (1989). DNA probes were labeled with [^{32}P]dCTP according to the manufacturer's instruction (Sigma, St. Louis). For Northern blot hybridization, total RNA was extracted from transgenic lines and wild-type Arabidopsis according to the method of Chomczynski and Sacci (1987). The probes and method of hybridization (55°C) are the same with Southern blots.

Example 1. Comparison of the amino acid sequences of different DHN1s in cultivated and wild barley.

The *Dhn1* gene coding region of wild barley was obtained from an accession (20-05) collected in Sede Boqer, located in the Negev desert in southern Israel, where the annual rain fall is normally less than 100 mm. The amino acid sequence predicted from *Dhn1* in 20-05 was compared with other DHN1s of barley cultivars: Dicktoo (Genebank Accession No. 4105103) (SEQ ID NO: 3), Himalaya137 (Genebank Accession No 296198) (SEQ ID NO: 4), Himalaya139 (Genebank Accession No 118483) (SEQ ID NO: 5), and Georgie139 (Genebank Accession No CAA66970) (SEQ ID NO: 6).

The results depicted in **Fig. 2A** indicate that the DNH1 of the wild barley (20-05) presents two amino-acid substitutions in comparison to the barley cultivars: one in the S segment (S replaced by P) and the other one in the Φ segment (T replaced by M). Overall, as shown in the cladogram of **Fig. 2B**, DHN1 in wild barley (20-05) resembles barley cv Himalaya137 most closely, both having two S

residues absent in the S segment (shown by the arrow) and a G substituted by D in the Y segment.

Example 2. Effect of drought stress on transgenic plants overexpressing DHN1.

Arabidopsis plants were transformed with the vector containing the *dhn1* gene under the control of the CaMV 35S promoter (pNEdh construct) by the floral dip transformation protocol as described in Materials and Methods. Preliminary data were obtained with T2 plants (selected on hygromycin resistance) with drought and salt stress.

To test the drought stress, plants (control and transgenic) were grown in a greenhouse for 3 weeks with normal watering, followed by withholding water for 5 weeks. As shown **in Fig. 3,** under drought stress, the transgenic plants overexpressing DHN1 (**Fig. 3B**) are more drought tolerant than wild-type (control) plants (**Fig. 3C**). The transgenic plant under drought stress behaved like the wild-type plant not subjected to drought stress (**Fig. 3A**). DHN1 seems to have a protective role as anthocyanins were accumulated in the leaf tissue of the wild-type plants under water stress (**Fig. 3C**). Moreover, the flowering time in wild-type plants were 15 days earlier than that of the transgenic plants under the same water stress conditions (**Fig. 3**).

To test the effect of salt during the vegetative stage, two-week-old seedlings growing in the soil were irrigated with 0, 50, 75, 100, 150, 200, and 250 mM NaCl, respectively. Similar results were also obtained with salt tolerance, in respect of the early flowering and anthocyanin accumulation in wild-type plants compared with that of the transgenic plants (data not shown).

Example 3. Effect of severe drought stress and re-watering on transgenic plants overexpressing DHN1.

In order to test the plants under more severe drought stress, the transgenic (line N9) and wild-type plants were transferred to dry soil to undergo even stronger water stress. Eight-week old plants with normal watering were transplanted to dry

soil and re-watered 24 hours after transplanting. The results are shown in **Fig. 4.** The photograph was taken 36 hours after re-watering. Both the transgenic (right) and the wild-type (left) plants wilted after transplanting. However, the N9 plant recovered and grew healthy following re-watering 24 hours after transplanting, whereas the wild-type plant (left) died consequently.

REFERENCES

Asghar R, Fenton RD, DeMason DA, Close TJ. Nuclear and cytoplasmic localization of maize embryo and aleurone dehydrin. *Protoplasma* 177:87-94(1994)

Badr A, Muller K, Schafer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F. On the origin and domestication history of Barley (Hordeum vulgare). *Mol Biol Evol.* 17(4):499-510 (2000).

Baek HJ, Beharav A, Nevo E. Ecological-genomic diversity of microsatellites in wild barley, Hordeum spontaneum, populations in Jordan. *Theor Appl Genet.* 106(3):397-410 (2003).

Barcelo P and Lazzeri PA. 'Transformation of cereals by microprojectile bombardment of immature inflorescence and scutellum tissues', in *Methods in Molecular Biology*, H. Jones, ed. 1995, Humana Press Inc.: Totowa, New Jersey, USA. pp. 113–22.

Becker D, Brettschneider R and Lorz H. Fertile transgenic wheat from microparticle bombardment of scutellar tissue. *Plant Journal* 5(2):299–307 (1994).

Becker D, Kemper E, Schell J, Masterson R. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol Biol* 20:1195-1197(1992).

Borovskii GB, Stupnikova IV, Antipina AI, Vladimirova SV and Voinikov VK. Accumulation of dehydrin-like proteins in the mitochondria of cereals in response to cold, freezing, drought and ABA treatment. *BMC Plant Biology.* 2:5 (2002).

Campbell SA, Close TJ. Dehydrins: genes, proteins, and associations with phenotypic traits. *New Phytol.* 137:61-74 (1997).

Cheng M, Fry JE, Pang S, Zhou H, Hironaka CM, Duncan DR, Conner TW and Wan Y, Genetic transformation of wheat mediated by *Agrobacterium tumefaciens. Plant Physiology* 115: 971–80 (1997).

Choi DW, Koag MC, Close TJ. Map location of barley Dhn gene determined by gene-specific PCR. *Theor Appl Genet.* 101: 350-354 (2000).

Choi DW, Zhu B, Close TJ. The barley (*Hordeum vulgare* L.) Dehydrin multigene family: sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv Dicktoo. *Theor Appl Genet* 98: 1234-1247 (1999).

Chomczynski P and Sacci N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162: 156-159 (1987).

Close TJ. Dehydrins: a commonalty in the response of plants in dehydration and low temperature. *Physiol Plant* 100: 291–296 (1997).

Close TJ. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiol Plant* 97:795-803 (1996).

Close TJ, Lammers PJ. An osmotic stress protein of cyanobacteria is immunologically related to plant dehydrins. *Plant Physiol.* 101:773-779 (1993).

Close TJ, Fenton RD, and F. Moonan. A view of plant dehydrins using antibodies specific to the carboxy terminal peptide. *Plant Mol Biol* 23:279-286 (1993).

Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16: 735-743(1998).

Cushman JC, Bohnert HJ. Genomic approaches to plant stress tolerance. *Current Opinion in Plant Biology*, 3:117-24 (2000).

Danyluk J, Perron A, Houde M, Limin A, Fowler B, Benhamou N, and F. Sarhan. Accumulation of an acidic dehydrin in the vicinity of the plasma membrane during cold acclimation of wheat. *Plant Cell.* 10:623-638 (1998).

Egerton-Warburton LM, Balsamo RA, Close TJ. Temporal accumulation and ultrastructural localization of dehydrins in *Zea mays* L. *Physiol Plant* 101: 545–555 (1997).

Gelvin SB and Schilperoort RA (Eds), Plant Molecular Biology Manual, 2nd Edition, Kluwer Academic Press, Dordrecht, Netherlands (1994).

Harwood WA, Ross SM, Cilento P, Snape JW. The effect of DNA/gold particle preparation technique, and particle bombardment device, on the transformation of barley (*Hordeum vulgare*). *Euphytica* 111: 67-76 (2000).

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF. Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* 280:104-6 (1998).

Kasuga M, Liu Q, Miura S,Yamaguchi-Shinozaki K and Shinozaki K. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nature Biotech* 17: 287-291 (1999).

Kaye C, Neven L, Hofig A, Li QB, Haskell D, and C. Guy. Characterization of a gene for spinach CAP160 and expression of two spinach cold-acclimation proteins in tobacco. *Plant Physiol* 116(4): 1367-1377 (1998).

Koag MC, Fenton RD, Wilkens S, and TJ Close. The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. *Plant Physiol.* 131: 309-316 (2003).

Li R, Brawley SH, Close TJ: Proteins immunologically related to dehydrins in fucoud algae. *J Phycol.* 34:642-650 (1998).

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, et al. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA-binding domain separate two cellular signal transduction pathways in drought- and low temperature-responsive gene expression, respectively, in Arabidopsis. *Plant Cell* 10:1391-406 (1998).

Mtwisha L, Brandt W, McCready S, Lindsey GG. HSP 12 is a LEA-like protein in Saccharomyces cerevisiae. *Plant Mol Biol.* 37:513-521(1998).

Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA *Nucleic Acids Res.* 8:4321-5 (1980).

Nehra NS, Chibbar RN, Leung N, Caswell K, Mallard C, Steinhauer L, Baga M and Kartha K. Self-fertile transgenic wheat plants regenerated from isolated scutellar tissues following microprojectile bombardment with two distinct gene constructs. *Plant Journal* 5(2): 285–97 (1994).

Nevo, E. Origin, evolution, population genetics and resources for breeding of wild barley, *Hordeum spontaneum*, in the Fertile Crescent. In: *Barley: Genetics, Mol Biol Biotech*, Shewry, P. (ed.). C.A.B. International, pp. 19-43. (1992).

Pan A, Hayes PM, Chen F, Chen THH, Blake T, Wright S, Karsai I, and Z. Bedo. Genetic analysis of the components of winter hardiness in barley (*Hordeum vulgare* L.) *Theor Appl Genet* 89: 900-910 (1994).

Sambrook J, Fritsch FF and Maniatis T. Molecular Cloning: A Laboratory Manual, 2d edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

Song W and Henry RJ. Molecular analysis of the DNA polymorphism of wild barley *(Hordeum spontaneum)* germplasm by the polymerase chain reaction. *Genetic Resources and Crop Evolution.* 42: 273-281(1995).

Tingay S, McElroy D, Kalla R, Fieg S, Wang M, Thornton S and Brettel R, *Agrobacterium tumefaciens*-mediated barley transformation. *Plant Journal* 11: 1369–76 (1997).

Turpeinen T, Tenhola T, Manninen O, Nevo E, Nissila E. Microsatellite diversity associated with ecological factors in Hordeum spontaneum populations in Israel. *Mol Ecol.* 0(6):1577-91 (2001).

Wan Y and Lemaux PG. Generation of large numbers of independently transformed fertile barley plants. *Plant Physiology* 104: 37–48 (1994).

Weeks JT, Andersen OD and Blechl AE, Rapid production of multiple independent lines of fertile transgenic wheat (*Triticum aestivum* L.). *Plant Physiology* 102(4): 1077–84 (1993).

Welin BV, Olson A, Nylander M and ET Palva. Characterization and Differential Expression of Dhn/Lea/Rab- Like Genes During Cold-Acclimation and Drought Stress in *Arabidopsis thaliana. Plant Mol. Biol.* 26(1): 131-144 (1994).

Weining S and Langridge P. Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction. *Theor Appl Genet* 82: 209-216 (1991).

Wise MJ and Tunnacliffe A. POPP the question: what *do* LEA proteins do? *Trends in Plant Science*, 9, 13-17 (2004).

Zhang L, Ohta A, Takagi M, Imai R. Expression of plant group 2 and group 3 LEA genes in *Saccharomyces cerevisiae* revealed functional divergence among LEA proteins. *J. Biochem.* 127: 611-616 (2000).

Zhu B, Choi DW, Fenton R, and Close TJ. Expression of the barley dehydrin multigene family and the development of freezing tolerance. *Mol. Gen. Genet.* 264:145-153 (2000).

317

CLAIMS

1. An isolated DNA having a nucleotide sequence coding for the dehydrin 1 (DHN1) protein of SEQ ID NO:2.

2. An isolated DNA according to claim 1, having a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 and a DNA having a nucleotide sequence differing from SEQ. ID NO: 1 in codon sequence due to the degeneracy of the genetic code.

3. An isolated DNA according to claim 2 comprising a DNA consisting of the nucleotide sequence of SEQ ID NO: 1, said DNA comprising the coding region of the dehydrin1 gene (*dhn1*) coding for the dehydrin 1(DHN1) protein of SEQ ID NO: 2.

4. An isolated DNA according to claim 1 comprising a cDNA consisting of the nucleotide sequence coding for the dehydrin 1(DHN1) protein of SEQ ID NO: 2.

5. A chimeric DNA construct capable of expression in plant cells, comprising: (a) a DNA sequence of SEQ ID NO: 1 coding for the DHN1 protein of SEQ ID NO: 2 or a cDNA coding for the DHN1 protein of SEQ ID NO: 2, and (b) DNA sequences enabling expression of the DHN1 protein in plant cells.

6. A chimeric DNA construct according to claim 5, comprising (a) a DNA sequence of SEQ ID NO: 1 and (b) DNA sequences enabling expression of the DHN1 protein in plant cells.

7. A recombinant vector comprising the DNA according to any one of claims 1 to 4. or the chimeric DNA construct of claim 5 or 6.

8. An expression vector comprising the chimeric DNA construct of claim 5 or 6.

9. A transgenic plant transformed with the DNA of SEQ ID NO: 1 or cDNA coding for the protein DHN1 of SEQ ID NO: 2, said DNA being operably linked to DNA sequences enabling expression of the DHN1 protein in plant cells and subsequent improvement of tolerance of the plant to abiotic stresses.

10. A transgenic plant which contains in its cells a chimeric DNA according to claim 6 such that the plant exhibits tolerance to abiotic stresses.

11. A transgenic plant according to claim 9 or 10 which is a cereal plant.

12. A transgenic plant according to claim 11 wherein said cereal is barley.

13. A transgenic plant according to any one of claims 9 to 12 wherein said abiotic stress is drought, cold/freezing or high salinity stress.



Fig. 1



Fig. 2A

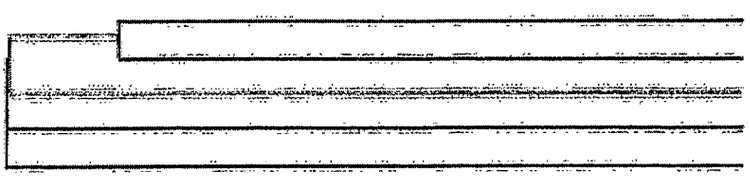


Fig. 2B

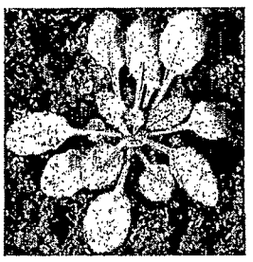


Fig. 3A

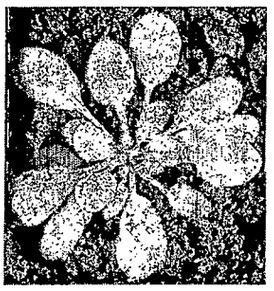


Fig. 3B

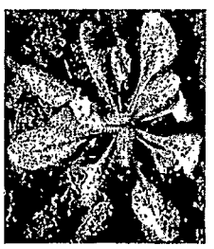


Fig. 3C

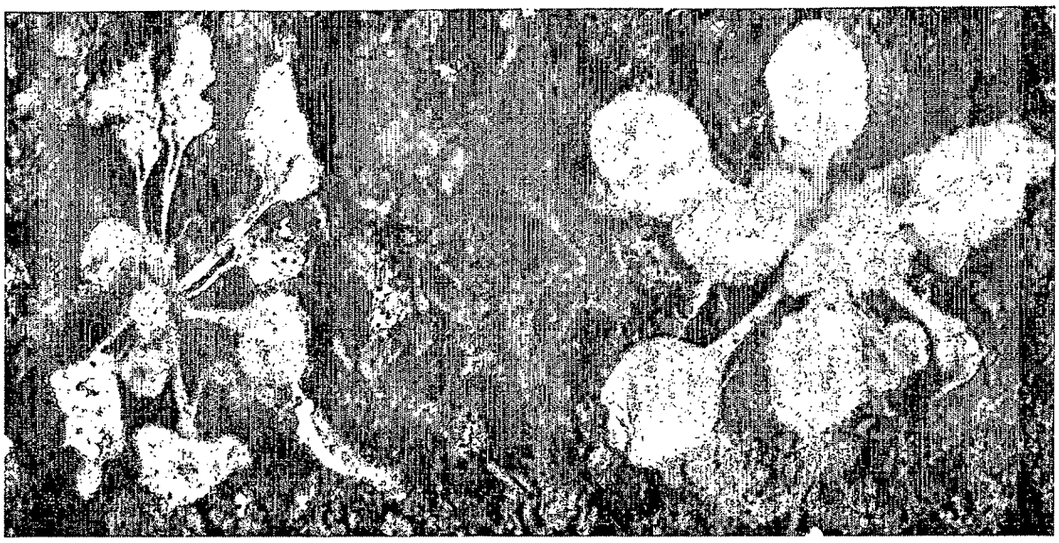


Fig. 4

SEQUENCE LISTING

<110> Carmel-Haifa University Economic Corp. Ltd
NEVO, Eviatar
SONG, Weining
DU, Xianghong

<120> TRANSGENIC PLANTS CONTAINING A DEHYDRIN GENE

<130> CARM-001 PCT

<150> US 60/580,726
<151> 2004-06-21

<160> 11

<170> PatentIn version 3.1

<210> 1
<211> 662
<212> DNA
<213> Wild barley (accession 20-05) Dhn1 coding region

<400> 1
gattatcgag atctgttcgt acttcgtagc accacaatgg agtaccaggg tcagcacggt 60

cacgccaccg acaaggtgga ggagtacggc cagcccgtgg ccgggcacgg cggtttcacc 120

gacgggccca cggggacgca cggagctgcc ggcgtcggcg gcgcgcagct ccaggcgacg 180

agggatggcc acaagactga cggcgtcctt cgacgctccg gcagccccag ctccgtaagc 240

tcccgggcag atctgtcgct gcatatattg agttgtacat gtcgctccac tctatcatct 300

taacttgctg tgcctgtgta cgtacgtatg tagtctgagg acgacggcgt gggcgggagg 360

aggaagaagg ggatgaagga gaagatcaag gagaagctcc ccggaggagc ccacaaggac 420

gccgcagggc agcagcagca gacggcgatg gcgggcgagt acgcgggcac acacggcatg 480

gaggccaccg gcgagaagaa gggcgtcatg gacaagatca aggagaagct tcccggcgga 540

cagcactgag ccagccatgt gtgtcttctc tatgttccga cgaagctcgg ctagctagca 600

tccgcagcca tgcgttacta cagataaata atggactctc cgatatatgt cgtggttacc 660

ag 662

<210> 2
<211> 137
<212> PRT
<213> Wild barely

<400> 2

Met Glu Tyr Gln Gly Gln His Gly His Ala Thr Asp Lys Val Glu Glu
1 5 10 15

Tyr Gly Gln Pro Val Ala Gly His Gly Gly Phe Thr Asp Gly Pro Thr
20 25 30

Gly Thr His Gly Ala Ala Gly Val Gly Gly Ala Gln Leu Gln Ala Thr
35 40 45

Arg Asp Gly His Lys Thr Asp Gly Val Leu Arg Arg Ser Gly Ser Pro
50 55 60

Ser Ser Ser Glu Asp Asp Gly Val Gly Gly Arg Arg Lys Lys Gly Met
65 70 75 80

Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly Ala His Lys Asp Ala
85 90 95

Ala Gly Gln Gln Gln Gln Thr Ala Met Ala Gly Glu Tyr Ala Gly Thr
100 105 110

His Gly Met Glu Ala Thr Gly Glu Lys Lys Gly Val Met Asp Lys Ile
115 120 125

Lys Glu Lys Leu Pro Gly Gly Gln His
130 135

<210> 3
<211> 139
<212> PRT
<213> Dicktoo (Genebank Accession No.4105103)

<400> 3

Met Glu Tyr Gln Gly Gln His Gly His Ala Thr Asp Lys Val Glu Glu
1 5 10 15

Tyr Gly Gln Pro Val Ala Gly His Gly Gly Phe Thr Gly Gly Pro Thr
20 25 30

Gly Thr His Gly Ala Ala Gly Val Gly Gly Ala Gln Leu Gln Ala Thr
35 40 45

Arg Asp Gly His Lys Thr Asp Gly Val Leu Arg Arg Ser Gly Ser Ser
50 55 60

Ser Ser Ser Ser Ser Glu Asp Asp Gly Val Gly Gly Arg Arg Lys Lys
65 70 75 80

Gly Met Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly Ala His Lys
85 90 95

Asp Ala Ala Gly Gln Gln Gln Gln Thr Ala Met Ala Gly Glu Tyr Ala
100 105 110

Gly Thr His Gly Thr Glu Ala Thr Gly Glu Lys Lys Gly Val Met Asp
115 120 125

Lys Ile Lys Glu Lys Leu Pro Gly Gly Gln His
130 135

<210> 4
<211> 137
<212> PRT
<213> Himalaya137 (Genebank Accession No.296198)

<400> 4

Met Glu Tyr Gln Gly Gln His Gly His Ala Thr Asp Lys Val Glu Glu
1 5 10 15

Tyr Gly Gln Pro Val Ala Gly His Gly Gly Phe Thr Asp Gly Pro Thr
20 25 30

Gly Thr His Gly Ala Ala Gly Val Gly Gly Ala Gln Leu Gln Ala Thr
35 40 45

Arg Asp Gly His Lys Thr Asp Gly Val Leu Arg Arg Ser Gly Ser Ser
50 55 60

Ser Ser Ser Glu Asp Asp Gly Val Gly Gly Arg Arg Lys Lys Gly Met
65 70 75 80

Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly Ala His Lys Asp Ala
85 90 95

Ala Gly Gln Gln Gln Gln Thr Ala Met Ala Gly Glu Tyr Ala Gly Thr
100 105 110

His Gly Thr Glu Ala Thr Gly Glu Lys Lys Gly Val Met Asp Lys Ile
115 120 125

Lys Glu Lys Leu Pro Gly Gly Gln His
130 135

<210> 5
<211> 139
<212> PRT
<213> Himalaya139 (Genebank Accession No.118483)

<400> 5

Met Glu Tyr Gln Gly Gln His Gly His Ala Thr Asp Lys Val Glu Glu
1 5 10 15

Tyr Gly Gln Pro Val Ala Gly His Gly Gly Phe Thr Gly Gly Pro Thr
20 25 30

Gly Thr His Gly Ala Ala Gly Val Gly Gly Ala Gln Leu Gln Ala Thr
35 40 45

Arg Asp Gly His Lys Thr Asp Gly Val Leu Arg Arg Ser Gly Ser Ser
50 55 60

Ser Ser Ser Ser Ser Glu Asp Asp Gly Val Gly Gly Arg Arg Lys Lys
65 70 75 80

Gly Met Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly Ala His Lys
85 90 95

Asp Ala Ala Gly Gln Gln Gln Gln Thr Ala Met Ala Gly Glu Tyr Ala

100 105 110

Gly Thr His Gly Thr Glu Ala Thr Gly Glu Lys Lys Gly Val Met Asp
115 120 125

Lys Ile Lys Glu Lys Leu Pro Gly Gly Gln His
130 135

<210> 6
<211> 139
<212> PRT
<213> Georgie139 (Genebank Accession No.CAA66970)

<400> 6

Met Glu Tyr Gln Gly Gln His Gly His Cys Thr Asp Lys Val Arg Ser
1 5 10 15

Thr Gly Gln Pro Val Ala Gly His Gly Phe Thr Gly Gly Pro Thr Gly
20 25 30

Thr His Gly Ala Ala Gly Val Gly Gly Ala Gln Leu Gln Ala Thr Arg
35 40 45

Asp Gly His Lys Thr Asp Gly Val Leu Arg Arg Ser Gly Ser Ser Ser
50 55 60

Ser Ser Ser Ser Glu Asp Asp Gly Val Gly Gly Arg Arg Lys Lys Gly
65 70 75 80

Met Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly Ala His Lys Asp
85 90 95

Ala Arg Ala Gly Gln Gln Gln Gln Thr Ala Met Ala Gly Glu Ile Ala
100 105 110

Gly Thr His Gly Thr Glu Ala Thr Gly Glu Lys Lys Gly Val Met Asp
115 120 125

Lys Ile Lys Glu Lys Leu Pro Gly Gly Gln His
130 135

<210> 7
<211> 1089
<212> DNA
<213> Wild barley (accession 20-05) Dhn1 promoter

<400> 7
gacaagctta gtgcaagcat cgtggcaact tttagcccgt aaaaaagtcg ttgaaacata 60

tcaaattatg atctagtttt gaaaatcttg tcgagacgga ttaatgatga aaatgtattt 120

ttaatcgaat tttttattta agagataaat atttgaagtc caaaaaaatc cgttgatgcc 180

atccgtttga tgtggcaaga tggatgataa cgaagacgtt tggaccatat tgctttgtac 240

cacatgtgcg acacttatta tttgggtttt gcaatgggtc ggcgtttaga cgaggtgtta 300

agtgtttaat gaacaatctt tctatgcatt gattggtagt cattttgtac aggtccgtat 360

atttagcatc gttttttctg gatcatcata atatcctctc ttctttttaa ttgatttgtc 420

atatcactat tttacttatg tgagacactc ttagcgcatc tacgcctagg tctactggga 480

agatgtaatc ttgatcagtt tggtcgtttt gtgtgtttgt cagttgtgat ctagatcagt 540

cgatgatttt ctttttcttt ttttaatttc aaagatggac tgacgtcttg atccttcgtt 600

ttaaaaaagt tagctaattc aaacatggcg ttcatttatc cagagaaaaa ggtgaatttt 660

gcgcctacgg gtcgaggagc aacgaagttt ccacgtagca cgcacgctgt accgaacatc 720

tctattcatg caggaatgcg tggatggagt agtcatgccc ttcgagacag accgaagacg 780

tacgtggcgc cccagacgag gccgcgtggc atgcatgccg acgttgttta ttggcagtag 840

atttatttgc ccacttcact ccaccataat ggaaccccat acatgcgcgc acacgtactc 900

cttccacctg caaagccagc gcatcccgaa ctcgaacaca cttacacagc cagcattata 960

taaataacca tatagctttg cacttctctc cgtcgcagtt cgcaacagat cagcacactt 1020

ccacagccac acgtataatc tgtagatcga tcactgttcc tacttcgtag caccacaatg 1080

gtagatctg 1089

<210> 8
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 8
atcgagatct gttcgtactt cgtagcacc 29

<210> 9
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 9
gctggtaacc acgacatata tcggaga 27

<210> 10
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 10
gacgagggat ggccacaaga ctga 24

<210> 11
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 11
agtaacgcat ggctgcggat gcta 24

XP 019322122

Theor Appl Genet (2005) 112: 114–130
DOI 10.1007/s00122-005-0112-2

ORIGINAL PAPER

Chenwei Lin · Lukas A. Mueller · James Mc Carthy
Dominique Crouzillat · Vincent Pétiard
Steven D. Tanksley

Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts

Received: 23 June 2005 / Accepted: 10 September 2005 / Published online: 5 November 2005
© Springer-Verlag 2005

Abstract An EST database has been generated for coffee based on sequences from approximately 47,000 cDNA clones derived from five different stages/tissues, with a special focus on developing seeds. When computationally assembled, these sequences correspond to 13,175 unigenes, which were analyzed with respect to functional annotation, expression profile and evolution. Compared with *Arabidopsis*, the coffee unigenes encode a higher proportion of proteins related to protein modification/turnover and metabolism—an observation that may explain the high diversity of metabolites found in coffee and related species. Several gene families were found to be either expanded or unique to coffee when compared with *Arabidopsis*. A high proportion of these families encode proteins assigned to functions related to disease resistance. Such families may have expanded and evolved rapidly under the intense pathogen pressure experienced by a tropical, perennial species like coffee. Finally, the coffee gene repertoire was compared with that of *Arabidopsis* and Solanaceous species (e.g. tomato). Unlike *Arabidopsis*, tomato has a nearly perfect gene-for-gene match with coffee. These results are consistent with the facts that coffee and tomato have a similar genome size, chromosome karyotype (tomato, $n=12$; coffee $n=11$) and chromosome architecture. Moreover, both belong to the Asterid I clade of dicot plant families. Thus, the biology of coffee (family Rubiacaeae) and tomato (family Solanaceae) may be united into one common network of shared discoveries, resources and information.

Communicated by R. Hagemann

C. Lin · L. A. Mueller · S. D. Tanksley (✉)
Department of Plant Breeding and Genetics, Department of Plant Biology, Cornell University, Ithaca, NY, 14853 USA
E-mail: sdt4@cornell.edu
Fax: +1-607-2556683

J. M. Carthy · D. Crouzillat · V. Pétiard
Nestlé Research Center, Tours, 101, Avenue Gustave Eiffel, 49716, 37097 Tours Cedex 2, France

Keywords *Coffea canephora* · Rubiaceae · Solanaceae · Seed development · Comparative genomics

Introduction

Coffee is an important international commodity, ranking among the five most valuable agricultural exports from developing countries (Food and Agriculture Organization, http://apps.fao.org). Moreover, production and processing of coffee employs more than 25 million people worldwide (O'Brien and Kinnaird 2003). Despite its economic importance, coffee has received little attention with respect to molecular genetics and genomics research. As of December 2004, only 1,570 nucleotide and 115 protein sequences from coffee had been deposited in GenBank with the majority of those sequences derived from leaf ESTs. Many of the remaining sequences correspond to enzymes in the caffeine biosynthesis pathway—the most extensively studied pathway in coffee (Moisyadi et al. 1998; Ogawa et al. 2001; Mizuno et al. 2003; Uefuji et al. 2003).

Commercial coffee production relies mainly on two closely related species: *Coffea arabica* and *Coffea canephora*, accounting for approximately 70 and 30% of worldwide coffee production, respectively (Herrera et al. 2002). Although *C. canephora* accounts for a lower total proportion of the coffee market than does *C. arabica*, it is the main source for soluble coffee, which is consumed widely throughout the world. *C. canephora* is a diploid ($2n=2x=22$), outcrossing and highly polymorphic species native to central Africa, but which has expanded, through cultivation, especially to western Africa, Indonesia and Vietnam (Wrigley 1988). In contrast, *C. arabica* is believed to be a recently derived tetraploid ($2n=4x=44$) native to a small region of what is now Ethiopia. *C. arabica* is now grown widely throughout the world.

The goal of the current project was to increase the genetic and molecular knowledge of coffee through the generation and annotation of an EST database using

BNSDOCID: <XP_____19322122A_I_>

high throughput single-pass 5′ sequencing of cDNAs derived from leaf, pericarp and seed tissues from a set of *C. canephora* varieties. Special emphasis was given to sequencing cDNAs from different stages of seed development, both to shed light on this important, but not well understood aspect of plant development and to capture as many genes as possible involved in determining the final chemical composition of seeds which constitute the commercial product. As a result, the EST database reported herein is, to our knowledge, the largest public database of seed-derived ESTs (White et al. 2000; Suh et al. 2003).

The EST database was used to derive a coffee unigene build, which was subsequently subjected to functional annotation. The coffee unigene build was compared with the model species *Arabidopsis*. *Arabidopsis* is the only fully sequenced dicot genome and hence provides a complete set of predicted genes against which to compare the EST-derived unigene set of coffee. However, coffee and *Arabidopsis* belong to different plant families (Rubiaceae and Brassicaceae, respectively) which are distantly related phylogenetically and which diverged from their last common ancestor approximately 94 million years ago (MYA) (Gandolfo et al. 1998; Crepet et al. 2004) (Fig. 1).

The plant family most closely related to coffee in which extensive sequencing has been conducted is Solanaceae (Fig. 1). In this family, comprehensive EST databases have been developed for tomato, potato, pepper, eggplant and petunia (http://www.sgn.cornell.edu/) (Hoeven et al. 2002; Ronning and Stegalkina 2003; Lee et al. 2004). Both Rubiaceae and Solanaceae belong to the Asterid I clade of dicots, and based on existing fossil evidence, are thought to have diverged from one another approximately 50 MYA (Gandolfo et al. 1998; Crepet et al. 2004) (Crepet personal communication) (Fig. 1). The closer taxonomic affinities of coffee and Solanaceae (e.g. tomato) are paralleled by a number of striking botanic and genetic similarities, including the production of fleshy berries, a similar genome content ($C = 950$ and 640 Mb for tomato and coffee, respectively) (Hoeven et al. 2002), similar basic chromosome number ($x = 12$ for tomato and most other Solanaceae; $x = 11$ for coffee) and similar chromosome architecture with highly condensed pericentric heterochromatin and decondensed euchromatin at the pachytene stage of meiosis (Rick 1971; Pinto-Maglio and Cruz 1998). For these reasons, the coffee unigene set was also compared against a series of Solanaceae EST-derived unigene sets.

Materials and methods

Library construction

Source of tissues

C. canephora was farm-grown in east Java. Plant tissue was frozen in liquid nitrogen, transported on dry ice and

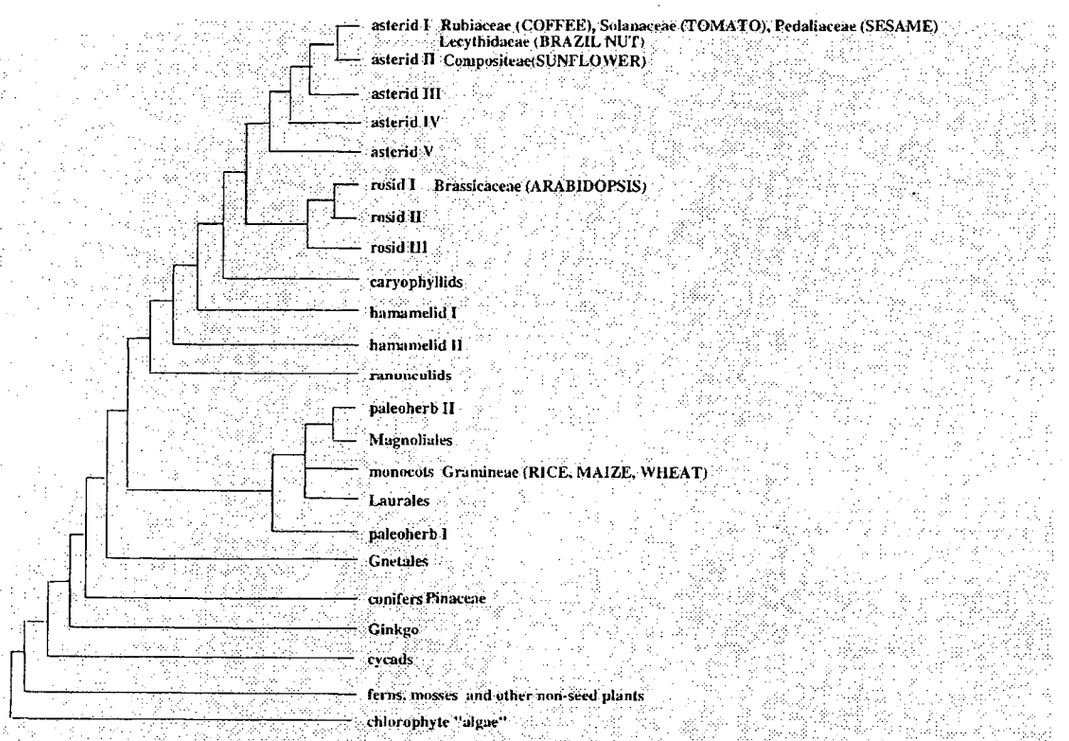


Fig. 1 Dendrogram depicting phylogenetic relationships of coffee to other higher plant taxa (based on Chase et al. 1993)

BNSDOCID: <XP_____19322122A__I_>

stored at –80°C until RNA extraction. The specified tissues (leaf, pericarp, young cherries with seed and pericarp mixed, middle stage seed and late stage seed) were collected from five different varieties Table 1). The maturation period of the *C. canephora* varieties under study is approximately 9–11 months, from pollination to ripening. The early stage cherry, middle stage seed and late stage seed were collected between 18 to 22 weeks, 30 weeks and 42 to 46 weeks after pollination, respectively.

RNA and mRNA isolation

Total RNA was extracted using phenol/chloroform (Rogers et al. 1999) and further treated with DNase I (RNase-free) and purified using an RNeasy Kit (Qiagen, Valencia, CA 91355). Messenger RNA was extracted from total RNA with PolyTrack mRNA Isolation Systems (Promega, Madison, WI 53711).

cDNA libraries

Directional cDNA libraries were constructed with 3–5 μg of mRNA with the ZAP-cDNA Gigapack III Gold Cloning Kit (Stratagene, La Jolla, CA 92037). The average insert length was estimated by PCR in 36 randomly selected clones from each library and ranged 1.2–1.5 kb (Table 1).

Sequencing

Bacteria, containing coffee cDNAs, were cultured in 384-well plates and cDNA inserts subjected to 5′ end sequencing at the BioResource Center at Cornell University (http://www.brc.cornell.edu). The average size of quality reads was 613 bp with a maximum of 1,037 bp.

Sequence quality processing

EST sequences were base-called and screened for vector sequences using PHRED software (Ewing et al. 1998). The longest stretch of overall high quality (PHRED score over 15 which corresponds to over 98% confidence) of each sequence was identified. PolyA repeats were trimmed to at most 20 bp and any sequence past the PolyA (mostly low quality sequence) was discarded. After the trimming, the sequences were screened against the *E.coli* K12 genome to remove any bacteria contamination. The remaining sequences were screened for minimum length (150 bp) and maximum allowed ambiguity (4%) and low complexity (60% of the sequence are of the same nucleotide, or 80% of same two nucleotides, which indicate error in sequencing).

Unigene assembly

Unigene sets were built by combining the sequences from all five coffee cDNA libraries. Clustering was performed using a program developed at the Sol Genomics Network (SGN at http://www.sgn.cornell.edu), which relied on a custom pre-clustering algorithm, and on the CAP3 program for contig generation (Huang and Madan 1999). The preclustering algorithm clustered sequences using a Smith Waterman type algorithm with initial word matching. The command line settings for CAP3 were as follows: -e 5000 -p 90 -d 10000 -b 60. The -e, -d and -b options are set such that the assembler disregards them or minimizes their effect. The -p option increases the sequence identity necessary on overlaps to 90 from a default of 75, which were found to be not stringent enough. Sequences were also checked for length, complexity and contamination. The builds were uploaded to the database, where each unigene was assigned a unique unigene ID.

Annotation

Protein prediction

The most likely coding frame and the corresponding transcribed peptide for each unigene were generated using ESTScan, a program that detects coding regions in EST sequences and corrects some sequencing errors (nucleotide substitution, deletion/insertion, erroneous stop codon, etc) (Iseli et al. 1999). As a hidden Markov

Table 1 Characteristics of the 5 cDNA libraries used to develop the coffee EST database

Library name	Tissue	Varieties	Average insert size, kb	Good quality ESTs
Leaf	Leaves, young	BP409	1.5 ± 0.6	8,942
Pericarp	Pericarp, all developmental stages	BP358, BP409, BP42, BP961, Q121	1.4 ± 0.5	8,956
Early stage cherry	Whole cherries, 18 and 22 week after pollination	BP358, BP409, BP42, Q121	1.4 ± 0.3	9,843
Middle stage seed	Endosperm and perisperm of seeds, 30 week after pollination	BP409, BP961, Q121	1.4 ± 0.3	10,077
Late stage seed	Endosperm and perisperm of seeds, 42 and 46 week after pollination	BP358, BP409, BP42, BP961, Q121	1.4 ± 0.3	9,096

BNSDOCID: <XP_____19322122A__I_>

Model (HMM) based program, ESTScan requires a training set of high quality coding sequences to generate a model. However, very few high quality coffee coding sequences are publicly available (less than 50 full length in GenBank). Therefore, we used as a training set of 483 nuclear genes from the closely related species, tomato (http://www.ebi.ac.uk/embl). Table 2 compares the coffee unigene assembly and translation (using ESTScan) with that of tomato, using the same training tomato set and parameters. The fact that the two give such similar results, gives confidence that the coffee EST-derived unigenes are being efficiently translated via the ESTScan software using the tomato training set (Table 2).

BLAST matches between coffee unigenes and other sequence databases

Sequence matches for coffee unigenes against other sequence databases were obtained using BLAST (Altschul et al. 1997) (see Results and discussion section for details of each BLAST analysis). The target databases were:

1. The GenBank non-redundant (NR) protein and dbest dataset (NCBI, http://www.ncbi.nlm.nih.gov).
2. The predicted *Arabidopsis* proteome and cDNA databases which were downloaded from the *Arabidopsis* Information Resource (TAIR, http://www.arabidopsis.org and ftp://ftp.arabidopsis.org/home/tair/).
3. The Solanaceae EST-derived unigene sets, including tomato (*Solanum. lycopersicum*) (184,860 ESTs, 30,576 unigenes), potato (*Solanum. tuberosum*) (97,425 ESTs, 24,932 unigenes), pepper (*Capsicum annuum*) (20,738 ESTs, 9,554 unigenes), petunia (*Petunia hybrida*) (3,181 ESTs, 1,841 unigenes) and eggplant (*Solanum melongena*) (11,479 ESTs, 5,135 unigenes), all of which can be accessed at SGN (http://www.sgn.cornell.edu). To estimate ribosomal, chloroplast and mitochondrial contamination, the coffee unigene set was also screened against the *Arabidopsis* complete mitochondrial genome (NC_001284), tobacco ribosomal rRNA gene (5.8s AJ012367, 18s AJ236016 and 25s S52185) and the complete tobacco chloroplast genome (Z00044) sequences from GenBank.

Table 2 Comparison of the coffee and tomato EST databases derived from use of ESTScan calibrated with the same tomato training set (see Materials and methods for details)

	Tomato	Coffee
Total unigenes	30,576	13,175
Average unigene length, bp	774	678
Unigenes with coding regions	96%	95%
Average length (bp) of predicated peptides	569	556
Average ESTScan score	409	346

Functional annotation based on predicted peptides

ESTScan-predicted coffee peptides were subjected to InterPro Scan annotation, which integrates the most commonly used protein signature databases (PROSITE, PRINTS, Pfam, ProDom, etc.) together with their associated scanning methods for protein domain analysis (Apweiler et al. 2001; Zdobnov and Apweiler 2001). Based on the domain annotation, GO accession of the unigenes were assigned using interpro2go conversion file from the GO consortium (http://www.geneontology.org, also available at http://www.ebi.ac.uk/interpro).

Functional categorization based on gene ontology

GO annotations were formatted for input into the GOSlim program and the output parsed to count the occurrence of each GO category. GOSlims are 'slimmed down' versions of the ontologies that allow a high-level view of gene functions. The GOSlim file and program were obtained from the Gene Ontology Consortium at http://www.geneontology.org.

Gene family analysis

The predicted protein sequences for the coffee unigene set and the *Arabidopsis* protein set were combined into a single file, formatted as a blast database using formatdb, and run with BLASTP (protein vs. protein sequence) against itself with option m8 for output. The resulting file was used as the input for the tribeML program. TribeML (Enright et al. 2002) formats the clusters, such that each cluster was given in a tab delimited file, one cluster per line. Simple scripts were used to parse the information to detect the largest gene families, coffee specific families, and families that showed large expansions in coffee.

Results and discussion

Generation of coffee EST database and unigene set

A total of 62,829 cDNA clones, derived from mRNA from five different tissues, were subjected to 5′ sequencing. After quality evaluation (see Materials and methods for details), the database was reduced to 46,914 high quality ESTs, averaging 613 bp in length. This corresponded to approximately 9,000 high quality ESTs from each of the five cDNA libraries (Table 1). These high quality ESTs were assembled into 13,175 unigenes with average length of 678 bp (ranging 150–2,714 bp), among which 7,272 (55%) were singletons and 5,903 (45%) were contigs (Fig. 2). Of the contigs, the majority (87%) was represented by two to ten ESTs while some (13%) were comprised of more than ten ESTs (Fig. 2). All trace files, sequences and derived unigenes/annotation can be found at the Solanaceae Genomics Network

BNSDOCID: <XP_____19322122A_I_>

(SGN) website (http://www.sgn.cornell.edu) and will be submitted to GenBank at the time of publication.

Differentiating between paralogs and alleles

The coffee ESTs were derived from five *C. canephora* varieties (Table 1), however, unigene assembly for EST reads does not automatically distinguish between polymorphic alleles and highly similar paralogs. *C. canephora* is known to be a highly allogamous, heterozygous and polymorphic species, raising the possibility that some of the EST-derived unigenes might actually represent different allelic forms of the same gene. In order to evaluate how well paralogs were distinguished from allelic polymorphism in the unigene assembly, BLASTN (nucleotide vs. nucleotide sequences) was performed amongst all of the coffee unigenes and the results parsed to identify the best non-self matches. As a control, BLASTN was performed in a similar manner on a complete set of *Arabidopsis* predicted genes, including both coding and UTR regions obtained from TAIR (http://www.arabidopsis.org). Since the *Arabidopsis* database was derived from the complete genome sequence of a single, inbred stock (*Columbia*), the entire gene complement is present without the complication of allelic variation (initiative). For both coffee and *Arabidopsis*, pair-wise combinations for which the homology extended over more than 70% of the sequence were recorded. The sequence similarity for each of these combinations was plotted for both coffee and *Arabidopsis* (Fig. 3).

The *Arabidopsis* plot showed two peaks, one with low identity (~87%) and the other with higher identity (over 99%) with 6.4 and 1.6% of genes falling into the low and high identity peaks, respectively. Like *Arabidopsis*, the coffee plot also showed two peaks, one with lower identity (around 91%) and the other with high identity (over 99%). These two peaks corresponded to 6.2 and 0.8% of the total unigenes, respectively. ESTs corresponding to ten pairs of coffee unigenes, from the >99% peak, were used as probes on genomic southern hybridizations to determine whether the matching pairs were truly duplicated in the coffee genome (paralogs) or rather allelic (single copy). For eight of the ten pairs, the paired ESTs hybridized the same single copy gene on southerns (data not show). Thus, a significant number (approximately 80%) of the unigenes in this peak are likely to be allelic. However, this category represented only a small portion of the coffee unigenes (0.8%). A similar experiment was performed with 11 ESTs from 11 pairs of unigenes in the second, lower homology peak (around 91% identity, see Fig. 3). In this case, the majority (8 out of 11) was determined to represent true paralogs (two or more copies in the genome) (data not shown). Thus for further discussions, it is assumed that the majority coffee EST-derived unigenes do in fact correspond to unique coffee genes.

Functional annotation of coffee EST-derived unigenes

Predicted coffee proteins

ESTScan (see Materials and methods for details) was able to identify protein-coding sequences in 12,534 coffee unigenes (95% of total unigenes), among which 1,515 (12%) were putatively full-length (starting with ATG and ending with a stop codon). Due to the cDNA library construction method, the unigenes were biased for the 3′ end—57% of the unigenes covered the 3′ end (ending with a stop codon) while only 36% covered the 5′ end

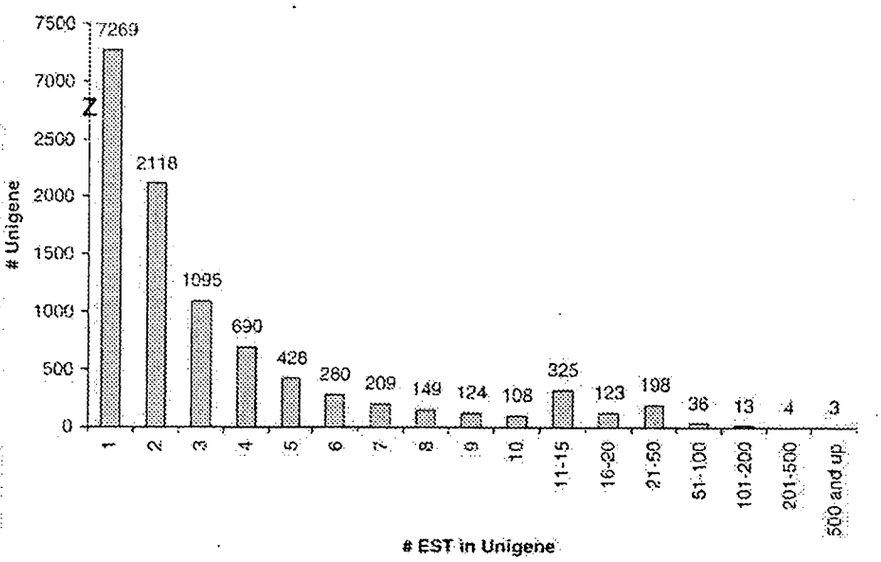


Fig. 2 Histogram depicting the distribution of EST content for all coffee unigenes. Numbers above *bars* equals the number of unigenes represented in each

BNSDOCID: <XP_____19322122A_I_>

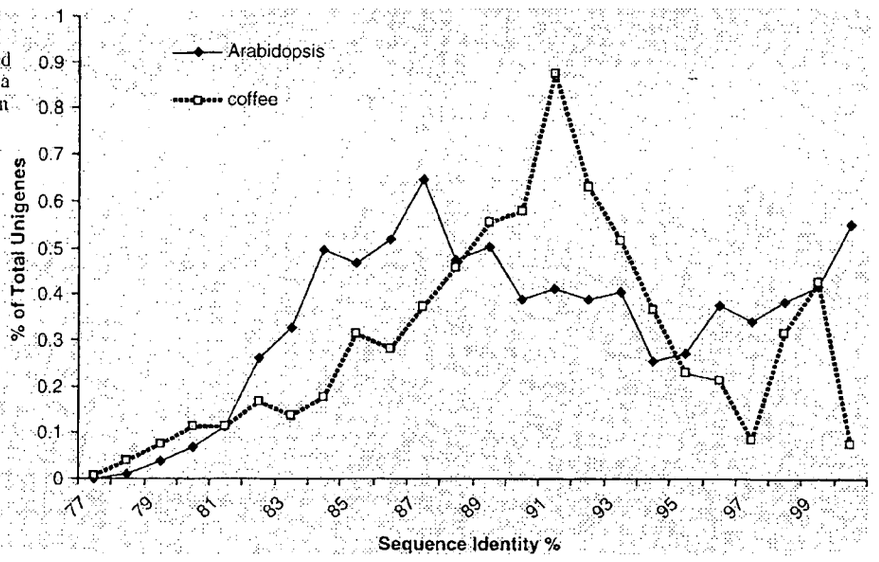


Fig. 3 Plot depicting the sequence identify of the most similar match for each coffee unigene as compared with all other coffee unigenes. As a control, a similar analysis is shown for *Arabidopsis* genes (see Results for details)

(starting with ATG). Of the 5% of unigenes from which a protein sequence could not be predicted, 81% were singletons and the majority (97%) did not match to any *Arabidopsis*, GenBank non-redundant (NR) or Solanaceae unigene sequences, suggesting that they are not bona fide gene transcripts.

Protein domain annotation

Peptides translated from ESTScan predicted coding sequences were subjected to InterPro Scan (see Materials and methods for details) for protein domain analysis. A total of 1,678 distinct domains were identified in the database, which were present in 4,414 (35%) unigenes. Table 3 lists the 20 most represented InterPro domains found in the coffee genes, as well as comparative statistics for tomato and *Arabidopsis* genes. Coffee and tomato were very similar with regard to the most frequent protein domains; however, a number of striking differences were noted when comparing coffee and *Arabidopsis*. Most notably, the following five domains were much more common in coffee encoded proteins than in *Arabidopsis*: proline-rich regions, tyrosine protein kinases, glucose/ribitol dehydrogenases, ubiquitins and ubiquitin-conjugating enzymes. The latter two categories, involving ubiquitinization, suggest a more active and/or complex system for protein turnover in coffee versus *Arabidopsis*.

Gene ontology annotation

Gene ontology (GO) annotations were made for each coffee unigene based on InterPro domain annotation (see Materials and methods for details). A total of 3,248 unigenes (25% of total unigenes and 74% of unigenes with InterPro domain match) could be assigned a GO annotation. In order to obtain an overview of representation of biological processes by the coffee unigene set, the GO annotations were mapped to the plant GOSlim, a group of selected higher-level categories of the GO ontology (see Materials and methods for details). Figure 4 depicts the biological process GOSlim categories for each of which at least 1% of the coffee unigenes were assigned. Metabolism is the most abundant category, both in terms of number of subcategories (14 subgroups) and number of unigenes in the subcategories (2,541 unigenes and ~80% of the GO annotated unigenes). In addition, signal transduction, cell growth and/or maintenance, response to stress and response to endogenous stimulus were among the most abundant categories.

A comparison was made between the GOSlim biological process of *Arabidopsis*, tomato and coffee (Fig. 4). For both the tomato and coffee unigene sets, the GO annotations were based on InterProScan results and approximately 25% of both unigene sets were assigned a GO annotation. In *Arabidopsis*, the genes are of full length, giving a higher chance of finding functional domains. Moreover, extensive experimental research and manual annotation has been carried out in *Arabidopsis*, resulting in a higher proportion of genes with assigned GO annotation. Therefore 83% of the *Arabidopsis* genes are assigned GO annotations. No significant differences were observed in the annotated categories for coffee versus tomato—possibly reflecting their close taxonomic affinity. However, for a number of categories, coffee had significantly different proportions of genes than *Arabidopsis*. The categories with the largest significant differences ($P < 0.001$, based on Chi-square test) are: carbohydrate metabolism, other metabolism, biosynthesis, catabolism, protein biosyn-

BNSDOCID: <XP_____.19322122A_I_>

Table 3 Twenty most abundant InterPro domains identified in coffee unigene set and comparative statistics for tomato and *Arabidopsis* genes

InterPro accession	Description	% of unigenes (ranking)		
		Coffee	Tomato	*Arabidopsis*
IPR000719	Protein kinase	1.6	1.20 (1)	3.0 (1)
IPR000694	Proline-rich region	1.3	0.91 (4)	0.003 (1763)
IPR002290	Serine/threonine protein kinase	0.85	1.10 (2)	0
IPR001245	Tyrosine protein kinase	0.69	1.0 (3)	0.15 (311)
IPR008271	Serine/threonine protein kinase, active site	0.61	0.68 (5)	2.6 (2)
IPR000504	RNA-binding region RNP-1 (RNA recognition motif)	0.55	0.60 (6)	0.59 (6)
IPR001680	G-protein beta WD-40 repeat	0.49	0.51 (8)	0.51 (8)
IPR001611	Leucine-rich repeat	0.48	0.59 (7)	0.59 (7)
IPR002048	Calcium-binding EF-hand	0.36	0.34 (13)	0.34 (13)
IPR000379	Esterase/lipase/thioesterase	0.33	0.43 (10)	0.43 (10)
IPR001806	Ras GTPase superfamily	0.32	0.26 (22)	0.43 (70)
IPR003579	Ras small GTPase, Rab type	0.29	0.23 (27)	0
IPR005123	2OG-Fe(II) oxygenase superfamily	0.27	0.26 (21)	0.47 (52)
IPR000626	Ubiquitin	0.27	0.22 (32)	0.40 (89)
IPR002401	E-class P450, group I	0.27	0.46 (8)	0.77 (24)
IPR002347	Glucose/ribitol dehydrogenase	0.26	0.23 (28)	0.33 (110)
IPR001005	Myb DNA-binding domain	0.26	0.34 (15)	1.34 (8)
IPR005225	Small GTP-binding protein domain	0.26	0.24 (25)	0.68 (27)
IPR000608	Ubiquitin-conjugating enzymes	0.26	0.21 (34)	0.19 (221)
IPR007090	Leucine-rich repeat, plant specific	0.25	0.40 (12)	1.07 (11)

thesis, protein modification and energy pathways. In all cases, coffee had a significantly higher proportion of genes in these categories than *Arabidopsis* (Fig. 4). It is interesting to note that many of these categories center around the synthesis, breakdown or modification of compounds. One of the hallmarks of coffee is its high-level diversity of primary and secondary compounds, which contributes to the sensory quality of brewed coffee

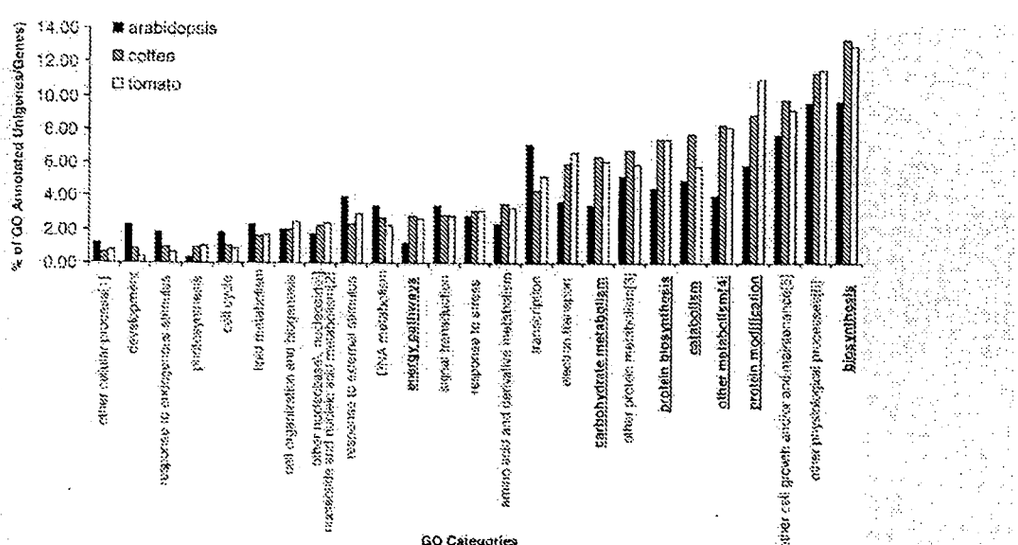


Fig. 4 Comparison of the gene ontology-based gene annotation categories for the coffee EST-derived unigene set, tomato EST-derived unigene set and the *Arabidopsis* proteome. Figure contains only categories in which more than 1% of the coffee unigenes were assigned. Categories for which coffee differs most significantly from *Arabidopsis* are shown in underline bold. (1) Cellular processes other than signal transduction and cell growth and/or maintenance. (2) Nucleobase/nucleoside/nucleotide and nucleic acid metabolism other than DNA metabolism and transcription. (3) Protein metabolism other than protein biosynthesis and protein modification. (4) Metabolism other than amino acid and derivative metabolism, biosynthesis, carbohydrate metabolism, catabolism, electron transport, lipid metabolism, nucleobase/nucleoside/nucleotide and nucleic acid metabolism and protein metabolism. (5) Cell growth and/or maintenance other than cell cycle and cell organization and biogenesis. (6) Physiological processes other than photosynthesis, response to stress, response to endogenous stimulus, response to external stimulus and metabolism

BNSDOCID: <XP_____19322122A_I_>

beans. The Rubiaceae family in general contains some of the most diverse species with regard to secondary metabolism and is an especially rich source of alkaloids—a number of which have pharmacological and/or psychotropic properties (Kutchan 1995; Facchini 2001). In fact, the most widely used psychotropic drug, caffeine, comes mainly from coffee. One can speculate that this metabolic diversity is reflected in the relatively high proportion of coffee genes with putative functions related to metabolism.

In silico analysis of unigene expression

Complexity and uniqueness of different stages/tissues

In an effort to determine which tissues gave rise to the more redundant or more complex pools of gene expression, three library/tissue parameters were calculated: (1) the percentage of total coffee unigenes represented by each library/tissue; (2) the percentage of more highly expressed unigenes (represented by ten or more ESTs from the library) in each library/tissue; (3) the percentage of unigenes from each tissue/stage which are unique to that library/tissue (not found in any other cDNA library). The results from these calculations are depicted in Fig. 5. The pericarp and leaf libraries represented more unigenes, contributed to more unique unigenes, and contained fewer highly abundant unigenes than did the early cherry, middle seed and late seed libraries—perhaps reflecting the overall lower complexity of these latter stages/tissues (Fig. 5). Among the seed stages, the middle stage of development gave more novel genes than did the early or late stages of seed development (Fig. 5).

Differential expression of genes across stages/tissues

The coffee cDNA libraries were non-normalized and deeply sampled, and thus could be used to compare gene expression amongst the five cDNA libraries using statistical method (Audic and Claverie 1997). In all the pair-wise comparison between the cDNA libraries, only a small proportion of the overall unigenes (approximately 5%) showed significantly different ($P < 0.05$) expression levels between two or more libraries (Table 4). The leaf and the early stage cherry libraries showed the highest divergence in gene expression (752 unigenes which is 5.7% of total unigenes, showed significant differential expression) while the leaf and the pericarp libraries had the lowest number of differentially expressed unigenes (384 unigenes and 2.9% of total). Interestingly, the three stages of seed development showed as much differential expression, relative to each other, as they did to non-seed tissues (leaf and pericarp). These results provide strong evidence that throughout the approximately 11 months from pollination to mature seed, a continuously changing cascade of molecular events is unfolding in the developing seed.

Highly expressed genes

Table 5 lists the blast annotation, library composition and the best *Arabidopsis*/Solanaceae gene match for the 20 most highly expressed unigenes from the coffee EST

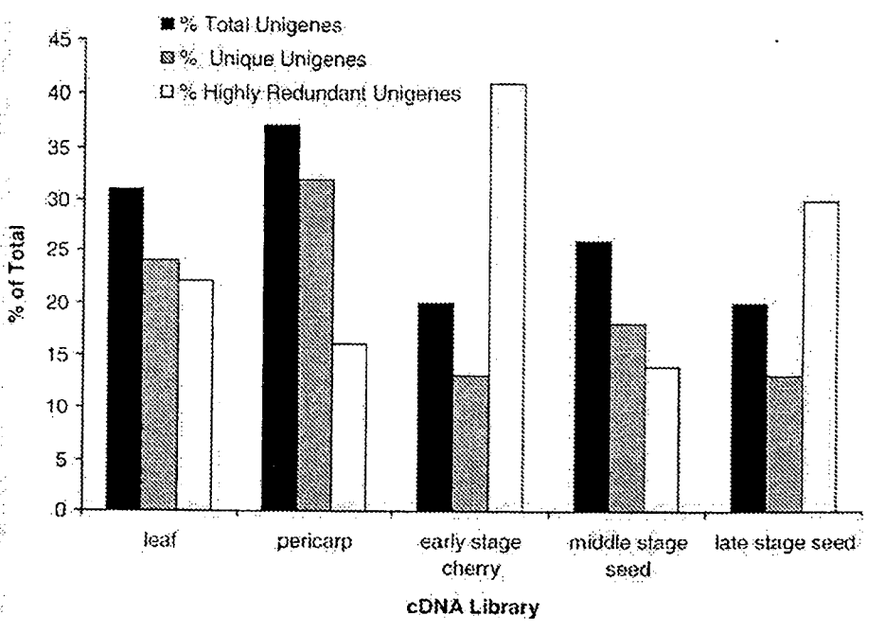


Fig. 5 Characteristics of each coffee cDNA library in comparison to the entire coffee EST-derived unigene set. The total unigene and highly expressed unigene categories sum to greater 100% since the same unigene may contain ESTs from more than one library

BNSDOCID: <XP_____19322122A_I_>

Table 4 Number of coffee unigenes showing significantly ($P < 0.05$) different expression in pairwise comparisons of cDNA libraries

Library	Pericarp	Early stage cherry	Middle stage seed	Late stage seed
Leaf	384	752	548	562
Pericarp		610	458	527
Early stage cherry			602	728
Middle stage seed				585

database. Some of these highly expressed genes are common to and highly expressed in all plant species—a good example of this is Unigene 122071, which encodes the small subunit of ribulose bisphosphate carboxylase—a key enzyme in photosynthesis and carbon fixation (Table 5). However, a number of these highly expressed genes are unique to coffee, and many have expression patterns confined to particular tissue/stages, providing a glimpse into some of the potential unique aspects of coffee biology. A brief description of some of these is given below.

Seed storage protein genes

Unigene 125230: a putative 2S seed storage protein Unigene 125230 is the most highly expressed gene across the entire coffee unigene set (1219 ESTs) and was the dominant transcript in the middle stage seed library, accounting for 10% of the ESTs at this stage (Table 5). This gene shows high homology to a tomato unigene derived from a developing seed cDNA library, but has no detectable homolog in *Arabidopsis* (Table 5). Other than the match with tomato, weak homology was also detected for 2S seed storage proteins from sesame, sunflower, and Brazil nut (in decreasing order of similarity). It is interesting to note that coffee, tomato, sesame and sunflower are fairly closely related taxonomically. All belong to the Asterid I/II clade of Eudicots (Fig. 1). This close phylogenetic relationship may explain why Unigene 125230 has homologous matches only in these species. Moreover, since Unigene 125230 shows homology to the 2S seed storage proteins in these related species, we conjecture that unigene 125230, its tomato unigene match, the sesame, Brazil nut and sunflower 2S storage protein gene all encode orthologous 2S seed storage proteins. This is the first time that a 2S seed storage protein has been identified in coffee or any Solanaceae species. Finally, a BLAST search of Unigene 125230 against the coffee unigene set revealed additional putative copies of the 2S seed storage protein. However, on close examination, all appear to be splicing variants or low quality sequences. Moreover, southern hybridization with a 2S cDNA probe on genomic DNA confirmed that the 2S gene is single copy in the coffee genome (data not shown).

Unigene 120912: 11S seed storage protein Unigene 120912 is the second most abundant unigene, containing 687 ESTs (Table 5). This gene is preferentially expressed during middle and late stage seed development and shares high similarity (over 98% identity) with a previously cloned *C. arabica* 11S seed storage protein (Marraccini et al. 1999; Rogers et al. 1999). This unigene also has a highly significant match to the *Arabidopsis* 12S storage protein and to a tomato unigene derived from seed ESTs (Pang et al. 1988) (Table 5). Given these results, we conclude that unigene 12912 is allelic with the previously described 11S seed storage protein gene from *C. arabica* and orthologous to 11S/12S seed storage proteins in both tomato and *Arabidopsis*. A BLAST search of Unigene 120912 against the coffee unigene set revealed additional putative copies of the 11S seed storage protein. However, like the 2s seed storage protein (Unigene 124230), all appears to be results of alternative splicing or low sequence quality. Moreover, southern hybridization with an 11S cDNA probe on genomic DNA confirmed that the 11S gene is single copy in the coffee genome (data not shown).

Other seed-specific genes

Early stage seed development Unigenes 122206, 119460 and 121265 were all highly expressed and specific to the early cherry stage. The early cherry library was derived from RNA from both pericarp and seed tissue while the pericarp library was derived from RNA coming from all stages of pericarp development. Thus, if the above genes were highly expressed in the pericarp of the early cherry, they should be present in the pericarp library as well. The fact that these genes showed little or no expression outside the early cherry stage, suggests that they are probably specific to early developing seed tissues and not pericarp tissues (Table 5). Unigene 122206 showed high homology to an *Arabidopsis* gene annotated as encoding the enzyme S-adenosyl-L-methionine (SAM) synthetase (Table 5). This enzyme synthesizes S-adenosyl-L-methionine from L-methionine and ATP and is often represented by multiple isozymes in plant species (Schroder et al. 1997). Thus, Unigene 122206 appears to be an SAM synthetase specific to early seed development (Table 5).

Unigene 119460 shows high homology to the highly conserved WRKY transcription factor family. The WRKY transcription factor is a large gene family having more than 70 members in the *Arabidopsis* genome (Dong et al. 2003). Previous studies showed that it is related to wounding, stress, pathogen infection and senescence in many plant species. In some recent studies, the WRKY

BNSDOCID: <XP_____19322122A_!_>

BNSDOCID: <XP_____19322122A_1>

Table 5 The 20 most highly expressed coffee unigenes: functional annotation and most similar *Arabidopsis* and Solanaceae homologs

Coffee unigene#: annotation	Best match (e value/score)		EST count					
	Arbidopsis	Solanaceae Unigene_species	Total ESTs	Leaf	Pericarp	Early stage cherry	Middle stage seed	Late stage seed
125230: putative 2s seed storage protein	ND	243065_tomato (e-103/238)	1,219	*7*	*15*	*21*	**1,037**	*139*
120912: 11s seed storage protein	At5g44120 (1e-88/324)	228376_tomato (0/802)	687	*0*	*3*	*28*	244	**412**
121707: unknown function	At1g29050 (1e-139/489)	246695_potato (e-163/283)	324	*2*	*3*	*1*	149	**169**
120118: unknown function	At5g59320 (2e-21/99.8)	221585_tomato (e-134/475)	292	*0*	*0*	*3*	58	**231**
124988: unknown function	ND	ND	204	58	**84**	*55*	*1*	*6*
120685: chitinase	At5g24090 (2e-43/172)	214596_tomato (1e-35/84.5)	202	**99**	*40*	*58*	*0*	*5*
124158: photoassimilate-responsive protein	At3g54040 (2e-36/149)	196924_pepper (2e-39/138)	182	*1*	*1*	*2*	**150**	28
119890: unknown function	ND	204426_pepper (5e-07/52.8)	183	*0*	*0*	*0*	**183**	0
123265: ADP-ribosylation factor	At2g47170 (1e-99/359)	238338_tomato (0/693)	182	58	*14*	**65**	*21*	*24*
124083: secretory peroxidase	At4g21960 (e-153/537)	196145_pepper (0/681)	161	**55**	*5*	*19*	49	*33*
124911: metallothionein	At5g02380 (0.32/32.3)	207464_petunia (2e-06/51.0)	163	*40*	**65**	*30*	*11*	*17*
119817: chitinase	At3g12500 (e-103/373)	248120_potato (e-148/521)	148	*0*	*22*	*0*	*0*	**126**
124815: unknown function	At3g29240 (1e-87/320)	227940_tomato (e-146/517)	145	*2*	*0*	*1*	*0*	**142**
122206: SAM synthase	At2g36880 (0/711)	270415_petunia (0/887)	142	*1*	*9*	**130**	*0*	*2*
119460: WRKY4 transcription factor	At1g80840 (3e-75/279)	237166_tomato (e-137/487)	123	*0*	*0*	**123**	*0*	*0*
123045: unknown function	At3g16000 (0.69/31.2)	218824_tomato (90.36/33.1)	123	**81**	*17*	*22*	*3*	*0*
120481: AdoMet synthase	At4g01850 (0/723)	243236_potato (0/886)	108	*19*	*27*	**31**	*30*	*11*
121265: Mobl/phocein	At5g45550 (e-119/425)	196814_pepper (e-146/513)	113	*1*	*0*	**112**	*0*	*0*
124791: plasmodesmal receptor	At5g15140 (1e-99/360)	203764_pepper (8e-86/314)	105	*0*	*0*	*3*	*26*	76
122071: rubiso small subunit	At1g67090 (9e-70/260)	207453_petunia (3e-89/297)	99	**76**	*8*	*9*	*6*	*0*

BLAST match values are given in parentheses
Bold numbers indicate library for which the highest number of ESTs were observed for each gene. Italic numbers indicate library for which the number of EST is significantly lower ($P < 0.05$) than the highest

338

protein family was found to be involved in sugar signaling in barley and seed development in *Arabidopsis* (Johnson et al. 2002; Sun et al. 2003). However, the function of the best *Arabidopsis* match to Unigene 119460 (At1g80840) has not been determined. Hence, understanding the function of this highly expressed, WRKY-like coffee gene awaits further study.

Unigene 121265 is highly homologous to a gene in *Arabidopsis* annotated as encoding a Mob1/phocein protein. Mob1/phocein proteins are found in virtually all eukaryotes. While they are conjectured to be involved in cell cycle control, there is still little experimental evidence demonstrating biological function (Pon 2004). Thus it seems premature to conjecture what role Unigene 121265 might have that is specific to the early development of coffee seeds.

Middle stage seed development As described earlier, Unigene 125230 is a putative 2S seed storage protein with peak expression during middle seed development. Also showing preferential expression during this same stage were Unigenes 121707, 124158 and 119890. Unigene 121707 is a gene of unknown function with high homology matches both in *Arabidopsis* and Solanaceae EST-derived unigenes (Table 5). Unigene 124158 is homologous to an *Arabidopsis* gene classified as a photoassimilate-responsive protein, which is related to pathogenesis (Herbers 1995). Finally, Unigene 119890, which is also specific to middle stage seed development, is apparently a gene unique to coffee, which will be discussed more in the following section.

Late stage seed development As discussed previously, Unigene 120912 corresponds to the 11S seed storage protein, which is largely expressed late in seed development. Other genes with preferential expression in late stage seeds are Unigenes 120118, 119817 and 124791. Unigene 119817 likely encodes a chitinase and is further discussed in the next section. Unigene 120118 shows high homology to genes in both *Arabidopsis* and Solanaceae EST-derived unigene sets; however, none have known function. Unigene 124791 gives a strong match to an *Arabidopsis* gene annotated as a plasmodesmatal receptor.

Two highly expressed genes with homology to chitinase

Unigenes 120685 and 119817 show high sequence similarity to a number of genes classified as chitinases in other organisms. Chitinases are a large and diverse class of proteins, some of which have been implicated in resistance to fungi in various plant species, including coffee (Rojas-Herrera 2002; Chen et al. 2003). The two unigenes differ in that Unigene 120685 is expressed in leaves, pericarp and early stage cherries, but not in mid or late stage seed development (Table 5). Unigene 119817, on the other hand, was found to be exclusively expressed in late stage developing seeds and pericarp tissue. As previously mentioned, early stage cherries contained both pericarp and seed tissues. The fact that Unigene 120685 was not found in the middle and late stages of seed development suggests that this gene may not be expressed in seeds, but rather in the maternally derived pericarp and leaf tissues. Based on these results, one can speculate that these two putative chitinase genes may be involved in pathogen defense in developing CHERRIES, with Unigene 120685 being expressed in early developing, post pollination pericarp and leaf tissues and Unigene 119817 being expressed primarily late in seed development, just prior to maturity.

Highly expressed genes unique to coffee

Unigene 124988 This highly expressed unigene had no significant matches in the *Arabidopsis* proteome, Solanaceae EST-derived databases, GenBank NR databases, or GenBank dbest. Moreover, the predicted protein encoded by Unigene 124988 has no recognizable domains, which might give clues to its function. ESTs for this unigene were detected in all five libraries, with highest expression being observed in the pericarp (Table 5).

Unigene 119890 Unigene 119890 also has no significant match in any of the tested databases, with the possible exception of a very weak match in the Solanaceae unigene sets (the best hit was from pepper with an e value of 5e-7, Table 5). Like Unigene 124988, its predicted protein has no recognizable domains. Unigene 119890 was highly and exclusively expressed in the middle stage of developing seeds (Table 5).

The fact that neither Unigene 124988 nor Unigene 119890 have counterparts in any other databases suggests that they may represent coffee-specific genes or genes that have been evolving at such a rapid rate that they no longer bear any recognizable homologies with proteins from other plants, including the closely related Solanaceous plants. We speculate that these genes may be related to chemical or morphological features unique to coffee.

Gene families unique or significantly expanded in coffee

The predicted protein sequences for the coffee unigene set and the *Arabidopsis* proteome were combined and organized into protein families using tribeML software (see Materials and methods for details). In most cases, the number of gene family members was higher in *arabidopsis* than in coffee (data not shown). However, this is to be expected since the entire *Arabidopsis* genome has been sequenced (thus revealing all genes), whereas in coffee the EST-derived unigene set is by nature incomplete and unlikely to contain all members of any given gene family. However, despite the incomplete nature of the coffee EST-derived unigene set, several cases was encountered in which coffee has more gene family

BNSDOCID: <XP_____19322122A_I_>

Table 6 Gene families expanded in coffee relative to *Arabidopsis*

Family #	# *Arabidopsis* family member	# Coffee family member	Longest coffee member	Annotation
266	1	21	122330	Retrotransposon gag protein, class I
180	5	14	124952	Polygalacturonase isoenzyme 1 beta subunit with BURP domain
632	1	12	123451	Acidic endochitinase
386	2	10	124158	Photoassimilate-responsive protein
382	4	8	119672	Hypersensitive-induced protein, band 7 protein
394	2	7	122791	E-class P450
483	2	6	120054	Bet v I allergen
623	3	6	119581	Root hair defective protein
1,182	1	5	126674	Unknown function
695	2	5	126974	Tyrosine decarboxylase
783	2	5	122423	Unknown function
1,117	2	5	119449	Trypsin inhibitor Kunitz

members than does *Arabidopsis*. To further investigate this topic, a list was made of the gene families for which coffee has significantly more members than *Arabidopsis* (Table 6). We refer to these as coffee-expanded gene families. In addition, gene families found in coffee, but for which no counterparts exist in *Arabidopsis* were also compiled (Table 7). We refer to these as coffee-unique gene families. All gene families listed in both Tables 6 and 7 were subjected to manual examination to remove any members determined to be results of alternative splicing and/or poor sequence quality—both of which could inflate the number of putative gene family members in coffee. Hence, the differences in copy number reported are likely to be due to true differences in gene copy number and not artifacts of EST-based gene predictions.

Coffee-expanded gene families

The most expanded gene family in coffee corresponds to a retrotransposon gag protein (Table 6). This result has two implications. First, the retroelement encoding this gag protein occurs at a higher frequency in coffee compared with *Arabidopsis*, although we cannot determine whether this difference is due to a true expansion of this element in coffee subsequent to divergence from *Arabidopsis*, or rather a loss of the element in the *Arabidopsis* lineage. Second, the fact that this retrotransposon gag protein element was discovered in an EST-database indicates that this particular retroelement is being transcribed in the coffee genome, and hence may represent an active retrotransposon.

Another gene family for which coffee has significantly more members than *Arabidopsis*, encodes proteins annotated in *Arabidopsis* as acid endochitinases and photoassimilate-responsive proteins (Table 6). As noted earlier, chitinases are associated with fungal resistance and are among the most highly expressed genes in coffee. The fact that chitinases are both highly expressed and represented by an expanded gene family in coffee may reflect a greater need for fungal resistance engendered both by the perennial nature of coffee and the fact that it is a tropical species for which a multiplicity of fungal pathogens is common. The reasons for the putative

Table 7 Gene families unique to coffee in comparison to *Arabidopsis*

Gene family #	# Family member	Longest member	Solanaceae hit	Annotation
243	27	122956	258190 potato	Retrotransposon gag protein, class II
687	11	120121	221585 tomato	Thaumatin, pathogenesis related
965	10	119718	249401 potato	Zn-finger, CCHC type
974	10	120244	2610402 potato	Disease resistance protein (TIR-NBS-LRR class)
852	9	119638	225732 tomato	Retrotransposon gag protein, classs III
360	8	121998	23671 tomato	Disease resistance protein
1,019	7	124574	222350 tomato	Leucine-rich repeat, disease resistance protein
1,607	7	122216	none	Unknown function
1,610	7	130519	none	Unknown function
1,676	7	126264	243065 tomato	Unknown function
708	6	123769	236157 tomato	ABA/WDS induced protein
1,852	5	120284	213688 tomato	Proline-rich region, extension-like protein
2,362	5	122218	237314 tomato	Unknown function
2,459	5	124466	267984 potato	Leucine-rich repeat, plant specific, receptor-related protein kinase

BNSDOCID: <XP_____19322122A__I_>

expansions of the other gene families listed in Table 6 remain for future studies to determine.

Coffee-unique gene families

Table 7 lists the top gene families (based on copy number), which occur in coffee, but not in *Arabidopsis*. Of the 15 gene families listed, four are of unknown function. For those that could be functionally annotated, five (45%) have putative functions related to disease resistance, such as TIR-NBS-LRR disease resistance proteins, LRR proteins and thaumatin pathogenesis-related proteins (Table 7). These finding are consistent with rapid evolution of genes/gene families related to disease resistance, likely driven by selection pressure from continuously changing pathogens and/or pathogens unique to the particular environments of a species (Meyers 1998; Michelmore and Meyers 1998). Also included in this list of coffee-unique gene families are two, which encode retrotransposon gag-proteins (Table 7).

Comparison of the coffee gene repertoire with that of *Arabidopsis* and Solanaceae species

Each coffee unigene was screened against the *Arabidopsis* proteome using BLASTX, which compares all six translated frames of the query with the target. In total, 21% of the coffee unigenes had no clear match in *Arabidopsis* (match score 0–30) (Fig. 6). These coffee genes, which lack an *Arabidopsis* counterpart, may have evolved rapidly since coffee and *Arabidopsis* diverged from their last common ancestor, and hence their counterparts in each genome are no longer recognizable by standard string searches like BLAST. To investigate this topic further, the 2,853 unigenes with no clear match in *Arabidopsis* (match score < 30, see Fig. 6) were screened against the Solanaceae EST-derived unigene sets using TBLASTX which compares all six translated frames of both query (coffee unigenes) and target (Solanaceae unigenes, see Materials and methods for details) sequences. The family Solanaceae is much more closely related to coffee than is *Arabidopsis* (Fig. 1), and hence coffee genes without a match in *Arabidopsis* may well have recognizable orthologs in the Solanaceae. The results show that a large portion (90%) of the coffee genes that lack a match in *Arabidopsis* do have a match in Solanaceae at the same stringency criteria (match score > 30). For many of these cases, however, the coffee-Solanaceae matches are still quite diverged (low match scores) suggesting that the proteins encoded by these genes have been evolving at a rapid rate in dicots.

There were, however, some exceptions to this trend represented by coffee genes having no match in *Arabidopsis*, yet a strong match in Solanaceae. These latter cases may represent genes that have been under differential selection pressure since the time that the Rosids (which includes *Arabidopsis*) diverged from the last common ancestor with the Asterids (which includes Solanaceae and coffee) (Fig. 1). Alternatively, these genes may have been lost in the lineage leading to *Arabidopsis*, but retained in the Asterids. In either case, such genes may have functional roles that are, in part, responsible for the morphological, developmental and chemical characteristics that differentiate *Arabidopsis* from Rosid species (e.g. coffee, Solanaceae). In an effort to shed more light on this issue, the 20 coffee unigenes having the highest match scores with Solanaceae, but lacking a counterpart in *Arabidopsis* were parsed and annotated (Table 8). Seven of these coffee genes have no match either in GenBank NR or dbest hence their function is unknown. Another unigene (Unigene 125230, the 2s seed storage protein, see above discussion) had weak match to ESTs of sesame, a species also belonging to the Asterid I clade (Fig. 1). Thus, these eight genes (40%) appear to be specific to the tomato/coffee/sesame

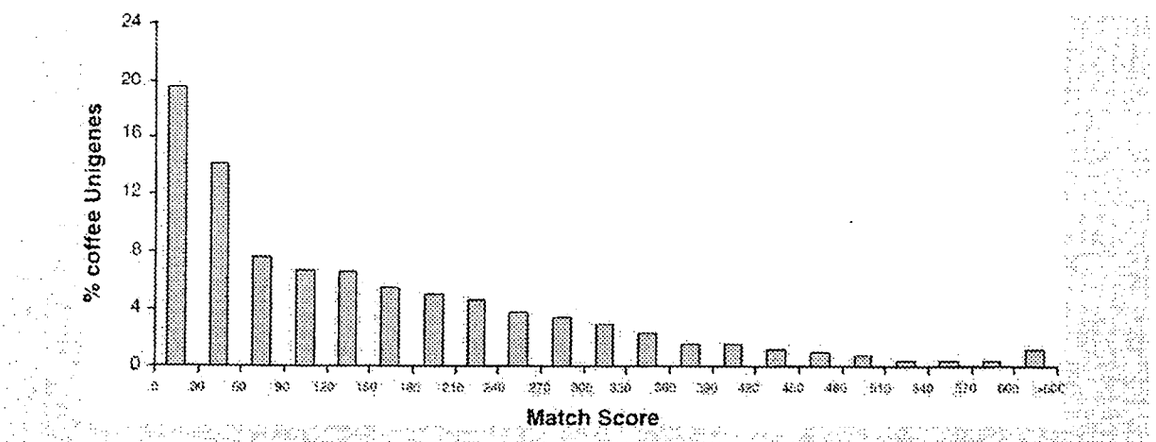


Fig. 6 Histogram showing match scores for each coffee unigene as compared with its best match in the *Arabidopsis* proteome

BNSDOCID: <XP_____19322122A_I_>

Asterid I clade (Fig. 1, Table 8). Elucidating their function may shed light on the molecular processes, which diversified early in the evolution of dicot species, especially those species belonging to the Asterid clade of dicots.

Surprisingly, 8 (40%) of the 20 coffee genes having no match in *Arabidopsis* did have matches in species phylogenetically distant from both coffee and *Arabidopsis*, including two to non-plant species (Drosophila and rat) (Table 8). Five of these matches were to rice genes, which is a monocot and highly divergent from coffee, *Arabidopsis* and other dicot species (Table 8). The fact that all of these species diverged from *Arabidopsis* and coffee long before the latter diverged from each other suggests that these genes may have been present in the last common ancestor of *Arabidopsis* and coffee/Solanaceae, but subsequently lost in the *Arabidopsis* lineage.

Coffee genes share greater similarity to genes in tomato/Solanaceae than to Arabidopsis

As discussed earlier, coffee is much more closely related to the Solanaceae than to *Arabidopsis* (Fig. 1). Hence, Solanaceae species may be better models for coffee genomics than *Arabidopsis*. The results for fast-evolving genes, presented above, are consistent with this prediction. To further investigate this assertion, the degree to which each coffee unigene matched *Arabidopsis* versus Solanaceae was investigated. In doing this analysis, one has to keep in mind that the entire gene repertoire of *Arabidopsis* is known, whereas the EST-derived unigene sets for Solanaceae do not represent the entire gene repertoire of these species. We estimate that the combined EST-derived unigene sets of Solanaceae species represent as much as three-quarters of the Solanaceae gene content (Hoeven et al. 2002). Moreover, *Arabidopsis* genes are of full length, while Solanaceae EST-derived unigenes are not necessarily of full length.

To mitigate against these factors, each coffee unigene was first screened, via TBLASTX, against the Solanaceae EST-derived unigene databases. Only those coffee unigenes matching a Solanaceae unigene with a match score greater than 100 were compared with the *Arabidopsis* proteome. Moreover, the match to *Arabidopsis* was confined to only the portion of the coffee unigene matching a Solanaceae EST-derived unigene. The ratio of the *Arabidopsis* match score divided by the Solanaceae match score was used as a measure of the degree to which the Solanaceae homolog shows greater similarity to the coffee unigene than to the closest *Arabidopsis* homolog. The results of this analysis are shown in Fig. 7. The ratio of the *Arabidopsis*/Solanaceae match score averaged 0.75, indicating a significantly better match in most cases with Solanaceae versus *Arabidopsis*. Moreover, in more than 95% of the cases, the ratio was less than one, indicating that, in most instances, coffee contains genes significantly more homologous to Solanaceae than to *Arabidopsis*. For those cases in which the ratio was greater than 1 (5%), we cannot rule out the possibility that the true ortholog was missing in the Solanaceae EST-derived unigene sets, hence the com-

Table 8 Coffee genes not found in *Arabidopsis*, but with conserved counterparts in tomato or other Solanaceous species

Coffee unigene	Solanaceae EST-derived unigene match	Score	GenBank (non-redundant and dbest) best match	Score	Annotation
124978	240871 tomato	454			Unknown function
121324	235756 tomato	429	gb|CB686389.1 [*Brassica* napus]	44	Unknown function
131820	213100 tomato	426	gi|50252229.1 [*Oryza* sativa]	73	Unknown function
121542	240321 tomato	416	ref|NP_922676.1 [*Oryza* sativa]	75	Unknown function
131934	219759 tomato	377	emb|CAE05735.1 [*Oryza* sativa]	297	TFIIH basal transcription factor p52 subunit
121140	236347 tomato	320			Unknown function
131445	225435 tomato	320			Unknown function
125230	243065 tomato	238	gi|13183175 [*Seasame* indicum]	45	2S albumin
131030	246364 potato	213	ref|NP_524404.1 [*Drosophila* melanogaster]	110	Phospyhatidyl inositol transfer protein
120120	237254 tomato	202	gb|CF349465.1 [Rose]	52	Unknown function
126635	237314 tomato	185			Unknown function
126575	237314 tomato	182			Unknown function
130675	209387 petunia	177	gb|CK093976.1 [*Populus* tremula]	438	Unknown function
128020	237314 tomato	167			Unknown function
123615	249253 potato	163	gb|AAO73272.1 [*Oryza* sativa]	140	Unknown function
126432	240551 tomato	163	gi|34878866 [*Rattus* norvegicus]	56	Phosphatidylinositolglycan class N
124384	197378 pepper	156	gb|CA815435.1 [*Vitis* vinifera]	1,009	Unknown function
122126	239632 tomato	153			Unknown function
131601	232010 tomato	145	gb|CK229938.1 [*Macaca* mulatta]	74	40S Ribosomal protein S21
119644	237150 tomato	143	ref|NP_921250.1 [*Oryza* sativa]	70	Helicase

The GenBanks Best match exclude those from Solanaceae, Coffea and Hedyotis (both members of the Rubiaceae family). Solanaceae EST-derived

BNSDOCID: <XP_____19322122A__1_>

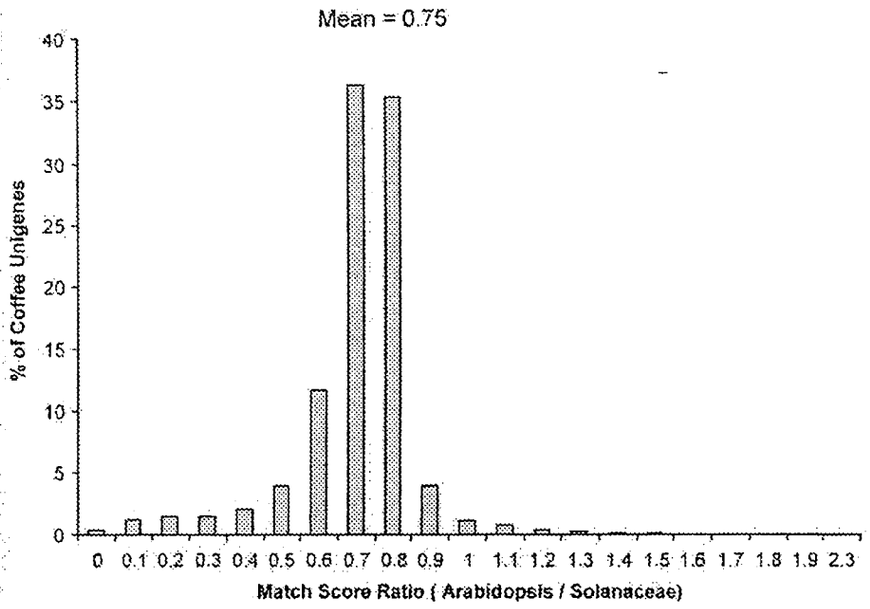


Fig. 7 Ratio of highest *Arabidopsis* match score to highest Solanaceae match score for individual coffee unigene. The analysis restricted to coffee unigenes with a Solanaceae match score > 100

parison was between paralogs. These results, combined with the results on fast evolving genes presented earlier, provide clear evidence that the Solanaceae provide a much better model than *Arabidopsis* for genomic and biological studies in coffee. This is especially relevant as sequencing of the tomato genome is currently underway (http://www.sgn.cornell.edu/help/about/tomato_sequencing.html).

Conclusions

Herein, we describe the development and analysis of a large EST database for coffee. The resulting 47,000 ESTs correspond to 13,175 unique genes (unigenes), a large portion of which are expressed during seed development—a stage important to coffee as a crop and one for which our understanding of molecular development is still rudimentary. To our knowledge, this is the largest public database for seed-derived ESTs. Hence, this EST database represents a new public resource, which can facilitate a better understanding of seed development, as well as genomic, molecular and breeding research in coffee. By comparisons with *Arabidopsis* and Solanaceous species, we have identified the two major seed storage proteins of coffee (2S and 11S) and demonstrated that these proteins are expressed at different times during seed development. Through in silico gene expression analysis, we have identified a number of highly expressed genes that show high specificity for different stages of seed development as well as for the pericarp tissue that surrounds the seeds. Many of these highly expressed genes are unique to coffee and/or the Asterid clade of higher plants. While the functions of most of these highly expressed, tissue/stage specific genes remain to be determined, the fact that they have been identified points the way to promoters, which can potentially be used to drive gene expression in specific stages/tissues of the coffee plant. Many of these genes are specific to defined periods of seed and/or pericarp development—both critically important for insect/pathogen resistance and in determining the quality of the coffee bean with respect to commercial coffee products.

Coffee, as a member of the family Rubiaceae, is distantly related to the model species *Arabidopsis*. A computational comparison of the coffee EST-derived unigene set with the sequence databases for *Arabidopsis* and Solanaceous species (e.g. tomato, pepper), indicate that the latter are much better genomic models for coffee than is *Arabidopsis*. These results are consistent with the fact that coffee and solanaceous species share very similar chromosome architecture and are closely related, both belonging to the Asterid I clade of dicot plant family. Moreover, the ability to identify orthologous genes between coffee and tomato opens the door to eventually developing detailed comparative maps for these two species and to the sharing of genomic and biological tools/discoveries—an outcome that should expedite research in both taxa.

Acknowledgements This work was supported by a grant from the Nestle Corporation to increase the knowledge base on coffee for the benefit of all stakeholders of the production chain, and funding from the National Science Foundation Plant Genome Program (no. 0116076). The authors wish to thank Dr. Ir Zaenudin Su, Dr. T. Wahyudi, Dr. Surip Mawardi and P. Priyono of the Indonesian Coffee and Cacao Research Institute (ICCRI) for supplying all the leaf and fruit samples used here from the ICCRI collection. We

BNSDOCID: <XP_____19322122A_I_>

thank V. Caillet and Dr. Pierre Marraccini for their respective efforts in preparing the RNA samples used in the cDNA library construction. Thanks to Dr. Anne Frary for critical reading and editing of the manuscript.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E et al (2001) The InterPro database, an intergrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res 29:37–40

Audic S, Claverie J-M (1997) The significance of digital gene expression profiles. Genome Res 7:986–995

Chase MW, Soltis DE et al (1993) Phylogenetics of seed plants—an analysis of nucleotide-sequences from the plastid gene rbcl. Ann Mo Bot Gard 80:528–580

Chen ZJ, Ribeiro R, Silva MC (2003) Heat shock-induced susceptibility of green coffee leaves and berries to Colletotrichum gloeosporioides and its association to PR and hsp70 gene expression. Physiol Mol Plant P 63:181–190

Crepet WL, Nixon KC, Gandolfo MA (2004) Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from cretaceous deposits. Am J Bot 91:1666–1682

Dong JX, Chen CH, Chen ZX (2003) Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. Plant Mol Biol 51:21–37

Enright AJ, Dongen SV, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575–1584

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. Genome Res 8:175–185

Facchini PJ (2001) Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. Annu Rev Plant Phys Plant Mol Biol 52:29–66

Gandolfo MA, Nixon KC, Crepet WL (1998) A new fossil flower from the Turonian of New Jersey: Dressiantha bicarpellata gen. et sp. nov. (Capparales). Am J Bot 85:964–974

Herbers KMG, Badur R, Sonnewald U (1995) A simplified procedure for the subtractive cDNA cloning of photoassimilate-responding genes: isolation of cDNAs encoding a new class of pathogenesis-related proteins. Plant Mol Biol 29:1027–1038

Herrera JC, Combes MC, Anthony F, Charrier A, Lashermes P (2002) Introgression into the Allotetraploid Coffee (*Coffea arabica* L.): segregation and recombination of the C. *canephora* genome in the tetreploid interspecific hybrid (C. *arabica*×C. *canephora*). Theor Appl Genet 104:661–668

Hoeven RVd, Ronning C, Giovannoni J, Matin G, Tanksley S (2002) Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell 14:1441–1456

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Initiative TAG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. American Association of Artificial Intelligence

Johnson CS, Kolevski B, Smyth DR (2002) TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. Plant Cell 14:1359–1375

Kutchan TM (1995) Alkaloid biosynthesis—the basis for metabolic engineering of medicinal plants. Plant Cell 7:1059–1070

Lee S, Kim SY, Chung E, Joung YH, Pai HS, Hur CG, Choi D (2004) EST and microarray analyses of pathogen-responsive genes in hot pepper (Capsicum annuum L.) non-host resistance against soybean pustule pathogen (Xanthomonas axonopodis pv. glycines. Funct Integr Genomics 4:196–205

Marraccini P, Deshayes A, Petiard V, Rogers WJ (1999) Molecular cloning of the complete 11S seed storage protein gene of *Coffea arabica* and promoter analysis in transgenic tobacco plants. Plant Physiol Biochem 37:273–282

Meyers BCCD, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

Michelmore RW, Meyers BB (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Mizuno K, Kuda A, Kato M, Yoneyama N, Tanaka H, Ashihara H, Fujimura T (2003) Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (*Coffea arabica* L.). FEBS Lett 534:75–81

Moisyadi S, Neupane KR, Stiles JI (1998) Cloning and characterization of a cDNA encoding xanthosine-N7-methyltransferase from coffee (*Coffea arabica*). Acta Hort 461:367–377

O'Brien TG, Kinnaird MF (2003) Caffeine and conservation. Science 300:587

Ogawa M, Herai Y, Koizumi N, Kusano T, Sano H (2001) 7-methylxanthine methyltransferase of coffee plants gene isolation and enzyme properties. J Biol Chem 276:8213–8218

Pang PP, Pruitt RR, Meyerowitz EM (1988) Molecular cloning, genomic organization, expression and evolution of 12s seed storage protein genes of Arabidopsis thaliana. Plant Mol Biol 11:805–820

Pinto-Maglio CAF, Cruz ND (1998) Pachytene chromosome morphology in Coffea L. II. C. arabica L. complement. Caryologia 51:19–35

Ponchon LDC, Kajava AV, Fesquet D, Padilla A (2004) NMR solution structure of Mobl, a mitotic exit network protein and its interaction with an NDR kinase peptide. J Mol Biol 337:167–182

Rick CM (1971) Some cytogenetic features of the genome in diploid species. Stadler Sym 1:153–174

Rogers WJ, Bezard G, Deshayes A, Meyer I, Petiard V, Marraccini P (1999) Biochemical and molecular characterization and expression of the 11S-type storage protein from *Coffea arabica* endosperm. Plant Physiol Biochem 37:261–272

Rojas-Herrera RL-VV (2002) Induction of a class III acidic chitinase in foliar explants of Coffea arabica L. during somatic embryogenesis and wounding. Plant Sci 163:705–711

Romning CM, Stegalkina SSAR, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD, Fry WE, van der Hoeven R, Tanksley S, Zhang PF, Jin HL, Yamamoto ML, Baker BJ, Buell CR (2003) Comparative analyses of potato expressed sequence tag libraries. Plant Physiol 131:419–429

Schroder G, Eichel J, Breinig S, Schroder J (1997) Three differentially expressed S-adenosylmethionine synthetases from Catharanthus roseus: molecular and functional characterization. Plant Mol Biol 33:211–222

Suh MCKM, Hur CG, Bae JM, Park YI, Chung CH, Kang CW, Ohlrogge JB (2003) Comparative analysis of expressed sequence tags from Sesamum indicum and Arabidopsis thaliana developing seeds. Plant Mol Biol 52:1107–1123

Sun CXPS, Olsson H, Boren M, Ahlandsberg S, Jansson C (2003) A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. Plant Cell 15:2076–2092

BNSDOCID: <XP_____19322122A__I_>

Uefuji H, Ogita S, Yamaguchi Y, Koizumi N, Sano H (2003) Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. Plant Physiol 132:372–380

White JATT, Newman T et al (2000) A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. Plant Physiol 124:1582–1594

Wrigley G (1988) Coffee. (Logman Scientific & Technical)

Zdobnov EM, Apweiler R (2001) InterProScan—an intergration for the signature-recognition methods in InterPro. Bioinformatics 17:847–848

BNSDOCID: <XP_____19322122A_I_>

XP009079141 (11)

Annals of Botany 97: 755–765, 2006
doi:10.1093/aob/mcl032, available online at www.aob.oxfordjournals.org

Isolation and Characterization of cDNA Encoding Three Dehydrins Expressed During *Coffea canephora* (Robusta) Grain Development

CÉCILE HINNIGER[1], VICTORIA CAILLET[1], FRANCK MICHOUX[1,†], MOHAMED BEN AMOR[1,‡], STEVE TANKSLEY[2], CHENWEI LIN[2] and JAMES MCCARTHY[1,*]

[1]*Nestlé Research Center, Tours, 101, Avenue Gustave Eiffel, BP 49716, 37097 Tours Cedex 2, France and* [2]*Department of Plant Breeding and Genetics, Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA*

Received: 4 July 2005 Returned for revision: 20 November 2005 Accepted: 10 January 2006 Published electronically: 27 February 2006

• *Background and Aims* Dehydrins, or group 2 late embryogenic abundant proteins (LEA), are hydrophilic Gly-rich proteins that are induced in vegetative tissues in response to dehydration, elevated salt, and low temperature, in addition to being expressed during the late stages of seed maturation. With the aim of characterizing and studying genes involved in osmotic stress tolerance in coffee, several full-length cDNA-encoding dehydrins (CcDH1, CcDH2 and CcDH3) and an LEA protein (CcLEA1) from *Coffea canephora* (robusta) were isolated and characterized.
• *Methods* The protein sequences deduced from the full-length cDNA were analysed to classify each dehydrin/LEA gene product and RT–PCR was used to determine the expression pattern of all four genes during pericarp and grain development, and in several other tissues of *C. arabica* and *C. canephora*. Primer-assisted genome walking was used to isolate the promoter region of the grain specific dehydrin gene (CcDH2).
• *Key Results* The CcDH1 and CcDH2 genes encode Y_3SK_2 dehydrins and the CcDH3 gene encodes an SK_3 dehydrin. CcDH1 and CcDH2 are expressed during the final stages of arabica and robusta grain development, but only the CcDH1 transcripts are clearly detected in other tissues such as pericarp, leaves and flowers. CcDH3 transcripts are also found in developing arabica and robusta grain, in addition to being detected in pericarp, stem, leaves and flowers. CcLEA1 transcripts were only detected during a brief period of grain development. Finally, over 1 kb of genomic sequence potentially encoding the entire grain-specific promoter region of the CcDH2 gene was isolated and characterized.
• *Conclusions* cDNA sequences for three dehydrins and one LEA protein have been obtained and the expression of the associated genes has been determined in various tissues of arabica and robusta coffees. Because induction of dehydrin gene expression is associated with osmotic stress in other plants, the dehydrin sequences presented here will facilitate future studies on the induction and control of the osmotic stress response in coffee. The unique expression pattern observed for CcLEA1, and the expression of a related gene in other plants, suggests that this gene may play an important role in the development of grain endosperm tissue. Genomic DNA containing the grain-specific CcDH2 promoter region has been cloned. Sequence analysis indicates that this promoter contains several putative regulatory sites implicated in the control of both seed- and osmotic stress-specific gene expression. Thus, the CcDH2 promoter is likely to be a useful tool for basic studies on the control of gene expression during both grain maturation and osmotic stress in coffee.

Key words: Dehydrins, late embryogenic abundant protein (LEA), seed development, *Coffea*, *C. canephora*, *C. arabica*, Rubiaceae.

INTRODUCTION

A group of proteins, called the late embryogenesis abundant proteins (LEA), have been shown to accumulate in a co-ordinated fashion during the latter stages of cotton seed development (Dure *et al.*, 1981). Dehydrin proteins (DHN) are a sub-group of the LEA proteins that have also been called the 'LEA D-11 family' or LEA type 2 proteins (Dure, 1993; Close, 1996; Ingram and Bartels, 1996). Expression of the DHN proteins have been associated with the protection of various types of plant cells from osmotic stresses, such as those caused by desiccation, salt, and low temperatures (Skriver and Mundy, 1990; Close, 1996; Ingram and Bartels, 1996; Allagulova *et al.*, 2003).

* For correspondence. E-mail james.mccarthy@rdto.nestle.com
† Present address: Department of Biological Sciences, Imperial College London, South Kensington, London SW7 2AZ, UK.
‡ Present address: Laboratoire Génome et Développement des Plantes (UMR 5096), Université de Perpignan, 52 ave Paul Alduy, 66860 Perpignan, France.

Over the last few years, direct experimental evidence linking higher expression of dehydrins and protection from osmotic stress has begun to appear in the literature. For example, arabidopsis plants engineered to overexpress a dehydrin fusion protein were found to have improved survival to low temperature exposures (Puhakainen *et al.*, 2004). Similarly, expression of a citrus dehydrin protein in transgenic tobacco has been shown to give increased tolerance to low temperatures (Hara *et al.*, 2003). Other supporting evidence for the linkage of dehydrins and tolerance to low temperature-induced stress are the observations that QTL (quantitative trait loci) for freezing tolerance and winter-hardiness map very closely to dehydrins (Close, 1996; Zhu *et al.*, 2000). DHN genes are also expressed significantly in seeds towards the end of maturation, a period when the seed undergoes a developmentally programmed reduction in water content (Choi and Close, 2000; Nylander *et al.*, 2001). The LEA/dehydrin proteins have been estimated to comprise up to 4 % of the total seed protein, and

© The Author 2006. Published by Oxford University Press on behalf of the Annals of Botany Company. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

BNSDOCID: <XP_____9079141A_I_>

XP009079141

are thought to be involved in protecting the embryo and/or other seed tissues from the osmotic stresses associated with the low water content of mature seed (Roberts *et al.*, 1993; Wise and Tunnacliffe, 2004).

Although a considerable number of dehydrin proteins have been isolated and studied, the precise physiochemical and/or structural mechanism(s) whereby these proteins function to protect cells from osmotic stress *in vivo* is unknown. The dehydrins are very hydrophilic proteins and exhibit an unusually low level of recognizable structure (Close, 1996; Soulages *et al.*, 2003). A key characteristic of the dehydrins is the presence of one or more lysine-rich stretches of 15 amino acids, called the K motifs, that are predicted to form class A amphipathic alpha-helices (Dure, 1993; Close, 1996, 1997). Dehydrins can also contain two other motifs, an N-terminal Y segment (consensus V/TDE/QYGNP) and a serine-rich S segment which can be phosphorylated and is thought to participate in nuclear localization (Godoy *et al.*, 1994; Close, 1997). It has been proposed that the short amphipathic K segments of dehydrin polypeptides interact with solvent-exposed hydrophobic patches on proteins undergoing partial denaturation, and thereby interfere with protein aggregate formation (Close, 1996). Amphipathic K helixes could also be involved in binding membrane lipids and thus could play a more specific role in protecting lipoproteins, proteins located in membranes, and/or the membrane structure itself (Close, 1996; Koag *et al.*, 2003). An alternative proposal for at least part of the protective effect of dehydrins is the ability of these very stable, but relatively unstructured proteins to tightly bind and organize water molecules (Soulages *et al.*, 2003). This latter effect could help slow the water loss from cells under dehydration conditions, and also possibly improve the stability of certain macromolecules by the development of dehydrin-based regions of more tightly bound 'ordered' water around these molecules.

The quality of a coffee beverage is directly related to the biochemical composition and the structural features of the mature grain. However, currently little is known concerning the influence of important grain components like the polysaccharides, proteins and lipids, on coffee quality. One approach to addressing this question is to study the expression of genes encoding these major components, or enzymes involved in their synthesis, and to determine if any correlation can be established between the expression of these genes and specific coffee qualities. Towards this goal, the work presented here describes the isolation and characterization of cDNA encoding a group of proteins that are relatively highly expressed during late grain development, the dehydrins.

The coffee dehydrin cDNAs were identified in the recently established coffee EST (expressed sequence tag) database which contains more than 47 000 ESTs generated from different EST libraries, including several grain-specific and pericarp-specific libraries (Lin *et al.*, 2005). Here, five full length cDNAs isolated from these libraries, representing three distinct dehydrin genes that are strongly expressed in coffee grain, are described. In addition, a cDNA is presented which encodes an LEA transcript that is highly expressed during a very limited phase of grain maturation and may be associated with tissues which undergo programmed cell death. Using RT–PCR, the expression of these coffee genes have been studied in several tissues, and during several stages of development in both the grain and pericarp tissues of the coffee cherry. A genomic fragment encoding the putative promoter region of the coffee dehydrin CcDH2, which is exclusively expressed during grain development, has also been isolated and sequenced. The potential uses of this promoter for future studies of coffee grain development and osmotic stress tolerance, and how the dehydrin proteins could influence coffee flavour generation during the coffee roasting process are discussed.

MATERIALS AND METHODS

Plant materials

Freshly harvested roots, young leaves, stem, flowers and fruit at different stages of development [small green fruit (SG), large green fruit (LG), yellow fruit (Y) and red fruit (R)] were obtained from *Coffea arabica* 'Catura T-2308' grown under greenhouse conditions (25 °C, 70 % RH) in Tours, France. Freshly harvested young leaves were obtained from *Coffea canephora* 'BP409' grown under greenhouse conditions (25 °C, 70 % RH) in Tours, France, and roots, stems, flowers and fruit from *Coffea canephora* 'BP409' were obtained from plants field grown in Indonesia (Indonesian Coffee and Cacao Research Institute, ICCRI). After harvesting, the fresh tissues were frozen immediately in liquid nitrogen until RNA extraction.

RNA extraction and synthesis of cDNA

RNA was extracted from frozen tissues and treated with DNase I as described elsewhere (Simkin *et al.*, 2006). cDNA was then prepared from 4 µg of DNase I-treated RNA using oligo (dT_{20}) according to the protocol of the Superscript™ II Reverse Transcriptase kit (Invitrogen, Carlsbad, CA, USA). The absence of contaminating genomic DNA was verified using primers for a ubiquitously expressed coffee chalcone isomerase gene that spanned an intron in this gene (Simkin *et al.*, 2006).

RT–PCR analysis

PCR reactions were carried out using the *Coffea arabica* and *Coffea canephora* cDNA samples described above. The primers used are described in Table 1. PCR reactions (50 µL) were set up containing 10 µL of a 1/100-fold dilution of the cDNAs (except for CcDH2 where 10 µL of a 1/1000-fold dilution of the cDNA set was used), 1 µM of each primer, 5 µL of 10× ThermoPol Buffer (New England Biolabs, Beverly, MA, USA), 1 µL of DMSO, 200 µM of dNTPs and 2 units of Taq polymerase (New England Biolabs). The cycling conditions were 2 min at 94 °C, 35 cycles (except for CcDH1 where 40 cycles was used) of 94 °C for 1 min, 60 °C for 1 min and 72 °C for 1.5 min. The final extension step was for 7 min at 72 °C. The RT–PCR products were resolved on 2 % (w/v) agarose gels and stained with ethidium bromide. The CcRPL39 gene

BNSDOCID: <XP._____9079141A__I_>

XP009079141

TABLE 1. *List of primers used for RT–PCR and quantitative RT–PCR experiments*

Oligo name	5′→3′ Sequence	Oligo name	5′→3′ Sequence
rpl39_forward	TGGCGAAGAAGCAGAGGCAGA	DH2_reverse	TCAAGCGCACAACAAGGTCC
rpl39_reverse	TTGAGGGGGAGGGTAAAAAG	DH3_forward	AGGTGGTGGTCAGAAGAAGAC
DH1_forward	GAAGAAGGGGATGAAGGAG	DH3_reverse	GACACACTGGAAAGCTGCTA
DH1_reverse	TACGGACAAACACACTACAG	LEA1_forward	CCAATAACAGCTCAAGAATCA
DH2_forward	CCTCCAACAACCACCACTG	LEA1_reverse	TTCCCTTCCATCCCACTCT

encodes the constitutively expressed coffee L39 protein (a 60S ribosomal large subunit protein). The expression of this gene was used as a semi-quantitative control to verify that each RNA sample was transcribed into cDNA at relatively similar efficiencies.

Isolation and characterization of the DH2 promoter region

The promoter sequence of CcDH2 was isolated using the method described in the Genome Walker® kit (BD Sciences-Clontech). The method used to isolate the genomic DNA from *Coffea canephora* 'BP409' was described by Crouzillat *et al.* (1996). The CcDH2 specific forward Genome Walker® primer used was: DH2aprimer1–5′-TGTGCTCCTGATGCTCTCTGTCCTTGTGC-3′. An approx. 2.1-kB fragment was isolated using *Hin*dIII-digested *Coffea canephora* 'BP409' genomic DNA ligated to the Genome Walker® adaptor sequence. The PCR was carried out in a 50-μL reaction using the Clontech Advantage® 2 PCR kit as described by the supplier using 0.5-μM final concentrations of DH2aprimer1 and the Genome Walker® AP1 primer. The PCR reaction was carried out using the following conditions: 94 °C for 2 s and 72 °C 3 min (seven cycles), 94 °C for 2 s and 67 °C 3 min (32 cycles), followed by 4 min at 67 °C. The major PCR fragment obtained was then cloned into the plasmid pCR4-TOPO® (Invitrogen). A plasmid (pJMc1) containing the appropriate insert was purified and its insert was completely sequenced. To verify that the inserts in pJMc1 and pcccs30w8a4 were from the same gene, the corresponding overlapping sequence of these two clones was re-amplified from genomic DNA using the primers DH2a geneup 5′-ATAGTGACCTTAATAGCGATCTTGTTGC-3′ and DH2a genelow 5′-CCAAATCAAATCAAACCAAGCAAATC-3′. The PCR reaction was performed with *Coffea canephora* 'BP409' genomic DNA and using Taq (New England Biolabs) as described by the supplier except 1 μM of the specific primers was used (DH2a geneup and DH2a genelow). The PCR conditions were 94 °C for 1 min, then 35 cycles of 94 °C for 1 min, 58 °C for 1.5 min and 72 °C for 3 min, followed by 7 min at 72 °C. The main PCR fragment produced was then cloned into pCR4-TOPO®. The resulting plasmid containing the appropriate insert (pVC1) was purified and its insert was completely sequenced.

Southern blot analysis

Genomic DNA was prepared as described previously (Crouzillat *et al.*, 1996). Five micrograms of genomic DNA from *C. canephora* BP 409 DNA was digested overnight with the appropriate enzymes (10 U μg^{-1}) according to the supplier's recommendations and the products were separated on a 0.8 % agarose gel. Southern blotting and hybridizations were carried out as described previously (Crouzillat *et al.*, 1996). The probe was generated by first PCR amplifying the insert of the CcDH2 clone cccs30w8a4 with the primers T3 + T7. This PCR product was then labelled with [^{32}P]dCTP using the 'rediprime™ II random prime labelling system' kit (Amersham).

RESULTS

Identification and characterization of coffee dehydrin cDNA

An EST database of >47 000 *C. canephora* EST sequences has recently been generated using RNA isolated from young leaves and from the grain and pericarp tissues of cherries harvested at different stages of development (Lin *et al.*, 2005). An annotated 'unigene' set (i.e. contig set) of 13 175 sequences was then obtained by clustering the most highly similar, overlapping ESTs and annotating the unigenes thus generated. The unigenes were screened for dehydrin sequences using various approaches, including a search of the unigene annotations with the keyword 'dehydrin' and by using various arabidopsis and tomato dehydrin protein sequences to do a tBlastn search of the coffee unigene set. The various search protocols yielded several candidate dehydrin unigenes, and the longest cDNA clone for each of these unigenes was isolated from the library and completely sequenced.

DNA sequence analysis of the selected full-length cDNA clones indicated that there were essentially four unique sequences representing three different dehydrin genes. The corresponding genes have been named CcDH1, CcDH2 and CcDH3. Two apparently allelic sequences of the gene CcDH1 were identified by sequencing two of the longest cDNAs in unigene #121870. The cDNA clones CcDH1a (pcccl26i7) and CcDH1b (pcccs30w27m8) were 880 and 901 bp long, respectively. The two ORF (open reading frame) sequences exhibited five single base changes, and CcDH1b had an insertion of nine bases, which together resulted in six amino acid changes. CcDH1a and CcDH1b encode proteins of 172 amino acids and 175 amino acids having predicted molecular weights of approx. 17.8 kDa and 18.1 kDa, respectively. Several other sequence differences also exist in the untranslated regions, including a 12-bp deletion in CcDH1a. Two distinct poly (A)-containing unigenes, #123406 and #123405, were found to encode the ORF for CcDH2. Unigene #123405 (pcccs46w30p1; CcDH2b) is composed of two ESTs and differs from the unigene

BNSDOCID: <XP_____9079141A_I_>

XP009079141

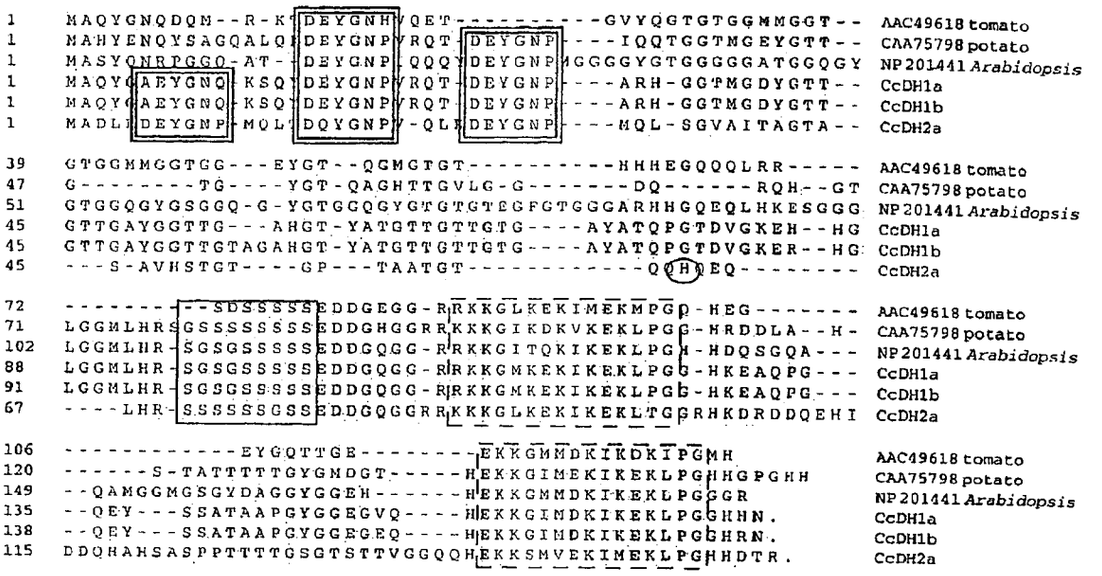


FIG. 1. The optimized alignment of *C. canephora* Y_3SK_2-type dehydrins with several of the closest plant homologues. This alignment was generated using the Clustal W program in the MegAlign software (DNASTAR) and then further optimized manually. Shaded blocks indicate identical amino acids. The doubled rectangles demarcate the Y-segments, the single dark rectangles demarcate the S-segments, and the rectangles with broken lines demarcate the K-segments. The black circle represents the single amino acid difference between CcDH2a and CcDH2b. Accession numbers: CcDH1a, DQ323987; CcDH1b, DQ323988; CCDH2a, DQ323989; *Lycopersicon esculentum* dehydrin TAS14, AAC49618; *Solanum commersonii* dehydrin Dhn1, CAA75798; *Arabidopsis thaliana* dehydrin RAB18, NP_201441.

sequence #123406 (pcccs30w8a4; CcDH2a) by the presence of an intron sequence. Interestingly, when the intron/exon borders of the cDNA for DH2b (cccs46w30p1; accession number DQ323990) were examined in detail, it was observed that the 3′ junction had the sequence ttatgg/TCG while other genomic sequences obtained by genome walking (see below) had the sequence ttatag/(T or A)CG. The 761-bp cDNA pcccs30w8a4 (CcDH2a) encodes a protein 162 amino acids long with the predicted molecular weight of 17.4 kDa. Finally, a third coffee gene, CcDH3, was encoded by a single unigene #123385. This gene is represented by the 835-bp insert of plasmid pcccwc22w11a5, which encodes a protein of 227 amino acids with the predicted approximate molecular weight of 25.1 kDa.

Alignments of the coffee dehydrins with the most homologous protein sequences found in the non-redundant protein database, and the analysis of the presence of the dehydrin-specific amino acid motifs Y, S and K (Close, 1996), are shown in Figs 1 and 2. The dehydrins fall into two classes, with CcCDH1 and CcDH2 having the structure Y_3SK_2 and CcDH3 having the structure SK_3. Of those protein sequences with the Y_3SK_2 structure, CcDH1a and CcDH1b showed absolute conservation in each of the three motifs, as well as in the two conserved regions that precede each of the two K motifs, indicating that these two sequences are allelic. In contrast, CcDH2 is clearly different from CcDH1. While CcDH2 has the structure Y_3SK_2, it exhibits punctual differences from CcDH1 in all but one of the Y, S and K motifs, as well as more significant differences outside these dehydrin-specific motifs. The CcDH1 and CcDH2 encode proteins with calculated pIs that are near neutral, and their hydrophilicity plots indicate that these proteins are very hydrophilic throughout (data not shown). The calculated pI of the protein encoded by CcDH3 is slightly acidic (5.47), and this protein is also very hydrophilic.

Characterization of a cDNA encoding a coffee LEA protein

Using a cDNA library constructed from RNA prepared from coffee grain 30 weeks after fertilization (M. Ben Amor *et al.*, unpubl. res.), a full-length cDNA clone (pDav1-59; 1232 bp) encoding an LEA protein was isolated and sequenced and renamed CcLEA1. Also the EST databas. was searched for 'unigenes' annotated as LEA proteins. This search produced nine gene sequences, one of which (unigene #119994) corresponded to the previously isolated sequence of CcLEA1 (Table 2). It was decided to focus on CcLEA1 in more detail in this work because of its unusual expression pattern. Based on its EST representation, this gene appears to be strongly and exclusively expressed during only one period of grain development (30 weeks after flowering). The protein sequence encoded by CcLEA1 is 357 amino acids and has a predicted molecular weight of 39.5 kDa. The calculated pI for CcLEA1 is slightly basic (8.17), and hydrophilicity plotting shows that, although this protein does not have large regions of hydrophobicity, it is clearly less hydrophilic than the three coffee dehydrins. It is noted that the first 30 N-terminal residues of CcLEA1 form one of the two small hydrophobic regions found in this protein. Figure 3 shows the alignment of CcLEA1 with the three most homologous sequences found in the non-redundant Genbank protein database. While the overall identity values of these aligned sequences only ranged from 34.7 % for the arabidopsis sequence to 47.9 % for

BNSDOCID: <XP_____9079141A__I_>

XP009079141

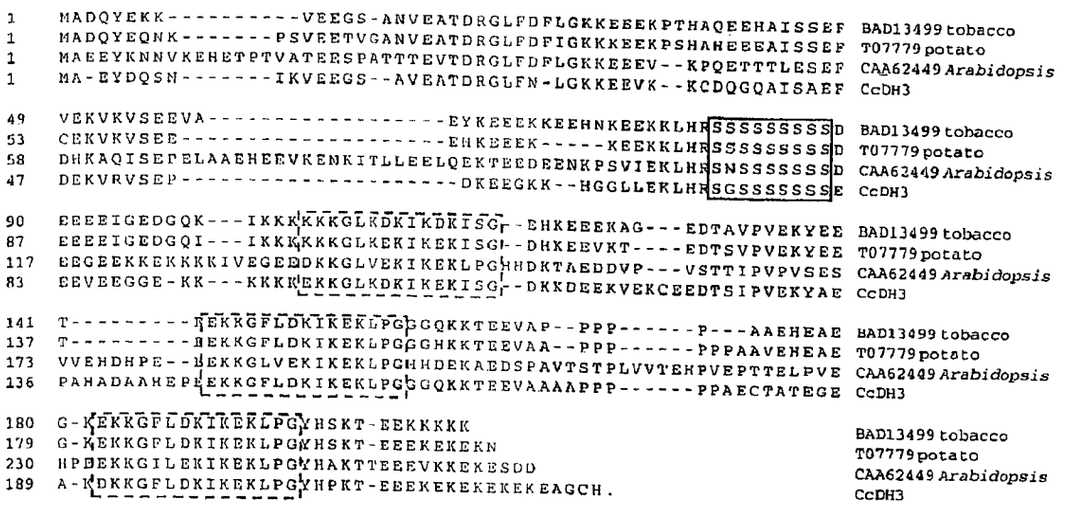


FIG. 2. The optimized alignment of the *C. canephora* SK_3-type dehydrin with several of the closest plant homologues. This alignment was generated as described for Fig. 1. Shaded blocks indicate identical amino acids. The single dark rectangles demarcate the S-segments, and the rectangles with broken lines demarcate the K-segments. Accession numbers: CcDH3, DQ333960; *Nicotiana tabacum* dehydrin, BAD13499; *Solanum tuberosum* dehydrin homologue CI7, T07779; *Arabidopsis thaliana* dehydrin, CAA62449.

TABLE 2. Coffea canephora *unigene sequences annotated as late embryogenesis abundant proteins (the corresponding unigene sequences can be found at http://www.sgn.cornell.edu)*

		In-silico expression (no. of ESTs in libraries)						
Unigene name	No. of ESTs	Whole cherries 22 weeks	Seed 18 weeks	Seed 30 weeks	Seed 42 weeks	Seed 46 weeks	Pericarp	Leaf
124784	26			17		9		
124426	26					9		17
119994	22			22				
124428	10	1				3	4	2
121426	8		1	5		2		
124123	5	3					1	1
120926	2					2		
131122	1					1		
127677	1					1		

the *Picea* (white spruce) sequence, a closer examination of the alignment shows that there are several short, but highly conserved regions in these proteins. It is also noted that all the related protein sequences had relatively similar hydrophilicity profiles to CcLEA1 and each had it's most significant hydrophobic patch located in the N-terminal 1–25 amino acids, possibly indicating this protein has a leader peptide sequence to direct synthesis into the endoplasmic reticulum (data not shown). Interestingly, CcLEA1 has a rather striking proline-rich segment in the N-terminal region that is clearly absent from the other proteins in Fig. 3. However, CcLEA1 does have the conserved set of cysteine residues that were previously defined in the maize root cap proteins.

RT–PCR expression analysis of CcDH1, CcDH2, CcDH3 and CcLEA-1 genes in different coffee tissues and during coffee grain development

RT–PCR analysis was carried out for each of the three coffee dehydrin genes, and the results of this experiment are shown in Fig. 4. CcDH1 was expressed significantly in arabica grain at all the stages examined, and in the three last stages examined for robusta. It has previously been shown that there are significant differences between the gene expression profiles observed between the small green arabica and small green robusta grain samples used here. For example, small green grain from arabica, but not from robusta, have detectable transcripts of genes associated with endosperm expansion such as the oleosins and the 11S storage protein (Simkin *et al.*, 2006). Close examination of the original gels indicated that very low levels of CcDH1 expression could also be detected in several of the other tissues tested, although no signal was detected for arabica in the small green pericarp and yellow pericarp samples, or for robusta in the root or young leaf samples. Among the tissues having expression of CcDH1, the arabica flower sample appeared to have the highest level of transcripts. A relatively high level of CcDH2 transcripts was also detected in all the grain-development stages of arabica and the last three stages of the robusta grain examined

BNSDOCID: <XP_____9079141A_!_>

```
                                                                                *
1   MEI-SKTLLLVISLVA----ATCFLQA------K--------AAG----------------VYCS  NP200248 Arabidopsis
1   MAANK-L--MNVVAVA--V--CVMVMVN------A----ASAVG----------K---AKCT  T09288 Picea glauca
1   MARLCALIPLATVLLA----AVAATAAPSDRPPKAQGPKPHKEKEKEKP--------KPMKVKCR  BAA75477 Zea mays
1   MQKMTPLRCINFIFLAFWVPAVLAVMAEKPLVPTYLIPKPPPPPSPVKPSVPVIPVKPRIVRCR  CcLEA1
                *       *   *                   *   *   *           *       *
31  N--PYTRCYRKYIRCPEECPSKTAMNSKNKVCYADCDRPTCKSQCRMRKPNCNRPGSACYDPRF  NP200248 Arabidopsis
33  D-KWYPRCYGYQYDCPANCPY----N------CDMDCK--TCKTVC-----PCDKPGGVCQDPRF  T09288 Picea glauca
54  PRKLYPYCPGKPMECPAECSQ-----S----CYADCS--SCKPVC-----VCSVPG-ACGDPRF  BAA75477 Zea mays
65  S-TLFPLCFNIPPVCPLDCLT----N-----CLVDCV--TCKAYC-----SCNFPGAVCQDPRF  CcLEA1
                o
93  IGGDGIVFYFHGKSNEEFSLVSDSDLQINGRFIGHRPA-GRARDFTWIQALGFLFNS-NKFSLE  NP200248 Arabidopsis
80  IGGDGIMFYFHGKRDQDFCLISDSNLHINAHFIGKRGQ-GMGRDFTWVQSIGVLLEDGRQFYLG  T09288 Picea glauca
101 IGGDGNAFYFHGRRDADFCVLSDRDLHINAHFIGKHGADGMSRDFTWIQAIAVLFD-GHELYVG  BAA75477 Zea mays
112 VGGDGNTFYFHGRKDQDFCLVSDTNLHVNGHFIGKRKPN-LRRDFTWVQAIGIMFDD-HRILVA  CcLEA1

155 AAKTASWDNEIDHLKFSYDGQDLSVPEETLSTWYSP-NKDIKIERVSMRNSVIVTIKDKAEIMI  NP200248 Arabidopsis
143 AKKVSTWDNSVDQLTMALNGQTLTLPPGEGATWAT-A-SGLNVTRSDRANEVVVQVEDKLKISA  T09288 Picea glauca
164 ARKTAAWDDDVDRMELTLDGEPVRLLPGTDAAWTSGAVPALSVTRTSAANGVLVSLDGRFTIRA  BAA75477 Zea mays
174 AKRTSTWDDNVDRLAISIDGNPISLPTEEGSKWQLPAPSNVSIMRTSNNNGLVVEAVNNFRITA  CcLEA1
                *
218 NVVPVTKEDDRIHSYKVPS-DDCFAHLEVQFRFFNLSPKVDGILGRTYRPDFQNPAKPGVAMPV  NP200248 Arabidopsis
205 RVVPISEEESRVHNYGIIAGEDCFAHLELSFKFYSLSPNVSGVLGQTYGAEYRSPVKMGVAMPI  T09288 Picea glauca
228 NAVPITEEESRVHRYGVTA-DDCLAHLDLAFKFGALTADVHGVVGQTYRSDYVNRFDVKASMPT  BAA75477 Zea mays
238 NVVPITAQESKVHGYDITD-EDCFTHLELGFKFFNITDSTDGVLGQTYRSDYVNKMKVNAVMPV  CcLEA1
                *               *               o
281 VGGEDSPKTSSLLSNDCKTCIF--S-ESQAEIDSVKSEIEYATLDCTRGASSGYGIVCRK      NP200248 Arabidopsis
269 MGGESNYVTSNLFAADCKVARFA----S----SSDDEYAITSALDCNSGRGSGHGIVSRR      T09288 Picea glauca
291 MGGDSNYTTSSLFAADCAVARYAPSGGSRDRDDGVAMVSEIAGITCSSGMG-GQGVVCKK      BAA75477 Zea mays
301 MGGDRKYLTSGLFSADCAVSRFG--GKVLEKANSASPVHEYPALNCKSGME-GNGLVCKK.      CcLEA1
```

FIG. 3. The optimized alignment of the *C. canephora* late embryogenesis abundant protein CcLEA1 with several of the closest plant homologues. This alignment was generated as described for Fig. 1. Shaded blocks indicate identical amino acids. Conserved cysteines are marked by an asterisk, and the positions of two less highly conserved cysteines are marked by a circle. Accession numbers: CcLEA1, DQ333961; *Arabidopsis thaliana* late embryogenesis abundant related protein, NP_200248; *Picea glauca* late embryogenesis abundant protein EMB7, T09288; *Zea mays* root cap protein 2, BAA75477.

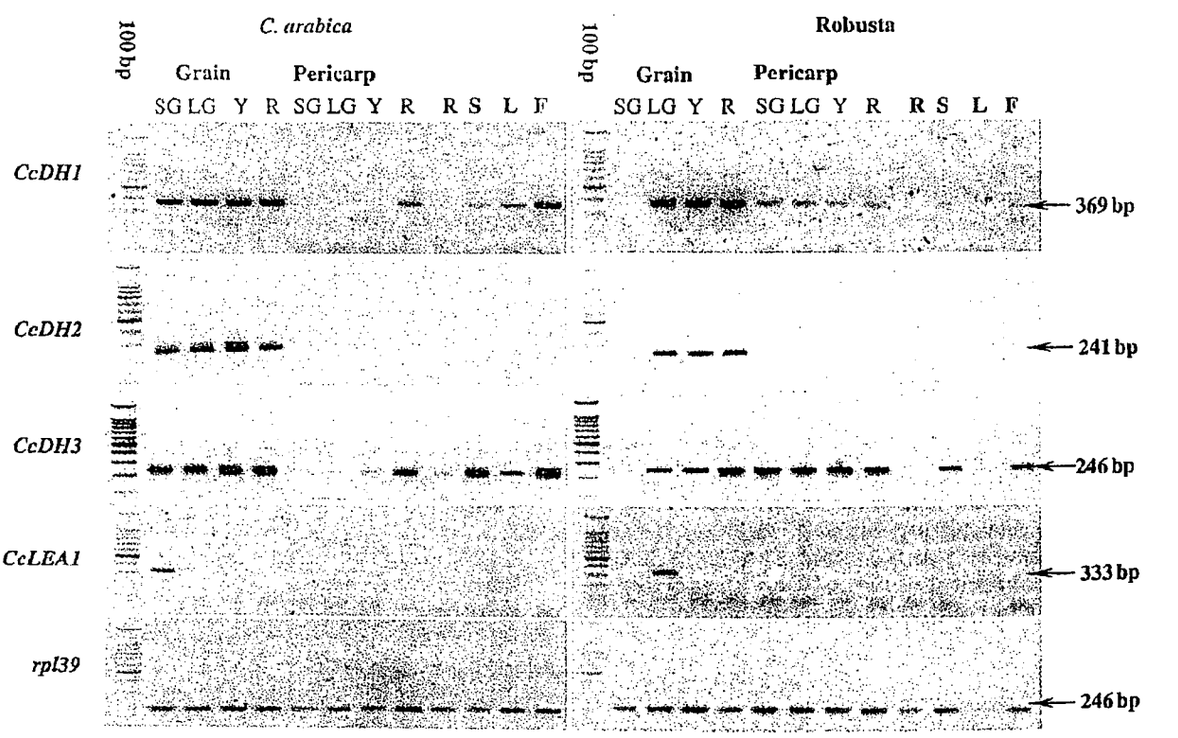


FIG. 4. RT–PCR expression analysis of the coffee dehydrins and CcLEA1 transcripts in different organs of *Coffea arabica* and *C. canephora*. 100 bp, 100-bp molecular weight marker ladder; SG, LG, Y, R, small green, large green, yellow and red for the grain and pericarp, respectively; R, S, L, F, root, stems, young leaves and flowers, respectively.

BNSDOCID: <XP_____9079141A__!_>

XP009079141

(Fig. 4). In contrast to CcDH1, no CcDH2 transcripts were detected by RT–PCR in the other tissues studied.

RT–PCR analysis of CcDH3 gene expression demonstrated that these transcripts were also detected in all the arabica grain samples, as well as in the last three stages of robusta grain development (Fig. 4). CcDH3 transcripts could also be clearly detected in some of the other tissues, such as the red pericarp, stem and flowers. Close examination of the original data showed that CcDH3 transcripts could actually be detected in all of the other arabica and robusta tissues examined. A significant difference was noted between the transcript levels for the first three pericarp stages of robusta and arabica. Whether this difference reflects an intrinsic difference between arabica and robusta cherries, or simply reflects some environmental difference between the samples remains to be determined. Also the expression of CcLEA1 was examined using RT–PCR. The data obtained confirms that this gene has a very unique expression pattern, with transcripts being detected only in the small green stage of arabica and the large green stage of robusta grain (Fig. 4). No expression was detected in any of the other arabica or robusta tissues sampled. These data are consistent with the distribution pattern of ESTs for this gene in the EST libraries, which indicated that this gene is expressed in robusta grain at 30 weeks after flowering (WAF) but not in any of the other grain libraries, or in the cherry, pericarp and leaf libraries.

Isolation and sequence analysis of the CcDH2 promoter

It is desirable to develop a repertoire of grain-specific promoters for future studies on transcriptional control in coffee, and for use in the construction of model recombinant genes to be expressed specifically in the grain. The data presented earlier demonstrate that the promoter of the CcDH2 gene is a relatively strong grain-specific promoter. Thus, it was decided to isolate a genomic fragment which incorporates this promoter by employing the Genome Walker technique. To ensure the strength of this promoter was associated with a single/low copy gene, first the copy number of the CcDH2 gene was estimated by Southern blotting. The data presented in Fig. 5 shows that each restriction enzyme digestion produced a single band, strongly suggesting that CcDH2 is encoded by a single gene in the *C. canephora* genome. Using a CcDH2-specific Genome Walker primer designed from the centre of the cDNA pcccs30w8a4 (DH2a primer 1), together with the API primer of the Genome Walker kit, a genomic fragment from *C. canephora* that contained over 1.5 kb of DNA upstream of the pcccs30w8a4 cDNA sequence (pJMc1) was successfully PCR amplified and cloned. To amplify a single complete genomic fragment containing the DH2 promoter region and transcribed region, a new round of PCR was performed as described in Materials and methods using *C. canephora* DNA. This experiment resulted in the generation of the genomic fragment cloned in pVC1 which extended 1.43 kb upstream of the corresponding cDNA sequence (pcccs30w8a4). There were six base changes between the genomic sequences of pJMc1 and pVC1; one was in the promoter region, two were in the intron,

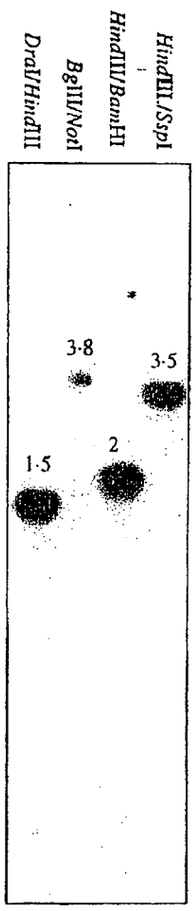


FIG. 5. Southern blot analysis of the CcDH2 dehydrin.

and three were in the protein-coding sequence (Thr/Ala, Thr/Ser and Gly/Gly). Examination of an alignment of the four sequences available for CcDH2 (two genomic and two cDNA) indicates that these sequences represent three different alleles. These alleles are manifest by the SNPs associated with the amino acid changes noted above. The cDNA sequence in CcDH2a is identical to the coding sequence of the genomic clone pJMc1, indicating that these sequences probably represent the same allele.

Sequence analysis of the composite sequence obtained from the plasmids containing both genomic and cDNA sequences showed that the CcDH2 gene contains a single intron (231 bp) located within the ORF region (Fig. 6). Further analysis of the promoter region of this gene indicates the presence of a putative TATA sequence 30 bp upstream of the 5′ end of the cDNA sequence. Several potential regulatory elements previously shown to be involved in the regulation of gene expression during seed development and during osmotic stress can also be identified in the 5′ upstream region of the CcDH2 gene. For example, three regions with similarity to the arabidopsis ABA-responsive element RYACGTGGYR (Iwasaki *et al.*, 1995) were found. Two other elements were found that shared significant similarity with the RY repeat (CATGCA(T/a) (A/g) of the 'legumin' box that is involved in regulating the

BNSDOCID: <XP_____9079141A_I_>

```
atagtgacct taatagcgat cttgttgctt ttgatcgtca gaaaagtagt ggacatgacg      60
gaagaggtcc taagatgagt tccagttcca gcatgaaggg ctctttggcg aagcctttct     120
tgaggcgtca cttttctttt ggatctaaag gcagtagatc aatgtcagag aatcattctt     180
cctggaagag gggattcttc tgggcaaaat cgagaaagga ttaagttctg tctagagtta     240
caaaggtgag caacagtcac ggtttttat tagggaatgg aaggattgga tccctttca      300
cgtagtgaac aacatatatt ttgcatggtt ggtcttagta cctataacac gaaaatgttc     360
ttcatccgtt ctattaatca ttaggcttta gtcatttaaa ttttttacat cccgcatttc     420
tcctcttgat tcttgttgat ttctgcagat tccacagttg ttcttcagat gggctacgaa     480
atgcatgcag ggagcaggca atcagccata aattcaaccc tgtcaaggaa gctggcattg     540
tctcgtgcaa atgtaggtta gcttttgaag atacactgca aagggaagac catacagatg     600
gggaaatgaa ttcattataa tataggaaaa aggaaagatg ataggggtca gggcgtccgt     660
gcatcatgaa actagttctc tttcattttg tacgatggct gtttactgtt taatttcatg     720
aaattagttt ggatatatgc gtagcgtttt accatcgcat ttctaaatcg atattctatg     780
ggccgaatta cgcgttggag acatcattgg gttgctcctc tcaatcccat ctctatctat     840
tgacggatcc ggatcatgat gttgaacctt tcaacttttg acttagatgg gatttgtgtt     900
cgcgtgttgt taacttgtta ctgaccgact cagaagacag cggattctga cttcaccacg     960
tgtctcttta gtgaaaattt aaaaggcatt tttcttctgt tcatagttta aaatgtaatg    1020
tgattattaa aagatcgttt ggtattattt caaggatgga tggattggat ggaagggata    1080
tctgatatat atcataccct tccaaaattc aggaccatga cgtatttaat atcccccagc    1140
ggaagacacg tgccttgatg tcttataggt ggcaatacac ttcagcttcc tctgctaata    1200
cgtgtgagga tcttcggtac catgcagaaa agaccgcggt gctccttcca ccgtcctcat    1260
ccctctcttg gctttttaa gtctcctgcg atatccaaaa tccaaacaaa gccgttatcg     1320
cagcdaaaat tcgtcaaccc caagtctcag gctaccttaa tttcagtgcc ctttttcttt    1380
atttttttct aataacagga gtcctggaaa atg gct gac ttg cgt gat gaa tat      1434
                              Met Ala Asp Leu Arg Asp Glu Tyr
                              1               5
gga aat cct atg cag ttg acc gac cag tat ggc aac ccg gtt cag ctc      1482
Gly Asn Pro Met Gln Leu Thr Asp Gln Tyr Gly Asn Pro Val Gln Leu
    10                  15                  20
aag gac gag tat ggc aac cca atg cag ctt agc ggt gta gct atc acc      1530
Lys Asp Glu Tyr Gly Asn Pro Met Gln Leu Ser Gly Val Ala Ile Thr
25                  30                  35                  40
gcc ggg acg gct agt act gtc cat tct act gga acc gga cca act gct      1578
Ala Gly Thr Ala Ser Thr Val His Ser Thr Gly Thr Gly Pro Thr Ala
                45                  50                  55
gcc act gga acc cag caa cat cag gag cag ctt cat cgg tct agc agc      1626
Ala Thr Gly Thr Gln Gln His Gln Glu Gln Leu His Arg Ser Ser Ser
            60                  65                  70
tca agc tct ggc tcg   gtgagatact tgccaagtta caatgtgtgt gtctgtgtgt    1681
Ser Ser Ser Gly Ser
        75
gtataatgcg ccatcataat tgtttgcttg acagatcctg ttaataatga accgtaattt    1741
gacgtaaagt gtacacgttt tgtttttctg ggactaacat aatatcgaat caggctcctg    1801
ttgaatttga atgttgttag ctaaaagaaa attttggtgg ctgagttgtt gaatttggtt    1861
tatag acg gag gat gat gga caa gga gga aga aga aag aaa aaa ggg ttg     1911
      Thr Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys Lys Gly Leu
            80                  85                  90
aaa gaa aag ata aag gag aaa cta acg ggc ggt agg cac aag gac aga      1959
Lys Glu Lys Ile Lys Glu Lys Leu Thr Gly Gly Arg His Lys Asp Arg
        95                  100                 105
gac gat cag gag cac atc gat gat cag cac gcg cac agc gcc tct cct      2007
Asp Asp Gln Glu His Ile Asp Asp Gln His Ala His Ser Ala Ser Pro
    110                 115                 120
cca aca acc acc act ggc agc ggg acg tct act aca gtc ggg ggt cag      2055
Pro Thr Thr Thr Thr Gly Ser Gly Thr Ser Thr Thr Val Gly Gly Gln
125                 130                 135                 140
cag cat gaa aag aag agc atg gtg gag aag att atg gaa aag ctc cct      2103
Gln His Glu Lys Lys Ser Met Val Glu Lys Ile Met Glu Lys Leu Pro
                145                 150                 155
ggc cat cac gac acc cgc tag ttaccta ccacaacata ctgtgatcat            2151
Gly His His Asp Thr Arg STOP
            160
cgtgtaaaat ctctcctgat gcctaggaaa tctagattat gttaggcatt ttgtttggta    2211
tgtatgtgtg attaagacct tgttgtgcgc ttgaatcttg aacgtgcatg ggatttgctt    2271
ggtttgattt gatttgg
```

G. 6. DNA sequence of the CcDH2a promoter and transcribed sequence from *Coffea canephora*. The nucleic acid sequence of pVC1 (accession number Q338457) insert is presented, and its corresponding amino acid sequence. The first base in the cDNA (C) is marked by a circle. The putative TATA box is own in bold and underlined. The RY repeat sequences are marked with grey, the ABA responsive elements are shown in boxes, and the DRE/CRT sequence is boxed with heavy lines.

BNSDOCID: <XP_____9079141A_I_>

XP009079141

expression of genes encoding the legumin type storage proteins (Shirsat *et al.*, 1989; Baumlein *et al.*, 1992). The presence of one dehydration-responsive element/C-Repeat *cis*-acting sequence motif (DRE/CRT; G/ACCGAC) has also been identified. A second DRE/CRT element is found in the first exon. DRE/CRT motifs have been shown to interact with DREBs/CBF transcription factors to control the response of linked genes to dehydration and other stresses in arabidopsis and rice (Dubouzet *et al.*, 2003). Finally, several E-box motifs (CANNTG), which are well defined components in the promoters of seed-specific genes such as the 2S storage protein (Chatthai *et al.*, 2004) and oleosins (Simkin *et al.*, 2006), have been identified in the CcDH2 promoter region.

DISCUSSION

The dehydrins are a group of specialized plant proteins that are involved in protecting cells from dehydration-related stresses (Close, 1996; Allagulova *et al.*, 2003). Dehydrins also participate in two important and rather specific aspects of seeds. (1) They are widely perceived to participate, with other LEA proteins, in the dehydration process that occurs during the late stages of seed maturation by assisting the acclimatization of seed tissues to the lower water content found in mature seeds (Close, 1996; Nylander *et al.*, 2001). (2) It is presumed that the dehydrins synthesized in seeds during maturation continue to stabilize the associated cellular structures during seed quiescence. In this latter context, it should also be noted that it has recently been proposed that dehydrins may also possess a radical-scavenging capability (Hara *et al.*, 2003) and have metal-binding properties (Alsheikh *et al.*, 2003), both characteristics that are likely to be useful during long periods of seed storage.

Despite the involvement of dehydrin proteins in plant resistance to osmotic stresses such as drought stress, and the probable importance of the dehydrins during grain development, little information is available on these genes in coffee. Thus it has been decided to take advantage of the new coffee EST collection to isolate cDNAs encoding highly expressed coffee dehydrins, and to establish the expression patterns of the dehydrins in different tissues, as well as during coffee grain and pericarp maturation. Five cDNAs from *C. canephora*, which represent three distinct dehydrin genes (CcDH1, CcDH2 and CcDH3) are presented here, as well as a cDNA representing an LEA gene (CcLEA1) with an extremely restricted expression pattern. The gene CcDH1 is represented by two cDNAs that differ by 21 bases and have over 95 % identity. The small sequence differences translate into six amino acid changes, three of which result from a 9 bp deletion in the ORF of CcDH1a. Because these two cDNAs show such a high level of identity, and the fact that the protein sequence changes fall outside the highly conserved regions in this protein sequence family (Fig. 1), it is considered that these two cDNAs represent different alleles of one gene.

Two distinct unigene sequences were found to encode an identical dehydrin protein (CcDH2a and CcDH2b). However, closer examination of full-length cDNA clones representing these two unigenes indicated that unigene #123405, which consisted of two ESTs, actually contained an intron within the predicted ORF sequence. This intron containing cDNA (CcCH2b; pcccs4630p1) also contained a poly (A) tail sequence indicating that the corresponding mRNA was polyadenylated but not spliced. Interestingly, the 3′ splice site of the intron in CcDH2b was slightly different from the equivalent genomic sequence of CcDH2a [Fig. 6; ttatgg/TCG versus genomic sequence ttatag/T(or A)CG]. The only other differences between the intron sequence of the cDNA and the intron sequences in three independently obtained genomic sequences were a few punctual base changes. None of these changes in the intron were conserved in each of the three genomic sequences. The fact that the two ESTs of unigene #123405 (CcDH2b) had the same splice site mutations and were found in two different grain specific libraries (30-week and 46-week libraries), strongly suggests that this intron-containing transcript is real and not a cloning artefact or sequencing error. Therefore, the available data strongly suggest that the absence of splicing in the two cDNAs of unigene #123405 were due to the single base pair change in the splice site (gg versus ag). As the 30-week and 46-week grain libraries were made from more than one variety, it is currently not known which varieties harbours this mutation. It is also not known if this mutation has any biological consequences. Interestingly, it has been been found recently that approx. 5 % of the cDNAs in a large set of full-length arabidopsis cDNAs harbour unspliced introns, which the authors have called 'retained introns' (Iida *et al.*, 2004). This result is consistent with the observation that cDNA with poly (A) tails and containing unspliced introns can be found relatively easily in the new coffee EST libraries (J. McCarthy *et al.*, unpubl. res.). Much further work is needed to determine to what extent such 'retained introns' are the result of problems such as slow primary transcript processing, or base changes causing the reduced/loss of splicing, or alternatively, whether more complex mechanisms/functions are involved.

Although their primary amino acid sequences are relatively different (CcDH1a shares 47.3 % identity with CcDH2a), both CcDH1 and CcDH2 can be classified as Y_3SK_2 dehydrins based on the distribution of the Y, S, and K motifs in the respective sequences. In addition to being relatively close in motif composition, CcDH1 and CcDH2 also exhibit some similarities in their patterns of expression. Both genes are strongly expressed during the later stages of grain development (Fig. 4). However, it appears that CcDH1 is also weakly expressed in several other tissues, albeit at different levels in arabica and robusta. In contrast, the expression of CcDH2 is limited to the grain in both species. It is noted that one CcDH2 EST was found in the leaf EST library, suggesting that this gene may, under some conditions, be expressed in the leaf. Further experiments are necessary to determine whether CcDH2 expression is inducible, for example, by dehydration stress or by exposure to cold temperatures. Given that the expression patterns for CcDH1 and CcDH2 are different, and the fact that the proteins vary slightly, even in the highly conserved Y, S and K motifs, could imply that these proteins may have

BNSDOCID: <XP_____9079141A_I_>

XP009079141

related, but not identical roles. For example, it is possible that the CcDH1 and CcDH2 dehydrins are involved in protecting different functional targets. Alternatively, they could have similar functional targets, but their expression is controlled differently to enable these genes to be induced differentially in particular tissues by specific developmental and/or environmental cues. The coffee dehydrin CcDH3 has the structure SK_3, and is also expressed during the late stages of grain development (Fig. 4). In addition, CcDH3 expression is detected in several other tissues, as was CcDH1 expression. This latter point suggests that the expression of CcDH1 and CcDH3 could be controlled by relatively similar signals, although differences in the absolute levels of transcripts indicates transcription induction and/or transcript stabilities are not identical for both genes. As for the majority of the dehydrins previously described from other plants, the precise functions of the coffee dehydrins described here are not yet known. Nonetheless, the significant expression of the CcDH1, CcDH2 and CcDH3 genes during the late stages of grain maturation clearly suggests that the three dehydrins play a significant role(s) in conditioning the coffee grain for maturation-induced dehydration, and probably are also important in protecting the mature grain tissues before germination. Future experiments will be aimed at determining whether one or more of these dehydrins are induced in coffee under conditions of water and low temperature stress.

The LEA gene described here (CcLEA1) has an unusual expression pattern, with transcripts being detected only during one relatively short period of grain development, and not in any other tissues. In robusta, CcLEA1 expression was only detected at the large green grain stage, and not in either the small green or yellow stages of grain development. This stage of grain development spans the period when the perisperm tissue undergoes a substantial size reduction and the endosperm expands significantly (unpubl. res.), thus suggesting that the CcLEA probably plays a specific role during the perisperm/endosperm transition. A Blast analysis of the protein database with the CcLEA1 protein uncovered three potentially related protein sequences from arabidopsis, *Picia* (white spruce), and maize. Although the highest level of identity was only 47.9 % for the *Picia* sequence, the protein alignment seen in Fig. 3 shows that the three related proteins did share several highly conserved blocks of homology, especially after the more variable N-terminal end of the protein. None of the homologous proteins have been assigned a function. However, the maize cDNA sequence, which is called 'root cap protein 2', was isolated with another highly similar cDNA called 'root cap protein 1' from maize root caps using a differential screen (Matsuyama *et al.*, 1999). Transcription of both genes was restricted to the outermost cells of the maize root cap. These cells are believed to be first associated with the production of mucilage and then, later, are sloughed off (Moore and McClelen, 1983). It also appears that, as they sloughed off, the root cap cells undergo cell death (Matsuyama *et al.*, 1999). Although expression of CcLEA1 was not detected in roots, it is quite possible that the root cap fraction of the total root samples tested here were very small and thus genes specifically expressed in root caps were below the level of detection. Further expression analysis using particular regions of the root should clarify whether CcLEA1 is expressed in the root cap region of coffee. It is interesting to speculate that this class of protein may play a similar role in both the root cap and the perisperm region as it develops into endosperm during grain maturation. Although no data concerning the putative arabidopsis LEA homologue (accession number NP200248) have been found in the literature, an examination of the arabidopsis MPSS expression database (http://mpss.udel.edu) shows that this arabidopsis gene is expressed in both the root and in germinating seeds. No expression was detected for this gene in any of the other tissues described in the arabidopsis MPSS database. The detection of transcripts for the arabidopsis homologue in root supports the idea that the arabidopsis protein is a homologue to the maize root cap protein and possibly CcLEA1. It will be interesting in the future to investigate whether the putative homologue expressed during the germination of arabidopsis seeds is associated with tissues undergoing cell death, and similarly, if the root cap proteins of maize are expressed during maize seed development and/or germination. Finally, one striking similarity between CcLEA1 and its putative homologues is the near absolute conservation of the cysteines, a feature which was first observed in the maize root cap proteins (Matsuyama *et al.*, 1999). However, the functional significance of these conserved cysteines is currently unknown.

Because the CcDH2 gene appears to be grain specific, the promoter of this gene was isolated for further study. The promoter sequence obtained (Fig. 6) has several previously recognized regulatory sequence motifs that are expected to be involved in the temporal and spatial control of CcDH2 dehydrin expression. Considering that the detailed testing of coffee grain-specific promoters can take more than 3–4 years due to the slow development and flowering of coffee, the possibility of testing the capacity of this promoter sequence is being examined to direct grain-specific expression in model plants such as arabidopsis or tomato. It is now well established that dehydrin proteins somehow participate in protecting plants from water-related environmental stresses (Allagulova *et al.*, 2003). Therefore the detailed information on the coffee dehydrins presented here, and the identification of other LEA cDNA in the coffee EST bank (Table 2), should open some new avenues for research on stress tolerance in coffee. For example, it is now possible to study the induction of these genes in different tissues in response to various stresses. It is also possible to examine whether there are variations in the expression of these genes in coffee varieties exhibiting dramatically different tolerances to drought and low temperatures. The information on the coffee dehydrin and LEA sequences presented here also opens up the possibility of investigating the potential relationship(s) between the poor storage capability of coffee grain and the expression levels of various dehydrin and LEA genes.

BNSDOCID: <XP_____9079141A_I_>

XP009079141

ACKNOWLEDGEMENTS

We thank Dr Vincent Petiard and the members of the coffee gene discovery group for their support and helpful discussions. In particular, we thank Maud Lepelley for her assistance with the DNA sequencing and bioinformatic analysis of the cDNA clones, and Dr Andy Simkin and Dr Isabelle Privat for comments on the manuscript. Finally, we thank Dr Ir Zaenudin Su and Dr Surip Mawardi of the Indonesian Coffee and Cacao Research Institute (ICCRI) for supplying samples of *C. canephora* 'BP409'.

LITERATURE CITED

Allagulova CR, Gimalov FR, Shakirova FM, Vakhitov VA. 2003. The plant dehydrins: structure and putative functions. *Biochemistry – Moscow* **68**: 945–951.

Alsheikh MK, Heyen BJ, Randall SK. 2003. Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. *Journal of Biological Chemistry* **278**: 40882–40889.

Baumlein H, Nagy I, Villarroel R, Inze D, Wobus U. 1992. Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *The Plant Journal* **2**: 233–239.

Chatthai M, Forward BS, Yevtushenko D, Stefanov I, Osuska L, Osusky M, *et al.* 2004. 2S storage protein gene of Douglas-fir: characterization and activity of promoter in transgenic tobacco seeds. *Plant Physiology and Biochemistry* **42**: 417–423.

Choi DW, Close TJ. 2000. A newly identified barley gene, Dhn12 encoding a YSK2 DHN, is located on chromosome 6H and has embryo-specific expression. *Theoretical and Applied Genetics* **100**: 1274–1278.

Close T. 1996. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. *Physiologia Plantarum* **97**: 795–803.

Close TJ. 1997. Dehydrins: a commonality in the response of plants to dehydration and low temperature. *Physiologia Plantarum* **100**: 291–296.

Crouzillat D, Lerceteau E, Petiard V, Morera J, Rodriguez H, Walker D, *et al.* 1996. *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. *Theoretical and Applied Genetics* **93**: 205–214.

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, *et al.* 2003. OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *The Plant Journal* **33**: 751–763.

Dure L. 1993. Structural motifs in LEA proteins of higher plants. In: Close TJ, Bray EA, eds. *Response of plants to cellular dehydration during environmental stress*. Rockville, MD: American Society of Plant Physiologists, 91–103.

Dure L, Greenway S, Galau G. 1981. *Biochemistry* **20**: 4162–4178.

Godoy JA, Lunar R, Torres-Schumann S, Moreno J, Rodrigo M, Pintor-Toro JA. 1994. Expression, tissue distribution and subcellular localization of dehydrin TAS14 in salt-stressed tomato plants. *Plant Molecular Biology* **26**: 1921–1934.

Hara M, Terashima S, Fukaya T, Kuboi T. 2003. Enhancement of cold tolerance and inhibition of lipid peroxidation by citrus dehydrin in transgenic tobacco. *Planta* **217**: 290–298.

Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, *et al.* 2004. Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic Acids Research* **32**: 5096–5103.

Ingram J, Bartels D. 1996. The molecular basis of dehydration tolerance in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 377–403.

Iwasaki T, Yamaguchi-Shinozaki K, Shinozaki K. 1995. Identification of a cis-regulatory region of a gene in *Arabidopsis thaliana* whose introduction by dehydration is mediated by abscisic acid and requires protein synthesis. *Molecular and General Genetics* **247**: 391–398.

Koag MC, Fenton RD, Wilkens S, Close TJ. 2003. The binding of maize DHN1 to lipid vesicles: gain of structure and lipid specificity. *Plant Physiology* **131**: 309–316.

Lin C, Mueller LA, McCarthy J, Crouzillat D, Petiard V, Tanksley SD. 2005. Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theoretical and Applied Genetics* **112**: 114–130.

Matsuyama T, Yasumura N, Funakoshi M, Yamada Y, Hoshimoto T. 1999. Maize genes specifically expressed in the outermost cells of the root cap. *Plant Cell Physiology* **40**: 469–476.

Moore R, McClelen C. 1983. Ultrastructural aspects of cellular differentiation in the root cap of *Zea mays*. *Canadian Journal of Botany* **61**: 1566–1572.

Nylander M, Svensson J, Palva ET, Welin BV. 2001. Stress-induced accumulation and tissue-specific localization of dehydrins in *Arabidopsis thaliana*. *Plant Molecular Biology* **45**: 263–279.

Puhakainen T, Hess MW, Makela P, Svensson J, Heino P, Palva ET. 2004. Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. *Plant Molecular Biology* **54**: 743–753.

Roberts J, DeSimone N, Lingle W, Dure L. 1993. Cellular concentrations and uniformity of cell-type accumulation of two LEA proteins in cotton embryos. *The Plant Cell* **5**: 769–780.

Shirsat A, Wilford N, Croy R, Boulter D. 1989. Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Molecular and General Genetics* **215**: 326–331.

Simkin AJ, Qian T, Caillet V, Michoux F, Ben Amor M, Lin C, Tanksley S, McCarthy J. 2006. Oleosin gene family of *Coffea canephora*: quantitative expression analysis of five oleosin genes in developing and germinating coffee grain. *J Plant Physiol* (in press).

Skriver K, Mundy J. 1990. Gene expression in response to abscisic acid and osmotic stress. *The Plant Cell* **2**: 503–512.

Soulages JL, Kim K, Arrese EL, Walters C, Cushman JC. 2003. Conformation of a group 2 late embryogenesis abundant protein from soybean: evidence of poly (L-proline)-type II structure. *Plant Physiology 131: 963–975.*

Wise M, Tunnacliffe A. 2004. POPP the question: what do LEA proteins do? *Trends in Plant Science* **9**: 13–17.

Zhu B, Choi DW, Fenton R, Close TJ. 2000. Expression of the barley dehydrin multigene family and the development of freezing tolerance. *Molecular and General Genetics* **264**: 145–153.

BNSDOCID: <XP_____9079141A_I_>

XP-002974277

28/04/1999 24
1-24 = 24 3



☐1: AB026642. Arabidopsis thali...[gi:4757398] Links

```
LOCUS       AB026642               77287 bp    DNA     linear   PLN 27-DEC-2000
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MED5.
ACCESSION   AB026642 BA000014
VERSION     AB026642.1  GI:4757398
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (sites)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. I.
            Sequence features of the regions of 4,504,864 bp covered by sixty
            P1 and TAC clones
  JOURNAL   DNA Res. 7 (2), 131-135 (2000)
  MEDLINE   20277480
   PUBMED   10819329
REFERENCE   2  (bases 1 to 77287)
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MED5
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MED16 and the 3' clone is F21A17.
FEATURES             Location/Qualifiers
     source          1..77287
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="MED5"
```


BNSDOCID: <XP_____2974277A__I_>

```
                /clone_lib="Mitsui P1"
     CDS        join(689..829,1056..1214)
                /note="gene_id:MED5.1
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB01785.1"
                /db_xref="GI:9293882"
                /translation="MCMAYLVVFRYFGCFWSSNRGKTEHEVRMPNDVFSKNTLSQPRS
                HARGRWTIQSRLHCISLDRKTEPLKSSLINDILDHWEAHFSPSDANDIFGRGIMT"
     CDS        complement(join(2226..2351,2434..2661,2796..3083,
                3159..3326))
                /note="gene_id:MED5.2"
                /codon_start=1
                /evidence=not_experimental
                /product="transposon protein-like"
                /protein_id="BAB01786.1"
                /db_xref="GI:9293883"
                /translation="MFTSFFVACVWRFTRDRGSRLVRKITRIFLQKFDAPFYNWSCVP
                VDKRERLFLEFAKTHHWDPLLTGTVQYYFNEIVKRRLKDMVSTARTTREQPPWIGETL
                WGTMCAYWDTEAAQKRSRTYSKARLSDRNGIGPHVHYSGPKSFQEIQDELEEKLGRPV
                HLGEVFIETHTKSDGSFVDQKSEKIAQAYQQNVRDRLSALEASASAVSDGSSRPPELT
                LDDYTAIFLESTEKDSRGNPYGLGCLKDTLGSANRNHSGSSSSFQALEERL"
     CDS        complement(join(4457..4848,4900..5115))
                /note="gb|AAD28667.1
                gene_id:MED5.3
                similar to unknown protein"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
     CDS        complement(6870..15564)
                /note="gene_id:MED5.4"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
                /product="Athila retroelement ORF1 protein-like"
     CDS        complement(join(16691..17082,17134..17349))
                /note="gb|AAD28667.1
                gene_id:MED5.5
                similar to unknown protein"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
     CDS        complement(join(18583..18807,18971..19069))
                /note="gene_id:MED5.6"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
                /product="transposon protein-like"
     CDS        complement(19942..23841)
                /note="gene_id:MED5.7"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
                /product="transposon protein-like"
     CDS        complement(join(26389..27150,27293..27596,27715..28429,
                28770..28877,29127..29173,29504..29571))
                /note="gene_id:MED5.8"
                /codon_start=1
                /evidence=not_experimental
                /product="Ac-like transposase"
                /protein_id="BAB01787.1"
                /db_xref="GI:9293884"
                /translation="MKICNQQYILALGVRRNRFIWFGVNGDQEPRRRRQTRQSEGPIN
                MEDNVDEKSTEKVMEDDVDRMNENREKRVEASSGYNDWRNAAQRLKRHETTHRHIISM
```


BNSDOCID: <XP_____2974277A_1_>

```
                THWMEIEVRLQKSKTIDNHMEEAIKKEKKYWRDVLVRIIAVVKTLAEQTKEKVDDPNS
                GNVLAFIRMIGGFDLVMKEQIRRANKGEIHCHFLSHKIQNELIELLANETKMMILKKI
                KDAKYFSVILDSSPDDSRKEQMTFLIRCVDVSTCSPKIEEFFSTFLHIKDKRGEGLFK
                TLQDALIDLKLNIDDIRGQGYDNGHNMMGKHKGVQKSEYIVYLQLQPTIGKCRETMNC
                ITLKLLSQTRWESRIDSLKPIRYQASKIRDALFYLAEYSDNPRHRSDVESLAESETHG
                IGGFEFLFGMVIWYDLLAAVNIVFPIKPKRLVKKKKPFGEDVEKIDESKTAKESFRIH
                YFMNIMDQAIMSIEVRFEQFQEYEKIFGFLFGVKRLKVAEDDELRTSCMKLEASLKHD
                VHSDVDGEDLFMELKLLKDVLPKEITKPVEVLDFLKRMDGCYPNTWITYRILLTIPVS
                IASAERTFPKLKLIKNYLRSTVPQERLNGLALISIEQELVKKLDYQKLINEFAGAKAR
                RIMFRHEEKLYFCLFFLLTIFLFNFLKQNKFHIILFCEKILGH"
CDS             complement(36504..37124)
                /note="gb|AAF24530.1
                gene_id:MED5.9
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB01788.1"
                /db_xref="GI:9293885"
                /translation="MFRIVLHACCIFALISGDLELLTSYLQEADLITTNHSTKRSSRS
                FRISRDHSTHRSSRGHHHFTSPLDHEIECPHLHHQTITRSLHSTMRSSAFTFTARPLL
                DLVTQPRGRVSPSQHLTAYSMTSFRAFFIPHSTRHSSTRKKRRLQIFTRPLTRPPGSS
                TVLNPSKYFVVLSIRVSGYLAISSMYSSFSQTRSFLPSFISADLVF"
CDS             complement(38893..45012)
                /note="gene_id:MED5.10"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
                /product="Athila retroelement ORF1 protein-like"
CDS             complement(46183..46803)
                /note="gb|AAF24530.1
                gene_id:MED5.11
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB01789.1"
                /db_xref="GI:9293886"
                /translation="MFRIVLHACCIFALISSDLELLTSYLEEADLITTNHSTKRSSRS
                FRISRDHSTHRSSRGHHHFTSPLDHEVECPHLHHQTITRSLHSTMRSSAFTFTARPLL
                DLVTQPRGRVSPSQHLTAYSMTSFRAFFIPHSTRHSSTRKKRRLQLFTRPLTRPPGSS
                TVLNPSKYFVVLSIRVSGYFAISSMYFSFSQTRSFLPSFISADLVF"
CDS             49533..51269
                /note="gene_id:MED5.12"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
                /product="reverse transcriptase-like protein"
CDS             complement(join(53826..54084,54364..54445,54944..55217,
                55368..55565))
                /note="gene_id:MED5.13
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB01790.1"
                /db_xref="GI:9293887"
                /translation="MATTFIHNDLGLTIKQSLLRTKRNKKNQTSHYNHKMPPQKAPQS
                PQPFTFQIFDEIDGDFCEEEVLSNTMQVFISANRLFRYQESLKRNPIYKLKKFLITPW
                KIIYKVSDHKHGIFFNDKTLMVLVPTANHQIDEQKIRFRGFEEFASIVYTNIDLFAVV
                SLSNTSSTRLFFDTDVDETNQFLTWKLASLALLQLWTYNHNMDGTTSNATFATINLEQ
                ETHQCIATIAKNTTTFALSGLTTMNDLNEVIFIQNDIAQIALSLLPQIRDRR"
CDS             complement(58666..60218)
                /note="gb|AAD48951.1
                gene_id:MED5.14
                similar to unknown protein"
                /pseudo
                /codon_start=1
```


BNSDOCID: <XP_..._2974277A_I_>

```
                     /evidence=not_experimental
     CDS             complement(72081..76294)
                     /note="gene_id:MED5.15"
                     /pseudo
                     /codon_start=1
                     /evidence=not_experimental
                     /product="reverse transcriptase-like protein"

ORIGIN
        1 gatccaagaa aaaatcgaaa catgccttta gtgcaaaagg gtagacttaa gatcattatt
       61 gggtttagag ataggaccat ctacttgtta ttcacggttt atgtcattgg gtagataggg
      121 tgctattctt gtatgcatag gttggcactt acctttagga ttcttctaaa gctcagtcct
      181 ttttcattga gagtcccctg ctattaaatt ctgaaccact aaaaagggaa cattttgtc
      241 tcataacctg tcattagcca aatgagttca ctagtgatgc atatcttgat tcatgttttt
      301 gaacttaatg aatgttaaag ggattggttg atcttgacgc attgtgcatt tgagtgtagg
      361 tttggatcat aacagagcat ggctaaagtt tttaagtaga agtcgatcac ctcgcatctt
      421 agaactgtta gctggtacat tgatcttaat tgtcatatct catgctttgg ttctgaatcc
      481 ccaacttcaa acctctcctt ctgcttatgt cttcttgttt gcttgagggc aagcaaagac
      541 taagtttggg gaagttgata tgtctataat ttgcatgatt taggcattca ttcttcacca
      601 ctttgtcatt gtttcatcca caaatcatct cattaggagt cgttttaggg tgttttacat
      661 acttaggacg attttgcatt aggattgtat gtgcatggca tatttggttg ttttcaggta
      721 ctttggatgc ttttggagtt caaacagagg aaaaacagaa catgaagtca gaatgcccaa
      781 tgacgtcttc agcaaaaata ctctatcgca gcctcggagc catgccaggg taaatcactc
      841 tatcacaacc ccggagaagc aagcagctca gtttcactca atcgcgaagc cggaggaatc
      901 accccatcag aactactcga tcaccactcc accaccagga agcctatctg ctctttctgc
      961 actcaaccag acaagctaag aacaaagaag agaagagaag aagaaaccac tcgatcacca
     1021 catcataaca tccactcgat cgcactggat cacagggtcg ctggaccatt caatctcgac
     1081 ttcactgcat ctcactcgat cgcaagacgg agcccttgaa gtccagtttg atcaatgaca
     1141 tactcgatca ctgggaagct catttcagcc cgtcagatgc caatgacata ttcggtcgcg
     1201 ggatcatgac gtgaagaacg tttaagttct cgcttatatc actacaagaa aacaggtgat
     1261 aaacgactta cacggcagtt gcatagtagt cgtaaaaaat gataatacga ctaatcagcg
     1321 actaacgacg ttagtcatta ttacatagtc gctgaatgag atagtcgtga attagtcgtt
     1381 gatcgacgac tatattacga caataacaca atcagtaaac aatggtcgtt gtatagtcgt
     1441 atcgtattgg cgactattta acgaccatag cacgattaat tgtcctgtac gaagcgtgtg
     1501 aaacacgtgg tcgtataata gtcgtataat tgtttccgac gattggacga ctaattctat
     1561 caatgtaagg aaataattct cgtagtcata taatagtcgt ataatgtttt acgactatta
     1621 taggtctcgt ggacgattat attgtgacta ggtgtcttag tcggataata atcggtaatt
     1681 agtgtgacta ttaaacgact attttacgtc taatttgcta cctaaatacg actactgagc
     1741 gactacttgt tttgtccgaa taagaatttg gtcatcgatt gttgggtttc ttgtttcttg
     1801 tttgttggtt atcgattgtt gttttcttgt tttcttgttt tctccttttt tatttgaaat
     1861 gctaaaaaga tgaaaagact tgtatatata gaaaatcaag tacaaataca catacaatta
     1921 tcttgtttat tacacaaaca acttattcaa ataaacatat gttcaagtag agtttgggca
     1981 aagataatct aaggatcatt aggtggagga cttggtatac gttctatggt tgtggtagag
     2041 tgacttgcca tgaagtccaa gaactgcgga tcggtttggc ggagatactt ctcgactaaa
     2101 gacaattgct cgagcttgtc cttttgctca gctgtgactc gggatgactc agcttctcgg
     2161 gcagcaatct cagcatctct cttctcatta tatgcagcct gctcttctat tttcctttga
     2221 gcttcctaca gccgttcttc tagggcttga aaggatgatg aggaaccgga gtgattgcgg
     2281 ttggcgctgc ctagagtgtc ttttagacat cctagtccat aaggattgcc tcttgaatcc
     2341 ttttctgtgg actgcaagaa atagagaaga agctcattag tgacagaata ttaattggac
     2401 tctaagactc ataattatac taaaacatct tacctcgaga aagatggctg tataatcatc
     2461 tagtgtgagc tccggaggtc gtgaagagcc atcagagaca gcagaagcag acgcctctag
     2521 tgctgacagc ctatctctca cattctgctg ataagcttga gcaatcttct ccgacttctg
     2581 atcaacaaat gagccatccg acttagtgtg tgtttcgatg aacacctcac caagatggac
     2641 cggtcttccc aacttctctt cctgcaaaca caaatcaata tcaacagtag tcactagaca
     2701 taagaactca acatactcac aagaaaacag actatattgg attgaaaaca gaaaacagaa
     2761 aactctaact tatgagaatc tgcagaaaca cttaccaatt catcttggat ttcttgaaaa
     2821 gattttggcc cagagtagtg gacgtgagga ccgataccgt tacggtcaga gagacgagct
     2881 ttggaatagg tccgactcct tttctgtgct gcttctgtgt cccagtaagc acacattgtt
     2941 ccccacagcg tttctccaat ccatggaggc tgctctcgag ttgtccttgc ggtgctaacc
     3001 atgtccttca agcgcctctt gacgatctca ttgaagtaat actggactgt ccctgttagt
     3061 aagggatccc agtggtgagt tttctacaaa aaaaaggaag aagttagtga aactacataa
     3121 acttaacttg aacaagagca gatatagagg aagcttaccg caaactctaa aaataatctt
     3181 tctcttttat caactggaac acatgaccag ttatagaagg gagcatcgaa tttttgtaag
     3241 aaaattctag tgatcttcct aaccaagcga gatcctctgt ctcgagtaaa cctccaaaca
     3301 catgcaacaa agaaagaagt aaacatgtta caaaagtgaa caaaattatt ttcccaaaag
     3361 aaatacaagt cacacacgac tataaaccat gttccacaat agtactgtct caacaatata
```


BNSDOCID: <XP_____2974277A__I_>

```
3421 gcatcattca caatcaaagc atcatcatga taataggttt ttaagatcaa ttaatcctaa
3481 agcactatca tgtcgttgta gtactgaagg ttgtcaatcc aaatgggtgt gatgctaaca
3541 atcaagatgt gatcagaagt cactaagtca agccaagcaa taacaggttt tggtgtgttc
3601 tagcagtcct aagtgaacat gcagaaaaca gaaaatcaaa cagaaacagt agaacactcg
3661 accaacacag ctctttgcag agcactgagt gaggtcagct ctttgcagag cactgagtga
3721 ggtcgagtga atggtcgagt aacagacaga aaatggaaca gggtactcga ttgcacagtc
3781 tgttgccaga caactgggtg gtcgagtacc aggtcgagta acagtaaaga attgcagaaa
3841 ctaagcaaca ggtgacagga tgcagtaaac agcagtagag taaacaataa aatcaatcag
3901 acaaagaatt cctaaggatg gggtaattga gtagtttcgt ttcctcagtt tacaggtgat
3961 tgacatgctc aataaattaa ccctagacaa caagttcatt aaccaaatct aagtgccacc
4021 gcaacagaga ctctcaagct aacctagtcc cagactcaat caccattgac gagagctaac
4081 ctaacaggca ttatgaatca acaagttaaa gccaaaacgc tccctgcaac caataccttg
4141 gtacagggtc acgaatctct ggaattagtg gttcagacat ttcatcgaac acctttggg
4201 cgcggaaatg tctgggctca aattccagtt gatcagaaag caataggcat taagaacaac
4261 taatccagaa gggatctatc aatcaatact caatcaccta gcatactgag aattctacac
4321 tactctaacc catcctcaga aacctattca ctactcagac atcatgacaa aaagcataaa
4381 cgttgaacag gttgaagaca tgataacagc agtatgataa cagaatgaac aacagtaaac
4441 gaaaatcaag aacagaaaca gaatactgaa tactgaatag ataaagagaa atccggatta
4501 caaaaagtct agaacacaag gtctgcggaa ataaaaacta ggtaaaaact tcttgttctc
4561 gaaaatgaga agtacatgct acttatagcc aaatatcccg aaaccctaat actcaaaacg
4621 acgaagtatt tggacgggtt aagcactgta ctcgaccaac gtggtcgagt gaatagctgg
4681 agtcttcttt tcttccttgt gctcgagtgt ctggttgagt gcagaatgaa gaaggctctg
4741 aagcttgtca tcgagtatgc agtcgagtgg tgtgatggag gtactcgacc tctaggtaga
4801 gtgaggaggt cgagtggttg tctgatgatg gaggtgaaga cactcgtcct catggtcgag
4861 tgaagtgatc gagtgactgt ctggtgatga gatgatggca ctcgacctcg gggtcgagtg
4921 gtggttgaat tggtgatgtc ctctactcga cctgttggtc gagtggtctc ttgagatctt
4981 gaagctccta ctcgacctct gggtcgagtg gtttgtcatg atcaggtccg ctccttccag
5041 atagctcgtc aacagctcca aatcacctga aatcaacgca aatatgcaac atgcatgcaa
5101 aactatccta aacatgcaaa gtgatgacaa atgatgtgaa atagtataaa acagtgatga
5161 tatgatgcta aaatgatgac aaatggatgc tcaaaacatg caaaatatac acttatcaac
5221 tcccccaaac ttagtctttg cttgccctca agcaaacaat caagaacaag ctggaggtga
5281 cgtttgaaag cggggacgca gagccaaagc attagataaa ccagatgaaa tcaatgtcca
5341 agttgaaagt tctaagctgc aagaccgaat tctacttaaa aacgttagcc atgcctcatt
5401 aatcaatcaa tccgactcat atgctcgacc tacataagtt ttcaaatcca ccaatccctt
5461 taacattcat tagttctgga acgtgaatca agcaatgcat catcaaagaa ctcatttggc
5521 taaggtaaaa ggtcaagaga caaagatggt cccttacaga ataggcttca gtaacagggg
5581 atctctcaaa aaatgattaa actttagtag aatgctaaag gtaagtccca agctatgcat
5641 acagagataa gatcccatct acccagagta catcaaatga agctctaatg gtaatcccat
5701 ctccaatccc aatacaagaa ctaaactcta tcctttgtat tcatgaatct tttaaaccct
5761 ttttctaatc attcttttct tttctcttaa ctctaggcga agcttcaaca ctttgatata
5821 tctagcggat ttggatttct ttaacaacta aggttcctgt ttttaacatt ttaaaaccaa
5881 gggcttttgc ccatcttcaa tagctaacaa gaactctttc ttttctttac aaatcccaaa
5941 acctttacta gacttttttct gcttctttag cttacttcac ctagagtcct ttacctttte
6001 tcatcaatga atcaaccctt ttttttcttt ttcttttaaa cacatcccaa cccaaagtat
6061 aagcacccac ctagtcctat ccagtccgaa aaatgcgaac tagaactaca tgagacacta
6121 tctctgtcgt tacgaaccta acactttcta ccaagcttaa tcggaataag acctcacaaa
6181 cactcacaag tgatataggg cttgaaagaa agtctaggtt tgggtttggg ttaactgctc
6241 taataaggag agtcagctac ttggattaac aagagtgatc aaaatgaata agagaatcca
6301 agtattttag ggtatccatg agttcaaatt tgatgcagac atgataatta aagatcagat
6361 tcaggttcaa attgataggg aatggtatca gatcatgaca gtgtggtcga gtagagcaat
6421 ctaggcatgg ttcagtttta tttcaaattc attgactacg caacagacaa ctagcaacga
6481 gggcactgag tttatccggt gaatcaaaat gcttacaagg ctatctcatc atcatccccc
6541 atgcatatgc aacaatccta aatgaaacac tctagactcg atcctaaata agatgcaact
6601 acatgaacac tctttttttt gtgaaatcat tttctgaaat ttttcaaatt ttttcggttt
6661 ttctatgcct tagatgaatg caggctcaaa attatacaca atgcaacaca cacttcctga
6721 gccctccccc aaacttaaat cacacagtcc actatgtgaa taaacttgga agagagtact
6781 caggacaaat ccaatcacag aaacaaaaaa tgaagaaaca aatcacaaca gacagtataa
6841 tacaatggtg aagtgaggcg cttacctcag tcgaagaatg aacgtagttg ctcgtcaatg
6901 gctgcctgcg attgctccca ggccatagac gaacctcctt gttctgtatc agctccagca
6961 ggtgggtact cgacctcatc agctctgtgg cggccagcac cgctctgagg aaactcgacc
7021 tctctccttc tgctgcggcc agcaacgcgg tcgagtgatc tcctcgagtg aatgaggcgt
7081 gatctgctgc tagatcgagt gagtgtagcc ttccttcttc tcttgtgttc ccgggattcg
7141 aatgacgaat gcctagcctg gacatgtggg actctctgct cgctaagctc aggcctgtgc
7201 gcaggtgcat taattctcct cgatggcatc tccataggag gttgtacctg tggaatagca
7261 gtggtggagg aagagcaact cagcttgtct gtcagaagct tgaaacactt ggcgatcagc
```


BNSDOCID: <XP_____2974277A_I_>

```
 7321 ttgtcctgct tcttgcacca cttctgcaat ttactgttat tcttatgagc ttcagtcaag
 7381 ctcttgctct gcctcgctgg aggtacatgt tcagcaaagt gatacatgct agtgtcgaac
 7441 ttcactgctc catcttcatc catcccatcc gaagcagctt cttcaagaag atcgtcctca
 7501 tctaatggtg gagcgttctc atagtagagg cgtccaatct caggcctaaa atcaatgtta
 7561 tctccctcga ttatccgtgt gacctcaggg ctagggagaa ggatgttagc tctcctgtct
 7621 gtagagtgct caaacctgaa cctgtggaaa tcgtcaacca ttgcaaactc caggaattcg
 7681 cagtgacgta ggtgctcgat atccattgct cgcggctcga gtccagcaga aatgagtggg
 7741 actccacaag ctatcagaat aggtgtcacc acgccaccta tgcacagagc gcctcgtgct
 7801 ctcttgcggt tattgctgac cgcccagttt ttgtatccac acagatggat cagaagaagt
 7861 atagagagag gtgtgtcatt gacttcaccc tggagggaca tgccattttt agtttggcga
 7921 agagttcctt tgagggccat tgcgatcatc tccatatcag agttagtgac agtccctgta
 7981 atctctcggg agtagagtac gttggctaca gaacgctgga agtacctgat gacagggctg
 8041 cgtatctgat tgctctttga ccaggaagca ttcaacggta cagagctgcc aatggtgatc
 8101 cacaagtctt tcaactcttc tctttcatac tttggcttag atcccgttcc actggggaaa
 8161 tcaaacaatc cttccaatcg cttgattgat aacctgtact catgaccata cacagaaaat
 8221 cgcaagaatc ccaatccttc acaatccaac tcatcagagg tcataccttg gtagagctct
 8281 acttgcagtg tggagaggaa ttgtattgtc tcatcttcat atgctacata cggataagcc
 8341 atcaaagtgt ccagatgaca actttggtac aggtgctgaa catcttccaa caatccgagt
 8401 tgtgcaagtg aggtcgagca aggatacctc gtgctacaga actcattcag cttgaaaagc
 8461 ttaacatatt cctcagtagg caatatgtcg ggcttgtgca gaagtttggt agctctgcga
 8521 gtctgttcag gcatgtactc gtcctccaga tcttcgtcta taagctcata cctctctgtc
 8581 atcgctctct ttcctcgggc tatctcagct ttccttctga aactttcgta ttctctctgt
 8641 tctctctctg gccttgaaga tgacgattct gtctcatcca tgttgtagtc aggatcaaca
 8701 gaagaactgc cactgtaatt gctcatgatt ctcctgtagg gatttagaac tcaaagctca
 8761 aatgtcaata ggttagtctc aaagatagca atgaaaccaa aattattgtc aataggcaat
 8821 gaaacagatt gaataggtcg agtacaattc cgtcctcgag taacctgagc tgttcctgag
 8881 cttagatgaa atgaatcgag tatcttctca gaggtcgagt tgggtgcttc agcaggtcga
 8941 gtgaaacgta gagatagagt gaatcgtgaa agctagtggt cgagtagctt ctctgaagtt
 9001 ggaaggctct actgaattca cgaacaagaa gccgagtaga ttattgcagg ggtcgagtgc
 9061 tttcaggagg tcgagtgatt tcttgaagtc ggctaggacg gcgtcggtcg agtgcctcta
 9121 cgagaccagg aagcgaggtt gagcggtttc ttttgaaaga gacggcggtt gatgccgatg
 9181 aattggtcga gtatacggaa ggaatcgagt gaagtggttt agtcggttga cggagtggat
 9241 gggctggtga cggcggcgag ggaggggtc gagtgaagtt ctcagtggtc gagtagctct
 9301 ttggacggcg atatgtggtg tagagagatt agagtgtcgg cggtgaaatg aaaagtggaa
 9361 ggtcgagtgg atttggagtg gtcgagtagc gcttcaaaat gagtggtcta ggttgagttt
 9421 tagattctct ctccagctta tcctcagcgc gtggtccaca cactgttaat tgagaaaatt
 9481 tctttgcaga aaacctactg taaactgtag gaaatttctg caagtctctg tagcttctgt
 9541 tggtcgagta aaatgatgat ttaaaaattt ttgggtttta aacatcacat atactcgacc
 9601 actgtactcg agtgcagggt ctcaggtgca acagaagatt ttcaaaattt tgtactccaa
 9661 cgggctgtac actcgatctc caggtcttct gtagagtaaa agttttaaat ttttactcta
 9721 tctccgggtt gttctcttgt ctctgaaaac cacaaacaca actaaaacaa atcaaaaata
 9781 tcaagaaggt aaaaattatt tacaaggata gtcatgggac ttcctcccaa gtgagcttgt
 9841 ttaaagtcat tagcttgact cctttactcc tttaatcatc aagaagctcc tagagatgaa
 9901 ccgatgtcac ctctggaagg atttggtctg ctaagtattt cttgagcctt tgaccattta
 9961 ctgtaaaatc tccactctta ccagctagag tgactgctcc ataaggacgg acctcagtga
10021 tgcagaaggg gccagaccat ctggatttaa gttttcctgg aaagagtttc aagcgagagt
10081 tgaatagcag cacctgatca ccaacctgga aatccttagt gatgatcttc ttgtcatgga
10141 aaagcttggt tctctccttg tagattttag aactttcata agcttctaga cggatctcat
10201 cgaggtcact gagttggatc aatcgcttct cctcagcagt ttttatgtca aagttcagaa
10261 gttttaccgc ccacattgct ttgtactcga gctcaacggg tagatgacat gattttccat
10321 agagaagatt gaaaggagtt gtacctatgg gggtcttgaa ggctgtcctg taagcccata
10381 atgcatcatc tagctttgca gaccagtctt tccttgtaat cccaacagtc ttttccagaa
10441 ttgtttttat ctccctattg gagatctcaa cctgcccgct tgtttgtaga tgataaggag
10501 tggcgacctt gtgctttacc ccatgcttct tcaagaggtt ctcaaaaacc ttgttgatga
10561 aatgctttcc gccatcactg attactaccc tgggaactcc aaatcttggg aagattatgg
10621 ttttgaacaa cttcagcaca acttttgcat cgttggtagg actagcaata gcttcgaccc
10681 atttcgatac gtagtctacg gcgaccagta tatatttatt accgtatgaa gatgggaatg
10741 gacccataaa atcaatcccc catacatcaa agatctcaac ttccaagatt gggttctaag
10801 gcatctcatt tcttctactg atgttgcctt ttctctggca agaatcacac tttgaaacaa
10861 actcctgagc atccttaaac attgttggcc accaaaagcc tgcttgaaga atcttggaca
10921 ctgtcttgaa cgtcgcgaag tgaccaccat atgctgagcc atggcaatgc agcaggatac
10981 cttccacttc atcttctgag acacatctcc tgtagatctt atctttacag agagtgtaaa
11041 gataaggctc atcccaatag taatggtgta tgtctctgaa aaacttcttc ctttcataac
11101 ttgtcaggtt gggaggctct actccgcatg ctaagtaatt gacgtgatca gcataccatg
11161 gagattctcc ttcaacagca ttcagttggt ggaactcttt tccactgtag ctttttccga
```


BNSDOCID: <XP_____2974277A__I_>

```
11221 agtggttttt caattctcat cctcgacaga tggtccgctg cgccattttc aatgcctttc
11281 ttatctacta tctccatgtc aaactcttgc aataaaagta tccatcttaa cagtcttggt
11341 ttagtatcct tcttggcata cagatgcctc aaagctgcat ggtctgtata gacagttacc
11401 tttgatccaa ccaaatagct tctgaacttc tcaaatgcga atacaacagc tagaagctcc
11461 ttctcagttg ttgcatatct tccctgagcg tcatccaacg ttcggctggc gtaatatatg
11521 acatgaagct tcttgtctat tttctggcct aaaacagctc ctactgcgta atctgatgca
11581 tcacacatga tttcaaatgg atagtcccaa ttaggtgctc gaacaaccgg agcagatacc
11641 aaagcatcct tgatggtttg aaaggatttg aggcagtcat catcgaattt aaattcggtt
11701 tccttgcaca atagtctggt taacggtctg gctatcttgg agaagtcttt aatgaatctt
11761 ctatagaatc cagcatgacc aaggaagctt ctaatgtcct tcaccgtttt aggtggctac
11821 agctgcatca tcacttcaac ttttcctttg tcaacctcta tacccttctc tgatatcttg
11881 tgacccaata ctatgccttc cttcaccatg aaatgacact tttcccaatt gagaacaaga
11941 ttcgtctctt cgcacctagt caatacccctg ccaagattca acaaacatga ggagaaagag
12001 gggccataga ccgaaaaatc gtccatgaaa acctccacca tctcctcgat taagtctgaa
12061 aatatagagg tcatacacct ccgaaatgtt gcaggagcgt tgcataaacc aaatggcatt
12121 ctcttatagg caaaagttcc ataaggacac ctgaaagtgg ttttttcttg atcattaggg
12181 tgaattggta tttgaaagaa accactgtat ccatcaagaa agcaatagta tggatattta
12241 gccaaacgtt caagcatttg gtcaatgaat ggtaaaggaa aatgatcttt cctagatgca
12301 gcattcaact tcctataatc tatgcacatt ctatgaccag ttatagttct agtagggatc
12361 agttcatctt tttcattttt gacaacagtc attccaccct ttttagggac gcaatgcact
12421 ggagaaaccc aagtactatc agagataggg tagatgacac cagcatcaag cagtttcaaa
12481 atttcttttt tcactacttc tttcaaattg ggatttaacc tcctctgtgg ttcaatgctg
12541 gaataggatt cgttttcaag atggatccta tggttgcata aactaggtga aattcccttа
12601 atgtcggata acgaataacc aattgctctc ctatactttc taagctcaga taatagcagg
12661 ttcacttcat cactatttaa ctcagcatta atgatcacag ggtaagtaga gtttggacca
12721 aggaatgtgt accttagacc ttttggaagt ggtttgagat caactttcgg tgccttaagt
12781 tcagaccagt catctgaagt tggaatgatc ggctcccaag gttcgtcggt agacaaggtc
12841 gagtggttga ttgagtacgt cctcgagagt gcaggttgaa ctcgagttga cccttcttcg
12901 ctcattacca ttacctctgt tgctggtcca ttcaattcct caaagggttc tgattcttcc
12961 atcgctttat gggagtctaa cagcttctgg tatcccaatg tttccaaatg cagaaaccca
13021 tcttcaccac ttttggttaa agcactgtta agatgatctt cttctgccag ctcttccagt
13081 aactcatcag ctaactgatc catttcttcg atccaaaaga gttgcccttc tatggtaggc
13141 ttcttcatcg cgtctttgac atcgaaggtc atcctaaagt ctttgccaag gttcagatca
13201 atcttccctt tcttgacatc aatcatagct cctgcagtgg ctaagaaagg tctccctaga
13261 atcaacgggt ccttgggctc ttcatccatc tccatgacta caaagtcagt tggtatctcc
13321 acagcaccga tcctgattgg gaggttttcg agcaaaccat gaggaattct tactgatctg
13381 tcagctagga tgagggatat attgcaggac ttgtattgag tgaaccccaa tcttttggcg
13441 acagatagcg gcatgagact gactgatgct cctaaatcgc ataggcatct gttgaaagcc
13501 aatggaccta aagagcacgg tagagtgaac gaaccaggat cactaagctt cttaggaatg
13561 atcttcttct gtatgatagc gctgcattca tgactcaata ccaccatccc ttgcacttct
13621 tgaattctct ccacaatcaa gtctttcaga aacttgtgag agtctgggat tagcgctaga
13681 gcgtcaacaa gaggtatccg taactcaacc tccttaatgt tcttggcaaa catagctcta
13741 tatttgtctg ccaacgcttt cttgtgacgg ccaggaaatg gaagctgcag tttgtaggga
13801 ggagggacaa agaccttttc cttgtttttg acagcaactg gttttggagc agttggtgat
13861 gctgctggga tgacaggccg agtggttggt cgagtaatgc tgtcgagtgg cggatcgagt
13921 gattgctcga gtgacaggtc gagtgcttgc tcgagtgaca ggtcgagtgg tttgtcagct
13981 tgatctttct cgagacttaa atcctcccca tcttgatctt cactgtcctc agtgaccgtc
14041 tttggttcct ccttagttgg aagtgctttt ccactacgca gggtgatagc atgagcatat
14101 tcttttggat tctgaactgc tttgcctggt agttggcttg tctgttttgt agctgaagaa
14161 gatgtagaat gtccctctaa gtaacggact ttggtattca gtgtttccat cttgacatgg
14221 agatcattgt agctacagtc cagcttgtta tgtaactcag atatcttctt tgccatctcc
14281 atggagccag aggcttgccc gtgaagcagc tgttgtagca tttgtttcat atcagcgtca
14341 ggagcagcag aaccttggtg ttgctgaggt gcaaacccag gtggttgttg ctggtagttc
14401 ccctggaatt gctgcttagg aatgaaacct tggttgtagg gaacaaaagg tttgttctga
14461 ctttgttgtt gctgtggagg atacacctga tcgtgagggt tagcaacgtt ggtgctgcgg
14521 taggagaggt tgggattgtt ggttttgaag ttgttgtatc ctttgtagcc accctggttg
14581 ttgttgatgt agcttacttc ttccaactgg ttaccctccc catcttggac ttcatactgc
14641 tcatcatcaa caaggaagtg cacatgctta tgctggctaa gaagaatcct gtccagcttg
14701 tcattcagcg ctttgatctc cttcctgtgt ttgtcatcag agtcagctgt tcctctgacg
14761 gtcctatcac agtcttcgtt gtaatttcca tctgattgag caaggttctc aaccaattcc
14821 cagccttctt caacatcctt gttctggaaa ttcccattgc tggcggtatc cagaagcatt
14881 ctgatgcgtg gtaggactcc tctgtaaaga gtgctgagca gagaggcttt agtaaagcca
14941 tgatgagggc attggttggt gtaacccttg aaacgctccc atgcttcaca gaggctttca
15001 ccagtcttct gtgagaaacc agaaatctca tttctgaatc ttgcagttct ggcattggag
15061 aagaactttg acagaaaagc cttcttacaa tcatcccagg tggtgattga gtcatggggc
```


BNSDOCID: <XP_____2974277A_!_>

```
15121 atggggcaga ttcttctccc agatgtgtgc tttgtctccc aaggagaatg gaaacaaacg
15181 gagcttgaat ccgtcttcac tgacaccatt gatctttgtt aggttgcaga gcctatcgaa
15241 ttcatcaagg tggtcgagtg gatcctccat tggcaaacca tggaatttgt tcccctgaat
15301 catcgagatg agaccactct taatctcgaa gttgttattc tggatagcag gaggtgcgat
15361 tcccttcctc tgacggtgat cacgtggtgc atctccagca ccaatgttgg catggccatt
15421 ttgctgattc tgttcttcag ccataacgag tggttgctga ggtacagggt tgactgtggt
15481 tctccttttt ctgatttcgc gggctgtgcg gttgatgttg tcgttgaaaa ggaggttctg
15541 gttgcctgtt gatcgagttc gcatacagca gaggtgtacc tggatcaaga agacaacaag
15601 aaaaacacac agtaagtaac tgaaaagaaa attgaaaaag aacttgatct aagcaagtct
15661 gaaatctcaa atgtaacaaa caaacaccca attggcaacg gcgccaaatt gataataggt
15721 ttttaagatc aattaatcgt aaagcactat catgtcgttg tagtactgaa ggttgtcaat
15781 ccaaatgggt gtgatgctaa caatcaagat gtgatcagaa gtcactaagt caagccaagc
15841 aataacaggt tttggtgtgt tctagcagtc ctaagtgaac atgcagaaaa cagaaaatca
15901 aacagaaaca gtagaacact cgaccaacac agctctttgc agagcactga gtgaggtcga
15961 gtgaatggtc gagtaacaga cagaaaatgg aacagggtac tcgattgcac agtctgttgc
16021 cagacaactg ggtggtcgag taccaggtcg agtaacagta aagaattgta gaaactaagc
16081 aacatgtgac aggatgcagt aaacagcagt agagtaaata ataaaatcaa tcagacaaag
16141 aattcctaag gatggggtaa ttgagtagtt tcgtttcctc agtttacagg tgattgacat
16201 gctcaataaa ttaaccctag acaacaagtt cattaaccaa atctaattgc caccgcaaca
16261 gagactctca agctaaccta gtcccagact caatcaccat tgacgagagc taacctaaca
16321 ggcattatga atcaacaagt taaagcaaaa acgctccctg caaccaatac cttggtacag
16381 ggtcacgaat ctctggaatt agtggttcag acatttcatc gaacaccttt tgggcgcgga
16441 aatgtctggg ctcaaattcc agttgatcag aaagcaatag gcattaagaa caactaatcc
16501 agaagggatc tatcaatcaa tactcaatca cctagcatac tgagaattct acactactct
16561 aacccatcct cagaaaccta ttcactactc agacatcatg acagaaagca taaacgttga
16621 acaggttgaa gacatgataa cagcagtatg ataacagaat gaacaacagt aaacgaaaat
16681 caagaacaga aactgaatac tgaatactga atagataaag agaaatccgg attacaaaaa
16741 gtctagaaca caaggtctgc ggaaataaaa actagggaaa aacttcttgt tctcgaaaat
16801 gagaagtaca tgctacttat agccaaatat cccgaaaccc taatactcaa aacgacgaag
16861 tatttggacg ggttaagcac tgtactcgac caacgtggtc gagtgaatag ctggagtctt
16921 cttttcttcc ttgtgctcga gtgtctggtt gagtgcagaa tgaagaaggc tctgaagctt
16981 gtcatcgagt atgcagtcga gtggtgtgat ggaggtactc gacctctagg tagagtgagg
17041 aggtcgagtg gttgtctgat gatggaggtg aagacactcg tcctcatggt cgagtgaagt
17101 gatcgagtga ctgtctggtg atgagatgat ggcactcgac ctcggggtcg agtggtggtt
17161 gaattggtga tgtcctctac tcgacctgtt ggtcgagtgg tctcttgaga tcttgaagct
17221 cctactcgac ctctgggtcg agtggtttgt catgatcagg tccgctcctt ccagatagct
17281 cgtcaacagc tccaaatcac ctaaaatcaa cgcaaatatg caacatgcat gcaaaactat
17341 cctaaacatg caaagtgatg acaaatgatg tgaaatggta taaaacagtg atgatatgat
17401 gctaaaatga tgacaaatgg atgctcaaaa catgcaaaat acacacttat cacatcattt
17461 gtgtaatata gaacaacata acacgagttt agaactgaac taaactaatg accaagaact
17521 aatgtaatca caacacaaca ctcaacgatg tttgaccgat taacaacaga gcatcaaaca
17581 acaataaaac aggcattact acataaacta atgagtctag aactaaactt gactaaatgc
17641 ctaacccgaa aacactagca acacaagagc caacaatgtt tcacaacaca acagccaaca
17701 aggttcacat tttcggtttg attccaaagc caacaaagaa agaagtaaac atgttacaaa
17761 agtgaacaaa gttatgttcc caaaagaaat acaagtcaca cacgactata aactatgttc
17821 cacaatagta ctgtctcaac aatatagcat cattcacaat caaagcatca tcatttgtgt
17881 aatatagaac aacataacac gagtttagaa ctgaactaaa ctaatgacca agaactaatg
17941 taatcacaac acaacactca tctatgtttg accgattaac aacagagcat caaacaacaa
18001 taaaacaggc attactacat aaactaatga gtctagaact aaacttgact aaatgcctaa
18061 cccgaaaaca ctagcaacac aagagccaac aatgtttcac aacacaacag ccaacaaggt
18121 tcacattttt ggtttgattc caaagccaac aaagaaagaa gtaaacatgt tacaaaagtg
18181 aacaaagtta tgttcccaaa agaaatacaa gtcacacacg actataaact atgttccaca
18241 atagtactgt ctcaacaata tagcatcatt cacaatcaaa gcatcatcat ttgtgtaata
18301 tagaacaaca taacacgagt ttagaactga actaaactaa tgaccaagaa ctaatgtaat
18361 cacaacacaa cactcatcta tgtttgaccg attaacaaca gagcatcaaa caacaataaa
18421 acaggcatta ctacataaac taatgagtct agaactaaac ttgactaaat gcctaacccg
18481 aaaacactag caacacaaga gccaacaatg tttcacaaca caacagccaa caaggttcac
18541 attttcggtt tgattcacaa cagagcatac agaacaagaa aacagggaat ttaaagtgaa
18601 cttaccattc ggtttttggc cggggaatgg gagacaacac acaacaccac gcttctctgc
18661 caggcacaga gagaatgtca ttcagagcaa ccacggagtc ttcttggagc tacggcaaca
18721 ctgcaggaat gccctcttgg aagttgttgc cttgagacga gggatgactg tgagatgcag
18781 tttctggatc cggcgaaaga ggagggtctt gctgagcagt ctcttgtggt ggcaacggat
18841 ctacacgctg tggttgattg gtggagtgtt ggaagagttg ctgcggtggt gggtagttac
18901 ggtagtgggg ggcagaggat gaggcgccgg cgccggctcc ttgaggcgga atcgatgcag
18961 aaacttgaac cgtcaccgcc gccggggtga aatcgtactg ctgcggtagc gaagagggtc
```


BNSDOCID: <XP_____2974277A__I_>

```
19021 ttccagcagg cgttgaacct ccggctcttt gagtcacgtt tggtgtggtc ttcctctttc
19081 aagatcctcc acgaccggcg ttaggcatcc tcttggagat gagagaatcg agatggatcg
19141 aagaggcaag ggagagtttg gcgaagaaga gagaatcgaa gaagagagaa tcgagatgga
19201 tcgatgtatt agggtttgtt ggtgaaagac ggatcgattt agggttaaga agaagaagaa
19261 gaagaagaag aagaagaaga agaagaagaa gaagaagaag aagaagaaga gaatcgattt
19321 agggttacga agaagaagaa gaagaagaag aagatcgaag aagatcgagg aagaagaaga
19381 agagagaaga agagaagaat cgaagaagtg gctaagtgtt tggaacgaag gggaaatgaa
19441 tcgaattagg gttttcaaca ctggagtagt cgtatatcag tcgcataata gtcactaaat
19501 caattgggtt tgggcttcga cgattaaaag cccaataact tacaccgttg tcgccccata
19561 gtcgccgaat cgtcaccatg taaattgggt tatgggctta gtaatttcga agcccaacaa
19621 catacacaag tgtcgctgtg tagtcgtata attgtcacca cttaagttgg gtttgggctt
19681 tgtgatttcg aagcccaaca gcttacaaga tagtcactgt gtagtcgtat attaatcacc
19741 acctgagttg ggtttgggct ttgttattcg aagcccaaca tcttacatgg tagtcgcata
19801 atagtcgtaa tttcgtcacc ttatgcgttt taaaattatt taataatttt ttaataatta
19861 ttcaaatact tattcaaata catcgtatat tacaaaattt attcaaataa ttaattatac
19921 atctatttca cacaattatt attcatcttc gtcttctatt tcatcttcat ccgatgacga
19981 taaattacat cggaattcat cttctggttc agcatctcca atatcgtaat cgaggatgat
20041 tagggttacg tctgctaccg gatagtctat tacaagatct tcaattgtag tctgcaaaac
20101 agcttcatca tcctcggttt gcaacagtgt cggatcagta ttttcatatt gtccagaaat
20161 gtttcccctt ggattgacct ttaaaacact gagccaaatg ttctttggtt tctttgtata
20221 tggatatgga atataacaaa cttgatctgc ttgagatgct gttcaaacat gttatacata
20281 tatcaacaaa aatatataca acattattta ttaatgaatt agaaatttta agtatactta
20341 attacgtacc aagaataaat ggctcgtatt tattgtattt cctcgtagat agaacatcta
20401 caacgccacc atggtttcgg cgagttcctc tcccgatctt tggatcatac cacttacact
20461 taaaaagtgt gatcctcaaa ttgacgatgc cctcatactc gagttctatg acatcagtta
20521 aggtaccgta gaaatctgct tcatcagatg aatcagcgta attttcacct ttaacacata
20581 ctccatagtt acatgtcttt cgtccagcac cgtgggtttg cgtatgaaac aaatatcctc
20641 ttgtaaaata cattggccaa gttttaacct tgttcaaagg tccttgcaca atctcttgaa
20701 cccaagtagg aaatggagaa ttagtgttcc aataagttac ctacataata tagataaaca
20761 atatcaatat gatcttaaat taaatcacgc tagctacaag aaaacagaca aatagttaca
20821 ttcatgaacc aaaattactg tagacttaca tattctttaa cccatagatg atattcttca
20881 gctcttctga cgtagagttc tttttcggac aaggatggat atttcgcaga aagataatcc
20941 tcgaacatac tgtctcaaca aacgaagaca aataaatcaa agtgatctta tctatgaata
21001 aatcaaagtg attaaaaata cctctcaaat ggtttaaaat agtcacaatt tcttaaaaca
21061 tatgcatgtg cattttggta atctttctct gaaagccagt ctacatgcat tgcaccactt
21121 ggacggccga cctgcataaa tatgttaggt actccaggga catgatatac aggaacttca
21181 ccttcattga atcttgttaa cctgccacat ataattgaat taagatcaaa gtagtatagc
21241 aacatgttaa ataaataaag gtatttacct tgatttcgtt tgtatatgat cggcaaagta
21301 gtgctctgaa aaataagcaa tctcgtcatt gatatatgac tcaacaattg agcctgcagc
21361 atatcttttg ttctttgctt ttcctttcaa ctttttgaaa aacctctcaa atggatacat
21421 ccacctatat tggacagggc cacccaattc agcttcgtat gggagatgga taggtaaatg
21481 ttccataaca tcgaaaaatg atggtggaaa gatcttttcc aagttacaga tgattaaaac
21541 tatgttctgt ttgagaattt gaacgcgact tgtctgtaaa gttctcgaac ataggtcccg
21601 aaaaaatgct ccaattccta aaatatatgt ggaagttaaa taattacata aaaaaatttc
21661 tccataaaac atagttagaa ttgtacctga taatgcaagg tggacattcc ggtcaagaag
21721 ttcagcaaag ataaatggaa gtagtcgctc cataaaaaca tgacaatcat gactcttcat
21781 gcctgaaaac ataccattct ccaaatctac acaactagcc aaatctgaag aataaccatc
21841 tggaaatcta acatcatgtt tcacacattg aaataaactc tgtttggctt cttctgtcaa
21901 tgtatacggt ggaaaaggag ctttacccgt actatcaata tgtaagcctg gccgagaaca
21961 atatttctca atatccaatc ttgacatgat attgtctttt gatttaccct taacacccaa
22021 gagagtgttc atgatgttgt caagaaaatt cttctctgta tgcataacat caatattatg
22081 tcgaagattg aggtccttcc agtaagacaa ctcccacaat atgctttctt tgtgccaatt
22141 atgttccttc ccatagcctc gcatcttctt ttcatgaccg ttcccaccaa catccttcgt
22201 tttgggtgga tttacactag ccaaccgctc gtaataaact tgctcccctg tcaaagattc
22261 gggtggatac tcgctagaag cgtctcttcc tttgagaaag tctttttttgt tcctacgaga
22321 tggatgacca tgaggaagaa atctacggtg acaatcaaac caacatgtct tccgtccatt
22381 gggtagataa aaagactttg tactttccat gcaaactgga caagacaatt taccatgagt
22441 agtccatcct gataacatgc tatacgcggg aaagtcgcta atcgtccaga gtagtactgc
22501 ttttagatta aaattctgac tcaatgaaac atcgtacgca tcgactccag tacaccacaa
22561 ctcttttagc tcctcaataa gaggttggag gaagacatca agactagctc gaggatgatt
22621 tggtccagaa ttcagaatgg taagaaataa gtactctgta ttcatgcaca tacgaggggg
22681 taaattatat ggagttagga tcacaggcca caatgaatga ttacgagaca tgctaaatgg
22741 attgaaccca tcagtacaca acccgagata aacgttacgg ggttcttcag caaaccgggg
22801 atgtaactct tggaaatatc tccactctgc cgcatctgaa ggatgattca tttctccctc
22861 ctttgattgg tgctcagcat gccatcgcat agccgcagct gtcttatggc tctgatacat
```


BNSDOCID: <XP_____2974277A_I_>

```
22921 tctcttcaac tggtctgcaa taggtagata ccacatacga ctatatggta ctttggttct
22981 tcgacggtcg tccttaggct tccatctttg tgcgccacaa aatcgacatt gatcttcttt
23041 ttcgtcttct ttccaaaata acatacaatt gtttttgcaa acatcaatcg tatgatatgg
23101 aagtcctaaa ttgcgcatca acttctcggt ctggtaatgt gaagtggtag cctggtttcc
23161 ttctggtaga aaatctgtaa acatctcgca aacggagtcc accaactttt cactcatatt
23221 atactccgct ttgttgtgca tgagtcgaga tgctaatgat aactgcgact gcccttcacg
23281 acaaccatcg tacaatggat tatttgcacc ttctaacaag tcgtagaact tatttgaatg
23341 gttacggacc ggctcttcca cattctgata actaccatca tgatgatagc aatcatcatg
23401 acccacgtta tcatcatacc ccacgttaaa attaaatgca tcgttcacca tatccacata
23461 tggatcttct acaacattac ccacttcttc atgattctca ctaaaatacg tcctatccgt
23521 gtaactagtt ccgatatcca tattcaattc ttctccatgc ttataccaaa cataataatc
23581 aggcataaat ccttgactaa acaaatgact actaactctt ctacccgaga tgatgtgttt
23641 ttcattcttg catacactac aaggacaatg aaatttacct cgacaactct gtactatagg
23701 ctgactatta gcgaatgtca tgaattcctc aactcctgcg acgtattcgg ctgacaaatt
23761 ccccgtcact tcatcgaatc tcttgtacat ccaatctcga tagaaaccgc cactaccgcc
23821 ataattgtag tttcccgcca ttatagaata aactaaattt ttattgtttt ctttttttt
23881 tcttttgatt gttgttttga tttgtttgtg aaagaagagt gtcgtgaatt ttgattatat
23941 agatgaattt ttatgactaa tttatgacta caacaaacaa attaggtgta cctacttggc
24001 aatgcggaaa acgttttcca acaacttaat gcgactattc tgcgactcag ttgcgacaat
24061 tgtacgattt cagacatgtg tcgtatttta gtcgccgatt gggcgactaa cttccgcagt
24121 cggaaactgt agtcgtccat ttgcgactat gtaacgacaa acgtttttct tatataatat
24181 aatcgttaaa tagtcgttgc tatacgagta tttgccgact aacatattta gtcgtaaaat
24241 aacgactaat tgactactaa tgttcatgtt gtcgtatatg cgtcgtaaaa cggtcgtgac
24301 gtagtcgcca aaaacgacga ctacatctgt tgtcgctttc tgtagtcgtt gtttgtctgt
24361 tttcttgtag tgtataaact gcctcttagg gttttcccga gggcatccgc tattatccgt
24421 tattatctcg cttttacatt ttcgcaactt ttatagccgc agtgtttctc ctgcaacttt
24481 gtattttcac tttcaaatat ttgatattca gttttcgttt aagtctttcg tcatttctgt
24541 tatcatgttt ttaatcgttg ctttgctatt gtctttgatt atgtctgagt agtgaactgt
24601 atttctgagg atgggataga atagttgtgg tttacttggt cgttttaggt ggttgagttt
24661 agaatactag aatcccttct ggattagttg cgcttattgc ttgtttcact ggctaatgtg
24721 gagtatgagt ccagacgtcc agcgctcagg aggcgtttaa tgggcgttgg atcctattaa
24781 ttccagagat atctacctct atcccatggg attggctgta gagagggtat tgattttaac
24841 atgttgattc gataatgcct gcttagttag agatcgctaa agggatttag gtcagtgtct
24901 agctttactt gaggatttct atcacgggag tgaattaatc tgttgttgaa cttgttgtct
24961 agggatagct tctgtgaagt tttgtcaacc gcctgaatag attaagtaga ccacactatt
25021 cgattcccca tccttaggaa tcgtttatcc atttgattgc ctgttatcat ttctgctttg
25081 tcatttactt gcctgctatt gccatttctg cttttagttt catttcattt ctgttattgc
25141 atactaggac tgttagaacc aaccaacttt ctgtttggct tgacttagag ctttttgatc
25201 atatctcatc tgtttgcatc acaccggatt ggattgacat cctaaaatag tacaacgata
25261 ggatagtgtt ttagggatta agtgagtaaa atcctactat caatctgtaa ttgaagtggt
25321 ttttttgtag aaaagaaaaa aaaaaggaaa aattgccaac acatgtcaca caaaaaaaaa
25381 aattccaaat gtcaagattt gagagagttt agaagaaacc ttctcatatt atggaagacg
25441 tctcaagtta attgtattgg acttcctcta tgtttatatt ctttgtattt tctcctctta
25501 ttagtcttag tagattgata attaaaattt catctcatat gaatttggtt tgttctaaat
25561 taacctcctt gcaagtgatg gatttgggtc acaagaaaat tctaccataa tatctaacac
25621 tccaaacact atgtgtgctc aagacatgct attgcataag ctttctatgc caatcagaaa
25681 atatcatgtc gtatctacaa atattgtaaa atccaaaaat aaaaaatggt aaactaaact
25741 atacacaaat tcggcatgaa gagtatccct tgatttgcaa ttttggtggt gcatatgcaa
25801 ccatatctca ttttcacaac ttttgtgtta aacctcagat ccgaaatgtt ggtcctaaat
25861 ctatagccca ttaaatttgt agaaccgttg ttatttaat acaaaagtca caactataat
25921 tttgccgttg tgtcgaacaa tacaaatttt tgaccaaaac aaaatttcga ttgcgcaaac
25981 gcaaaatatt tgttttctac attaaatatt ctccttcagg ctttgaaata taacaaaaaa
26041 aaaatcacat aacgccatct ccatgtgcag ttaatttatg caaatcttag agcgccacaa
26101 gtactctagt ataggtgaac tacttggcga tattgcttag ctttaaattg tgatccctta
26161 ggttggctca ataacccgat cgtatcaaag atcaattgat actgtcactt ggtactcaag
26221 gttgttctgt cattacgaac atgtcttatt tgccaaaaac atttccatca tgtcacactc
26281 gcaactagta tattttaggt tgcaaatgga aataagaaaa cgtccaaaga gagagaaaga
26341 ttaccagtgc cggtccgaga atatttgatg cttaaagcaa attcaaaatt aatgccccaa
26401 aattttctcg caaaataaaa taatatgaaa tttatttttgt ttaagaaaat tgaacagaaa
26461 aattgttaac aagaaaaaaa gacagaaata taatttttct tcatggcgaa acataatcct
26521 ccttgctttc gctcctgcaa attcattaat taatttttgg taatcgagct tcttaaccag
26581 ttcttgttct atagatatca gcgccaaacc attcaatcgt tcttgaggca cagtcgaacg
26641 tagataattt tttatgagct ttaattttgg aaaagttctc tccgcagaag caattgaaac
26701 cggaatggtg agtagtatac gataagtgat ccatgtgttt ggataacaac catccattct
26761 tttcaaaaaa tccaacactt cgaccggctt tgtaatctct tttggcaaca catcttttaa
```


BNSDOCID: <XP_____2974277A_I_>

```
26821 aagtttcagc tccataaata gatcttctcc atcaacatct gaatgaacat catgcttcaa
26881 agaagcttca agcttcatac acgatgtcct taactcatca tcttctgcta ccttcaatct
26941 ctttacacca aacaaaaatc caaatatttt ttcatattct tggaattgct cgaatcttac
27001 ctcaatagac attatagctt gatccattat attcatgaag taatgaattc taaaactctc
27061 ctttgcagtc ttactttcat caatcttctc aacatcctca ccaaatggct tcttcttttt
27121 caccaagcgc ttcggtttta taggaaagac ctgctcaatc tccatctcaa tggcaatatg
27181 tgtagcttta gctttggctt tctcagtagt ttttcaaata agtaacaagc cctcctaact
27241 gaaaaatagc aacctcaaga tccatatcct caaactgtaa ggatttgctt actatattca
27301 cagcagctag caaatcgtac caaattacca ttccaaataa aaattcaaac cctccaattc
27361 catgtgtttc actctctgca agagactcaa catcactcct gtgtcttgga ttatcactgt
27421 attctgccaa ataaaaaagt gcatctcgta tctttgaagc ttgatatctt attggtttaa
27481 gactatcaat gcggctttcc caacgcgttt gagacaacag tttaagtgta atacaattca
27541 ttgtttctct acacttccct attgttggtt gaagctgcaa atagacaata tattcgctga
27601 acaattctaa aaaaagaaac agctttcttg gattctaaaa aaagaaacag tgtaaggtta
27661 agactctgac atccatatgg agtgtaaaat gctcttggat ttgtatcaag caacctttc
27721 tgaactcctt tatgttttcc catcatatta tgtccgttgt cgtatccttg tcctcttata
27781 tcatcaatat tcaacttaag atcaatcaat gcatcttgga gagtcttaaa aagtccctca
27841 cccctcttat ctttgatgtg tagaaacgtc gagaagaact cctcaatctt aggtgaacaa
27901 gttgacacat ccacacatcg aattaagaag gtcatttgtt ctttacgact atcatccgga
27961 gaggaatcaa gaattactga aaaatacttc gcatctttaa tcttctttaa gatcatcatc
28021 ttggtctcat tagcaagtaa ttctatcaat tcattctgaa tcttatgact gagaaaatgg
28081 caatgaattt caccttttatt agcccgtcga atttgttcct tcattaccaa atcaaaacct
28141 ccaatcattc taataaatgc caaaacattt ccactattag gatcatcaac cttctcctta
28201 gtttgttcag ctagagtctt taccactgca attattctca ccaatacatc tctccaatat
28261 tttttctcct tcttaattgc ctcttccata tggttatcaa ttgttttact cttttgcaat
28321 ctcacttcta tctccatcca atgagtcatg gaaataatat gacgatgagt tgtttcatgt
28381 cttttcagcc tttgcgcagc atttctccag tcattatacc cactacttgc taacaaagta
28441 ggatttgttt cttaatcgaa taatttacag caaaagcaaa agaccttatc tttaactttc
28501 gagtaaatta accatctcta acttgtttct ctccattact cagtcttctt atttaagatg
28561 aataagtgaa acatctatca atatcatctc ttggaaaacg gtaatcagct cgtggtcttg
28621 gaggtgatcc tttctcaacc aaataatctc tccttttgg tcaatcttct cccaatttgc
28681 tggattagtc atatccacat tatcataaaa aaaaactata atctcatcaa ctttcttatc
28741 aactctctca tcttctttca tgttttcacc ctcgacccgt ttttctctat tctcattcat
28801 acgatctaca tcatcctcca tcacttttc tgtacttttc tcatctacat tatcctccat
28861 attaataggc ccttcacctt ccttatcaac ctcaggtcta gcagtctcat caattccaga
28921 attttgagtt cttgtaactt accttagcat atcaccaatt tgagattttt ctcttcctat
28981 tttgacttcc acttggctga cttctacctt ccggtggcat tgtactacaa ataaagaata
29041 actagatttt agaatgagta tacaataata gaaacaacaa aaagattaag tgataacata
29101 aacacaaaac aacgatatct tatcactttg tctagtttgt cgccgtcgcc gtggctcttg
29161 atcaccgtta acccttgtca aaaaccctaa aattttaatg ggttaagcct tgtgaatttg
29221 ggttggcttt gtaaattttt taattgggtt tgtagatttg ttaatgtatt tatatagtgt
29281 aataagtttg ggctttgttt taagcccatt ttcctctttt tttttgtgta attcggttta
29341 aattaaccaa aataaacatt cagtttaatt ttgaaattaa atttaagccc tataaccgaa
29401 ccgaataacc cgtgtaccga cagaaccgac ccgaaccgaa attaaatttc ggttcaattc
29461 ggatagattt ttcaaaaact gaatttctcc aaaatcgaat aacccgaacc aaataaaccg
29521 atttctccga actcccaggg ctagtatata ctgttgattg cagatcttca tgttttgttc
29581 acaatattga accacaagcg gttcatgaat atttttttaat attaaaataa atttcacatt
29641 acattgaatt tataatgttg aatgtaacaa agactcagtt catctgtatc aaacaatacg
29701 accgctttat ttccactctc cactcacagc ctatacgttg aatggtaaaa acggatcaga
29761 cgacatagat aaaatattta tattataaaa aacttaatat ttacataaac tatacatact
29821 aaaactttta tattaatttt caatattctt ttacatcttt ttaacatgta attattaggt
29881 aaacgtaaaa atattttaat attattgact gtttaataaa aagttatgat tgcataaaat
29941 gtttaatcaa cataaaaata attatgtcat ttttaatttt ctaaaattaa taattttagt
30001 aaatgatctc aaatatatat tttggattta tatatccatt ggagtaaatg aacatctttg
30061 aatatctcaa acatatatta gtgtaaagca tacaaatata agtacataaa tatttttaa
30121 aaggaaaata attataaaaa gtaattgaaa agtggagcaa atacaaaaca caacaaaaag
30181 gtaacgtatc tactgtagta gacatatcca aatatgtcat gacaaatctc tctccgcttt
30241 aactatacct ctaaacaata ttatttttta atttaaaatt gaaattatct gcttgctata
30301 aaaattaacg tgaatggtga aacataaaaa cagtacgtcc gctttgtctt tgaaccgtgg
30361 ttgaaccatg atttcattca caaccaatgt gttttaataa ttttgtagat tttgtgttta
30421 tgtgaagatt ttaatatata tatatatata tatatattta tctgttgttt gttttctatg
30481 atattctatg aggatgtcca aaacgattga gttcgaaaaa catgaagggt aaaaacatgt
30541 taagagaagt ggtggatcat ggaaatcgac gtaagcattg aacttaaatg gaattgttat
30601 attttgtttt tatttttta aaaatacaaa ttaaatgaac ccatcaaata ctcgtaatgc
30661 atattgtgta tatatagtca ctttactatg atgttctact tttggatgaa ttacgtaaac
```


BNSDOCID: <XP_____. _2974277A__I_>

```
30721 tggaaatata aagaaaaaaa tatatttcca accaaaacca aaatcaaaaa ctttagaaga
30781 atatatgtat attatatata tatatatgat atttttatta attaattttt atcttttaat
30841 tgacaaaata ttgttggact tttttaacaa gtcacatggt ttcatcccat ctatatatca
30901 cattgacttt ttccactttc atcctgacac ttgttcgttg ggggttaaag ggaagaatct
30961 gtatggatcg ctcaaagaca actttgagca tctgaaccat tttttaatga acgagggcat
31021 gattcaagag gaagactaca agtgtattgt agaaaaagaa aacaaggaaa gaggaagcag
31081 ctcaaagtca aaacaaaaca aaaagaaaat gagttccaag aaggaaaata ccaaaactga
31141 tccatgtttt ggaaatatga ggttaattat caaatatatt aattattata tctttttaa
31201 aggtaatcag atatataata tatgttttat ttttccagga taaaaaggtt aacaaaatct
31261 ggaatttgga aactatgaga gagattgatg aaagcaagtt gattgagtta ctatagagag
31321 gaccagttac tgtgaatata gaagtattag atgaattaag ctatatacgg ggtgagtaaa
31381 ttacagatat agacatatac ttttcaatta ctatacacta tgattttcat taaaatattt
31441 cttataacca aaacgacagg atgaagttct tagaggatat ctaagaaatg atctagaaga
31501 tcatttgatt gagttgacgg gttatgacac tacttccgaa ggcattaact aatttaaatt
31561 taaaaattca tacggtcttc gctcgaggaa aatggatttg gaaaaaataa ttcgaaagag
31621 tagccgaaag aaacacgagc catctttagt caatgtttac atgtatcgag aggtatacat
31681 aatattttat tttttttac ttggtctatg agtagtcagg tcacagttat actcattttt
31741 ggttaatttt tgtgtgcagc ttctggacaa acgcaaaggc agagagagag gaagaagcca
31801 ttgatgagcg ttaacaaaat gaaaataaac atattttaag ttttttttc tctaaactct
31861 cgttttttt gttgacaatt taattattgt agtaaagaat ttacggctag ctagatcata
31921 atcagcatgt gcggaatatc atatttttaa aaaaaacaaa attctagatt gatcaaaata
31981 tttcttctag tttgattact aatgcttaaa tagattaagt ttttcttttt ttggcaagat
32041 atagattcag ttttagtttg gattatttca tgtatattaa ggtaggcgta actattttgg
32101 aatgagacac tcacccgtca tgtatctgtc ataaatttct cgtcccataa tcagttccgt
32161 atatgggaaa taaaagtatt aacatcccat atccgtcata aatggaaaac gtgtaacaaa
32221 atacatactt gcatgattat gtgtgtaaac caatttattt tatgtagtct ttataacccg
32281 aatttatacc agttgggtaa ttagttagat tttcagtttg gtttaaaact atgttcacag
32341 tctattgatt atttgttata tccttttagt tatcttcctt ggtggactaa tattgagtag
32401 atgttgatct caaatcagtc gtatgctaat acttatgcta ttagaatcat tcataactcc
32461 tctaatctgc aaatatttat gatctttcaa gttagagaca acttattgtt ttgtaaataa
32521 aatattatat ttttatagat ttataatgtg tcccttagct aaattttgca tattacattg
32581 tttaatcata atcaatttga tttttgaata aaaataattt ttttgagata catctataaa
32641 aataatttat tattgatgca tagtttgttt taaataaaaa cagtataaac tagtttacat
32701 tatttcgaaa agaaagatca tgtgaaaaag gaaaaaataa aactacaaat attttgaatg
32761 tataatttat tgtttagata tttttggacc tttatcaaaa tgaaggcaat aatgataatt
32821 actttcaatt ttctattatt agattttaat tcatagtaaa tataaatctt ggttcataca
32881 taatctcgtt tttaagtaat tgtgaaaaaa tattataaca aaattgaatc ttacaaatat
32941 ttttatttca gttttgtaaa aaatttacaa ttaaatctta aaatcctcaa ttgataaaaa
33001 aattttataa ctaaagtttt atttttgaa ttataatgat aatattattt tcttttgaat
33061 taaatgataa tcagaccttc acacatttgg taatgacccc aaaattctaa aatctaattt
33121 atttatcaat caatttttta ttattattag tagtatcaag ctttgttcaa taaaacaaaa
33181 aacaattaat ggtgtcaagc aagttttacg cttttaatcc ttttaataaa aattagtttc
33241 caaaatttta attaaataat atagtacaat tggatattaa aacttagttt tccaattctc
33301 attgattaac aatttattag aaattaattt caaattttcg tccacattta ataagagtta
33361 aatatttttg aaaactaatt actttttta attttatttt atttcttatt tatatatatt
33421 aagtttttga ttactattta gtgaattcca ataaatagta tacatgagta tgtgacctaa
33481 ccataactaa ataaatgtaa aaaaaattca ttgtcttaaa atgtcagaat taacaagctt
33541 tcttgatttt attaaaaaat ttcaaaatga aagattgaga atgttcttct gagttttttc
33601 aaacactctt tatatttttt ttatcaaatt gacaagtgaa taataataaa atcataggaa
33661 atgtcaaatc ttcttacaaa aattcatctc ttaaaaccaa aagaaaaaga aaactttcac
33721 agtttaaatg atcaatcaaa tcacttacac gttgataatg tcaaaggttt tggaaatgat
33781 tcatattttc tttctcttcg ttctctcact ttttttttc tctttctttg gttttgcttt
33841 ctctctcttt cgtctagtct tcacatacat atagctccga atggagaaag cgtatcaggc
33901 ggagcagatt tagcgtatca agtagaacaa aagcggatca agtagatcta actgtacaag
33961 tgcggatcaa gtggttcaaa cacaagtggc tccgaatgga gaaagcctct ttcgtggaga
34021 ggatcaagca tttcaagtag aactaaagcg gatcaaacgg atcaggcgtt tcaagcacaa
34081 ctaaagcgga tcaagcggtt ggaacgcttc aaaatatgta ttgaatggta aaaacggatc
34141 aaccgatgca gatgaaataa tttattgtaa aaatattaat ttaaaataaa ttataataca
34201 taataaaaaa atattactaa atttacaata ttagtctaac atatttttaa acatgtaatt
34261 atataattca gttttaaagt atttatatag ttttattcat tattttggca ttataaatac
34321 caaataaaat gctaataagt ataaaaacaa attaaatttt aaggaatgtt ccactattag
34381 gcacttttaa aaaatgtttt cccgtttgag gcactttata gggaatagac acttttttata
34441 ctaattcttt ttctttttgga ctttttcacc ctcaattcaa tctttatgtt gattattctc
34501 aatttctctt tcttcttaat cttcttccac tatctttacc caaaataaat gaaatttttc
34561 tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc
```


BNSDOCID: <XP_____2974277A__I_>

```
34621 tctctctctc tctctctctc tctctcattc catgatttgt agcgcctctc tctctccacg
34681 tctccataac ttttgggctc gattcctctt atcgattatg gcagcggttc accatcggtg
34741 gtaccgattc tgcaaacaaa tgggagaagc ttcttaaaca attcgatctc gattatctca
34801 ggaatttggt aaaagttcta tgtccataag gtcttggagg tggaaggtct ctgcaaacag
34861 taagaaaaat caacaaataa tttttgggtg tccataacaa ctagtctatt taaatatatc
34921 cattaagaaa acttttact gattcatatc cattctatat accatgtcca taataacagg
34981 aaattcgatt aatcgtggtt tgtaaagaaa ccaaagttcc atgtccataa ggtcttggaa
35041 gtggaggtct cttcaaactg aaagaaaaat caacaaacaa ttttttggtg tccataataa
35101 cgagtccatt taaatatatc cattcagaaa cctttctact gatatatatc cattctatat
35161 accatgtcaa taatatcatg agattcgatt aacagtgttt tgtaaagaaa ccaaagttac
35221 atgtccataa tgtcttgaaa gtggaggtct ctgcaaattg aaagaaaaat caacaaacaa
35281 tgttttagtg tccataataa cgagtccatt taaatatatc cattcggaaa cctttttact
35341 gatctatatc cattctctat accatgtcca taataacaga aaattcgatt aaccggaatc
35401 tcgatgctac atagatgaaa cgagtttgac acatgagaga gggcaaaact caaataaagc
35461 cgccattgtt gaagaatgat aataagtttt gatgtcaatt atcctaaaac accaatcatg
35521 tcgttgtagt attttaggtt gtcaatccaa atgggtgtga tgctaacaat caagatgtga
35581 tcagaagtca ctaagtcaag ccaagcataa catgttttgg tgtgttctag cagtcctaag
35641 tgaacatgca gaaaacagaa aatcaaacag aaacagtaaa acactcgacc aacacagctc
35701 tttgcagagc actgggtgag gtcaagtgac agatcgagta acagacagaa aatgaaacag
35761 ggatactcga ctgcacagtc tgttaccaga caactgggtg gtcgagtacc aggtcgagta
35821 acagtaaaga attgcagaaa ctaagcaaca gatgacagga tgcagtaaac agcagtagag
35881 taaacaataa aatcaatcag acaaagaatt cctaaggatg gggtaattga gtaatttcgt
35941 ttcctcagtt tacaggtgat tgacatgctc aataaataaa ccctagacaa caagttcatc
36001 aaccaaatct aagtgccacc gcaacagaga ctctcaagct aacctagtcc cagactcaat
36061 caccattgac gagagctaac ctaacaggca ttacgaatca acaagttaaa gccaaaacgc
36121 tccctacaac taataccttg gtacagggtc acgaatctct ggaattagtg gttcagacat
36181 ttcatcgaaa acctttttggg cgcggaaatg tctgggctaa aattccagtt gatcagaaag
36241 caataggcat taagaacaac taatccagaa gggatctatc aatcaatact caatcaccta
36301 gcatactgag aattctacac tactctaacc catcctcaga aacctattca ctactcagac
36361 atcataacag aaggcataaa cattgaacaa gttgaaaaca tgctaacaac agtgtaacag
36421 caggatgata acagaatgaa caacagtaaa ggaaatcaag aacaaaagct gaatactgaa
36481 gagaaatccg gattacaaaa agtctagaac acaaggtctg cggaaataaa actaggtaaa
36541 aaacttcttg tctgcgaaaa tgaggagtac atgctactta tagccaaata tcccgaaacc
36601 ctaatactca aaacgacgaa gtatttggac gggttaagaa ctgtactcga cccaggtggt
36661 cgagtgagag gtcgagtgaa tatctggagt cttcttttct tccttgtgct cgagtgtctg
36721 gttgagtgcg gaatgaagaa ggctctgaag cttgtcatcg agtatgcagt caagtgttgt
36781 gatggcgata ctcgacctct aggttgagtg acgaggtcga gtagtggtct ggcagtgaag
36841 gtgaaggcac tcgacctcat ggtcgagtga agtgatcgag tgatggtctg gtgatggaga
36901 tgagggcact cgatctcgtg gtcgagtggt gaagtgaagt ggtgatgtcc tctgctcgac
36961 ctgtgggtcg agtggtctct tgagatcctg aagctcctac tggacctctt ggtcgagtgg
37021 tttgtcgtga tcaggtccgc ttcttgcaaa tagctcgtca acagctccaa atcacctgaa
37081 atcaacgcaa atatgcaaca tgcatgcaaa actatcctaa acatgcaaag tgatgacaaa
37141 tgatgtgaaa tggtataaaa cagtgatgat atgatgctaa agtgatgata aatggatgct
37201 caaacatgca aaatacacac ttatcaattc ccccaaactt agtctttgct tgccctcaag
37261 caaacattca agaacaagct ggaagtgagg tttgaaagcg gggactcaga gccaaagcac
37321 cagatagacg agatgaaatc aatgtccaag ttaaaagttc taagctgcaa gatcgaattc
37381 tacttaaaaa cgttagccat gcctcatcaa tcaatcaatc cgactcatat gctcgaccta
37441 cacaagtttt caaatctatc aatcccttta acattcatta gttctggaac gtgaatcaag
37501 caatgcatca tcaaagaact catttggcta aggtaaaagg tcaagagaca aagatggtcc
37561 cttacagaat aggcttcagt aacaggggat ctctcaaaaa atgattaagc tttagtagaa
37621 tgctaaaggt aagtcccaag ctatgcatac agagataaga tcccatctac ccagagtaca
37681 tcaaatgaag ctctaatggt aatcccatct ccaatcccaa tacaagaact aaactctacc
37741 cttttcattc atgaatcttt taagcacttt ttctaatcat tctttttctt tctcttaact
37801 ctaggagaag ctttctcaac actttgttac atctagcgga tttggatttc tttaacaact
37861 aaggttcctg ttttttaacat tttaaagcca agggcttttg cccatctcca atagctaaca
37921 agaactcttt ctttttcttta caaatcccaa aacctttact agacttttttc tgcttcttta
37981 gcttacttca cctagagtct tctacctttt cttatcattg aatcaacccc ctttttttt
38041 tttttttctt ttaaacacat cccaacccaa agtataaaca cccacctagt cctatccagt
38101 ccaaaaaatg cgaactagaa ctacatgaga cactatctct gtcgttacga acctaacact
38161 ttctaccaag cttaatcgga ataagacctc acaaacactc acaagtgata tacggcttga
38221 aagaaagtct aggtttgggt ttgggttaac tgctctaata aggagagtca gctacttgga
38281 ttaacaagag tgatcaaaat gaataagaga atccaagtat tttagggtat ccatgagttc
38341 aaatttgatg cagacatgat aattaaagat cagattcagg ttcaaattga tagggaatgg
38401 tatcagatca tgacagtgtg gtcgagtaga gcaatctagg catggttcag ttttatttca
38461 aactcattga ttacgcaaca gacaactagc aacgagggca ctgagtttat ccggtgaatc
```


BNSDOCID: <XP_____ 2974277A__1_>

```
38521 aaaatgctta caaggctatc tcatcatcat cccctatgca tatgcaacaa tcctaaatga
38581 aacactctag actcgatcct aaatgtgatg caactacatg aacactcttt ttttttttt
38641 ttatgaaatc attttctgaa atttttcaaa ttttttggt ttttctatg ccttagatga
38701 atgcagactc aaaattatac acaatgcaac acacacttcc tgagccctcc cccaaactta
38761 aatcacacag tccactgtgt gaataaactt ggaagagagt actcaggaca aatccaatca
38821 cagaaacaaa aaatgaagaa acaaatcaca accgacggta taatacaatg gtgaagtgag
38881 gcgcttacct cagtcgaaga atgaacgtag ttgctcgtca atggctgcct gcgattgctc
38941 ccaggccata gacgaaccgc cttgttctgt atcagctcca gcaggtgggt actcgacctc
39001 atcagctctg tggcggccag caccgctctg aggaaactcg acctctctcc ttctgctgcg
39061 gccagcaccg cggtcgcgta aactacgaga ctgcagaagt cgtccgctcc tactgttaga
39121 tcgagtgagt gtggtcctcc tcttcctccc agattcacga ggctcgaatg atgaatgcct
39181 cgcagggaca tgcagaatct cctgctctct aggctcaggc ctgcgccgac ttggctcggg
39241 cgcgtcatat ctcctcgaag gcatgtcctg agggagatgt ccctgcggaa tagctgtagt
39301 ggaagaggag cagctgagct tgtccttcag aaacttgata gcccttaagc acttggcgag
39361 taacttgtcc tgtttcttgc accacttctg cagcttgctg ttgttcctgt gagcttcgct
39421 caaactcttg ctttccctcg caggaggtat atgctcactg aaatgataca tgctcgtgtc
39481 aaactccgcc tcatctatat catccgtctc atcaacatct tctgttgtag ctccttcggt
39541 gtgactaaac tcctcagttg gcggactgcc ctcgaaataa agatcttcaa gcgcaggctt
39601 gaagtcaatg ttcctgccct caagaatcct agtagcgtca atgcagggga gaagaatgtt
39661 ggcgattcta atcgatgaat gctcgaacct gtagcggtga aagtcgccag ccatgtcaaa
39721 ctccaagaac tcacatctgc gcaaatgatc caagtccatc atcctcggtg ctaaacctgg
39781 ttccttgagg gggactccac aagcttccag aattggcgtc acaacgccac ccacgcatag
39841 tgctcctctc accttcttct tcccgtttgt caacgcccac ttcctgtatc cacacagatg
39901 gatcaacaga ggcattactg gtggtgggtt gttgagatcg cctctcaaga cattcttccc
39961 cttcgtctgg cggagaatcc ctgtaagcgc ataatctatc atctccatgt ccgtgttaga
40021 cacggtgcct gtagactccc tggagtaaaa gatgttggcg actgatcgct gaaagtagcg
40081 gatcacagga ctttggattt ggttgctctt agaccgcgcc gagtttagcg gtagattgtt
40141 ccctatagtg gcccacaaat ccttaagctc ctctctgtca aatctgggtt tggtttcctt
40201 tccactaggg aaaccaaaca gttcttctaa cttcttgatc gaaagctgat accgctgttc
40261 attcactgag aacgtcaaga agcctaatcc catggtatcc agctcaaagt ctgtgagtgc
40321 ttcatacatc tccacttgca gcgtggaaag aaactcgatt gtctcctcct tataggcagg
40381 gtaagggtaa gacatgagtg tctccagatg acacttctcg aacaaatgct gcacgtcctc
40441 cagtaacccc agctgggcta aagtctcata acacggatac ctcgttcccc agaagtcatt
40501 catctcgaag aatctgatat actcctccac tgtaacctcg tcgggtttgt tcagcagctt
40561 cgtctccctg cgccatgact ctggctcata ctcagtctcg atatcctcat caatcagctc
40621 atatctactt gacatcgccc tctttcctct ggcgatttca gctcttctct gattacgacg
40681 ggctgtagag cgttcgaatt catccctgaa gctctggtaa acatgttgct ctctcggtct
40741 agcggaccag gactcagctt catccacatt gtaatcaggg tccatagagg attctccgct
40801 ggtgttactc attacggaac ctgaaaacaa aattaaactc aaagatagac gttaatacgt
40861 tagtcccaac aaaaccaaaa acacaaacca acagagtaaa acgatcgagt gtaccaaaat
40921 tttttatttt ttttttttcg gtttggtgaa ctcgcggtct agtgattcct tggttgtgct
40981 tggcgatcga gtgaccttga gaggcgagag atggtgaagc ggtttgaaag aggtcgagtt
41041 tggttgctta atcgagtgaa ccggcgatga cgagctttcg tggggatgag agcagcggcg
41101 atggaaacga tggagaggcg ttgagaagcg gcgagatcga gtgactacgc aaggtcgagt
41161 atgttgagag atggcggtga gagcgatgaa gctgtgaagg agcagcggtc gagtgaactt
41221 gagctcgcgg aagagaagaa acggtggtgt tgtcgagtga tggctggact tgagagatga
41281 gagaacgatg gcaccttggt cgagtgatgt cgtagatgga gagatgatga gcttaaaggt
41341 gataagtgtt caacaatgaa gagaccgagt gatggtggag gtcgggtgat ggagggacaa
41401 ggcgataggg tcgagtgatg aaacgatgaa gaagatggtg atgtgaagcg gtcgagtggt
41461 agtggagagg tgatgagagc ttcggccgag tgatggttgt gggattggcg aggagtggtc
41521 gagtggtttt gagcgagaag tatggggacg gagaacgagt gtttgttgag cttgagagtg
41581 agagttgatg agtgaagtga gagtcgagtg gtgttgggat gaaaataaaa gggagaagtc
41641 gagtgttttt ttttttaagg aaggagaagg cggagaggtc gagtggatgg gagtttgttt
41701 tgtcggtggt gatgaatgag agtgatgggt cgagtttatg gggttgtgga ggagaaaacg
41761 aggcgatgag gtcgagttaa gagggatgga ggtcgagtga tggcgttgac gtcgagtgga
41821 ggagaagaga tgaggagaga tttcctgtcc tttggtctgt tgcatggagg acatgggcgt
41881 cagcggcggg tcccagaatg tcttcttgtc tcttctctgt attctccgtt cactagaaaa
41941 actccaggaa atcccgcaat cctctataaa cttgaaagat tccctggaag actttgtagc
42001 ttatgttggt cgagtaaaat gatgattaaa atttttttta ggtttttaaa catcactcat
42061 actcgatcac tatactcgat ctctaggcca tcagttgagt aaaagttttg aatttttact
42121 ctaccttcac gttacccctt tcttcctgag aaccacaaac ataaccaaaa caaatcaaaa
42181 ataacaagaa agtaaaaatt atttacaagg atagtcatgg gacttcctcc caagtgagct
42241 tgttttaagt ctctagcttg actcctcaca cttatcaaga agaaaatgaa ctgtagagag
42301 ggtcagcata atcctctctg ataccatcat catattcagc gtcctccctg gagaggtcta
42361 cagcagaata ctcaatcccc ctttcttcaa aataggcttc cttctcttct ggagggtaat
```


BNSDOCID: <XP_____2974277A_I_>

```
42421 ctgcatcttg agaccatctt gaagtgaatt cttttctcgc gatagtgtta atcccgcctt
42481 taggttgagc ttcctttctgc atctgattca gcttactcct aagttcctca atagtgtggg
42541 ctagcttctc tatagtttca tcttgcttat cagacctctt tttaagttct ttcactcttt
42601 cagtctcata tccttctcct gagatcgctc tatcatttcc agaggttctt ccatcttctg
42661 ttgacccttg tggagttctg ctgatgttaa agtgcaactt gatgtgcttt ccaaggttga
42721 gacctatttt cccttctctg acatctatca ctgctcccac tgaagctaag aatggtcttc
42781 ccagtattaa gggatcctta tgctctactt ccatatccag caccacgaag tctgtaggca
42841 cttctactcc attaatcatg actggtaggt ctttcagcat gccgaagggt tttctagaag
42901 acctatccgc aaggatcaaa gtcaggtcac aaggtttata ttgaatgaac tctaaccttc
42961 ttgccacaga gagtggcatc aagctaacag aggctccgag atcgcacaga caatcgctga
43021 aggcaaattc ttcaatggaa catggtagag tgaaagatcc aggatcttct aacttctctt
43081 gaacaatcct cttatcagca gctcgagatg gagtggcatc actttcatca tccgaatcgt
43141 gatacatctt gatcctttcc agtatcaaat ttctcacatc tttgtgagga tcaggaatca
43201 ggttaagaac ttccattagt ggcatcactc cttcaatctc gtccaactgc ttcgctgcaa
43261 tagacttgta tctctccatc catgctttcc tgaatgtcca cgggagagca cgtgagggta
43321 gtggagtagg tttgaatcgt tttaccattc tctcgattat ggcggccttc tcttccgagg
43381 tggattgagt gagatggcta gaaccagcgg aatggttgaa ttcaacaaca gagacttcaa
43441 cctgagtaca agcctcccct tcttgaacat cactgtcccc agtgacgaaa tcagggacag
43501 gtcgagttgg cagctctcga tcatggcaaa tggtgatggc gtgagcagtt gcatactctt
43561 ttggattctg tatggacttc cctggaagtc ctgtaacttt tggtgtagag gtagatgctg
43621 attggccttc caagtatctg actttggtat tcaatgcctc cactttggca ttcaggtcat
43681 tgtagctaca gtccagctta tggtgcaatt cagataattt cttagccatt tccattgagc
43741 tagatgcttg tccctgtagc agctgctgga ccatctgttt catttcagca tcaggaccaa
43801 caggaccttg gttttgctga ggtgcaaacc caggtggtgg tggttgctgg tagttcccct
43861 ggaattgctg cttaggaacg aaaccttggt tgtagggaac aaaaggtttg ttctgacctt
43921 gttgttgctg tggaggatac acctgatcct gagcattagc aacgttggtg ctgcggtagg
43981 agaggttggg attgttggtt ttgaagttgt tgtatccttt gtagccaccc tggttgttgt
44041 tgatgtagct gacttcttcc aactggttac cctccccatc ttggacttga tactgctcgt
44101 catcaacaag gaagtgcaca tgcttctgct ggctaagaag aatcctgtcc agcttgtcat
44161 tcagcgcttt gatctccttc ctgtgtttgt catcagagtc agctgttcct cttacggtcc
44221 tatcacagtc ttcattgtaa ttgccatctg attgagcaag gttctcaacc aattcccagc
44281 cttcttcaac atccttgttc tggaaattcc cattgctggc ggtatccaga agcattctga
44341 tgcgtggtag gactcctctg taaagagtgc tgagcagaga ggctttagta aagccatgat
44401 gagggcattg gttggtgtaa cccttgaaac gctcccatgc ttcacagaag ctttcaccag
44461 tcttctgtga gaaaccagaa atctcatttc tgagtcttgc agttctggca ttggagaaga
44521 actttgacag aaaagccttc ttacaatcat cccaggtggt gattgagtca tggggcagat
44581 tcttctccca gatgtgtgct ttgtctccca aggagaatga aacaaacgga gcttgaatcc
44641 gtcttcactg acaccattga tctttgttag gttgcagagc ctatcgaatt catcaaggtg
44701 gtcgagtgga tcctccattg gcaaaccatg gaatttgttc ccctgaatca tcgagatgag
44761 cccactctta atctcaaaat tgttgttctg gatagcagga ggtgcaattc ccttcctctg
44821 acggggatca cgtggtgcat ctccagcacc aatgttggca gggccatttt gttgattctg
44881 gtcgccagcc atttcgagtg gttgctgagg tacagggttg actgtggttc tccttttttct
44941 gagttcgcga gctgtgcggg gtatgttgtc gttgaaaagg agcttctggt tgcctgttga
45001 tcgagttcgc atacagcaga gctgtaccaa tcaagaagac aacaagaaaa acacacagtt
45061 agtaaccgaa aagaaaattg aaaaaaaact tgatctaagc aagtctgaaa tcttaaatgt
45121 aacaaacaaa cacccaattg gcaacggcgc caaattgata ataggttttg atgtcaatta
45181 tcctaaaaca ccaatcatgt cgttgtagta ttttaggttg tcaatccaaa tgggtgtgat
45241 gctaacaatc aagatgtgat cagatgtcac taagtcaagc caagcataac atgttttggt
45301 gtgttctagc agtcctaagt gaacatgcag aaaacagaaa atcaaacata aacagtaaaa
45361 cactcgacca acacagctct ttgcagagca ctgggtgagg tcgagtgaca ggtcgagtaa
45421 cagacagaaa atgaaacagg gatactcgac tgcacagtct gttgccagac aactaggtgg
45481 tcgagtacca agtcgagtaa cagtaaagaa ttgcagaaac taagcaacag atgacaggat
45541 gcagtaaaca gcagtagagt aaacaataaa atcaatcaga caaagaattc ctaaggatgg
45601 ggtaattgag taatttcgtt tcctcagttt acaggtgatt gacatgctca ataaataaac
45661 cctagacaac aagttcatca accaaatcta agtgccaccg caacagagac tctcaagcta
45721 acctagtccc agactcaatc accattgacg agagctaacc taacaggcat tacgaatcaa
45781 caagttaaag ccaaaacgct ccctacaact aataccttgg tacagggtca cgaatctctg
45841 gaattagtgg ttcagacatt tcatcgaaca ccttttgggc gcggaaatgt ctgggctcaa
45901 attccagttg atcagaaagc aataggcatt aagaacaact aatccagaag ggatctatca
45961 atcaatactc aatcatctag catactgaga attctacact actctaaccc atcctcagaa
46021 acctattcac tcctcaaaca tcataacaga aggcataaac attgaacaag ttgaaaacat
46081 gctaacaaca gtgtaacagc aggatgataa cagaatgaac aacagtaaac aaaatcaaga
46141 acaaaagctg aatactgaag agaaatccgg attacaaaaa gtctagaaca caaggtctgc
46201 ggaaataaaa ctaggtaaaa aacttcttgt ctgcgaaaat gagaagtaca tgctacttat
46261 agcaaaatat cccgaaaccc taatactcaa aacgacgaag tatttggacg ggttaagaac
```


BNSDOCID: <XP_____2974277A__I_>

```
46321 tgtactcgac ccaggtggtc gagtgagagg tcgagtgaat agctggagtc ttcttttctt
46381 ccttgtgctc gagtgtctgg ttgagtgcgg aatgaagaag gctctgaagc ttgtcatcga
46441 gtatgcagtc aagtgttgtg atggcgatac tcgacctcta ggttgagtga cgaggtcgag
46501 tagtggtctg gcagtgaagg tgaaggcact cgacctcatg gtcgagtgaa gtgatcgagt
46561 gatggtctgg tgatggagat gagggcactc gacctcgtgg tcgagtggtg aagtgaagtg
46621 gtgatgtcct ctgctcgacc tgtgggtcga gtggtctctt gagatcctga agctcctact
46681 cgacctcttg gtcgagtggt ttgtcgtgat caggtccgct tcttccaaat agctcgtcaa
46741 cagctccaaa tcacttgaaa tcaacgcaaa tatgcaacat gcatgcaaaa ctatcctaaa
46801 catgcaaagt gatgacaaat gatgtgaaat ggtataaaac agtgatgata tgatgctaaa
46861 gtgatgacaa atggatgctc aaacatgcaa aatacacact tatcaaagaa gaagaaggtt
46921 cttcgagaac tgaatctcaa atttatgtca aaacaaagat ttaaattgca aaataaatta
46981 caatcttaaa gaaaacaagg gctcgtattc tcgtattggt ccagagagaa gagaaagcta
47041 tcgcgtagag tggatttgaa gatattatgg gaaaggtttt tttctttgat aaagtgagaa
47101 aaaaaaagaa aaagagagag aagagatggg aaaaggaaag agaagagatg gccgatgaaa
47161 aaatatagtt aaacggtaaa taagtcaaat aattaagaga atattacatg tagaggcact
47221 ttttagaaag tttggccgca aataggcact tcataagcat tagtaacttt cagtgatcga
47281 aattacccct ttaccctgaa atcgtcggtt tatgaaaacc taagaaatat aaccggtttt
47341 ctctatcttc ttttatcacg atcataaaaa aaacacgaac ccagtaaaaa ccctaattta
47401 ctgtttctcc acattgcgac gagaagcgaa tccggccact cctctgactc cagcgactcc
47461 tccgtcactg tccaccaatc cgcgactccg gcgactctat cctctctgaa aataaaaacc
47521 aaatagtcat ttcccccaaa tctctacaat aatcgaattc gatttcgatt tattgttttc
47581 gtaagccaga tttcaatatc ataccagttg attctgactg atcggaatac tctatggaag
47641 gtacaaatca tttccaatca tctgattttg gtccaattga ctcagaagaa gaaatcctct
47701 tgtcgatgaa tgtttctgca aatttttttt ttaaaatcga aaaaggtttt ataattgttt
47761 gtagctaagc acagacgatt atggttttag tattttgtcc aaatactcgt taatctaatc
47821 gttatctctt aaactggttt cttgccaatg aacagggtcg acatctcagg aaccatcttc
47881 ttagtgggat tcaaaatttc caaatagact cttcgccacc gatcaatatc cgaggggaag
47941 gctaaactta tactctagac aagatattct atcctttatc aggaaggttt ttgatgggac
48001 agaggaatta gatattattc ttcagtcttg ctttggatct ctattccatt tgccggttag
48061 tagggttgcg acttcgggta aggtcattca tgccttgtta tgtagacagt tgcttactgt
48121 aacgcccgtg aaccagaaaa tcggttaggg tttgttttta aatcatccgg tttaattggt
48181 tgggttaatt ggaaacccta gagtgtgcct ataaaagaag aaaactcgaa attagggctt
48241 tagttagcca cttttcgttt tataaaggag aaaaagagag aaagggagaa ggagcagaat
48301 cgttagggtt tgggagaaac cagtttcgtc ttatcaaatc gtggtaaaag gtaacgaaat
48361 ttggtagagt tattcccaag tcgtttatcg ttattctcca gtttacagaa ctggatttgt
48421 attaatattg gttgagttat tcgcagtttc gtgagacacg ttttctcaga cgcgaattgg
48481 gaccagcgga gattgaaact aagatcgtgg tctcgtctgt tttccgatcg gcacgagatt
48541 ttgtgggcag ctttgtgacg tctgcggata agatttcacc ggagcgattg aatagtaaac
48601 ggttattttg tattttaggg tttcgtcttt gagactgctt gtttctggtt tttcctgttc
48661 agtttgcttc aaggttgctt ttggttgtgg acttgttgta ggacgtttct ggaatgtttc
48721 agacgcaagg agagtctctc ctggatttaa ttattgttac gatcaatttg ttaaggtgag
48781 tgtgtgacca tgagcttatc tgagcgattg ggttgtatga tttgttttgt gtgatggtat
48841 gtaggtgtgg tgaatgtgtt ttgggtgaat tgttcaatga ctggtttgtt tgtattgtat
48901 tcgtgattat acctttttat ttgcttgtgt gtatagctcg tagatgggag gatcgcctca
48961 ctaagtattt atgataatac ttacgcatct cattgtgttg ttgatgcagg taaagataaa
49021 gtatgatcgt ggaatcatgg caaggaagag aggatgatct aggaactcgt ttgttttatt
49081 tttggtgttt ttggttattg gaaaccggtt aggattgtca gttattcttg gttttgttgg
49141 tcgaattatt tatttattca taaaatgata tttttgactt acggtttaat ttttatgtta
49201 ttggttggtt taattagata tttttgggaa tatttaatta tatttataaa aaaaaaaaaa
49261 aatggggtcg ggttgtttca atttggtatc agagcgatta cggttctagg atgtgtagaa
49321 aaattaattg atgaattatt ttcctgatga ttgatgtggg atgagctagg gtgttagcat
49381 attgaggatt gattatggag tgtttagtca attgtgtgtt gttggagtcc tagcttcatc
49441 ttctccgagc cttaaatgag attccacggt aagttgtttg tgatatattc gtttgtttta
49501 gtttcttagc catttgttct tgagtagcgc agatgggtag aggagctgga gttcgtggtc
49561 gtggtcgtgg tcgcggtagg agacgagtcc tcgagggtac cggcgagagt aatggccaca
49621 gtgccacagt tgagcagagt gtgggttcgc agcatgggtt tgtggagcct ggggttagga
49681 acggtcttgg tgccgatata gctggtgcag ccggagtggg ggctggtgga gctggtgtcg
49741 gtgccggtgt gcatgtcgtt ggtgctgagg gcccaagagt gatgggtgcc gcaaccagag
49801 gagcccaggt tccagaggtt ggtttggcgg gcctactgag acagttgttg gagcggttac
49861 caggtgtggt accagtgcag gttccagtgg cgccacgagt ggcaggggtg cagcagaggg
49921 ctgcggttgc tgaggaggtt ccatcttatt tgaggatgat ggagcacttg caaaggattg
49981 gcaccgggta tttttctggt ggtactagtc ctgaggagac cgacagttag aggtcgcggg
50041 ttgagcggaa cttcggttcg agcagatgcc cggcggagta tcgggtggat ctagcagtgc
50101 attttctaga gggggatgca catttgtggt ggcggagtgg gactgccagg aggaggcagg
50161 cggatatgtc ttgggcagat ttcgtggctg agtttaatgc caagtacttt ccgcaggagg
```


BNSDOCID: <XP_____.__2974277A__I_>

```
50221 cgctcgaccg tatggaggcg cgttttcttg agctgacaca gggtgagcgg tcggttcggg
50281 agtatgaccg ggagtttaac cgactcttgg tgtatgcggg tgggggcatg gaggatgacc
50341 agacccaaat tagaaggttc ttgaggggac tttgaccgga tttgagggtg cggtgtagag
50401 tgtcgcagta cgccatgaag gcgacactgg tggagacggc tgctgaggta gaggagaacc
50461 ttcagaggca ggtggtggga gtgagtccag cggtacaact gaagaagact cagcagcagg
50521 tggcacctag caagggcggc aagcctgcgc aggggcagaa gaggaagtgg gatcatcctt
50581 ccagagctgg acagggtggt cgtgcaggat gtttttcttg tgggagcctt gaccacaagg
50641 tggcggattg cacgcagcga gctgagacga gggagtgcta ccactgcagg gagagaggac
50701 atctcaggcc aatttgtcct aagctgcagc ggatggcaat ggcagtggta cagccggcgg
50761 tgcagcacgg agcgtaggtg tagcaggaag tatagcagtt agcacatatt gcagcggcac
50821 cgcagggtta cactacgcgt gagataggcg gtaccagcaa cagagcgatt actggtatga
50881 tttctgagac tcagactttg tgaattgttt ctgttgtttt ggatatcatt gttttgtgat
50941 gaatgttagg gttcttagca catgaggttt gtgtagggac cttattagtg ggcggtgtag
51001 aggtccatgt gttgtttgac tctggagcat cacattgctt cattaccccg gagagtgcct
51061 cgcgcggcaa catccgtggg gatcctagtg agcagcttgg agccgtcaag gttgccggag
51121 ggcagtttct agcggttctg gagagagcga agggtgtaga tattcagatt gtaagggagt
51181 caatgccagc agatttgatt atcagccctg tagagttgta tgatattagc cgccatgggt
51241 ttatttggtt agtgggaagc tatggtttct gatttgatta gaattagaga gaacgaggtt
51301 aagggttaca tgtaacggtt tttggatatc taatatttat ttggagttac ttggtcaatg
51361 gtttaatagg tcagtttata taagtggaga agtgaaaggg tgtttgattc agggttgagt
51421 tttctgtttt ttcagtcatt ccaaccaaca ctcctgagtg taatgaataa tccataacag
51481 ttggaaacat aggtgtaggt gtatggtttg ctgcatagat gggcttgaaa atgttgacaa
51541 tcatcatata tctggtatca caatatacat tgagtggttg ttttgctttt taaatttagt
51601 acttttttcc gaagaatatg tttatagaaa cgctaaacag attttttttt cgttttttgt
51661 tttttgtact taatcacata tgggtaagca aactaaatat ttatatttga caattctgac
51721 acttagaaat gtaaatgtag tgtaaatgcg ttatttataa aggaacaaaa tattacagat
51781 taagaaagag tccttttatt gtcattttaa gttgaccatt acatagtttt ttagtttgtc
51841 acatacgtta gtttggaatg gtgaacatag gagaggactt gtttgataag tttattcttt
51901 cagttttaca gtacctgagt ttagttgtta aacaatatat attaaacaaa cggaatatcg
51961 acttttatat tttctagagt gtgaagggtg ttttcaaaag taaaaggaag ctgcatcttt
52021 tggtcaaaac taaaaaatag caaaaacttt cttttctttg gggtggcaat gtggaaattc
52081 cattacgaaa ttgaaaagaa actaagaaaa aaaacacaat taaaaagtag acgtaaaaga
52141 caatgaaaga aacataagta gtgacatttt agttgtcgag acggggacgt ttgtccttct
52201 ttccatttga tgaggctgag ttagatgttt ctgatttgcg gtcagctttt ttagttgtgg
52261 caataatctg acagtcaaca tctccatctt tggcgttttc ttcacctgtt tcaataggaa
52321 tggcgatttg ggtatcaatg atacgggaga tagtttggcg agaagtcttg aaattgtatt
52381 ttgtgagttt gatttggaat ttgcatgagc agccaatgat atgctttaaa cagatttgag
52441 ttgtggtgtt tattcattca catgaatatt ttcctgcagc ttgagattga gagaaattaa
52501 ttttagttga aggtgttagt ttataaaaga aagagtggtg gtaaggagaa gtatatctcc
52561 atgtatttat ttaaaacatc acgaacgctt cttccttcaa ttttggtgcc ttcaccatca
52621 aagcaacaaa agttccaacg tctctgtttt tatcacgcac tctcatttcg aaacggtacc
52681 tgtggatttg agagtttggt acaacgctgt gtaatgtttt gtgatatatc cttatttta
52741 ctatttttag ctatgattta ggtattagtt ttagagtcta tttgttattt ctagagttct
52801 ttatagtctt gttcaggtct ttacatgttt gggattcatt tggagataat ggagtatttt
52861 ggaggaaaac agagatttta tgctgcagaa gaaatccaaa gactaaaaat cgcataggta
52921 tccaacgaca ctgcccctgg caattcaaat atgcaactta ttgaagtttt gaagatttct
52981 aaattcgact ctgaactttc tcttatttct aaagaaagcc catcacgttt taggtcatat
53041 atatacatag ttttaaaggt cattgtttag acctaagttt tggggaggtt attgtctttc
53101 tctattgttc ttgggagcaa aaccctgtag acactccctt tagagaagat tcttctaact
53161 cttttctctt gatatattca tcattgattt cttattcaaa catgttctgt tcttcgatga
53221 tcatgcttga atagttttct tgttaggttt agggtttctc attagggtta taaatgattt
53281 agtagattgc ctcttgctag gttatctgta ggattgttca tctttaattg ttcttaatac
53341 tagttctaaa gtagctaagt agaacttgat tccagataat cgatataccc gccataggta
53401 tcttttatcg gaactattat ttgatgagcc tggttaatag gtgtgtaata agaatcggca
53461 tctctactaa ggtgaactta tctaactcgt ttagcatgca tgctttggaa tagggtttca
53521 cggaacgccc ccgggaatct tcgctctagg catagtgagt ttcacggaac cccaccggtt
53581 attcttagat tagagtttga gtttaagaac ggtttgataa caacctagct atcactagat
53641 ctactaaccg tttgtttcct gagaataccc taaccctgac gttttcttat tttattttac
53701 atccccatca tctgctattg cttgcttttc tttcttttat tttctttatt attttgttta
53761 gcttaatttg ataaccttag tctattgtgt gatccgagag ctttgtggaa ttagaccttt
53821 aagtgctaac gacgatctct tatttgaggg agtaggctta aggcaatttg agctatatca
53881 ttttgtatga atataacctc atttaggtcg ttcatagttg tcaaacctga taaggcaaat
53941 gttgttgtat tctttgcaat ggttgcaata cattgatgtg tctcctgttc caagtttatt
54001 gttgcaaatg ttgcattaga tgtagtacca tccatattgt ggttgtatgt ccacaactgt
54061 agcaaagcaa gtgaagctaa cttcctacat tggtaatgtt aggtttctta aagcagtaag
```


BNSDOCID: <XP___ ___2974277A__1_>

```
54121 aaatggtggt taatgtgatt tataaataaa gagagaatac tttgaatata acatgtggtt
54181 ttcattatgt agaaactggt atatttcatt gattgtgact gtctccagtt ttgttacatt
54241 agaggaggag cttgcgatag ttggcgtttt ttaccactat ctcctaacct ggtggtggta
54301 gtgacttagg catgtggaat atagttgttt gaggagattc ttatattaga aagagataca
54361 taccaagtaa ggaactggtt ggtttcatcg acgtcggtgt caaagaatag acgtgtggat
54421 gatgtgttgc tgagagacac aacagctgaa gagattatgc agattatatt ttagaattat
54481 agtataggtt aatgttacga agcattgtgg gacaatggta aacaagtgat ttatacctcc
54541 gattgttttc gagttgacag tagtgataag gaggactatt ggtttgactt cactttggtt
54601 ccactattct ctgaatgttg ccgcaatgtt ttcccagaga tttatgcgta tggtctgatc
54661 gctgtaaatt aagagttttt ttttttttta ataattggaa tgaatgtttt tgaaaagaag
54721 atgataagat tcattactct ttgagtagta ggtataagaa aactcgatcg cttgatctag
54781 ggccttcaac atgggggggct gtttccattg tggtaaggga gtgtaggttg tcaccactaa
54841 ttaggcgtag ttatccaatg acatctacaa aacacagaaa tagaaggagt agttgtcgtg
54901 aaatattatt tgaatagccc aggatgtatt actagtagct taccaaagag gtcaatattt
54961 gtgtacacaa ttgaagcaaa ttcttcgaaa ccacgaaatc taatcttttg ttcatcaatt
55021 tgatggtttg ccgtaggcac gagtaccatc aacgttttgt cgttgaagaa gatgccatgt
55081 ttgtggtcgg acactttgta tattattttc caggggggtga tcagaaattt tttgagcttg
55141 tatatggggt tgcgtttaag actctcttgg taacggaata gacggtttgc cgaaatgaat
55201 acttgcattg tgtttgactg taggacgaat gattagtata aattcgtgtt gggtatgtaa
55261 attcagcata ttagatatat ctgggagtcg attagcagta attcaagatc catcaatatg
55321 ttgtcttttt tgaaatttcg ggattcctag taatgaatta gtcttactaa gacttcttcc
55381 tcacagaaat ctccatcaat ctcatcgaaa atctggaagg tgaatggttg gggagactgt
55441 ggagctttt gtggtggcat tttgtgattg taatgagagg tctggttttt tttgtttctc
55501 tttgtcctga ggaggctttg ttttatagtg agacccaaat cgttgtgaat gaaggttgtg
55561 gccattgatg ctttttttaat gtgaatgtat gggagagtaa agatgaggaa gtggagagaa
55621 ttaaagatgg cttgtcgggt ttttttagg aacctgaagc gataaaactt cagaggatag
55681 agacgacgac gagtgttggt gacgaagtag gttttttttt cttaattttt ttttgacgaa
55741 gtagggttta tattgtatta ttagtagatt cttttgggcc aacgtttttc aaaagatgtg
55801 gatttgtatt tcttttgggc caaggggttt tactagtatt tggtccatta gagttgctga
55861 ggtgtacttt tttatgctta ttttggcatc cgtttaattc ggagaaagta gctgctgacg
55921 tgtctattag ggagagtact tttgagaagg ctgatgtgga caccttgaga ggcgtcaagg
55981 agttcctttt attagtagta gatttgtttt attcaataaa aaaaattgta ttatgaacaa
56041 aataataata aaaaactata tttcaagact ttatttttgat ttgaattctt taaagatttt
56101 tatgtcgaac cagtttttag aaactaccaa tttgagactt ttttttaag tgacagtcat
56161 taaaatttttg aaagtaacaa caatttaaaa gtcattaatt tttaaaaaat taacatcaaa
56221 tttctagtta aagccatcaa attttgtaat ctatcaacca tgtatatggt aagtgcatag
56281 tatatcggtt taaaaagtac ataaataaaa aattttacgt cattacaata taactaacat
56341 taaaacctgg aaattaactt gtggaaagca ctaaaatttc acgtcatcct gtaaagtaca
56401 aaataaatag tttttaagga gaatatatat aatgactgag atcaaggctg tctcaataag
56461 atatgatggt taaagaaaat ataatttttt ggcctattaa aagaaagttg aaaaaactgc
56521 aaagaattgt actgttggag acccaaaatc gtaatacttt atgaagaaaa ataagaatta
56581 accattctac cactagagtt tttagtaaat tgatgctcct agtttttaat actaatttga
56641 tgcctaaagc ccttacttca cggactttaa ctcctggtcg ggccagactg agatcaaatg
56701 aagaggacta aaggaagaag aacataaatg aatagagaaa aaatgagaag aaagctgaaa
56761 tctcatgaac ttgtggaaag attcttttta gaccacttta attcaattta tttacctatt
56821 aagaatatca ttttttcata tcacctactt tactttggct ccgactagag aaagtggatt
56881 tagcagtcta agcggatcaa gtagataaag tgaggatcaa gcggttttgg tggaaagagt
56941 gcggatcaag cggttatggc aggtctaaaa tgtgtgtgaa tggtgaaaga ggataaaaaa
57001 aacgcatatc aatatgaggt gaatgcattt cacgaaatac cattgtcctt cagcaaaaac
57061 tctatggatg ttaatatcat aaatgtaaga agctatgcat caatcactat ttcctaaaaa
57121 attacataaa caacattagt ttcatctctt atagtcttca tatattagat tcaacaaaga
57181 taagaagttt catcttacca tttcatcgca tagaatcatg tagtatcttc ttattctttc
57241 atgagagagt gaacgtcaaa gaacttccac ttttttgttt ttggagaaat ttgagagatt
57301 tatgagattt ctttgtacaa acttcctcta tcgtcttttg ccatgtttat gcttttctcc
57361 aattaatagt gtagaagaga gctgtacaag gttgaagata gtaatgtata taaagtaaat
57421 attggataca tatgtgaaag tgtatatttt tgattattta gtgtatattt gttttacatc
57481 taaattaaaa attgaaacac gatttctaag atatgcaaat cttgtcaaga tttgttttgg
57541 attctcaata aaaatgtata aagctttata tggttacttt caaaaagaat tagtaaaaac
57601 taaaaaatat gacataagaa atcatatctg tttgtaaagt tgtcaaatat agttcataat
57661 ttaaaatgca aaaattttgt tgaagaaata tatttttgtt tatcatatag atagttgtta
57721 tattcctcca aatcaatgag tcatacaatc tttgattttg ttaagatttc tataaccgta
57781 atatacataa ttcaaaaact tcatttaatc atatagttaa ttttttggca tacttatata
57841 gtttgtaaga tattatattt atatttataa ttaatttagt ttatatcgta aacctaaaat
57901 aaattcttca tttaatcata tagttaattg ttatgtttaa aaaactaaga cagttaaacc
57961 aaaatatgta ttttaaaatt gagtatatat taataagtaa attaaaaata aaataacata
```


BNSDOCID: <XP_____2974277A_I_>

```
58021 tataaaaata atatatgtcc tacttggatt tttattgaag tacaaacatg tgtgtttaaa
58081 aggtttgggt tttatttggt ttaagtattg gtttaaaatg actactttta tttgggccta
58141 cattaattat gaatagattt tacaatccaa ttaaaataag ttacgaatcc catatattta
58201 gattgttaga tatgactatt tcacaacata aagctaaaaa aaatgttagt tgttatgtaa
58261 aagagtctaa aaaaatgaaa cctaacatgt attcattaaa aagtgaaatt ataaggatgg
58321 atcattcaaa catgataaat agaagatgca tatcataaaa aaaattaaaa aactaaattc
58381 caaatagaat aacatataaa aaatacaaat tagaacttgc atattcaagc aacaaatact
58441 aaataactta tataaaaagg aaataatcca aaataataca aaataaaaga aagacttcat
58501 attttccatc cgaattagac ttcaatttgt taaaatataa attaaagagt ttagagatca
58561 ttcttcttgg gcaagtcact acacaaatct tttataatct ctccattagt attgttttca
58621 ttgactttat tttctattcc cattttatag caatgcttct tacttgtaga acactggcca
58681 tttccattcc tattttcaat ttccttacgt tttgatgaag gagtggaagt tgatgacaaa
58741 tcaacttcat aaacctcaac accattagtg agaaaaagcg aaaaccaaat agaattaata
58801 gataacgaaa ttttaataag tataaatgat gaaatacata tttataagac gattgaactt
58861 gttcgatcaa gagaaaccac tgacgaagga ttcacataaa tatctgagtc ttcaagaacg
58921 atttcatcaa caatattttc atcatacagt ctagttacat tgtagtagtc atatctacta
58981 aagatgtttt cgttctgaag atataagaga aactgaaacc atttaaaaaa attgaagata
59041 aaccatggtg aacatttaaa ctttttcata attaatacct tagtagtagc aatcacttcc
59101 atgatagttt gcctaggata cttagcatca tcatcatcct caatttttga tgcaaactat
59161 cttctataga acgaatttgt tcctcctaat gttattccat cgcttgaaat gttgcataaa
59221 aatatttaat gtttatacaa ttataaattt aacaaaaata attgtggttg ctttgattca
59281 attacctttt cataaaagca gccaatgaat ccaacgttgt ataatcaatg atttaggaaa
59341 catcaaatgc attcactcga caattatatt ttgacaataa tataatagtt ggaaaatagt
59401 atacttcagt tgagtttact ttgaaaagat agattttcac aaagcaaatc aaagaaacaa
59461 gataccatca atagattctt ggcatgcttg atcaacaatc tttgaaaact ttcttaaaag
59521 agtacacgcc aagcgattat cccttgaagg tttaaaaata aagatgaata gagggtaaat
59581 aaacatcagc caacttaatg aggaaaaaat aagggtaaag tgaatatact cattgtctct
59641 gatatggaaa tcaatttttt tttgttgatt tgttgttagc attagatttc atctcgaaca
59701 acaacttgaa caatcacatc ttgacaaaat aataaatata gataagattc aacgacatat
59761 ttctaaaatt aataaatacc aaatataaat ttaaagaaaa accttaccga ctaagagact
59821 tgtgttagca cgactggaaa gaatagcatc atacatggca agggataaaa acaactcatc
59881 aaagatgtta tcatatggag agacaatggt tttaaaaaga aatccaatct tgtacttcaa
59941 gtcaacaact caatattgcc cacttgttgg attcaatctt ccattcaccc tcgactaaac
60001 aatttttaaa ttgtttatgt gatctctgct aacaacagca tagatcttat caccctacaa
60061 aaaaccttat attgttaaca agttatatgt gaataagatt gtgtatataa taatataata
60121 cagatctatg gagattaaaa actatttaaa gcttgatcag cgaaaatgat ttccaatgat
60181 tcaccagttt ttgtggagga ttgtctccaa ccatgaagca atttaaccta gatattccat
60241 gcatttttgt aagcctcgag agatttcaag gggggaaatg gaagaagatt agcaaacatc
60301 atttttttct cttttttta tttgttttgt atgaggaaga atatgcaaaa ttggtgaatg
60361 aaacatgcaa atggggaaga cggtatttat aagagttgaa gggctaggtc caaaatattc
60421 tatttggatt atgtcttgga aaaaaagttt gtgattaatc taaatcaaaa aaatataggg
60481 aaatattaag gtgattgtct taaaattagt ttgggaaata tagaatatgt taggaaacgg
60541 tttaatgatt cgagagattt ttgaacaata atattgattt agatttttaa aatatatatt
60601 tgattactta tagttgggta aaaaatttaa atttacttgc attaaagtaa gataattcaa
60661 aaacgattgg gaacattcct tttccagggc tgagattata agttatgacc cgttcattat
60721 cattcgggta aaaagatccg cacatttgtt aatggattaa aattaaatca aaggttcata
60781 ataaagaatt ttgatggcta agattagaac ttactattcg actgcaacaa aacgggttaa
60841 aatagatcca cgtccattaa aacgattttt tttatgtcat aaattcttta acagaatgac
60901 attctttgta aatactaatc aactttcagg gtagctccat aaacatactg ctcaaataat
60961 agtatagatg agctgaataa actgaattag atgaataaga taaaaattat gctaattggt
61021 gacaatctag ctaaatagtt gaatgtagag aaatattatt taaaataaaa tctaatgttg
61081 ctattgagaa ctacgtcgac ttgccaccat atttgctcag caatttgaca acaggcgata
61141 ttcatataat tttcatcttg tgtttcttct tccatcccaa cactttgatc atcattgtgt
61201 ttataagtag agaggaaact tacctttcac tcatgactat aattattttt aaaacaagaa
61261 aaggagttta gctcataata ttttctatga agataaatct tgatgggcag agtatattta
61321 taggaactga tattagttgt tggttttaat aaaataaaaa tatgaattgt tcaaaacagt
61381 taaatagaga aatgatttgt aggaactact attagatgtt ggtgtctgta cttgaatatc
61441 ttctatttga ttacaaagta tccaaagatt aaactaaatt agtttagtat ttaaattttt
61501 gagattctct aaacaaacaa taaatttata taaaataatt taaataaatg ataaaataaa
61561 ataagtaagt tttcattttg ttcaatttgt tcctctccta atttgaggat gaacaatttc
61621 caactaaaaa gtgtaattat ccaactttaa aattggtttg aattttgttt taaggcaaat
61681 tgtgactttt tggaggaaca aagttgaatc aaggtttcaa ctctttcaac ttagccgatt
61741 aaggcatcca tttctagctt atgtttgttg taatttttca atatctataa tcttattact
61801 ctatatatat aatatttata tatatatata atatgtatat atatatatat tgttttgatt
61861 aaatagaatt atgcatcata aataatctat tgattattct tcttttatta tctaaaggac
```


BNSDOCID: <XP_____2974277A__I_>

```
61921 ataaaaatat tgatgttata aaagagtgaa gtatagagaa gggactaata attgaagaag
61981 actttccgtt gattttcttt gcctaaattg tgtttaattt ttacaattga gtgaaaatgc
62041 atatgtatac tacaaaatat agacaacatt actgtaaata acaaatgaca atgggctcca
62101 tatcatgtta attgtaaata acagccgcta ccgtaataat tctcaggatc gttctgagaa
62161 ggttggggga cagctgggga agtgtaattc cgaattggac cagagcttcg ggaagcggga
62221 aggttaatct ggcataatcg aattgtagag tacacacgga ccagctatat cgttcaggtg
62281 agaacagagg tcgacagacg agcgaggagg gttctagtcg tcaaaaggtg agttgtgcct
62341 ttccgccgaa gagtcgaagc ttcataatgg ggatgaagga ttcataattg cagaaaaagt
62401 agaaaatага ttcttttcga ttatttgtcc taaagaggaa gagtaagtat aagaagtgaa
62461 agagaagaaa ggagtatata ccttacacga acttgtggct gagaacgaaa cggccacgag
62521 agtcgttgga aaagtggaaa gtcaaacaat atttaaggag ggggggatgg gtaattgttg
62581 gtgttggata ccctacggaa gatgcaataa gtcaacaaga gtgaaagtca aacaatattt
62641 gagtgtgcga atccgacatg cacaaccgcc acagaacatc agccccaaaa tcagtcttct
62701 cttaacacac tagtaagaac tcaggttcct cgcccagatt tggatttgag taccaatgac
62761 aataatctct gtacttgatc aatcctttca gtttttcatt tgtttactgt gaaactgaat
62821 gtttagtatc ttttttgga ttgtaaaact taaaacctta aacttatgta tcggaatgta
62881 ttttggattg atttaggatt tgtgttggaa tttgtgaatt gtttgttgta tttaattgta
62941 tttagattca gaataaacga aacattgttc agtggcgacc attgaacaaa aagagattgc
63001 cactattatt tcaaaacaaa attagtggct aagtttttgc aaagatggtg gctaaatcag
63061 ttccaaatat tcaccaatag tggttacgat tacagccacc actaattatg gtgacaaaat
63121 agtgccattt tgctatgata atagttagta actcaatagc taattgctac aaaaaaaaat
63181 agtacccaag cagtagctac attagatacc atcattttgt ttctgattca gcagttaact
63241 actacataac caatggttgc caattggtca ccatatagcc actagtaagt gaagtagcaa
63301 cgtagtagct aatcagccac acatggtgat tacggcgact gaatttcatt tagcttattt
63361 ccataagcaa ccgaataaaa ttagtgcagt aacagtggca aatttaaata tagcgaccga
63421 tttgatatat taaagactgg aaatgtagtg ttctacgcaa tctcaggctc gtccaatcac
63481 aaagacgaaa ccaattttag gctcatcaaa tcaattggca tgaaatcaat atcaggctca
63541 tccaagcatt tggcataaaa ccataaaacc aatctcaggt ttgtcaaata atccatgtac
63601 gaaacaaatc acaggttcgc cgatcatgag acatgaaatg caatcccaag ctcgtgccgc
63661 caacacgtaa actcaaatta tatcatcatg ttatatataa agtattccta tcatacatca
63721 tatcttcctt ctatgtcaca caatttgaat caatagacat acatcaaggg acaaagtggc
63781 taagcctata cacgcaattc aagaagaaaa tacttataaa gtaaccacaa tgtttccgca
63841 aatcccgcct cacttttgcc ttaatttagt tagaaatagt gactagcaag ttgcaaatgc
63901 gactagaaac agcgacaatg ttctggaaac ttgcagctaa agaaataaat tcaaacaata
63961 tattagggtt tctttttta acaagcacaa ttgtttcttg aacaacaagc aaacgtcgtg
64021 attgctaggg tgtacctttt ccggcaactt tgaagactta ggtcgctttc ctcacacctc
64081 ctctgtcaat aacaattgta ggtctcgcct agtttgctct aacttgaatc tgataatttt
64141 gttttaaaaa gttaggtttt ctagaaatca aaacaagtga tttttcatca acttccaaaa
64201 caatttctga atgcaatata cctcctttta acaaaacata aactgctatg gattccttga
64261 agcgagatgt atcgagagac gtcaaaatat tgatctaagt ctttctgttt cgacgggaaa
64321 tctgtccagt aaatccttac ttggcacaat ttctcttctt cttcttcaca ataaaaactt
64381 ttatttaatc caaagatcca acagatcaaa acaaggagtt gtagaagaag aagaagaaat
64441 aattaggtcg ggtttcctta tccaaaagaa ctagtacaat gggtctcgat tctggtgggt
64501 gatgtgaacg acggtgacta cgtgaagctc cgctccaatg gtgataatgt tgcctttact
64561 ttcaccacaa tttgcctttg taattggact agaattttat cttggcggct taaactagat
64621 tttctaaaga agaaggatat gttgacacac atacacatgt atatggccac atacaaaacc
64681 ttgtcgtttt atgattttgc tttacttggg ttttatgggc ttcaacttaa ttacgattga
64741 ggatgttaca atttttcaag cttatggaat tcgaggacga attcggcaat gaattatatt
64801 ttcatttata tatgatttga actaacataa cactaaaaat aattcagttg acctttaaaa
64861 actcaaatat caagtagtgt agcaataact aaatgaagca taaatttgaa aatccaaata
64921 caataagaaa caaaaatcca aatataaatc aagtctcgaa acaaattatt tgtctcttcc
64981 caaagccgat cgttttggag gtgatgacaa ttgttcctca gtttggttga gttgttttgt
65041 cttgcaatag cgtaatgtct gggccgccca cacctatgga aatgataaat tcctcttggg
65101 gaattctctc ttttatgccc caattagcca catttcatgc aactgcctga ggttatctaa
65161 caaacttcgc ggtgtagttt tccacaagta tcacacacat gagctttctc tcgcatttgg
65221 ttatccaagt cctcttgtga gaatccatcc cctttgcagc ttcaaccgtc tctttttcta
65281 gcccaatttc ttaattagca gctatttcca ctagctcaat catgtttggg tagtagttcc
65341 tctcacaaag gattttgaag tttaatcctc atatccctta taaacatgcg aacaagctcc
65401 tcctcatcca agaatcggct tgaaaacctg ctcaattgat cgaactccac cacatattcc
65461 cttatagtcc atattttcct ggcggagttc tcagaactct ctcttcatct ggtacattgt
65521 gatgtagaca tcaaccttac ttccataact gctgaactac attatgtacc acatactgta
65581 ctacatcctg taccagaagc cccagaaggt cgcataacac gttcaaaggc taaaaaattg
65641 aagaaaaggt ttaactttgt tgtgctctat attatgagct ccttagagtt tgagggagta
65701 aactggtcag tagcacccat taccggatat gaactaatag ctgaagatga gcttgtcaag
65761 cattcaccaa aatgagtctt gaacgaacta ctcccctagg tatgaaatct catttgacag
```


BNSDOCID: <XP_____2974277A__I_>

```
65821 gttcaagagc ccttaaaaac caagtgaaga ctttggagct gcaagagaca aacgattcaa
65881 gcgtcacatc tgaaatcaag agagtcaaga tcaagaccaa cacatgaagg acgtggaaga
65941 tctcgaagaa actaattaaa cttctcaaga ggaactccag aaagaaccct acttcatgga
66001 actagccagc aaacaagcag aagacgtgat ccagaagtac ctcaacaagt actataagaa
66061 gtgacttact cgacatctat tgggccaaaa acgcatccag ataaggttta ttcaattaat
66121 actatctttc caagttcaca cagcccaaaa aacggctaga ctcatcatca aacgtgctcc
66181 tcaagagaag ttaagtagtt tccttatata gtttattgtc tttctaaaat caatggcttt
66241 tgtgctttga ttttttttt ttccttgtta cattttgca tttcttttt gaccctaggg
66301 ctagctgaag gagtcttata aagtctgcat ggagaaagtt gtacaatgtt tcaaccttaa
66361 tggtcaataa agaagccgca ttttcttatg ttttttctct taaggtttct tgagagaata
66421 agcttgctct ttgtcttgtg tttgtgattc accaaggtta agtttgttta ctatatcaga
66481 caccaatcac atatttgtgg tgtacttcaa tccgccagcg ttccccctcta tttgatccca
66541 cacgagagaa gggatcgcca tcgtgtcaga tctacatcat ggaccatttg tccttcacag
66601 gagagaagcg atcgccactt tgtgaagcca ccatctctat ctaatacaag gctcaccacc
66661 atgtcatgta tcatctggta tcagagcgca gttacgcttt ccttgccagg tttcattgtt
66721 cccttatttt ctcatctgct ttacttttg aataagtttt tgtatcagtt ttcttgagaa
66781 tatccctaag cctaacgtct taattacttg catttacaaa cttaaaattt gcagtttgct
66841 aatctgcttt cttttatttt cattacttat tgatttagct taatttcgat actataacct
66901 tttgtgtgaa caagagagct tcgtggaaat caatccctaa ataattacaa ttggcctctt
66961 atttgagata gtagtttagg gtaatttaat catatctcaa attggcaccg ttgggaattt
67021 catgggaatc accatttggt ttagtttaga gattttgttt aaatttttct tttccttttg
67081 atactaacgc gtttcttgtt tttcttctct tatgtgcttc aggtgtatct gatatagctc
67141 aaattatcct taagcttact ccctcaaata agagatcgtc attagcactt aagtgttgaa
67201 tttcactaag ctctcgttgt tcacacaata aataagtgat aattaatatt aagctagaca
67261 aagtggatga aagtaaagac aagaaaatag taaatcaagt tctaaattca ggagattaat
67321 gcttcaggtc tagggatagt cttagggaat agctaaattg tagatctagc tatgactatg
67381 attgttatcg acccgttctc aaacacaaac cactattata gaataaccca acgttggctt
67441 ggcattcctc cctatatcta agctaaggat tctctacgat gggttggaaa atctctaatc
67501 ctaaatatgc atgctaaacg agtttgataa cttcacctaa gtaaataagt caattctaat
67561 tacacaccta aaagttttta aggggttgtg ttatgatggg gcactttaca agtttgtaac
67621 tccatttaga gcactttgga gggagtaggc acttttgga ggtttttcac tatttgataa
67681 aaatcaatac taatttaccc ttctatcatg tgagcaatta acccaactat gactaacttg
67741 tttattctca gccattggat atgtatgaaa gagaatagat ctaatggctt tggatgtatc
67801 tctcttctca ctttattttc tctctctttc ttttataatt ttatttttct ttagtttgtc
67861 acattcttta ttttattaaa tatcttttaa ccatttttgt aatatatata ataataacat
67921 tttttccttc attttgaatc ccatattact atatttattc gtattattat ttattgtatc
67981 atcaacccat taacagattc agaaaacaaa aaatccatat gatctaatcc atatggacaa
68041 caatgaaata tataatcttt aatgagaacc atgtaactat aaactctaaa atcattataa
68101 acaatcatag ttccttaaca ctaaatccta aaccctaaac attagatact aaactctaaa
68161 cattaaacct taactcctat atttaaaccc taaaccctaa attgtaaccc atagacacta
68221 agcccttaac catatacata taccctatac catataccat aaatcctaaa cactaaaccc
68281 taaaccctaa acctaaaacc ttaacccaaa accctaaacc ctaaacctaa aaccttaacc
68341 caaaacccta ttacattacc ttaaacatag tcctaaacct agactctaag ccatatacca
68401 taaacactat accctaaata ctatgcacta aaccctaacc ccaaaaccct ataacattac
68461 cctaaaccct aaatcctaat ctctaaaccc taaaccacta attttctaaa ctataaactc
68521 taacgctaaa ctctaaaatc cagacaataa taattatacc ctaaacccta aaccttaaac
68581 cctaaaatat aattcctaaa cagtatttcc ctaaacccta aatccaaatt aataaataaa
68641 ctattaatac ttatcattaa accaattaca cataattagc tttaggttat ccatcataac
68701 aaagccacgt ggtttgacaa cccgaaaaat acttatatat tagggcatat attgttcaac
68761 tatacatatc aacgtggtca tatattttgg gttggaaaaa tctagagatt ttgtatactg
68821 gtccagcctt gatagtctct acaatccttt tttttggaaa tgtgtaaagg caaaaaattc
68881 ttatgtattg tatggagaaa tccttataac ctaaacccta aaccctaaac cctaaaccct
68941 aaaccttaaa ccctaaaccc ttaaacccta aacccctaaa ccctaaacca taaagaataa
69001 acccttaaca atatacatat accatatcac gtaccctaaa ccctaaatcc taaattctaa
69061 gctcttaact ttaaaccttc ttgttgttgt ccatcaatat taaacataaa tctcctaaat
69121 ataaacaaat ttaaatagaa aatccatatg atctaatcca tatggacaac aatggaatat
69181 ataatctcaa acgagaaaca tgtaaccata aaccctaaat cattgtagaa aaccatgatt
69241 cttcaaccca aaaccctaaa tcatctaaac taaacgataa atccaaaaca ataaagccta
69301 aatcctaaaa tgtaaaccct agatactaaa tcctaaaccc tagacattaa accctagccc
69361 ttgactttaa agcctaaacc ctaaatccta agttttatat gtacataatc tataaccatc
69421 acttctggtc catgataatc taaagcattc ttcttaatgt gaaagcgtaa aatattattg
69481 agatcgatgt gttcataaac aatcatacaa tatacactcg tatttattat ttgttaatat
69541 gtaatcgtgg ataaataaat attaacattt tcgtaaataa acttttgtgg atatgaccaa
69601 cagaacttgt ttcaccatat ttatcaacat ctttttcata acttaacacc tgttttatgg
69661 atagtattat caaaacgcaa tcactttgaa ttatcgacca tcaacacaca tccaccaaac
```


BNSDOCID: <XP_____2974277A__I_>

```
69721 tctgaccatg tggtcatttc aatcattttg aactatcgac catcgatgca catccaccaa
69781 actctgacca cgtggtttac atacaaataa gaatgacaaa atagtaattt gtatacatat
69841 catcctttcc ttgaaaccct agccgcctat cactctgcta tcatcttctt cctccgactc
69901 attaacctat tcaaccacca tgtcttcttt gattcatcaa aaccataacc attattacca
69961 ccatcatcat catccctctt ccacctctca catttcacct ctccatcaac tatcacaaaa
70021 acaaattaaa gtctttacct tattcatcaa aattaggtca aattaaacaa agaaaagtaa
70081 aaaaaaaatt gaaaatattt gattgtatca ccggatctta acaattttga ctaatcaagt
70141 tgaatcaaac aaattggttt ataaaatcta taaaatttaa tttctatttg aatcaaacaa
70201 gttgaaactt tgctcaaaaa tttgagacat gtaattgaga tttcagagag aggaagaaga
70261 aagaaggaaa atagatttga gagagatgtc acgtcatttg ggagagagga agaagaaaga
70321 aggaaaagag aaggaagaag gaagaaagaa aagagtagga agaaaaatat tatttaattc
70381 ctcaactaca ccaaagtaac tatggttaac ttacttttga gtttatggta aaaaaacttg
70441 aaaaagtgcc tacttctatg aaagtgctac attgcctaat tagaactcaa aaagtgtctt
70501 tctgcctaat attctttaaa aataaatatg ttactaaaaa tatgcatggt ctgttttaca
70561 aataagacaa acttctggga cttgttaaag acttgatgaa acgtgggaat tacttttttca
70621 aataagggtg tcgattgaca tagatgtgat gtcgatcgac accatcgtgt gagtaacgtt
70681 tctactagtc ttctttcctg gaaagcttcc aaagttcctg attatgctcc atagctctca
70741 aaagaagtta tgtacctata atggctgcaa aacatacgaa aaggactaga aagactcaaa
70801 aaattatcct aatactatgt taaaaatcag taaaaaccat gatatatcaa ctcccacaga
70861 cttagcttat gcttgccctc aagcaaaaca taaatccgaa catgtggaag aagtttgaaa
70921 atgtaaaaac tcccgatctc tctagcttat cagcttgatc attcaaacca tatccacgta
70981 cttaacagtt ctattgaagt cctaaccacc atgtagttat ccgcaggctt acacaacatc
71041 ctaaatttca taccaaattt cacaactata tccattaggc tttcatcgca gggtctcatc
71101 aatcaagtta gaactcaaga acctactaga ctcattcatg tactatacca tgataattta
71161 accattaaaa catttttttt gcatttttte tcttcccaaa cttattcttt tcagagatta
71221 gacaaggagt actcgctcat cggataatgg gatcaaagga aatattccta gccctttaca
71281 ctacagaaaa tcatcttcat tactctcttt ttctgtaatt tgttatatgg gaagagtaat
71341 acttattctt tttcacattg gtcttattat caggtatgaa caaggagtca aatcctatga
71401 acaattaaca tttttgatga ggtaaaagga agagcagtga cattagctcc agcctttatg
71461 gattatgttg gaaagatgga tgtccttaat tctgttggct tgccctctgg attttccatt
71521 agtcgggttg aatcgaatat gcagtactta atcaatacat gcaataaaat cgagagaaag
71581 atcataattt cagacaaaat tttgattcaa taaaacgaca ctctattaaa aataacgcaa
71641 ctgactctta agaaaagata agataaagtt agcaatagac tccctagact tagataacat
71701 tgtccgaaga gttattaagc tacagaaatc aaaagaaaaa tttattaaga gttaaaaagt
71761 tagagaaaca aagttgccac attaaccatt aagcagtgtc gatcgatacc aagttggtgt
71821 caatcgacac tctagtcgtt actgtgttgc tttgcgattt cagctcagca ttgctgagct
71881 tatcttagtg ttttaacatg gatcctttag ttgaaactgt taaatcacga taacacacat
71941 caataaggtt caacaaagca taagaaaata agtgcaacaa gtttcgaagc aaataacaaa
72001 ctgtctcaaa gaaacaaaaa atgtataaag agaaagaaaa actaagtgaa aacaagaagt
72061 aaagataggg agaatggcta atcctcatgc tcaaagcctg aagtgggtat tggagagcaa
72121 cgacaacaat gagtaggggc aatacaacca gaacaagaaa gtgtagctat cgcaatccaa
72181 acgctcttaa aactgtcaac tacatgctgc tgaaaggctg atggctccat cggaatgggt
72241 ggatgagtag gaagctcgaa agtcaaggct gaagaagatg acccaaaaca aagacccccа
72301 acagggtcaa tgaactcatc gactgtctgt gttggtgcgg atcccatggt tgtgcctctc
72361 tatcgcctag tgtgtggtgc agcgtgtagc catctcccga tcaaggtgga aaccaagctc
72421 ctcaagctgt ccttgtaatc tctgtaagct tagtacaggg caactgattc atatggttcc
72481 gaacaccaca aatgaagaag ataagagctc cttcctgagc atttcgcacc ctgtcatccg
72541 cataataacc ctaacagatg ccataaattg ccacactagc tctggataca aatcatactg
72601 cctagtacaa atcgttctca agcctatatc atctagcaat cctaaacaga cttctccatc
72661 cgcatacaat gcattaaatc cgaatccgca aaccgagtag gttgggtatc cacattcaac
72721 caagcattgt agaacaatgt atcatattca tctcaatccc gagaaaaagg gtgatgaaca
72781 tagtcacttt ttgtgaattc gctcattctc cattttgggg caagggaaac gtgaaaacga
72841 ccgtgaaaag ctacaggagg ttgggtctgc ggtggtggtg agtcgccggc ggtagtggtg
72901 cgccactatg gctgggagga ggaagaagaa cctgcactgc caccttтctt ttgtcgcttt
72961 ttctcgccaa ccattcttca gatcaataaa attaatactg ggaatgaata gagagcaaaa
73021 ttctaatgaa tttagacttt gaaatcaaag ataacagtga agaaacgaga aatctagaca
73081 gaaaaaacaa aaattgtcga tgagaaaata aagaagaaga aattgaagct taagggtttt
73141 aacgcaacat acctcgaaat taacggtgtc gatcgacact agcttagtgt cgattgatac
73201 cattgagctt gctcgacacc aaaactcgta attaagcctt ttttttcttt ttcttttgaa
73261 taatcaaaga tagaaaataa ttaagtgaaa taaaaaaact atatataact taccagtggg
73321 ttgcctctca ttaagcgctt gtttagagtc aatagcttaa ctttggagca gttcaatcaa
73381 taggtggatc ctgaagatcc aatggtctcg ccatagaaat ttcggtttca accaaataat
73441 gcttgactct ttgtccattt actgaaaatt tattgccatt tttcccaata agaacgatag
73501 ttctatatgg ttgaactgaa cgaatctagg gacctgacca acgagaagac aacttgccat
73561 gaaatatatt tagtcaagaa ttataaagaa gtacctggtc attaggtcga gggatcgaga
```


BNSDOCID: <XP_____2974277A__I_>

```
73621 gatgatcttc ctgtcattat aggccttggt gcatctgcaa caaggtagtt aactacgtca
73681 gcataccatg gataatcgtt ttcaactact gtggcagcgc agttctttgt gaaaatccag
73741 tcgttcgtct gattagtcag gtgtcgatcg acccttgaa tggtatcggt cgacaccctc
73801 gagcatggtc gaatctcatt ctcaactgta acggaggtgt cgattgacac tttgaatgtt
73861 ttttatctat cgacaatgca gctcgattca tcaactcaac aattttgtag tcgtctttct
73921 tagttgtatc aatcatgtat atattttctt caagcaagaa gtcattgatg ggaacatcat
73981 cttcaatcct aattcgtgac aaatgatctg caacaccatt ctcaactctt ttcttatctc
74041 gcacctcgat gccaaattcc tgaagaagca aaatccatct tagtagcctt ggctgagcat
74101 tcttcttttg catcaaataa ttcatggcta catgatttgt atgcaaaatc accttagatc
74161 caacaagata agatctgaat ttttcaaaag cgataacaac agctaacagc tccttctctg
74221 ttgttgcata atttctctgc gcatcatcaa gtgtcttact agcaaaatag atagcatgaa
74281 gtttcttatc ctttctctgc cctagaactg tcgcaactgc aaagtcacta gcatcacaca
74341 ttactttaaa aggtattgtt gaactattgg tgcgcttatc aaggcttgtt atatctgacg
74401 gtatgcatca tgacactcta ctgtaaactc aaacttgaca tccttgtaga gcaaagctgt
74461 caacggtctg gcgatttttc tgaaatcctt gataaatctc ctttagaaac cagcatgacc
74521 aaggaaattt ctcatggctt tgacgttatc aggtggctga aggcttgtca taaccttaat
74581 cttagctctg tcaacctcta tgccatgtgt tgaaatcctg tgaccaagaa ctattccatc
74641 ttgcaccatg aagtgacatt tctcccaatt caaaactaga tgtttctctt cacatcttgc
74701 caacactttg tagagattct cttaacaacc ctcaaacaat gatccataaa ctgaaaaatt
74761 gtccataaag accaccatga agtcctcaat aatgtctgta aagatagaca tcatacctct
74821 ctaaaaaggt tgtagaagca ttgcaacgac cgaaaggcat cctgcgataa gtgaatgtac
74881 catatgggca agtgaatgtg gtcttctctt ggtaatcggg atgaatgggg atttagaaga
74941 aacccgaata tccatcaagg aaacagtagt ccttgtgatt cgctatcctt tccagcatct
75001 gatcaataaa tggtagaagg aaaatgatct ttcctggtag aagcattcag cttccggtaa
75061 tcgatacaca tacgatggcc tttaatcgtc cttgtaggaa tcaattcatc cttgtcgttc
75121 ttgaccactg taacacctcc tttcttagga actacatcga ttggactgac cgaactgcaa
75181 tttgaaatag gatagatgat ccgtgcttcg aacagcttca tgatctcttt cttaacaaca
75241 tctttcaaat tcagatttag tcttctctaa tgctcaattg aagacttagt ctcatcctcg
75301 agatagattc tatgcatacc tgtaatatca tcaagagagt tgccaagagc attgcgatat
75361 atgcgcagct tgcttagcaa tagggtgagt tctgcgtgac ttaaagaggc attaacaaaa
75421 acagggtaag aagatttttt accaagaaat tcgtacgtta gccctgaaag aagtggcttt
75481 agctcaatct taggggcttt atctagaccc ctatcggcgt tagaaatcag ttttggggtc
75541 gtagatggaa gatcttcttg gactattagt tgcatcacat gttgggtctc atctagcagc
75601 tttgttgtcc aattctgcga tatcctctat tgataatacc agtctctatc tctgaaatac
75661 tttttgaagc tattggtgcg aaagaatcta tcagaaacgt tttaccttct ataatcggct
75721 tcttaaccaa cttgttcata tcaaattgca cagtaagatc accaacattc agccatattc
75781 ttcccttgtg aatgtctatg atagctttag ctgtgtgcaa gaaaggtcta ccaaggatta
75841 aaggatcctt aagttttctt tcataagata aaatcacata atatgtggga atcatgctat
75901 tctcaatcat cacaagaaca tcatatagga caccatctag aatacgggtt gaacaatgtg
75961 ctaatacaag agtgatatta gtaggcccaa gattgaaata cccaatatgt gttgctacag
76021 aacgaggcat aaggtttaca ctagaaccaa gatcgcatag agctcttggg aaggtagctg
76081 gtaccctaaa ccctaaaccc taaactcaat agttgctccg agtaaaaagg taccaagatc
76141 ttgtagcttt cttggagtcc ggttctgaga tcatcataac gctgcattca ttcttcagat
76201 tcttggtgac caatgtcttg atatattttt gtatcgtaag agagattaga atagcttcca
76261 caaaaggcat ttcaaaagtc agtttgtcaa tcatagcctt gcatctgaca tcatctaaag
76321 cttttcttgc cttcttagga tccttagaaa aaggaatact tcgatttgct gattttgctt
76381 cagctgtagt tgggattttg tctgaatgaa ctgggtgttg attgacatgg gtaagtgtcg
76441 atcgacatta tcgggcattg gttcctcctc cttcatgaaa ttttcgattc agctcgacac
76501 aggaataaca gcttcttcaa gattaattgg tgttggtcga caccctgttg tatcgattga
76561 cactgaaggt ttttctgcaa ctggagaaag cttagctctt gatgaatgtt ctctctaaag
76621 aactgggttg actagctttc cactacatag cgtaactaca ttgacggact ctgttgtaga
76681 ggaagactct tggagtttcc tgatgtctgt ttgcatttga gataaaccga tttttagcta
76741 tctgagtagt ccctctagca tggctttagt tttattctta gtgacatgga acaaatttac
76801 cctaagctta ccgtcttaaa taaagagttc atggttagta cttagtgttg aatttcaaaa
76861 agctcttttg ttcacacaat atgcctgtcg aaatcaagct agatagataa taagaaagta
76921 agaaagaaga taaagcaaag caaatagcag attatgggtt gtaagataag ataataaaaa
76981 ctttaggttt agggtattct taggaaacaa atgattagta gatctagtaa aagctaggat
77041 gttatcgaac ctttcttaaa ctcaaatcac tgttagatgg taacaggtga ttcgtcatat
77101 tactcactat gtctaaggct aaagtttccc agcgttggct ttgcaaatcc ccaatttaag
77161 acatgcatgc taaacgagtt tgataagttc acctaagtag ataggccgat tcttattaca
77221 cacctataaa cctggctcat caaaaagtat ttctggtaac aaataggcat acctattatt
77281 taagatc
//
```


BNSDOCID: <XP_____2974277A__I_>

Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

2/5/04 5:27 PM

BNSDOCID: <XP_____2974277A__I_>

(12) SOLICITUD INTERNACIONAL PUBLICADA SEGÚN EL TRATADO DE COOPERACION DE PATENTES (PCT)

(19) Organización Mundial de Propiedad Intelectual
Oficina Internacional

(43) Fecha de Publicación Internacional
11 de Enero de 2007 (11.01.2007)

PCT

(10) Número de Publicación Internacional
WO 2007/005980 A3

(51) Clasificación Internacional de Patentes:
C12N 15/82 (2006.01) *C12N 5/04* (2006.01)
C07K 14/415 (2006.01)

(21) Número de Solicitud Internacional: PCT/US2006/026234

(22) Fecha de Registro Internacional:
5 de Julio de 2006 (05.07.2006)

(25) Idioma de Registro: Inglés

(26) Idioma de Publicación: Inglés

(30) Información de Prioridad:
60/696,890 6 de Julio de 2005 (06.07.2005) US

(71) Solicitantes (*para todos los Estados designados excepto Estados Unidos de América*): **CORNELL RESEARCH FOUNDATION** [US/US]; 20 Thornwood Drive, Suite 105, Ithaca, NY 14850 (US). **NESTEC S.A.** [CH/CH]; Avenue Nestlé 55, CH-1800 Vevey XX (CH).

(72) Inventores; e
(75) Inventores/Solicitantes (*para Estados Unidos de América solamente*): **TANKSLEY, Steven, D.** [US/US]; 17 Bridle Lane, Dryden, NY 13053 (US). **LIN, Chenwei** [CN/US]; 600 Warren Rd, #4-3a, Ithaca, NY 14850 (US). **BEN AMOR, Mohamed** [TN/TN]; Avenue Nestlé 55, Tn-tebag-korba, XX 8070 (TN). **MCCARTHY, James, Gerard** [US/FR]; 298 Vallee De Vautruchot, F-37210 Noizay XX (FR). **PETIARD, Vincent** [FR/FR]; 28 Allee Cote Fleurie, F-37100 Tours XX (FR).

(74) Agente: **REED, Janet, E.**; DRINKER BIDDLE & REATH LLP, ONE LOGAN SQUARE, 18th And Cherry Streets, Philadelphia, PA 19103-6996 (US).

(81) Estados Designados (*a menos que sea indicado de otra manera, para cada tipo de protección nacional disponible*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

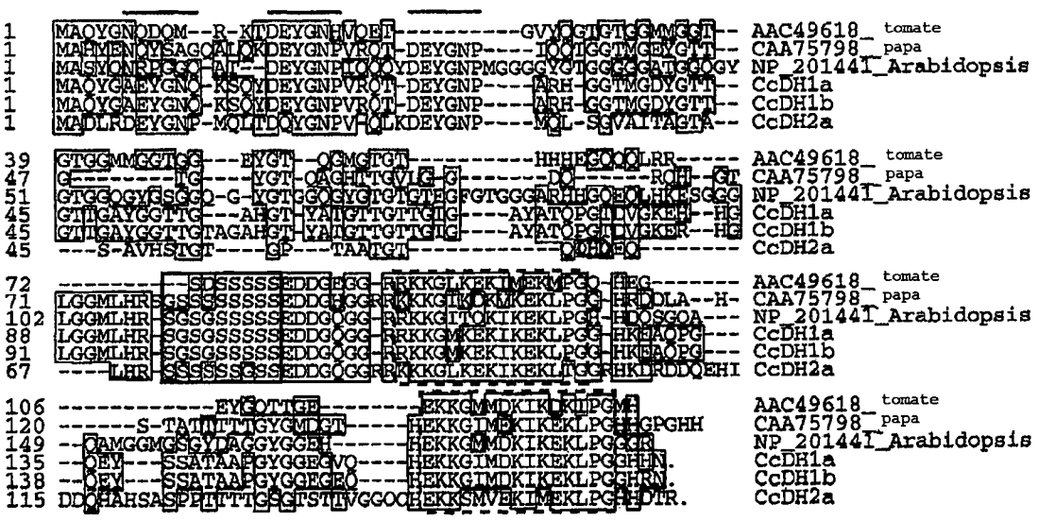
(84) Estados Designados (*a menos que sea indicado de otra manera, para cada tipo de protección regional disponible*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Euroasiática (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), Europea (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publicada:
— *con reporte de búsqueda internacional*
— *antes de la expiración del límite de tiempo para modificar las reivindicaciones y ser republicada de nuevo en el caso de recibo de modificaciones*

(88) Fecha de publicación del reporte de búsqueda internacional:
16 de Agosto de 2007

Para los códigos de dos letras y otras abreviaturas, refiérase a "Las Notas Guías sobre Códigos y Abreviaturas" que aparecen al comienzo de cada emisión regular de la Gaceta PCT.

(54) Título: GENES PARA DESHIDRINA Y PROMOTORES DEL CAFÉ



(57) Resumen: Se describen polinucleótidos que codifican para deshidrina -una proteína abundante embriogénica tardía (LEA)- de plantas de café. También se describen una secuencia promotora de un gen para deshidrina de café y métodos para utilizar estos polinucleótidos y secuencias promotoras para regulación del gen y manipulación del sabor, aroma, tolerancia a la tensión y otras características de los granos de café.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 January 2007 (11.01.2007)

PCT

(10) International Publication Number
WO 2007/005980 A3

(51) International Patent Classification:
C12N 15/82 (2006.01) *C12N 5/04* (2006.01)
C07K 14/415 (2006.01)

(21) International Application Number:
PCT/US2006/026234

(22) International Filing Date: 5 July 2006 (05.07.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/696,890 6 July 2005 (06.07.2005) US

(71) Applicants *(for all designated States except US)*: **CORNELL RESEARCH FOUNDATION** [US/US]; 20 Thornwood Drive, Suite 105, Ithaca, NY 14850 (US). **NESTEC S.A.** [CH/CH]; Avenue Nestle 55, CH-1800 Vevey XX (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANKSLEY, Steven, D.** [US/US]; 17 Bridle Lane, Dryden, NY 13053 (US). **LIN, Chenwei** [CN/US]; 600 Warren Rd, #4-3a, Ithaca, NY 14850 (US). **BEN AMOR, Mohamed** [TN/TN]; Avenue Nestle 55, Tn-tebag-korba, XX 8070 (TN). **MCCARTHY, James, Gerard** [US/FR]; 298 Vallee De Vautruchot, F-37210 Noizay XX (FR). **PETIARD, Vincent** [FR/FR]; 28 Allee Cote Fleurie, F-37100 Tours XX (FR).

(74) Agent: **REED, Janet, E.**; DRINKER BIDDLE & REATH LLP, ONE LOGAN SQUARE, 18th And Cherry Streets, Philadelphia, PA 19103-6996 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

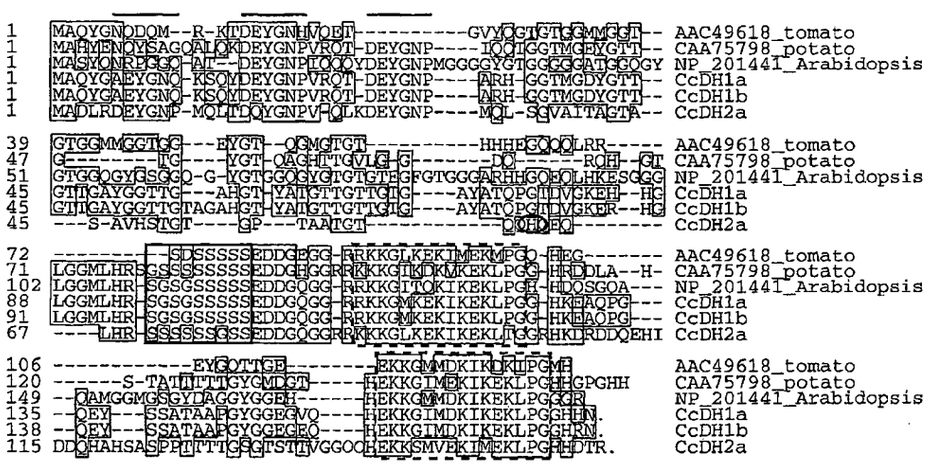
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
16 August 2007

[Continued on next page]

(54) Title: DEHYDRIN GENES AND PROMOTERS FROM COFFEE



(57) Abstract: Dehydrin- and late embryogenic abundant (LEA) protein-encoding polynucleotides from coffee plants are disclosed. Also disclosed are a promoter sequence from a coffee dehydrin gene, and methods for using these polynucleotides and promoter sequences for gene regulation and manipulation of flavor, aroma, stress tolerance and other features of coffee beans.

WO 2007/005980 A3

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TRATADO DE COOPERACIÓN DE PATENTES

De la
AUTORIDAD DE BÚSQUEDA INTERNACIONAL

Para: ver forma PCT/ISA/220	**PCT** **OPINIÓN ESCRITA DE LA AUTORIDAD DE BÚSQUEDA INTERNACIONAL** (PCT Regla 43*bis*.1)
	Fecha de correo *(día/mes/año)* ver forma PCT/ISA/210 (segunda hoja)

Registro de referencia del solicitante o agente ver forma PCT/ISA/220	**PARA FUTURAS ACCIONES** Ver párrafo 2 a continuación	
Solicitud Internacional No. PCT/US2006/026234	Fecha de registro internacional *(día/mes/año)* 05.07.2006	Fecha de prioridad *(día/mes/año)* 06.07.2005
Clasificación Internacional de Patentes (IPC) o clasificación nacional e IPC INV. C07K14/415 C12N15/82 C12N5/04		
Solicitante CORNELL RESEARCH FOUNDATION		

1. Esta opinión contiene indicaciones relacionadas con los siguientes ítems:

- ☒ Casilla No. I Bases de la opinión
- ☐ Casilla No. II Prioridad
- ☒ Casilla No. III No establecimiento de opinión respecto a novedad, nivel inventivo y aplicabilidad industrial
- ☒ Casilla No. IV Falta de unidad de invención
- ☒ Casilla No. V Afirmación razonada de conformidad con la Regla 43*bis*.1(a)(i) con respecto a novedad, nivel inventivo o aplicabilidad industrial; citaciones y explicaciones que sustentan dicha afirmación
- ☒ Casilla No. VI Ciertos documentos citados
- ☐ Casilla No. VII Ciertos defectos en la solicitud internacional
- ☐ Casilla No. VIII Ciertas observaciones sobre la solicitud internacional

2. **ACCIONES FUTURAS**

Si se hace una demanda para examen preliminar internacional, esta opinión generalmente será considerada como una opinión escrita de la Autoridad Examinadora Preliminar ("IPEA"). Sin embargo, esto no aplica cuando el solicitante escoge una Autoridad diferente a ésta para que sea la IPEA y la IPEA escogida le ha notificado a la Oficina Internacional de conformidad con la Regla 66.1*bis*(b) que no se considerarán las opiniones escritas de esta Autoridad de Búsqueda Internacional.

Si esta opinión es, según se estipula anteriormente, considerada una opinión escrita de la IPEA, se invita al solicitante a presentar ante la IPEA una respuesta escrita junto, cuando sea apropiado, con modificaciones, antes de la fecha de expiración de tres meses desde la fecha de envío por correo de la Forma PCT/ISA/220 o antes de la fecha de expiración de 22 meses desde la fecha de prioridad, la que expire más tarde.

Para opciones futuras, ver la Forma PCT/ISA/220.

3. Para detalles adicionales, ver las notas a la Forma PCT/ISA/220.

Nombre y dirección de correo de la ISA:	Fecha de terminación de esta opinión	Funcionario Autorizado
Oficina Europea de Patentes D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	ver forma PCT/ISA/210	LOUBRADOU-BOURGES, N Teléfono No. +49 89 2399-7342

Casilla No. I Bases de la opinión

1. Respecto al **idioma**, esta opinión ha sido establecida sobre las bases de:
   ☒ la solicitud internacional en el idioma en el cual se presentó

   ☐ una traducción de la solicitud internacional en , el cual es el idioma de una traducción proporcionada para los propósitos de búsqueda internacional (Reglas 12.3(a) y 23.1(b)).

2. Respecto a cualquier **nucleótido y/o secuencia aminoácida** revelada en la solicitud internacional y necesaria para la invención reivindicada, esta opinión se ha establecido según:

   a. tipo de material:

   ☒ un listado de secuencia
   ☐ tabla(s) relacionada con el listado de secuencia

   b. formato de material:

   ☒ en formato escrito
   ☒ en formato legible por computador

   c. tiempo de presentación/suministro:

   ☒ contenido en la solicitud internacional según se registró.
   ☒ registrada junto con la solicitud internacional en formato legible por computador.
   ☐ proporcionada subsiguientemente a esta Autoridad para la búsqueda..

3. ☒ Además, en el caso en que más de una versión o copia de un listado de secuencia y/o tabla relacionada con ello se haya registrado o proporcionado, las declaraciones requeridas de que la información en las copias subsiguientes o adicionales es idéntica a aquella en la solicitud según se presentó o no va más allá de la solicitud según se registró, según lo apropiado, fueron suministradas.

4. Comentarios adicionales:

OPINIÓN ESCRITA DE LA AUTORIDAD DE BÚSQUEDA INTERNACIONAL

Solicitud Internacional No. PCT/US2006/026234

Casilla No. III No establecimiento de opinión respecto a novedad, nivel inventivo y aplicabilidad industrial

Las preguntas de si la invención reivindicada parece ser novedosa, comprender un nivel inventivo (ser no obvia) o ser aplicable industrialmente no se han examinado respecto a:

☐ toda la solicitud internacional,
☒ reivindicaciones Nos. 1-39, 41-47 en parte, 40 todas (falta de unidad)

porque:

☐ dicha solicitud internacional, o dichas reivindicaciones Nos. se refieren a la siguiente materia objeto la cual no requiere un examen preliminar internacional (*especificar*):

☐ la descripción, reivindicaciones o dibujos (*indique los elementos particulares a continuación*) de dichas reivindicaciones Nos. son tan poco claros que no se podría formar una opinión significativa (*especificar*):

☐ las reivindicaciones, o dichas reivindicaciones Nos. son sustentadas tan inadecuadamente por la descripción que no se podría formar una opinión significativa.

☒ no se ha establecido ningún reporte de búsqueda internacional para toda la solicitud o para dichas reivindicaciones Nos. 1-39, 41-47 en parte, 40 todas (falta de unidad)

☐ no pudo establecerse una opinión significativa sin el listado de secuencia; el solicitante, dentro del límite de tiempo prescrito, no:

- ☐ suministró un listado de secuencias en papel para cumplir con el estándar para ello provisto en el Anexo C de las Instrucciones Administrativas, y dicho listado no estuvo disponible para la Autoridad Internacional de Examen Preliminar de forma y manera aceptable para ésta.
- ☐ suministró un listado de secuencias en formato electrónico para cumplir en el estándar para ello provisto en el Anexo C de las Instrucciones Administrativas y tal listado no estuvo disponible para la Autoridad Internacional de Examen Preliminar de forma y manera aceptable para ésta.
- ☐ pagó la tarifa de suministro tardío para el suministro de un listado de secuencia en respuesta a una invitación de acuerdo con las Reglas 13*ter*.1(a) o (b).

☐ no pudo establecerse una opinión significativa sin las tablas relacionadas con los listados de secuencias; el solicitante, dentro del límite de tiempo prescrito, no suministró dichas tablas en formato electrónico para cumplir con los requisitos técnicos para ello provistos en el Anexo C-*bis* de las Instrucciones Administrativas y dichas tablas no estuvieron disponibles para la Autoridad Internacional de Examen Preliminar de forma y manera aceptable para ésta.

☐ las tablas que se relacionan con el listado de secuencias de aminoácidos y/o nucleótidos, si están solamente en formato electrónico, no cumplen con los requisitos técnicos para ello provistos en el Anexo C-*bis* de las Instrucciones Administrativas.

☐ Ver Casilla Suplementaria para detalles adicionales

Casilla No. IV Falta de unidad de invención

1. ☒ En respuesta a la invitación (Forma PCT/ISA/206) para pagar derechos adicionales, el solicitante ha, dentro del límite de tiempo aplicable:

   ☒ pagado derechos adicionales.

   ☐ pagado derechos adicionales bajo protesta y, donde aplique, el derecho de protesta.

   ☐ pagado derechos adicionales bajo protesta pero el derecho de protesta aplicable no pagado.

   ☐ no pagado derechos adicionales.

2. ☐ Esta Autoridad encontró que el requisito de unidad de invención no se ha cumplido y escogió no invitar al solicitante para que pague derechos adicionales.

3. Esta Autoridad considera que el requisito de unidad de invención de acuerdo con las Reglas 13.1, 13.2 y 13.3 es

   ☐ se ha cumplido.

   ☒ no se ha cumplido por las siguientes razones:

   ver hoja separada

4. Por consiguiente, este reporte ha sido establecido con respecto a las siguientes partes de la solicitud internacional:

   ☐ todas las partes.

   ☒ las partes que se relacionan con las reivindicaciones Nos. 1-39, 41-47 en parte

Casilla No. V Afirmación razonada bajo la Regla 43*bis*.1(a)(i) respecto a novedad, nivel inventivo o aplicabilidad industrial; citaciones y explicaciones que sustentan dicha afirmación

1. Declaración

Novedad (N)	Si:	Reivindicaciones 4, 25-30
	No:	Reivindicaciones 1-3, 5-24, 31-39, 41-47
Nivel Inventivo (IS)	Si:	Reivindicaciones 4
	No:	Reivindicaciones 1-3, 5-39, 41-47
Aplicabilidad Industrial (IA)	Si:	Reivindicaciones 1-39, 41-47
	No	Reivindicaciones

2. Citaciones y explicaciones

ver hoja separada

Casilla No. VI Ciertos documentos citados

1. Ciertos documentos publicados (Reglas 43*bis*.1 y 70.10)

   y/o

2. Descripciones no escritas (Reglas 43*bis*.1 y 70.9)

   ver forma 210

Se hace referencia a los siguientes documentos:

D1: KASUGA MIE ET AL: "Una combinación del gen Arabidopsis DREB1A y del promotor mejorado de sequedad - y bajo - tolerancia de tensión a la temperatura de tensión-inducible rd29A en tabaco por transferencia de gen." PLANTA & FISIOLOGÍA CELULAR MARZO DE 2004, vol. 45, no 3, Marzo de 2004 (2004-03), páginas 346-350, XP002421281 ISSN: 0032-0781

D2: FERNANDEZ DIANA ET AL: "Café (Café arábigo L.) genes expresados tempranamente durante la infección por el hongo de la roya (Hemileia vastatrix)" PATOLOGÍA DE PLANTA MOLECULAR, VOL. 5, NO 6, NOVIEMBRE 2004 (2004-11), PÁGINAS 527-536, XP002421282 ISSN: 1464-6722

D3: WO 01/77384 A (EPIGENOMICS AG [DE]; OLEK ALEXANDER [DE]; PIEPENBROCK CHRISTIAN [DE];) Octubre 18 de 2001 (2001-10-18)

D4: WO 03/027249 A (SYNGENTA PARTICIPATIONS AG [CH]; SU WENPEI [US]; ANDON NANCY [US]; HAY) Abril 3 de 2003 (2003-04-03)

D5: KAYE CLAUDIA ET AL: "Caracterización de un gen CAP160 para espinaca y expresión de dos proteínas de climatización fría de espinaca en tabaco" FISIOLOGÍA DE LA PLANTA (ROCKVILLE), vol. 116, no. 4, Abril de 1998 (1998-04), páginas 1367-1377, XP002421536 ISSN: 0032-0889

D6: ALLAGULOVA CH R ET AL: "Las deshidrinas de la planta: estructura y funciones putativas" BIOQUIMICA. BIOKHIMII DE SEP, 2003, vol. 68, no. 9, Septiembre de 2003 (2003-09), páginas 945-951, XP009079176 ISSN: 0006-2979

D7: WISE MICHAEL J ET AL: "La pregunta "POPP": ¿Que hacen las proteínas "LEA"?" TENDENCIAS EN LA CIENCIA DE LA PLANTA, VOL. 9, no. 1, Enero de 2004 (2004-01), páginas 13-17, XP002421537 ISSN: 1360-1385

D8: SATOS ET AL: "Análisis estructural del cromosoma arabidopsis thaliana 3.1. Secuencia de rasgos de las regiones de 4,504,864 bp cubiertos por sesenta P1 y clones TAC" BANCO DE GENES, 28 de Abril de 2000 (2000-04-28), XP002974277

Re Ítem III.

La presente opinión está restringida a las invenciones 1 y 4 como se definen a continuación las cuales han sido sujetas a búsqueda.

Re Ítem IV.

Las invenciones/grupos de invenciones a parte son:

-Invención 1, reivindicación 1-39,41-47 parte
molécula de ácido nucleico que tiene una secuencia de codificación que codifica la secuencia como se indica en SEQ ID N°7 u 8 que codifica para dos variantes alélicas de un primer gen deshidrinas llamado CcDH1A y CcDH1b, y materia derivada: mRNA, cDNA (SEQ ID N°1 o 2), oligonucleótido, vector, célula central, planta transgénica, método, promotor.
-Invención 2, reivindica 1-39,41-47 parte; 40 todos
molécula de ácido nucleico teniendo una secuencia de codificación que codifica la secuencia como se indica en SEQ ID N°9 o 10 las cuales son dos codificaciones unigenes distintas para un segundo gen deshidrina diferente llamado CcDH2a y CcDH2b, y materia derivada: mRNA, cDNA (SEQ ID N°3 o 4), oligonucleótido, vector, célula central, planta transgénica, método, promotor (teniendo la secuencia específica SEQ ID N°13).
-Invención 3, reivindicaciones 1-39,41-47 parte
molécula de ácido nucleico teniendo una secuencia de codificación que codifica la secuencia como se indica en SEQ ID N°11 la cual codifica para un tercer gen deshidrina diferente llamado CcDH3, y materia derivada: mRNA, cDNA (SEQ ID N°5), oligonucleótido, vector, célula central, planta transgénica, método, promotor.
-Invención 4, reivindicación 1-32 parte
molécula de ácido nucleico teniendo una secuencia de codificación que codifica la secuencia como se indica en SEQ ID N°12 la cual codifica para gen LEA llamado CcLEA1, y materia derivada: mRNA, cDNA (SEQ ID N°6), oligonucleótido, vector, célula central, planta transgénica, método.

Ellas no están tan vinculadas como para formar un único concepto inventivo general (Regla 13.1 PCT) por las siguientes razones:

El problema técnico definido por la primera invención es la previsión de una molécula de ácido nucleico teniendo una secuencia de codificación codificando la secuencia como se indica en SEQ ID N°7 u 8 el cual codifica para dos variantes alélicas de una primer gen deshidrina llamado CcDH1A y CcDH1b, y materia derivada: mRNA, cDNA (SEQ ID N°1 o 2), oligonucleótido, vector, célula central, planta transgénica, método, promotor.

El problema técnico definido por la segunda invención es la provisión de una molécula de ácido nucleico teniendo una secuencia de codificación codificando la secuencia como se indica en SEQ ID N°9 o 10 las cuales son dos codificaciones unigenes distintas para un segundo gen deshidrina diferente llamado CcDH2a y CcDH2b, y materia derivada: mRNA, cDNA (SEQ ID N°3 o 4), oligonucleótido, vector, célula central, planta transgénica, método, promotor (teniendo la secuencia específica SEQ ID N°13).
El problema técnico definido por la tercera invención es la provisión de una molécula de ácido nucleico teniendo una secuencia de codificación codificando la secuencia como se indica en SEQ ID N°11 la cual codifica para un tercer gen deshidrina diferente llamado CcDH3, y materia derivada: mRNA, cDNA (SEQ ID N°5), oligonucleótido, vector, célula central, planta transgénica, método, promotor.
El problema técnico definido por la cuarta invención es la provisión de una molécula de ácido nucleico teniendo una secuencia de codificación codificando la secuencia como se indica en SEQ ID N°12 la cual codifica para gen LEA llamado CcLEA1, y materia derivada: mRNA, cDNA (SEQ ID N°6), oligonucleótido, vector, célula central, planta transgénica, método.

Las soluciones provistas son como arriba se definieron. Estas soluciones no resuelven un problema técnico en particular, y por lo tanto las cuatro invenciones definidas no son consideradas para ser unitarias a priori.
Los cuatro diferentes genes mencionados arriba no comparten ningún rasgo técnico particular: Las proteínas deshidrina y LEA pertenecen a grupos de proteínas acumuladoras de manera coordinada durante las últimas etapas del desarrollo de la semilla, de las cuales numerosos miembros son conocidos (ver por ejemplo D1 Kasuga et al, 2004 y descripción p.2-4). El rasgo "aislado del café" no es apropiado como a tal para constituir un rasgo técnico especial, puesto que dicho rasgo no es un rasgo técnico distintivo para una proteína aislada definida por su primera secuencia. Es más, el deshidrina como una proteína, ha sido identificado en el café (ver D2, Fernández, 2004, Tabla 2). Cada miembro de esta familia es por lo tanto considerado para constituir una invención separada.

En consecuencia, ni los problemas objetivos subrayando la materia de las invenciones reivindicadas, ni sus soluciones definidas por los rasgos técnicos especiales permiten que se establezca una relación entre las invenciones mencionadas las cuales involucran un único concepto inventivo general.

Re Ítem V
Afirmación razonada con respecto a la novedad, nivel inventivo o aplicación industrial; referencias y explicaciones que apoyan tal afirmación

I. INVENCIÓN 1

I.1. La presente solicitud no cumple los criterios del Artículo 33(1) PCT, puesto que el presente asunto de la reivindicación 10 no es nuevo ni inventivo en el sentido del Artículo 33(2) y (3) PCT.

D3 (ver referencias mencionadas en el reporte de búsqueda) revela varios oligonucleótidos claramente comprendidos dentro del ámbito del reivindicación 10.

I.2. La presente solicitud no cumple los criterios del Artículo 33(1) PCT, puesto que el presente asunto de las reivindicaciones 1-3, 5-9, 11-19, 20-24 no es nuevo ni inventivo en el sentido del Artículo 33(2) PCT y (3) PCT.

Un DNA o una secuencia de proteína no son novedosos simplemente porque éstas sean aisladas de orígenes diferentes comparados con el arte anterior el cual revela un DNA idéntico o una secuencia de proteína.

El documento D1 (Kasuga et al.) revela secuencias que tienen más del 60% de identidad al nivel del ácido amino, así como correspondientes secuencias de ácido nucleico con secuencias del así llamado CcDH1a y CcDH1b de la presente solicitud (SEQ ID N°7 u 8, SEQ ID N°1 y 2), anticipando por lo mismo el presente asunto de al menos las reivindicaciones 1-3, 5-9.
El documento D4 (WO03/027249) revela la secuencia del maíz DHN (como se indica adelante en SEQ ID N°34) teniendo más del 60% de identidad en el nivel del aminoácido, así como de la secuencia correspondiente del ácido nucleico con las secuencias de la así llamada CcDH1a y CcDH1b de la presente solicitud (SEQ ID N°7 y 8, SEQ ID N°1 Y 2), en consecuencia anticipando por lo tanto el presente asunto de por lo menos los reivindicaciones 1-3, 5-9. Vectores, células centrales, planta transgénica inter alia café son reveladas, en consecuencia anticipando el presente asunto de las reivindicaciones 11-18, 20-24.

I.3. La presente solicitud no cumple los criterios del Artículo 33(1) PCT, puesto que el presente asunto de las reivindicaciones 31-32 no es nuevo ni inventivo en el sentido del Artículo 33(2) PCT y (3) PCT.
El documento D5 (Kaye et al.) revela el método para incrementar resistencia al estrés en una planta comprendiendo incrementos de producción de una deshidrina en donde una proteína deshidrina (de espinaca) ha sido introducida dentro de una planta transgénica (una planta de tabaco). La resistencia verificada al estrés es particularmente más tolerante al frío. Sin embargo, es considerado que la resistencia osmótica al estrés es meramente una continuación fenotípica más característica de un sistema de tolerancia al estrés general. Por consiguiente, se considera que D5 anticipa el presente asunto de las reivindicaciones 31-32.
El documento D1 revela un método en donde la resistencia de la planta al estrés en el tabaco ha sido mejorado a través de la sobre-expresión del gen Dreb1a de Arabidopsis. En respuesta a la mencionada sobre-expresión, los genes objetivo inducibles por estrés fueron expresados fuertemente, dentro de ellos el gen deshidrina. Por lo tanto, el presente asunto de reivindicación 31 está claramente anticipado.

I.4. La presente solicitud no cumple los criterios del Artículo 33(1) PCT, puesto que el presente asunto de las reivindicaciones 25-30 no es inventivo en el sentido del Artículo 33(3) PCT: Las reivindicaciones mencionadas meramente relatan la materia obviamente derivada en vista del extensamente documentado estado del arte anterior con respecto a los deshidrinas (ver D1-D2, D4-D7, y literatura citada p. 1-4 de descripción). Podría ser obvio para las personas preparadas afrontar la existencia de las deshidrinas en el café así como en otras plantas, para manipular dicho gen vía ingeniería genética para efectos de ajustar su expresión. El ajuste de los elementos encontrados naturalmente es fácilmente afrontado para tener consecuencias en el aroma o el sabor dentro de las semillas.

I.5. La presente solicitud no cumple con los criterios del Artículo 33(1) PCT, puesto que el presente asunto de las reivindicaciones 33-39, 41-47 no es novedoso ni inventivo en el sentido del Artículo 33(2) y (3) PCT:

Dicho presente asunto (restringido a la invención 1) es relacionado con el promotor del gen CcDH1. Sin embargo, en virtud de la ausencia de cualquier definición por su secuencia, la persona calificada no está en condición para distinguir el promotor reivindicado de cualquier promotor presente en el arte anterior. Más particularmente, aun como se define el promotor con secuencias regulatorias específicas (ver reivindicación 39) es anticipado por los promotores conocidos de deshidrinas teniendo los mismos elementos regulatorios (ver por ejemplo los promotores descritos en D1). La mera definición de la secuencia regulada la cual corresponde del resultado a ser logrado, no puede de manera alguna permitirse para una definición de naturaleza técnica de dicho promotor.
Como consecuencia, el presente asunto de las reivindicaciones 41-47 las cuales son reivindicaciones relacionadas es también considerado de carecer de novedad y actividad inventiva.

I.6. El presente asunto de la reivindicación 4, revela la molécula del ácido nucleico codificando la proteína deshidrina consistente en la secuencia del aminoácido como se indica en SEQ ID N°7 u 8 y llamado CcDH1a y CcDH1b, no es descrito en el arte anterior. La novedad es en consecuencia reconocida (Artículo 33(2) PCT).

Esta molécula fue identificada de bases de datos de más de 47,000 etiquetas de secuencias expresadas (ESTs por sus siglas en ingles) de varios "Coffea canephora" (robusta) librerías cDNA hechas con RNA aislados de hojas jóvenes y del grano y tejidos del pericarpio de cerezas cosechadas en diferentes etapas de desarrollo, usando RT-PCR. La molécula cDNA de CcDH1 fue encontrada para ser expresada a un nivel relativamente alto. Las transcripciones CcDH1 fueron también detectadas en otros tejidos tales como el pericarpio, hojas y flores. La función fue asignada por homologias con deshidrinas conocidas, y motivos específicos estuvieron presentes. Debido a que la inducción de expresión es asociada con estrés osmótico en otras plantas, y puesto que la estructura de expresión de CcDH1 es específica y particular, tal información habilitará dicha proteína deshidrina en particular para ser genéticamente manipulada, con el propósito de mejorar el aroma y el sabor del café, así como conceder otras ventajas fenotipicas asociadas con el mejoramiento de la resistencia al estrés osmótico.
En consecuencia, puesto que la provisión de CcDH1 provee ventajas y efectos inesperados, la actividad inventiva es reconocida (Artículo 33(3) PCT).

II. INVENCIÓN 4

II.1. La presente solicitud no cumple con los criterios del Artículo 33(1) PCT, puesto que el presente asunto de las reivindicaciones 1-3, 5-9, 11-19, 20-24 no es nuevo ni inventivo en el sentido del Artículo 33(2) PCT y (3) PCT:

Un DNA o la secuencia de una proteína no son novedosos solamente porque ésta es aislada de un origen diferente comparado con el arte anterior, el cual revela un idéntico DNA o secuencia de proteína.

El documento D8 (Sato et al) revela secuencias teniendo más del 50% de identidad al nivel del aminoácido, así como a la correspondiente secuencia del ácido nucleico con las secuencias del así llamado CcLEA1 de la presente solicitud (SEQ ID N°6, SEQ ID N°12), en consecuencia anticipando de este modo el presente asunto de por lo menos las reivindicaciones 1-3, 5-9.

Los vectores y las células centrales son revelados, anticipando de esta manera el presente asunto de las reivindicaciones 11-18, 20-22.
Las plantas transgénicas como están reveladas en la reivindicación 23-24 no pueden ser consideradas para ser inventivas puesto que ellas aparecen por ser obviamente materia derivada (Artículo 33(3) PCT.

II.2. La presente solicitud no cumple con los criterios del Artículo 33(1) PCT, debido a que el presente asunto de las reivindicaciones 31-32 (restringidos a LEA) no es nuevo y/o no inventivo en el sentido del Artículo 33(2) PCT y (3) PCT.

El documento D1 revela un método en donde la resistencia al estrés de las plantas en tabacos ha sido mejorada a través de la sobreexpresión del gen Dreb1a de Arabidopsis. En respuesta de la mencionada sobreexpresión, los genes objetivo del estrés inducido fueron fuertemente expresados, entre ellos el gen LEA. Por lo tanto, el presente asunto de la reivindicación 31 está claramente anticipado.
Por las razones tal como están escritas arriba (§ I.3) concerniente a D5, parece que el presente asunto de la reivindicación 31-32 no puede ser considerado para ser inventivo puesto que la persona calificada en el arte puede obviamente aplicar el efecto del conocimiento concerniente de la deshidrina a la proteína LEA.

II.3. La presente solicitud no cumple con el criterio del Artículo 33(1) PCT, debido a que el presente asunto de las reivindicaciones 25-30 no es inventivo en el sentido del Artículo 33(2) PCT:

Dicha reivindicación meramente relata las materias obviamente derivadas en vista del extensamente documentado arte anterior con respecto de la deshidrina o LEA (ver D1-D2, D4-D7, y literatura citada p. 1-4 de descripción). Sería obvio para la persona calificada, enfrentar la existencia de LEA en café así como en otras plantas, para manipular dicho gen vía ingeniería genética con el propósito de modular sus expresiones. La modulación de los elementos naturalmente encontrados es fácilmente confrontable para tener consecuencias en el aroma o el sabor dentro de las semillas.

II.4. El presente asunto de reivindicación 4, revelando la molécula de ácido nucleico poniendo en código la proteína LEA consistente en la secuencia del aminoácido como se indica en SEQ ID N°12 y llamada CcLEA, no está descrito en el arte anterior. Lo mismo aplica para el presente asunto de reivindicación 10 (restringido a SEQ ID N°6). La novedad es consecuentemente reconocida (Art. 33(2) PCT).

La molécula de ácido nucleico fue identificada de bases de datos de más de 47,000 etiquetas de secuencias expresadas (ESTs por sus siglas en ingles) de varias "Coffea canephora" (robusta) librerías cDNA hechas con RNA aislados de hojas jóvenes y del grano y tejidos del pericarpio de cerezas cosechadas en diferentes etapas de desarrollo, usando RT-PCR. La molécula cDNA de CcLEA fue encontrada para ser expresada durante solo un período del desarrollo del grano (30 semanas después de florecidas). La función fue asignada por homologias con una LEA conocida. Desde que el patrón de expresión de CcLEA es específico y particular, tal información permitirá a dicha proteína LEA en particular ser genéticamente manipulada, con el propósito de mejorar el aroma y el sabor del café, así como conceder otras ventajas fenotipicas asociadas con el mejoramiento de la resistencia al estrés osmótico.
En consecuencia, desde que la provisión de CcLEA provee ventajas y efectos inesperados, la actividad inventiva es reconocida (Art. 33(3) PCT).

Re Ítem VI

Ciertos documentos citados:

D9: WO 2005/122697 A (CARMEL HAIFA UNIVERSITY ECONOM [ILLINOIS]; NEVO EVIATAR [ILLINOIS]; SONG WEINING [) 29 de Dicembre de 2005 (2005-12-29)

Este documento revela la proteína deshidrina de la cebada llamada DHN1, teniendo más de 60% de identidad en el nivel de secuencia del amino ácido con CcDH1 como se revela en la presente solicitud con SEQ ID N°7 u 8, planta transgénica transformada con el gen DHN1 teniendo una mejor resistencia a la sequedad que el tipo salvaje.

PATENT COOPERATION TREATY

OPIN 20.09.07

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)
20/06/07

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2006/026234	05.07.2006	06.07.2005

International Patent Classification (IPC) or both national classification and IPC
INV. C07K14/415 C12N15/82 C12N5/04

Applicant
CORNELL RESEARCH FOUNDATION

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☒ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of [redacted] from the date of mailing of Form PCT/ISA/220 or before the expiration of [redacted] date, whichever expires later.

06.07.05 20.09.07
06.05.07

For further options, see Form PCT/ISA/220.

3. For further details, see notes to Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	see form PCT/ISA/210	LOUBRADOU-BOURGES, N Telephone No. +49 89 2399-7342

Form (PCT/ISA/237) (Cover Sheet) (April 2005)

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application and necessary to the claimed invention, this opinion has been established on the basis of:

   a. type of material:

      ☒ a sequence listing

      ☐ table(s) related to the sequence listing

   b. format of material:

      ☒ on paper

      ☒ in electronic form

   c. time of filing/furnishing:

      ☒ contained in the international application as filed.

      ☒ filed together with the international application in electronic form.

      ☐ furnished subsequently to this Authority for the purposes of search.

3. ☒ In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

4. Additional comments:

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of

☐ the entire international application

☒ claims Nos. 1-39,41-47 part, 40 all (lack of unity)

because:

☐ the said international application, or the said claims Nos. relate to the following subject matter which does not require an international search *(specify)*:

☐ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. are so unclear that no meaningful opinion could be formed *(specify)*:

☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no international search report has been established for the whole application or for said claims Nos. 1-39,41-47 part, 40 all (lack of unity)

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not, within the prescribed time limit:

- ☐ furnish a sequence listing on paper complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in a form and manner acceptable to it.
- ☐ furnish a sequence listing in electronic form complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in a form and manner acceptable to it.
- ☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rules 13*ter*.1(a) or (b).

☐ a meaningful opinion could not be formed without the tables related to the sequence listings; the applicant did not, within the prescribed time limit, furnish such tables in electronic form complying with the technical requirements provided for in Annex C-*bis* of the Administrative Instructions, and such tables were not available to the International Searching Authority in a form and manner acceptable to it.

☐ the tables related to the nucleotide and/or amino acid sequence listing, if in electronic form only, do not comply with the technical requirements provided for in Annex C-*bis* of the Administrative Instructions.

☐ See Supplemental Box for further details

Box No. IV Lack of unity of invention

1. ☒ In response to the invitation (Form PCT/ISA/206) to pay additional fees, the applicant has, within the applicable time limit:

   ☒ paid additional fees

   ☐ paid additional fees under protest and, where applicable, the protest fee

   ☐ paid additional fees under protest but the applicable protest fee was not paid

   ☐ not paid additional fees

2. ☐ This Authority found that the requirement of unity of invention is not complied with and chose not to invite the applicant to pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rule 13.1, 13.2 and 13.3 is

   ☐ complied with

   ☒ not complied with for the following reasons:

   see separate sheet

4. Consequently, this report has been established in respect of the following parts of the international application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. 1-39,41-47 part

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	4,25-30
	No:	Claims	1-3,5-24,31-39,41-47
Inventive step (IS)	Yes:	Claims	4
	No:	Claims	1-3,5-39,41-47
Industrial applicability (IA)	Yes:	Claims	1-39,41-47
	No:	Claims	

2. Citations and explanations

see separate sheet

Box No. VI Certain documents cited

1. Certain published documents (Rules 43*bis*.1 and 70.10)

   and / or

2. Non-written disclosures (Rules 43*bis*.1 and 70.9)

   see form 210

Q52

Reference is made to the following documents:

D1: KASUGA MIE ET AL: "A combination of the Arabidopsis DREB1A gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer." PLANT & CELL PHYSIOLOGY MAR 2004, vol. 45, no. 3, March 2004 (2004-03), pages 346-350, XP002421281 ISSN: 0032-0781

D2: FERNANDEZ DIANA ET AL: "Coffee (Coffea arabica L.) genes early expressed during infection by the rust fungus (Hemileia vastatrix)" MOLECULAR PLANT PATHOLOGY, vol. 5, no. 6, November 2004 (2004-11), pages 527-536, XP002421282 ISSN: 1464-6722

D3: WO 01/77384 A (EPIGENOMICS AG [DE]; OLEK ALEXANDER [DE]; PIEPENBROCK CHRISTIAN [DE];) 18 October 2001 (2001-10-18)

D4: WO 03/027249 A (SYNGENTA PARTICIPATIONS AG [CH]; SU WENPEI [US]; ANDON NANCY [US]; HAY) 3 April 2003 (2003-04-03)

D5: KAYE CLAUDIA ET AL: "Characterization of a gene for spinach CAP160 and expression of two spinach cold-acclimation proteins in tobacco" PLANT PHYSIOLOGY (ROCKVILLE), vol. 116, no. 4, April 1998 (1998-04), pages 1367-1377, XP002421536 ISSN: 0032-0889

D6: ALLAGULOVA CH R ET AL: "The plant dehydrins: structure and putative functions." BIOCHEMISTRY. BIOKHIMII A SEP 2003, vol. 68, no. 9, September 2003 (2003-09), pages 945-951, XP009079176 ISSN: 0006-2979

D7: WISE MICHAEL J ET AL: "POPP the question: What do LEA proteins do?" TRENDS IN PLANT SCIENCE, vol. 9, no. 1, January 2004 (2004-01), pages 13-17, XP002421537 ISSN: 1360-1385

D8 : SATOS ET AL: "Structural analysis of arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and TAC clones" GENBANK, 28 April 2000 (2000-04-28), XP002974277

Re Item III.

The present opinion is restricted to invention 1 and invention 4 as below defined which have been subjected to search.

Re Item IV.

The separate inventions/groups of inventions are:

- invention 1, claims 1-39,41-47 part
nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°7 or 8 which code for two allelic variants of a first dehydrin gene named CcDH1a and CcDH1b, and derived matter : mRNA, cDNA (SEQ ID N°1 or 2), oligonucleotide, vector, host cell, transgenic plant, method, promoter.
- invention 2, claims 1-39,41-47 part ; 40 all
nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°9 or 10 which are two distinct unigenes coding for a second different dehydrin gene named CcDH2a and CcDH2b, and derived matter : mRNA, cDNA (SEQ ID N°3 or 4), oligonucleotide, vector, host cell, transgenic plant, method, promoter (having the specific sequence SEQ ID N°13).
- invention 3, claims 1-39,41-47 part
nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°11 which codes for a third different dehydrin gene named CcDH3, and derived matter : mRNA, cDNA (SEQ ID N°5), oligonucleotide, vector, host cell, transgenic plant, method, promoter.
- invention 4, claims 1-32 part
nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°12 which codes for a LEA gene named CcLEA1, and derived matter : mRNA, cDNA (SEQ ID N°6), oligonucleotide, vector, host cell, transgenic plant, method.

They are not so linked as to form a single general inventive concept (Rule 13.1 PCT) for the following reasons:

The technical problem defined by the first invention is the provision of a nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°7 or 8 which code for two allelic variants of a first dehydrin gene named CcDH1a and CcDH1b, and derived matter : mRNA, cDNA (SEQ ID N°1 or 2), oligonucleotide, vector, host cell, transgenic plant, method, promoter.

The technical problem defined by the second invention is the provision of a nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°9 or 10 which are two distinct unigenes coding for a second different dehydrin gene named CcDH2a and CcDH2b, and derived matter : mRNA, cDNA (SEQ ID N°3 or 4), oligonucleotide, vector, host cell, transgenic plant, method, promoter (having the specific sequence SEQ ID N°13).
The technical problem defined by the third invention is the provision of a nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°11 which codes for a third different dehydrin gene named CcDH3, and derived matter : mRNA, cDNA (SEQ ID N°5), oligonucleotide, vector, host cell, transgenic plant, method, promoter.
The technical problem defined by the fourth invention is the provision of a nucleic acid molecule having a coding sequence encoding the sequence as set forth in SEQ ID N°12 which codes for a LEA gene named CcLEA1, and derived matter : mRNA, cDNA (SEQ ID N°6), oligonucleotide, vector, host cell, transgenic plant, method.

The solutions provided are as above defined. These solutions do not solve a single technical problem, and therefore the four defined inventions are not considered to be unitary a priori.
The four above mentioned different genes do not share any special technical features : the dehydrins and LEA proteins belong to a group of proteins accumulating in a coordinate fashion during the latter stages of seed development, from which numerous members are known (see for example D1 Kasuga et al, 2004 and description p.2-4). The feature "isolated from coffee" is not suitable as such for constituting a special technical feature, since said feature is not a distinguishing technical feature for an isolated protein defined by its primary sequence. Moreover, a dehydrin like protein has been identified in coffee (see D2, Fernandez, 2004, Table 2). Each member of this family is therefore considered to constitute a separate invention.

Consequently, neither the objective problems underlying the subject of the claimed inventions, nor their solutions defined by the special technical features allow for a relationship to be established between the said inventions which involves a single general inventive concept.

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

I. INVENTION 1

I.1. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claim 10 is not new and not inventive in the sense of Article 33(2) and (3) PCT.

D3 (see references mentioned in the search report) discloses several oligonucleotides clearly falling onto the scope of claim 10.

I.2. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 1-3, 5-9, 11-19, 20-24 is not new and not inventive in the sense of Article 33(2) PCT and (3) PCT.

A DNA or a protein sequence is not novel simply because it is isolated from a different origin compared to the prior art which discloses the identical DNA or protein sequence.

The document D1 (Kasuga et al.) discloses sequences having more than 60 % identity at the amino acid level, as well as corresponding nucleic acid sequences with the sequences of the so-called CcDH1a and CcDH1b of the present application (SEQ ID N°7 or 8, SEQ ID N°1 and 2), thereby anticipating the subject-matter of at least claims 1-3,5-9.
The document D4 (WO03/027249) discloses the maize DHN1 sequence (as set forth in SEQ ID N°34) having more than 60 % identity at the amino acid level, as well as corresponding nucleic acid sequences with the sequences of the so-called CcDH1a and CcDH1b of the present application (SEQ ID N°7 or 8, SEQ ID N°1 and 2),

thereby anticipating the subject-matter of at least claims 1-3,5-9. Vectors, host cells, transgenic plant inter alia coffee are disclosed, thereby anticipating the subject-matter of claims 11-18, 20-24.

I.3. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 31-32 is not new and not inventive in the sense of Article 33(2) PCT and (3) PCT.
The document D5 (Kaye et al.) discloses method to increase resistance to stress in a plant comprising increasing production of a dehydrin wherein a dehydrin protein (from spinach) has been introduced within a transgenic plant (a tobacco plant). The stress resistance checked is more particularly cold tolerance. However, it is considered that the osmotic stress resistance is merely a further phenotypic trait of the more general stress tolerance system. Thereby, it is considered that D5 anticipates the subject-matter of claims 31-32.
The document D1 discloses a method wherein plant stress resistance in tobacco has been improved through the overexpression of the Dreb1a gene from Arabidopsis. In response of said overexpression, target stress-inducible genes were strongly expressed, among them a dehydrin gene. Thus, the subject-matter of claim 31 is clearly anticipated.

I.4. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 25-30 is not inventive in the sense of Article 33(3) PCT:
Said claims merely relate to obvious derived matter in view of the largely documented prior art with respect to dehydrins (see D1-D2, D4-D7, and literature cited p.1-4 of description). It would be obvious for the skilled person, envisaging the existence of dehydrins in coffee as well as in other plants, to manipulate said gene via genetic engineering in order to modulate its expression. The modulation of elements naturally found is easily envisaged to have consequences on aroma or flavour within seeds.

I.5. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 33-39, 41-47 is not novel and not inventive in the sense of Article 33(2) and (3) PCT:
Sais subject-matter (restricted to invention 1) relates to the promoter of the gene CcDH1. However, in view of the absence of any definition by its sequence, the skilled

person in not in position to distinguish the claimed promoter form any promoter present in the prior art. More particularly, the promoter even as defined with specific regulatory sequences (see claim 39) is anticipated by the known promoters of dehydrins having the same regulatory elements (see for example promoters described in D1). The mere definition of the regulated sequence which corresponds of the result to be achieved can in no way allow for a definition of a technical nature of said promoter.
As a consequence, the subject-matter of claims 41-47 which are related claims is also considered to lack novelty and inventive activity.

I.6. The subject-matter of claim 4, disclosing the nucleic acid molecule encoding the dehydrin protein consisting of the amino acid sequence as set forth in SEQ ID N°7 or 8 and called CcDH1a and CcDH1b is not described in the prior art. Novelty is therefore acknowledged (Art.33(2) PCT).

This molecule was identified from databases of over 47,000 expressed sequence tags (ESTs) from several Coffea canephora (robusta) cDNA libraries made with RNA isolated from young leaves and from the grain and pericarp tissues of cherries harvested at different stages of development, using RT-PCR. The cDNA molecule of CcDH1 was found to be expressed at a relatively high level. CcDH1 transcripts were also detected in other tissues such as pericarp, leaves and flowers. Function was assigned by homologies with known dehydrins, and specific motives were present. Because induction of expression is associated with osmotic stress in other plants, and since expression pattern of CcDH1 is specific and particular, such information will enable said particular dehydrin protein to be genetically manipulated, with the goal of improving the aroma and flavour of the coffee, as well as imparting other phenotypic advantages associated with improved osmotic stress resistance. Consequently, since the provision of CcDH1 provides advantages and unexpected effects, inventive activity is recognized (Art.33(3) PCT).

II. INVENTION 4

II.1. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 1-3, 5-9, 11-19, 20-24 is not new and not inventive in the sense of Article 33(2) PCT and (3) PCT.

A DNA or a protein sequence is not novel simply because it is isolated from a different origin compared to the prior art which discloses the identical DNA or protein sequence.

The document D8 (Sato et al.) discloses sequences having more than 50 % identity at the amino acid level, as well as corresponding nucleic acid sequences with the sequences of the so-called CcLEA1 of the present application (SEQ ID N°6, SEQ ID N°12), thereby anticipating the subject-matter of at least claims 1-3,5-9.

Vectors, host cells are disclosed, thereby anticipating the subject-matter of claims 11-18, 20-22.
Transgenic plants as disclosed in claim 23-24 cannot be considered to be inventive since they appear to be obviously derived matter (Art.33(3) PCT.

II.2. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 31-32 (restricted to LEA) is not new and/or not inventive in the sense of Article 33(2) PCT and (3) PCT.

The document D1 discloses a method wherein plant stress resistance in tobacco has been improved through the overexpression of the Dreb1a gene from Arabidopsis. In response of said overexpression, target stress-inducible genes were strongly expressed, among them a LEA gene. Thus, the subject-matter of claim 31 is clearly anticipated.
For the reasons as written above (§ I.3) concerning D5, it appears that the subject-matter of claims 31-32 cannot be considered to be inventive since the skilled person in the art would obviously apply the knowledge concerning effect of the dehydrins to LEA proteins.

II.3. The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 25-30 is not inventive in the sense of Article 33(3) PCT:

Said claims merely relate to obvious derived matter in view of the largely documented prior art with respect to dehydrins or LEA (see D1-D2, D4-D7, and literature cited p.1-4 of description). It would be obvious for the skilled person, envisaging the existence of LEA in coffee as well as in other plants, to manipulate said gene via genetic engineering in order to modulate its expression. The modulation of elements naturally found is easily envisaged to have consequences on aroma or flavour within seeds.

II.4. The subject-matter of claim 4, disclosing the nucleic acid molecule encoding the LEA protein consisting of the amino acid sequence as set forth in SEQ ID N°12 and called CcLEA is not described in the prior art. The same applies for the subject-matter of claim 10 (restricted to SEQ ID N°6). Novelty is therefore acknowledged (Art.33(2) PCT).

The nucleic acid molecule was identified from databases of over 47,000 expressed sequence tags (ESTs) from several Coffea canephora (robusta) cDNA libraries made with RNA isolated from young leaves and from the grain and pericarp tissues of cherries harvested at different stages of development, using RT-PCR. The cDNA molecule of CcLEA was found to be expressed during only one period of grain development (30 weeks after flowering). Function was assigned by homologies with known LEA. Since expression pattern of CcLEA is specific and particular, such information will enable said particular LEA protein to be genetically manipulated, with the goal of improving the aroma and flavour of the coffee, as well as imparting other phenotypic advantages associated with improved osmotic stress resistance. Consequently, since the provision of CcLEA provides advantages and unexpected effects, inventive activity is recognized (Art.33(3) PCT).

Re Item VI.

Certain cited document:

D9: WO 2005/122697 A (CARMEL HAIFA UNIVERSITY ECONOM [IL]; NEVO EVIATAR [IL]; SONG WEINING [) 29 December 2005 (2005-12-29)

This document discloses dehydrin protein from barley named DHN1, having more than 60% identity at the amino acid sequence level with CcDH1 as disclosed in the present application with SEQ ID N°7 or 8, transgenic plant transformed with DHN1 gene having an improved drought resistance than wild type.

GENES PARA DESHIDRINA Y PROMOTORES DEL CAFÉ

RESUMEN-OBJETO Y FINALIDAD DE LA INVENCIÓN

Se describen polinucleótidos que codifican para deshidrina -una proteína abundante embriogénica tardía (LEA)- de plantas de café. También se describen una secuencia promotora de un gen para deshidrina de café y métodos para utilizar estos polinucleótidos y secuencias promotoras para regulación del gen y manipulación del sabor, aroma, tolerancia a la tensión y otras características de los granos de café.

GENES PARA DESHIDRINA Y PROMOTORES DEL CAFÉ

CAMPO DE LA INVENCIÓN

La presente invención se relaciona con el campo de biotecnología agrícola. En particular, la invención describe polinucleótidos que codifican para deshidrina a partir de plantas de café, secuencias promotoras a partir de genes para deshidrina de café y métodos para utilizar estos polinucleótidos promotores para regulación de genes y manipulación de sabor, aroma y otras características de los granos de café.

ANTECEDENTES DE LA INVENCIÓN

Las diversas publicaciones, que incluyen patentes, solicitudes publicadas y artículos escolares se mencionan a través de la especificación. Cada una de estas publicaciones se incorporan como referencia en la presente en su totalidad. Las citas que no se establecen completamente dentro de la especificación se pueden encontrar al final de la especificación.

El aroma y el sabor del café son componentes claves en la preferencia del consumidor por variedades y marcas de café. El aroma y el sabor característicos del café se basan en una serie compleja de reacciones químicas que involucran precursores de sabor (reacciones de

Maillard) que se producen durante el tostado del grano. Los precursores de sabor incluyen compuestos químicos y biomoléculas presentes en el grano de café verde. Hasta ahora se han identificado más de 800 sustancias químicas y biomoléculas que contribuyen al sabor y aroma del café (Montavon *et al.*, J. Agric. Food Chem., 51:2328-34 (2003)).

Debido a que los consumidores de café se vuelven cada vez más sofisticados, es deseable producir café con aroma y sabor mejorados con el fin de satisfacer las preferencias del consumidor. Tanto el aroma como el sabor se pueden impartir artificialmente en productos de café por medios químicos. Véanse, por ejemplo, la patente de E.U.A. No. 4,072,761 (aroma) y la patente de E.U.A. No. 3,962,321 (sabor). No obstante, hasta ahora ha habido pocos datos respecto a la influencia de los componentes de grano de café natural tales como polisacáridos, proteínas y lípidos en el aroma y sabor del café. Una solución es seleccionar variedades con el plasma germinal existente que tengan características de sabor superiores. Una desventeja de esta solución es que, con frecuencia, las variedades de calidad más alta también producen rasgos agronómicos negativos significativos tales como un bajo rendimiento y poca resistencia a enfermedades y tensiones ambientales. También es posible seleccionar variedades nuevas a partir de ensayos de crianza en las cuales las variedades con rasgos

industriales y agronómicos diferentes se cruzan y su progenie se analiza para buscar una alta calidad y buen desempeño agronómico. No obstante, esta última solución consume mucho tiempo, en donde un experimento de cruz y selección durante tres estaciones de crecimiento requiere un mínimo de 7-8 años. Por lo tanto, una solución alternativa para mejorar la calidad del café sería el utilizar técnicas de biología molecular para mejorar aquellos elementos responsables del sabor y aroma que se encuentran naturalmente en los granos de café y agregar elementos que mejoren el aroma y el sabor y que no se encuentren de manera natural en los granos de café. La manipulación genética es particularmente adecuada para obtener estos fines. Por ejemplo, las proteínas de café de diferentes especies de café se pueden barrer. En la alternativa, la expresión de genes que codifican para proteínas de café que se encuentran de manera natural y que contribuyen positivamente al sabor del café se pueden mejorar. Inversamente, se puede suprimir la expresión de genes que codifiquen para proteínas de café que se presentan de manera natural pero que perjudican el sabor del café. Otra aplicación de las técnicas modernas es utilizar información molecular respecto a la asociación de alta calidad con alelos específicos para la obtención de variedades nuevas para determinar la presencia o ausencia

de los mismos bajo crianza asistida por marcadores.

Los cafés de diferentes variedades y orígenes presentan variaciones en la calidad del sabor y aroma significativas cuando las muestras de granos verdes se tuestan y se procesan de la misma manera. Las diferencias en calidad son una manifestación de las variaciones químicas y físicas dentro de las muestras de grano que resultan principalmente de diferencias en las condiciones de crecimiento y procesamiento y también a partir de diferencias en los antecedentes genéticos tanto de la planta materna como del grano. A nivel de composición química, por lo menos parte de la calidad del sabor se puede asociar con variaciones en las concentraciones de metabolitos pequeños tales como azúcares, ácidos, sustancias fenólicas y cafeína que se encuentra asociada con el grano de variedades diferentes. Se acepta que existen otras moléculas menos caracterizadas de sabor y precursores de sabor. Además, es probable que las variaciones estructurales dentro del grano probablemente también contribuyan a las diferencias en la calidad del café. Una solución para encontrar componentes nuevos en los granos de café relacionados con la calidad del café es estudiar los genes y proteínas que se expresan de manera diferencial durante la maduración de muestras de grano en diferentes variedades que poseen características de calidad

diferentes.

Se ha demostrado que un grupo de proteínas denominadas proteínas abundantes de embriogénesis tardía (LEA) se acumulan de una manera coordinada durante las últimas etapas del desarrollo de la semilla de algodón (Dure, L, *et al*. Biochemistry 20: 4162-4178 (1981)). Las proteínas deshidrina (DHN) son un subgrupo de las proteínas LEA que también se les ha denominado la familia "LEA D-11" o proteínas LEA tipo 2 (Close, T, Physiol. Plant 97:795-803 (1996); Ingram, J. Annu. Rev. Plant Physiology Plant Mol Biol 47: 377-403 (1996)). La expresión de las proteínas DHN se ha relacionado con la protección de diversos tipos de células vegetales de tensiones osmóticas, tales como aquellas causadas por la desecación, concentración de sal y baja temperatura (Skriver, K. *et al*. Plant Cell 2: 503-512 (1990); Allagulova, CR, *et al*. Biochemistry-Moscú 68: 945-951 (2003)).

En años recientes las pruebas experimentales directas han relacionado la expresión aumentada de deshidrinas con la protección de tensión osmótica. Por ejemplo, las plantas de *Arabidopsis* manipuladas para sobreexpresar una proteína de fusión de deshidrina se ha encontrado que tienen una supervivencia mejorada cuando se exponen a baja temperatura (Puhakainen, T, *et al*. Plant Molecular Biology 54: 743-753 (2004)). De manera similar,

la expresión de una proteína de deshidrina cítrica en tabaco transgénico ha demostrado que proporciona tolerancia aumentada a baja temperatura (Hara, M, *et al*. Planta 217: 290-298 (2003)). Otras pruebas de soporte para el enlace deshidrinas y tolerancia a tensiones inducidas por baja temperatura son las observaciones de que los loci QTL para tolerancia al congelamiento y la resistencia al invierno se encuentran en el mapa muy cercanos a las deshidrinas (Close, T, 1996; Zhu, B, *et al*. Molecular and General Genetics 264: 145-153 (2000)). Los genes para DHN también se expresan de manera robusta en semillas hacia el final de la maduración, un período cuando la semilla experimenta una reducción programada por el desarrollo en el contenido de agua (Nylander M. *et al*. Plant Molecular Biology 45: 263-279 (2001); Choi, DW, *et al*. Theoretical and Applied Genetics 100: 1274-1278 (2000)). Las proteínas LEA/deshidrina se ha calculado que comprenden hasta 4% de la proteína de semilla total y se considera que están involucrados en la protección del embrión u otros tejidos de la semilla de las tensiones osmóticas relacionadas con un bajo contenido de agua de la semilla madura (Roberts, J, *et al*. Plant Cell 5: 769-780 (1993); Wise, M, *et al*. Trends Plant Sci. 9: 13-17 (2004)).

Las deshidrinas se perciben ampliamente que participan, con otras proteínas LEA, en el proceso de

deshidratación que se produce durante las etapas tardías de la maduración de la semilla por ayuda de la aclimatación de los tejidos de semillas a un contenido de agua menor encontrado en semillas maduras (Close, Tm 1996); Nylander M, 2001). Además, se considera que las semillas sintetizan deshidrinas durante la maduración y también continúan estabilizando las estructuras celulares asociadas durante el período de reposo de la semilla. En este último contexto, recientemente se ha propuesto que las deshidrinas también pueden tener una capacidad eliminadora de radicales (Hara, M, 2003) y propiedades de unión de metal (Alsheikh, MK, *et al.* J. Biol. Chem. 278: 40882-40889 (2003)), es probable que ambas características sean útiles durante períodos largos de almacenamiento de semilla.

Se han aislado y estudiado un número considerable de proteínas deshidrina y se encuentran bajo investigación los mecanismos fisicoquímicos o estructurales precisos y las proteínas funcionan para proteger a las células de tensión osmótica *in vivo*. Las deshidrinas son proteínas muy hidrofílicas y presentan una concentración inusualmente baja de estructura reconocible (Close T, 1996; Soulages, JL, *et al.* Plant Physiology 131: 963-975 (2003)). Un elemento clave de las deshidrinas se considera que es la presencia de cadenas ricas en lisina de uno o más de 15 aminoácidos, denominadas los "motivos K" los cuales se

predice que forman las hélices α-amfifáticas clase A (Close, T, 1996; Close, TJ Physiol. Plant 100: 291-296 (1997)). Las deshidrinas también contienen otros dos motivos, un "segmento Y" en la parte N terminal (consenso V/TDE/QYGNP) y un "segmento S" rico en serina, este último el cual puede ser fosforilado y se considera que participa en la localización nuclear (Close TJ, 1997; Godoy, JA, *et al*. Plant Mol. Biol. 1921-1934 (1994)). Se ha propuesto que los segmentos K anfipáticos cortos de polipéptidos de deshidrina interactúan funcionalmente con los parches hidrofóbicos expuestos a solvente de las proteínas que experimentan desnaturalización parcial y de esta manera bloquean la formación de agregado de proteína (Close T, 1996). Las hélices K anfipáticas también se han involucrado en la unión de lípidos de membrana y por lo tanto pueden jugar un papel más específico en proteger a las lipoproteínas, proteínas localizadas en membranas o a la estructura de la membrana misma (Close T, 1996; Koag, MC, *et al*. Plant Physiology 131: 309-316 (2003)). Una propuesta alternativa para al menos parte del efecto protector de las deshidrinas es la capacidad de que estas proteínas muy estables, pero relativamente no estructuradas se unan firmemente y organicen moléculas de agua (Soulages JL, 2003). Este último efecto puede llevar a una pérdida reducida de agua de las células y también puede mejorar la

estabilidad de ciertas macromoléculas por el desarrollo de la región basada en deshidrina de agua "ordenada" unida firmemente alrededor de estas moléculas.

Pese a la relación de las proteínas de deshidrina en la resistencia de la planta a tensiones osmóticas tales como sequía y tensión por sal y la probable importancia de las deshidrinas durante el desarrollo del grano, está disponible poca información respecto a estos genes en el café. En el café se entiende poco acerca del número de deshidrinas, su estructura proteínica, sus niveles de expresión y la distribución en tejidos diferentes de la planta de café y entre las especies de café así como durante la maduración del grano y pericarpio de café y la regulación de su expresión a nivel molecular. Por lo tanto, existe la necesidad de identificar, aislar y caracterizar proteínas de deshidrina de café, genes y elementos reguladores genéticos. Dicha información permitirá que las proteínas de deshidrina de café sean manipuladas genéticamente con el objetivo de mejorar el aroma y el sabor del café e igualmente para impartir otras ventajas fenotípicas relacionadas con resistencia a la tensión osmótica mejorada.

Las deshidrinas, las cuales se expresan a concentraciones relativamente altas al final de la maduración del grano son de interés debido a los papeles

potencialmente importantes que tienen el organizar moléculas de agua en el grano de café y en estabilizar macromoléculas y organelos dentro del grano deshidratado maduro. Por lo menos parte de este efecto protector se considera que se debe a la capacidad de las deshidrinas y otras proteínas LEA para estabilizar diferentes límites de agua/proteína/lípidos. Debido a que las concentraciones de agua pueden influir en el espectro de productos formados en la reacción de Maillard (Turner, J, *et al*. J. Agric. Food Chem 50: 5400-5404 (2002)), la capacidad y organización de las moléculas de agua en el grano de café pueden influir en el sabor que genera la reacción de Maillard que se presenta durante el tostado del café.

DESCRIPCIÓN BREVE DE LA INVENCIÓN

La invención descrita en la presente menciona polinucleótidos que codifican para deshidrina a partir de plantas de café, sus polipéptidos codificados, secuencias promotoras a partir de genes para deshidrina de café y métodos para utilizar estos polinucleótidos, polipéptidos y promotores para la regulación y manipulación de genes para el sabor, aroma y otras características de los granos de café.

Un aspecto de la invención describe una molécula de ácido nucleico aislada de café (género *Coffea*) que tiene

una secuencia codificante que codifica para una deshidrina o una proteína abundante embriogénica tardía (LEA). En algunas modalidades, la deshidrina codificada tiene un peso molecular de entre aproximadamente 17 kDa y aproximadamente 26 kDa. En algunas modalidades, la deshidrina codificada o la proteína LEA tiene una secuencia de aminoácidos que es por lo menos 46% o más idéntica a cualquiera de las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 7-12. En algunas modalidades, la secuencia codificante es 45% o más idéntica con cualquiera de las secuencias codificantes que se establecen en las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 1-6.

En algunas modalidades, la molécula de ácido nucleico es un gen que tiene un marco de lectura abierto que comprende la secuencia codificante. De manera alternativa, puede comprender una molécula de ARNm producida por transcripción de ése gen o una molécula de ADNc producida por transcripción inversa de la molécula de ARN. La invención también describe un oligonucleótido de entre 8 y 100 bases de longitud el cual es complementario a un segmento de la molécula de ácido nucleico mencionada antes.

Otro aspecto de la invención describe un vector que comprende la molécula de ácido nucleico que codifica para deshidrina o para LEA descrita antes. En algunas

modalidades, el vector es un vector de expresión que se selecciona del grupo de vectores que consisten de plásmido, fagémido, cósmido, baculovirus, bácmido, vectores bacterianos, de levaduras y virales. En algunas modalidades, el vector contiene la secuencia codificante de la molécula de ácido unida operablemente a un promotor constitutivo. En otras modalidades, la secuencia codificante está unida operablemente a un promotor inducible. En otras modalidades, la secuencia modificante de la molécula de ácido nucleico está unida operablemente a un promotor específico de tejido, tal como un promotor específico de semillas, preferiblemente un promotor específico de semilla de café. En las modalidades específicas, el promotor específico de semilla es un promotor para el gen de deshidrina de café tal como el promotor contenido en la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 13.

De acuerdo con otro aspecto de la invención, se proporciona una célula hospedera transformada con el vector transformado antes. La célula hospedera puede ser una célula vegetal, bacteriana, micótica, de insecto o de mamífero. En algunas modalidades, la célula hospedera es una célula vegetal que se selecciona de cualquiera de café, tabaco, Arabidopsis, maíz, trigo, arroz, soya, cebada, centeno, avena, sorgo, alfalfa, trébol, canóla, azafrán,

girasol, cacahuate, cacao, tomate tomatillo, papa, pimienta, berenjena, remolacha, zanahoria, pepino, lechuga, chícharo, aster, begonia, crisantemo, delfinio, zinia y césped. La invención también describe una planta transgénica fértil producida al regenerar la célula vegetal transformada. En una modalidad específica, la planta transgénica fértil es de la especie *Coffea*.

Otro aspecto de la invención describe un método para modular sabor o aroma de granos de café. El método comprende modular la producción de una o más deshidrinas o proteínas LEA con semillas de café. En algunas modalidades, el método comprende implementar la producción de una o más deshidrinas o proteínas LEA, por ejemplo al incrementar la expresión de uno o más genes que codifican para deshidrina o proteína LEA, endógenos dentro de las semillas de café o al introducir un transgen que codifica para deshidrina o proteína LEA en la planta. En otras modalidades, el método comprende disminuir la producción de una o más deshidrinas o proteínas LEA, por ejemplo, al introducir una molécula de ácido nucleico en el café que inhibe la expresión de uno o más de los genes que codifican para deshidrina o proteína LEA.

Otro aspecto de la invención describe un método para incrementar la resistencia a tensión osmótica en una planta. Este método comprende incrementar la producción de

una o más proteínas deshidrinas o LEA dentro de la planta, por ejemplo, al introducir un transgen que codifica para deshidrina o proteína LEA dentro de la planta.

De acuerdo con otro aspecto de la invención, se proporciona un promotor aislado a partir del gen de planta de café que codifica para deshidrina. En algunas modalidades, el gen de café que codifica para deshidrina codifica para una proteína deshidrina que tiene una o más de las características descritas en lo anterior. En algunas modalidades, el promotor comprende una o más secuencias reguladoras que se seleccionan del grupo que consiste de una caja TATA, un elemento sensible a ácido abscísico, una secuencia repetida RY (CATGCA(T/a)(A/g) de una caja de leguminina para regular la expresión de proteínas tipo leguminina por lo menos un elemento sensible a deshidratación/motivo de secuencia de acción cis repetido C (G/ACCGAC y por lo menos un motivo de caja E (CANNTG). En una modalidad específica, el promotor comprende la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 13.

La invención también describe un gen quimérico que comprende un promotor de un gen que codifica para deshidrina de café, unido operablemente a una o más secuencias codificantes. Un vector para transformar una célula, que comprende un gen quimérico, también se proporciona así como células transformadas con el vector y

plantas transgénicas fértiles producidas por regeneración de la célula vegetal transformada con el vector.

Otras características y ventajas de la presente invención se comprenderán a partir de los dibujos, la descripción detallada y ejemplos que siguen.

DESCRIPCIÓN BREVE DE LOS DIBUJOS

Figura 1. Alineación óptima de deshidrinas tipo Y_3SK_2 de *Coffea canephora* con varios homólogos vegetales cercanos. La alineación se genera utilizando el programa Clustal W en el software MegAlign (DNASTAR) y después se optimiza de manera adicional manualmente. Los aminoácidos idénticos se encuentran dentro de rectángulos. Las líneas continuas indican los segmentos Y, los rectángulos oscuros solos indican los segmentos S y los rectángulos con líneas discontinuas indican los segmentos K. El círculo negro representa el aminoácido diferente entre CcDH2a y CcDH2b. Números de acceso: deshidrina TAS14 de *Lycopersicon esculentum* (AAC49618) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 14); deshidrina Dhn1 de *Solanum commersonii* (CAA75798) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 15); deshidrina RAB18 de *Arabidopsis thaliana* (NP_201441) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 16).

Figura 2. Alineación óptima de la deshidrina tipo SK_3 de *Coffea canephora* con varios homólogos vegetales

cercanos. La alineación se genera como se describe para la figura 1. Los aminoácidos idénticos se encuentran en rectángulos. Los rectángulos oscuros solos indican los segmentos S y los rectángulos con líneas discontinuas indican los segmentos K. Números de acceso: deshidrina de *Nicotiana tabacum* (BAD13499) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 17); homólogo CI7 de deshidrina de *Solanum tuberosum* (T07779) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 18); deshidrina de *Arabidopsis thaliana* (CAA62449) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 19).

Figura 3. Alineación óptima de proteína abundante de embriogénesis tardía de *Coffea canephora* CcLEA1 con varios homólogos vegetales cercanos. La alineación se genera como se describe para la figura 1. Los aminóacidos idénticos están dentro de los rectángulos. Las cisteínas conservadas se marcan con un asterisco y la posición de las dos cisteínas menos altamente conservadas se marcan con un círculo. Números de acceso: La proteína relacionada abundante con embriogénesis tardía de *Arabidopsis thaliana* (NP_200248) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 20), proteína abundante de embriogénesis tardía de *Picea glauca* EMB7 (T09288) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 21), proteína 2 de tapa de raíz de *Zea mays* (BAA75477) (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 22).

Figura 4. Análisis de expresión por RT-PCR de las

deshidrinas de café y los transcritos de CcLEA1 en diferentes órganos de *Coffea arabica* y *Coffea canephora*. 100 pares de bases representan una escalera marcadora de peso molecular de 100 pb: SG, LG, YG y RG, representan el grano verde pequeño, verde grande, verde amarillo y rojo para el grano del pericarpio, respectivamente; R, S, L, F, representan raíces, tallos, hojas y flores.

Figura 5. Análisis de expresión cuantitativa por RT-PCR de CcDH2 en diferentes órganos de *Coffea canephora* y *Coffea arabica*. GSG, GLG, GYG y GRG representan los granos verde pequeño, verde grande, amarillo y rojo, respectivamente; PSG, PLG, PYG y PRG representan el pericarpio verde pequeño, verde grande, amarillo y rojo, respectivamente; R, S, L, F representa raíz, tallo, hojas y flores. Las desviaciones estándar se reportan en la gráfica para cada reacción.

Figura 6. Análisis de transferencia Southern de deshidrina CcDH2. El autorradiograma se expone durante tres días.

Figura 7. Secuencia de ADN del promotor CcDH2a y secuencia transcrita de *Coffea canehpora*. Se presenta la secuencia de ácido nucleico del injerto pVC1, junto con la secuencia de aminoácidos correspondiente. La primera base en ADNc (C) está marcada con un círculo. La caja TATA putativa está subrayada. Las secuencias repetidas RY se

marcan con un subrayado doble y los elementos sensibles a ABA se encuentran dentro de un rectángulo y la secuencia DRE/CRT se encuentra en un rectángulo con líneas gruesas.

Figura 8. Gráficas de hidrofilicidad de Kyte-Doolittle de polipéptidos codificados: CcDH1a (173 aminoácidos, SECUENCIA DE IDENTIFICACIÓN NÚMERO: 7); CcDH1b (176 aminoácidos, SECUENCIA DE IDENTIFICACIÓN NÚMERO: 8); CcDH2a (163 aminoácidos SECUENCIA DE IDENTIFICACIÓN NÚMERO: 9); CcDH2b (163 aminoácidos SECUENCIA DE IDENTIFICACIÓN NÚMERO: 10); CcDH3 (228 aminoácidos SECUENCIA DE IDENTIFICACIÓN NÚMERO: 11); CcLEA1 (358 aminoácidos, SECUENCIA DE IDENTIFICACIÓN NÚMERO: 12).

Figura 9. Análisis de expresión por RT-PCR cuantitativa en tiempo real de la expresión de CcDH1 en las hojas de planta control y tensionadas por sequía. Las plantas 1-3 son plantas control que se humedecen regularmente, las plantas 4-6 no se les proporciona agua desde el inicio del experimento ("0") hasta la semana 6.

Figura 10. Análisis de expresión RT-PCR cuantitativo en tiempo real de la expresión de CcDH2 en las hojas de las plantas control y tensionadas con sequía. A las plantas 1-3 son plantas control a las que se les humedecen regularmente; las plantas 4-6 no se les proporciona agua desde el inicio del experimento ("0")

hasta la semana 6.

DESCRIPCIÓN DETALLADA DE LAS MODALIDADES ILUSTRATIVAS

Definiciones:

Diversos términos en relación a las moléculas biológicas y otros aspectos de la presente invención se utilizan a través de la especificación y las reivindicaciones.

El término "aislado" significa alterado "por la mano del hombre" de su estado natural. Si una composición o sustancia se encuentra en la naturaleza, se ha "aislado" si ha cambiado o se ha retirado de su ambiente general, o ambas cosas. Por ejemplo, un polinucleótido o un polipéptido presente de manera natural en una planta o en un animal vivo no está "aislada", pero el mismo polinucleótido o polipéptido separado de los materiales coexistentes de su estado natural está "aislado" como se utiliza el término en la presente.

El término "polinucleótido", también denominado como "molécula de ácido nucleico" generalmente se refiere a cualquier ribonucleótido o polidesoxirribonucleótido el cual pueda ser ARN o ADN no modificado o ARN o ADN modificado. El término "polinucleótidos" incluye, sin limitación, ADN de cadena sencilla y doble, ADN que es una

mezcla de las regiones de cadena sencilla y doble, ARN de cadena sencilla y doble y ARN que es una mezcla de regiones de cadena sencilla y doble, moléculas híbridas que comprenden ADN y ARN que puede ser de cadena sencilla o, de manera más típica, de cadena doble o una mezcla de regiones de cadena sencilla y doble. Además, el término "polinucleótido" se refiere a regiones de cadena triple que comprenden ARN o ADN o tanto ARN como ADN. El término polinucleótido también incluye los ADN o ARN que contienen una o más bases modificadas y ADN o ARN cuyas estructuras principales se han modificado por estabilidad o por otras razones. Las bases "modificadas" incluyen, por ejemplo, bases tritiladas y bases poco comunes tales como inosina. Se pueden realizar una diversidad de modificaciones a ADN y ARN; de esta manera, el término "polinucleótidos" abarca formas modificadas químicas, enzimática o metabólicamente de polinucleótidos, como se encuentran típicamente en la naturaleza así como las formas de químicas de ADN y ARN características de virus y células. El término "polinucleótidos" también abarca polinucleótidos relativamente cortos, con frecuencia denominados como oligonucleótidos.

El término "polipéptido" se refiere a cualquier péptido o proteína que comprende dos o más aminóacidos unidos entre sí por enlaces peptídicos o enlaces peptídicos

modificados, es decir, isósteros de péptido. El término "polipéptido" se refiere a ambas cadenas cortas, denominadas comúnmente como péptidos, oligopéptidos u oligómeros y a cadenas más largas, que generalmente se denominan como proteínas. Los polipéptidos pueden contener aminoácidos diferentes de los 20 aminoácidos codificados por genes. Los "polipéptidos" incluyen secuencias de aminoácidos modificadas ya sea por procedimientos naturales, tal como procesamiento postraduccional o por técnicas de modificación química las cuales son bien conocidas en el arte. Dichas modificaciones se describen bien en textos básicos y en monografías más detalladas así como en literatura de investigación voluminosa. Las modificaciones se pueden presentar en cualquier parte en un polipéptido e incluyen a la estructura principal peptídica, las cadenas laterales de aminoácidos y las partes terminales amino o carboxilo. Se apreciará que el mismo tipo de modificación puede estar presente en el mismo grado o en grados variables en varios sitios en un polipéptido dado. Además, un polipéptido dado puede contener muchos tipos de modificaciones. Los polipéptidos se pueden ramificar como resultado de su condición ubicua y pueden ser cíclicos, con o sin ramificación. Los poliéptidos cíclicos, ramificados y ramificados cíclicos pueden resultar de procedimientos postraduccionales naturales o se

pueden elaborar por métodos sintéticos. Las modificaciones incluyen acetilación, acilación, ADP-ribosilación, amidación, unión covalente de flavina, unión covalente de una porción hemo, unión covalente de un nucleótido o derivado nucleotídico, unión covalente de un lípido o derivado lipídico, unión covalente de fosfatidilinositol, reticulado, ciclización, formación de enlace disulfuro, desmetilación, formación de reticulaciones covalentes, formación de cistina, formación de piroglutamato, formilación, γ-carboxilación, glucosilación, formación de ancla GPI, hidroxilación, yodación, metilación, miristoilación, oxidación, procesamiento proteolítico, fosforilación, prenilación, racemización, selenoilación, sulfatación, adición de aminoácidos a proteínas mediada por ARN de transferencia tales como arginilación y ubiquitinación. Véase, por ejemplo, Proteins - Structure and Molecular Properties, 2a. Ed., T. E. Creighton, W. H. Freeman and Company, Nueva York, 1993 y Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 en Posttranslational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, Nueva York, 1983; Seifter *et al.*, "Analysis for Protein Modifications and Nonprotein Cofactors", Meth Enzymol (1990) 182:626-646 y Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", Ann

NY Acad Sci (1992) 663:48-62.

El término "variante" como el término utilizado en la presente es un polinucleótido o un polipéptido que difiere de un polinucleótido o polipéptido de referencia respectivamente pero que retiene propiedades esenciales. Una variante típica de un polinucleótido difiere en la secuencia nucleotídica de otro polinucleótido de referencia. Los cambios en la secuencia nucleotídica de la variante pueden o no alterar la secuencia de aminoácido de un polipéptido codificado por el polinucleótido de referencia. Los cambios nucleotídicos pueden resultar en sustituciones, adiciones, supresiones, fusiones y corte de aminoácidos en el polipéptido codificado por la secuencia de referencia, como se expone en lo siguiente. Una variante típica de un polipéptido difiere en la secuencia de aminoácidos de otro polipéptido de referencia. Generalmente, las referencias se limitan de manera que las secuencias del polipéptido de referencia y las variantes son cercanamente similares en general y en muchas regiones, son idénticas. Un polipéptido variante y de referencia puede diferir en la secuencia de aminoácidos por una o más sustituciones, adiciones o supresiones en cualquier combinación. Un residuo aminoácido sustituido o insertado puede o no estar codificado por el código genético. Una variante de un polinucleótido o polipéptido puede ser una

que se presente de manera natural tal como una variante alélica o puede ser una variante que no se conozca que se produce de manera natural. Las variantes que no se presentan de manera natural de los polinucleótidos y polipéptidos se pueden elaborar por técnicas de mutagénesis o por síntesis directa.

En referencia a plantas mutantes, los términos "mutante nulo" o "mutante con pérdida de función" se utilizan para designar un organismo o secuencia de ADN genómico con una mutación que provoca que un producto de gen sea no funcional o que esté ausente en gran medida. Dichas mutaciones se pueden presentar en las regiones codificante o reguladora del gen y pueden ser cambios de residuos individuales o inserciones o supresiones de regiones de ácidos nucleicos. Estas mutaciones también se pueden presentar en regiones codificantes o reguladoras de otros genes las cuales pueden regular o controlar un gen o proteína codificada de manera que provocan que la proteína sea no funcional o esté ausente en gran medida.

El término "sustancialmente la misma" se refiere a secuencias de ácido nucleico o de aminoácidos que tienen variaciones de secuencia que no afectan materialmente la naturaleza de la proteína (es decir, la estructura, características de estabilidad, especificidad de sustrato o actividad biológica de la proteína). Con referencia

particular a secuencias de ácido nucleico, el término "sustancialmente el mismo" se pretende que se refiera a la región codificante y a las secuencias conservadas que gobiernan la expresión y se refiere principalmente a codones degenerados que codifican para el mismo aminoácido o codones alternados que codifican para aminoácidos de sustitución conservadora en el polipéptido codificado. Con referencia a las secuencias de aminoácidos, el término "sustancialmente el mismo" se refiere de manera general a sustituciones o variaciones conservadoras en las regiones del polipéptido no involucradas en la determinación de la estructura o función.

Los términos "porcentaje idéntico" y "porcentaje similar" también se utilizan en la presente en comparaciones con secuencias de aminoácidos y ácidos nucleicos. Con referencia a las secuencias de aminoácidos, los términos "identidad" o "porcentaje idéntico" se refieren al porcentaje de aminoácidos de la secuencia de aminoácidos objeto que se han hecho coincidir con aminoácidos idénticos en la secuencia de aminoácidos comparada por un programa de análisis de secuencia. El término "porcentaje similar" se refiere al por ciento de aminoácidos de la secuencia de aminoácidos objeto que se han hecho coincidir con aminoácidos idénticos o conservados. Los aminoácidos conservados son aquellos los

cuales difieren en estructura pero son similares en propiedades físicas de manera que un intercambio de uno por otro no cambia apreciablemente la estructura terciaria de la proteína resultante. Las sustituciones conservadoras se definen en Taylor (1986, J. Theor. Biol. 119:205). Cuando se hace referencia a moléculas de ácido nucleico, el "porcentaje idéntico" se refiere al porcentaje de nucleótidos de la secuencia de ácido nucleico sujeto que se han hecho coincidir con nucleótidos idénticos por un programa de análisis de secuencia.

Los términos "identidad" y "similitud" se pueden calcular fácilmente por métodos conocidos. Las secuencias de ácido nucleico y las secuencias de aminoácidos se pueden comparar utilizando programas de computadora que alinean las secuencias similares de los ácidos nucleicos o de aminoácidos y de esta manera definen las diferencias. En las metodologías preferidas, se emplean los programas BLAST (NCBI) y los parámetros utilizados en ése programa y se utiliza el sistema DNAstar (Madison, WI) para alinear fragmentos de secuencia de secuencias de ADN genómico. No obstante, las alineaciones equivalentes y las determinaciones de similitud/identidad se pueden obtener mediante el uso de cualquier programa de alineación estándar. Por ejemplo, el GCG Winsconsin Package versión 9.1, disponible de Genetics Computer Group en Madison,

Wisconsin, y los parámetros implícitos o utilizados (castigo por creación de separación = 12, castigo por extensión de separación = 4) por ése programa también se pueden utilizar para comparar la identidad y similitud de secuencias.

Los "anticuerpos", como se utilizan en la presente, incluyen anticuerpos policlonales y monoclonales, anticuerpos quiméricos, de cadena sencilla y humanizados así como fragmentos de anticuerpo (por ejemplo Fab, Fab', $F(ab')_2$ y F_v), que incluye a los productos de un Fab u otra biblioteca de expresión de inmunoglobulina. Con respecto a los anticuerpos, el término "inmunológicamente específico" o "específico" se refiere a anticuerpos que se unen a uno o más epítopos de una proteína de interés, pero los cuales no reconocen sustancialmente ni se unen a otras moléculas en una muestra que contiene una población mixta de moléculas biológicas antigénicas. Los análisis de cribado para determinar la especificidad de unión de un anticuerpo son bien conocidos y se practican de manera habitual en la técnica. Para una discusión extensa de dichos análisis véase Harlow *et al.* (Eds.), ANTIBODIES A LABORATORY MANUAL; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY (1988), Capítulo 6.

El término "sustancialmente puro" se refiere a una preparación que comprende por lo menos 50-60% en peso

del compuesto de interés (por ejemplo, ácido nucleico, oligonucleótido, proteína, etc.). De manera más preferible, la preparación comprende por lo menos 75% en peso y de manera mucho más preferible 90-99% en peso del compuesto de interés. La pureza se mide por métodos apropiados para el compuesto de interés (por ejemplo métodos cromatográficos, electroforesis en gel de agarosa o poliacrilamida, análisis por CLAP y similares).

Con respecto a las moléculas de ácido nucleico de cadena sencilla, el término "que hibridiza específicamente" se refiere a la asociación entre dos moléculas de ácido nucleico de cadena sencilla de secuencia complementaria suficiente para permitir dicha hibridación bajo condiciones predeterminadas utilizadas generalmente en la técnica (algunas veces denominada "sustancialmente complementaria"). En particular, el término se refiere a la hibridación de un oligonucleótido con una secuencia sustancialmente complementaria contenida dentro de una molécula de ADN o ARN de cadena sencilla, con la exclusión sustancial de hibridación del oligonucleótido con ácidos nucleicos de cadena sencilla de secuencia no complementaria.

Una "secuencia codificante" o "región codificante" se refiere a una molécula de ácido nucleico que tiene información de secuencia necesaria para producir

un producto de gen tal como un aminoácido o polipéptido cuando se expresa la secuencia. La secuencia codificante puede comprender secuencias no traducidas (por ejemplo intrones o regiones 5' ó 3' no traducidas), dentro de regiones traducidas o pueden carecer de dichas secuencias no traducidas intermedias (por ejemplo como en el ADNc).

El término "intrón" se refiere a secuencias polinucleotídicas en un ácido nucleico que no codifican información relacionada con la síntesis de proteínas. Dichas secuencias se transcriben en el ARNm pero son retiradas antes de la traducción del ARNm en una proteína.

El término "unido operablemente" o "insertado operablemente" significa que las secuencias reguladoras necesarias para expresión de la secuencia codificante se colocan en una molécula de ácido nucleico en las posiciones apropiadas en relación a la secuencia codificante de manera que permiten la expresión de la secuencia codificante. A modo de ejemplo, un promotor está unido operablemente con una secuencia codificante cuando el promotor es capaz de controlar la transcripción o expresión de dicha secuencia codificante. Las secuencias codificante se pueden unir operablemente a promotores o secuencias reguladoras en una orientación directa o antisentido. El término "unido operablemente" algunas veces se aplica a la distribución de otros elementos de control de transcripción (por ejemplo

mejoradores) en un vector de expresión.

Las secuencias de control transcripcional y traduccional son secuencias reguladoras de ADN, tales como promotores, mejoradores, señales de poliadenilación, terminadores y similares que proporcionan la expresión de una secuencia codificante en una célula hospedera.

Los términos "promotor", "región promotora" o "secuencia promotora" se refieren de manera general a regiones reguladores transcripcionales de un gen, las cuales se pueden encontrar en el lado 5' ó 3' de la región codificante o dentro de la región codificante, o dentro de intrones. Típicamente, un promotor es una región reguladora de ADN capaz de unir ARN polimerasa en una célula e iniciar la transcripción de una secuencia codificante hacia el extremo 3' (en la dirección 3'). La secuencia promotora 5' típica está unida en su parte 3' terminal por el sitio de inicio de transcripción y se extiende hacia arriba (dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Dentro de la secuencia promotora está un sitio de inicio de transcripción (definido convenientemente por el mapeo con la nucleasa S1), así como dominios de unión de proteína (secuencias de consenso) responsables de inhibir la ARN polimerasa.

Un "vector" es un replicón tal como un plásmido, fago, cósmido o virus al cual se le puede insertar operablemente otro segmento de ácido nucleico de manera que lleve a cabo la replicación o expresión del segmento.

El término "construcción de ácido nucleico" o "construcción de ADN" algunas veces se utiliza para referirse a una secuencia o secuencias codificantes unidas operablemente a secuencias reguladoras apropiadas e insertadas dentro de un vector para transformar una célula. Este término se puede utilizar de manera intercambiable con el término "ADN transformante" o "transgen". Dicha construcción de ácido nucleico puede contener una secuencia codificante para un producto de gen de interés, junto con un gen marcador seleccionable o un gen indicador.

Un "gen marcador" o "gen marcador seleccionable" es un gen cuyo producto de gen codificado confiere una característica que permite a una célula que contiene al gen que se selecciona de entre células que no contienen el gen. Los vectores utilizados para manipulación genética típicamente contienen uno o más genes marcadores seleccionables. Los tipos de genes marcadores seleccionables incluyen: (1) genes de resistencia a antibióticos, (2) genes que proporcionan tolerancia o resistencia a herbicidas y (3) genes marcadores metabólicos u auxotróficos que permiten que las células transformadas

sinteticen un componente esencial, habitualmente un aminoácido, el cual de otra manera las células no pueden producir.

Un "gen indicador" también es un tipo de gen marcador. Típicamente codifica para un producto de gen que es susceptible de ser analizado o que es detectable por un medio de laboratorio estándar (por ejemplo actividad enzimática, fluorescencia).

El término "que expresa", "expresado" o "expresión" de un gen se refiere a la biosíntesis de un producto de gen. El procedimiento involucra la transcripción del gen en ARNm y después traducción del ARNm en uno o más polipéptidos y abarca todas las modificaciones postraduccionales que se presenten de manera natural.

El término "endógeno" se refiere a cualquier constituyente, por ejemplo, un gen o ácido nucleico o un polipéptido que se puede encontrar de manera natural dentro del organismo especificado.

Una región "heteróloga" de una construcción de ácido nucleico es un segmento (o segmentos) identificable de la molécula de ácido nucleico con una molécula más grande que no se encuentra en asociación con la molécula más grande en la naturaleza. De esta manera, cuando la región heteróloga comprende un gen, el gen habitualmente estará flanqueado por ADN que no flanquea al ADN genómico

en el genoma del organismo fuente. En otro ejemplo, una región heteróloga es una construcción en donde la secuencia codificante en sí misma no se encuentra en la naturaleza (por ejemplo un ADNc en donde la secuencia codificante genómica contiene intrones o secuencias sintéticas que tienen codones diferentes que el gen nativo). Las variaciones alélicas o los eventos mutacionales que no se presentan de manera natural no dan lugar a una región heteróloga de ADN como se define en la presente. El término "construcción de ADN", como se define en lo anterior, tambíen se utiliza para referirse a una región heteróloga, particularmente una construida para uso en la transformación de una célula.

Una célula ha sido "transformada" o "transfectada" por ADN exógeno o heterólogo cuando dicho ADN ha sido introducido dentro de la célula. El ADN transformante puede o no estar integrado (unido covalentemente) dentro del genoma de la célula. En procariotas, levaduras y células de mamífero, por ejemplo, el ADN transformante se puede mantener en un elemento episómico tal como un plásmido. Con respecto a células eucarióticas una célula transformada de manera estable es aquella en la cual el ADN transformante se ha integrado al cromosoma de manera que es heredado por las células descendientes a través de replicación del cromosoma. Esta

estabilidad se demuestra por la capacidad de la célula eucariótica para establecer líneas de células o clones constituidos de una población de células descendientes que contienen el ADN transformante. Un "clón" es una población de células derivadas de una célula única o de ancestros comunes mediante mitosis. Una "línea celular" es un clón de una célula primaria que es capaz de crecimiento estable *in vitro* durante muchas generaciones.

Los términos "grano", "semilla" o "frijol", se refiere a la unidad de reproducción de la planta en florecimiento capaz de desarrollarse en otra planta similar. Como se utiliza en la presente, especialmente con respecto a las plantas de café, los términos se utilizan de manera sinónima e intercambiable.

Como se utiliza en la presente, el término "plantas" incluye referencia a plantas completas, órganos de plantas (por ejemplo hojas, tallos, vástagos, raíces), semillas, polen, células de las plantas, organelos de células de las plantas y descendencia de los mismos. Las partes de plantas transgénicas se va a entender dentro del alcance de la invención que comprenden, por ejemplo, células vegetales, protoplastos, tejidos, callos, embriones así como flores, tallos, semillas, polen, frutas, hojas o raíces que se originan en plantas transgénicas o en su descendencia.

El término "tensión osmótica" se refiere a cualquier tensión en la planta que interrumpe la concentración normal de agua, azúcar o electrolitos en una célula vegetal o en una planta en su totalidad. La tensión osmótica puede relacionarse ambientalmente, por ejemplo, con condiciones de una cantidad baja de agua o sequía prolongadas, temperaturas bajas, congelamiento, temperaturas de congelamiento, un contenido de sal elevado en el suelo y similares. La tensión osmótica también se puede presentar de manera natural, como se esperaría para el desarrollo y maduración de una semilla.

DESCRIPCIÓN

En uno de sus aspectos, la presente invención describe moléculas de ácido nucleico a partir de café que codifica para una diversidad de proteínas de deshidrina así como otra proteína LEA (abundante embriogénica tardía). Los ejemplos representativos de moléculas de ácido nucleico que codifican para deshidrina y LEA se identifican a partir de bases de datos de más de 47,000 etiquetas de secuencia expresada (EST) a partir de varias bibliotecas de ADNc de *Coffea canephora* (robusta) elaboradas con ARN aislado de hojas jóvenes y de granos y tejido del pericarpio de cerezos cosechados en etapas diferentes de desarrollo. Las EST superpuestas se identifican como "agrupadas" en genes

únicos (contiguos) que comprenden secuencias codificantes completas. Las secuencias de genes únicos se anotan al realizar una búsqueda BLAST de cada secuencia individual contra una base de datos de proteína no redundante de NCBI (National Center for Biotechnology Information). El análisis de secuencia de ADN muestra cinco secuencias únicas que representan tres genes diferentes para deshidrina y un gen para LEA. Estos ADNc se denominan en la presente como CcDH1a (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 1), CcDH1b (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 2), CcDH2a (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 3), CcDH2b (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 4) y CcDH3 (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 5). Se encontró que CcDH1a y CcDH1b son variantes alélicas entre sí mientras que se encontraron dos genes únicos distintos que codifican para un marco de lectura abierto para CcDH2. Además, el análisis de una biblioteca de ADNc construida a partir de ARN aislado de granos de café a 30 semanas después de la fertilización muestra un clón de ADNc de longitud completa que codifica para la proteína LEA de café. Este ADNc se denomina en la presente como CcLEA1 (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 6).

Las secuencias de aminoácidos deducidas de CcDH1a- CcDH3 se establecen en la presente como las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 7-11. Las proteínas

tienen masas moleculares de aproximadamente 17.8 kDa (CcDH1a, SECUENCIA DE IDENTIFICACIÓN NÚMERO: 7), 18.1 kDa (CcDH1b SECUENCIA DE IDENTIFICACIÓN NÚMERO: 8), 17.4 kDa (CcDH2a y CcDH2b, SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 9 y 10) y 21.5 kDa (CcDH3, SECUENCIA DE IDENTIFICACIÓN NÚMERO: 11). Estas proteínas se encuentra que contienen motivos de aminoácidos de deshidrina de signatura o firma y se clasifican de acuerdo con estos motivos. CcDH1a, CcDH1b y CcDH2 (a y b) tienen la estructura Y_3SK_2, y CcDH3 tiene la estructura SK_3. CcDH1a y CcDH1b muestran conservación absoluta en cada uno de los tres motivos y en las dos regiones conservadas que preceden a cada uno de los dos motivos K. En contraste, CcDH2 muestra diferencias puntuales en la totalidad excepto uno de los motivos Y, S y K y diferencias más significativas fuera de estos motivos específicos para deshidrina. El graficado de hidrofilicidad muestra que la totalidad de las proteínas de deshidrina de café identificadas en la presente son muy hidrofílicas en su totalidad (véase la figura 8).

La secuencia deducida de aminoácidos codificada por CcLEA1 se establece en la presente como la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 12. Esta proteína tiene una masa molecular de aproximadamente 39.5 kDa. La gráfica de hidrofilicidad de esta proteína indica que esta proteína es menos hidrofílica que las moléculas de deshidrina y que

existen dos regiones hidrofóbicas pequeñas, una de las cuales se localiza en sus primeros 30 residuos en la parte N terminal. La parte N terminal de CcLEA1 también se encuentra que contiene un segmento rico en prolina sorprendente.

Otro aspecto de la invención describe secuencias promotoras y elementos relacionados que controlan la expresión de genes para deshidrina en café. Como se describe con mayor detalle en los ejemplos, se identifica una secuencia promotora (contenida en la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 13), a partir de CcDH2a identificada por paseo de cebador asistido por PCR. El promotor CcDH2 se ha demostrado que contiene varios elementos reguladores análogos a los caracterizados previamente en otras especies que están involucradas en la regulación de expresión de un gen durante el desarrollo de semillas. Estos elementos reguladores de CcDH2 se muestran en la figura 7 y se describen en los ejemplos. Utilizando este promotor unido al gen indicador GUS, se ha determinado que el promotor es específico para semillas, silicuas, cotiledones, hipocotilos y primeras hojas verdaderas de plántulas en desarrollo. Además, como se describe en los ejemplos, la expresión de los genes para CcDH1 y CcDH2 también se ha demostrado que se induce por tensión de sequía y otras condiciones de tensión.

Aunque los polinucleótidos que codifican para las proteínas deshidrinas y LEA de *Coffea canephora* se describen y ejemplifican en la presente, esta invención se pretende que abarque ácidos nucleicos y proteínas codificadas de otras especies de *Coffea* que sean suficientemente similares para ser utilizadas de manera intercambiable con los polinucleótidos de *C. canephora* y proteínas para los propósitos descritos en lo siguiente. En consecuencia, cuando se utilizan en la presente los términos "deshidrina" o "proteínas abundantes de embriogénesis tardías (LEA)", se pretende que abarquen todas las proteínas deshidrinas o LEA de *Coffea* que tengan las características generales físicas, bioquímicas y funcionales descritas en la presente así como los polinucleótidos que las codifican.

Considerados en términos de sus secuencias, los polinucleótidos que codifican para las proteínas de deshidrina y abundantes de embriogénesis tardía de la invención incluyen variantes alélicas y mutantes naturales de las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 1-6, las cuales probablemente se encuentren en variedades diferentes de *C. canephora* y homólogos de las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 1-6 que probablemente se encuentren en especies de café diferentes. Debido a que dichas variantes y homólogos se espera que posean ciertas

diferencias en la secuencia nucleotídica y de aminoácidos, esta invención proporciona moléculas de ácido nucleico que codifican para las proteínas deshidrina o LEA aisladas que codifican para polipéptidos respectivos que tienen por lo menos aproximadamente 40%, 45%, 50% ó 55%, de manera preferible por lo menos aproximadamente 60, 65 ó 70% y de manera más preferible por lo menos aproximadamente 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 90% ó 80%, incluso de manera más preferible 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89% e incluso de manera más preferible 90%, 91%, 92%, 93%, 94%, 95% y de manera mucho más preferible 96%, 97%, 98% y 99% o más de identidad con cualquiera de las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 7-12 y que comprendan una secuencia nucleotídica que tenga intervalos equivalentes de identidad con cualquiera de las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 1-6. Debido a la variación de secuencia natural que probablemente exista entre las deshidrinas y las proteínas LEA y los genes que los codifican en diferentes variedades y especies de café, una persona experta en la técnica esperaría encontrar este nivel de variación mientras aún se mantiene las propiedades únicas de los polipéptidos y polinucleótidos de la presente invención. Dicha esperanza se debe en parte a la degeneración del código genético así como al éxito evolutivo conocido de las variaciones conservadoras en las

secuencias de aminoácidos las cuales no alteran de manera apreciable la naturaleza de la proteína codificada. En consecuencia, dichas variantes y homólogos se consideran sustancialmente los mismos que los otros y se incluyen dentro del alcance de la presente invención.

Como se ha mencionado, los inventores han demostrado que la expresión de algunos de los genes para las proteínas deshidrinas o LEA son semillas, silicuas y plántulas específicas en café así como el hecho de ser inducibles por sequía y otras formas de tensión. En consecuencia, las secuencias reguladoras de genes asociadas con los genes que codifican para la proteína deshidrina y LEA son de utilidad práctica y se les considera dentro del alcance de la presente invención. El promotor DH2 de *C. canephora* se ejemplifica en la presente. La región hacia el extremo 5' de la secuencia genómica de DH2 de *C. canephora* se establece en la presente como la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 13 y contiene parte o la totalidad de un promotor ejemplar de la invención, aunque otras porciones del promotor se pueden encontrar en otras ubicaciones en el gen, como se explica en la definición de "promotor" como se establece en lo anterior. No obstante, los promotores y otras secuencias reguladoras de gen de los genes para deshidrina y la proteína LEA de cualquier especie de café se pueden obtener por los métodos descritos

en lo siguiente y se pueden utilizar de acuerdo con la presente invención. Los promotores y elementos reguladores que gobiernan la especificidad de tejido y la especificidad temporal de la expresión del gen para deshidrina y para proteína LEA se pueden utilizar para aventajar para alterar o modificar la tolerancia a la tensión osmótica de diversas especies de café, entre otras utilidades.

Las siguientes secciones establecen los procedimientos generales involucrados en la práctica de la presente invención. En la medida en que se mencionan materiales específicos, es simplemente con propósitos de ilustración y no se pretende que limiten la invención. A menos que se especifique de otra manera, los procedimientos generales bioquímicos y de biología molecular, tales como los que se establecen en Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor Laboratory (1989) o Ausubel *et al.* (eds), Current Protocols in Molecular Biology, John Wiley & Sons (2005) son los que se utilizan.

Moléculas de ácido nucleico, proteínas y anticuerpos:

Las moléculas de ácido nucleico de la invención se pueden preparar por dos métodos generales: (1) pueden ser sintetizadas a partir de trifosfatos de nucleótidos apropiados, o (2) se pueden aislar a partir de fuentes biológicas. Ambos métodos utilizan protocolos bien

conocidos en la técnica.

La disponibilidad de información de secuencia nucleotídica tal como el ADNc que tiene las SECUENCIAS DE IDENTIFICACIÓN NÚMEROS: 1-6 o la secuencia reguladora de la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 13, habilita la preparación de una molécula aislada de ácido nucleico de la invención mediante síntesis oligonucleotídica. Los oligonucleótidos sintéticos se pueden preparar por el método de fosforoamidita utilizado en el equipo Applied Biosystems 38A DNA Synthetizer o en dispositivos similares. La construcción resultante se puede purificar de acuerdo con métodos conocidos en la técnica tal como cromatografía líquida de alta resolución (CLAR). Los polinucleótidos de cadena doble largos tales como una molécula de ADN de la presente invención se deben sintetizar en etapas, debido a las limitaciones de tamaño inherentes en los métodos de síntesis oligonucleotídicos actuales. Así, por ejemplo, una molécula larga de cadena doble se puede sintetizar como varios segmentos más pequeños de complementariedad apropiada. Los segmentos de complementariedad producidos de esta manera se pueden reasociar de manera que cada segmento posee términos cohesivos apropiados para unión de un segmento adyacente. Los segmentos adyacentes se pueden ligar por reasociación de las partes terminales cohesivas en presencia de ADN ligasa para construir la totalidad de

la molécula larga de cadena doble. Una molécula de ADN sintética construida de esta manera después se puede clonar y amplificar en un vector apropiado.

De acuerdo con la presente invención, los ácidos nucleicos que tengan la homología de secuencia de nivel apropiado con parte o la totalidad de las secuencias codificante o reguladora de los polinucleótidos que codifican para deshidrina o la proteína LEA se pueden identificar mediante la utilización de hibridación y condiciones de lavado de rigurosidad apropiada. Se apreciará por aquellos expertos en la técnica que la estrategia mencionada antes, cuando se aplica a secuencias genómicas, además de habilitar el aislamiento de las secuencias que codifican para deshidrina o la proteína LEA, también será capaz de aislamiento de promotores y otras secuencias reguladoras de genes relacionadas con deshidrina o genes de proteína LEA, aunque las secuencias reguladoras mismas pueden no compartir homología suficiente para habilitar la hibridación adecuada. Además, la anotación de por lo menos una secuencia codificante parcial permitirá que un experto en la técnica determine la secuencia codificante remanente así como las secuencias reguladoras promotoras y de otros genes relacionadas con la deshidrina o proteína LEA de interés mediante la técnica de caminado genómico hacia el extremo 5' o hacia el extremo 3'. Dichas

técnicas se establecen en el arte. (Mishra RN *et al.*, 2002; Rishi AS *et al.*, 2004).

Como una ilustración típica, las hibridaciones se pueden realizar, de acuerdo con el método de Sambrook *et al.*, utilizando una solución de hibridación que comprende: 5X SSC, 5X de reactivo de Denhardt, SDS 1.0%, 100 μg/ml de ADN de esperma de salmón fragmentado y desnaturalizado, pirofosfato de sodio 0.05% y hasta 50% de formamida. La hibridación se lleva a cabo a 37-42°C durante por lo menos seis horas. Después de la hibridación los filtros se lavan como sigue: (1) 5 minutos a temperatura ambiente en 2X SSC y SDS 1%; (2) 15 minutos a temperatura ambiente en 2X SSC y SDS 0.1%; (3) 30 minutos-1 hora a 37°C en 2X SSC y SDS 0.1%; (4) 2 horas a 45-55°C en 2X SSC y SDS 0.1%, cambiando la solución cada 30 minutos.

Una fórmula común para calcular las condiciones de rigurosidad necesarias para obtener hibridación entre moléculas de ácido nucleico de una homología de secuencia especificada (Sambrook *et al.*, 1989):

$$Tm = 81.5°C + 16.6\text{Log}\,[Na+] + 0.41\,(\%\ G+C) - 0.63\,(\%\ \text{de formamida}) - 600/\#\text{pares de bases en dúplex}$$

Como una ilustración de la fórmula anterior, utilizando [Na+] = [0.368] y formamida al 50%, con un

contenido de GC de 42% y un tamaño de sonda promedio de 200 bases, la Tm es de 57°C. La Tm de un ADN dúplex disminuye en 1-1.5°C con cada disminución de 1% en la homología. Por lo tanto, los objetivos con más de aproximadamente 75% de identidad de secuencia se observarán utilizando una temperatura de hibridación de 42°C. En una modalidad, la hibridación es a 37°C y el lavado final es a 42°C; en otra modalidad, la hibridación es a 42°C y el lavado final es a 50°C; y en otra modalidad adicional la hibridación es a 42°C y el lavado final es a 65°C con las soluciones de hibridación y lavado anteriores. Las condiciones de alta rigurosidad incluyen hibridación a 42°C en la solución de hibridación anterior y un lavado final a 65°C en 0.1X SSC y SDS 0.1% durante 10 minutos.

Los ácidos nucleicos de la presente invención se pueden mantener como ADN en cualquier vector de clonación conveniente. En una modalidad preferida, los clones se mantienen en un plásmido de un vector de clonación/expresión tal como pGEM-T (Promega Biotech, Madison, WI), pBluescript (Stratagene, La Jolla, CA), pCR4-TOPO (invitrogen, Carlsbad, CA) o pET28a+ (Novagen, Madison, WI), la totalidad de los cuales se puede propagar en una célula hospedera *E. coli* adecuada.

Las moléculas de ácido nucleico de la invención incluyen ADNc, ADN genómico, ARN y fragmentos de los mismos

los cuales pueden ser de cadena sencilla, doble o incluso triple. De esta manera, esta invención proporciona oligonucleotidos (cadenas directas o antisentido de ADN o ARN) que tienen secuencias capaces de hibridizar con por lo menos una secuencia de una molécula de ácido nucleico de la presente invención. Dichos oligonucleotidos son útiles como sondas para detectar genes que codifican para deshidrina o proteína LEA o ARNm en muestras de prueba o en tejido vegetal, por ejemplo por amplificación por PCR o para la regulación de expresión positiva o negativa de genes que codifican para deshidrina o proteína LEA en o antes de la traducción del ARNm en proteínas. Los métodos en los cuales los oligonucleotidos o polinucleótidos que codifican para deshidrina o proteína LEA se pueden utilizar como sondas para dichos análisis e incluyen, pero no se limitan a: (1) hibridación *in situ*; (2) hibridación Southern; (3) hibridación Northern y (4) reacciones de amplificación diversas tal como reacciones en cadena de polimerasa (PCR, que incluye RT-PCR) y reacción en cadena de ligasa (LCR).

Los polipéptidos codificados por ácidos nucleicos de la invención se pueden preparar en una diversidad de maneras, de acuerdo con métodos conocidos. Si se producen in situ, los polipéptidos se pueden purificar a partir de fuentes apropiadas, por ejemplo semillas, pericarpios y otras partes de la planta.

De manera alternativa, la disponibilidad de moléculas de ácido nucleico que codifican para polipéptidos que habilitan la producción de las proteínas utilizando métodos de expresión in vitro conocidos en la técnica. Por ejemplo, un ADNc o un gen se puede clonar en un vector de trascripción in vitro apropiado tal como pSP64 o pSP65 para trascripción in vitro, seguido por traducción sin células en un sistema de traducción sin células tal como reticulocitos de germen de trigo o de conejo. Los sistemas de trascripción y traducción in vitro están disponibles comercialmente, por ejemplo, de Promega Biotech, Madison, WI, BRL, Rockville, MD o Invitrogen, Carlsbad, CA.

De acuerdo con una modalidad preferida, las cantidades más grandes de polipéptidos de deshidrinas o proteína LEA se pueden elaborar por expresión en un sistema adecuado procariótico o eucariótico. Por ejemplo, parte o la totalidad de una molécula de ADN, tal como los ADNc que tienen las SECUENCIAS DE IDENTIFICACION NUMEROS: 1-6 se pueden insertar en el vector plasmídico adaptado para expresión en una célula bacteriana (tal como *E. coli*) o una célula de levadura (tal como *Saccharomyces cerevisiae*) o en un vector de baculovirus para expresión en una célula de insecto. Dichos vectores comprenden los elementos reguladores necesarios para la expresión del ADN en la célula hospedera, colocado de manera tal que permite la

expresión del ADN en la célula hospedera. Dichos elementos reguladores necesarios para expresión incluyen secuencias promotoras, secuencias de inicio de trascripción y, opcionalmente, secuencias mejoradoras.

Las deshidrinas o proteínas LEA producidas por expresión de genes en un sistema procariótico o eucariótico recombinante se pueden purificar de acuerdo con métodos conocidos en la técnica. En una modalidad preferida, se puede utilizar un sistema de expresión/secreción disponible comercialmente, por lo que la proteína recombinante se expresa y posteriormente se secreta de la célula hospedera, para que se purifique fácilmente del medio circundante. Si no se utilizan vectores de expresión/secreción, una solución alternativa involucra purificar la proteína recombinante por separación por afinidad, tal como interacción inmunológica con anticuerpos que se unen específicamente a la proteína recombinante. Dichos métodos se utilizan comúnmente por los expertos en la técnica.

Las deshidrinas y proteínas LEA de la invención, preparadas por los métodos mencionados antes se pueden analizar de acuerdo con procedimientos estándar.

Las deshidrinas y proteínas LEA purificadas a partir de café o producidas recombinantemente se pueden utilizar para generar anticuerpos policlonales o monoclonales, fragmentos de anticuerpo o derivados como se

definen en la presente, de acuerdo con métodos conocidos. Los anticuerpos que reconocen y unen fragmentos de las deshidrinas o proteínas LEA de la invención también se contemplan, con la condición de que los anticuerpos sean específicos par deshidrinas o proteínas LEA. Por ejemplo, si los análisis de las proteínas o los análisis Southern y de clonación (véase abajo) indican que los genes clonados pertenecen a una familia de genes múltiples, entonces los anticuerpos específicos de miembro elaborados para péptidos sintéticos que corresponden con regiones no conservadas de la proteína se pueden generar.

Los equipos que comprenden un anticuerpo de la invención para cualquiera de los propósitos descritos en la presente también se incluyen dentro del alcance de la invención. En general, dicho equipo incluye un antígeno de control para el cual el anticuerpo es inmunoespecífico.

Las deshidrinas, y probablemente también las proteínas LEA están involucradas en proteger los componentes celulares de tensiones osmóticas (deshidratación, bajas temperaturas/congelamiento, concentración de sal). En consecuencia, las deshidrinas y proteínas LEA de café descritas y ejemplificadas en la presente se espera que encuentren utilidad en una diversidad de aplicaciones alimenticias y cosméticas. Por ejemplo, las deshidrinas o proteínas LEA se pueden utilizar

para alterar la nucleación de hielo en alimentos congelados o para facilitar el secado de proteínas de una manera que permite la rehidratación rápida en una etapa posterior. Como otro ejemplo, las deshidrinas o proteínas LES se pueden utilizar para hidratar productos de crema cutáneos. Además, las propiedades recién descubiertas antioxidantes o de unión de iones de las deshidrinas pueden demostrar ser ventajosas tanto en productos alimenticios como cosméticos. En conexión con aplicaciones alimenticias, resulta notable que las deshidrinas son una proteína muy estructurada y altamente soluble y que no se conocen que tengan enlaces disulfuro. Como un resultado, es probable que estas proteínas presenten una antigenicidad muy baja y probablemente se digieran fácilmente por proteasas en el intestino.

Una o más de las aplicaciones mencionadas antes para las deshidrinas o proteínas LEA se pueden buscar al aprovechar la disponibilidad de polinucleótidos que codifican para deshidrina y la proteína LEA que se describen en la presente para generar cantidades significativas de proteína pura utilizando organismos recombinantes (por ejemplo en la levadura *Pichia pastoris* o en lactobacilos compatibles con alimentos o en células vegetales) y después probar las proteínas en análisis establecidos de antemano para formación de hielo, efectos

en el secado, rehidratación y potencial antioxidante. Si se encuentra que las proteínas purificadas específicas son particularmente útiles también se pueden aislar versiones naturales de estas proteínas de granos de café que se determina que son ricos estas deshidrinas o proteínas LEA particulares.

Vectores, Células, Tejidos y Plantas:

También se describen de acuerdo con la presente invención vectores y equipos para producir células hospederas transgénicas que contienen polinucleótidos u oligonucleótidos que codifican para deshidrina o proteína LEA, u homólogos, análogos o variantes de los mismos en una orientación directa o antisentido, o un gen indicador y otras construcciones bajo el control de los promotores del gen que codifica para deshidrina o proteína LEA y otras secuencias reguladoras. Las células hospederas adecuadas incluyen, pero no se limitan a células vegetales, células bacterianas, levaduras y otras células micóticas, células de insecto y células de mamífero. Los vectores para transformación de una amplia variedad de estas células hospederas son bien conocidos por aquellos expertos en la técnica. Estos incluyen, pero no se limitan a plásmidos, fagémidos, cósmidos, baculovirus, bácmidos, cromosomas artificiales bacterianos (BAC), cromosomas artificiales de

levadura (YAC) así como otros vectores bacterianos, de levaduras virales. Típicamente, los equipos para producir células hospederas transgénicas contendrán uno o más vectores apropiados e instrucciones para producir las células transgénicas utilizando el vector. Los equipos pueden incluir adicionalmente uno o más componentes adicionales tales como medio de cultivo para cultivar las células, reactivo para realizar la transformación de las células y reactivos para probar las células transgénicas para expresión del gen, por mencionar algunos.

La presente invención incluye plantas transgénicas que comprenden una o más copias de un gen que codifica para deshidrina o proteína LEA o secuencias de ácido nucleico que inhibe la producción o la función de deshidrinas o proteínas LEA endógenas de la planta. Esto se lleva a cabo por transformación de las células vegetales con un transgen que comprende parte o la totalidad de una secuencia que codifica para deshidrina o proteína LEA, o un mutante, antisentido o variante del mismo, que incluye ARN, controlado ya sea por secuencias reguladoras nativas o recombinantes, como se describe en lo siguiente. Se prefieren especies de plantas de café transgénicas que incluyen, sin limitación, *C. abeokutae, C. arabica, C. arnoldiana, C. aruwemiensis, C. bengalensis, C. canephora, C. congenis C. dewevrei, C. excelsa, C. eugenioides* y *C.*

heterocalyx, C. kapakata, C. khasiana, C. liberica, C. moloundou, C. rasemosa, C. salvatrix, C. sessiflora, C. stenophylla, C. travencorensis, C. wightiana y C. zanguebariae. También se incluyen en la invención plantas de cualquier especie; estas incluyen pero no se limitan a tabaco, *Arabidopsis* y otras especies "fáciles de manipular de laboratorio", cosechas de cereal tal como maíz, trigo, arroz, soya, cebada, centeno, avenas, sorgo, alfalfa, trébol y similares, plantas productoras de aceite tales como nebina, azafrán, girasol, cacahuate, cacao y similares, cosechas vegetales tales como tomate tomatillo, papa, pimienta, berenjena, remolacha, zanahoria, pepino, lechuga, chícharo y similares, plantas hortícolas tales como aster, begonia, crisantemo, delfinium, petunia, zinia, lawn y turfgrasses y similares.

Las plantas transgénicas se pueden generar utilizando métodos de transformación vegetal estándar conocidos por aquellos expertos en la técnica. Estas incluyen, pero no se limitan a vectores de *Agrobacterium*, tratamiento de protoplastos con polietilenglicol, suministro de ADN biolístico, micro-haces láser UV, vectores de virus gemini y otros vectores virales vegetales, tratamiento de protoplastos con fosfato de calcio, electroporación de protoplastos aislados, agitación de suspensión celulares en solución con microesferas

recubiertas con el ADN transformante, agitación de suspensión celular en solución con fibras de silicio recubiertas con ADN transformante, captación directa de ADN, captación de ADN mediada por liposomas y similares. Dichos métodos se han publicado en la técnica. Véase, por ejemplo, Methods for Plant Molecular Biology (Weissbach & Weissbach, eds., 1988(; Methods in Plant Molecular Biology (Schuler & Zielinski, eds., 1989); Plant Molecular Biology Manual (Gelvin, Schilperoort, Verma, eds., 1993); y Methods in Plant Molecular Biology - A Laboratory Manual (Maliga, Klessig, Cashmore, Gruissem & Varner, eds., 1994).

El método de transformación depende de la planta que se va a transformar. Los vectores de *Agrobacterium* con frecuencia se utilizan para transformar especies dicotiledoneas. Los vectores binarios de *Agrobacterium* incluyen, pero no se limitan a BIN19 y derivados de los mismos, serie del vector pBI y vectores binarios pGA482, pGA492, pLH7000 (GenBank, Acceso AY234330) y cualquiera adecuado de los vectores pCAMBIA (derivados a partir de vectores pPZP construidos por Hajdukiewicz, Svab & Maliga, (1994) Plant Mol Biol. 25:989-994 disponible de CAMBIA, GPO Box 3200, Canberra ACT 2601, Australia o por medio de la red mundial en CAMBIA.org). Para transformación de especies monocotiledoneas, con frecuencia es útil para transformación nuclear el bombardeo biolístico con

partículas recubiertas con ADN transformante y fibras de silicio recubiertas con el ADN transformante. De manera alternativa, se han utilizado con éxito vectores (superbinarios) de *Agrobacterium* para la transformación de arroz, maíz y otras especies monocotiledoneas.

Las construcciones de ADN para transformar una planta seleccionada comprenden una secuencia codificante de interés unida operablemente a secuencias reguladoras 5' apropiadas (por ejemplo promotores y secuencias reguladoras traduccionales) y secuencias reguladoras 3' (por ejemplo terminadores). En una modalidad preferida, se utiliza una secuencia codificante para deshidrina o proteína LEA bajo el control de sus elementos reguladores 5' y 3' naturales. En otras modalidades, las secuencias reguladoras que codifican para deshidrina o proteína LEA se barren (por ejemplo la secuencia codificante CcLEA1 unida operablemente al promotor CcDH2) para alterar el contenido de agua o proteína de la semilla de la planta transformada para una mejoría genotípica, por ejemplo en sabor, aroma u otra característica.

En una modalidad alternativa, la región codificante del gen se coloca bajo un promotor constitutivo poderoso tal como el promotor 35S del virus de mosaico de coliflor (CaMV) o el promotor 35S del virus de mosaico de escrofularia. Otros promotores constitutivos contemplados

para uso en la presente invención incluyen, pero no se limitan a: manopina sintetasa T-DNA, nopalina sintasa y promotores de octopina sintasa. En otras modalidades se utiliza un promotor monocotiledoneo fuerte, por ejemplo el promotor de ubiquitina de maíz, el promotor de actina de arroz o el promotor de tubulina de arroz (Jeon *et al.*, Plant Physiology. 123: 1005-14, 2000).

Se contempla que estén dentro del alcance de la presente invención plantas transgénicas que expresan deshidrina o secuencias codificantes de proteína LEA bajo un promotor inducible. Los promotores vegetales inducibles incluyen un promotor controlado por un represor/operador de tetraciclina, los promotores de genes de choque térmico, promotores inducidos por tensión (por ejemplo heridas), promotores de un gen de respuesta de defensa (por ejemplo genes para fenilalanina amoníaco liasa), promotores de genes inducidos por heridas (por ejemplo genes para la proteína de la pared celular ricos en hidroxiprolina), promotores de genes inducibles químicamente (por ejemplo genes para nitrato reductasa, genes para glucanasa, genes para quitinasa, etcétera) y promotores de genes inducibles en la oscuridad (por ejemplo gen para asparagina sintetasa) por mencionar sólo algunos.

Los promotores específicos de tejido y específicos de desarrollo también se contemplan para uso en

la presente invención, además de los promotores para deshidrina o proteína LEA específicos de semilla de la invención. Los ejemplos no limitantes de otros promotores específicos de semilla incluyen Cim1 (mensaje inducido por citocinina), cZ19B1 (zeína de 19 kDa de maíz), milps (myo-inositol-1-fosfato sintasa) y celA (celulosa sintasa) (Solicitud de E.U.A. No. de Serie 09/377,648), β-faseolina de frijol, napina, β-conglicinina, lectina de soya, cruciferina, zeína de 15 kDa de maíz, zeína de 22 kDa, zeína de 27 kDa, zeina g, ceroso, contraído 1, contraído 2 y globulina 1, legúmina 11S de soya (Bäumlein *et al.*, 1992) y proteína de almacenamiento de semilla 11S *C. canephora* (Marraccini *et al.*, 1999, Plant Physiol Biochem. 37:273-282). Véase también el documento WO 00/12733, en donde se describen promotores de semilla preferidos de los genes end1 y end2. También se pueden utilizar otros promotores específicos de semilla de *Coffea* que incluyen pero que no se limitan a un promotor de gen de oleosina descrito en la Solicitud PCT Copendiente y de Cesión Común No. [AUN NO ASIGNADO]. Los ejemplos de otros promotores específicos de tejido incluyen, pero no se limitan a: los promotores del gen de la subunidad pequeña de ribulosa bisfosfato carboxilasa (RuBisCo) (por ejemplo el promotor de la subunidad pequeña de café como se describe por Marracini *et al.*, 2003) o los promotores del gen para clorofila a/b

unión proteína (CAB) para la expresión en tejido fotosintético y los promotores del gen para glutamina sintetasa específicos de raíz en donde se desea la expresión en raíces.

La región codificante está unida operablemente también a una secuencia reguladora 3' apropiada. En modalidades en donde la secuencia reguladora 3' nativa no está en uso se puede utilizar la región de poliadenilación de nopalina sintetasa. Otras regiones reguladoras 3' útiles incluyen, pero no se limitan a la región de poliadenilación de octopina sintasa.

La región codificante seleccionada bajo el control de elementos reguladores apropiados se une operablemente a un marcador de resistencia de medicamento nuclear tal como resistencia a canamicina. Otros sistemas marcadores seleccionables útiles incluyen genes que confieren resistencia a antibióticos o herbicidas (por ejemplo resistencia a higromicina, sulfonilurea, fosfinotricina o glifosato) o genes que confieren crecimiento selectivo (por ejemplo fosfomanosa isomerasa, que permite el crecimiento de células vegetales en manosa). Los genes marcadores seleccionables incluyen, sin limitación, genes que codifican para resistencia a antibióticos tales como aquellos que codifican par neomicina fosfotransferasa II (NEO), dihidrofolato

reductasa (DHFR) e higromicina fosfotransferasa (HPT) así como genes que confieren resistencia a compuestos herbicidas tales como EPSPS resistente a glifosato u glifosato óxidoreductasa (GOX), *Bromoxynil nitrilasa* (BXN) para resistencia a bromoxinilo, genes AHAS para resistencia a imidazolinonas, genes de resistencia a sulfonilurea y genes de resistencia a 2,4-diclorofenoxiacetato (2,4-D).

En algunas modalidades, los promotores y otras secuencias reguladoras de expresión abarcadas por la presente invención están unidos operablemente a genes indicadores. Los genes indicadores contemplados para uso en la invención incluyen, pero no se limitan a genes que codifican para la proteína fluorescente verde (GFP) proteína fluorescente roja (DsRed), proteína fluorescente cyan (CFP), proteína fluorescente amarilla (YFP), proteína fluorescente naranja Cerianthus (cOFP) fosfatasa alcalina (AP), β-lactamasa, cloranfenicolacetiltransferasa (CAT), adenosina desaminasa (ADA), aminoglucósido fosfotransferasa (neo^r, $G419^r$) dihidrofolato reductasa (DHFR), higromicina-B-fosfotransferasa (HPH), timidina cinasa (TK), lacZ (que codifica para α-galactosidas) y xantina guanina fosforribosiltransferasa (XGPRT), β-glucuronidasa (gus), fosfatasa alcalina placentaria (PLAP), fosfatasa alcalina embriónica secretada (SEAP) o luciferasa de luciérnaga o bacteriana (LUC) respecto a muchos de los procedimientos

convencionales seleccionados con la práctica de la invención los expertos en la técnica tendrán conocimiento de secuencias adicionales que pueden servir para la función de un marcador o indicador.

Se conocen en la técnica modificaciones adicionales de secuencia para incrementar la expresión de genes en un hospedador celular. Estas modificaciones incluyen eliminación de secuencias que codifican para señales superfluas de poliadenilación, señales de sitio de empalme exon-intron, secuencias repetidas similares a transposón y otras de dichas secuencias bien caracterizadas que pueden ser perjudiciales para la expresión del gen. De manera alternativa, si es necesario, se puede ajustar el contenido de G/C de la secuencia codificante a concentraciones promedio para una célula hospedera de planta de café dada, calculado con referencia a los genes conocidos que se expresan en una célula de planta de café. También, cuando es posible, la secuencia codificante se modifica para evitar las estructuras de ARNm secundarias de orquilla predichas. Otra alternativa para incrementar la expresión de un gen es el uso de secuencias líder 5'. La secuencias líder de transducción son bien conocidas en la técnica e incluye el derivado en acción cis (omega') de la secuencia líder 5' (omega) del virus del mosaico de tabaco, las secuencias líder 5' del virus de mosaico de bromo,

virus de mosaico de alfalfa y virus del mosaico del nuevo amarillo.

Las plantas se transforman y posteriormente se criban para una o más propiedades, que incluyen la presencia del producto de transgen, ARNm que codifica para el transgen o un fenotipo alterado con expresión del transgen. Debe reconocerse que la cantidad de expresión así como el patrón de expresión específico de tejido y específico temporal de los transgenes en plantas transformadas puede variar dependiendo de la posición de su inserción en el genoma nuclear. Dichos efectos de posición son bien conocidos en la técnica. Por esta razón, varios transformantes nucleares se pueden regenerar y probar para expresión del transgen.

Métodos

Los ácidos nucleicos y polipéptidos de la presente invención se pueden utilizar en cualquiera de numerosos métodos por los cuales los productos de proteína se pueden expresar en plantas de café con el fin de que las proteínas puedan jugar un papel en la tolerancia a la tensión en la planta y en un mejoramiento en el sabor o aroma de la bebida de café o de los productos de café producidos finalmente a partir de la semilla de la planta de café que expresa la proteína.

Con respecto a la tolerancia a la tensión, ahora se ha establecido bien que las proteínas deshidrinas participan en la protección de plantas de tensiones osmóticas o ambientales tales como deshidratación y congelamiento (Allagulova *et al.*, 2003). Por ejemplo, las deshidrinas juegan un papel en la retención de agua y la regulación osmótica para proteger a la planta contra la pérdida de agua, especialmente en condiciones de sequía. Además, dado que las deshidrinas presentan la capacidad de interactuar con cationes, las proteínas deshidrinas pueden servir para unir el exceso de sales durante períodos de agua limitada y pueden servir con una función quelante. Las deshidrinas también juegan un papel en la integridad estructural de material nuclear tal como cromatina durante la desecación de la célula en la semilla en desarrollo y se ha encontrado que tienen un papel en la protección contra la formación de cristales de hielo y degradación de almidón en temperaturas de congelamiento. La función de una proteína de deshidrina dada se puede relacionar con la ubicación de la proteína dentro de la célula. Por ejemplo, las deshidrinas que se localizan hacia el exterior de la membrana celular pueden servir para estabilizar lípidos y proteínas de membrana (Allagulova *et al.*, 2003). Por lo tanto, la capacidad de manipular la producción de deshidrina o proteína LEA en una planta o incluso el uso de

polinucleótidos o proteínas de la invención para monitorear dicha expresión de gen permitirá el estudio y manipulación de la tolerancia osmótica a la sequía, al frío o a la concentración salina en el café. Dicha manipulación se puede extender desde la planta en germinación a la planta en crecimiento y al desarrollo de frutas y semillas, a la estabilidad al almacenamiento, después de la recolección, de los granos de café. Este conocimiento permite la generación de plantas de café modificadas que están mejor equipadas para un crecimiento sano y producción de cosechas bajo condiciones de tensiones osmóticas o ambientales prolongadas tales como las que se encuentran bajo períodos de sequía, carencia de agua, congelamiento o condiciones de congelamiento prolongadas. Por lo tanto, un aspecto de la invención describe métodos para proteger plantas, preferiblemente plantas de café al incrementar la resistencia a la tensión osmótica al modular la expresión de deshidrinas o proteínas LEA en la planta.

Con respecto al sabor y aroma de grano de café tostado, se espera que las deshidrinas y proteínas LEA relacionadas ejerzan cierta influencia en la generación de sabores de café vía la reacción de Maillard que se produce durante el tostado. Las proteínas, y particularmente los productos de degradación de proteína (péptidos y aminoácidos) representan un grupo importante de precursores

de sabor (Spanier *et al.*, 2004). Por lo tanto, las proteínas relativamente abundantes tales como las deshidrinas y proteínas LEA se puede esperar que representen cierta contribución a las secciones generadoras de sabor que se producen durante el tostado de café. En este contexto, es posible que estas proteínas relativamente no estructuradas y muy hidrofílicas y altamente solubles reaccionen de manera diferente de otras proteínas celulares durante la reacción de Maillard, ya sea debido a su estructura inusual o debido a sus interacciones inusuales con moléculas de agua. Es bien sabido que las concentraciones de agua presentes durante las reacciones de cocinado, tal como la etapa de tostado del café también pueden influir fuertemente en las vías de las reacciones químicas inducidas por calor (Turner *et al.*, 2002). Debido a que las deshidrinas contribuyen a la organización de moléculas de agua en el grano, una diversidad de diferencias específicas en las concentraciones y distribución de las deshidrinas puede alterar el desarrollo de sabor durante el procedimiento de tostado. La capacidad para monitorear (por ejemplo mediante crianza asistida por marcador) o manipulación de los perfiles de expresión de deshidrina y proteína LEA se proporcionan por los polinucleótidos de la presente invención de acuerdo con los métodos descritos en este documento.

Así, un aspecto de la presente invención describe métodos para alterar la deshidrina o el perfil de proteína LEA en una planta, preferiblemente café, que comprende incrementar o disminuir una cantidad de actividad de una o más deshidrinas o proteínas LEA en la planta. Por ejemplo, en una modalidad de la invención, un gen que codifica para deshidrina bajo el control de sus propias secuencias controladoras de expresión se utiliza para transformar una planta con el propósito de incrementar la producción de dicha deshidrina en la planta. De manera alternativa, una región codificante para deshidrina o proteína LEA se une operablemente a regiones controladoras de expresión heterólogas tales como promotores constitutivos o inducibles.

La organización de moléculas de agua o la estabilización de macromoléculas u organelos en el grano de un planta también se puede alterar al disminuir la producción de una o más deshidrinas o proteínas LEA en la planta o al cribar variantes que se presentan de manera natural para expresión disminuida de deshidrina o proteína LEA. Por ejemplo, se pueden generar plantas mutantes carentes de la función (nulificadas) o se pueden seleccionar de poblaciones de plantas mutantes disponibles actualmente. También se apreciará por aquellos expertos en la técnica que las poblaciones de plantas mutantes también

se pueden cribar en busca de mutantes que sobreexpresen una deshidrina particular, utilizando uno o más de los métodos descritos en la presente. Las poblaciones mutantes se pueden elaborar por mutagénesis química, mutagénesis por radiación y transposón o inserciones de ADN T o direccionamiento de lesiones locales inducidas en genomas (TILLING, véase, por ejemplo Henikoff *et al.*, 2004, Plant Physiol. 135(2): 630-636; Gilchrist & Haughn, 2005, Curr. Opin. Plant Biol. 8(2): 211-215). Los métodos para producir poblaciones mutantes son bien conocidos en la técnica.

Los ácidos nucleicos de la invención se pueden utilizar para identificar mutantes de deshidrina o proteína LEA en diversas especies vegetales. En especies tales como maíz o *Arabidopsis*, en donde están disponibles líneas de inserción de transposón, los cebadores oligonucleotídicos se pueden diseñar para cribar líneas para inserción en los genes para deshidrina o proteína LEA. Mediante crianza después se puede desarrollar una línea de plantas que sea heterocigota u homocigota para el gen interrumpido.

La planta también se puede manipular para mostrar un fenotipo similar al observado en mutantes nulificados criados por técnicas mutagénicas. Un mutante nulificado transgénico se puede generar por la expresión de una forma mutante de la deshidrina o proteína LEA seleccionada para criar un "efecto negativo dominante". Aunque la invención

no se limita a algún mecanismo en particular, esta proteína mutante competirá con la proteína natural para interactuar con las proteínas u otros factores celulares. Los ejemplos de este tipo de efecto "dominante negativo" son bien conocidos para los sistemas tanto de insecto como de vertebrados (Radke *et al.*, 1997, Genetics 145: 163-171; Kolch *et al.*, 1991, Nature 349: 426-428).

Otra clase de mutante nulificado transgénico se puede crear al inhibir la traducción de ARNm que codifica para deshidrina o proteína LEA mediante "bloqueo de expresión de gen postranscripcional". El gen que codifica para deshidrina o proteína LEA de las especies dirigidas para regulación por disminución o un fragmento de las mismas se pueden utilizar para controlar la producción de la proteína codificada. Para este propósito se pueden utilizar moléculas antisentido de longitud completa. De manera alternativa se pueden utilizar oligonucleótidos antisentido dirigidos a regiones específicas del ARNm que son críticas para la traducción. El uso de moléculas antisentido para disminuir los niveles de expresión de un gen predeterminado se conocen en la técnica. Las moléculas antisentido se pueden proporcionar *in situ* por células vegetales transformadas con construcción de ADN las cuales, por transcripción, producen las secuencias de ARN antisentido. Dichas construcciones se pueden diseñar para

producir secuencias antisentido de longitud completa o parciales. Este efecto de bloqueo de expresión del gen se puede incrementar por sobreproducción transgénica del ARN tanto directo como antisentido de la secuencia codificante del gen de manera que se produzca una alta cantidad de ARNds (por ejemplo véase Waterhouse *et al.*, 1998, Proc. Natl. Acad. Sci. U.S.A. 95: 13959-13964). A este respecto, las secuencias que contienen ARNds que corresponden a parte o la totalidad de por lo menos un intrón se han encontrado particularmente eficaces. En una modalidad, parte o la totalidad de la cadena antisentido de la secuencia codificante para deshidrina o proteína LEA se expresa por un transgen. En otra modalidad las cadenas directas y antisentido hibridizantes de parte o la totalidad de la secuencia que codifica para deshidrina o proteína LEA se expresan transgénicamente.

En otra modalidad, se puede bloquear la expresión de genes para deshidrina o LEA mediante el uso de una diversidad de otras técnicas post-transcripcionales para el bloqueo de expresión de genes (bloqueo de expresión de ARN) que actualmente están disponibles para sistemas vegetales. El bloqueo de expresión de ARN involucra el procesamiento de ARN de cadena doble (ARNds) en fragmentos pequeños de 21-28 nucleótidos por una enzima ARNasa basada en H ("Dicer" o "similar a Dicer"). Los productos de separación,

los cuales son ARNsi (ARN de interferencia pequeña) o ARNmi (micro ARN) se incorporan en los complejos efectores de proteína que regulan la expresión del gen de una manera específica de secuencia (para revisiones de núcleo de expresión de ARN en plantas, véase Horiguchi, 2004, Differentiation 72: 65-73; Baulcombe, 2004, Nature 431: 356-363; Herr, 2004, Biochem. Soc. Trans. 32: 946-951).

Los ARN de interferencia pequeña se pueden sintetizar o transcribir químicamente y se pueden amplificar *in vitro* y después se suministran a las células. El suministro puede ser a través de microinyección (Tuschl T *et al.*, 2002), transfección química (Agrawal N *et al.*, 2003), electroporación o transfección catiónica medida por liposomas (Brummelkamp TR *et al.*, 2002; Elbashir SM *et al.*, 2002), o cualquier otro medio disponible en la técnica, como lo apreciará una persona experta en la técnica. De manera alternativa, el ARNsi se puede expresar intracelularmente al insertar plantillas de ADN para ARNsi dentro de las células de interés, por ejemplo por medio de un plásmido (Tuschl T *et al.*, 2002) y se pueden dirigir específicamente a células seleccionadas. Los ARN de interferencia pequeños se han introducido con éxito en plantas (Klahre U *et al.*, 2002).

Un método preferido de bloqueo de expresión de ARN en la presente invención es el uso de ARN de horquilla

corta (ARNsh). Un vector que contiene una secuencia de ADN que codifica para una secuencia de ARNsi deseada particular se suministra en una célula objetivo por un medio común. Una vez en la célula, la secuencia de ADN se transcribe continuamente en moléculas de ARN que forman un bucle de regreso sobre si mismas y forman estructuras de horquilla a través de apareamiento de bases intramoleculares. Estas estructuras de horquilla, una vez procesadas por la célula, son equivalentes a moléculas ARNsi y se utilizan por la célula para mediar el bloqueo de expresión de ARN de la proteína deseada. Se describen por Horiguchi, 2004, *supra* diversas construcciones de utilidad particular para bloqueo de expresión de ARN en plantas. Típicamente, dicha construcción comprende un promotor, una secuencia del gen objetivo del cual se va a bloquear la expresión en la orientación "directa", un separador, el antisentido de la secuencia del gen objetivo y un terminador.

Otro tipo adicional de mutante nulificado sintético también se puede generar por la técnica de "cosupresión" (Vaucheret et al., 1998, Plant J. 16(6):651-659). Las células vegetales se transforman con una copia del gen endógeno dirigido para represión. En muchos casos, esto resulta en la represión completa del gen nativo así como del transgen. En una modalidad, el gen que codifica para deshidrina o proteína LEA de la especie vegetal de

interés se aísla y utiliza para transformar células de esa misma especie.

Las plantas mutantes o transgénicas producidas por cualquiera de los métodos anteriores también se describen de acuerdo con la presente invención. Preferiblemente, las plantas son fértiles por lo que son útiles para propósitos de crianza. De esta manera, los mutantes de las plantas que presentan uno o más de los fenotipos deseables mencionados antes se pueden utilizar para crianza de plantas o directamente en aplicaciones agrícolas u hortícolas. También son de utilidad como herramientas de investigación para la dilucidación adicional de la participación de las deshidrinas y proteínas LEA en el sabor, aroma y otras características de las semillas de café relacionadas con el contenido y organización del agua. Las plantas que contienen un transgen o una mutación especificada también se pueden cruzar con plantas que contienen un transgen o genotipo complementario con el fin de producir plantas con fenotipos mejorados o combinados.

La presente invención también describe composiciones y métodos para producir de una manera preferida para las semillas o específica para la semilla cualquier producto de gen heterólogo seleccionado en una planta. Una secuencia codificante de interés se coloca bajo

el control de un promotor para deshidrina o proteína LEA de café específica de semilla u otro promotor específico de semilla y otras secuencias reguladoras apropiadas para elaborar el gen quimérico específico de semilla. El gen quimérico se introduce en la célula vegetal por cualquiera de los métodos de transformación descritos en la presente o conocidos en la técnica. Estos genes quiméricos y métodos se pueden utilizar para elaborar una diversidad de productos de genes de interés en la planta que incluyen pero que no se limitan a: (1) productos de gen detectables tales como GFP o GUS, como se menciona en lo anterior; (2) productos de gen que confieren un beneficio agronómico u hortícola, tales como aquellos cuyas actividades de enzima resultan en la producción de micronutrientes (por ejemplo provitamina A, también conocido como β-caroteno) o antioxidantes (por ejemplo ácido ascórbico, ácidos grasos ω, licopeno, isoprenos, terpenos); o (3) productos de gen para controlar patógenos o plagas, tales como los derivados de Mourgues et al., (1998), TibTech 16:203-210 u otros conocidos como protectores de las semillas de plantas o dañinos para los patógenos.

Además, se pueden utilizar algunos de los promotores para deshidrina o gen LEA para producir proteínas recombinantes tanto en las semillas como en las silicuas. Además, dado que algunos de los genes para

deshidrina también se activan bajo condiciones de sequia u otras condiciones de tensión, estos promotores pueden demostrar ser útiles para la expresión directa del gen en otros tejidos tales como hojas maduras cuando son tensionadas osmóticamente. Esta última característica indica que es factible el uso de estos promotores para expresar proteínas recombinantes específicamente en las hojas de las plantas (por ejemplo tabaco) al final de la maduración conforme experimentan senescencia y comienzan a secarse.

Se considera que la deshidrinas son parte de la defensa de la planta contra la deshidratación. Por lo tanto, la inducción de los genes CcDH1 y CcDH2 se puede utilizar como una medida de la tensión de deshidratación existente en una planta; tanto el tiempo de inducción de la tensión de agua así como el nivel de tensión de agua. Por lo tanto, la expresión de deshidrina se puede utilizar para cribar poblaciones de plantas para determinar sus capacidades de respuesta ante la tensión osmótica.

Los siguientes ejemplos se proporcionan para describir la invención con mayor detalle. Se pretende que los ejemplos ilustren y no que limiten la invención.

Ejemplo 1

Material vegetal para extracción de ARN

Las raíces, hojas jóvenes, tallos, flores y frutos recién cosechados en etapas diferentes de desarrollo se obtienen de *Coffea arabica L. cv. Caturra T-2308* y *Coffea canephora* var. BP409 que crecen bajo condiciones de invernadero (25°C, 70% de humedad relativa) y también de *Coffea canephora* BP409 que crece en el campo en Java Oriente, Indonesia. Las etapas de desarrollo se definen como sigue: fruto verde pequeño (SG), fruto verde grande (LG), fruto amarillo (Y) y fruto rojo (R). Los tejidos frescos se congelan de inmediato en nitrógeno líquido, se almacenan a -80°C hasta que se utiliza para extracción de ARN.

Ejemplo 2

Protocolos para extracción de ARN total,

Generación de ADNc y condiciones de reacción por PCR

Las muestras de tejido almacenadas a -80°C se muelen hasta volverse un polvo y se extrae en la ARN total a partir de este polvo utilizando el método descrito previamente (Rogers et al., 1999). Las muestras son tratadas con DNasa utilizando el equipo "Qiagen RNase-Free DNase" de acuerdo con las instrucciones del fabricante para

eliminar la contaminación de ADN. Todas las muestras de ARN se analizan por electroforesis en gel de formaldehído agarosa e inspección visual de las bandas de ARN ribosómico por tinción con bromuro de etidio. Utilizando oligo (dT_{20}) como un cebador, se prepara ADNc a partir de aproximadamente 4 μg de ADN total de acuerdo con el protocolo en el equipo Superscript II Reverse Transcriptase (invitrogen, Carlsbad, CA). Para determinar la presencia de ADN genómico contaminante en las preparaciones de ADNc se diseña un par cebador que abarca un intrón conocido de un ADNc que se expresa de manera ubicua específico, calcona isomerasa. La ausencia del fragmento genómico en las reacciones de PCR indican la ausencia de contaminación de ADN genómico detectable. Las reacciones por PCR se llevan a cabo utilizando ADNc de *Coffera arabica* y *Coffea canephora* preparado como se describe en lo anterior. Los cebadores específicos de gen se describen en la tabla 1.

Tabla 1. Lista de cebadores utilizados para RT-PCR y RT-PCR cuantitativa

Nombre del oligonucleótido (SECUENCIA DE IDENTIFICACIÓN NÚMERO:)	Secuencia 5' --> 3'	Nombre del oligonucleótido	Secuencia 5'--<3'
Rp139_directo (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 23)	TGGCGAAGAAGCAGAGGCAGA	LEA1_directo (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 31)	CCAATAACAGCTCAAGAATCA
Rpl39_inverso (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 24)	TTGAGGGGGAGGGTAAAAAG	LEA1_inverso (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 32)	TTCCCCTTCCATCCCATCT
DH1_directo (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 25)	GAAGAAGGGGATGAAGGAG	rpl39_F1 (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 33)	GAACAGGCCCATCCCTTATTG
DH1_inverso (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 26)	TACGGACAAACACACTACAG	npl39_R1 (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 34)	CGGCGCTTGGCATTGTA
DH2_directo (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 27)	CCTCCAACAACCACCACTG	rpl39-MGB (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 35)	ATGCGCACTGACAACA
DH2_inverso (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 28)	TCAAGCGCACAACAAGGTC	DH2a-F1 (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 36)	GGGAGGCACAAGGACAGAGA
DH3_directo (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 29)	AGGTGGTGGTCAGAAGAAGAC	DH2a-R1 (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 37)	GCTGTGCGCGTGCTGAT
DH3_inverso (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 30)	GACACACTGGAAAGCTGCTA	DH2a-MGB (SECUENCIA DE IDENTIFICACIÓN NÚMERO: 38)	GAGGAGCACATCGAT

Se establecen reacciones de PCR (50 µl) que contienen 10 µl de una dilución en 100 veces de los ADNc excepto par CcDH2, en donde se utilizan 10 µl de una dilución de 1000 veces del ADNc. 1 µM de cada cebador, 5 µl de amortiguador ThermoPol 10X (New England Biolabs Beverly, MA), 1 µl de DMSO, 200 µM de dNTP y 2 unidades de

polimerasa Taq (New Engl and Biolabs Beverly, MA). Las condiciones de los ciclos son 2 min a 94°C, 35 ciclos (excepto para CcDH1 en donde se utilizaron 40 ciclos) de 94°C durante 1 minuto, 60°C durante 1 minuto y 72°C durante 1.5 minutos. La etapa de extensión final es durante 7 minutos a 72°C. Los productos de RT-PCR se resuelven en geles de agarosa 2% (p/v) y se tiñen con bromuro de etidio. El gen CcRL39, el cual codifica para la proteína L39 de café que se expresa de manera constitutiva (una proteína de subunidad grande ribosómica 60S) se utiliza como el control semicuantitativo para verificar que cada muestra de ARN ha sido transcrita en ADNc a eficiencias relativamente similares. Se utiliza como control positivo la amplificación del gen para RPL39, para la transcripción inversa con los cebadores que se muestran en la tabla 1.

Se llevan a cabo las pruebas cuantitativas TaqMan-PCR de CcDH2 y CcRPL139 de acuerdo con el protocolo del fabricante (Applied Biosystems, Perkin-Elmer) utilizando ADNc de *Coffea arabica* y *Coffea canephora* y las sondas TaqMan que se muestran en la tabla 1. Reacciones de 25 μl que contienen 12.5 μl de TaqMan^{MR} Universal Master Mix 2X, 20 nM de TaqMan^{MR}-MGB probe, 80 nM de cebadores específicos TaqMan^{MR} y 5 μl de ADNc diluido 1000 veces. Las condiciones de los sitios son 50°C durante 2 minutos, 95°C durante 10 minutos y después 40 ciclos de 94°C durante 15

segundos y 60°C durante 1 minuto. Cada reacción se repite 3 veces. La expresión del gen para DH2 en cada muestra de ADNc se normaliza para la expresión del gen RPL39 en la misma muestra.

Ejemplo 3

Protocolo para aislamiento de la región promotora DH2

La secuencia promotora de CcDH2 se aísla utilizando el equipo Genome Walker de acuerdo con las especificaciones del fabricante (GD Sciences-Clontech). El ADN genómico se aísla de *Coffea canephora* var. BP409 como se ha descrito (Crouzillat et al., 1996). El cebador de Genome Walker directo específico para CcDH2 utilizado es cebador DH2a 1-5' TGTGCTCCTGATGCTCTCTGTCCTTGTGC 3' (SECUENCIA DE IDENTIFICACION NÚMERO: 39). Se aísla un fragmento de aproximadamente 2.1 kb utilizando ADN genómico de *coffea canephora* var. BP409 digerido con HindIII, ligado a la secuencia adaptadora Genome Walker. La amplificación por PCR se lleva a cabo en una reacción de 50 μl utilizando el equipo Clontech Advantage 2 PCR de acuerdo con el protocolo del fabricante utilizando concentraciones finales de 0.5 μM del cebador 1 de DH2a y el cebador Genome Walker AP1. La reacción por PCR se lleva a cabo bajo las siguientes condiciones: 94°C durante 2 segundos y 72°C 3 minutos (7 ciclos), 94°C durante 2 segundos y 67°C 3

minutos (32 ciclos) seguido por 4 minutos a 67°C. El fragmento principal de PCR obtenido después se clona en el plásmido pCR4-TOPO (Invitrogen).

Se purifica un plásmido pJMc1 que contiene el inserto apropiado y el inserto se secuencia por completo. Para verificar que los insertos en pJMc1 y (pcccs30w8a4) provienen del mismo gen, la secuencia superpuesta correspondiente de estos dos clones se vuelve a amplificar a partir de ADN genómico utilizando los cebadores DH2a geneup 5' ATAGTGACCTTAATAGCGATCTTGTTGC 3' (SECUENCIA DE IDENTIFICACION NÚMERO: 40) y DH2a genelow 5' CCAAATCAAATCAAACCAAGCAAATC 3' (SECUENCIA DE IDENTIFICACION NÚMERO: 41). La reacción de PCR se realiza con el ADN genómico de *coffea canephora* var. B409 y utilizando Taq (New Englands Biolabs) y 1 μM de cebadores específicos (DH2a geneup y DH2a genelow). La reacción de PCR se lleva a cabo con las siguientes condiciones: 94°C durante 1 minuto después 35 ciclos de 94°C 1 minuto, 58°C, 1.5 minutos y 72°C 3 minutos, seguido por 7 minutos a 72°C. El fragmento principal de PCR producido después se clona en pCR4-TOPO. Un plásmido pVC1 (figura 7) que contiene el inserto apropiado se purifica y su inserto se secuencia completamente. Existen 5 cambios de bases entre los fragmentos genómicos de pJMc1 y pVC1. Un cambio es en la región promotora, dos cambios son el intrón y dos cambios

son en la secuencia codificante de la proteína. De estos cambios en la secuencia codificante de la proteína, un cambio es neutro y el otro resulta en un aminoácido diferente.

Ejemplo 4

Protocolo de transferencia Southern

Se prepara ADN genómico como se ha descrito previamente (Crouzillat et al., 1996). Se digieren durante la noche 5 µg de ADN genómico de ADN *C. canephora* BP 409 con las enzimas apropiadas (10 U/µg) de acuerdo con las recomendaciones del proveedor y los productos se separan en geles de agarosa 0.8%. Se lleva a cabo la transferencia Southern e hibridaciones como se ha descrito previamente (Crouzillat et al., 1996). La sonda se genera al amplificar primero por PCR el inserto del clon CcDH2 cccs30w8a4 con los cebadores T3 + T7. Este producto de PCR se marca con [^{32}P]dCTP utilizando el equipo "rediprime^{MR} II random prime labeling systems" (Amersham).

Ejemplo 5

Identificación y caracterización del ADNc para deshidrina de café

Se generan más de 47,000 secuencias EST a partir de varias bibliotecas de café elaborados con ARN aislado de

hojas jóvenes y a partir de tejido de grano y pericarpo de cerezas cosechadas en etapas de desarrollo diferentes. Las EST que se superponen posteriormente se "agrupan" en "genes únicos" (es decir, secuencias contiguas) y las secuencias de genes únicos se anotan al realizar una investigación BLAST de cada secuencia individual contra una base de datos de proteína no redundante NCBI. Los genes únicos se criban para secuencias de deshidrina utilizando diversos enfoques que incluyen una búsqueda de las anotaciones de gen único con la palabra clave "deshidrina" y mediante la utilización de diversas secuencias de proteína para deshidrina de *Arabidopsis* y tomate en una búsqueda tBlastn del conjunto de gen único para café. Los diversos protocolos de búsqueda proporcionan diversos genes únicos candidatos para deshidrina y el clon de ADNc potencialmente más largo para cada uno de estos genes únicos se aísla a partir de la biblioteca y se secuencía por completo (tabla 2).

Tabla 2. ADNc de longitud completa de *Coffea canephora* que codifica para deshidrina y proteínas LEA y el número de EST encontrados para cada secuencia. Los números de genes únicos y los pesos moleculares se proporcionan entre paréntesis para cada plásmido: ccc126i7 (Unigene 121870; 17.8 kDa); ccCs30w27m8 (Unigene 121870; 18.1 kDa); cccs30w8a4 (Unigene 123406; 17.4 kDa); cccs46w30p1 (Unigene

123405; 17.4 kDa); cccwc22w11a5 (Unigene 123385; 25.1 kDa); Dav1-59 (Unigene 119994; 39.5 kDa).

Nombre del plásmido	Nombre del gen	Estructura	Expresión in silico (número de EST en bibliotecas)						
			Cerezas completas 22 semanas	Grano, 18 semanas	Grano 30 semanas	Grano 42 semanas	Grano 46 semanas	Pericarpio	Hoja
ccc126i7	CcDH1a	Y_3SK_2		1	6		18		18
ccc30w27m8	CcDH1b								
ccc30w8a4	CcDH2a	Y_3SK_2			20		3		1
ccc46w30p1	CcDH2b				1		1		
cccwc22w11a5	CcDH3	SK_3	2				7		3
Dav1-59	CcLEA1	proteína LEA			22				

El análisis de secuencia de ADN de los clones de ADNc de longitud completa seleccionados muestran cuatro secuencias únicas que representan tres genes diferentes para deshidrina. Los genes correspondientes se denominan CcDH1, CcDH2 y CcDH3. Dos secuencias aparentemente alélicas del gen CcDH1 se identifican por secuenciado de los dos ADNc más largos en el gen único #121870. Los clones de ADNc para CcDH1a y CcDH1b son de 836 y 896 pares de bases de largo, respectivamente. Las dos secuencias con ORF presentan cinco cambios de base únicos y CcDH1b presenta una inserción de 9 bases. Estas diferencias se traducen en seis diferencias en aminoácidos. CcDH1a y CcDH1b codifican para proteínas de 172 aminoácidos y 175 aminoácidos los

cuales tienen pesos moleculares predichos de aproximadamente 17.8 kDa y 18.1 kDa, respectivamente. Se encontró que dos genes únicos distintos, #123406 y #123405 codifican para el ORF para CcDH2. El gen único "123405 (CcDH2b) está compuesto de 2 EST y difiere de la secuencia del gen único #123406 (CcDH2a) por la presencia de una secuencia de intrón. Cuando se examina con mayor detalle los límites intrón/exón del ADNc par DH2b (cccs46w30p1), se observa que la unión 3' tiene la secuencia ttatgg/TCG mientras que otras secuencias genómicas obtenidas mediante paseo genético (genome walkin véase abajo) tiene la secuencia ttatag/T(A)CGG. En general, las secuencias de intrón del ADNc cccs46w30p1 y las secuencias de intrón en el ADN genómico son casi idénticas. Debido a que ninguno de los cambios únicos aparece en la totalidad de las tres secuencias de intrón genómico disponibles, parece que una alteración en la secuencia de sitio de empalme 3' de CcDH2b puede ser la causa del empalme aberrante de este ADNc. El ADNc de 756 pares de bases pcccs30w8a4 (CcDH2a) codifica para una proteína de 162 aminoácidos de largo con un peso molecular predicho de 17.4 kDa. Se encuentra un gen único para CcDH3 (#123385). La secuencia proteínica codificada por CcDH3 demuestra que este ADNc de 833 pares de bases codifica para una proteína de 227 aminoácidos con un peso molecular aproximado predicho de 25.1 kDa.

Las secuencias de proteínas de las deshidrinas de café se alinean con las secuencias de proteínas más homólogas que se encuentran en la base de datos de proteínas no redundantes y también se analizan para determinar la presencia de motivos de aminoácidos específicos para deshidrina Y, S y K. Las figuras 1 y 2 muestran que las tres deshidrinas se encuentran en dos clases, en donde CcDH1a, CcDH1b y CcDH2a tienen la estructura Y_3SK_2 y CcDH3 tiene la estructura SK_3. Una de estas secuencias proteínicas son la estructura Y_3SK_2, CcDH1a y CcDH1b muestran conservación absoluta en cada uno de los tres motivos, así como en las dos regiones conservadas que proceden cada uno de los dos motivos K. Esta observación, y el hecho de que las diferencias de secuencia pequeñas que existen se presentan en por lo menos las regiones conservadas de las secuencias alineadas concuerda con la idea de que CcDH1a y CcDH1b son alélicos. En contraste, CcDH2a es claramente diferente de CcDH1. Aunque CcDH2a tiene la estructura Y_3SK_2, también presenta diferencias puntuales en la totalidad excepto en los motivos Y, S y K así como diferencias más significativas fuera de estos motivos específicos para deshidrina. Las proteínas que codifican para CcDH1 y CcDH2 se calculan sus pI que están cerca de la parte neutra y sus gráficas de hidrofilicidad indican que estas proteínas son muy hidrofílicas en su

secuencia. El pI calculado de la proteína codificada por CcDH2 es ligeramente ácido (5.47) y esta proteína también es muy hidrofílica, como se muestra por la gráfica de hidrofilicidad de Kyte-Doolittle en la figura 8.

Ejemplo 6

Caracterización de ADNc que codifica para la proteína LEA de café

Se construye una biblioteca de ADNc a partir de ARN preparado a partir de granos de café 30 semanas después de la fertilización. A partir de esta biblioteca se aísla y secuencia un clon de ADNc de longitud completa (Dav1-59) que codifica para una proteína LEA. Este clon de ADNc se le cambia el nombre a CcLEA1. Además, la base de datos EST se examina en busca de "genes únicos" anotados como proteínas LEA. Esta búsqueda produjo 9 secuencias de genes únicos, uno de los cuales (gen único #119994) corresponde a la secuencia aislada previamente de CcLEA1 (tabla 3). El análisis de EST indica que CcLEA1 se expresa de manera fuerte y exclusiva durante solo un período del desarrollo del grano (30 semanas después de la fluoración).

Tabla 3. Secuencias del gen único de *Coffea canephora* anotadas como proteínas LEA

Nombre del gen único	Número de los EST	Expresión in silico (número de EST en bibliotecas)						
		Cerezas completas 22 semanas	Grano, 18 semanas	Grano 30 semanas	Grano 42 semanas	Grano 46 semanas	Pericarpio	Hoja
124784	26			17		9		
124426	26					9		17
119994	22			22				
124428	10	1				3	4	2
121426	8		1	5		2		
124123	5	3					1	1
120926	2					2		
131122	1					1		
127677	1					1		

La proteína codificada por CcLEA1 es de 358 aminoácidos y tiene un peso molecular predicho de 39.5 kDa. El pI calculado para CcLEA1 es ligeramente básico (8.17) y una gráfica de hidrofilicidad muestra que aunque esta proteína no tiene regiones significativas de hidrofobocidad, es menos hidrofílica que las tres deshidrinas de café. Los primeros 30 residuos de la parte N terminal de CcLEA1 forman una de dos regiones hidrofóbicas pequeñas en esta proteína. La figura 3 muestra la alineación de CcLEA1 con las tres secuencias más homologas que se encuentran en la base de datos de proteínas de GenBank no redundantes. Los valores de identidad total de estas secuencias alineadas únicamente varían desde 34.7%

para la secuencia de *Arabidopsis* a 47.9% para la secuencia de *Picea* (abeto blanco). No obstante, existen regiones cortas altamente conservadas en estas proteínas. La totalidad de las secuencias de proteína relacionadas tienen perfiles de hidrofilicidad relativamente similares a los de CcLEA1 y generalmente el parche hidrofóbico más significativo de las proteínas se puede encontrar en los aminoácidos 1-25 de la parte N terminal. También se encontró que CcLEA1 contiene un segmento rico en prolina en la región N terminal el cual no se encuentra en las otras proteínas en la figura 3.

Ejemplo 7

Expresión por RT-PCR de los genes para CcDH1, CcDH2, CcDH3 y CcLEA-1 en diferentes tejidos de café durante el desarrollo del gano de café

Debido a que las bibliotecas EST no están normalizadas y fueron muestreadas profundamente, el número de EST encontrados en cada gen único proporciona un cálculo aproximado del nivel de expresión de ese gen en cada tejido muestreado. La tabla 2 (anterior) muestra que CcDH1 y CcDH2 se expresan fuertemente en el grano a las 30 y 46 semanas después de la fertilización pero no se detectan en la biblioteca de pericarpio ni se detectan en las cerezas completas (grano en desarrollo + pericarpio) a las 22

semanas después de la fertilización. Tanto CcDH1 como CcDH2 se expresan en las hojas, aunque CcDH1 se puede expresar en hojas jóvenes a una concentración mayor que CcDH2. La expresión de CcDH3 se detecta en el grano a las 46 semanas después de la fertilización y en hojas jóvenes, aunque también se observaron 2 EST en muestras de cerezas completas de 22 semanas.

Para extender los datos de expresión, se lleva a cabo un análisis por RT-PCR para cada uno de los tres genes para deshidrina de café. Los resultados de este análisis se muestran en la figura 4. CcDH1 se expresa de manera significativa en granos de arabica y la totalidad de las etapas examinadas y en las tres últimas etapas examinadas para robusta. La expresión de CcDH1 también se puede detectar en otros tejidos probados, aunque no se detecta señal para arabica en el pericarpio verde pequeño y las muestras de pericarpio amarillo o para robusta en las muestras de raíz y hoja. Entre los tejidos que tienen expresión de CcDH1, la muestra de flor de arabica parece tener la concentración más alta de transcritos.

Se detecta un nivel relativamente alto de transcritos CcDH2 en todas las etapas de desarrollo de los granos de la especie arábica y por lo menos tres etapas de los granos del género robusta (figura 4). En contraste con CcdH1, no se detectaron transcritos CcdH2 por RT-PCR en los

otros tejidos estudiados. La ausencia de niveles significativos de transcritos CcDH2 en tejidos diferentes a los de grano, así como la inducción posterior de este gen en robusta se confirma por RT-PCT cuantitativa TaqMan (figura 5). Se compara la expresión de CcDH2 con el transcrito expresado de manera constitutiva del gen para CcRPL39 (RPL39 codifica para la proteína #39 de la subunidad ribosómica grande). La comparación demuestra que la expresión de CcDH2 incrementa gradualmente en robusta, comenzando desde la etapa verde grande hasta la etapa roja madura. En contraste, los datos de RT-PCR cuantitativos para arábica muestran que el nivel más alto de los transcritos de CcDH2 se detecta en la etapa verde grande y que los niveles de transcrito disminuyen en cierta medida conforme progresa la maduración.

El análisis por RT-PCR de la expresión del gen para CcDH3 demuestra que los niveles significativos de estos transcritos también se detectan en todas las muestras de grano arábica así como en las últimas tres etapas del desarrollo de grano robusta (figura 4). Los niveles de los transcritos CcDH3 en parte de los otros tejidos, tal como el pericarpio rojo, el tallo y las flores es casi tan alto como en el grano.

El examen de los datos originales muestra que los transcritos de CcDH3 se pueden detectar en la totalidad de

los otros tejidos arábica y robusta examinados. No obstante, se observa una diferencia significativa entre los niveles de transcrito para las primeras tres etapas de pericarpio de robusta y arábica.

La expresión de CcLEA1 también se evaluó por RT-PCR. Los datos obtenidos confirman que este gen presenta un patrón de expresión único, con transcritos que se detectan únicamente en la etapa verde pequeña de arábica y la etapa verde grande del grano de robusta (figura 4). No se detecta expresión en cualquier otro de los tejidos de arábica o robusta muestreados. Estos datos concuerdan con el patrón de distribución de EST para este gen que se observa en la tabla 3, lo cual indica que este gen se expresa en granos de robusta a 30 WAF pero no en ninguna otra de las bibliotecas EST de grano, cereza, pericarpio u hoja.

Ejemplo 8

Análisis de Secuencia del Promotor CcDH2

CcDH2 es un gen específico de grano que se expresa de manera moderada. Se lleva a cabo transferencia Southern para determinar si el nivel de expresión que se obtiene de un gen con un número de copia único/bajo o un gen de copia múltiple. Como se muestra en la figura 6, cada digestión produce únicamente una banda, lo que sugiere que es probable que CcDH2 sea codificado por un gen único en el

genoma de *C. canephora*.

Se utiliza la técnica de paseo de cebador para aislar un fragmento genómico que incorpora el promotor CcDH2. La amplificación por PCR del ADN se lleva a cabo utilizando un cebador Genome Walker específico para CcDH2 diseñado desde el centro del ADNc cccs30w8a4 (cebador 1 de DH2a) y el cebador AP1 del equipo Genome Walker. Se genera y clona un fragmento genómico de 2.1 kb que abarca 1.43 kb corriente arriba de la secuencia de ADNc cccs30w8a4 *(C. canephora)*. El análisis de secuencia de la secuencia compuesta obtenida a partir de los plásmidos superpuestos que contienen secuencias de ADNc y genómico muestran que el gen CcDH2 contiene un intrón único (231 pares de bases) que se localiza dentro de la región de ORF (figura 7). Un análisis adicional de la región promotora de este gen indica la presencia de una secuencia TATA putativa de 30 pares de bases corriente arriba del extremo 5' de la secuencia de ADNc.

Varios elementos reguladores potenciales muestran propiamente que están involucrados en la regulación de la expresión de un gen durante el desarrollo de semillas y también se identificaron en la región hacia el extremo 5' del gen para CcDH2. Por ejemplo se encontraron tres regiones con similitud con el elemento sensible ABA de Arabidopsis RYACGTGGYR (SECUENCIA DE IDENTIFICACION NUMERO:

42). Se encontraron dos elementos que comparten similitud significativa con la secuencia repetida RY (CATGCA (T/a)(A/g) de la región del núcleo en la caja "legumina" que están involucrados en la regulación de la expresión de las proteínas de almacenamiento de tipo legúmina. La presencia de dos elementos sensibles a deshidratación/secuencia en acción cis C-repetida (DRE/CRT) de los motivos (G/ACCGAC) también se identificaron. Los motivos DRE/CRT se ha demostrado previamente que interactúan con los factores de transcripción DREBs/CBF para controlar la respuesta a genes relacionados con la deshidratación y otras tensiones en arabidopsis y arroz. (Dubouzet JG, Plant J. 33: 751-763 (2003)). Además, varios motivos de caja E (CANNTG), los cuales son componentes bien definidos en los promotores de proteína de almacenamiento tal como la proteína 2S (Chatthai M, Plant Physiol Biochem 42: 417-423 (2004)), se identificaron en la región promotora CcDH2.

Ejemplo 9

Análisis Funcional del Promotor CcDH2 de deshidrina de Café en Arabidopsis thaliana

Se examinó la expresión en *arabidopsis thaliana* de un gen indicador que codifica para β-glucurodinasa (GUS) bajo el control del promotor CcDH2.

Materiales y Métodos:

La secuencia promotora de DH2 de deshidrina a partir de pVC1 se amplifica utilizando polimerasa Pfu1 bajo las condiciones descritas por el proveedor (Stratagene) y los cebadores:

TG - TG698 ttgaagcttGTGGACATGACGGAAGAGGT (SECUENCIA DE IDENTIFICACION NUMERO: 43) y

TG - TG743 gcagatctaccatggAGGACTCCTGTTATTAGAAAA (SECUENCIA DE IDENTIFICACION NUMERO: 44).

El fragmento de PCR obtenido de esta manera después se corta con HindIII y BgIII y se clona en los sitios HindIII/BglII del vector de transformación vegetal pCAMBIA1301. Esto coloca el fragmento de aproximadamente 1.3 kb que contiene la secuencia promotora de deshidrina y la región 5' no traducida completa del ADNc para deshidrina (aproximadamente 80 pares de bases) dentro de 2 pares de bases de ATG para el GUS (primer exon de GUS). Se verifica la colocación correcta del promotor por secuenciado. El nuevo promotor de deshidrina CcDH2 que contiene el vector se denomina pCAMBIA1301UCD2.4.

Transformación de planta. El vector de transformación pCAMBIA1301UCD2.4 después se transforma en *Agrobacterium tumefaciens* cepa EHA105 utilizando procedimientos estándar. El gen de resistencia a

higromicina, activado por el promotor 2 x 35S, es el marcador seleccionable vegetal en pCAMBIA1301. La transformación mediada por *Agrobacterium tumefaciens* de *Arabidopsis* (con el plásmido pCAMBIA1301UCD2.4) se realiza por el método de goteo floral (Clough and Bent, 1998).

Las plantas transformadas se identifican por siembra en placa en agar 0.8% que contiene nitrato de sodio 1 mM y 50 μg por ml de higromicina. Las plántulas transformadas se identifican 7 días después de la siembra como plantas con una raíz primaria extendida. Las plántulas se transfieren a agar 0.8% que contiene 0.5X de sales M&S. Las plantas posteriormente se transfieren al suelo cuando el segundo par de hojas se desarrolla y se permite que maduren y que produzcan semillas (T1). En algunos casos, las semillas T1 se hacen germinar y después se permite que crezcan y que produzcan semillas (T2).

Tinsión con GUS. Las plántulas y silicuas examinadas para tinsión con GUS son de las semillas T1 o T2 y se encuentran en etapas diferentes de desarrollo. La solución de tinsión GUS se prepara al disolver 5 mg de X-Gluc en 50 μl de dimetilformamida y después al agregar esto a 10 ml de $NaPO_4$ 50 mM, pH 7.0. Con un fórceps fino, se transfieren las plántulas de las placas de germinación a un tubo de microcentrífuga de 1.5 ml que contiene 1.0 ml de tinsión GUS. Los tubos se transfieren a un desecador y se

colocan bajo vacío durante 10 minutos y se incuban a 37°C (en la oscuridad) durante 24 ó 48 horas. La mancha se retira y se sustituye con una solución de eliminación de manchado (EtOH 70%). Se acelera el aclarado al colocar los tubos a 37°C. Dependiendo de la cantidad de pigmento en el tejido se requieren varios cambios de EtOH 70%. Las plántulas manchadas y otros tejidos se observan bajo un microscopio de disección y las imágenes se registran digitalmente. En el caso de las silicuas, las silicuas se retiran de las plantas y se abren con un bisturí para permitir la penetración de la mancha. La mancha GUS utilizada en el procedimiento se modifica para incluir Triton X100 0.5%. Después de la tinsión, las silicuas se destiñen al incubarlas en EtOH:ácido acético (2:1) y después se incuba en el medio Light de Hoyer (hidrato de cloral 100 g en 60 ml de agua). Las silicuas con semillas más jóvenes se preincuban en una solución de etanol:ácido acético durante 4 horas y con semillas más viejas durante 8 horas. Las silicuas se aclaran en medio Light de Hoyer durante 24 horas a varios días.

Resultados:

Se observa la expresión de GUS en *Arabidopsis thaliana* transformada con pCam1301UCD2-4. Se encuentra que la expresión de GUS es abundante en cotiledones y en el hipocotilo de las plántulas de una semana de edad. En

plántulas de dos semanas de edad, la tinsión con GUS aún es abundante en los primeros dos cotiledones y en un nivel menor en las primeras hojas verdaderas. La actividad de GUS no se detecta de manera significativa en la raíz o en el segundo par de hojas en desarrollo. No se detecta expresión en hojas maduras. La expresión de GUS también se detecta en la pared de silicua y en las semillas en desarrollo. Una tinsión con GUS de 48 horas de las semillas de una línea T1 resulta en que algunas semillas son positivas para actividad de GUS y otras son negativas.

Resumiendo, los datos presentados en este ejemplo confirman que el promotor CcDH2 de deshidrina de café activa la expresión del gen relacionado (en este caso, GUS) fuertemente en semillas, silicuas y en los primeros cotiledones e hipocotilos de las plantas en germinación. Este resultado demuestra que la secuencia promotora CcDH2 descrita aquí contiene todos los elementos funcionales necesarios para impulsar la expresión del gen específico en plantas. Los datos obtenidos también indican que el promotor CcDH2 se puede utilizar para impulsar la expresión de genes en tejidos inmaduros tales como embriones derivados de los primeros dos cotiledones de plántulas. Además, los datos indican que el promotor CcDH2 se activa en otros tejidos destinados a experimentar desecación tal como las silicuas. Finalmente, dada la distancia evolutiva

relativamente grande entre *Arabidopsis* y *Coffee*, los datos presentados aquí muestran que el promotor CcDH2 de *Coffee* funciona en *Arabidopsis*, implica que este promotor debe ser activo en una variedad relativamente amplia de la plantas.

Ejemplo 10

Expresión inducida por tensión osmótica de genes que codifican para las deshidrinas CcDH1 y CcDH2 de café

Se sabe que los genes para deshidrinas son inducidas bajo formas diferentes de tensión osmótica. Por lo tanto se llevó a cabo una evaluación para determinar si los genes para CcDH1 y CcDH2 de café son inducidos por tensiones osmóticas diferentes.

Materiales y Métodos:

Se llevaron a cabo experimentos de deshidratación utilizando árboles catimor de *Coffea arabica* propagados clonalmente pequeños que han crecido en un invernadero. Los árboles son de aproximadamente 3 años de edad y han crecido en el suelo. Varias semanas antes de los experimentos los árboles se cultivan juntos en el invernadero con una temperatura de aproximadamente 25°C con una humedad relativa de aproximadamente 70% y se les suministra agua diariamente utilizando irrigación automática. Al inicio del experimento tres árboles fungieron como controles y fueron

tratados diariamente. Los otros tres árboles no recibieron agua y por lo tanto experimentaron deshidratación progresiva. Se llevó a cabo un muestreo de dos hojas jóvenes (con un tamaño de 5-8 cm y se tomaron de crecimiento que germina en la parte superior de la planta) cada semana para cada árbol y las muestras se congelaron directamente en nitrógeno líquido.

Extracción de ARN y síntesis de ADNc. La extracción de muestras de tejido sometidas a diversos tratamientos de tensión en los controles se realizaron utilizando el miniequipo RNEASY^{MR} Plant de Qiagen GmbH (Hilden, Alemania). Las muestras de tejido congelado inicialmente se muelen en un almirez y mano utilizando nitrógeno líquido con el fin de obtener un polvo. El ARN en este polvo congelado después se extrae de acuerdo con el protocolo de miniequipo RNEASY^{MR} Plant. Brevemente, se mezcla un máximo de 100 mg de polvo congelado con el amortiguador de lisis celular y β-mercaptoetanol. Para los tejidos que mostraron necrosis significativa también se agregó PMSF 2 μm. Con el fin de eliminar concentraciones bajas de ADN genómico contaminante se utilizó un tratamiento con desoxirribonucleasa libre de fibronucleasa contenido en el miniequipo RNEASY^{MR} Plant (como se describe por el proveedor), es decir, un tratamiento de 15 min a temperatura ambiente en la columna. Al final el ARN se

eluye de la columna en 50 µl de agua libre de ribonucleasa. Se determina la cantidad de ARN por medición espectrofotométrica a 260 nm y se calcula la calidad de ARN al calcular la relación de absorbancia 260 nm/280 nm. La calidad de los ARN también se verifica por electroforesis en geles de agarosa 1%. Las secciones de transcripción inversa para estas muestras ARN se llevan a cabo como sigue: aproximadamente 1 µg de ARN total y 12.4 µM de oligo-dT [2.3 µl de oligo-dT 70 µM (Proligo)] con agua libre de ribonucleasa hasta un volumen final de 13 µl. Esta mezcla se incuba a 65°C durante 5 min. Después se agregan 7 µl de una mezcla de amortiguador 5X (amortiguador de reacción TRANSCRIPTOR^{MR} RT) 20U de inhibidor de ribonucleasa, 1 mM de las cuatro dNTP (250 µm de cada una) y 10U de transcriptasa inversa TRANSCRIPTOR^{MR} (Roche, Nutley, NJ). Esta mezcla se incuba a 55°C durante 40 minutos, después se agregan 0.5 µl de ribonucleasa H (Invitrogen, Carlsbad, CA) a la mezcla de 20 µl y la reacción se incuba adicionalmente durante 30 minutos a 37°C. Los ADNc que se generan se purifican utilizando el equipo de purificación SNAP^{MR} Gel de Invitrogen (Carlsbad, CA) de acuerdo con las instrucciones proporcionadas por el proveedor.

Cebadores y diseño de la sonda MGB. Los conjuntos de cebadores y sonda MGB se diseñan utilizando el programa

SB

PRIMER EXPRESS^{MR} (Applied Biosystems, Foster City, CA). La temperatura de hibridación de los cebadores son de aproximadamente 60°C mientras que la temperatura de la sonda MGB es cercana a los 70°C. El tamaño de los amplicones es de aproximadamente 80 pares de bases. Los cebadores se sintetizan por PROLIGO y las sondas MGB se sintetizan de acuerdo con las instrucciones del proveedor (Applied Biosystems, Foster City, CA). Las secuencias de los cebadores y las sondas para CcDH2 y Ccrpl39 se han presentado en lo anterior en la tabla 1. Los cebadores para CcDH1 son 5'CACTGGCACTACTGGAGCCTATG 3' (SECUENCIA DE IDENTIFICACION NUMERO: 45) y 5' GCTGGGTGGCGTATGCA 3'. La sonda MGB para CcDH1 es 5' CTGGAGCACATGGGA 3' (SECUENCIA DE IDENTIFICACION NUMERO: 46).

RT-PCR cuantitativa en tiempo real. El ADNc utilizado para estos experimentos se prepara como se describe en lo anterior. Se lleva a cabo TaqMan-PCR como lo indica el fabricante (Applied Biosystems, Perkin-Elmer) y como se describe en lo anterior. Brevemente, todas las reacciones tienen un volumen de 25 μl y contienen 5 μl de ADNc, amortiguador TaqMan 1x (Applied Biosystems), $MgCl_2$, 5 mM, 200 μM de cada uno de dATP, dCTP, dGTP y dUTP y 0.625 unidades de ADN polimerasa Amplitaq Gold. El amortiguador de reacción de Applied Biosystems contiene AmpErase^{MR} UNG (Uracilo-N-glucosilasa) y colorante de referencia pasivo

(ROX TM) y componentes de amortiguador optimizados. Las reacciones de PCR se llevan a cabo utilizando 800 nM de los cebadores específicos de gen, directo e inverso y con 200 nM de la sonda TaqMan y 5 μl de una dilución 100 veces de ADNc, lo cual corresponde a aproximadamente 0.01 μg de ARN total. Las reacciones se incuban durante 2 minutos a 50°C, después 10 minutos a 95°C seguido por 40 ciclos de amplificación durante 15 segundos y a 95°C/1 minuto a 60°C. Las reacciones se llevan a cabo y se analizan utilizando un sistema de detección GeneAmp 7500 Sequence (Applied Biosystems). Cada muestra se realiza 3 veces y se calcula el valor promedio. La cuantificación se lleva a cabo utilizando el método de cuantificación relativa en donde el ARN expresado de manera constitutiva para la proteína ribosómica rpl39 que actúa como referencia interna en cada muestra. Con el fin de utilizar el método de cuantificación relativa, es necesario mostrar que la eficiencia de amplificación para las secuencias de gen de prueba diferentes son equivalentes de manera general a la eficiencia de amplificación de la secuencia de referencia (secuencia de ADNc para rpl39) utilizando los conjuntos de cebador y sonda definidos específicamente. Para determinar esta equivalencia relativa, el AND plasmídico que contiene las secuencias apropiadas de ADNc se diluyen 1/1000, 1/10,000, 1/100,000 y 1/1,000,000 de veces utilizando las

condiciones Q-PCR descritas en lo anterior, se calcula la pendiente de la curva Ct = f(logaritmo de la cantidad de ADN) para cada conjunto de plásmido/cebador/sonda TaqMan. Los conjuntos de plásmidos/cebador/sonda TaqMan proporciona curvas con pendientes cercanas a 3.32, lo cual representa una eficiencia de 100% lo que se considera aceptable. Los conjuntos de plásmidos/cebador/sonda TaqMan utilizados proporcionan todos valores aceptables para CT = f(logaritmo de la cantidad de ADN).

La ausencia de algún nivel significativo de ADN genómico residual en las preparaciones de ADNc se verifica al medir el nivel de la señal de amplificación de PCR cuantitativa para un conjunto genómico específico de cebador/sonda para el gen GOS versus la señal para una sonda de ADNc para el gen GOS.

Resultados:

La figura 9 muestra la inducción de la expresión del gen para CcDH1 en las hojas de árboles pequeños que crecen en invernadero cuando se suspende el suministro de agua (condiciones de sequía). Después de dos semanas, la expresión de DH1 es inducida de manera significativa en una de las plantas sometidas a tensión por agua (planta # 4). Después de tres semanas la expresión de CcDH1 ha sido inducida en la totalidad de las tres plantas tensionadas

por agua (plantas # 4-6). La inducción es muy significativa, alcanzando un RQ superior a 50 para la planta #4. Considerando que el gen control rpl39 se expresa de manera relativamente alta, un RQ de 50 representa un nivel muy alto de inducción de expresión de un gen y sugiere fuertemente que CcDH1 juega un papel importante en la respuesta a la sequía en las hojas de café. La figura muestra la inducción de la expresión del gen para CcDH2 en el mismo conjunto de muestras. La inducción de CcDH2 muestra varias diferencias con la respuesta a CcDH1. La primera diferencia es que CcDH2 es inducido claramente de manera posterior a CcDH1, con la planta aparentemente más estresada (planta #4) mostrando una inducción en la semana 3, una semana después que para CcDH1. A la semana 4, CcDH2 ha sido inducido en la totalidad de las tres plantas sometidas a tensión. La segunda diferencia entre la inducción de CcDH2 versus CcDH1 es que el nivel de inducción de CcDH2 es significativamente menor que el observado para CcDH1. En general, estos resultados indican que CcDH1 y CcDH2 no son inducidas precisamente de la misma manera, aunque las señales probablemente se superpongan, es decir, una o más de las señales necesarias para la inducción de CcDH1 son necesarias para CcDH2, pero CcDH2 puede necesitar una o varias señales adicionales que aparecen únicamente cuando aumenta la tensión por agua.

Sosteniendo el argumento de que CcDH2 es inducido por condiciones que se aproximan a una pérdida de agua extrema, la expresión en las plantas = 5 y # 6 continúa aumentando conforme empeora la tensión por agua con el tiempo, sin agua. De manera alternativa, puede ser que CcDH2 se exprese de manera más significativa que lo indicado en la figura 10, pero esta inducción se localiza en tejidos específicos. Es importante hacer notar que las tres plantas control a las que se les administra agua regularmente no muestran inducción de CcDH1 o CcDH2 (figuras 9, 10; planta #1-3).

Los resultados presentados en lo anterior indican que los promotores asociados con CcDH1 y CcDH2 pueden ser útiles para inducir y activar la expresión del gen en tejidos tensionados osmóticamente. Por ejemplo, estos promotores se pueden utilizar para activar la expresión de genes que son capaces de proporcionar cierta protección para tensión osmótica en el período preciso en el momento en que se produce esta tensión, pero no en la mayor parte de los tejidos bajo condiciones normales. También es evidente de la investigación anterior que, si el objetivo es inducir un gen con baja tensión por agua, debe utilizarse óptimamente el promotor para CcDH1, mientras que la inducción del gen con una tensión mayor de agua, el uso de CcDH2 puede ser más adecuado cuando el objetivo es inducir un gen recombinante únicamente bajo condiciones de

tensión de agua relativamente alta.

Para examinar adicionalmente el efecto de las condiciones de tensión osmótica diferentes examinamos el efecto de la temperatura fría y concentraciones elevadas de NaCl en la expresión de CcDH1 y CcDH2. También se realizaron pruebas para el efecto de una hormona asociada con señalización de tensión osmótica (ácido abscíciso-ABA). Para estos experimentos se utilizaron microcortes de café que crecen en medio sólido *in-vitro* (medio sólido B0.3 en cajas de Petri). Para el experimento con frío y con ABA se generaron microcortes para robusta variedad de ThB4 en placas B0.3 (período de iluminación de 16 horas). Cuando estos microcortes fueron suficientemente grandes, se transfirieron a medio/cajas nuevas y se incubaron 7 días adicionales a 24°C (período de iluminación de 16 horas). En T=0, un conjunto de cajas se transfirieron a 5°C durante 7 días (período de iluminación a 16 horas). Otro conjunto de estos microcortes se colocaron en medio con ABA (medio B0.3 + ABA 100 μM) durante 7 días a 24°C (período de iluminación de 16 horas). Las muestras tomadas en T=0 y T=7 días (5°C y ABA) se congelan a -80°C y después se extrae a ARN para análisis QRT-PCR como se describe en lo anterior. Los resultados de expresión obtenidos para la muestra T=0 del material inicial indican que CcDH1 tiene un RQ = 1.17. Otros experimentos han mostrado que las concentraciones

basales de CcDH1 pueden variar de RQ 0.05 a aproximadamente RQ1 en los microcortes. Dicho espectro relativamente amplio sugiere que el mayor nivel representa la sección de una ligera tensión osmótica en el material inicial para estos experimentos (posiblemente relacionado con que también el material sea fijado a medio sólido o que también las plantas se han sellado lo cual altera la humedad relativa y la edad precisa del material de inicio). No obstante, las muestras que se mantuvieron a 5°C durante 7 días no mostraron inducción de CcDH1 y, de hecho, las concentraciones de CcDH1 en realidad cayeron a un RQ = 0.16, un nivel más cercano al nivel de expresión observado con mayor frecuencia en microcortes no tensionados).

En contraste con los datos presentados en lo anterior para CcDH1, no se detectó expresión de CcDH2 en T=0. No obstante, después de 7 días a 5°C se detectó un RQ = 0.022 para CcDH2. Esta última observación sugiere que CcDH2 es inducido muy ligeramente por condiciones frías. Se hace notar que 5°C es una temperatura muy baja para el café y por lo tanto es posible que si la temperatura de tensión por frío es ligeramente mayor (8-15°C), la inducción de CcDH2 puede ser más significativa. Finalmente, los microcortes tratados con ABA 100 μM durante 7 días mostraron un incremento significativo en la expresión de CcDH1 (RQ, 6.99). Este último resultado demuestra que ABA

está involucrado en la señalización a la inducción de CcDH1. En contraste, ABA no produce ninguna expresión detectable de CcDH2 indicando que esta hormona sola no es suficiente para inducir CcDH2 en algún grado detectable.

Para probar el efecto de la concentración de sal, se agregó NaCl 250 mM en el medio sólido B0.3. En T=0, se coloca el conjunto de microcortes de robusta FRT 35 en medio B0.3 o el medio B0.3 con NaCl 250 mM y se colocó a 24°C (período de iluminación de 16 horas). A T=0 y a los días 4 y 7 se tomaron muestras y se congelaron a -80°C para un análisis posterior con QRT-PCR como se describe en lo anterior. Los resultados obtenidos muestran que en los controles las concentraciones de CcDH1 son muy bajas al inicio y permanecen bajas durante el experimento (control - T=0, RQ = 0.2; T=4 días, RQ=0.07; T=7 días, RQ=0.38). En el tratamiento de prueba, la concentración de CcDH1 se incrementó significativamente los días 4 y 7 del tratamiento (NaCl 250 mM - T=4 días, RQ= 3.31; T=7 días, RQ=4.46). Este experimento muestra que el incremento en la concentración de sal puede inducir la expresión de DH1 a niveles significativos aunque no tan altos como los observados en las hojas de plantas bajo tensión por agua. No se observa inducción significativa para la expresión de CcDH2 en presencia de NaCl 250 mM en cualquiera de las muestras de T=4 o T=7 días.

Referencias:

Agrawal N, Dasaradhi PVN, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK: RNA interference: Biology, mechanism, and applications. Microbiol. Mol Biol. Rev. 67:657-685 (2003).

Allagulova, CR, Gimalov, FR, Shakirova, FM, Vakhitov, VA: The plant dehydrins: Structure and putative functions. Biochemistry-Moscow 68: 945-951 (2003).

Alsheikh, MK, Heyen, BJ, Randall, SK: Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. J. Biol. Chem. 278: 40882-40889 (2003).

Baumlein H, NIVRIDWU: Cis-analysis of a seed protein gene prometer: the conservative RY repeat CA TGCA TG within fue legumin box is essential for tissue-specific expression of a legumin gene. Plant J. 2: 233-239 (1992).

Brummelkamp TR, Bernards R, Agami R: A system for stable expression of short interfering RNAs in mammalian cells. Science 296:550-553 (2002).

Chatthai M, FBYDSIOLOMMS: 2S storage protein gene of Douglas-fir: characterization and activity of promoter in transgenic tobacco seeds. Plant Physiol Biochem 42: 417-423 (2004).

Choi, DW, Close, TJ: A newly identified barley gene, Dhn12 encoding a YSK2 DHN, is located on chromosome

6R and has embryo-specific expression. Theoretical and Applied Genetics 100: 1274-1278 (2000).

Close, T: Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol. Plant 97: 795-803 (1996).

Close, TJ: Dehydrins: a commonality in the response of plants to dehydration and low temperature. Physiol. Plant 100: 291-296 (1997).

Clough, SJ and Bent AF: Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant Journal 16; 735-743 (1998).

Crouzillat, D, Lerceteau, E, Petiard, V, Morera, J, Rodriguez, R, Walker, D, Philips, WRR, Schnell, J, Osei, J, Fritz, P: *Theobroma cacao* L: a genetic linkage map and quantitative trait loci analysis. Theor. Appl. Genet. 93: 205-214 (1996).

Dubouzet JG, SYIYKMDEMSSMSKY-SK: OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J. 33: 751-763 (2003).

Dure, L, Greenway, S, Galau, G: Biochemistry 20: 4162-4178 (1981).

Dure, L: Structural motifs in LEA proteins of higher plants. In: Close, T. J., Bray, E, and A. (eds), Response of Plants to Cellular Dehydration During

Environmental Stress, pp. 91-103. American Society of Plant Physiologists, Rockville, MD. (1993).

Elbashir SM, Harborth J, Weber K, Tuschl T: Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods 26:199-213 (2002).

Godoy, JA, Lunar, R, Torres-Schumann, S, Moreno, J, Rodrigo, M, Pintor-Toro, JA: Expression, tissue distribution and subcellular localization of dehydrin TAS14 in salt-stressed tomate plants. Plant Mol. Biol. 1921-1934 (1994).

Hara, M, Terashima, S, Fukaya, T, Kuboi, T: Enhancement of cold tolerance and inhibition of lipid peroxidation by citrus dehydrin in transgenic tobacco. Planta 217: 290-298 (2003).

Hara, M, Fujinaga, M, Kuboi, T: Radical scavenging activity and oxidative modification of citrus dehydrin. Plant Physiology and Biochemistry 42: 657-662 (2004).

Iida, K, Seki, M, Sakurai, T, Satou, M, Akiyama, K, Toyoda, T, Konagaya, A, Sinozaki, K: Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. Nuc1eic Acids Research 32: 5096-5103 (2004).

Ingram, J, Bartels, D: The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant

Physiology Plant Mol Biol 47: 377-403 (1996).

Iwasaki, T, Yamaguchi-Shinozaki, K, Shinozaki, K: Identification ofa cis-regulatory region of a gene in Arabidopsis thaliana whose introduction by dehydration is mediated by abscisic acid and requires protein synthesis. Molecular and General Genetics 247: 391-398 (1995).

Klahre D, Crete P, Leuenberger AS, Iglesias VA, Meins F: High molecular weight RNAs and small interfering RNAs induce systemic post-transcriptional gene silencing in plants. Proc. Natl. Acad. Sci. USA 99:11981-11986 (2002).

Koag, MC, Fenton, RD, Wilkens, S, Close, TJ: The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. Plant Physiology 131: 309-316 (2003).

Marraccini P., Deshayes A., Pétiard V. and Rogers W. J. 1999. Molecular cloning of the complete 11S seed storage protein gene of Coffea arabica and promoter analysis in fue transgenic tobacco plants. *Plant Physiol. Biochem.* 37:273-282.

Marraccini P, Courjault C, Caillet V, Lausanne F, LePage B, Rogers W, Tessereau S, and Deshayes A. 2003. Rubisco small subunit of *Coffea arabica:* cDNA sequence, gene cloning and promoter analysis in transgenic tobacco plants. *Plant Physiol. Biochem.* 41: 17-25.

Matsuyama, T, Yasumura, N, Funakoshi, M, Yamada,

Y, Hashimoto, T: Maize genes specifically expressed in the outermost cells of the root cap. Plant Cell Physiol. 40: 469-476 (1999).

Mishra RN, Singla-Pareek SL, Nair S, Sopory SK, Reddy MK: Directional genome walking using PCR. Biotechniques. 33:830-834 (2002).

Moore, R, McClelen, C: Ultrastructural aspects of cellular differentiation in fue root cap of Zea mays. Can. J. Bot. 61: 1566-1572 (1983).

Nylander, M, Svensson, J, Palva, ET, Welin, BV: Stress-induced accumulation and tissue-specific localization of dehydrins in Arabidopsis tbaliana. Plant Molecular Biology 45: 263-279 (2001).

Puhakainen, T, Hess, MW, Makela, P, Svensson, J, Heino, P, Palva, ET: Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. Plant Molecular Biology 54: 743-753 (2004).

Rishi AS, Nelson ND, Goyal A: Genome walking of large fragments: an improved method. J. Biotechnol. 111:9-15 (2004).

Roberts, J, DeSimone, N, Lingle, W, Dure, L: Cellular concentrations and unifomity of cell-type accumulation oftwo LEA proteins in cotion embryos. Plant Cell 5: 769-780 (1993).

Rogers, WJ., Bézard, G., Deshayes, A. , Meyer,

I., Pétiard, V., Marraccini, P. (1999). Biochemical and molecular characterisation and expression of the 11S-type storage protein from *Coffea arabica* endosperm. *Plant Physiol. Biochem.* 37(4): 261-272.

Shirsat A, WNCRBD: Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Molecular and General Genetics 215: 326-331 (1989).

Skriver, K, Mundy, J: Gene expression in response to abscisic acid and osmotic stress. Plant Cell 2: 503-512 (1990).

Soulages, JL, Kim, K, Arrese, EL, Walters, C, Cushman, JC: Conformation of a group 2 late embryogenesis abundant protein from soybean. Evidence of poly (L-proline)-type II structure. Plant Physiology 131: 963-975 (2003).

Spanier, AM, Flores, M, Toldra, F, Aristoy, MC, Bett, KL, Bystricky, P, Bland, J: Meat flavor: contribution of proteins and peptides to the flavor of beef. Adv. Exp. Med. Biol. 542: 33-49 (2004).

Tumer, J, Linforth, R, Taylor, A: Real-time monitoring of thermal flavor generation in skim milk powder using atmospheric pressure chemical ionization mass spectrometry. J. Agric. Food Chem 50: 5400-5404 (2002).

Tuschl T, Borkhardt A: Small interfering RNAs: A

revolutionary tool for the analysis of gene function and gene therapy. Mol. Interventions. 2:158-167 (2002).

Wise, M, Tunnacliffe, A: POPP fue question: what do LEA proeins do? Trends P1ant Sci. 9: 13-17 (2004).

Zhu, B, Choi, DW, Fenton, R, Close, TJ: Expression of the barley dehydrin multigene family and the development of freezing tolerance. Molecular and General Genetics 264: 145-153 (2000).

SECUENCIAS de ácidos nucleicos y polipéptidos reclamados

ADNc de CcH1a (véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 1) y proteína codificada (véase SECUENCIA DE IDENTIFICACIÓN NÚMERO: 7)

```
gtgggaagaa gtcctatcgg tctctgatct ttcacctttc gttaatttgt gttcgatatt      60

ctactcccgc tagtagttga aatttggcaa ttaag atg gcg caa tac ggg gct        113
                                        Met Ala Gln Tyr Gly Ala
                                        1               5

gaa tat ggc aac caa aag agc cag tac gat gag tac gga aac cca gtt       161
Glu Tyr Gly Asn Gln Lys Ser Gln Tyr Asp Glu Tyr Gly Asn Pro Val
            10                  15                  20

cgt cag aca gac gaa tat ggt aac cct gcc cgc cat gga ggt acc atg       209
Arg Gln Thr Asp Glu Tyr Gly Asn Pro Ala Arg His Gly Gly Thr Met
        25                  30                  35

ggt gat tat gga acc act ggc act act gga gcc tat ggt ggc aca act       257
Gly Asp Tyr Gly Thr Thr Gly Thr Thr Gly Ala Tyr Gly Gly Thr Thr
    40                  45                  50

gga gca cat ggg act tat gca act gga acc acc ggc act acc ggt acc       305
Gly Ala His Gly Thr Tyr Ala Thr Gly Thr Thr Gly Thr Thr Gly Thr
55                  60                  65                  70

ggt gca tac gcc acc cag cct ggc act gat gtg ggg aag gag cac cat       353
Gly Ala Tyr Ala Thr Gln Pro Gly Thr Asp Val Gly Lys Glu His His
                75                  80                  85
```

```
ggc ctt ggt ggc atg ctt cat cgc tct ggc agc ggt agc tct agc tcg      401
Gly Leu Gly Gly Met Leu His Arg Ser Gly Ser Gly Ser Ser Ser Ser
            90                  95                  100

tcc gag gat gat ggg caa ggc ggg agg agg aag aag ggg atg aag gag      449
Ser Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys Gly Met Lys Glu
        105                 110                 115

aag ata aag gag aaa ctg cct ggc ggt cac aag gag gct caa cct gga      497
Lys Ile Lys Glu Lys Leu Pro Gly Gly His Lys Glu Ala Gln Pro Gly
    120                 125                 130

caa gaa tat tcg agt gct act gca gct cct gga tac ggc ggg gaa gga      545
Gln Glu Tyr Ser Ser Ala Thr Ala Ala Pro Gly Tyr Gly Gly Glu Gly
135                 140                 145                 150

gtg cag cac gag aag aaa gga att atg gat aaa atc aag gag aaa tta      593
Val Gln His Glu Lys Lys Gly Ile Met Asp Lys Ile Lys Glu Lys Leu
                155                 160                 165

cca ggg ggt cac cac aac tga agatctaatt ctaataaata ttggtccgat        644
Pro Gly Gly His His Asn
            170

tatgatattg tgtacccctg ttttcaatct caatctcgtt cgtgtcgcgt ttgtgttttc   704

tgagatttga gtgtgtggac gtcttgagtt tctgtaattg gaataaaaga tgattcgtct   764

tcgtcttcgt ggactctgta gtgtgtttgt ccgtatattc ggcgtcttgt actcgggtca   824

tctggtcatg ta                                                      836
```

ADNc de CcDH1b (véase SECUENCIA DE IDENTIFICACIÓN NÚMERO: 2)
y proteína codificada (véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 8)

```
gtgggaagaa gtcttatcgg tctctgatcc ttcacctttc gttaatctgt gttctatatt    60

ctacttccgc tagtagttga aatttggcaa ttaag atg gcg caa tac ggg gct      113
                                      Met Ala Gln Tyr Gly Ala
                                      1               5

gaa tat ggc aac caa aag agc cag tac gat gag tac gga aac cca gtt      161
Glu Tyr Gly Asn Gln Lys Ser Gln Tyr Asp Glu Tyr Gly Asn Pro Val
            10                  15                  20

cgt cag aca gac gaa tat ggt aac cct gcc cgc cat gga ggt acc atg      209
Arg Gln Thr Asp Glu Tyr Gly Asn Pro Ala Arg His Gly Gly Thr Met
        25                  30                  35

ggt gat tat gga acc act ggc act act gga gcc tat ggt ggc aca act      257
Gly Asp Tyr Gly Thr Thr Gly Thr Thr Gly Ala Tyr Gly Gly Thr Thr
    40                  45                  50

ggg aca gct gga gca cat ggg act tat gca act gga acc acc ggc act      305
Gly Thr Ala Gly Ala His Gly Thr Tyr Ala Thr Gly Thr Thr Gly Thr
55                  60                  65                  70
```

```
acc ggt acc ggt gca tat gcc acc cag cct ggc act gat gtg ggg aag      353
Thr Gly Thr Gly Ala Tyr Ala Thr Gln Pro Gly Thr Asp Val Gly Lys
                75                  80                  85

gag cgc cat ggc ctt ggt ggc atg ctt cat cgc tct ggt agc ggt agc      401
Glu Arg His Gly Leu Gly Gly Met Leu His Arg Ser Gly Ser Gly Ser
            90                  95                  100

tct agc tcg tcc gag gat gat ggg caa ggc ggg agg agg aag aag ggg      449
Ser Ser Ser Ser Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys Gly
        105                 110                 115

atg aag gag aag ata aag gag aaa ctg cct ggc ggt cac aag gag gct      497
Met Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly His Lys Glu Ala
    120                 125                 130

caa cct gga caa gaa tat tcg agt gct act gca gct cct gga tac ggc      545
Gln Pro Gly Gln Glu Tyr Ser Ser Ala Thr Ala Ala Pro Gly Tyr Gly
135                 140                 145                 150

ggg gaa gga gag cag cac gag aag aaa gga att atg gat aaa atc aag      593
Gly Glu Gly Glu Gln His Glu Lys Lys Gly Ile Met Asp Lys Ile Lys
                155                 160                 165

gag aaa tta cca ggg ggt cac cgc aac tga agatctaatt ctaataaata        643
Glu Lys Leu Pro Gly Gly His Arg Asn
            170                 175

ttggatccaa ttatgatatc gtgtacccct gttttcaatc tcaatctcgt tcgtgtcgcg    703

tttgtgtctt ctgagatttg agtgtgtggg cgtcttgagt ttctgtaatc ggaataaaga    763

tgattcgtct tcgtcttcgt cttcgtcttc gtggactctg tagtgtgttt gtccgtatat    823

tcggcgtctt gtactcgggt catctggtca tgtatgtaac atgttatata tcaaatacgt    883

gaagttttgc gttaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa    943

aaaaaaaaaa aaa                                                      956
```

CcDH2a (véase SECUENCIA DE IDENTIFICACIÓN NÚMERO: 3) y proteína codificada (véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 9)

```
ctaaaattcg tcaaccccaa gtctcaggct accttaattt cagtgccctt tttctttatt     60

tttttctaat aacaggagtc ctggaaa atg gct gac ttg cgt gat gaa tat gga     114
                              Met Ala Asp Leu Arg Asp Glu Tyr Gly
                              1               5

aat cct atg cag ttg acc gac cag tat ggc aac ccg gtt cag ctc aag      162
Asn Pro Met Gln Leu Thr Asp Gln Tyr Gly Asn Pro Val Gln Leu Lys
10                  15                  20                  25

gac gag tat ggc aac cca atg cag ctt agc ggt gta gct atc acc gcc      210
Asp Glu Tyr Gly Asn Pro Met Gln Leu Ser Gly Val Ala Ile Thr Ala
                30                  35                  40
```

530

```
ggg acg gct agt gct gtc cat tct act gga acc gga cca act gct gcc      258
Gly Thr Ala Ser Ala Val His Ser Thr Gly Thr Gly Pro Thr Ala Ala
            45                  50                  55

act gga acc cag caa cat cag gag cag ctt cat cgg tct agc agc tca      306
Thr Gly Thr Gln Gln His Gln Glu Gln Leu His Arg Ser Ser Ser Ser
        60                  65                  70

agc tct ggc tcg tcg gag gat gat gga caa gga gga aga aga aag aaa      354
Ser Ser Gly Ser Ser Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys
    75                  80                  85

aaa ggg ttg aaa gaa aag ata aag gag aaa cta acg ggc ggg agg cac      402
Lys Gly Leu Lys Glu Lys Ile Lys Glu Lys Leu Thr Gly Gly Arg His
90                  95                  100                 105

aag gac aga gac gat cag gag cac atc gat gat cag cac gcg cac agc      450
Lys Asp Arg Asp Asp Gln Glu His Ile Asp Asp Gln His Ala His Ser
                110                 115                 120

gcc tct cct cca aca acc acc act ggc agc ggg acg tct act aca gtc      498
Ala Ser Pro Pro Thr Thr Thr Thr Gly Ser Gly Thr Ser Thr Thr Val
            125                 130                 135

ggg ggt cag cag cat gaa aag aag agc atg gtg gag aag att atg gaa      546
Gly Gly Gln Gln His Glu Lys Lys Ser Met Val Glu Lys Ile Met Glu
        140                 145                 150

aag ctc cct ggc cat cac gac acc cgc tag ttacctacca caacatactg      596
Lys Leu Pro Gly His His Asp Thr Arg
    155                 160

tgatcatcgt gtaaaatctc tcctgatgcc taggaaatct agattatgtt aggcattttg   656

tttggtatgt atgtgtgatt aagaccttgt tgtgcgcttg aatcttgaac gtgcatggga   716

tttgcttggt ttgatttgat ttggtgaaat aagttgtact aaaaaaaaaa aaaaaaaa     774
```

ADNc de CcDH2b (véase SECUENCIA DE IDENTIFICACIÓN NÚMERO: 4) y proteína codificada (véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 10)

```
ctaaaattcg tcaaccccaa gtctcaggct accttaattt cagtgccctt tttctttatt    60

tttttctaat aacaggagtc ctggaaa atg gct gac ttg cgt gat gaa tat gga    114
                                Met Ala Asp Leu Arg Asp Glu Tyr Gly
                                1               5

aat cct atg cag ttg acc gac cag tat ggc aac ccg gtt cag ctc aag      162
Asn Pro Met Gln Leu Thr Asp Gln Tyr Gly Asn Pro Val Gln Leu Lys
10                  15                  20                  25

gac gag tat ggc aac cca atg cag ctt agc ggt gta gct atc acc gcc      210
Asp Glu Tyr Gly Asn Pro Met Gln Leu Ser Gly Val Ala Ile Thr Ala
                30                  35                  40

ggg acg gct agt gct gtc cat tct act gga acc gga cca act gct gcc      258
Gly Thr Ala Ser Ala Val His Ser Thr Gly Thr Gly Pro Thr Ala Ala
            45                  50                  55
```

act gga acc cag caa ctt cag gag cag ctt cat cgg tct agc agc tca 306
Thr Gly Thr Gln Gln Leu Gln Glu Gln Leu His Arg Ser Ser Ser Ser
60 65 70

agc tct ggc tcg gtgagatact tgccaagtta caatgtgtgt gtctgtgtgt 358
Ser Ser Gly Ser
75

gtataatgcg ccatcataat tgtttgcttg acagatcctg ttaataatga accgtaattt 418

gacgtaaagt gtacacgttt tgtttttctg ggacttacat aatatcgaat caggctcctg 478

ttgaatttga atgttgttag ctaaaagaaa attttggtgg ctgagttgtt gaatttggtt 538

tatgg tcg gag gat gat gga caa gga gga aga aga aag aaa aaa ggg ttg 588
Ser Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys Lys Gly Leu
80 85 90

aaa gaa aag ata aag gag aaa cta acg ggc ggg agg cac aag gac aga 636
Lys Glu Lys Ile Lys Glu Lys Leu Thr Gly Gly Arg His Lys Asp Arg
95 100 105

gac gat cag gag cac atc gat gat cag cac gcg cac agc gcc tct cct 684
Asp Asp Gln Glu His Ile Asp Asp Gln His Ala His Ser Ala Ser Pro
110 115 120

cca aca acc acc act ggc agc ggg acg tct act aca gtc ggg ggt cag 732
Pro Thr Thr Thr Thr Gly Ser Gly Thr Ser Thr Thr Val Gly Gly Gln
125 130 135 140

cag cat gaa aag aag agc atg gtg gag aag att atg gaa aag ctc cct 780
Gln His Glu Lys Lys Ser Met Val Glu Lys Ile Met Glu Lys Leu Pro
145 150 155

ggc cat cac gac acc cgc tagttaccta ccacaacata ctgtgatcat 828
Gly His His Asp Thr Arg
160

cgtgtaaaat ctctcctgat gcctaggaaa tctagattat gttaggcatt ttgtttggta 888

tgtatgtgtg attaagacct tgttgtgcgc ttgaatcttg aacgtgcatg ggatttgctt 948

ggtttgaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1008

aaaaaaaaaa aa 1020

ADNc de CcDH3 (véase SECUENCIA DE IDENTIFICACIÓN NÚMERO: 5) y proteína codificada (véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 11)

atttttgct gtttcctgtt acattctgct ttacgtacat ccatcagcaa a atg gcc 57
Met Ala
1

gag tat gat cag agt aac atc aag gtt gag gag gga tca gct gtc gag 105
Glu Tyr Asp Gln Ser Asn Ile Lys Val Glu Glu Gly Ser Ala Val Glu
5 10 15

```
gcc acg gat cgc gga ctc ttc aac ttg ggc aag aaa gag gaa gtg aag      153
Ala Thr Asp Arg Gly Leu Phe Asn Leu Gly Lys Lys Glu Glu Val Lys
    20                  25                  30

aag tgt gat caa ggc cag gcc atc tct gcg gag ttt gat gag aaa gtg      201
Lys Cys Asp Gln Gly Gln Ala Ile Ser Ala Glu Phe Asp Glu Lys Val
35                  40                  45                  50

cgt gtt tct gaa cca gac aag gag gag gga aag aag cat ggt ggt ctt      249
Arg Val Ser Glu Pro Asp Lys Glu Glu Gly Lys Lys His Gly Gly Leu
                55                  60                  65

ctc gag aag ctc cac cga tct ggt agc agc tcc agc agc tca agt gag      297
Leu Glu Lys Leu His Arg Ser Gly Ser Ser Ser Ser Ser Ser Ser Glu
            70                  75                  80

gaa gaa gta gaa gag ggt ggt gag aag aag aag aaa aag aag gaa aag      345
Glu Glu Val Glu Glu Gly Gly Glu Lys Lys Lys Lys Lys Lys Glu Lys
        85                  90                  95

aag ggt ttg aag gac aag atc aag gag aag ata tcg ggt gat aag aag      393
Lys Gly Leu Lys Asp Lys Ile Lys Glu Lys Ile Ser Gly Asp Lys Lys
    100                 105                 110

gac gaa gaa aag gtt gaa aaa tgt gag gaa gac acg tct atc cca gtt      441
Asp Glu Glu Lys Val Glu Lys Cys Glu Glu Asp Thr Ser Ile Pro Val
115                 120                 125                 130

gag aaa tat gcc gaa ccg gcc cat gca gat gct gct cat gaa cca gag      489
Glu Lys Tyr Ala Glu Pro Ala His Ala Asp Ala Ala His Glu Pro Glu
                135                 140                 145

gag aaa aag ggc ttc tta gat aag atc aag gag aaa cta cca ggt ggt      537
Glu Lys Lys Gly Phe Leu Asp Lys Ile Lys Glu Lys Leu Pro Gly Gly
            150                 155                 160

ggt cag aag aag act gag gaa gtc gca gca gca gca ccg cct cct cct      585
Gly Gln Lys Lys Thr Glu Glu Val Ala Ala Ala Ala Pro Pro Pro Pro
        165                 170                 175

ccg gca gag tgc acc gcc act gaa ggt gag gcc aag gat aag aag gga      633
Pro Ala Glu Cys Thr Ala Thr Glu Gly Glu Ala Lys Asp Lys Lys Gly
    180                 185                 190

ttc ttg gac aag atc aag gag aag ctc cct ggc tac cat ccc aag act      681
Phe Leu Asp Lys Ile Lys Glu Lys Leu Pro Gly Tyr His Pro Lys Thr
195                 200                 205                 210

gaa gaa gag aag gaa aag gag aag gaa aaa gaa aag gag gct gga tgc      729
Glu Glu Glu Lys Glu Lys Glu Lys Glu Lys Glu Lys Glu Ala Gly Cys
                215                 220                 225

cat taa taaaagagca aagcaaatta atagcagctt tccagtgtgt cataattttg      785
His

catttggatt aatacatttt ggagtggcaa tgatcttttt attttaaa               833
```

ADNc de CcLEA1 (véase SECUENCIA DE IDENTIFICACIÓN NÚMERO: 6) y proteína codificada (véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 12)

```
aagcagtggt aacaacgcag agtacgcggg acgaacatca tcggtaccag tttcctattc      60

atacatcatc ttactagcac tg atg caa aaa atg act cct ctg aga tgc atc      112
                         Met Gln Lys Met Thr Pro Leu Arg Cys Ile
                         1               5                  10

aat ttc att ttt ctg gcc ttt tgg gtt cct gct gtc ctc gcg gtg atg      160
Asn Phe Ile Phe Leu Ala Phe Trp Val Pro Ala Val Leu Ala Val Met
                    15                  20                  25

gcc gaa aaa ccc cta gtt cct aca tac cta atc ccc aaa ccc cct cca      208
Ala Glu Lys Pro Leu Val Pro Thr Tyr Leu Ile Pro Lys Pro Pro Pro
            30                  35                  40

ccg cca tcg cca gtg aaa cca tca gta ccc gtg ata cct gtc aaa ccc      256
Pro Pro Ser Pro Val Lys Pro Ser Val Pro Val Ile Pro Val Lys Pro
        45                  50                  55

cga atc gtg aga tgc cgc tcc aca ttg ttt cct ctc tgc ttc aat atc      304
Arg Ile Val Arg Cys Arg Ser Thr Leu Phe Pro Leu Cys Phe Asn Ile
    60                  65                  70

ccc ttc gtt tgc ccc tta gac tgt ctt acc aac tgt tta gtg gac tgt      352
Pro Phe Val Cys Pro Leu Asp Cys Leu Thr Asn Cys Leu Val Asp Cys
75                  80                  85                  90

gtc acc tgc aag gct tac tgc agt tgc aac ttt ccc ggc gct gtt tgt      400
Val Thr Cys Lys Ala Tyr Cys Ser Cys Asn Phe Pro Gly Ala Val Cys
                95                  100                 105

cag gat cca cga ttc gtt ggg ggc gat ggc aac aca ttt tac ttc cat      448
Gln Asp Pro Arg Phe Val Gly Gly Asp Gly Asn Thr Phe Tyr Phe His
            110                 115                 120

ggc cgc aag gat cag gac ttc tgc ctg gtt tcg gat acc aat ctt cat      496
Gly Arg Lys Asp Gln Asp Phe Cys Leu Val Ser Asp Thr Asn Leu His
        125                 130                 135

gta aat ggt cat ttc att ggc aaa aga aaa cct aat ttg cgc aga gac      544
Val Asn Gly His Phe Ile Gly Lys Arg Lys Pro Asn Leu Arg Arg Asp
    140                 145                 150

ttc act tgg gtg cag gcc att gga ata atg ttc gac gac cac aga atc      592
Phe Thr Trp Val Gln Ala Ile Gly Ile Met Phe Asp Asp His Arg Ile
155                 160                 165                 170

ctc gtg gcc gca aaa agg act tca acg tgg gac gac aat gtg gat cga      640
Leu Val Ala Ala Lys Arg Thr Ser Thr Trp Asp Asp Asn Val Asp Arg
                175                 180                 185

ctc gct ata tcc att gat gga aat ccg att tcc ctc ccc act gaa gaa      688
Leu Ala Ile Ser Ile Asp Gly Asn Pro Ile Ser Leu Pro Thr Glu Glu
            190                 195                 200

gga tcc aaa tgg caa ctt ccg gcc ccg tcc aat gtc agt atc atg aga      736
Gly Ser Lys Trp Gln Leu Pro Ala Pro Ser Asn Val Ser Ile Met Arg
        205                 210                 215
```

```
aca agc aac aat aac gga ctt gtg gtt gaa gcc gtg aac aat ttc agg      784
Thr Ser Asn Asn Asn Gly Leu Val Val Glu Ala Val Asn Asn Phe Arg
    220                 225                 230

atc acc gcc aat gtg gtt cca ata aca gct caa gaa tca aaa gtt cat      832
Ile Thr Ala Asn Val Val Pro Ile Thr Ala Gln Glu Ser Lys Val His
235                 240                 245                 250

ggt tat gac att act gat gag gat tgc ttt acc cat ttg gag ctt ggg      880
Gly Tyr Asp Ile Thr Asp Glu Asp Cys Phe Thr His Leu Glu Leu Gly
                255                 260                 265

ttc aaa ttc ttc aac atc acc gat tca act gat gga gtt ttg gga caa      928
Phe Lys Phe Phe Asn Ile Thr Asp Ser Thr Asp Gly Val Leu Gly Gln
            270                 275                 280

acc tat agg agc gat tac gtg aac aaa atg aag gtg aat gcg gta atg      976
Thr Tyr Arg Ser Asp Tyr Val Asn Lys Met Lys Val Asn Ala Val Met
        285                 290                 295

cca gtc atg ggc ggt gac cgt aag tac ttg act tcg gga ctt ttt agt     1024
Pro Val Met Gly Gly Asp Arg Lys Tyr Leu Thr Ser Gly Leu Phe Ser
    300                 305                 310

gcc gat tgt gct gtt tct cgc ttt ggt ggg aag gtt ctt gag aaa gcc     1072
Ala Asp Cys Ala Val Ser Arg Phe Gly Gly Lys Val Leu Glu Lys Ala
315                 320                 325                 330

aat tct gct tct cct gtg cat gag tat cca gcc ttg aac tgc aag agt     1120
Asn Ser Ala Ser Pro Val His Glu Tyr Pro Ala Leu Asn Cys Lys Ser
                335                 340                 345

ggg atg gaa ggg aat ggc ttg gtt tgc aaa aaa taa ttaagttgct          1166
Gly Met Glu Gly Asn Gly Leu Val Cys Lys Lys
            350                 355

acagagcatg ttgtatgctg aatgatgagc tataaataat tgagtttcag aaaagtctta   1226

ttagaaatga agtgatcata gctttacatg caaaaaaaaa aaaaaaaaaa aaaaaaaaaa   1286

a                                                                   1287
```

Secuencia completa de CcDH2a, que incluye promotor (bases 1-1324; véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 13), secuencia codificante (véase la SECUENCIA DE IDENTIFICACIÓN NÚMERO: 3) y proteína codificada (véase SECUENCIA DE IDENTIFICACIÓN NÚMERO: 9) Intrón I (1642)..(1866)

```
atagtgacct taatagcgat cttgttgctt ttgatcgtca gaaaagtagt ggacatgacg     60

gaagaggtcc taagatgagt tccagttcca gcatgaaggg ctctttggcg aagcctttct    120

tgaggcgtca cttttctttt ggatctaaag gcagtagatc aatgtcagag aatcattctt    180

cctggaagag gggattcttc tgggcaaaat cgagaaagga ttaagttctg tctagagtta    240

caaaggtgag caacagtcac ggttttttat tagggaatgg aaggattgga tccctttcca    300

cgtagtgaac aacatatatt ttgcatggtt ggtcttagta cctataacac gaaaatgttc    360

ttcatccgtt ctattaatca ttaggcttta gtcatttaaa ttttttacat cccgcatttc    420
```

tcctcttgat tcttgttgat ttctgcagat tccacagttg ttcttcagat gggctacgaa 480

atgcatgcag ggagcaggca atcagccata aattcaaccc tgtcaaggaa gctggcattg 540

tctcgtgcaa atgtaggtta gcttttgaag atacactgca aagggaagac catacagatg 600

gggaaatgaa ttcattataa tataggaaaa aggaaagatg ataggggtca gggcgtccgt 660

gcatcatgaa actagttctc tttcattttg tacgatggct gtttactgtt taatttcatg 720

aaattagttt ggatatatgc gtagcgtttt accatcgcat ttctaaatcg atattctatg 780

ggccgaatta cgcgttggag acatcattgg gttgctcctc tcaatcccat ctctatctat 840

tgacggatcc ggatcatgat gttgaacctt tcaacttttg acttagatgg gatttgtgtt 900

cgcgtgttgt taacttgtta ctgaccgact cagaagacag cggattctga cttcaccacg 960

tgtctcttta gtgaaaattt aaaaggcatt tttcttctgt tcatagttta aaatgtaatg 1020

tgattattaa aagatcgttt ggtattattt caaggatgga tggattggat ggaagggata 1080

tctgatatat atcataccct tccaaaattc aggaccatga cgtatttaat atcccccagc 1140

ggaagacacg tgccttgatg tcttataggt ggcaatacac ttcagcttcc tctgctaata 1200

cgtgtgagga tcttcggtac catgcagaaa agaccgcggt gctccttcca ccgtcctcat 1260

ccctctcttg gctttttaa gtctcctgcg atatccaaaa tccaaacaaa gccgttatcg 1320

cagctaaaat tcgtcaaccc caagtctcag gctaccttaa tttcagtgcc ctttttcttt 1380

attttttttct aataacagga gtcctggaaa atg gct gac ttg cgt gat gaa tat 1434
Ala Asp Leu Arg Asp Glu Tyr
1 5

gga aat cct atg cag ttg acc gac cag tat ggc aac ccg gtt cag ctc 1482
Gly Asn Pro Met Gln Leu Thr Asp Gln Tyr Gly Asn Pro Val Gln Leu
10 15 20

aag gac gag tat ggc aac cca atg cag ctt agc ggt gta gct atc acc 1530
Lys Asp Glu Tyr Gly Asn Pro Met Gln Leu Ser Gly Val Ala Ile Thr
25 30 35

gcc ggg acg gct agt act gtc cat tct act gga acc gga cca act gct 1578
Ala Gly Thr Ala Ser Thr Val His Ser Thr Gly Thr Gly Pro Thr Ala
40 45 50 55

gcc act gga acc cag caa cat cag gag cag ctt cat cgg tct agc agc 1626
Ala Thr Gly Thr Gln Gln His Gln Glu Gln Leu His Arg Ser Ser Ser
60 65 70

tca agc tct ggc tcg gtgagatact tgccaagtta caatgtgtgt gtctgtgtgt 1681
Ser Ser Ser Gly Ser
75

gtataatgcg ccatcataat tgtttgcttg acagatcctg ttaataatga accgtaattt 1741

gacgtaaagt gtacacgttt tgtttttctg ggactaacat aatatcgaat caggctcctg 1801

```
ttgaatttga atgttgttag ctaaaagaaa attttggtgg ctgagttgtt gaatttggtt   1861

tatag acg gag gat gat gga caa gga gga aga aga aag aaa aaa ggg ttg   1911
      Thr Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys Lys Gly Leu
                  80                  85                  90

aaa gaa aag ata aag gag aaa cta acg ggc ggt agg cac aag gac aga     1959
Lys Glu Lys Ile Lys Glu Lys Leu Thr Gly Gly Arg His Lys Asp Arg
            95                  100                 105

gac gat cag gag cac atc gat gat cag cac gcg cac agc gcc tct cct     2007
Asp Asp Gln Glu His Ile Asp Asp Gln His Ala His Ser Ala Ser Pro
        110                 115                 120

cca aca acc acc act ggc agc ggg acg tct act aca gtc ggg ggt cag     2055
Pro Thr Thr Thr Thr Gly Ser Gly Thr Ser Thr Thr Val Gly Gly Gln
    125                 130                 135

cag cat gaa aag aag agc atg gtg gag aag att atg gaa aag ctc cct     2103
Gln His Glu Lys Lys Ser Met Val Glu Lys Ile Met Glu Lys Leu Pro
140                 145                 150                 155

ggc cat cac gac acc cgc tagttaccta ccacaacata ctgtgatcat            2151
Gly His His Asp Thr Arg
                160

cgtgtaaaat ctctcctgat gcctaggaaa tctagattat gttaggcatt ttgtttggta   2211

tgtatgtgtg attaagacct tgttgtgcgc ttgaatcttg aacgtgcatg ggatttgctt   2271

ggtttgattt gatttgg                                                  2288
```

La presente invención no se limita a las modalidades descritas y ejemplificadas en lo anterior, sino que es capaz de variación y modificación dentro del alcance de las reivindicaciones anexas.

REIVINDICACIONES

1. Una molécula de ácido nucleico aislada de café (género *Coffea*) que tiene una secuencia codificante que codifica para una deshidrina o una proteína abundante embriogénica tardía (LEA).

2. La molécula de ácido nucleico como se describe en la reivindicación 1, en donde la secuencia codificante codifica para una deshidrina que tiene un peso molecular de entre aproximadamente 17 kDa y aproximadamente 26 kDa.

3. La molécula de ácido nucleico como se describe en la reivindicación 2, en donde la deshidrina o proteína LEA tiene una secuencia de aminoácidos que es 46% o más idéntica a cualquiera de las SECUENCIAS DE IDENTIFICACION NUMEROS: 7-12.

4. La molécula de ácido nucleico como se describe en la reivindicación 3, en donde la deshidrina o proteína LEA tiene una secuencia de aminoácidos de cualquiera de las SECUENCIAS DE IDENTIFICACION NUMEROS: 7-12.

5. La molécula de ácido nucleico como se describe en la reivindicación 1, en donde la secuencia codificante es 50% o más idéntica a cualquiera de las secuencias codificantes que se establecen en las SECUENCIAS

DE IDENTIFICACION NUMEROS: 1-6.

6. La molécula de ácido nucleico como se describe en la reivindicación 5, en donde la secuencia codificante comprende cualquiera de las SECUENCIAS DE IDENTIFICACION NUMEROS: 1-6.

7. La molécula de ácido nucleico como se describe en la reivindicación 1, la cual es un gen que tiene un marco de lectura abierto que comprende la secuencia codificante.

8. Una molécula de ARNm producida por transcripción del gen como se describe en la reivindicación 7.

9. Una molécula de ADNc producida por transcripción inversa de la molécula de ARNm de la reivindicación 8.

10. Un oligonucleótido entre 8 y 100 bases de longitud el cual es complementario a un segmento de la molécula de ácido nucleico de la reivindicación 1.

11. Un vector que comprende la molécula de ácido nucleico de la reivindicación 1.

12. El vector como se describe en la reivindicación 11, el cual es un vector de expresión que se selecciona del grupo de vectores que consisten de los vectores plasmídicos, fagémidos, cósmidos, baculovirus, bácmidos, bacterianos, de levadura y virales.

13. El vector como se describe en la reivindicación 11, en donde la secuencia codificante de la molécula de ácido nucleico está unida operablemente a un promotor constitutivo.

14. El vector como se describe en la reivindicación 11, en donde la secuencia codificante de la molécula de ácido nucleico está unida operablemente a un promotor inducible.

15. El vector como se describe en la reivindicación 11, en donde la secuencia codificante de la molécula de ácido nucleico está unida operablemente a un promotor específico de tejido.

16. El vector como se describe en la reivindicación 15, en donde el promotor específico de tejido es un promotor específico de semilla.

17. El vector como se describe en la reivindicación 16, en donde el promotor específico de semilla es promotor específico de semilla de café.

18. El vector como se describe en la reivindicación 17, en donde el promotor específico de semilla de café es un promotor del gen para deshidrina.

19. El vector como se describe en la reivindicación 18, en donde el promotor de gen para deshidrina comprende la SECUENCIA DE IDENTIFICACION NUMERO: 13.

20. Una célula hospedera transformada con el vector como se describe en la reivindicación 11.

21. La célula hospedera como se describe en la reivindicación 20, que se selecciona del grupo que consiste de células vegetales, células bacterianas, células micóticas, células de insecto y células de mamífero.

22. La célula hospedera como se describe en la reivindicación 21, la cual es una célula vegetal que se selecciona del grupo de plantas que consisten de café, tabaco, *Arabidopsis*, maíz, trigo, arroz, soya, cebada, centeno, avena, sorgo, alfalfa, trébol, nabina, azafrán, girasol, cacahuate, cacao, tomate tomatillo, papa, pimienta, berenjena, remolacha, zanahora, pepino, lechuga, chícharo, aster, begonia, crisantemom, delfinium, zinia y trufgrasses.

23. Una planta transgénica fértil producida por degeneración de la célula hospedera de la reivindicación 22.

24. La planta transgénica fértil como se describe en la reivindicación 23, la cual es del género *Coffea*.

25. Un método para modular el sabor o aroma de granos de café que comprende modular la producción de una o más deshidrinas o proteínas LEA dentro de las semillas de café.

26. El método como se describe en la reivindicación 25, que comprende incrementar la producción de una o más deshidrinas o proteínas LEA.

27. El método como se describe en la reivindicación 26, que comprende incrementar la expresión de uno o más genes endógenos que codifican para deshidrina o proteína LEA dentro de las semillas de café.

28. El método como se describe en la reivindicación 27, que comprende introducir un transgén que codifica para deshidrina o proteína LEA dentro de la planta.

29. El método como se describe en la reivindicación 25, que comprende disminuir la producción de una o más deshidrinas o proteínas LEA.

30. El método como se describe en la reivindicación 29, que comprende introducir una molécula de ácido nucleico en el café que inhibe la expresión de uno o más genes que codifican para deshidrina o proteína LEA.

31. Un método para incrementar la resistencia a la tensión osmótica en una planta que comprende incrementar la producción de una o más deshidrinas o proteínas LEA dentro de la planta.

32. El método como se describe en la reivindicación 31, que comprende introducir un transgén que codifica para deshidrina o proteína LEA dentro de la

planta.

33. Un promotor aislado del gen de planta de café que codifica para una deshidrina.

34. El promotor como se describe en la reivindicación 33, en donde el gen que codifica para una deshidrina tiene un peso molecular de entre aproximadamente 17 kDa y aproximadamente 26 kDa.

35. El promotor como se describe en la reivindicación 34, en donde el gen que codifica para una deshidrina tiene una secuencia de aminoácidos que es 50% o más idéntica a cualquiera de las SECUENCIAS DE IDENTIFICACION NUMEROS: 7-11.

36. El promotor como se describe en la reivindicación 35, en donde el gen que codifica para una deshidrina tiene una secuencia de aminoácidos en cualquiera de las SECUENCIAS DE IDENTIFICACION NUMEROS: 7-11.

37. El promotor como se describe en la reivindicación 33, en donde el gen comprende un marco de lectura abierto que es 50% o más idéntico a cualquiera de las secuencias establecidas en las SECUENCIAS DE IDENTIFICACION NUMEROS: 1-5.

38. El promotor como se describe en la reivindicación 37, en donde el gen comprende un marco de lectura abierto que tiene cualquiera de las SECUENCIAS DE IDENTIFICACION NUMEROS: 1-5.

39. El promotor como se describe en la reivindicación 33, que comprende una o más secuencias reguladoras que se seleccionan del grupo que consiste de una caja TATA, un elemento sensible a ácido abscísico, una secuencia repetida RY (CATGCAT(T/a)(A/g) de una caja de leguminina para regular la expresión de proteína de tipo de leguminina, por lo menos un motivo de un elemento sensible a la deshidratación/secuencia de acción cis repetida C (G/ACCGAC y por lo menos un motivo de caja E (CANNTG).

40. El promotor como se describe en la reivindicación 39, que comprende la SECUENCIA DE IDENTIFICACION NUMERO: 13.

41. Un gen quimérico que comprende el promotor como se describe en la reivindicación 33, unido operablemente a una o más secuencias codificantes.

42. Un vector para transformar una célula, que comprende el gen quimérico de la reivindicación 41.

43. Una célula transformada con el vector de la reivindicación 42.

44. La célula transformada de la reivindicación 43, la cual es una célula vegetal.

45. La célula vegetal transformada como se describe en la reivindicación 44, la cual es una célula del género *Coffea*.

46. Una planta transgénica fértil producida al regenerar la célula vegetal transformada de la reivindicación 44.

47. La planta transgénica fértil como se describe en la reivindicación 46, la cual es del género *Coffea*.

NESTEC Dehydrin.ST25.txt
LISTADO DE SECUENCIAS

<110> Nestec S.A.
Cornell Research Foundation
BEN AMOR, Mohamed
MCCARTHY, James Gérard
PETIARD, Vincent
TANKSLEY, Steven D.
LIN, Chenwei

<120> Genes para deshidrina y promotores del café

<130> NO 7972/WO

<150> U.S. 60/696890
<151> 2005-07-06

<160> 46

<170> PatentIn version 3.3

<210> 1
<211> 880
<212> ADN
<213> Coffea Canephora

<400> 1
gtgggaagaa gtcctatcgg tctctgatct ttcacctttc gttaatttgt gttcgatatt 60

ctactcccgc tagtagttga aatttggcaa ttaagatggc gcaatacggg gctgaatatg 120

gcaaccaaaa gagccagtac gatgagtacg gaaacccagt tcgtcagaca gacgaatatg 180

gtaaccctgc ccgccatgga ggtaccatgg gtgattatgg aaccactggc actactggag 240

cctatggtgg cacaactgga gcacatggga cttatgcaac tggaaccacc ggcactaccg 300

gtaccggtgc atacgccacc cagcctggca ctgatgtggg gaaggagcac catggccttg 360

gtggcatgct tcatcgctct ggcagcggta gctctagctc gtccgaggat gatgggcaag 420

gcgggaggag gaagaagggg atgaaggaga agataaagga gaaactgcct ggcggtcaca 480

aggaggctca acctggacaa gaatattcga gtgctactgc agctcctgga tacggcgggg 540

aaggagtgca gcacgagaag aaaggaatta tggataaaat caaggagaaa ttaccagggg 600

gtcaccacaa ctgaagatct aattctaata aatattggtc cgattatgat attgtgtacc 660

cctgttttca atctcaatct cgttcgtgtc gcgtttgtgt tttctgagat ttgagtgtgt 720

ggacgtcttg agtttctgta attggaataa aagatgattc gtcttcgtct tcgtggactc 780

tgtagtgtgt ttgtccgtat attcggcgtc ttgtactcgg gtcatctggt catgtatgta 840

acatgttata tatcaaatac gtgaagtttt gcgttaaaaa 880

<210> 2
<211> 901
<212> ADN
<213> Coffea canephora

<400> 2
gtgggaagaa gtcttatcgg tctctgatcc ttcacctttc gttaatctgt gttctatatt 60

ctacttccgc tagtagttga aatttggcaa ttaagatggc gcaatacggg gctgaatatg 120

546

gcaaccaaaa gagccagtac gatgagtacg gaaacccagt tcgtcagaca gacgaatatg 180

gtaaccctgc ccgccatgga ggtaccatgg gtgattatgg aaccactggc actactggag 240

cctatggtgg cacaactggg acagctggag cacatgggac ttatgcaact ggaaccaccg 300

gcactaccgg taccggtgca tatgccaccc agcctggcac tgatgtgggg aaggagcgcc 360

atggccttgg tggcatgctt catcgctctg gtagcggtag ctctagctcg tccgaggatg 420

atgggcaagg cgggaggagg aagaagggga tgaaggagaa gataaaggag aaactgcctg 480

gcggtcacaa ggaggctcaa cctggacaag aatattcgag tgctactgca gctcctggat 540

acggcgggga aggagagcag cacgagaaga aaggaattat ggataaaatc aaggagaaat 600

taccaggggg tcaccgcaac tgaagatcta attctaataa atattggatc caattatgat 660

atcgtgtacc cctgttttca atctcaatct cgttcgtgtc gcgtttgtgt cttctgagat 720

ttgagtgtgt gggcgtcttg agtttctgta atcggaataa agatgattcg tcttcgtctt 780

cgtcttcgtc ttcgtggact ctgtagtgtg tttgtccgta tattcggcgt cttgtactcg 840

ggtcatctgg tcatgtatgt aacatgttat atatcaaata cgtgaagttt tgcgttaaaa 900

a 901

<210> 3
<211> 761
<212> ADN
<213> Coffea canephora

<400> 3
ctaaaattcg tcaaccccaa gtctcaggct accttaattt cagtgccctt tttctttatt 60

tttttctaat aacaggagtc ctggaaaatg gctgacttgc gtgatgaata tggaaatcct 120

atgcagttga ccgaccagta tggcaacccg gttcagctca aggacgagta tggcaaccca 180

atgcagctta gcggtgtagc tatcaccgcc gggacggcta gtgctgtcca ttctactgga 240

accggaccaa ctgctgccac tggaacccag caacatcagg agcagcttca tcggtctagc 300

agctcaagct ctggctcgtc ggaggatgat ggacaaggag gaagaagaaa gaaaaaaggg 360

ttgaaagaaa agataaagga gaaactaacg ggcgggaggc acaaggacag agacgatcag 420

gagcacatcg atgatcagca cgcgcacagc gcctctcctc caacaaccac cactggcagc 480

gggacgtcta ctacagtcgg gggtcagcag catgaaaaga agagcatggt ggagaagatt 540

atggaaaagc tccctggcca tcacgacacc cgctagttac ctaccacaac atactgtgat 600

catcgtgtaa aatctctcct gatgcctagg aaatctagat tatgttaggc attttgtttg 660

gtatgtatgt gtgattaaga ccttgttgtg cgcttgaatc ttgaacgtgc atgggatttg 720

cttggtttga tttgatttgg tgaaataagt tgtactaaaa a 761

<210> 4
<211> 959
<212> ADN
<213> Coffea canephora

<400> 4

547

ctaaaattcg tcaaccccaa gtctcaggct accttaattt cagtgccctt tttctttatt 60

tttttctaat aacaggagtc ctggaaaatg gctgacttgc gtgatgaata tggaaatcct 120

atgcagttga ccgaccagta tggcaacccg gttcagctca aggacgagta tggcaaccca 180

atgcagctta gcggtgtagc tatcaccgcc gggacggcta gtgctgtcca ttctactgga 240

accggaccaa ctgctgccac tggaacccag caacttcagg agcagcttca tcggtctagc 300

agctcaagct ctggctcggt gagatacttg ccaagttaca atgtgtgtgt ctgtgtgtgt 360

ataatgcgcc atcataattg tttgcttgac agatcctgtt aataatgaac cgtaatttga 420

cgtaaagtgt acacgttttg tttttctggg acttacataa tatcgaatca ggctcctgtt 480

gaatttgaat gttgttagct aaaagaaaat tttggtggct gagttgttga atttggttta 540

tggtcggagg atgatggaca aggaggaaga agaaagaaaa aagggttgaa agaaaagata 600

aaggagaaac taacgggcgg gaggcacaag gacagagacg atcaggagca catcgatgat 660

cagcacgcgc acagcgcctc tcctccaaca accaccactg gcagcgggac gtctactaca 720

gtcgggggtc agcagcatga aaagaagagc atggtggaga agattatgga aaagctccct 780

ggccatcacg acacccgcta gttacctacc acaacatact gtgatcatcg tgtaaaatct 840

ctcctgatgc ctaggaaatc tagattatgt taggcatttt gtttggtatg tatgtgtgat 900

taagaccttg ttgtgcgctt gaatcttgaa cgtgcatggg atttgcttgg tttgaaaaa 959

<210> 5
<211> 835
<212> ADN
<213> coffea canephora

<400> 5
atttttgct gtttcctgtt acattctgct ttacgtacat ccatcagcaa aatggccgag 60

tatgatcaga gtaacatcaa ggttgaggag ggatcagctg tcgaggccac ggatcgcgga 120

ctcttcaact tgggcaagaa agaggaagtg aagaagtgtg atcaaggcca ggccatctct 180

gcggagtttg atgagaaagt gcgtgtttct gaaccagaca aggaggaggg aaagaagcat 240

ggtggtcttc tcgagaagct ccaccgatct ggtagcagct ccagcagctc aagtgaggaa 300

gaagtagaag agggtggtga gaagaagaag aaaaagaagg aaaagaaggg tttgaaggac 360

aagatcaagg agaagatatc gggtgataag aaggacgaag aaaaggttga aaaatgtgag 420

gaagacacgt ctatcccagt tgagaaatat gccgaaccgg cccatgcaga tgctgctcat 480

gaaccagagg agaaaaaggg cttcttagat aagatcaagg agaaactacc aggtggtggt 540

cagaagaaga ctgaggaagt cgcagcagca gcaccgcctc ctcctccggc agagtgcacc 600

gccactgaag gtgaggccaa ggataagaag ggattcttgg acaagatcaa ggagaagctc 660

cctggctacc atcccaagac tgaagaagag aaggaaaagg agaaggaaaa agaaaaggag 720

gctggatgcc attaataaaa gagcaaagca aattaatagc agctttccag tgtgtcataa 780

ttttgcattt ggattaatac attttggagt ggcaatgatc tttttatttt aaaaa 835

548

NESTEC Dehydrin.ST25.txt

<210> 6
<211> 1232
<212> ADN
<213> coffea canephora

<400> 6
acgaacatca tcggtaccag tttcctattc atacatcatc ttactagcac tgatgcaaaa 60
aatgactcct ctgagatgca tcaatttcat ttttctggcc ttttgggttc ctgctgtcct 120
cgcggtgatg gccgaaaaac ccctagttcc tacataccta atccccaaac cccctccacc 180
gccatcgcca gtgaaaccat cagtacccgt gatacctgtc aaaccccgaa tcgtgagatg 240
ccgctccaca ttgtttcctc tctgcttcaa tatccccttc gtttgccccc tagactgtct 300
taccaactgt ttagtggact gtgtcacctg caaggcttac tgcagttgca actttcccgg 360
cgctgtttgt caggatccac gattcgttgg gggcgatggc aacacatttt acttccatgg 420
ccgcaaggat caggacttct gcctggtttc ggataccaat cttcatgtaa atggtcattt 480
cattggcaaa agaaaaccta atttgcgcag agacttcact tgggtgcagg ccattggaat 540
aatgttcgac gaccacagaa tcctcgtggc cgcaaaaagg acttcaacgt gggacgacaa 600
tgtggatcga ctcgctatat ccattgatgg aaatccgatt tccctcccca ctgaagaagg 660
atccaaatgg caacttccgg ccccgtccaa tgtcagtatc atgagaacaa gcaacaataa 720
cggacttgtg gttgaagccg tgaacaattt caggatcacc gccaatgtgg ttccaataac 780
agctcaagaa tcaaaagttc atggttatga cattactgat gaggattgct ttacccattt 840
ggagcttggg ttcaaattct tcaacatcac cgattcaact gatggagttc tgggacaaac 900
ctataggagc gattacgtga acaaaatgaa ggtgaatgcg gtaatgccag tcatgggcgg 960
tgaccgtaag tacttgactt cgggactttt tagtgccgat tgtgctgttt ctcgctttgg 1020
tgggaaggtt cttgagaaag ccaattctgc ttctcctgtg catgagtatc cagccttgaa 1080
ctgcaagagt gggatggaag ggaatggctt ggtttgcaaa aaataattaa gttgctacag 1140
agcatgttgt atgctgaatg atgagctata aataattgag tttcagaaaa gtcttattag 1200
aaatgaagtg atcatagctt tacatgcaaa aa 1232

<210> 7
<211> 2288
<212> ADN
<213> coffea canephora

<220>
<221> Promotor
<222> (1)..(1324)
<223> promotor de CcDH2a

<220>
<221> exon
<222> (1414)..(1641)
<223> Exon I de cDH2a

<220>
<221> Intron
<222> (1642)..(1866)

549

<220>
<221> exon
<222> (1867)..(2121)
<223> Exon II de CcDH2a

<400> 7
atagtgacct taatagcgat cttgttgctt ttgatcgtca gaaaagtagt ggacatgacg 60

gaagaggtcc taagatgagt tccagttcca gcatgaaggg ctctttggcg aagcctttct 120

tgaggcgtca cttttctttt ggatctaaag gcagtagatc aatgtcagag aatcattctt 180

cctggaagag gggattcttc tgggcaaaat cgagaaagga ttaagttctg tctagagtta 240

caaaggtgag caacagtcac ggtttttttat tagggaatgg aaggattgga tccctttca 300

cgtagtgaac aacatatatt ttgcatggtt ggtcttagta cctataacac gaaaatgttc 360

ttcatccgtt ctattaatca ttaggcttta gtcatttaaa ttttttacat cccgcatttc 420

tcctcttgat tcttgttgat ttctgcagat tccacagttg ttcttcagat gggctacgaa 480

atgcatgcag ggagcaggca atcagccata aattcaaccc tgtcaaggaa gctggcattg 540

tctcgtgcaa atgtaggtta gcttttgaag atacactgca aagggaagac catacagatg 600

gggaaatgaa ttcattataa tataggaaaa aggaaagatg ataggggtca gggcgtccgt 660

gcatcatgaa actagttctc tttcattttg tacgatggct gtttactgtt taatttcatg 720

aaattagttt ggatatatgc gtagcgtttt accatcgcat ttctaaatcg atattctatg 780

ggccgaatta cgcgttggag acatcattgg gttgctcctc tcaatcccat ctctatctat 840

tgacggatcc ggatcatgat gttgaacctt tcaacttttg acttagatgg gatttgtgtt 900

cgcgtgttgt taacttgtta ctgaccgact cagaagacag cggattctga cttcaccacg 960

tgtctcttta gtgaaaattt aaaaggcatt tttcttctgt tcatagttta aaatgtaatg 1020

tgattattaa aagatcgttt ggtattattt caaggatgga tggattggat ggaagggata 1080

tctgatatat atcataccct tccaaaattc aggaccatga cgtatttaat atcccccagc 1140

ggaagacacg tgccttgatg tcttataggt ggcaatacac ttcagcttcc tctgctaata 1200

cgtgtgagga tcttcggtac catgcagaaa agaccgcggt gctccttcca ccgtcctcat 1260

ccctctcttg gctttttaa gtctcctgcg atatccaaaa tccaaacaaa gccgttatcg 1320

cagctaaaat tcgtcaaccc caagtctcag gctaccttaa tttcagtgcc ctttttcttt 1380

atttttttct aataacagga gtcctggaaa atg gct gac ttg cgt gat gaa tat 1434
Ala Asp Leu Arg Asp Glu Tyr
1 5

gga aat cct atg cag ttg acc gac cag tat ggc aac ccg gtt cag ctc 1482
Gly Asn Pro Met Gln Leu Thr Asp Gln Tyr Gly Asn Pro Val Gln Leu
10 15 20

aag gac gag tat ggc aac cca atg cag ctt agc ggt gta gct atc acc 1530
Lys Asp Glu Tyr Gly Asn Pro Met Gln Leu Ser Gly Val Ala Ile Thr
25 30 35

gcc ggg acg gct agt act gtc cat tct act gga acc gga cca act gct 1578
Ala Gly Thr Ala Ser Thr Val His Ser Thr Gly Thr Gly Pro Thr Ala
40 45 50 55

550

```
gcc act gga acc cag caa cat cag gag cag ctt cat cgg tct agc agc     1626
Ala Thr Gly Thr Gln Gln His Gln Glu Gln Leu His Arg Ser Ser Ser
                60                  65                  70

tca agc tct ggc tcg gtgagatact tgccaagtta caatgtgtgt gtctgtgtgt     1681
Ser Ser Ser Gly Ser
            75

gtataatgcg ccatcataat tgtttgcttg acagatcctg ttaataatga accgtaattt   1741

gacgtaaagt gtacacgttt tgtttttctg ggactaacat aatatcgaat caggctcctg   1801

ttgaatttga atgttgttag ctaaaagaaa attttggtgg ctgagttgtt gaatttggtt   1861

tatag acg gag gat gat gga caa gga gga aga aga aag aaa aaa ggg ttg   1911
      Thr Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys Lys Gly Leu
                80                  85                  90

aaa gaa aag ata aag gag aaa cta acg ggc ggt agg cac aag gac aga     1959
Lys Glu Lys Ile Lys Glu Lys Leu Thr Gly Gly Arg His Lys Asp Arg
            95                  100                 105

gac gat cag gag cac atc gat gat cag cac gcg cac agc gcc tct cct     2007
Asp Asp Gln Glu His Ile Asp Asp Gln His Ala His Ser Ala Ser Pro
        110                 115                 120

cca aca acc acc act ggc agc ggg acg tct act aca gtc ggg ggt cag     2055
Pro Thr Thr Thr Thr Gly Ser Gly Thr Ser Thr Thr Val Gly Gly Gln
    125                 130                 135

cag cat gaa aag aag agc atg gtg gag aag att atg gaa aag ctc cct     2103
Gln His Glu Lys Lys Ser Met Val Glu Lys Ile Met Glu Lys Leu Pro
140                 145                 150                 155

ggc cat cac gac acc cgc tagttaccta ccacaacata ctgtgatcat           2151
Gly His His Asp Thr Arg
                160

cgtgtaaaat ctctcctgat gcctaggaaa tctagattat gttaggcatt ttgtttggta   2211

tgtatgtgtg attaagacct tgttgtgcgc ttgaatcttg aacgtgcatg ggatttgctt   2271

ggtttgattt gatttgg                                                 2288


<210>  8
<211>  172
<212>  PRT
<213>  Coffea canephora

<400>  8

Met Ala Gln Tyr Gly Ala Glu Tyr Gly Asn Gln Lys Ser Gln Tyr Asp
1               5                   10                  15


Glu Tyr Gly Asn Pro Val Arg Gln Thr Asp Glu Tyr Gly Asn Pro Ala
            20                  25                  30


Arg His Gly Gly Thr Met Gly Asp Tyr Gly Thr Thr Gly Thr Thr Gly
        35                  40                  45


Ala Tyr Gly Gly Thr Thr Gly Ala His Gly Thr Tyr Ala Thr Gly Thr
    50                  55                  60
```


551

Thr Gly Thr Thr Gly Thr Gly Ala Tyr Ala Thr Gln Pro Gly Thr Asp
65 70 75 80

Val Gly Lys Glu His His Gly Leu Gly Gly Met Leu His Arg Ser Gly
85 90 95

Ser Gly Ser Ser Ser Ser Ser Glu Asp Asp Gly Gln Gly Gly Arg Arg
100 105 110

Lys Lys Gly Met Lys Glu Lys Ile Lys Glu Lys Leu Pro Gly Gly His
115 120 125

Lys Glu Ala Gln Pro Gly Gln Glu Tyr Ser Ser Ala Thr Ala Ala Pro
130 135 140

Gly Tyr Gly Gly Glu Gly Val Gln His Glu Lys Lys Gly Ile Met Asp
145 150 155 160

Lys Ile Lys Glu Lys Leu Pro Gly Gly His His Asn
165 170

<210> 9
<211> 175
<212> PRT
<213> Coffea canephora

<400> 9

Met Ala Gln Tyr Gly Ala Glu Tyr Gly Asn Gln Lys Ser Gln Tyr Asp
1 5 10 15

Glu Tyr Gly Asn Pro Val Arg Gln Thr Asp Glu Tyr Gly Asn Pro Ala
20 25 30

Arg His Gly Gly Thr Met Gly Asp Tyr Gly Thr Thr Gly Thr Thr Gly
35 40 45

Ala Tyr Gly Gly Thr Thr Gly Thr Ala Gly Ala His Gly Thr Tyr Ala
50 55 60

Thr Gly Thr Thr Gly Thr Thr Gly Thr Gly Ala Tyr Ala Thr Gln Pro
65 70 75 80

Gly Thr Asp Val Gly Lys Glu Arg His Gly Leu Gly Gly Met Leu His
85 90 95

Arg Ser Gly Ser Gly Ser Ser Ser Ser Ser Glu Asp Asp Gly Gln Gly
100 105 110

Gly Arg Arg Lys Lys Gly Met Lys Glu Lys Ile Lys Glu Lys Leu Pro
115 120 125

552

Gly Gly His Lys Glu Ala Gln Pro Gly Gln Glu Tyr Ser Ser Ala Thr
130 135 140

Ala Ala Pro Gly Tyr Gly Gly Glu Gly Glu Gln His Glu Lys Lys Gly
145 150 155 160

Ile Met Asp Lys Ile Lys Glu Lys Leu Pro Gly Gly His Arg Asn
165 170 175

<210> 10
<211> 162
<212> PRT
<213> Coffea canephora

<400> 10

Met Ala Asp Leu Arg Asp Glu Tyr Gly Asn Pro Met Gln Leu Thr Asp
1 5 10 15

Gln Tyr Gly Asn Pro Val Gln Leu Lys Asp Glu Tyr Gly Asn Pro Met
20 25 30

Gln Leu Ser Gly Val Ala Ile Thr Ala Gly Thr Ala Ser Ala Val His
35 40 45

Ser Thr Gly Thr Gly Pro Thr Ala Ala Thr Gly Thr Gln Gln His Gln
50 55 60

Glu Gln Leu His Arg Ser Ser Ser Ser Ser Ser Gly Ser Ser Glu Asp
65 70 75 80

Asp Gly Gln Gly Gly Arg Arg Lys Lys Lys Gly Leu Lys Glu Lys Ile
85 90 95

Lys Glu Lys Leu Thr Gly Gly Arg His Lys Asp Arg Asp Asp Gln Glu
100 105 110

His Ile Asp Asp Gln His Ala His Ser Ala Ser Pro Pro Thr Thr Thr
115 120 125

Thr Gly Ser Gly Thr Ser Thr Thr Val Gly Gly Gln Gln His Glu Lys
130 135 140

Lys Ser Met Val Glu Lys Ile Met Glu Lys Leu Pro Gly His His Asp
145 150 155 160

Thr Arg

<210> 11
<211> 162
<212> PRT
<213> Coffea canephora

<400> 11

Met Ala Asp Leu Arg Asp Glu Tyr Gly Asn Pro Met Gln Leu Thr Asp
1 5 10 15

Gln Tyr Gly Asn Pro Val Gln Leu Lys Asp Glu Tyr Gly Asn Pro Met
20 25 30

Gln Leu Ser Gly Val Ala Ile Thr Ala Gly Thr Ala Ser Ala Val His
35 40 45

Ser Thr Gly Thr Gly Pro Thr Ala Ala Thr Gly Thr Gln Gln Leu Gln
50 55 60

Glu Gln Leu His Arg Ser Ser Ser Ser Ser Ser Gly Ser Ser Glu Asp
65 70 75 80

Asp Gly Gln Gly Gly Arg Arg Lys Lys Lys Gly Leu Lys Glu Lys Ile
85 90 95

Lys Glu Lys Leu Thr Gly Gly Arg His Lys Asp Arg Asp Asp Gln Glu
100 105 110

His Ile Asp Asp Gln His Ala His Ser Ala Ser Pro Pro Thr Thr Thr
115 120 125

Thr Gly Ser Gly Thr Ser Thr Thr Val Gly Gly Gln Gln His Glu Lys
130 135 140

Lys Ser Met Val Glu Lys Ile Met Glu Lys Leu Pro Gly His His Asp
145 150 155 160

Thr Arg

<210> 12
<211> 227
<212> PRT
<213> Coffea canephora

<400> 12

Met Ala Glu Tyr Asp Gln Ser Asn Ile Lys Val Glu Glu Gly Ser Ala
1 5 10 15

Val Glu Ala Thr Asp Arg Gly Leu Phe Asn Leu Gly Lys Lys Glu Glu
20 25 30

Val Lys Lys Cys Asp Gln Gly Gln Ala Ile Ser Ala Glu Phe Asp Glu
35 40 45

Lys Val Arg Val Ser Glu Pro Asp Lys Glu Glu Gly Lys Lys His Gly
50 55 60

Gly Leu Leu Glu Lys Leu His Arg Ser Gly Ser Ser Ser Ser Ser Ser
65 70 75 80

Ser Glu Glu Glu Val Glu Glu Gly Gly Glu Lys Lys Lys Lys Lys Lys
85 90 95

Glu Lys Lys Gly Leu Lys Asp Lys Ile Lys Glu Lys Ile Ser Gly Asp
100 105 110

Lys Lys Asp Glu Glu Lys Val Glu Lys Cys Glu Glu Asp Thr Ser Ile
115 120 125

Pro Val Glu Lys Tyr Ala Glu Pro Ala His Ala Asp Ala Ala His Glu
130 135 140

Pro Glu Glu Lys Lys Gly Phe Leu Asp Lys Ile Lys Glu Lys Leu Pro
145 150 155 160

Gly Gly Gly Gln Lys Lys Thr Glu Glu Val Ala Ala Ala Ala Pro Pro
165 170 175

Pro Pro Pro Ala Glu Cys Thr Ala Thr Glu Gly Glu Ala Lys Asp Lys
180 185 190

Lys Gly Phe Leu Asp Lys Ile Lys Glu Lys Leu Pro Gly Tyr His Pro
195 200 205

Lys Thr Glu Glu Glu Lys Glu Lys Glu Lys Glu Lys Glu Lys Glu Ala
210 215 220

Gly Cys His
225

<210> 13
<211> 357
<212> PRT
<213> Coffea canephora

<400> 13

Met Gln Lys Met Thr Pro Leu Arg Cys Ile Asn Phe Ile Phe Leu Ala
1 5 10 15

Phe Trp Val Pro Ala Val Leu Ala Val Met Ala Glu Lys Pro Leu Val
20 25 30

Pro Thr Tyr Leu Ile Pro Lys Pro Pro Pro Pro Pro Ser Pro Val Lys
35 40 45

Pro Ser Val Pro Val Ile Pro Val Lys Pro Arg Ile Val Arg Cys Arg
50 55 60

Ser Thr Leu Phe Pro Leu Cys Phe Asn Ile Pro Phe Val Cys Pro Leu
65 70 75 80

Asp Cys Leu Thr Asn Cys Leu Val Asp Cys Val Thr Cys Lys Ala Tyr
85 90 95

Cys Ser Cys Asn Phe Pro Gly Ala Val Cys Gln Asp Pro Arg Phe Val
100 105 110

Gly Gly Asp Gly Asn Thr Phe Tyr Phe His Gly Arg Lys Asp Gln Asp
115 120 125

Phe Cys Leu Val Ser Asp Thr Asn Leu His Val Asn Gly His Phe Ile
130 135 140

Gly Lys Arg Lys Pro Asn Leu Arg Arg Asp Phe Thr Trp Val Gln Ala
145 150 155 160

Ile Gly Ile Met Phe Asp Asp His Arg Ile Leu Val Ala Ala Lys Arg
165 170 175

Thr Ser Thr Trp Asp Asp Asn Val Asp Arg Leu Ala Ile Ser Ile Asp
180 185 190

Gly Asn Pro Ile Ser Leu Pro Thr Glu Glu Gly Ser Lys Trp Gln Leu
195 200 205

Pro Ala Pro Ser Asn Val Ser Ile Met Arg Thr Ser Asn Asn Asn Gly
210 215 220

Leu Val Val Glu Ala Val Asn Asn Phe Arg Ile Thr Ala Asn Val Val
225 230 235 240

Pro Ile Thr Ala Gln Glu Ser Lys Val His Gly Tyr Asp Ile Thr Asp
245 250 255

Glu Asp Cys Phe Thr His Leu Glu Leu Gly Phe Lys Phe Phe Asn Ile
260 265 270

Thr Asp Ser Thr Asp Gly Val Leu Gly Gln Thr Tyr Arg Ser Asp Tyr
275 280 285

Val Asn Lys Met Lys Val Asn Ala Val Met Pro Val Met Gly Gly Asp
290 295 300

Arg Lys Tyr Leu Thr Ser Gly Leu Phe Ser Ala Asp Cys Ala Val Ser
305 310 315 320

Arg Phe Gly Gly Lys Val Leu Glu Lys Ala Asn Ser Ala Ser Pro Val
325 330 335

His Glu Tyr Pro Ala Leu Asn Cys Lys Ser Gly Met Glu Gly Asn Gly
340 345 350

Leu Val Cys Lys Lys
355

<210> 14
<211> 130
<212> PRT
<213> Lycopersicon esculentum

<400> 14

Met Ala Gln Tyr Gly Asn Gln Asp Gln Met Arg Lys Thr Asp Glu Tyr
1 5 10 15

Gly Asn His Val Gln Glu Thr Gly Val Tyr Gln Gly Thr Gly Thr Gly
20 25 30

Gly Met Met Gly Gly Thr Gly Thr Gly Gly Met Met Gly Gly Thr Gly
35 40 45

Gly Glu Tyr Gly Thr Gln Gly Met Gly Thr Gly Thr His His His Glu
50 55 60

Gly Gln Gln Gln Leu Arg Arg Ser Asp Ser Ser Ser Ser Ser Glu Asp
65 70 75 80

Asp Gly Glu Gly Gly Arg Arg Lys Lys Gly Leu Lys Glu Lys Ile Met
85 90 95

Glu Lys Met Pro Gly Gln His Glu Gly Glu Tyr Gly Gln Thr Thr Gly
100 105 110

Glu Glu Lys Lys Gly Met Met Asp Lys Ile Lys Asp Lys Ile Pro Gly
115 120 125

Met His
130

<210> 15
<211> 157
<212> PRT
<213> Solanum commersonii

<400> 15

Met Ala His Tyr Glu Asn Gln Tyr Ser Ala Gly Gln Ala Leu Gln Lys
1 5 10 15

Asp Glu Tyr Gly Asn Pro Val Arg Gln Thr Asp Glu Tyr Gly Asn Pro
20 25 30

Ile Gln Gln Thr Gly Gly Thr Met Gly Glu Tyr Gly Thr Thr Gly Thr
35 40 45

557

Gly Tyr Gly Thr Gln Ala Gly His Thr Thr Gly Val Leu Gly Gly Asp
50 55 60

Gln Arg Gln His Gly Thr Leu Gly Gly Met Leu His Arg Ser Gly Ser
65 70 75 80

Ser Ser Ser Ser Ser Ser Ser Glu Asp Asp Gly His Gly Gly Arg Arg
85 90 95

Lys Lys Lys Gly Ile Lys Asp Lys Val Lys Glu Lys Leu Pro Gly Gly
100 105 110

His Arg Asp Asp Leu Ala His Ser Thr Ala Thr Thr Thr Thr Thr Gly
115 120 125

Tyr Gly Met Asp Gly Thr His Glu Lys Lys Gly Ile Met Glu Lys Ile
130 135 140

Lys Glu Lys Leu Pro Gly His His Gly Pro Gly His His
145 150 155

<210> 16
<211> 186
<212> PRT
<213> Arabidopsis thaliana

<400> 16

Met Ala Ser Tyr Gln Asn Arg Pro Gly Gly Gln Ala Thr Asp Glu Tyr
1 5 10 15

Gly Asn Pro Ile Gln Gln Gln Tyr Asp Glu Tyr Gly Asn Pro Met Gly
20 25 30

Gly Gly Gly Tyr Gly Thr Gly Gly Gly Gly Gly Ala Thr Gly Gly Gln
35 40 45

Gly Tyr Gly Thr Gly Gly Gln Gly Tyr Gly Ser Gly Gly Gln Gly Tyr
50 55 60

Gly Thr Gly Gly Gln Gly Tyr Gly Thr Gly Thr Gly Thr Glu Gly Phe
65 70 75 80

Gly Thr Gly Gly Gly Ala Arg His His Gly Gln Glu Gln Leu His Lys
85 90 95

Glu Ser Gly Gly Gly Leu Gly Gly Met Leu His Arg Ser Gly Ser Gly
100 105 110

Ser Ser Ser Ser Ser Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys
115 120 125

```
Gly Ile Thr Gln Lys Ile Lys Glu Lys Leu Pro Gly His His Asp Gln
    130                 135                 140

Ser Gly Gln Ala Gln Ala Met Gly Gly Met Gly Ser Gly Tyr Asp Ala
145                 150                 155                 160

Gly Gly Tyr Gly Gly Glu His His Glu Lys Lys Gly Met Met Asp Lys
                165                 170                 175

Ile Lys Glu Lys Leu Pro Gly Gly Gly Arg
            180                 185
```

<210> 17
<211> 208
<212> PRT
<213> Nicotiana tabacum

<400> 17

```
Met Ala Asp Gln Tyr Glu Lys Lys Val Glu Glu Gly Ser Ala Asn Val
1               5                   10                  15

Glu Ala Thr Asp Arg Gly Leu Phe Asp Phe Leu Gly Lys Lys Glu Glu
            20                  25                  30

Glu Lys Pro Thr His Ala Gln Glu Glu His Ala Ile Ser Ser Glu Phe
        35                  40                  45

Val Glu Lys Val Lys Val Ser Glu Glu Val Ala Glu Tyr Lys Glu Glu
    50                  55                  60

Glu Lys Lys Glu Glu His Asn Lys Glu Glu Lys Lys Leu His Arg Ser
65                  70                  75                  80

Ser Ser Ser Ser Ser Ser Ser Ser Asp Glu Glu Glu Glu Ile Gly Glu
                85                  90                  95

Asp Gly Gln Lys Ile Lys Lys Lys Lys Lys Lys Gly Leu Lys Asp Lys
            100                 105                 110

Ile Lys Asp Lys Ile Ser Gly Glu His Lys Glu Glu Glu Lys Ala Gly
        115                 120                 125

Glu Asp Thr Ala Val Pro Val Glu Lys Tyr Glu Glu Thr Glu Glu Lys
    130                 135                 140

Lys Gly Phe Leu Asp Lys Ile Lys Glu Lys Leu Pro Gly Gly Gly Gln
145                 150                 155                 160

Lys Lys Thr Glu Glu Val Ala Pro Pro Pro Pro Pro Ala Ala Glu His
                165                 170                 175
```

Glu Ala Glu Gly Lys Glu Lys Lys Gly Phe Leu Asp Lys Ile Lys Glu
180 185 190

Lys Leu Pro Gly Tyr His Ser Lys Thr Glu Glu Lys Lys Lys Lys Lys
195 200 205

<210> 18
<211> 209
<212> PRT
<213> Solanum tuberosum

<400> 18

Met Ala Asp Gln Tyr Glu Gln Asn Lys Pro Ser Val Glu Glu Thr Val
1 5 10 15

Gly Ala Asn Val Glu Ala Thr Asp Arg Gly Leu Phe Asp Phe Ile Gly
20 25 30

Lys Lys Lys Glu Glu Lys Pro Ser His Ala His Glu Glu Glu Ala Ile
35 40 45

Ser Ser Glu Phe Cys Glu Lys Val Lys Val Ser Glu Glu Glu His Lys
50 55 60

Glu Glu Glu Lys Lys Glu Glu Lys Lys Leu His Arg Ser Ser Ser Ser
65 70 75 80

Ser Ser Ser Ser Ser Asp Glu Glu Glu Glu Ile Gly Glu Asp Gly Gln
85 90 95

Ile Ile Lys Lys Lys Lys Lys Lys Gly Leu Lys Glu Lys Ile Lys Glu
100 105 110

Lys Ile Ser Gly Asp His Lys Glu Glu Val Lys Thr Glu Asp Thr Ser
115 120 125

Val Pro Val Glu Lys Tyr Glu Glu Thr Glu Glu Lys Lys Gly Phe Leu
130 135 140

Asp Lys Ile Lys Glu Lys Leu Pro Gly Gly Gly His Lys Lys Thr Glu
145 150 155 160

Glu Val Ala Ala Pro Pro Pro Pro Pro Pro Ala Ala Val Glu His Glu
165 170 175

Ala Glu Gly Lys Glu Lys Lys Gly Phe Leu Asp Lys Ile Lys Glu Lys
180 185 190

Leu Pro Gly Tyr His Ser Lys Thr Glu Glu Glu Lys Glu Lys Glu Lys
195 200 205

Asn

560

<210> 19
<211> 265
<212> PRT
<213> Arabidopsis thaliana

<400> 19

```
Met Ala Glu Glu Tyr Lys Asn Asn Val Lys Glu His Glu Thr Pro Thr
1               5                  10                  15

Val Ala Thr Glu Glu Ser Pro Ala Thr Thr Thr Glu Val Thr Asp Arg
            20                  25                  30

Gly Leu Phe Asp Phe Leu Gly Lys Lys Glu Glu Glu Val Lys Pro Gln
        35                  40                  45

Glu Thr Thr Thr Leu Glu Ser Glu Phe Asp His Lys Ala Gln Ile Ser
    50                  55                  60

Glu Pro Glu Leu Ala Ala Glu His Glu Glu Val Lys Glu Asn Lys Ile
65                  70                  75                  80

Thr Leu Leu Glu Glu Leu Gln Glu Lys Thr Glu Glu Asp Glu Glu Asn
                85                  90                  95

Lys Pro Ser Val Ile Glu Lys Leu His Arg Ser Asn Ser Ser Ser Ser
            100                 105                 110

Ser Ser Ser Asp Glu Glu Gly Glu Glu Lys Lys Glu Lys Lys Lys Lys
        115                 120                 125

Ile Val Glu Gly Glu Glu Asp Lys Lys Gly Leu Val Glu Lys Ile Lys
    130                 135                 140

Glu Lys Leu Pro Gly His His Asp Lys Thr Ala Glu Asp Asp Val Pro
145                 150                 155                 160

Val Ser Thr Thr Ile Pro Val Pro Val Ser Glu Ser Val Val Glu His
                165                 170                 175

Asp His Pro Glu Glu Glu Lys Lys Gly Leu Val Glu Lys Ile Lys Glu
            180                 185                 190

Lys Leu Pro Gly His His Asp Glu Lys Ala Glu Asp Ser Pro Ala Val
        195                 200                 205

Thr Ser Thr Pro Leu Val Val Thr Glu His Pro Val Glu Pro Thr Thr
    210                 215                 220

Glu Leu Pro Val Glu His Pro Glu Glu Lys Lys Gly Ile Leu Glu Lys
225                 230                 235                 240
```

```
Ile Lys Glu Lys Leu Pro Gly Tyr His Ala Lys Thr Thr Glu Glu Glu
                245                 250                 255

Val Lys Lys Glu Lys Glu Ser Asp Asp
            260                 265
```

<210> 20
<211> 337
<212> PRT
<213> Arabidopsis thaliana

<400> 20

```
Met Glu Ile Ser Lys Thr Leu Leu Leu Val Ile Ser Leu Val Ala Ala
1               5                  10                  15

Thr Cys Phe Leu Gln Ala Lys Ala Ala Gly Val Tyr Cys Ser Asn Pro
            20                  25                  30

Tyr Thr Arg Cys Tyr Arg Lys Tyr Ile Arg Cys Pro Glu Glu Cys Pro
        35                  40                  45

Ser Lys Thr Ala Met Asn Ser Lys Asn Lys Val Cys Tyr Ala Asp Cys
    50                  55                  60

Asp Arg Pro Thr Cys Lys Ser Gln Cys Arg Met Arg Lys Pro Asn Cys
65                  70                  75                  80

Asn Arg Pro Gly Ser Ala Cys Tyr Asp Pro Arg Phe Ile Gly Gly Asp
                85                  90                  95

Gly Ile Val Phe Tyr Phe His Gly Lys Ser Asn Glu Glu Phe Ser Leu
            100                 105                 110

Val Ser Asp Ser Asp Leu Gln Ile Asn Gly Arg Phe Ile Gly His Arg
        115                 120                 125

Pro Ala Gly Arg Ala Arg Asp Phe Thr Trp Ile Gln Ala Leu Gly Phe
    130                 135                 140

Leu Phe Asn Ser Asn Lys Phe Ser Leu Glu Ala Ala Lys Thr Ala Ser
145                 150                 155                 160

Trp Asp Asn Glu Ile Asp His Leu Lys Phe Ser Tyr Asp Gly Gln Asp
                165                 170                 175

Leu Ser Val Pro Glu Glu Thr Leu Ser Thr Trp Tyr Ser Pro Asn Lys
            180                 185                 190

Asp Ile Lys Ile Glu Arg Val Ser Met Arg Asn Ser Val Ile Val Thr
        195                 200                 205
```

Ile Lys Asp Lys Ala Glu Ile Met Ile Asn Val Val Pro Val Thr Lys
210 215 220

Glu Asp Asp Arg Ile His Ser Tyr Lys Val Pro Ser Asp Asp Cys Phe
225 230 235 240

Ala His Leu Glu Val Gln Phe Arg Phe Phe Asn Leu Ser Pro Lys Val
245 250 255

Asp Gly Ile Leu Gly Arg Thr Tyr Arg Pro Asp Phe Gln Asn Pro Ala
260 265 270

Lys Pro Gly Val Ala Met Pro Val Val Gly Gly Glu Asp Ser Phe Lys
275 280 285

Thr Ser Ser Leu Leu Ser Asn Asp Cys Lys Thr Cys Ile Phe Ser Glu
290 295 300

Ser Gln Ala Glu Ile Asp Ser Val Lys Ser Glu Ile Glu Tyr Ala Thr
305 310 315 320

Leu Asp Cys Thr Arg Gly Ala Ser Ser Gly Tyr Gly Ile Val Cys Arg
325 330 335

Lys

<210> 21
<211> 320
<212> PRT
<213> Picea glauca

<400> 21

Met Ala Ala Asn Lys Leu Met Asn Val Val Ala Val Ala Val Cys Val
1 5 10 15

Met Val Met Val Asn Ala Ala Ser Ala Val Gly Lys Ala Lys Cys Thr
20 25 30

Asp Lys Trp Tyr Pro Arg Cys Tyr Gly Tyr Gln Tyr Asp Cys Pro Ala
35 40 45

Asn Cys Pro Tyr Asn Cys Asp Met Asp Cys Lys Thr Cys Lys Thr Val
50 55 60

Cys Pro Cys Asp Lys Pro Gly Gly Val Cys Gln Asp Pro Arg Phe Ile
65 70 75 80

Gly Gly Asp Gly Ile Met Phe Tyr Phe His Gly Lys Arg Asp Gln Asp
85 90 95

Phe Cys Leu Ile Ser Asp Ser Asn Leu His Ile Asn Ala His Phe Ile
100 105 110

Gly Lys Arg Gly Gln Gly Met Gly Arg Asp Phe Thr Trp Val Gln Ser
115 120 125

Ile Gly Val Leu Leu Glu Asp Gly Arg Gln Phe Tyr Leu Gly Ala Lys
130 135 140

Lys Val Ser Thr Trp Asp Asn Ser Val Asp Gln Leu Thr Met Ala Leu
145 150 155 160

Asn Gly Gln Thr Leu Thr Leu Pro Pro Gly Glu Gly Ala Thr Trp Ala
165 170 175

Thr Ala Ser Gly Leu Asn Val Thr Arg Ser Asp Arg Ala Asn Glu Val
180 185 190

Val Val Gln Val Glu Asp Lys Leu Lys Ile Ser Ala Arg Val Val Pro
195 200 205

Ile Ser Glu Glu Glu Ser Arg Val His Asn Tyr Gly Ile Ile Ala Gly
210 215 220

Glu Asp Cys Phe Ala His Leu Glu Leu Ser Phe Lys Phe Tyr Ser Leu
225 230 235 240

Ser Pro Asn Val Ser Gly Val Leu Gly Gln Thr Tyr Gly Ala Glu Tyr
245 250 255

Arg Ser Pro Val Lys Met Gly Val Ala Met Pro Ile Met Gly Gly Glu
260 265 270

Ser Asn Tyr Val Thr Ser Asn Leu Phe Ala Ala Asp Cys Lys Val Ala
275 280 285

Arg Phe Ala Ser Ser Ser Asp Asp Glu Tyr Ala Ile Thr Ser Ala Leu
290 295 300

Asp Cys Asn Ser Gly Arg Gly Ser Gly His Gly Ile Val Ser Arg Arg
305 310 315 320

<210> 22
<211> 349
<212> PRT
<213> Zea mays

<400> 22

Met Ala Arg Leu Gly Ala Leu Ile Pro Leu Ala Ile Val Leu Leu Ala
1 5 10 15

Ala Val Ala Ala Thr Ala Ala Pro Ser Asp Arg Pro Pro Lys Ala Gln

20 25 30

Gly Pro Lys Pro His Lys Glu Lys Glu Lys Glu Lys Pro Lys Pro Met
35 40 45

Lys Val Lys Cys Arg Pro Arg Lys Leu Tyr Pro Tyr Cys Pro Gly Lys
50 55 60

Pro Met Glu Cys Pro Ala Glu Cys Ser Gln Ser Cys Tyr Ala Asp Cys
65 70 75 80

Ser Ser Cys Lys Pro Val Cys Val Cys Ser Val Pro Gly Ala Cys Gly
85 90 95

Asp Pro Arg Phe Ile Gly Gly Asp Gly Asn Ala Phe Tyr Phe His Gly
100 105 110

Arg Arg Asp Ala Asp Phe Cys Val Leu Ser Asp Arg Asp Leu His Ile
115 120 125

Asn Ala His Phe Ile Gly Lys His Gly Ala Asp Gly Met Ser Arg Asp
130 135 140

Phe Thr Trp Ile Gln Ala Ile Ala Val Leu Phe Asp Gly His Glu Leu
145 150 155 160

Tyr Val Gly Ala Arg Lys Thr Ala Ala Trp Asp Asp Asp Val Asp Arg
165 170 175

Met Glu Leu Thr Leu Asp Gly Glu Pro Val Arg Leu Leu Pro Gly Thr
180 185 190

Asp Ala Ala Trp Thr Ser Gly Ala Val Pro Ala Leu Ser Val Thr Arg
195 200 205

Thr Ser Ala Ala Asn Gly Val Leu Val Ser Leu Asp Gly Arg Phe Thr
210 215 220

Ile Arg Ala Asn Ala Val Pro Ile Thr Glu Glu Glu Ser Arg Val His
225 230 235 240

Arg Tyr Gly Val Thr Ala Asp Asp Cys Leu Ala His Leu Asp Leu Ala
245 250 255

Phe Lys Phe Gly Ala Leu Thr Ala Asp Val His Gly Val Val Gly Gln
260 265 270

Thr Tyr Arg Ser Asp Tyr Val Asn Arg Phe Asp Val Lys Ala Ser Met
275 280 285

Pro Thr Met Gly Gly Asp Ser Asn Tyr Thr Thr Ser Ser Leu Phe Ala

290 295 300

Ala Asp Cys Ala Val Ala Arg Tyr Ala Pro Ser Gly Gly Ser Arg Asp
305 310 315 320

Arg Asp Asp Gly Val Ala Met Val Ser Glu Ile Ala Gly Ile Thr Cys
325 330 335

Ser Ser Gly Met Gly Gly Gln Gly Val Val Cys Lys Lys
340 345

<210> 23
<211> 21
<212> ADN
<213> Artificial

<220>
<223> cebador delant de RPL39

<400> 23
tggcgaagaa gcagaggcag a 21

<210> 24
<211> 20
<212> DNA
<213> Artificial

<220>
<223> cebador rev. de RPL39

<400> 24
ttgaggggga gggtaaaaag 20

<210> 25
<211> 19
<212> ADN
<213> Artificial

<220>
<223> cebador delant. de DH1

<400> 25
gaagaagggg atgaaggag 19

<210> 26
<211> 20
<212> ADN
<213> Artificial

<220>
<223> cebador rev. de DH1

<400> 26
tacggacaaa cacactacag 20

<210> 27
<211> 19
<212> ADN
<213> Artificial

<220>
<223> cebador delant. de DH2

<400> 27
cctccaacaa ccaccactg 19

<210> 28
<211> 19
<212> ADN
<213> Artificial

<220>
<223> cebador rev. de DH2

<400> 28
tcaagcgcac aacaaggtc 19

<210> 29
<211> 21
<212> ADN
<213> Artificial

<220>
<223> cebador delant de DH3

<400> 29
aggtggtggt cagaagaaga c 21

<210> 30
<211> 20
<212> DNA
<213> Artificial

<220>
<223> cebador rev de DH3

<400> 30
gacacactgg aaagctgcta 20

<210> 31
<211> 21
<212> ADN
<213> Artificial

<220>
<223> cebador delant de LEA1

<400> 31
ccaataacag ctcaagaatc a 21

<210> 32
<211> 19
<212> DNA
<213> Artificial

<220>
<223> cebador rev de LEA1

<400> 32
ttcccttcca tcccactct 19

<210> 33

567

<211> 21
<212> ADN
<213> Artificial

<220>
<223> rpl39-F1

<400> 33
gaacaggccc atcccttatt g 21

<210> 34
<211> 17
<212> ADN
<213> Artificial

<220>
<223> rpl39-R1

<400> 34
cggcgcttgg cattgta 17

<210> 35
<211> 16
<212> ADN
<213> Artificial

<220>
<223> rpl39-MPG

<400> 35
atgcgcactg acaaca 16

<210> 36
<211> 19
<212> ADN
<213> Artificial

<220>
<223> DH2a-F1

<400> 36
ggaggcacaa ggacagaga 19

<210> 37
<211> 17
<212> ADN
<213> Artificial

<220>
<223> DH2a-R1

<400> 37
gctgtgcgcg tgctgat 17

<210> 38
<211> 15
<212> ADN
<213> Artificial

<220>
<223> DH2a-MGB

<400> 38

568

caggagcaca tcgat 15

<210> 39
<211> 29
<212> ADN
<213> Artificial

<220>
<223> cebador 1 de DH2

<400> 39
tgtgctcctg atgctctctg tccttgtgc 29

<210> 40
<211> 28
<212> ADN
<213> Artificial

<220>
<223> DH2a geneup

<400> 40
atagtgacct taatagcgat cttgttgc 28

<210> 41
<211> 26
<212> ADN
<213> Artificial

<220>
<223> DH2a genelow

<400> 41
ccaaatcaaa tcaaaccaag caaatc 26

<210> 42
<211> 10
<212> ADN
<213> Arabidopsis thaliana

<400> 42
ryacgtggyr 10

<210> 43
<211> 29
<212> ADN
<213> Artificial

<220>
<223> TG-TG698

<400> 43
ttgaagcttg tggacatgac ggaagaggt 29

<210> 44
<211> 36
<212> ADN
<213> Artificial

<220>
<223> TG-TG743

569

<400> 44
gcagatctac catggacgac tcctgttatt agaaaa 36

<210> 45
<211> 23
<212> ADN
<213> Artificial

<220>
<223> cebador 1 CcDH1

<400> 45
cactggcact actggagcct atg 23

<210> 46
<211> 15
<212> ADN
<213> Artificial

<220>
<223> cebador 2 de CcDH1

<400> 46
ctggagcaca tggga 15

570

AAC49618_tomate
CAA75798_papa
NP_201441_Arabidopsis
CcDH1a
CcDH1b
CcDH2a

FIG. 1

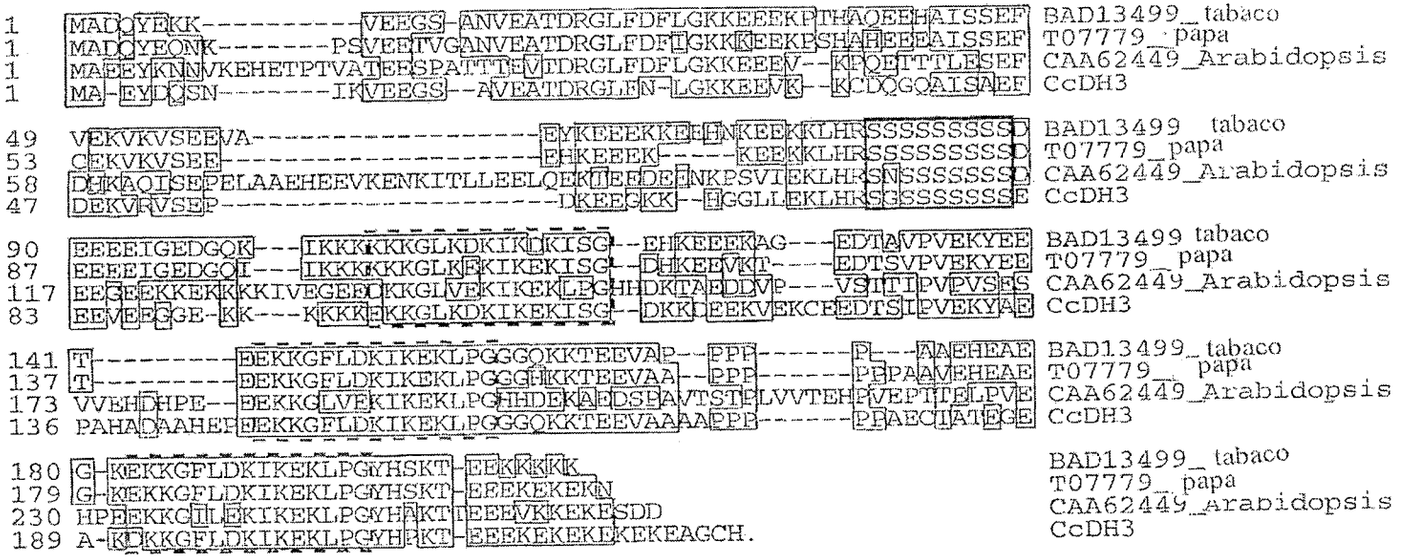


FIG. 2

NP200248_Arabidopsis
T09288_Picea_glauca
BAA75477_Zea_mays
CcLEA1

FIG. 3

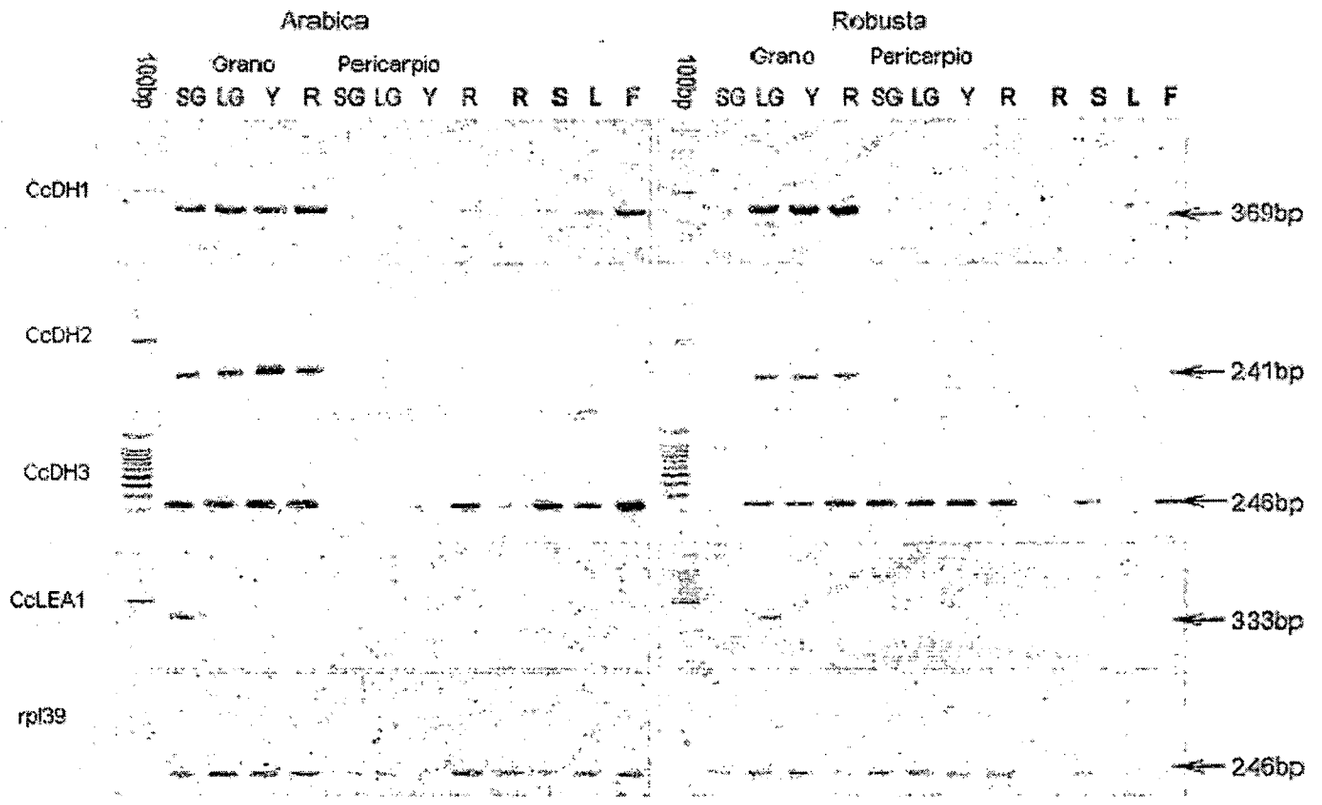

Arabica
Robusta
Grano
Pericarpio
100bp
SG LG Y R SG LG Y R R S L F
CcDH1
CcDH2
CcDH3
CcLEA1
rpl39
369bp
241bp
246bp
333bp
246bp

FIG. 4

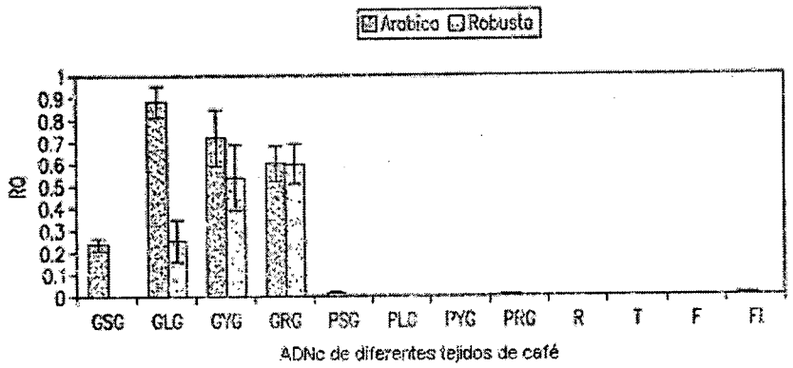

Arabica
Robusta
RQ
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
GSG
GLG
GYG
GRG
PSG
PLG
PYG
PRG
R
T
F
Fl
ADNc de diferentes tejidos de café

FIG. 5

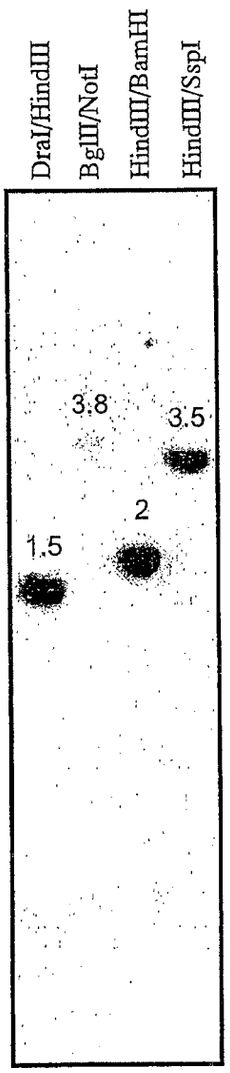

DraI/HindIII
BglII/NotI
HindIII/BamHI
HindIII/SspI
3.8
3.5
2
1.5

FIG. 6

```
atagtgacct taatagcgat cttgttgctt ttgatcgtca gaaaagtagt ggacatgacg      60
gaagaggtcc taagatgagt tccagttcca gcatgaaggg ctctttggcg aagcctttct     120
tgaggcgtca cttttctttt ggatctaaag gcagtagatc aatgtcagag aatcattctt     180
cctggaagag gggattcttc tgggcaaaat cgagaaagga ttaagttctg tctagagtta     240
caaaggtgag caacagtcac ggtttttat tagggaatgg aaggattgga tccctttca      300
cgtagtgaac aacatatatt ttgcatggtt ggtcttagta cctataacac gaaaatgttc     360
ttcatccgtt ctattaatca ttaggcttta gtcatttaaa ttttttacat cccgcatttc     420
tcctcttgat tcttgttgat ttctgcagat tccacagttg ttcttcagat gggctacgaa     480
atgcatgcag ggagcaggca atcagccata aattcaaccc tgtcaaggaa gctggcattg     540
tctcgtgcaa atgtaggtta gcttttgaag atacactgca aagggaagac catacagatg     600
gggaaatgaa ttcattataa tataggaaaa aggaaagatg ataggggtca gggcgtccgt     660
gcatcatgaa actagttctc tttcattttg tacgatggct gtttactgtt taatttcatg     720
aaattagttt ggatatatgc gtagcgtttt accatcgcat ttctaaatcg atattctatg     780
ggccgaatta cgcgttggag acatcattgg gttgctcctc tcaatcccat ctctatctat     840
tgacggatcc ggatcatgat gttgaacctt tcaacttttg acttagatgg gatttgtgtt     900
cgcgtgttgt taacttgtta ctgaccgact cagaagacag cggattctga cttcaccacg     960
tgtctcttta gtgaaaattt aaaaggcatt tttcttctgt tcatagttta aaatgtaatg    1020
tgattattaa aagatcgttt ggtattattt caaggatgga tggattggat ggaagggata    1080
tctgatatat atcatacccc tccaaaattc aggaccatga cgtatttaat atcccccagc    1140
ggaagacacg tgccttgatg tcttataggt ggcaatacac ttcagcttcc tctgctaata    1200
cgtgtgagga tcttcggtac catgcagaaa agaccgcggt gctccttcca ccgtcctcat    1260
ccctctcttg gctttttaa gtctcctgcg atatccaaaa tccaaacaaa gccgttatcg     1320
cagctaaaat tcgtcaaccc caagtctcag gctaccttaa tttcagtgcc ctttttcttt    1380
atttttttct aataacagga gtcctggaaa atg gct gac ttg cgt gat gaa tat      1434

                                    1               5
gga aat cct atg cag ttg acc gac cag tat ggc aac ccg gtt cag ctc      1482
Gly Asn Pro Met Gln Leu Thr Asp Gln Tyr Gly Asn Pro Val Gln Leu
    10                  15                  20
aag gac gag tat ggc aac cca atg cag ctt agc ggt gta gct atc acc      1530
Lys Asp Glu Tyr Gly Asn Pro Met Gln Leu Ser Gly Val Ala Ile Thr
25                  30                  35                  40
gcc ggg acg gct agt act gtc cat tct act gga acc gga cca act gct      1578
Ala Gly Thr Ala Ser Thr Val His Ser Thr Gly Thr Gly Pro Thr Ala
                45                  50                  55
gcc act gga acc cag caa cat cag gag cag ctt cat cgg tct agc agc      1626
Ala Thr Gly Thr Gln Gln His Gln Glu Gln Leu His Arg Ser Ser Ser
            60                  65                  70
tca agc tct ggc tcg   gtgagatact tgccaagtta caatgtgtgt gtctgtgtgt     1681
Ser Ser Ser Gly Ser
        75
gtataatgcg ccatcataat tgtttgcttg acagatcctg ttaataatga accgtaattt     1741
gacgtaaagt gtacacgttt tgtttttctg ggactaacat aatatcgaat caggctcctg     1801
ttgaatttga atgttgttag ctaaaagaaa attttggtgg ctgagttgtt gaatttggtt     1861
tatag acg gag gat gat gga caa gga gga aga aga aag aaa aaa ggg ttg      1911
      Thr Glu Asp Asp Gly Gln Gly Gly Arg Arg Lys Lys Lys Gly Leu
            80                  85                  90
aaa gaa aag ata aag gag aaa cta acg ggc ggt agg cac aag gac aga       1959
Lys Glu Lys Ile Lys Glu Lys Leu Thr Gly Gly Arg His Lys Asp Arg
        95                  100                 105
gac gat cag gag cac atc gat gat cag cac gcg cac agc gcc tct cct       2007
Asp Asp Gln Glu His Ile Asp Asp Gln His Ala His Ser Ala Ser Pro
    110                 115                 120
cca aca acc acc act ggc agc ggg acg tct act aca gtc ggg ggt cag       2055
Pro Thr Thr Thr Thr Gly Ser Gly Thr Ser Thr Thr Val Gly Gly Gln
125                 130                 135                 140
cag cat gaa aag aag agc atg gtg gag aag att atg gaa aag ctc cct       2103
Gln His Glu Lys Lys Ser Met Val Glu Lys Ile Met Glu Lys Leu Pro
                145                 150                 155
ggc cat cac gac acc cgc tag ttaccta ccacaacata ctgtgatcat             2151
Gly His His Asp Thr Arg STOP
            160
cgtgtaaaat ctctcctgat gcctaggaaa tctagattat gttaggcatt ttgtttggta     2211
tgtatgtgtg attaagacct tgttgtgcgc ttgaatcttg aacgtgcatg ggatttgctt     2271
ggtttgattt gatttgg
```

FIG. 7

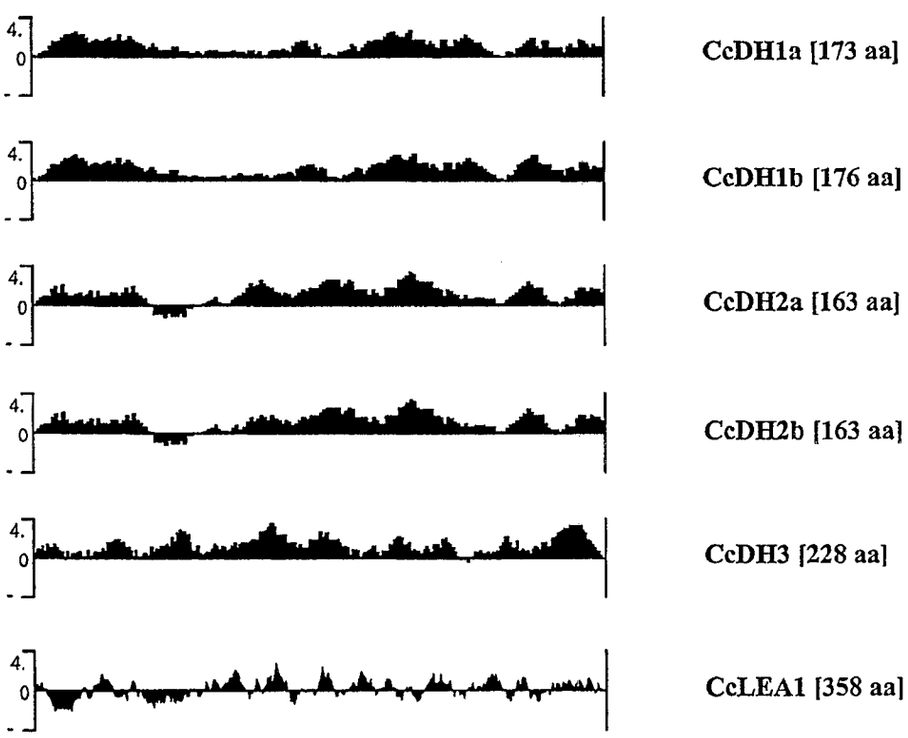

4.
0
-
CcDH1a [173 aa]
CcDH1b [176 aa]
CcDH2a [163 aa]
CcDH2b [163 aa]
CcDH3 [228 aa]
CcLEA1 [358 aa]

FIG. 8

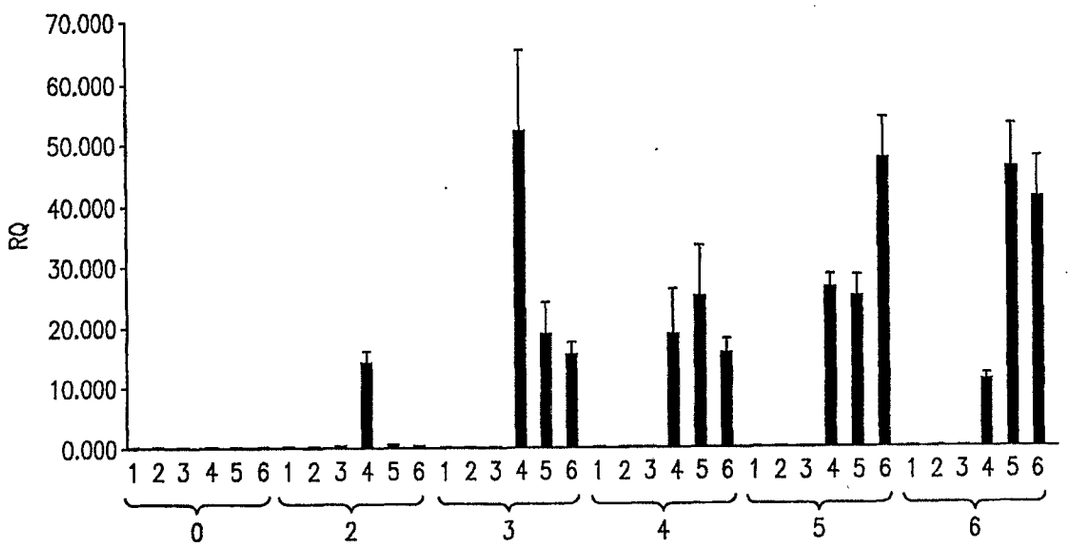

70.000
60.000
50.000
40.000
30.000
20.000
10.000
0.000
RQ
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6
0
2
3
4
5
6

FIG. 9

577

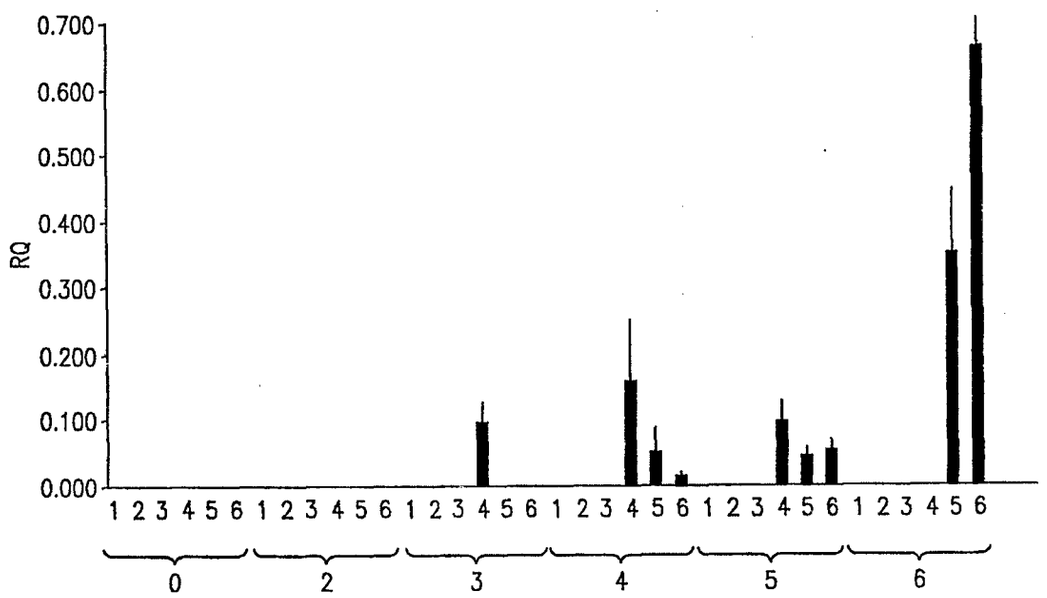

0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
RQ
1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6
0
2
3
4
5
6

FIG. 10

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

No. 07-137519- -00000-0000

Fecha: 2007-12-28 17:29:00 Dep: 2020 NUECREACIONI
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 578

Industria y Comercio
SUPERINTENDENCIA

REQUISITOS MINIMOS PARA ADMISION A TRAMITE
PCT

PATENTE DE INVENCION [x] MODELO DE UTILIDAD [ ]

- Indicación de que se solicita la concesión de una patente [x] 1
- Datos de identificación del solicitante o de la persona que presenta la solicitud [x] 1
- Descripción de la invención [x] 412
- Dibujos de ser estos pertinentes [x] 5+1
- Comprobante de pago de las tasas establecidas [x] 2

COMPLETA [x] INCOMPLETA [ ]

DISEÑO INDUSTRIAL [ ]

- Indicación de que se solicita el registro de Diseño Industrial [ ]
- Datos de identificación del solicitante o de la persona que presenta la solicitud [ ]
- Representación gráfica y fotográfica del Diseño Industrial o muestra del material que incorpora el diseño [ ]
- Comprobante de pago de las tasas establecidas [ ]

COMPLETA [ ] INCOMPLETA [ ]

ESQUEMA DE TRAZADO [ ]

- Indicación de que se solicita el registro de un Esquema de Trazado [ ]
- Datos de identificación del solicitante o de la persona que presenta la solicitud [ ]
- Representación gráfica del esquema trazado [ ]
- Comprobante de pago de las tasas establecidas [ ]

COMPLETA [ ] INCOMPLETA [ ]

Firma Responsable

Fecha_____

Carrera 13 No. 27-00 Piso 5, 7 y 10
Conmutador: 3 82 08 40
Fax: 350 52 20
E-mail: info@sic.gov.co
www.sic.gov.co
Bogotá, D.C., Colombia