

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO	
RAD: 13-67186- -15	FECHA: 2016-03-08 12:14:02
DEP: 2020 DIRECCION DE NUEVAS CREACIONES	EVE: 378 FASE NACIONAL INCLUIDO CAPITULO I
TRA: 11 PCT-PATENTE DE INVENCION CAPITULOS I Y II	FOLIOS: 9
ACT: 519 TRASLADOINFORME	

DIRECCIÓN DE NUEVAS CREACIONES

2020
Bogotá, D.C.,

Abogado(a)/Señor(a)
LUZ CLEMENCIA DE PÁEZ

Asunto: Radicación: 13-67186- -15
Trámite: 11
Evento: 378
Actuación: 519
Oficio: 2319
Folios: 9

Patente de Invención	X			Patente de Modelo de Utilidad	
Vía Nacional					
Vía PCT (Fase Nacional)	X	Capítulo I	X	Capítulo II	
No. Solicitud Internacional	PCT/EP2011/004383				
Fecha de Presentación Internacional	31/08/2011				

Como resultado del examen de fondo realizado, con fundamento en el Artículo 45 de la Decisión 486 de la Comisión de la Comunidad Andina, se comunica:

1. Documentos estudiados

Descripción:	Radicado N° 13-67186-00000-0000	3/Abril/2013
Reivindicaciones:	Radicado N° 13-67186-00000-0000	3/Abril/2013
	Radicado N° 13-67186-00006-0000	7/Julio/2014
	Radicado N° 13-67186-00009-0000	9/Octubre/2014
Dibujos:	Radicado N° 13-67186-00000-0000	3/Abril/2013
Secuencias:	Radicado N° 13-67186-00000-0000	3/Abril/2013
Prioridad:	EP10009180.0	3/Septiembre/2010

2. Análisis de la Prioridad

Se acepta la reivindicación de prioridad porque ésta solicitud fue presentada dentro del plazo establecido (Art. 9 D 486) y su contenido técnico corresponde con el de la prioridad reivindicada.

3. Modificaciones aceptadas (Art 34 D 486)

Se aceptan las modificaciones a las reivindicaciones en los folios 11 a 12 del Radicado N° 13-67186-00009-0000, ya que no amplían la materia inicialmente presentada, y sobre dicho capítulo se realizará el análisis de patentabilidad.

4. Objeto de la invención

De acuerdo con lo establecido en el numeral 1.2.2.2 del Título 10 de la Circular Única para efectos del examen de patentabilidad, se tendrá en cuenta la tasa pagada por el solicitante al momento de comenzar el examen para las reivindicaciones adicionales; esto es que cuando la solicitud contenga más de diez (10) reivindicaciones y no se haya pagado la tasa complementaria, sólo se estudiarán las cubiertas por la tasa pagada. En el presente caso se estudiarán las reivindicaciones 1 a 8 del capítulo reivindicatorio del Radicado N° 13-67186-00009-0000.

Título de la solicitud: "POLINUCLEÓTIDOS Y POLIPÉPTIDOS QUE CODIFICAN TRANSPORTADORES ABC QUE MODIFICAN SU ACTIVIDAD EN PLANTAS" (Modificado en el Radicado N° 13-67186-00006-0000)

El problema planteado en esta solicitud consiste en la necesidad de reducir el contenido de metales pesados, tal como el cadmio, de los vegetales o productos vegetales consumidos por animales y humanos.

Para resolver este problema técnico la solicitud en estudio proporciona un ARN bicatenario, que comprende al menos 19 nucleótidos contiguos de ARN NtMPR, el cual en contacto con una planta puede reducir los niveles de cadmio en al menos una parte de la planta.

El objeto de la presente solicitud se relaciona con la Clasificación Internacional de Patentes (CIP): C07K 14/41.

5. No es una invención (Art 15 D 486 literal (b))

El objeto de las reivindicaciones 2 - 4 no se considera una invención, por cuanto en su encabezado se reclama un polipéptido NtMRP aislado, expresado por un polinucleótido seleccionado del grupo constituido por las SEC ID Nos: 1, 2, 27, 28 o 29 o 51; 3 a 23 o 30 a 50 o 53, las cuales corresponden a secuencias de ADN tal y como se encuentra en la naturaleza o aquel que pueda ser aislado (ver alineamientos). Adicionalmente como se evidencia en el capítulo descriptivo (Ejemplos 1 y 5, Figs. 73 – 74, Fig. 1 y 6), dichas secuencias son aisladas del genoma de la planta de tabaco. Por lo anterior no pueden ser objeto de patente de conformidad con lo dispuesto en el artículo 15 lit. b) de la referida Decisión 486.

Query: SEQ ID NO: 3

Sbjct: *Nicotiana tomentosiformis* ABC transporter C family member 14-like (LOC104102849), transcript variant X2, mRNA ref|XM_009610689.1|

Range 1: 564 to 665 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
189 bits(102)	7e-45	102/102(100%)	0/102(0%)	Plus/Plus
Query 1	GTTAGAACCAATCTTTGGTTTAAGCTGTCTCTGATTTTGTCTAGCTATTTTAGCCTTATCT	60		
Sbjct 564	GTTAGAACCAATCTTTGGTTTAAGCTGTCTCTGATTTTGTCTAGCTATTTTAGCCTTATCT	623		
Query 61	TCTATAGTTTATGCATTTTGGTTATTGTGGGAAATTCCCAG	102		
Sbjct 624	TCTATAGTTTATGCATTTTGGTTATTGTGGGAAATTCCCAG	665		

Query: SEQ ID NO: 44
Sbjct: *Nicotiana sylvestris* ABC transporter C family member 3-like (LOC104231799), mRNA ref|XM_009784852.1|

Range 1: 823 to 987 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
305 bits(165)	1e-79	165/165(100%)	0/165(0%)	Plus/Plus
Query 1	GCATCGACAGATCAAAGTGCAATTGATCTGAATGTTCCCATTCAGTTGGATCCTTTGCC	60		
Sbjct 823	GCATCGACAGATCAAAGTGCAATTGATCTGAATGTTCCCATTCAGTTGGATCCTTTGCC	882		
Query 61	TTCACAATAATACAGCTTTTAGGGATTATTGGAGTAATGTCACAAGTTGCATGGCAGGTC	120		
Sbjct 883	TTCACAATAATACAGCTTTTAGGGATTATTGGAGTAATGTCACAAGTTGCATGGCAGGTC	942		
Query 121	TTCATTGTCTTTATTCCGGTCATTGCAGTTTGCATCTGGTTGGAG	165		
Sbjct 943	TTCATTGTCTTTATTCCGGTCATTGCAGTTTGCATCTGGTTGGAG	987		

Query: SEQ ID NO: 50
Sbjct: *Nicotiana sylvestris* ABC transporter C family member 3-like (LOC104231799), mRNA ref|XM_009784852.1|

Range 1: 2108 to 2232 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
231 bits(125)	2e-57	125/125(100%)	0/125(0%)	Plus/Plus
Query 1	GGCTCATTGCTGAATACGGCACTCCAGCCAGGTTGTTAGAGAACGAATCCTCATTGTTTG	60		
Sbjct 2108	GGCTCATTGCTGAATACGGCACTCCAGCCAGGTTGTTAGAGAACGAATCCTCATTGTTTG	2167		
Query 61	CTAAGCTCGTGGCAGAGTATAGTATGAGGTCAAATTCAAGTTTTGAGAATGTTTCAGACA	120		
Sbjct 2168	CTAAGCTCGTGGCAGAGTATAGTATGAGGTCAAATTCAAGTTTTGAGAATGTTTCAGACA	2227		
Query 121	TGTGA	125		
Sbjct 2228	TGTGA	2232		

6. De la solicitud de patente

Reivindicaciones (Art 30 D 486, Art 22 PCT)

1- La reivindicación 1 no es clara porque define un método para reducir los niveles de cadmio en al menos una parte de una planta, en función de las características técnicas del producto, y no se define a partir de sus características técnicas esenciales,

especificando una secuencia lógica de etapas y condiciones específicas que permitan a la persona normalmente versada en la materia reducir los niveles de expresión de un polinucleótido NtMRP. Adicionalmente la etapa b) del método se encuentra definida en términos de su resultado alcanzar que da cuenta de su mecanismo de acción, al indicar que se reduce la expresión de un polinucleótido NtMRP y la actividad de la proteína codificada por el mismo o la actividad de la proteína codificada por el mismo, en comparación con una planta de control no puesta en contacto con el dsRNA. En este caso las reivindicaciones incluirían no sólo la solución propuesta por el solicitante, sino todas las alternativas presentes o futuras que lleguen a ese resultado. Se sugiere a la solicitante definir el método únicamente a partir de una secuencia lógica de etapas y condiciones específicas que permitan la implementación del mismo.

- 2- Las reivindicaciones 5 y 6 no son claras porque no definen un ARN bicatenario en términos de sus características técnicas esenciales, especificando las secuencias de polinucleótidos con su respectivo SEC ID No., que caracterizan la hebra sentido y la hebra antisentido, en donde al menos una de las secuencias comprende 19 nucleótidos.
- 3- Las reivindicaciones 1, 5 - 6, no son claras porque el término “al menos”, es impreciso y no limitativo, por lo cual no permiten distinguir la invención del estado de la técnica. Se sugiere sustituirlo por un término o rango concreto.

7. Determinación del Estado de la Técnica

El estado de la técnica se determinó con base en la fecha de presentación de la solicitud de prioridad: 3 de septiembre de 2010, y se encontraron los siguientes documentos que anticipan el objeto de esta solicitud:

Ítem	Documento	Título	Fecha de Publicación
D5	US20090183280	“Transgenic Plants Modified for Reduced Cadmium Transport, Derivative Products, and Related Methods”	16/Julio/2009
D6	Bovet <i>et al.</i> , <i>Plant, Cell and Environment</i> , 2003. Vol. 26, Pág. 371–381.	“Transcript levels of AtMRPs after cadmium treatment: induction of AtMRP3”	14/Marzo/2003
D7	US20050091709	“Transgenic organism epressing fungal mrp-like abc transporters”	28/Abril/2005

El documento D5¹ divulga plantas de tabaco transgénicas y semillas derivadas modificadas genéticamente, para impedir el transporte de Cadmio (Cd) del sistema de la raíz a la porciones aéreas de dichas plantas, mediante la reducción de los niveles de expresión de transportadores relacionados con HMA (Abstract).

El documento D6² divulga un estudio que analiza el patrón de transcripción de los transportadores de casete que enlazan ATP (ABC), a través de microarreglos de ADNc, indicando que los niveles de transcripción de algunos de estos transportadores, que comprenden miembros de la subfamilia de proteínas MRP, se incrementan después del tratamiento con Cd²⁺ (Pág. 372, Col. 1).

El documento D7³ divulga secuencias de ADN aisladas que codifican proteínas MRP (proteínas asociadas a resistencia de fármacos), una subfamilia de los transportadores de

1 <http://worldwide.espacenet.com/publicationDetails/biblio?CC=US&NR=2009183280A1&KC=A1&FT=D>

2 <http://onlinelibrary.wiley.com/doi/10.1046/j.1365-3040.2003.00968.x/abstract>

3 <http://worldwide.espacenet.com/publicationDetails/biblio?CC=US&NR=2005091709A1&KC=A1&FT=D>

casete que enlazan ATP (ABC); permitiendo el desarrollo de organismos con resistencia a la acumulación de materiales tóxicos tales como metales pesados y herbicidas (Abstract).

8. Examen de Patentabilidad (Art 14 D486)

Nivel inventivo (Art18 D486)

La reivindicación 1 consiste en “*Un método para reducir niveles de cadmio en al menos una parte de una planta, que comprende los pasos de: (a) poner en contacto la planta con un ARN bicatenario que comprende al menos dos secuencias que son al menos parcialmente complementarias entre sí y en cuyo caso una hebra sentido comprende una primera secuencia y una hebra antisentido comprende una segunda secuencia y en cuyo caso al menos una de las secuencias comprende al menos 19 nucleótidos contiguos de ARN NtMRP, y (b) de este modo, reducir la expresión de un polinucleótido NtMRP y la actividad de la proteína codificada por el mismo o la actividad de la proteína codificada por el mismo en comparación con una planta de control no puesta en contacto con el dsRNA; en donde el ARN bicatenario comprende una primer cadena compuesta en una secuencia seleccionada del grupo que contiene: SEQ ID NOS: 3 a 23 y 30 a 50 y 53 y en donde una segunda cadena está compuesta en el complemento inverso de la secuencia correspondiente a SEQ ID NOS: 3 a 23 y 30 a 50 y 53*” (Radicado N° 13-67186-00009-0000).

Comparación de las características técnicas esenciales, con las del estado de la técnica:

Características esenciales	D5	D6
Método para reducir niveles de cadmio en al menos una parte de una planta	Método para reducir los niveles de cadmio en al menos una parte de una planta (Pág. 62, Rv. 13).	No registra
Poner en contacto la planta con un ARN bicatenario	ARNs de doble - hebra que inhiben la expresión del gen <i>NtHMA</i> en una célula de una planta (Ej. 13, Párr. [0127]).	No registra
Comprende al menos dos secuencias que son al menos parcialmente complementarias entre si	Comprende dos hebras de ARN que son suficientemente complementarias para hibridarse y formar una estructura doble (Ej. 13, Párr. [0127]).	No registra
Una hebra sentido comprende una primera secuencia	Hebra sentido que comprende una región complementaria a la hebra antisentido (Ej. 13, Párr. [0127]).	No registra
Una hebra antisentido comprende una segunda secuencia	Hebra antisentido que comprende una región complementaria a una secuencia diana, derivada del ARNm (Ej. 13, Párr. [0127]).	No registra
Una de las secuencias comprende al menos 19 nucleótidos contiguos de ARN NtMRP	ARNs de doble - hebra comprende entre 19 y 24 pares de base de longitud (Ej. 13, Párr. [0127]).	Regulación del gen <i>AtMRP3</i> (Pág. 375, Fig. 2b).
ARN bicatenario comprende una primer cadena compuesta en una secuencia seleccionada del grupo que contiene: SEQ ID NOS: 3 a 23 y 30 a 50 y 53 y en donde una segunda cadena está compuesta en el complemento inverso de la secuencia correspondiente a SEQ ID NOS: 3 a 23 y 30 a 50 y 53	Constructo de ARNi que comprende una primera secuencia que presenta al meno un 90% de identidad con una porción de la SEC ID No. 3 o 47, y una segunda secuencia (Pág. 62, Rv. 14).	No registra

El documento D5 se considera el estado de la técnica cercano a la invención definida en la reivindicación 1 porque divulga un método para reducir los niveles de cadmio (Cd) en al menos una parte de una planta, que comprende reducir los niveles de expresión del ARNm que codifica una proteína NtHMA en una planta, a través de la expresión del constructo que contiene el ARNi. Dicho constructo comprende una primera secuencia que presenta la menos un 90% de identidad con una porción de la SEC ID No. 3 o 47; una segunda secuencia; y una tercera secuencia que contiene una cadena complementaria inversa de la primera secuencia, posicionada en la misma orientación de la primera secuencia, y en donde la segunda secuencia está ligada de modo operable a la primera y a la tercera secuencia (Pág. 62, Rv. 13 - 14). Se desarrollan moléculas de ARNs de doble hebra que comprenden una hebra antisentido que consiste en una región complementaria con al menos una parte del ARNm formado en la expresión del gen *NtHMA*, y una hebra sentido la cual es complementaria con la hebra antisentido, de tal forma que las dos hebras se hibridan para formar una estructura doble, constituida típicamente entre 19 y 24 pares de bases de longitud (Ej. 13, Párr. [0127]). En algunas realizaciones se proporcionan plantas transgénicas de tabaco, en las cuales se reduce los niveles de expresión de miembros de la familia de genes *NtHMA*, limitando el transporte de cadmio a la lamina foliar (Pág. 10, Párr. [0086]).

La diferencia entre el método, definido en la reivindicación 1 y el método para reducir niveles de cadmio en una planta divulgado en el documento D5, consiste en que el método se encuentra dirigido a la regulación del gen *NtMRP* que codifica proteínas transportadoras del tipo ABC.

El efecto que logra la regulación del gen *NtMRP*, consiste en disminuir la expresión de un polinucleótido *NtMRP* y la actividad de las proteínas transportadoras ABC involucradas en el transporte de metales pesados, reduciendo los niveles de cadmio en al menos una parte de la planta.

Por lo tanto, el problema técnico objetivo que pretende resolver esta invención se puede formular así: Cómo modificar el método para reducir los niveles de cadmio en al menos una parte de una planta divulgado en D5, para regular la expresión del gen *NtMRP*.

Sin embargo el documento D6 ya sugería la regulación de los genes *AtMRPs*, luego de realizar un tratamiento con cadmio; en donde se evidenció que los niveles de transcripción de los genes que codifican para transportadores del tipo ABC debían ser modulados directamente por metales pesados y no a través de estrés oxidativo, estableciendo una correlación directa entre los metales pesados y la transcripción de los transportadores del tipo ABC, sugiriendo que tanto en plantas como levaduras estos transportadores están implicados en el flujo de metales pesados (Pág. 379, Col. 2). Específicamente se determinó que de todo el set de genes *AtMRPs*, el gen *AtMRP3* fue el que presentó la mayor inducción en los niveles de transcripción en las raíces de *A. thaliana* (Pág. 375, Fig. 2b). También se estableció algún tipo de inducción en los niveles de transcripción del gen *AtMRP4* (Pág. 374, Col. 1).

Por lo tanto, la persona del oficio normalmente versada en la materia, incluiría las sugerencias de regular la transcripción de genes *MRP* en plantas, específicamente los genes *AtMRP3* y *AtMRP4* (de los cuales ya se tiene evidencia de su participación en la regulación fisiológica del transporte de metales pesados) divulgadas en el documento D6, en el método para reducir los niveles de cadmio (Cd) en al menos una parte de una planta, que comprende el desarrollo de plantas transgénicas, por ejemplo de tabaco a partir de un

constructo de ARNi divulgado en el documento D5, para llegar así al objeto de la solicitud, por lo cual se considera obvia.

En consecuencia, la reivindicación 1 de acuerdo con lo divulgado en los documentos D5 y D6, carece de nivel inventivo (Art. 18).

La reivindicación 5 consiste en “*Un ARN bicatenario que comprende al menos dos secuencias que son al menos parcialmente complementarias entre si y en cuyo caso una hebra sentido comprende una primera secuencia y una hebra antisentido comprende una segunda secuencia y en cuyo caso al menos una de las secuencias comprende al menos 19 nucleótidos contiguos de NtMRP RNA*” (Radicado N° 13-67186-00009-0000).

Comparación de las características técnicas esenciales, con las del estado de la técnica:

Características esenciales	D5	D7
ARN bicatenario	ARNs de doble - hebra (Ej. 13, Párr. [0127]).	No registra
Comprende al menos dos secuencias que son al menos parcialmente complementarias entre si	Comprende dos hebras de ARN que son suficientemente complementarias para hibridarse y formar una estructura doble (Ej. 13, Párr. [0127]).	No registra
Una hebra sentido comprende una primera secuencia	Hebra sentido que comprende una región complementaria a la hebra antisentido (Ej. 13, Párr. [0127]).	No registra
Una hebra antisentido comprende una segunda secuencia	Hebra antisentido que comprende una región complementaria a una secuencia diana, derivada del ARNm (Ej. 13, Párr. [0127]).	No registra
Una de las secuencias comprende al menos 19 nucleótidos contiguos de ARN NtMRP	ARNs de doble - hebra comprende entre 19 y 24 pares de base de longitud (Ej. 13, Párr. [0127]).	Regulación de genes que codifican proteínas del tipo MRP (Pág. 21, Rv. 1).

El documento D5 se considera el estado de la técnica cercano a la invención definida en la reivindicación 1 porque divulga moléculas de ARNs de doble hebra que comprenden una hebra antisentido que consiste en una región complementaria con al menos una parte del ARNm formado en la expresión del gen *NtHMA*, y una hebra sentido la cual es complementaria con la hebra antisentido, de tal forma que las dos hebras se hibridan para formar una estructura doble. Dicha estructura presenta una estructura típica entre 19 y 24 pares de bases de longitud (Ej. 13, Párr. [0127]). Se divulga un constructo que contiene un ARNi que comprende una primera secuencia que presenta la menos un 90% de identidad con una porción de la SEC ID No. 3 o 47; una segunda secuencia; y una tercera secuencia que tiene una cadena complementaria inversa de la primera secuencia, posicionada en la misma orientación de la primera secuencia, y en donde la segunda secuencia esta ligada de modo operable a la primera y a la tercera secuencias (Pág. 62, Rv. 13 - 14). El constructo reduce la expresión del ARNm que codifica una proteína NtHMA disminuyendo los niveles de cadmio (Cd) en al menos una parte de una planta (Ej. 13, Párr. [0127] - Pág. 62, Rv. 13 - 14).

La diferencia entre el ARN bicatenario, definido en la reivindicación 5 y el ARN bicatenario divulgado en el documento D5, consiste en que el ARN bicatenario se encuentra dirigido a la regulación del gen *NtMRP* que codifica proteínas transportadoras del tipo ABC.

El efecto que logra la regulación del gen *NtMRP*, consiste en disminuir la expresión de un polinucleótido *NtMRP* y la actividad de la proteínas transportadoras ABC involucradas en el transporte de metales pesados, reduciendo los niveles de cadmio en al menos una parte de la planta.

Por lo tanto, el problema técnico objetivo que pretende resolver esta invención se puede formular así: Cómo modificar el ARN bicatenario divulgado en D5, para regular la expresión del gen *NtMRP*.

Sin embargo el documento D7 ya sugería la regulación de genes que codifican proteínas del tipo MRP (proteínas asociadas a la resistencia de múltiples fármacos), un transportador de casete que enlaza ATP (MRP – como proteína transportadora ABC) (Pág. 21, Rv. 1); en donde dicho gen corresponde al gen YCF1, el cual confiere tanto resistencia a materiales nocivos, como a la acumulación de los mismos. Los materiales nocivos son seleccionados de un grupo que consiste en cadmio, plomo, arsénico y herbicidas (Pág. 2, párr. [0030]). Adicionalmente se divulga el desarrollo de organismos transgénicos eucarióticos tales como plantas o levaduras, los cuales incluyen la transformación de una planta de tabaco (Pág. 3, párr. [0037 – 0038]).

Por lo tanto, la persona del oficio normalmente versada en la materia, incluiría las sugerencias de regular genes que expresan proteínas trasportadoras ABC involucradas en el transporte de metales pesados divulgadas en el documento D7, en el procedimiento para preparar un ARN bicatenario que regule la expresión del ARNm que codifica una proteína NtHMA disminuyendo los niveles de cadmio (Cd) en al menos una parte de una planta divulgado en el documento D5, para llegar así al objeto de la solicitud, por lo cual se considera obvia.

En consecuencia, la reivindicación 5 de acuerdo con lo divulgado en los documentos D5 y D7, carece de nivel inventivo (Art. 18).

Las reivindicaciones dependiente 6 a 8 no mencionan alguna característica que haga inventivo el ARN bicatenario de la reivindicación 5, por lo cual se considera que tampoco tienen nivel inventivo.

9. Conclusión

Los requisitos de Patentabilidad de la presente solicitud están afectados así:

Ítem	Documento	Reivindicaciones afectadas	Requisito que afecta
D5	US20090183280	1, 5 - 8	Nivel inventivo
D6	Bovet <i>et al.</i> , 2003. Plant, Cell and Environment. Vol. 26, Pág. 371–381.	1	Nivel inventivo
D7	US20050091709	5 - 8	Nivel inventivo

En cumplimiento del artículo 45 de la Decisión 486 de la Comisión de la Comunidad Andina, se le indica que debe dar respuesta dentro del término de sesenta (60) días contados a partir del día siguiente de la fecha de notificación de la presente comunicación. En caso de no dar respuesta al presente requerimiento, o si a pesar de la respuesta subsistieran los impedimentos para la concesión, se denegará la patente.

Si como respuesta a este requerimiento el solicitante llegara a modificar: i) el capítulo reivindicatorio, ii) la descripción de la invención a proteger, iii) o presente un nuevo capítulo reivindicatorio, no deberá pagar la tasa de modificación voluntaria pero sí la correspondiente a la tasa de examen de patentabilidad para que se estudie la materia modificada. En todo caso, el número de requerimientos por no patentabilidad estará supeditado a que el término del estudio de patentabilidad no sea superior a 18 meses, contados desde la publicación de la solicitud de patente de invención. Adicionalmente, se le recuerda al solicitante que si como respuesta a este requerimiento llegara a modificar el capítulo reivindicatorio y éste contenga más de diez (10) reivindicaciones deberá pagar la tasa por reivindicación adicional.

Se invita al solicitante se sirva indicar, si así lo considera, su renuncia al término faltante de sesenta (60) días para dar respuesta a este requerimiento y presentarla por escrito.

Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Única No. 10 de 2001.



JOSÉ LUIS SALAZAR LÓPEZ
Director de Nuevas Creaciones (c)

El anterior requerimiento fue notificado por fijación en lista, de fecha 2016-03-15

Elaboró: Mario Andres Forero

Revisó: Dayron Alexander Ramirez Reyes

INFORME DE BÚSQUEDA DE PATENTES PARA EL Radicado N° 13-67186-00009-0000						
CRITERIOS PARA LA BÚSQUEDA (anexar pantallazo o reporte en PDF, junto con los boleanos utilizados)						
Palabras clave utilizadas		MRP proteins, accumulation cadmio, transgenic tobacco plants , gen NtMRP, reduction plant, modulating heavy metal content of plants				
CLASIFICACIONES INTERNACIONALES UTILIZADAS PARA LA BÚSQUEDA		C07K 14/41				
BASES DE DATOS CONSULTADAS.						
Patentscope		Google/patents				
NCBI		PatBase				
Espacenet		ScienceDirect				
Lens						
INFORMES DE EXÁMENES REALIZADOS EN OTRAS OFICINAS (documentos patentes, literatura no patente etc...)						
Oficina	Fecha del examen o de reporte de búsqueda de documentos	Anterioridades encontradas para el examen o reportados en el examen de búsqueda(relacionarlos todos)	Reiv. afectadas	RELEVANCIA (requisito que afecta)	FUERON TENIDOS EN CUENTA PARA EL EXAMEN EN LA OFICINA COLOMBIANA	
					SI	NO
PCT INTERNATIONAL SEARCH REPORT EPO	14/Mayo/2012	Database EMBL (online), XP002665516, database accession no. FH224591	10, 11	Novedad/ Nivel inventivo		X
		Wojas <i>et al.</i> , 2009. Ectopic expression of Arabidopsis ABC transporter MRP7 modifies cadmium root-to-shoot transport and accumulation.	1 - 15	Nivel inventivo		X
		WO2009074325	5, 6, 12 - 14	Nivel inventivo		X
		Markus <i>et al.</i> , 2004. Disruption of AtMRP4, a guard cell plasma membrane ABCC-type ABC transporter, leads to deregulation of stomatal opening and increased drought susceptibility.	2, 3	Nivel inventivo		X
		Sánchez – Fernández <i>et al.</i> , 1998. Cloning and expression analyses of ATMRP4, a novel MRP-like gene from <i>Arabidopsis thaliana</i> .	1 - 15	Nivel inventivo		X
		WO03033705	1 – 4, 7 – 11, 15	Nivel inventivo		X
		Bovet <i>et al.</i> , 2003. Transcript levels of AtMRPs after cadmium treatment: induction of AtMRP3.	5, 6, 12 - 14	Nivel inventivo	X	
DOCUMENTOS ADICIONALES ENCONTRADOS POR EL EXAMINADOR DE ACUERDO A LOS CRITERIOS DE BÚSQUEDA.						
US6166290						
US2005091709						
US2011023177						
US6677502						
US5998203						

WO2006091194					
WO2008070274					
Mizuno <i>et al.</i> , 2010. Isolation of multidrug resistance associated protein like gene from lead hyperaccumulator common buckwheat and its lead detoxification ability					
Zientara <i>et al.</i> , 2009. Activity of the AtMRP3 promoter in transgenic Arabidopsis thaliana and Nicotiana tabacum plants is increased by cadmium, nickel, arsenic, cobalt and lead but not by zinc and iron.					
BÚSQUEDA DE SECUENCIAS					
Número de depósito / patente y N° de secuencia % de identidad					
SEQ ID NO (de la solicitud)	NCBI	The Lens	EMBL-EBI	DDJB	Genome Quest
SEQ ID NO: 2	JP 2013542719 dbj HW415857.1 Secuencia de polinucleótidos con un 100% de identidad				
SEQ ID NO: 27	Nicotiana tomentosiformis ABC transporter C family member 14-like (LOC104102849), transcript variant X2, mRNA ref XM_009610689.1 Secuencia de polinucleótidos con un 99% de identidad				
SEQ ID NO: 28	JP 2013542719 dbj HW415880.1 Secuencia de polinucleótidos con un 100% de identidad				
SEQ ID NO: 29	JP 2013542719 dbj HW415881.1 Secuencia de polinucleótidos con un 100% de identidad				
SEQ ID NO: 51	JP 2013542719 dbj HW415903.1 Secuencia de polinucleótidos con un 100% de identidad				
SEQ ID NO: 3	Nicotiana tomentosiformis ABC transporter C family member 14-like (LOC104102849), transcript variant X2, mRNA ref XM_009610689.1				
SEQ ID NO: 44	Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA ref XM_009784852.1				
SEQ ID NO: 50	Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA ref XM_009784852.1				

Búsqueda de Secuencias

SEC ID NO: 24

Sequences producing significant alignments:						
Select: All None Selected:0						
Alignments Download GenPept Graphics Distance tree of results Multiple alignment						
	Description	Max score	Total score	Query cover	E value	Ident
☐	PREDICTED: ABC transporter C family member 14-like [Nicotiana tomentosiformis]	4187	4496	91%	0.0	99%
☐	PREDICTED: ABC transporter C family member 14-like [Nicotiana sylvestris]	4062	4319	91%	0.0	97%
☐	PREDICTED: ABC transporter C family member 4 [Solanum tuberosum]	3776	3968	90%	0.0	91%
☐	PREDICTED: ABC transporter C family member 4 [Solanum pennellii]	3758	3941	90%	0.0	91%
☐	PREDICTED: ABC transporter C family member 4 [Solanum lycopersicum]	3741	3933	90%	0.0	91%
☐	PREDICTED: ABC transporter C family member 4 [Sesamum indicum]	2996	3117	90%	0.0	76%
☐	hypothetical protein CISIN_1g000438mg [Citrus sinensis]	2982	3126	90%	0.0	76%
☐	hypothetical protein CICLE_v10016482mg [Citrus clementina]	2982	3126	90%	0.0	76%
☐	ABC transporter C family member 4 [Gossypium arboreum]	2974	3103	90%	0.0	76%
☐	ABC transporter C family member 4 [Gossypium arboreum]	2974	3103	90%	0.0	76%
		Accession				
		XP_009608983.1				
		XP_009757900.1				
		XP_006359383.1				
		XP_015087637.1				
		XP_004247427.1				
		XP_011079676.1				
		KDQ87266.1				
		XP_006444306.1				
		KHG13944.1				
		KHG13943.1				

Sequences producing significant alignments:						
Select: All None Selected:0						
Alignments Download GenPept Graphics Distance tree of results Multiple alignment						
	Description	Max score	Total score	Query cover	E value	Ident
☐	unnamed protein product [Oryza sativa]	1991	1991	82%	0.0	60%
☐	unnamed protein product [Arabidopsis thaliana]	1615	2758	87%	0.0	70%
☐	unnamed protein product [Oryza sativa Japonica Group]	1552	1598	77%	0.0	57%
☐	unnamed protein product [Arabidopsis thaliana]	1543	2699	87%	0.0	69%
☐	Sequence 128654 from patent US 7214786	1251	1251	42%	0.0	69%
☐	unnamed protein product [Oryza sativa]	1016	1574	74%	0.0	55%
☐	unnamed protein product [Noccaea caerulescens]	953	953	70%	0.0	48%
☐	unnamed protein product [Oryza sativa Japonica Group]	927	1826	81%	0.0	53%
☐	Sequence 200 from patent US 6677502	761	761	29%	0.0	64%
☐	Sequence 158514 from patent US 9012723	723	792	33%	0.0	62%
		Accession				
		CAW45323.1				
		CAW45283.1				
		CAW45303.1				
		CAW45275.1				
		ABT41184.1				
		CAW45324.1				
		CAW45315.1				
		CAW45302.1				
		AAS29847.1				
		AKY71573.1				

SEC ID NO: 25

Sequences producing significant alignments:						
Select: All None Selected:0						
Alignments Download GenPept Graphics Distance tree of results Multiple alignment						
	Description	Max score	Total score	Query cover	E value	Ident
☐	PREDICTED: ABC transporter C family member 14-like [Nicotiana tomentosiformis]	342	342	7%	2e-90	100%
☐	PREDICTED: ABC transporter C family member 14-like [Nicotiana sylvestris]	330	330	7%	7e-87	98%
☐	PREDICTED: ABC transporter C family member 4 [Solanum tuberosum]	326	326	7%	2e-85	96%
☐	PREDICTED: ABC transporter C family member 4 [Solanum pennellii]	323	323	7%	1e-84	95%
☐	PREDICTED: ABC transporter C family member 4 [Solanum lycopersicum]	323	323	7%	1e-84	95%
☐	hypothetical protein JCGZ_13916 [Jatropha curcas]	320	320	7%	9e-84	94%
☐	PREDICTED: ABC transporter C family member 4 [Jatropha curcas]	320	320	7%	1e-83	94%
☐	PREDICTED: ABC transporter C family member 14-like [Musa acuminata subsp. malaccensis]	319	319	7%	1e-83	94%
☐	hypothetical protein SOVF_051440 [Spinacia oleracea]	319	319	7%	2e-83	94%
☐	multidrug resistance-associated protein 2, 6 (mrp2, 6), abc-transporter, putative [Ricinus communis]	317	317	7%	8e-83	94%
		Accession				
		XP_009608983.1				
		XP_009757900.1				
		XP_006359383.1				
		XP_015087637.1				
		XP_004247427.1				
		KDP28145.1				
		XP_012082740.1				
		XP_009381892.1				
		KNA20530.1				
		XP_002523063.1				

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	unnamed protein product [Arabidopsis thaliana]	310	310	7%	9e-83	90%	CAW45275.1
☐	Sequence 45470 from patent US 7834146	308	308	7%	3e-82	91%	ADT24065.1
☐	unnamed protein product [Arabidopsis thaliana]	301	301	7%	4e-80	90%	CAW45283.1
☐	unnamed protein product [Oryza sativa]	301	301	7%	7e-80	86%	CAW45323.1
☐	Sequence 200 from patent US 6677502	297	297	7%	4e-79	89%	AAS29847.1
☐	Sequence 159497 from patent US 7214786	286	286	7%	1e-76	85%	ABT72027.1
☐	Sequence 34273 from patent US 8299318	272	272	7%	2e-72	78%	AFX40518.1
☐	Sequence 175729 from patent US 7214786	264	264	7%	3e-70	81%	ABT88259.1
☐	unnamed protein product [Oryza sativa Japonica Group]	268	268	7%	4e-70	81%	CAW45302.1
☐	Sequence 17216 from patent US 8299318	258	258	7%	1e-68	80%	AFX32307.1

SEC ID NO: 26

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Sequence 27 from patent US 6846622	39.3	39.3	1%	0.44	63%	AAX03124.1
☐	Sequence 6386 from patent US 6562958	36.6	36.6	2%	3.8	53%	AAQ29414.1
☐	Sequence 163354 from patent US 9012723	35.7	35.7	1%	6.0	58%	AKY76413.1
☐	Sequence 3111 from patent US 7989676	35.7	35.7	2%	7.0	52%	AEL86512.1
☐	Sequence 1043 from patent US 6753314	35.7	35.7	1%	7.2	52%	AAV20783.1
☐	Sequence 495 from patent US 6900016	35.7	35.7	1%	7.2	52%	AAY75592.1

SEC ID NO: 52

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Sequence 150332 from patent US 9012723	3072	3072	97%	0.0	72%	AKY63391.1
☐	Sequence 99531 from patent US 9012723	3007	3007	95%	0.0	72%	AKY12224.1
☐	unnamed protein product [Arabidopsis thaliana]	2782	2782	97%	0.0	68%	CAW45305.1
☐	unnamed protein product [Arabidopsis thaliana]	2744	2744	97%	0.0	67%	CAW45284.1
☐	unnamed protein product [Arabidopsis thaliana]	2496	2496	97%	0.0	64%	CAW45277.1
☐	unnamed protein product [Arabidopsis thaliana]	2491	2491	97%	0.0	64%	CAW45326.1
☐	unnamed protein product [Arabidopsis thaliana]	2472	2472	96%	0.0	64%	CAW45278.1
☐	unnamed protein product [Arabidopsis thaliana]	2466	2466	96%	0.0	64%	CAW45325.1
☐	unnamed protein product [Arabidopsis thaliana]	2442	2442	95%	0.0	64%	CAW45330.1
☐	unnamed protein product [Triticum aestivum]	2284	2284	87%	0.0	65%	CAW45314.1

SEC ID NO: 2

Sequences producing significant alignments:

Select: All None Selected:0

Alignments						
Description	Max score	Total score	Query cover	E value	Ident	Accession
☐ PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 14-like (LOC104102849), transcript variant X2, mRNA	3882	8212	59%	0.0	99%	XM_009610689.1
☐ PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 14-like (LOC104102849), transcript variant X1, mRNA	3882	8212	59%	0.0	99%	XM_009610688.1
☐ PREDICTED: Nicotiana sylvestris ABC transporter C family member 14-like (LOC104210647), mRNA	3410	7442	59%	0.0	96%	XM_009759598.1
☐ Solanum pennellii chromosome ch09, complete genome	2689	6758	85%	0.0	87%	HG975448.1
☐ Solanum lycopersicum chromosome ch09, complete genome	2673	6715	85%	0.0	87%	HG975521.1
☐ PREDICTED: Solanum tuberosum ABC transporter C family member 4 (LOC102595270), transcript variant X2, mRNA	2627	5624	54%	0.0	90%	XM_015314403.1
☐ PREDICTED: Solanum tuberosum ABC transporter C family member 4 (LOC102595270), transcript variant X1, mRNA	2627	5624	54%	0.0	90%	XM_006359321.2
☐ PREDICTED: Solanum pennellii ABC transporter C family member 4 (LOC107030934), mRNA	2575	5523	54%	0.0	89%	XM_015232151.1
☐ PREDICTED: Solanum lycopersicum ABC transporter C family member 4 (LOC101249208), mRNA	2564	5484	54%	0.0	89%	XM_004247379.2
☐ Citrus clementina hypothetical protein (CICLE_v10018482mg) mRNA, complete cds	701	1055	24%	0.0	76%	XM_006444243.1
☐ PREDICTED: Citrus sinensis ABC transporter C family member 14 (LOC102623173), transcript variant X2, mRNA	699	1054	24%	0.0	76%	XM_006479876.2
☐ PREDICTED: Citrus sinensis ABC transporter C family member 14 (LOC102623173), transcript variant X1, mRNA	699	1054	24%	0.0	76%	XM_006479877.2
☐ Theobroma cacao Multidrug resistance-associated protein 4 isoform 3 (TCM_004638) mRNA, complete cds	676	956	22%	0.0	76%	XM_007050838.1
☐ Theobroma cacao Multidrug resistance-associated protein 4 isoform 3 (TCM_004638) mRNA, complete cds	676	956	22%	0.0	76%	XM_007050837.1

Sequences producing significant alignments:

Select: All None Selected:0

Alignments						
Description	Max score	Total score	Query cover	E value	Ident	Accession
☐ JP 2013542719-A/2: Heavy Metal Reduction in Planta	13899	13899	100%	0.0	100%	HW415857.1
☐ JP 2013542719-A/1: Heavy Metal Reduction in Planta	13899	13899	100%	0.0	100%	HW415856.1
☐ JP 2013542719-A/24: Heavy Metal Reduction in Planta	3893	8223	59%	0.0	100%	HW415879.1
☐ JP 2013542719-A/14: Heavy Metal Reduction in Planta	3186	3186	22%	0.0	100%	HW415869.1
☐ Sequence 7545 from patent US 7635798	1808	1808	14%	0.0	96%	GV611389.1
☐ JP 2013542719-A/23: Heavy Metal Reduction in Planta	1576	1576	11%	0.0	100%	HW415878.1
☐ JP 2013542719-A/4: Heavy Metal Reduction in Planta	1216	1216	8%	0.0	100%	HW415859.1
☐ JP 2013542719-A/17: Heavy Metal Reduction in Planta	1147	1147	8%	0.0	100%	HW415872.1
☐ JP 2013542719-A/8: Heavy Metal Reduction in Planta	1018	1018	7%	0.0	100%	HW415863.1
☐ JP 2013542719-A/9: Heavy Metal Reduction in Planta	944	944	6%	0.0	100%	HW415864.1
☐ JP 2013542719-A/5: Heavy Metal Reduction in Planta	800	800	5%	0.0	100%	HW415860.1
☐ JP 2013542719-A/49: Heavy Metal Reduction in Planta	713	713	5%	0.0	100%	HW415904.1
☐ JP 2013542719-A/15: Heavy Metal Reduction in Planta	586	586	4%	3e-163	99%	HW415870.1
☐ Sequence 7546 from patent US 7635798	577	577	5%	2e-160	92%	GV611390.1
☐ Sequence 31 from patent US 7635798	568	568	5%	9e-158	92%	GV603875.1
☐ JP 2013542719-A/21: Heavy Metal Reduction in Planta	566	566	4%	3e-157	100%	HW415876.1

SEC ID NO: 27

Sequences producing significant alignments:

Select: All None Selected:0

Alignments						
Description	Max score	Total score	Query cover	E value	Ident	Accession
☐ PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 14-like (LOC104102849), transcript variant X2, mRNA	8338	8338	100%	0.0	99%	XM_009610689.1
☐ PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 14-like (LOC104102849), transcript variant X1, mRNA	8338	8338	100%	0.0	99%	XM_009610688.1
☐ PREDICTED: Nicotiana sylvestris ABC transporter C family member 14-like (LOC104210647), mRNA	7550	7550	99%	0.0	97%	XM_009759598.1
☐ PREDICTED: Solanum tuberosum ABC transporter C family member 4 (LOC102595270), transcript variant X2, mRNA	6074	6074	99%	0.0	91%	XM_015314403.1
☐ PREDICTED: Solanum tuberosum ABC transporter C family member 4 (LOC102595270), transcript variant X1, mRNA	6074	6074	99%	0.0	91%	XM_006359321.2
☐ PREDICTED: Solanum pennellii ABC transporter C family member 4 (LOC107030934), mRNA	5956	5956	99%	0.0	91%	XM_015232151.1
☐ PREDICTED: Solanum lycopersicum ABC transporter C family member 4 (LOC101249208), mRNA	5917	5917	99%	0.0	90%	XM_004247379.2
☐ Solanum pennellii chromosome ch09, complete genome	2575	5703	93%	0.0	89%	HG975448.1
☐ Solanum lycopersicum chromosome ch09, complete genome	2564	5664	93%	0.0	89%	HG975521.1
☐ Theobroma cacao Multidrug resistance-associated protein 4 isoform 1 (TCM_004638) mRNA, complete cds	2213	2213	83%	0.0	77%	XM_007050835.1
☐ PREDICTED: Gossypium raimondii ABC transporter C family member 4-like (LOC105803022), transcript variant X2, mRNA	2191	2191	83%	0.0	77%	XM_012634956.1
☐ PREDICTED: Gossypium raimondii ABC transporter C family member 4-like (LOC105803022), transcript variant X1, mRNA	2191	2191	83%	0.0	77%	XM_012634955.1
☐ Citrus clementina hypothetical protein (CICLE_v10018482mg) mRNA, complete cds	2082	2082	85%	0.0	77%	XM_006444243.1
☐ PREDICTED: Citrus sinensis ABC transporter C family member 14 (LOC102623173), transcript variant X2, mRNA	2080	2080	85%	0.0	77%	XM_006479876.2
☐ PREDICTED: Citrus sinensis ABC transporter C family member 14 (LOC102623173), transcript variant X1, mRNA	2080	2080	85%	0.0	77%	XM_006479877.2

SEC ID NO: 28

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
☐ PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104233618), mRNA	4608	5979	53%	0.0	99%	XM_009787043.1
☐ Solanum lycopersicum chromosome ch07, complete genome	4069	6146	93%	0.0	87%	HG975519.1
☐ Solanum pennellii chromosome ch07, complete genome	4026	6206	93%	0.0	87%	HG975446.1
☐ PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 3-like (LOC104119292), mRNA	4024	8303	81%	0.0	96%	XM_009630753.1
☐ PREDICTED: Solanum lycopersicum ABC transporter C family member 3 (LOC101250322), mRNA	2857	5423	74%	0.0	88%	XM_010325975.1
☐ PREDICTED: Solanum tuberosum ABC transporter C family member 3 (LOC102590403), mRNA	2848	5783	78%	0.0	88%	XM_006346095.2
☐ PREDICTED: Solanum pennellii ABC transporter C family member 3 (LOC107026039), mRNA	2819	5419	69%	0.0	88%	XM_015226868.1
☐ PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA	1173	4802	42%	0.0	100%	XM_009784852.1
☐ PREDICTED: Nicotiana sylvestris uncharacterized LOC104234923 (LOC104234923), ncRNA	957	1075	11%	0.0	95%	XR_713090.1
☐ PREDICTED: Vitis vinifera ABC transporter C family member 3-like (LOC100252075), mRNA	913	1215	40%	0.0	75%	XM_002265569.3
☐ PREDICTED: Fragaria vesca subsp. vesca ABC transporter C family member 3-like (LOC101311724), mRNA	911	911	27%	0.0	76%	XM_004309769.2
☐ Vitis vinifera contig VV78X216589.14, whole genome shotgun sequence	898	1552	40%	0.0	75%	AM483903.1
☐ PREDICTED: Malus x domestica ABC transporter C family member 3-like (LOC103448329), mRNA	887	887	27%	0.0	76%	XM_008387573.1
☐ Theobroma cacao Multidrug resistance-associated protein 3 (TCM_030693) mRNA, complete cds	881	881	22%	0.0	78%	XM_007020505.1

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
☐ JP_2013542719-A/25: Heavy Metal Reduction in Planta	11431	11431	100%	0.0	100%	HW415880.1
☐ JP_2013542719-A/26: Heavy Metal Reduction in Planta	9941	9941	86%	0.0	100%	HW415881.1
☐ JP_2013542719-A/48: Heavy Metal Reduction in Planta	4102	8168	71%	0.0	100%	HW415903.1
☐ JP_2013542719-A/38: Heavy Metal Reduction in Planta	4096	4096	35%	0.0	100%	HW415893.1
☐ JP_2013542719-A/40: Heavy Metal Reduction in Planta	1170	1170	10%	0.0	100%	HW415895.1
☐ JP_2013542719-A/44: Heavy Metal Reduction in Planta	566	566	4%	3e-157	100%	HW415899.1
☐ JP_2013542719-A/42: Heavy Metal Reduction in Planta	545	545	4%	4e-151	100%	HW415897.1
☐ JP_2013542719-A/46: Heavy Metal Reduction in Planta	444	444	3%	1e-120	100%	HW415901.1
☐ JP_2013542719-A/43: Heavy Metal Reduction in Planta	398	398	3%	1e-106	100%	HW415898.1
☐ JP_2013542719-A/41: Heavy Metal Reduction in Planta	305	305	2%	7e-79	100%	HW415896.1
☐ JP_2013542719-A/30: Heavy Metal Reduction in Planta	268	268	2%	9e-68	100%	HW415885.1
☐ JP_2013542719-A/31: Heavy Metal Reduction in Planta	254	254	2%	2e-63	100%	HW415886.1

SEC ID NO: 29

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
☐ PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104233618), mRNA	4301	5672	57%	0.0	99%	XM_009787043.1
☐ Solanum lycopersicum chromosome ch07, complete genome	4036	5841	97%	0.0	87%	HG975519.1
☐ Solanum pennellii chromosome ch07, complete genome	4002	5869	97%	0.0	87%	HG975446.1
☐ PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 3-like (LOC104119292), mRNA	3921	7637	84%	0.0	96%	XM_009630753.1
☐ PREDICTED: Solanum tuberosum ABC transporter C family member 3 (LOC102590403), mRNA	2841	5335	78%	0.0	88%	XM_006346095.2
☐ PREDICTED: Solanum lycopersicum ABC transporter C family member 3 (LOC101250322), mRNA	2824	4986	74%	0.0	88%	XM_010325975.1
☐ PREDICTED: Solanum pennellii ABC transporter C family member 3 (LOC107026039), mRNA	2819	5419	80%	0.0	88%	XM_015226868.1
☐ PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA	1173	4183	42%	0.0	100%	XM_009784852.1
☐ PREDICTED: Vitis vinifera ABC transporter C family member 3-like (LOC100252075), mRNA	913	1215	46%	0.0	75%	XM_002265569.3
☐ PREDICTED: Fragaria vesca subsp. vesca ABC transporter C family member 3-like (LOC101311724), mRNA	911	911	31%	0.0	76%	XM_004309769.2
☐ Vitis vinifera contig VV78X216589.14, whole genome shotgun sequence	898	1552	47%	0.0	75%	AM483903.1

Sequences producing significant alignments:						
Select: All None Selected:0						
Alignments Download GenBank Graphics Distance tree of results						
	Description	Max score	Total score	Query cover	E value	Ident
☐	JP 2013542719-A/26: Heavy Metal Reduction in Planta	9941	9941	100%	0.0	100%
☐	JP 2013542719-A/25: Heavy Metal Reduction in Planta	9941	9941	100%	0.0	100%
☐	JP 2013542719-A/48: Heavy Metal Reduction in Planta	4102	8168	82%	0.0	100%
☐	JP 2013542719-A/38: Heavy Metal Reduction in Planta	4096	4096	41%	0.0	100%
☐	JP 2013542719-A/40: Heavy Metal Reduction in Planta	1170	1170	11%	0.0	100%
☐	JP 2013542719-A/44: Heavy Metal Reduction in Planta	566	566	5%	2e-157	100%
☐	JP 2013542719-A/42: Heavy Metal Reduction in Planta	545	545	5%	3e-151	100%
☐	JP 2013542719-A/46: Heavy Metal Reduction in Planta	444	444	4%	1e-120	100%
☐	JP 2013542719-A/43: Heavy Metal Reduction in Planta	398	398	3%	9e-107	100%
						Accession
						HW415881.1
						HW415880.1
						HW415903.1
						HW415893.1
						HW415895.1
						HW415899.1
						HW415897.1
						HW415901.1
						HW415898.1

SEC ID NO: 51

Sequences producing significant alignments:						
Select: All None Selected:0						
Alignments Download GenBank Graphics Distance tree of results						
	Description	Max score	Total score	Query cover	E value	Ident
☐	PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 3-like (LOC104119292), mRNA	7395	7466	99%	0.0	97%
☐	PREDICTED: Solanum tuberosum ABC transporter C family member 3 (LOC102590403), mRNA	5570	5570	98%	0.0	90%
☐	PREDICTED: Solanum pennellii ABC transporter C family member 3 (LOC107026039), mRNA	5537	5537	98%	0.0	90%
☐	PREDICTED: Solanum lycopersicum ABC transporter C family member 3 (LOC101250322), mRNA	5515	5515	98%	0.0	90%
☐	PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104233618), mRNA	5349	5427	67%	0.0	99%
☐	PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA	3934	4111	50%	0.0	99%
☐	Solanum pennellii chromosome ch07, complete genome	2761	5366	94%	0.0	89%
☐	Solanum lycopersicum chromosome ch07, complete genome	2756	4923	87%	0.0	89%
☐	PREDICTED: Sesamum indicum ABC transporter C family member 3-like (LOC105174952), mRNA	2061	2061	95%	0.0	76%
☐	PREDICTED: Vitis vinifera ABC transporter C family member 3-like (LOC100254999), mRNA	2041	2041	96%	0.0	76%
☐	PREDICTED: Vitis vinifera ABC transporter C family member 3-like (LOC100265300), mRNA	1977	1977	75%	0.0	78%
						Accession
						XM_009630753.1
						XM_006346095.2
						XM_015226868.1
						XM_010325975.1
						XM_009787043.1
						XM_009784852.1
						HG975446.1
						HG975519.1
						XM_011097219.1
						XM_003634707.2
						XM_003634705.2

Sequences producing significant alignments:						
Select: All None Selected:0						
Alignments Download GenBank Graphics Distance tree of results						
	Description	Max score	Total score	Query cover	E value	Ident
☐	JP 2013542719-A/48: Heavy Metal Reduction in Planta	8106	8106	100%	0.0	100%
☐	JP 2013542719-A/26: Heavy Metal Reduction in Planta	4102	8168	100%	0.0	100%
☐	JP 2013542719-A/25: Heavy Metal Reduction in Planta	4102	8168	100%	0.0	100%
☐	JP 2013542719-A/38: Heavy Metal Reduction in Planta	4096	4096	50%	0.0	100%
☐	JP 2013542719-A/40: Heavy Metal Reduction in Planta	1170	1170	14%	0.0	100%
☐	JP 2013542719-A/44: Heavy Metal Reduction in Planta	566	566	6%	2e-157	100%
☐	JP 2013542719-A/42: Heavy Metal Reduction in Planta	545	545	6%	3e-151	100%
☐	JP 2013542719-A/46: Heavy Metal Reduction in Planta	444	444	5%	9e-121	100%
☐	JP 2013542719-A/43: Heavy Metal Reduction in Planta	398	398	4%	7e-107	100%
☐	JP 2013542719-A/41: Heavy Metal Reduction in Planta	305	305	3%	5e-79	100%
						Accession
						HW415903.1
						HW415881.1
						HW415880.1
						HW415893.1
						HW415895.1
						HW415899.1
						HW415897.1
						HW415901.1
						HW415898.1
						HW415896.1

SEC ID NO: 3

Download GenBank Graphics					Next Previous Descriptions
PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 14-like (LOC104102849), transcript variant X2, mRNA					
Sequence ID: ref XM_009610689.1 Length: 5024 Number of Matches: 1					
Range 1: 564 to 665 GenBank Graphics Next Match Previous Match					
Score	Expect	Identities	Gaps	Strand	
189 bits(102)	7e-45	102/102(100%)	0/102(0%)	Plus/Plus	
Query	1	GTTAGAACCAATCTTTGGTTAAGCTGTCCTGATTTTGTGAGCTATTTAGCCTTAICT	60		
Sbjct	564	GTTAGAACCAATCTTTGGTTAAGCTGTCCTGATTTTGTGAGCTATTTAGCCTTAICT	623		
Query	61	TCTATAGTTTTATGCATTTTGGTTATTGTGGGAAATCCCG	102		
Sbjct	624	TCTATAGTTTTATGCATTTTGGTTATTGTGGGAAATCCCG	665		
Related Information					
Gene - associated gene details					

SEC ID NO: 44

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

AlignmentsDownloadGenBankGraphicsDistance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA	305	305	100%	1e-79	100%	XM_009784852.1
☐	PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104235949), partial mRNA	289	289	100%	1e-74	98%	XM_009789786.1
☐	PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 3-like (LOC104119292), mRNA	278	278	100%	3e-71	97%	XM_009630753.1
☐	PREDICTED: Solanum pennellii ABC transporter C family member 3 (LOC107026039), mRNA	239	239	100%	1e-59	93%	XM_015226868.1
☐	Solanum pennellii chromosome ch07, complete genome	239	239	100%	1e-59	93%	HG975446.1
☐	PREDICTED: Solanum tuberosum ABC transporter C family member 3 (LOC102590403), mRNA	233	233	100%	6e-58	92%	XM_006346095.2
☐	PREDICTED: Solanum lycopersicum ABC transporter C family member 3 (LOC101250322), mRNA	228	228	100%	3e-56	92%	XM_010325975.1
☐	Solanum lycopersicum chromosome ch07, complete genome	228	228	100%	3e-56	92%	HG975519.1

DownloadGenBankGraphics

NextPreviousDescriptions

PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA

Sequence ID: [ref|XM_009784852.1|](#) Length: 2567 Number of Matches: 1

Range 1: 823 to 987 GenBankGraphics

Next MatchPrevious Match

Score	Expect	Identities	Gaps	Strand
305 bits(165)	1e-79	165/165(100%)	0/165(0%)	Plus/Plus
Query 1	GCATCGACAGATCAAAGTGCAATTGATCTGAATGTTCCCATTCAGTTGGATCCTTTGGCC	60		
Sbjct 823	GCATCGACAGATCAAAGTGCAATTGATCTGAATGTTCCCATTCAGTTGGATCCTTTGGCC	882		
Query 61	TTCAACAATAACAGCTTTTAGGGATTATTGGAGTAATGTCACAAGTTGCATGGCAGGTC	120		
Sbjct 883	TTCAACAATAACAGCTTTTAGGGATTATTGGAGTAATGTCACAAGTTGCATGGCAGGTC	942		
Query 121	TTCAATTGCTTTATTCCGGTCATTGCAGTTTGCATCTGGTTGGAG	165		
Sbjct 943	TTCAATTGCTTTATTCCGGTCATTGCAGTTTGCATCTGGTTGGAG	987		

Related Information

Gene - associated gene details

SEC ID NO: 50

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

AlignmentsDownloadGenBankGraphicsDistance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA	231	231	100%	2e-57	100%	XM_009784852.1
☐	PREDICTED: Nicotiana sylvestris uncharacterized LOC104234923 (LOC104234923), ncRNA	226	226	100%	7e-56	99%	XR_713090.1
☐	PREDICTED: Nicotiana tomentosiformis ABC transporter C family member 3-like (LOC104119292), mRNA	215	215	100%	2e-52	98%	XM_009630753.1

DownloadGenBankGraphics

NextPreviousDescriptions

PREDICTED: Nicotiana sylvestris ABC transporter C family member 3-like (LOC104231799), mRNA

Sequence ID: [ref|XM_009784852.1|](#) Length: 2567 Number of Matches: 1

Range 1: 2108 to 2232 GenBankGraphics

Next MatchPrevious Match

Score	Expect	Identities	Gaps	Strand
231 bits(125)	2e-57	125/125(100%)	0/125(0%)	Plus/Plus
Query 1	GGCTCATTTGCTGAATACGGCACTCCAGCCAGGTTGTTAGAGAACGAATCCTCATTGTTTG	60		
Sbjct 2108	GGCTCATTTGCTGAATACGGCACTCCAGCCAGGTTGTTAGAGAACGAATCCTCATTGTTTG	2167		
Query 61	CTAAGCTCGTGGCAGAGTATAGTATGAGGTCAAATTCAGTTTGGAGATGTTTCAGACA	120		
Sbjct 2168	CTAAGCTCGTGGCAGAGTATAGTATGAGGTCAAATTCAGTTTGGAGATGTTTCAGACA	2227		
Query 121	TGTGA	125		
Sbjct 2228	TGTGA	2232		

Related Information

Gene - associated gene details

Búsqueda en PatBase

PatBase

Username: SIC3@patbase.com
Subaccount: Mario Forero

Menu - Search - History - Session - Folder - Order Help - Logoff

Search 7: TAC=(((NICOTIANA~ TABACUM~) AND (EXPRESSION~ AND MRP4~)))
Results 0

Explore search Edit and resend search

Information Shortcuts

Command Reference

Bibliographic data
Textual data
Update commands
Operators
Navigation / shortcuts

Search tools

#	Search query	Results	Options
7	TAC=(((NICOTIANA~ TABACUM~) AND (EXPRESSION~ AND MRP4~)))	0	
6	TAC=((MRP~ PROTEINS~) AND TOBACCO~)	1	View Browse Hits Optimise more...
5	TAC=(((ACCUMULATION~ AND CADMIO~) AND (PLANTS~ AND TOBACCO~)))	1	View Browse Hits Optimise more...
4	TAC=(((GENE~ AND NTMRP~) AND (REDUCTION~ AND PLANT~)))	1	View Browse Hits Optimise more...
3	TAC=(((HEAVY~ METAL~) AND (NTMRP~ AND PLANTS~)))	1	View Browse Hits Optimise more...
2			
1	TAC=(((HEAVY~ AND METAL~) AND (GENE~ AND MRP~)))	10	View Browse Hits Optimise more...

PatBase

Username: SIC3@patbase.com
Subaccount: Mario Forero

Menu - Search - History - Session - Folder - Order Help - Logoff

Search 6: "TAC=((MRP~ PROTEINS~) AND TOBACCO~)" 1 of 1

Display format: Classic Sorted by: Publication date asc Print Save/export Snapshot Visual explorer Analytics

1) Family number: 28515604 (US2005091709A) Similar • Translate • Full-text • Status • Citations

Title: [EN] TRANSGENIC ORGANISM EPRESSING FUNGAL MRP-LIKE ABC TRANSPORTERS

Abstract: Source: US2005091709A [EN] The present invention relates to an isolated DNA encoding fungal MRP (multidrug resistance associated protein) subfamily of ATP-binding cassette (ABC) transporter protein which confers organisms with resistance to and/or accumulation of toxic materials such as heavy metals and herbicides, vectors containing the isolated DNA, and organisms transformed with the isolated DNA. Organisms transformed with the fungal MRP subfamily of ABC transporters of the present invention can be used for remediation of environment polluted with toxic materials. For example, the transformed plants can be used for cleaning polluted soil or water, and thereby providing an environmentally-friendly way to remediate polluted resources with low cost.

Classifications: Classification Explorer

International (IPC 8-9): A01H5/00 A01K67/027 C07K14/395 C07K14/705 C12N1/19 C12N15/09 C12N15/29 C12N15/82 C12N5/10 (Advanced/Invention); C12N15/29 (Advanced/Non-invention); A01H5/00 A01K67/027 C07K14/395 C12N1/19 C12N15/09 C12N15/29 C12N15/82 C12N5/10 (Core/Invention)

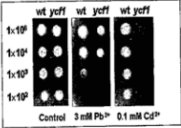
International (IPC 1-7): A01H1/00 A01H5/00 A01K67/027 C07H21/04 C07K14/415 C12N1/19 C12N15/09 C12N15/29 C12N15/82 C12N5/04 C12N5/10

CPC: C07K14/395 A01K2217/05 C07K14/705 C12N15/8259 C12N15/8271 C12N15/8274

European: C07K14/395 C07K14/705 C12N15/82C4E C12N15/82C8B C12N15/82C8B4 K01K217/05 M07K207/00

US: 435/419 435/419S 435/468 435/468S 435/69.1 435/69.100 435/69.100S 530/370 530/370S 536/23.6 536/23.600 536/23.600S 800/278 800/278P 800/295 800/298 800/306

FIG.1



PatBase

Username: SIC3@patbase.com
Subaccount: Mario Forero

Menu - Search - History - Session - Folder - Order Help - Logoff

Search 5: "TAC=(((ACCUMULATION~ AND CADMIO~) AND (PLANTS~ AND TOBACCO~)))" 1 of 1

Display format: Classic Sorted by: Publication date asc Print Save/export Snapshot Visual explorer Analytics

1) Family number: 43386941 (US2011023177A) Similar • Translate • Full-text • Status • Citations

Title: [EN] GENETICALLY MODIFIED PLANTS DISPLAYING REDUCED ACCUMULATION OF CADMIUM

Abstract: Source: US2011023177A [EN] The present invention relates to methods and means such as polynucleotides, recombinant expression cassettes and recombinant vectors, for reducing the accumulation of cadmium in a plant, and a transgenic plant expressing a variant of a plant P1B-type ATPase of the Zn2+/Co2+/Cd2+/Pb2+ subgroup, said transgenic plant displaying a reduced accumulation of cadmium in the aerial parts.

Classifications: Classification Explorer

International (IPC 8-9): A01H5/00 A01H5/12 C07H21/00 C12N15/82 C12N5/10 C12N9/14 (Advanced/Invention); A01H5/00 C12N15/82 C12N9/14 (Core/Invention)

International (IPC 1-7): C12N15/82

CPC: C12N15/8261 C12N9/14 C12N15/8243

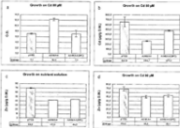
European: C12N15/82C4B C12N15/82C8 C12N9/14

US: 435/195 435/195S 435/320.1 435/320.100S 435/419 435/419S 536/23.2 536/23.200S 800/278 800/278P 800/298 800/298S 800/317.3 800/317.300S

Family: Family Explorer

Publication number	Publication date	Application number	Application date	Links
BRPI0722297 A2	20140603	BR2007PI22297	20071212	 

FIG. 1



Search 3: "TAC=(((HEAVY~ METAL~) AND (NTMRP~ AND PLANTS~)))" 1 of 1

Display format: Classic Sorted by: Publication date asc Print Save/export Snapshot Visual explorer Analytics

1) Family number: 50755649 (US2015232867A)

Similar • Translate • Full-text • Status • Citations

Title: [EN] **HEAVY METAL** REDUCTION IN PLANTA

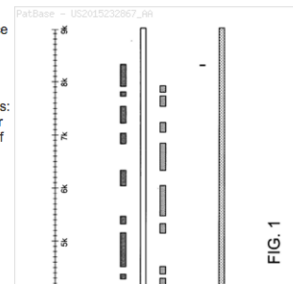
Abstract: Source: US2015232867A [EN] There is described a mutant, non-naturally occurring or transgenic **plant** or **plant** cell comprising (a) a polynucleotide selected from the group consisting of: (i) a polynucleotide comprising, consisting or consisting essentially of a sequence having at least 71 percent sequence identity to SEQ ID Nos: 1, 2, 27, 28 or 29 or 51; or (ii) a polynucleotide comprising, consisting or consisting essentially of a sequence having at least 65 percent sequence identity to any of SEQ ID Nos: 3 to 23 or 30 to 50; or (iii) a polynucleotide encoding a **NIMPR** polypeptide comprising, consisting or consisting essentially of a sequence having at least 65 percent sequence identity to any of SEQ ID Nos: 24 to 26 or 52, and wherein the polypeptide has **heavy metal** transporter activity; or (b) a polynucleotide comprising, consisting or consisting essentially of a sequence having at least 65 percent sequence identity to a region of any of SEQ ID Nos: 1 to 23 or 27 to 51; or (c) a double-stranded RNA comprising at least two sequences that are at least partially complementary to each other and wherein a sense strand comprises a first sequence and an antisense strand comprises a second sequence and wherein at least one of the sequences comprises at least 10 contiguous nucleotides of **NIMPR** RNA; or (d) an expression vector comprising the polynucleotide as set forth in (i), (ii) or (iii) or the polynucleotide construct as set forth in (b).

Classifications: [Classification Explorer](#)

International (IPC 8-9): A01G7/06 A01H5/00 A01H5/04 A01H5/10 A01H5/12 C07K14/41 C07K14/415 C07K16/16 C12N15/09 C12N15/113 C12N15/115 C12N15/29 C12N15/82 C12N5/04 C12N5/10 C12Q1/68 C7K14/41 (Advanced/Invention)

CPC: C07K14/415 C12N15/8242 C12N15/8271 C12N15/8218 C12Q1/6895 C07K16/16 C12N15/1138 C12Q2600/158

European: C07K14/415 C12N15/82C4 C12N15/82C8B





US 20090183280A1

(19) **United States**(12) **Patent Application Publication****Hayes et al.**(10) **Pub. No.: US 2009/0183280 A1**(43) **Pub. Date: Jul. 16, 2009**(54) **TRANSGENIC PLANTS MODIFIED FOR REDUCED CADMIUM TRANSPORT, DERIVATIVE PRODUCTS, AND RELATED METHODS**(76) Inventors: **Alec J. Hayes**, Chesterfield, VA (US); **Chengalrayan Kudithipudi**, Midlothian, VA (US); **Rutger S. van der Hoeven**, Midlothian, VA (US)*C12N 15/74* (2006.01)*C12N 5/04* (2006.01)*C07H 21/04* (2006.01)*A24F 47/00* (2006.01)*C07K 2/00* (2006.01)*C07K 16/00* (2006.01)(52) **U.S. CL. 800/285; 800/298; 536/23.1; 435/320.1; 435/425; 536/24.5; 131/347; 530/300; 530/387.9**

Correspondence Address:

BUCHANAN, INGERSOLL & ROONEY PC
POST OFFICE BOX 1404
ALEXANDRIA, VA 22313-1404 (US)(21) Appl. No.: **12/333,681**(22) Filed: **Dec. 12, 2008****Related U.S. Application Data**

(60) Provisional application No. 60/996,982, filed on Dec. 13, 2007.

Publication Classification(51) **Int. Cl.***C12N 15/82* (2006.01)*A01H 5/00* (2006.01)*C07H 21/02* (2006.01)(57) **ABSTRACT**

Various embodiments are directed to transgenic plants, including transgenic tobacco plants and derivative seeds, genetically modified to impede the transport of Cadmium (Cd) from the root system to aerial portions of transgenic plants by reducing the expression levels of HMA-related transporters. Various embodiments are directed to transgenic tobacco plants genetically modified to stably express a RNAi construct encoding RNAi polynucleotides that enable the degradation of endogenous NtHMA RNA variants. Reduced expression of NtHMA transporters in transgenic plants results in substantially reduced content of Cadmium (Cd) in the leaf lamina. Various consumable products that are substantially free or substantially reduced in Cd content can be produced by incorporating leaves derived from transgenic tobacco plants modified to reduce the expression of NtHMA transporters.

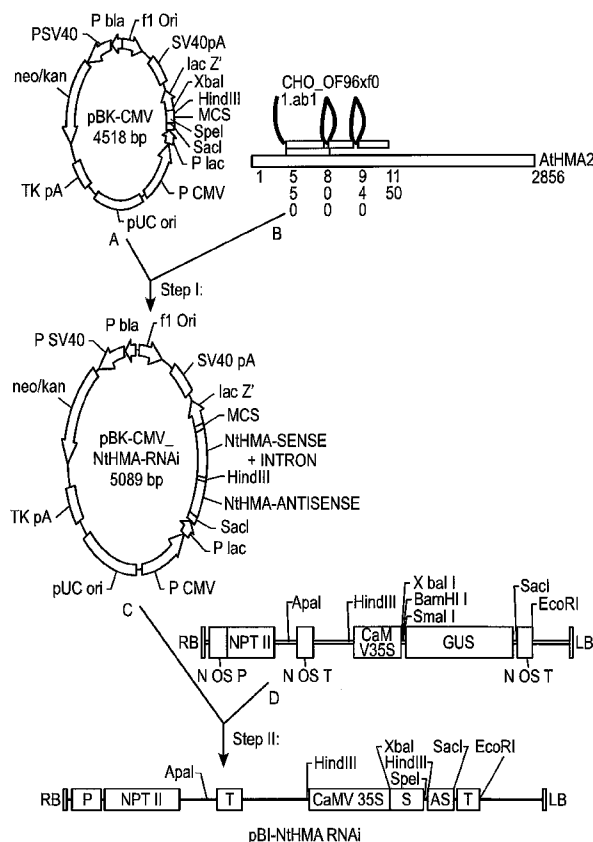




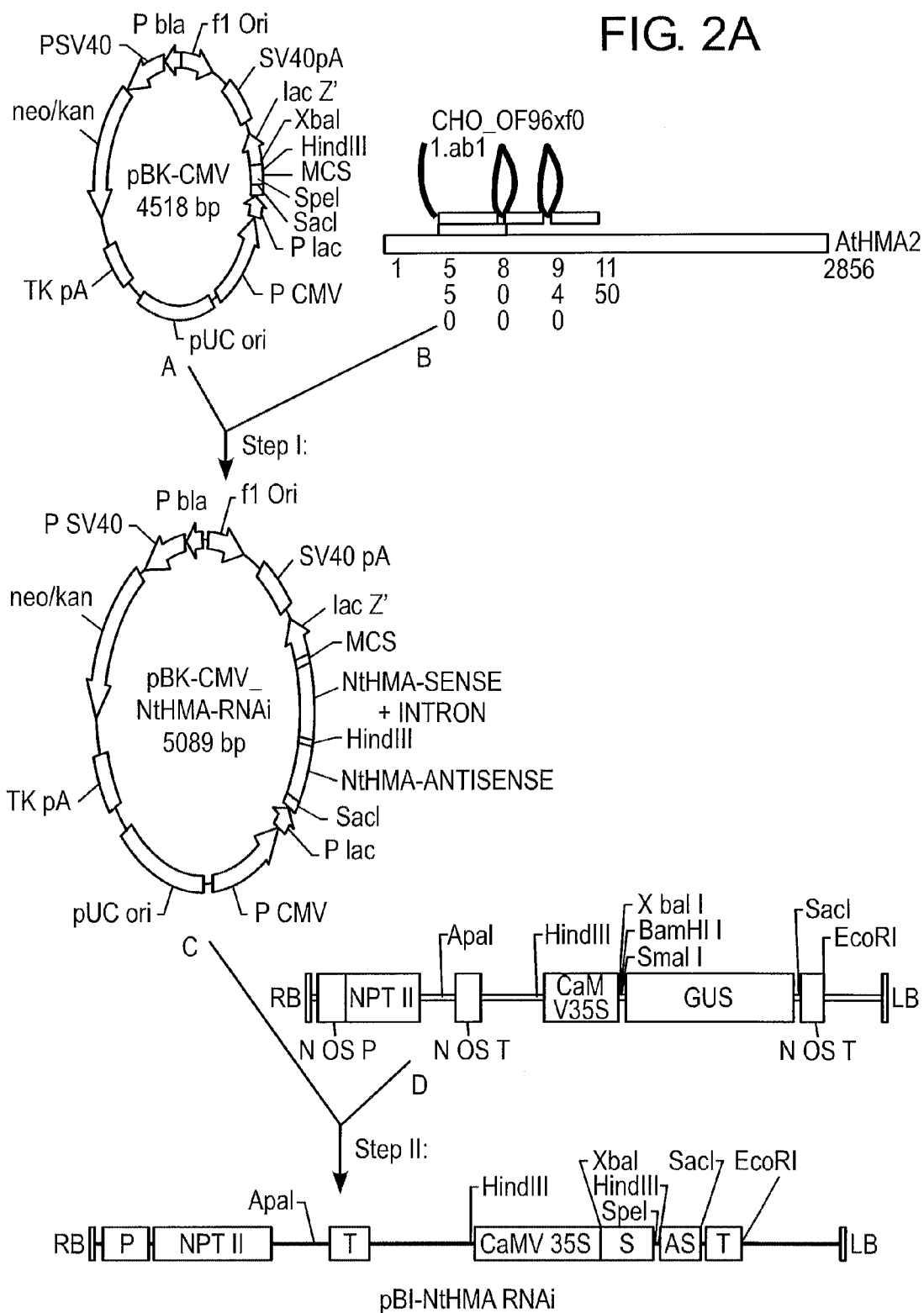
FIG. 1A

Table 1

Exon	Nucleotides	Position
Exon 1	1-303	724-1026
Exon 2	304-561	3245-3502
Exon 3	562-659	7364-7461
Exon 4	660-915	11525-11780
Exon 5	916-1056	11866-12007
Exon 6	1057-1381	12317-12644
Exon 7	1382-1584	13108-13310
Exon 8	1585-1787	13456-13658
Exon 9	1788-3285	14278-15775
Exon 10	3286-3618	16097-16429
Exon 11	3619-4392	16650-17423

FIG. 1B

FIG. 2A



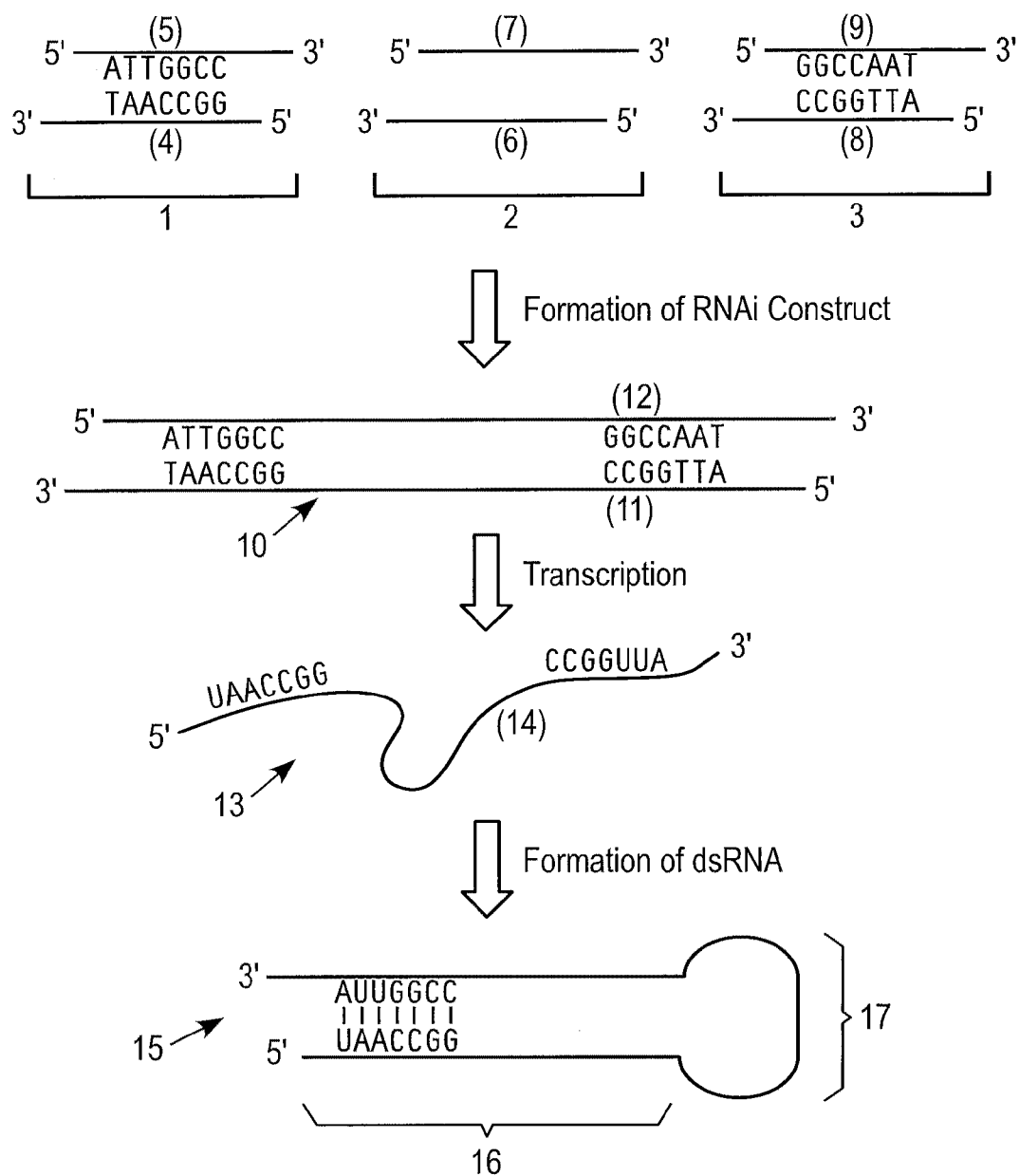


FIG. 2B

FIG. 3A

RNAi Construct NtHMA (660-915) Encoding RNAi Polynucleotide
SEQ ID NO:41

Sense Sequence SEQ ID NO:38

5'-ATTTGTAGTGCCAGCCCAGACCGTTGAATCTATTTGCTTAGAAAC
TGGAACGACTCGCCTGTCAGTGTTCCTCGTCCACGTCACATTCCC
CTTCATTACAACCTCCATCAATAGGTATAGTTTCACCAGCTTTAACAG
CAAGAATGCTATTCAACTTGACTTCATCAACATTTACGACTTCTCCAC
TTTCAGCTAAACTGCTGTTGGAGGGACTATATTGACCAGTGATGAC
ATAGCAGCAGTAGCCTACATAACCA-3'

Spacer Sequence SEQ ID NO:39

5'-AGCCTGAAGAATTGAGCAAATAACATTAACAAACAATACTTGAAG
TTTCAGCACTAAATAAATGAAGCATGAAGGAATACTACACTACCATT
AGA-3'

Reverse Complementary Sequence SEQ ID NO:40

5'-TGGTTATGTAGGCTACTGCTGCTATGTCATCACTGGTCAATATAG
TCCCTCCAACAGCAGTTTTAGCTGAAAGTGGAGAAGTCGTAAATGTT
GATGAAGTCAAGTTGAATAGCATTCTTGCTGTAAAGCTGGTGAAC
TATACCTATTGATGGAGTTGTAATGGAAGGGGAATGTGACGTGGAC
GAGAAAACACTGACAGGCGAGTCGTTTCCAGTTTCTAAGCAAATAGA
TTCAACGGTCTGGGCTGGCACTACAAAT-3'

FIG. 3B

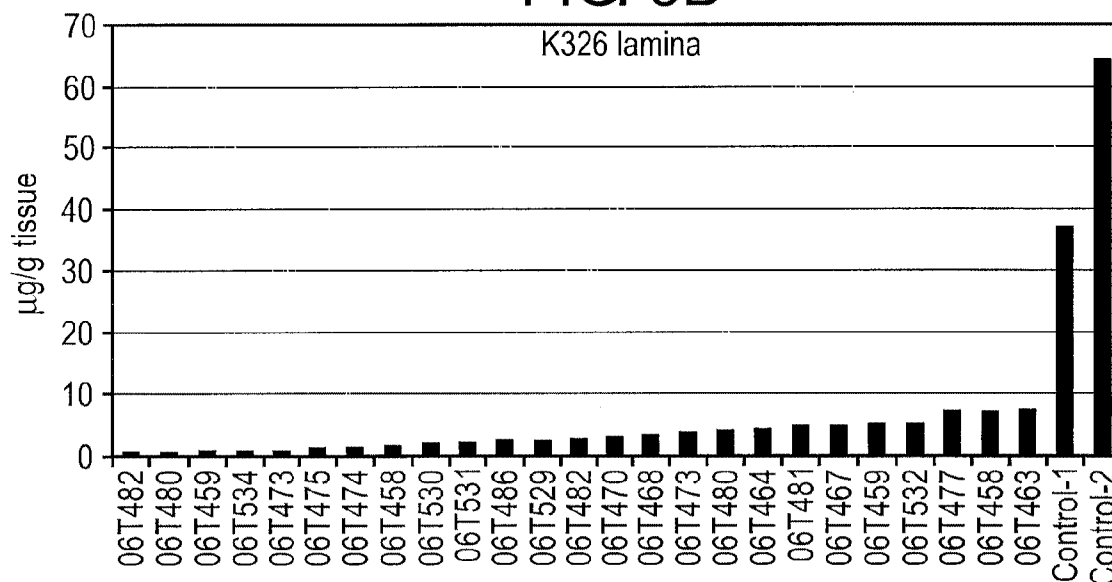


FIG. 3C

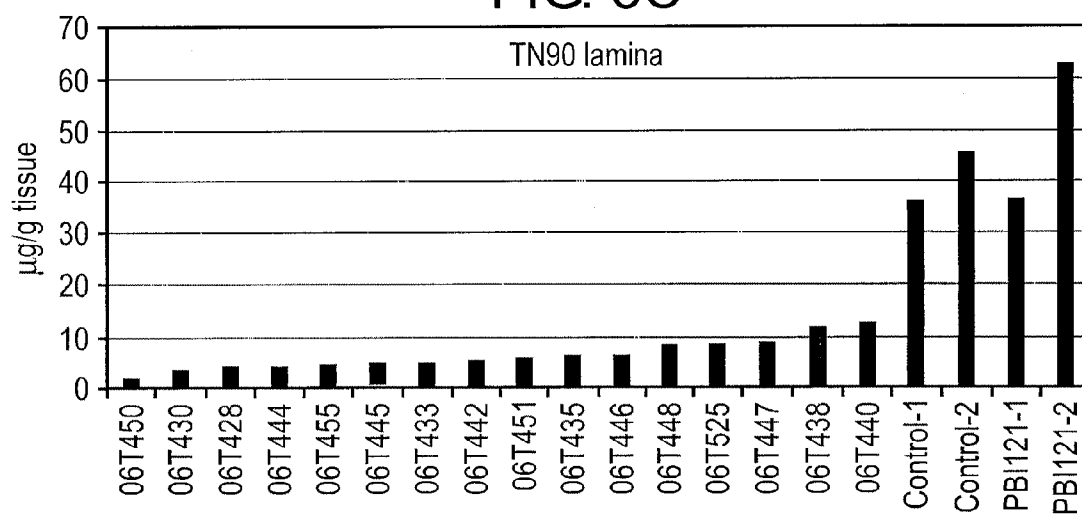


FIG. 3D

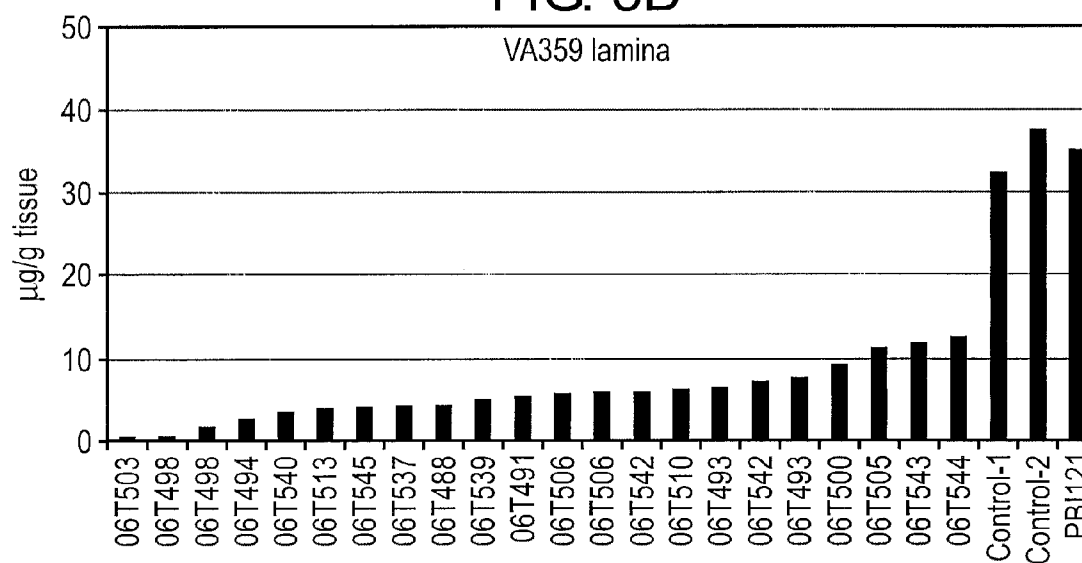


FIG. 4A

**RNAi Construct NtHMA (1382-1584) Encoding RNAi Polynucleotide
SEQ ID NO:45**

Sense Sequence SEQ ID NO:42

5'-TGAGAGCAAGTCAGGTCATCCGATGGCAGCCGCTCTGGTGGAC
TATGCACAATCAAATTCGTTGAGCCAAAGCCTGATAGAGTTGAGCA
GTTTCAAAATTTTCCTGGTGAAGGGATATTTGGAAGAATTGATGGAAT
GGAAATCTATGTCGGAATAGGAAAATTTCTTCAAGAGCTGGATGTA
CCACAGG-3'

Spacer Sequence SEQ ID NO:43

5'-TAAATGGTTGAATCATTTCTTATGCTCATAGTAGAGATAAAACATC
AGAGTTATAATTATAAGTATATGATTTCTCAGTTAATTTTGCTGTTAG
ATTTTCTTTGACCTGTTTAGCACTAATGCGGTGGATGTTTGAA-3'

Reverse Complementary Sequence SEQ ID NO:44

5'-CCTGTGGTACATCCAGCTCTTGAAGAAATTTTCCTATTCCCGACA
TAGATTTCCATTCCATCAATTCTTCCAAATATCCCTTCACCAGGAAAA
TTTTGAAACTGCTCAACTCTATCAGGCTTTGGCTCAACGGAATTTGAT
TGTGCATAGTCCACCAGAGCGGCTGCCATCGGATGACCTGACTTGC
TCTCAATGC-3'

FIG. 4B

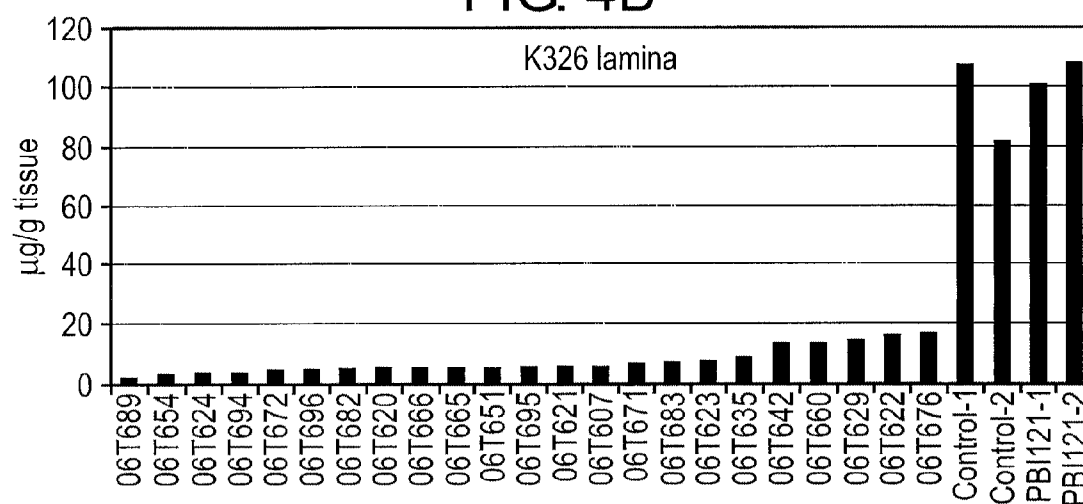


FIG. 4C

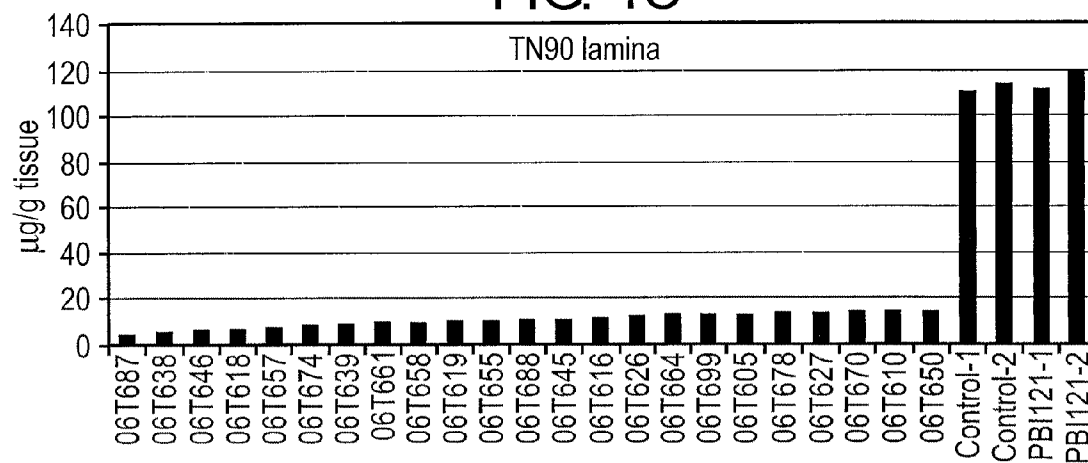
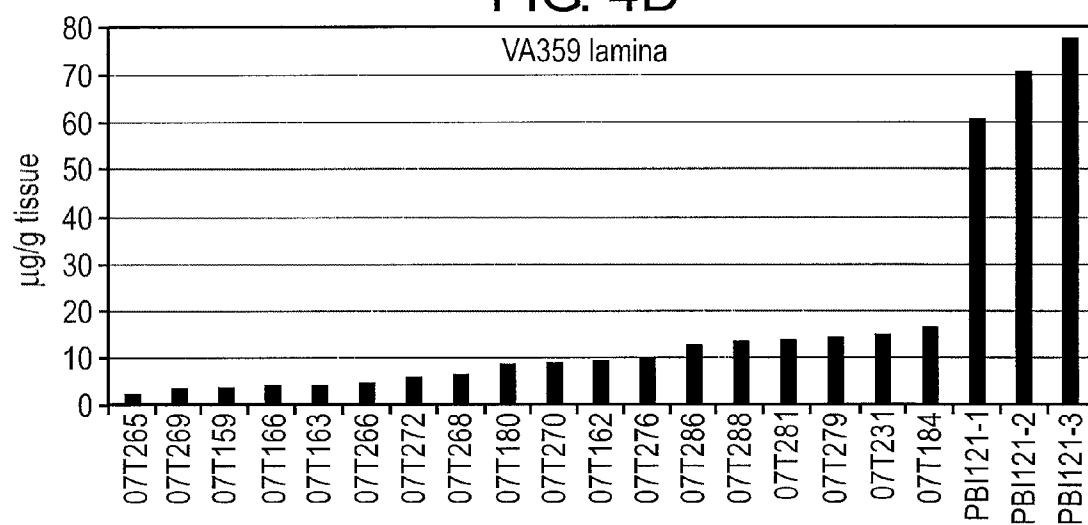


FIG. 4D



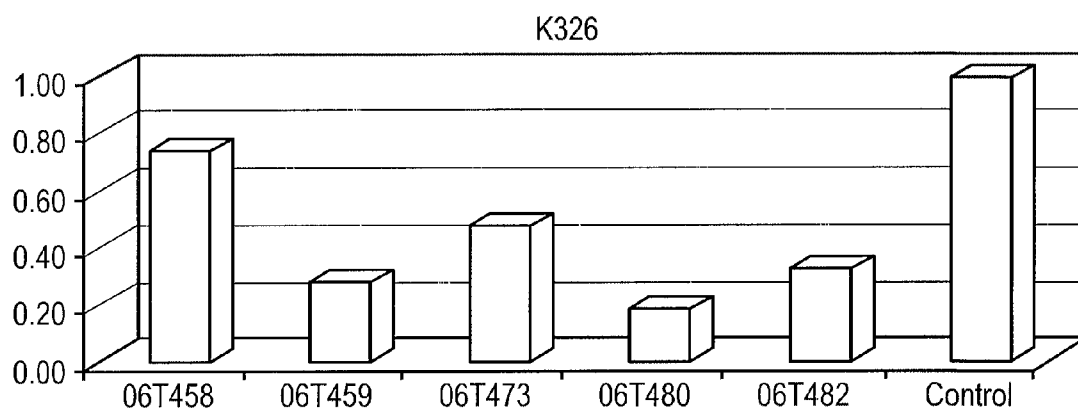


FIG. 5A

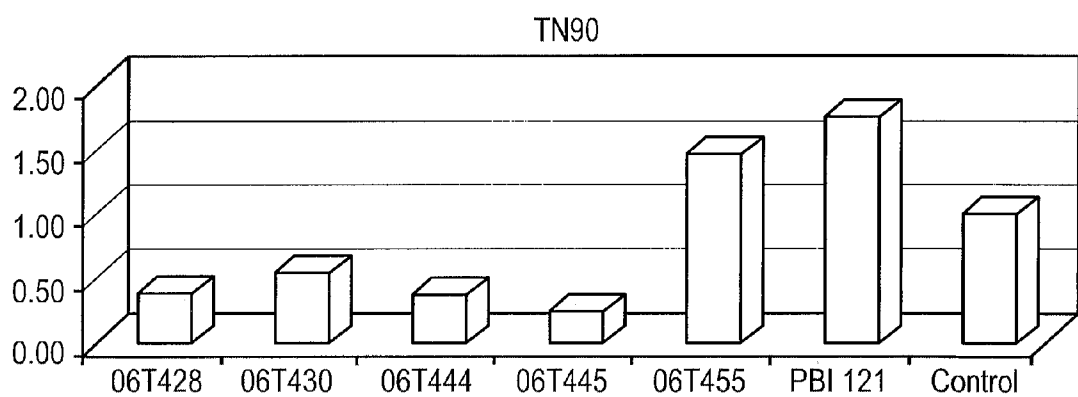


FIG. 5B

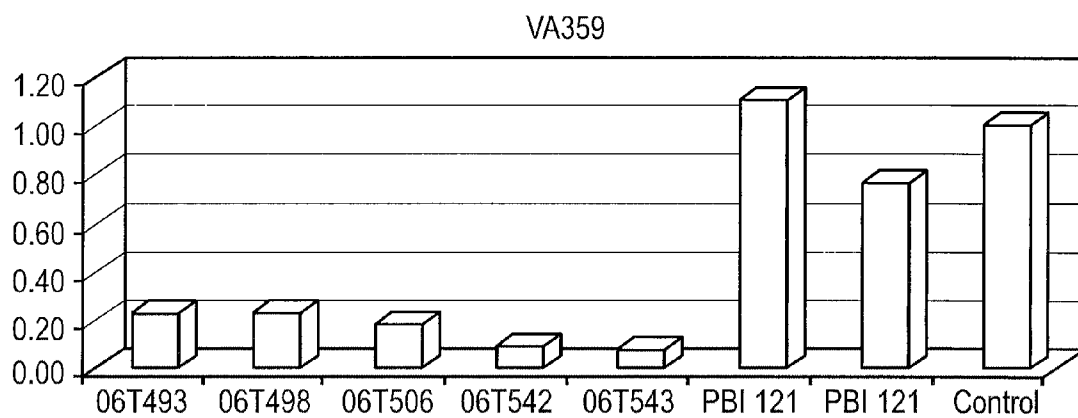


FIG. 5C

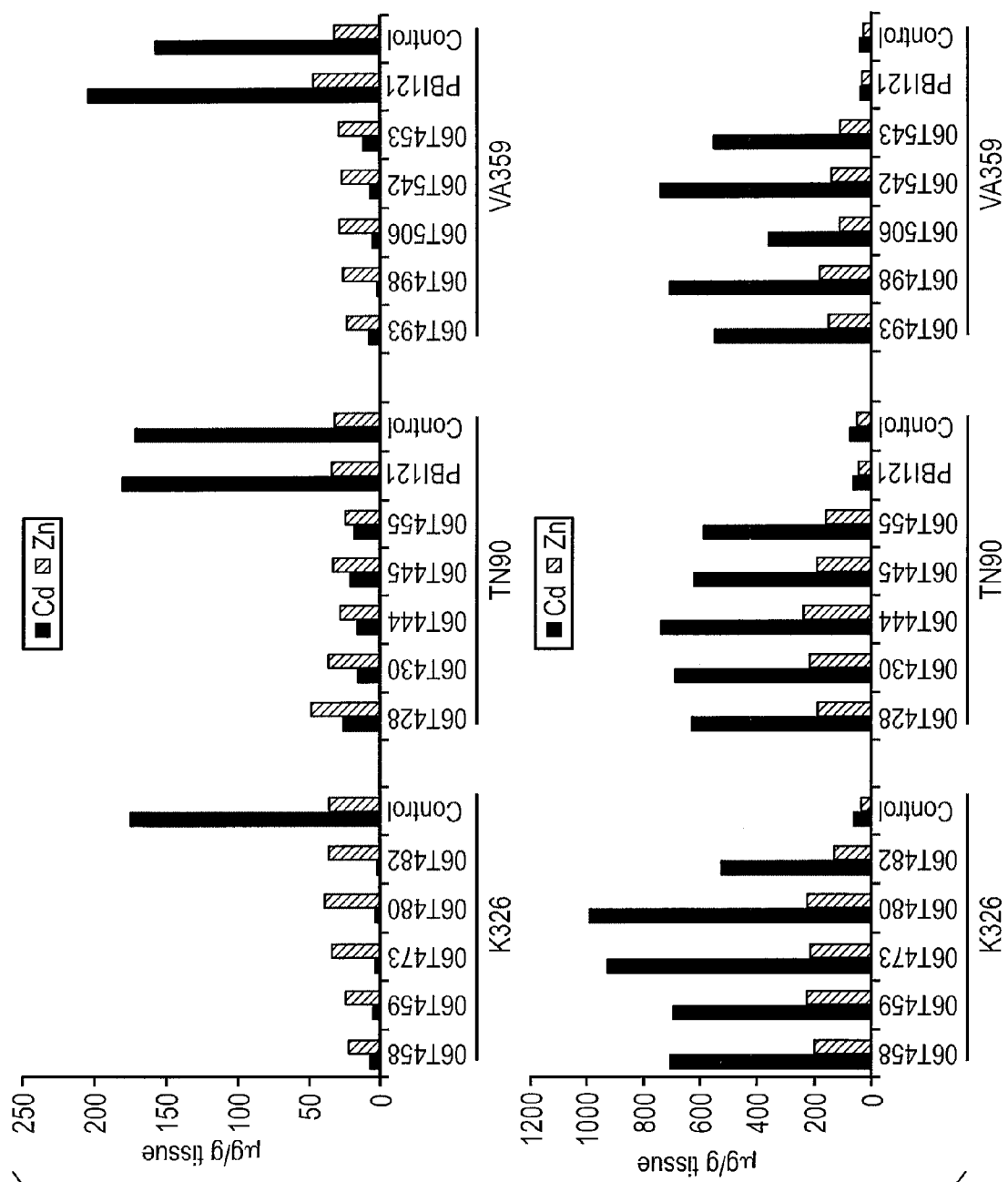


FIG. 6

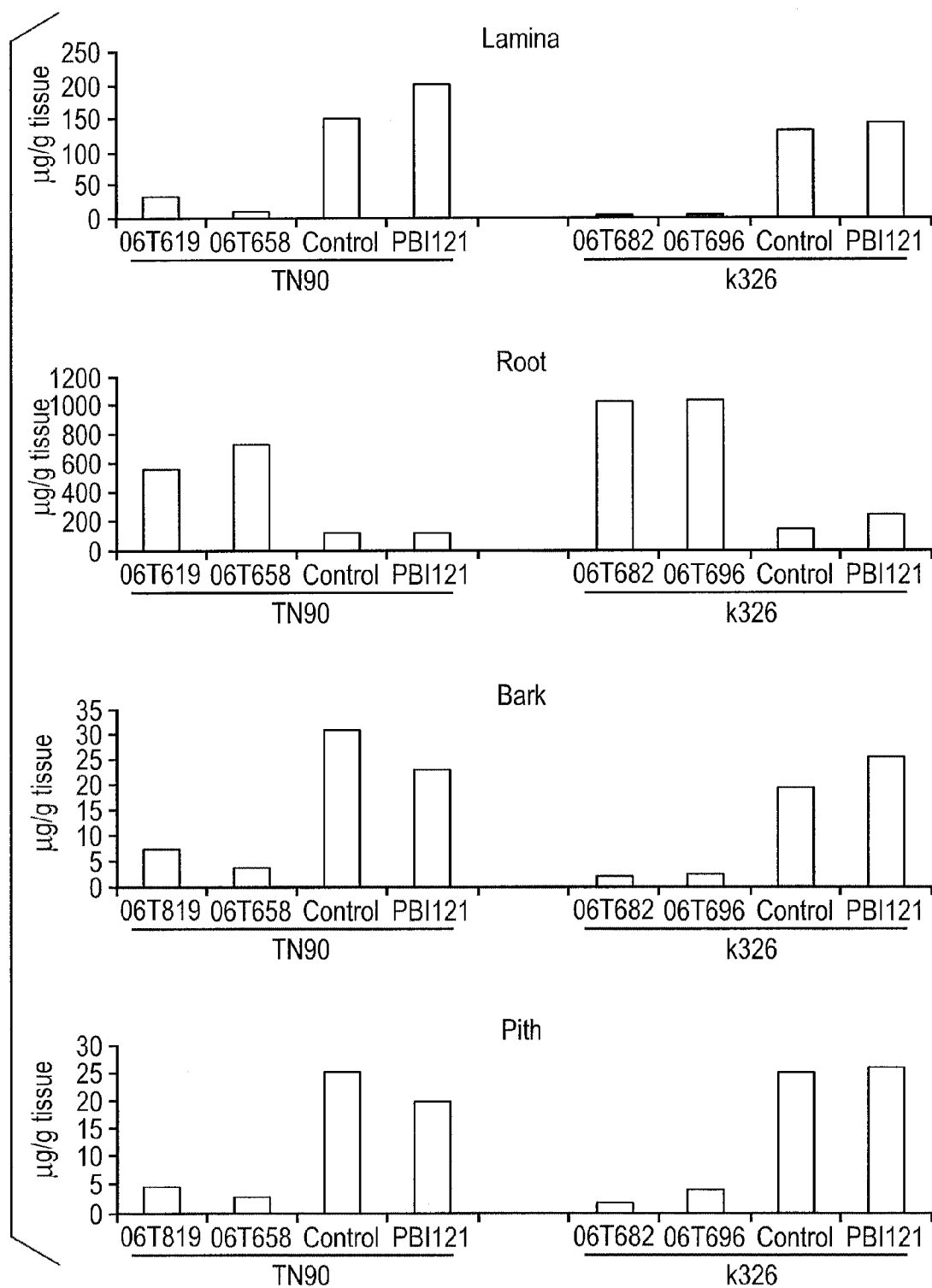


FIG. 7

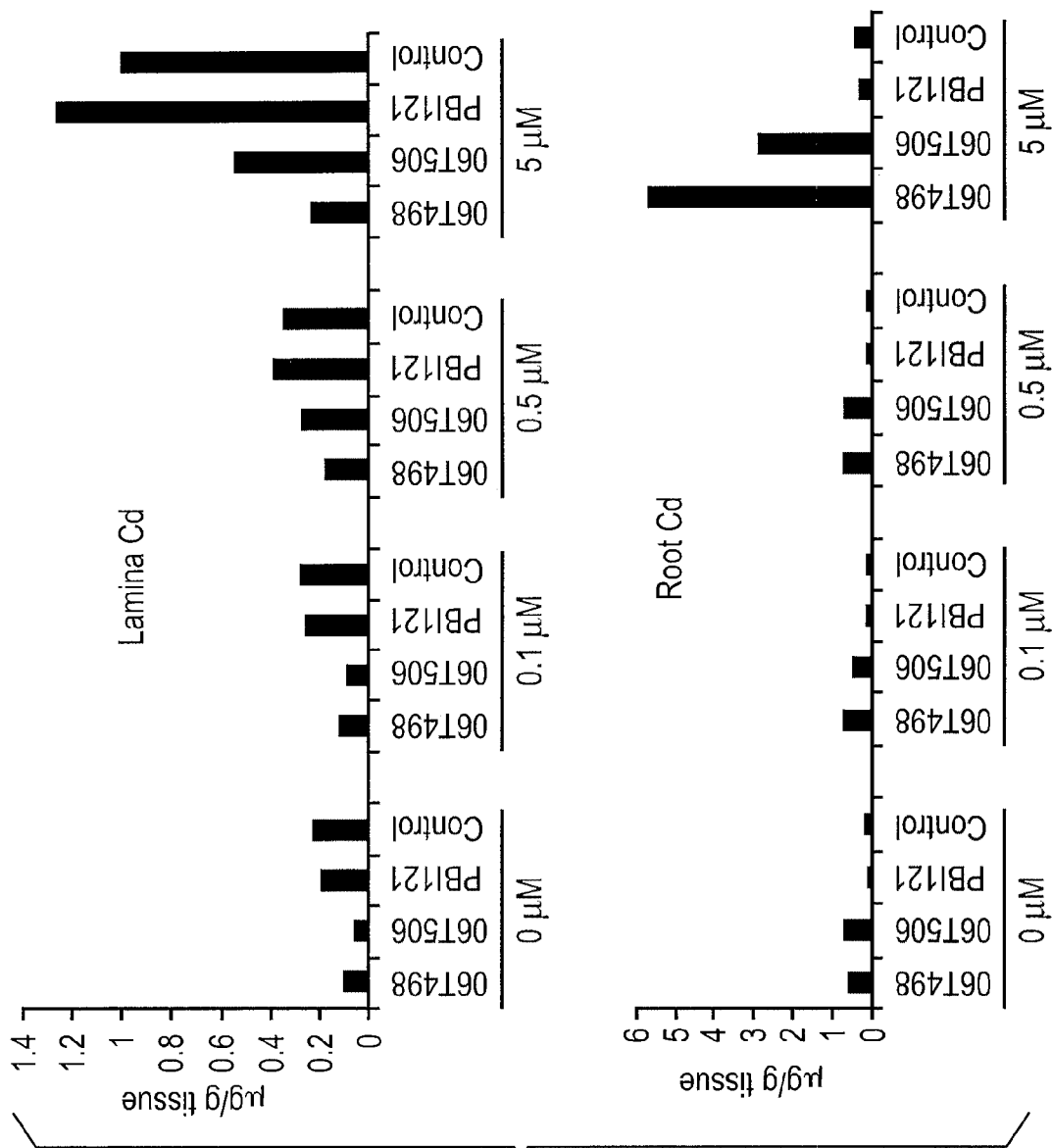


FIG. 8

TRANSGENIC PLANTS MODIFIED FOR REDUCED CADMIUM TRANSPORT, DERIVATIVE PRODUCTS, AND RELATED METHODS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority under 35 U.S.C. § 119 to U.S. Provisional Application No. 60/996,982, filed Dec. 13, 2007, the entire content of which is hereby incorporated by reference.

SEQUENCE LISTING

[0002] This application hereby incorporates by reference the text file filed electronically herewith having the name "sequencelisting.txt" created on Nov. 20, 2008 with a file size of 125,091 bytes.

TECHNICAL FIELD

[0003] Compositions, expression vectors, polynucleotides, polypeptides, transgenic plants, transgenic cell lines, and transgenic seeds, and methods for making and using these embodiments to produce various plants that can reduce the transport of heavy metals into aerial portions.

BACKGROUND

[0004] Plants obtain essential heavy metals, such as Zn, Ni, and Cu, by absorbing metal ion substrates from their environment by various transport mechanisms mediated by transmembrane transporters expressed on the surface of root cells and other vascular tissues. Transporters classified as P-type ATPases, such as P1B-type ATPases, are transporters that translocate positively charged substrates across plasma membranes by utilizing energy liberated from exergonic ATP hydrolysis reactions. P1B-type ATPases are also referred to as heavy metal ATPases ("HMAs") or CPx-type ATPases. HMAs have been grouped by substrate specificity into two subclasses, the Cu/Ag and Zn/Co/Cd/Pb groups. The first P1B-type ATPase to be characterized in plants is AtHMA4, cloned from *Arabidopsis*. Substrate selectivity by HMAs is not strictly limited to the transport of essential metals in that several non-essential metals can be recognized indiscriminantly as substrates, resulting in the accumulation of many non-essential metals, such as Cd, Pb, As, and Hg.

SUMMARY

[0005] Various embodiments are directed to compositions and methods for producing transgenic plants, including transgenic tobacco plants, genetically modified to impede Cadmium (Cd) transport from the root system to the leaf lamina by reducing the expression levels of transporters of the HMA family. A HMA homologue ("NtHMA") has been identified in tobacco, which can be utilized for constructing various RNAi constructs, encoding NtHMA RNAi polynucleotides of interest that can facilitate the degradation of endogenous NtHMA RNA transcripts. Transgenic plants that can express NtHMA RNAi polynucleotides can be utilized for reducing steady-state levels of NtHMA RNA transcripts, and consequently, for reducing the number of functionally active NtHMA transporters available for transporting metals across cellular membranes.

[0006] Various embodiments are directed to recombinant expression vectors comprising various NtHMA RNAi constructs, transgenic plants and seeds genetically modified to exogenously express NtHMA RNAi polynucleotides, cell lines derived from transgenic plants and seeds, and consumable products incorporating leaves derived from transgenic plants produced according to the disclosed methods.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] FIG. 1A is a schematic of a NtHMA genomic clone comprising 11 exons. FIG. 1B provides a list of nucleotide positions mapped to each exon within the isolated NtHMA genomic clone ("Table 1").

[0008] FIG. 2A illustrates an exemplary subcloning strategy for constructing a NtHMA RNAi expression vector that enables the constitutive expression of NtHMA RNAi polynucleotides of interest, as described in Example 2.

[0009] FIG. 2B illustrates a hypothetical double-stranded RNA duplex formed (as "stem-loop-stem" structure) from intra-molecular, base-pair interactions within NtHMA RNAi polynucleotide produced as a product transcribed from an exemplary NtHMA RNAi construct.

[0010] FIG. 3A shows an exemplary RNAi sequence, NtHMA (660-915), for producing NtHMA RNAi polynucleotides of interest, as described in Example 2.

[0011] FIGS. 3B-3D show Cd reduction in leaf lamina of multiple first generation (T0) transgenic lines, representing three varieties, that have been genetically modified to express NtHMA RNAi polynucleotides (660-915), as described in Example 5.

[0012] FIG. 4A shows an exemplary RNAi sequence, NtHMA (1382-1584), for producing NtHMA RNAi polynucleotides of interest, as described in Example 3.

[0013] FIGS. 4B-4D show Cd reduction in leaf lamina of multiple first generation (T0) transgenic lines, representing three varieties, that have been genetically modified to express NtHMA RNAi polynucleotides (1382-1584), as described in Example 5.

[0014] FIGS. 5A-C show normalized NtHMA RNA transcript levels in various first generation (T0) transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest, as determined by quantitative realtime PCR analysis of leaf lamina extracts, as described in Example 6.

[0015] FIG. 6 shows the distribution of Cd and Zn between the leaf lamina and the root of various first generation (T0) transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest, as presented in Table 2 and described in Example 7.

[0016] FIG. 7 shows Cd distribution among the bark, leaf lamina, pith, and root tissues of various first generation (T0) transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest, as presented in Table 3 and described in Example 8.

[0017] FIG. 8 shows Cd distribution between the leaf lamina and the root of various second generation (T1) transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest, as described in Example 9.

DETAILED DESCRIPTION

I. Isolation of Tobacco NtHMA Genes and Gene Products

[0018] FIG. 1A is a schematic of a NtHMA genomic clone comprising 11 exons encoding a heavy metal transporter

related to the HMA family of transporters. Example 1 further describes the identification of the NtHMA genomic clone (_HO-18-2) and 4 NtHMA cDNA clones. FIG. 1B. provides nucleotide positions corresponding to exon and intron subregions mapped within the NtHMA genomic clone (_HO-18-2).

[0019] A. NtHMA Polynucleotides

[0020] The term “polynucleotide” refers to a polymer of nucleotides comprising at least 10 bases in length. The polynucleotides may be DNA, RNA or a DNA/RNA hybrid, comprising ribonucleotides, deoxyribonucleotides, combinations of deoxyribo- and ribo-nucleotides, and combinations of bases and/or modifications, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine, hypoxanthine, isocytosine, and isoguanine. The term includes single- and double-stranded forms of DNA or RNA. The term “DNA” includes genomic DNAs, cDNAs, chemically-synthesized DNAs, PCR-amplified DNAs, and combinations/equivalents thereof. The term “isolated polynucleotide” refers to a polynucleotide not contiguous with any genome of origin, or separated from a native context. The term includes any recombinant polynucleotide molecule such as NtHMA RNAi constructs, NtHMA RNAi expression vectors, NtHMA genomic clones, and fragments and variants thereof.

[0021] As shown in FIG. 1A, the NtHMA genomic clone, designated as SEQ ID NO:1, comprises: intron 1 (SEQ ID NO:4), exon 1 (SEQ ID NO:5), intron 2 (SEQ ID NO:6), exon 2 (SEQ ID NO:7), intron 3 (SEQ ID NO:8), exon 3 (SEQ ID NO:9), intron 4 (SEQ ID NO:10), exon 4 (SEQ ID NO:11), intron 5 (SEQ ID NO:12), exon 5 (SEQ ID NO:13), intron 6 (SEQ ID NO:14), exon 6 (SEQ ID NO:15), intron 7 (SEQ ID NO:16), exon 7 (SEQ ID NO:17), intron 8 (SEQ ID NO:18), exon 8 (SEQ ID NO:19), intron 9 (SEQ ID NO:20), exon 9 (SEQ ID NO:21), intron 10 (SEQ ID NO:22), exon 10 (SEQ ID NO:23), intron 11 (SEQ ID NO:24), exon 11 (SEQ ID NO:25), and intron 12 (SEQ ID NO:26). Various embodiments are directed to isolated polynucleotides representing genomic fragments isolated at the NtHMA locus, comprising SEQ ID NO:1, fragments of SEQ ID NO:1, or variants thereof. Various embodiments are directed to isolated NtHMA polynucleotide variants comprising at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:1, or fragments of SEQ ID NO:1.

[0022] Various embodiments are directed to isolated polynucleotides having sequences that complements that of NtHMA polynucleotide variants comprising at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:1, or fragments of SEQ ID NO:1. Various embodiments are directed to isolated polynucleotides that can specifically hybridize, under moderate to highly stringent conditions, to polynucleotides comprising SEQ ID NO:1, or fragments of SEQ ID NO:1.

[0023] Various embodiments are directed to isolated polynucleotides of NtHMA cDNA (Clone P6663), comprising SEQ ID NO:3, fragments of SEQ ID NO:3, or variants thereof. Various embodiments are directed to isolated NtHMA polynucleotide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:3, or fragments of SEQ ID NO:3. Various embodiments are directed to isolated NtHMA polyribonucleotide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence

identity to SEQ ID NO:3, or fragments of SEQ ID NO:3, and in which Ts have been substituted with Us (e.g., RNAs). Various embodiments are directed to isolated polynucleotides that can specifically hybridize, under moderate to highly stringent conditions, to polynucleotides comprising SEQ ID NO:3, or fragments of SEQ ID NO:3. Various embodiments are directed to isolated polynucleotides having a sequence that complements that of NtHMA polynucleotide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:3, or fragments of SEQ ID NO:3.

[0024] Various embodiments are directed to isolated polynucleotides of NtHMA cDNA (Clone P6643), comprising SEQ ID NO:47, fragments of SEQ ID NO:47, or variants thereof. Various embodiments are directed to isolated NtHMA polynucleotide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:47, fragments of SEQ ID NO:47. Various embodiments are directed to isolated NtHMA polyribonucleotide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:47, fragments of SEQ ID NO:47, and in which Ts have been substituted with Us (e.g., RNAs). Various embodiments are directed to isolated polynucleotides that can specifically hybridize, under moderate to highly stringent conditions, to polynucleotides comprising SEQ ID NO:47, fragments of SEQ ID NO:47. Various embodiments are directed to isolated polynucleotides having a sequence that complements that of NtHMA polynucleotide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:47, fragments of SEQ ID NO:47.

[0025] Various embodiments are directed to biopolymers that are homologous to NtHMA polynucleotides and NtHMA polypeptides (“NtHMA homologues”), which can be identified from different plant species. For example, NtHMA homologues can be experimentally isolated by screening suitable nucleic acid libraries derived from different plant species of interest. Alternatively, NtHMA homologues may be identified by screening genome databases containing sequences from one or more species utilizing a sequence derived from NtHMA polynucleotides and/or NtHMA polypeptides. Such genomic databases are readily available for a number of species (e.g., on the world wide web ([www](http://www.tigr.org/tdb;genetics.wisc.edu;stanford.edu/about.ball;hiv-web.lanl.gov;ncbi.nlm.nih.gov;ebi.ac.uk;andpasteur.fr/other/biology)) at tigr.org/tdb;genetics.wisc.edu;stanford.edu/about.ball;hiv-web.lanl.gov;ncbi.nlm.nih.gov;ebi.ac.uk;andpasteur.fr/other/biology). For example, degenerate oligonucleotide sequences can be obtained by “back-translation” from NtHMA polypeptide fragments. NtHMA polynucleotides can be utilized as probes or primers to identify/amplify related sequences, or to obtain full-length sequences for related NtHMAs by PCR, for example, or by other well-known techniques (e.g., see PCR Protocols: A Guide to Methods and Applications, Innis et. al., eds., Academic Press, Inc. (1990)).

[0026] B. NtHMA Polypeptides

[0027] The term “NtHMA polypeptide” refers to a polypeptide comprising an amino acid sequence designated as SEQ ID NO:2; polypeptides having substantial homology (i.e., sequence similarity) or substantial identity to SEQ ID NO:2; fragments of SEQ ID NO:2; and variants thereof. The NtHMA polypeptides include sequences having sufficient or substantial degree of identity or similarity to SEQ ID NO:2, and that can function by transporting heavy metals across cell membranes.

[0028] NtHMA polypeptides include variants produced by introducing any type of alterations (e.g., insertions, deletions, or substitutions of amino acids; changes in glycosylation states; changes that affect refolding or isomerizations, three-dimensional structures, or self-association states), which can be deliberately engineered or isolated naturally. NtHMA polypeptides may be in linear form or cyclized using known methods (e.g., H. U. Saragovi, et al., *Bio/Technology* 10, 773 (1992); and R. S. McDowell, et al., *J. Amer. Chem. Soc.* 114:9245 (1992), both incorporated herein by reference). NtHMA polypeptides comprise at least 8 to 10, at least 20, at least 30, or at least 40 contiguous amino acids.

[0029] Various embodiments are directed to isolated NtHMA polypeptides encoded by polynucleotide sequence, SEQ ID NO:1, comprising SEQ ID NO:2, fragments of SEQ ID NO:2, or variants thereof. Various embodiments are directed to isolated NtHMA polypeptide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:2, or fragments of SEQ ID NO:2.

[0030] Various embodiments are directed to isolated NtHMA polypeptides (Clone P6663), comprising SEQ ID NO:2, fragments of SEQ ID NO:2, or variants thereof. Various embodiments are directed to isolated NtHMA polypeptide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:2, or fragments of SEQ ID NO:2.

[0031] Various embodiments are directed to isolated NtHMA polypeptides (Clone P6643), comprising SEQ ID NO:49, fragments of SEQ ID NO:49, or variants thereof. Various embodiments are directed to isolated NtHMA polypeptide variants comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:49, or fragments of SEQ ID NO:49.

II. Compositions and Related Methods for Reducing NtHMA Gene Expression and/or NtHMA-Mediated Transporter Activity

[0032] Suitable antagonistic compositions that can down-regulate the expression and/or the activity of NtHMA and NtHMA variants include sequence-specific polynucleotides that can interfere with the transcription of one or more endogenous NtHMA gene(s); sequence-specific polynucleotides that can interfere with the translation of NtHMA RNA transcripts (e.g., dsRNAs, siRNAs, ribozymes); sequence-specific polypeptides that can interfere with the protein stability of NtHMA, the enzymatic activity of NtHMA, and/or the binding activity of NtHMA with respect to substrates and/or regulatory proteins; antibodies that exhibit specificity for NtHMA; and small molecule compounds that can interfere with the protein stability of NtHMA, the enzymatic activity of NtHMA, and/or the binding activity of NtHMA. An effective antagonist can reduce heavy metal (e.g., Cd) transport into leaf lamina structures by at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%.

A. DEFINITIONS

[0033] Throughout this disclosure and the appended claims, the terms “a” and “the” function as singular and plural referents unless the context clearly dictates otherwise. Thus, for example, a reference to “an RNAi polynucleotide”

includes a plurality of such RNAi polynucleotides, and a reference to “the plant” includes reference to one or more of such plants.

[0034] The term “orientation” refers to a particular order in the placement of a polynucleotide relative to the position of a reference polynucleotide. A linear DNA has two possible orientations: the 5'-to-3' direction and the 3'-to-5' direction. For example, if a reference sequence is positioned in the 5'-to-3' direction, and if a second sequence is positioned in the 5'-to-3' direction within the same polynucleotide molecule/strand, then the reference sequence and the second sequence are orientated in the same direction, or have the same orientation. Typically, a promoter sequence and a gene of interest under the regulation of the given promoter are positioned in the same orientation. However, with respect to the reference sequence positioned in the 5'-to-3' direction, if a second sequence is positioned in the 3'-to-5' direction within the same polynucleotide molecule/strand, then the reference sequence and the second sequence are orientated in anti-sense direction, or have anti-sense orientation. Two sequences having anti-sense orientations with respect to each other can be alternatively described as having the same orientation, if the reference sequence (5'-to-3' direction) and the reverse complementary sequence of the reference sequence (reference sequence positioned in the 5'-to-3') are positioned within the same polynucleotide molecule/strand.

[0035] The term “NtHMA RNAi expression vector” refers to a nucleic acid vehicle that comprises a combination of DNA components for enabling the transport and the expression of NtHMA RNAi constructs. Suitable expression vectors include episomes capable of extra-chromosomal replication such as circular, double-stranded DNA plasmids; linearized double-stranded DNA plasmids; and other functionally equivalent expression vectors of any origin. A suitable NtHMA RNAi expression vector comprises at least a promoter positioned upstream and operably-linked to a NtHMA RNAi construct, as defined below.

[0036] The term “NtHMA RNAi construct” refers to a double-stranded, recombinant DNA fragment that encodes “NtHMA RNAi polynucleotides” having RNA interference activity. A NtHMA RNAi construct comprises a “template strand” base-paired with a complementary “sense or coding strand.” A given NtHMA RNAi construct can be inserted into a NtHMA RNAi expression vector in two possible orientations, either in the same (or sense) orientation or in the reverse (or anti-sense) orientation with respect to the orientation of a promoter positioned within a NtHMA RNAi expression vector.

[0037] The term “NtHMA RNAi polynucleotides” can target NtHMA RNA for enzymatic degradation, involving the formation of smaller fragments of NtHMA RNAi polynucleotides (“siRNAs”) that can bind to multiple complementary sequences within the target NtHMA RNA. Expression levels of one or more NtHMA gene(s) can be reduced by the RNA interference activity of NtHMA RNAi polynucleotides.

[0038] The term “template strand” refers to the strand comprising a sequence that complements that of the “sense or coding strand” of a DNA duplex, such as NtHMA genomic fragment, NtHMA cDNA, or NtHMA RNAi construct, or any DNA fragment comprising a nucleic acid sequence that can be transcribed by RNA polymerase. During transcription, RNA polymerase can translocate along the template strand in the 3'-to-5' direction during nascent RNA synthesis.

[0039] The terms “sense strand” or “coding strand” refer to the strand comprising a sequence that complements that of the template strand in a DNA duplex. For example, the sequence of the sense strand (“sense sequence”) for the identified NtHMA genomic clone is designated as SEQ ID NO:1. For example, the sense sequence for NtHMA cDNA, identified as clone P6663, is designated as SEQ ID NO:3. For example, the sense sequence for NtHMA cDNA, identified as clone P6643, is designated as SEQ ID NO:46. For example, if the sense strand comprises a hypothetical sequence 5'-TAATCCGGT-3', then the substantially identical corresponding sequence within a hypothetical target mRNA is 5'-UAAUCCGGU-3'.

[0040] The term “reverse complementary sequence” refers to the sequence that complements the “sense sequence” of interest (e.g., exon sequence) positioned within the same strand, in the same orientation with respect to the sense sequence. For example, if a strand comprises a hypothetical sequence 5'-TAATCCGGT-3', then the reverse complementary sequence 5'-ACCGGATTA-3' may be operably-linked to the sense sequence, separated by a spacer sequence.

[0041] The terms “NtHMA RNA transcript” or “NtHMA RNA,” in the context of RNA interference, refer to polyribonucleic acid molecules produced within a host plant cell of interest, resulting from the transcription of endogenous genes of the HMA family, including the isolated NtHMA gene (SEQ ID NO:1). Thus, these terms include any RNA species or RNA variants produced as transcriptional products from HMA-related genes that may be distinct from the isolated NtHMA gene (SEQ ID NO:1) but having sufficient similarity at structural and/or functional levels to be classified within the same family. For example, if a host plant cell selected for genetic modification according to the disclosed methods is tobacco, then target NtHMA RNA transcripts include: (1) pre-mRNAs and mRNAs produced from the transcription of the isolated NtHMA gene (SEQ ID NO:1); (2) pre-mRNAs and mRNAs produced from the transcription of any genes having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to the sequence of the isolated NtHMA gene (SEQ ID NO:1) (i.e. other distinct genes substantially identical to the identified NtHMA gene and encoding related isoforms of HMA transporters); and (3) pre-mRNAs and mRNAs produced from the transcription of alleles of the NtHMA gene (SEQ ID NO:1). The NtHMA RNA transcripts include RNA variants produced as a result of alternative RNA splicing reactions of heteronuclear RNAs (“hnRNAs”) of a particular NtHMA gene, mRNA variants resulting from such alternative RNA splicing reactions, and any intermediate RNA variants.

[0042] The terms “homology” or “identity” or “similarity” refer to the degree of sequence similarity between two polypeptides or between two nucleic acid molecules compared by sequence alignment. The degree of homology between two discrete nucleic acid sequences being compared is a function of the number of identical, or matching, nucleotides at comparable positions. The percent identity may be determined by visual inspection and mathematical calculation. Alternatively, the percent identity of two nucleic acid sequences can be determined by comparing sequence information using the GAP computer program, version 6.0 described by Devereux et al. (Nucl. Acids Res. 12:387, 1984) and available from the University of Wisconsin Genetics Computer Group (UWCGC). Typical default parameters for the GAP program include: (1) a unary comparison matrix

(containing a value of 1 for identities and 0 for non-identities) for nucleotides, and the weighted comparison matrix of Gribskov and Burgess, Nucl. Acids Res. 14:6745, 1986, as described by Schwartz and Dayhoff, eds., Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, pp. 353-358, 1979; (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps. Various programs known to persons skilled in the art of sequence comparison can be alternatively utilized.

[0043] The term “upstream” refers to a relative direction/position with respect to a reference element along a linear polynucleotide sequence, which indicates a direction/position towards the 5' end of the polynucleotide sequence. “Upstream” may be used interchangeably with the “5' end of a reference element.”

[0044] The term “operably-linked” refers to the joining of distinct DNA elements, fragments, or sequences to produce a functional transcriptional unit or a functional expression vector.

[0045] The term “promoter” refers to a nucleic acid element/sequence, typically positioned upstream and operably-linked to a double-stranded DNA fragment, such as a NtHMA RNAi construct. For example, a suitable promoter enables the transcriptional activation of a NtHMA RNAi construct by recruiting the transcriptional complex, including the RNA polymerase and various factors, to initiate RNA synthesis. “Promoters” can be derived entirely from regions proximate to a native gene of interest, or can be composed of different elements derived from different native promoters and/or synthetic DNA segments. Suitable promoters include tissue-specific promoters recognized by tissue-specific factors present in different tissues or cell types (e.g., root-specific promoters, shoot-specific promoters, xylem-specific promoters), or present during different developmental stages, or present in response to different environmental conditions. Suitable promoters include constitutive promoters that can be activated in most cell types without requiring specific inducers. Examples of suitable promoters for controlling NtHMA RNAi polypeptide production include the cauliflower mosaic virus 35S (CaMV/35S), SSU, OCS, lib4, usp, STLS1, B33, nos or ubiquitin- or phaseolin-promoters. Persons skilled in the art are capable of generating multiple variations of recombinant promoters, as described in a number of references, such as Okamuro and Goldberg, Biochemistry of Plants, Vol. 15:pp 1-82 (1989).

[0046] Tissue-specific promoters are transcriptional control elements that are only active in particular cells or tissues at specific times during plant development, such as in vegetative tissues or reproductive tissues. Tissue-specific expression can be advantageous, for example, when the expression of polynucleotides in certain tissues is preferred. Examples of tissue-specific promoters under developmental control include promoters that can initiate transcription only (or primarily only) in certain tissues, such as vegetative tissues, e.g., roots or leaves, or reproductive tissues, such as fruit, ovules, seeds, pollen, pistils, flowers, or any embryonic tissue. Reproductive tissue-specific promoters may be, e.g., anther-specific, ovule-specific, embryo-specific, endosperm-specific, integument-specific, seed and seed coat-specific, pollen-specific, petal-specific, sepal-specific, or combinations thereof.

[0047] Suitable leaf-specific promoters include pyruvate, orthophosphate dikinase (PPDK) promoter from C4 plant

(maize), cab-m1Ca+2 promoter from maize, the *Arabidopsis thaliana* myb-related gene promoter (Atmyb5), the ribulose biphosphate carboxylase (RBCS) promoters (e.g., the tomato RBCS 1, RBCS2 and RBCS3A genes expressed in leaves and light-grown seedlings, RBCS1 and RBCS2 expressed in developing tomato fruits, and/or ribulose biphosphate carboxylase promoter expressed almost exclusively in mesophyll cells in leaf blades and leaf sheaths at high levels).

[0048] Suitable senescence-specific promoters include a tomato promoter active during fruit ripening, senescence and abscission of leaves, a maize promoter of gene encoding a cysteine protease. Suitable anther-specific promoters can be used. Such promoters are known in the art or can be discovered by known techniques; see, e.g., Bhalla and Singh (1999) Molecular control of male fertility in *Brassica*, Proc. 10th Annual Rapeseed Congress, Canberra, Australia; van Tunen et al. (1990) Pollen- and anther-specific chi promoters from petunia: tandem promoter regulation of the chiA gene. Plant Cell 2:393-40; Jeon et al. (1999) Isolation and characterization of an anther-specific gene, RA8, from rice (*Oryza sativa* L.). Plant Molecular Biology 39:35-44; and Twell et al. (1993) Activation and developmental regulation of an *Arabidopsis* anther-specific promoter in microspores and pollen of *Nicotiana tabacum*. Sex. Plant Reprod. 6:217-224.

[0049] Suitable root-preferred promoters known to persons skilled in the art may be selected. See, for example, Hire et al. (1992) Plant Mol. Biol. 20(2):207-218 (soybean root-specific glutamine synthetase gene); Keller and Baumgartner (1991) Plant Cell 3(10):1051-1061 (root-specific control element in the GRP 1.8 gene of French bean); Sanger et al. (1990) Plant Mol. Biol. 14(3):433-443 (root-specific promoter of the mannopine synthase (MAS) gene of *Agrobacterium tumefaciens*); and Miao et al. (1991) Plant Cell 3(1):11-22 (full-length cDNA clone encoding cytosolic glutamine synthetase (GS), which is expressed in roots and root nodules of soybean). See also Bogusz et al. (1990) Plant Cell 2(7):633-641, where two root-specific promoters isolated from hemoglobin genes from the nitrogen-fixing nonlegume *Parasponia andersonii* and the related non-nitrogen-fixing nonlegume *Trema tomentosa* are described.

[0050] Suitable seed-preferred promoters include both seed-specific promoters (those promoters active during seed development such as promoters of seed storage proteins) and seed-germinating promoters (those promoters active during seed germination). See, e.g., Thompson et al. (1989) BioEssays 10: 108, herein incorporated by reference. Such seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); milps (myo-inositol-1-phosphate synthase); mZE40-2, also known as Zm-40 (U.S. Pat. No. 6,403,862); nuc1 (U.S. Pat. No. 6,407,315); and celA (cellulose synthase) (see WO 00/11177). Gama-zein is an endosperm-specific promoter. Glob-1 is an embryo-specific promoter. For dicots, seed-specific promoters include, but are not limited to, bean β -phaseolin, napin, β -conglycinin, soybean lectin, cruciferin, and the like. For monocots, seed-specific promoters include, but are not limited to, a maize 15 kDa zein promoter, a 22 kDa zein promoter, a 27 kDa zein promoter, a g-zein promoter, a 27 kDa γ -zein promoter (such as gzw64A promoter, see Genbank Accession #S78780), a waxy promoter, a shrunken 1 promoter, a shrunken 2 promoter, a globulin 1 promoter (see Genbank Accession # L22344), an ltp2 promoter (Kalla, et al., Plant Journal 6:849-860 (1994)); U.S. Pat. No. 5,525,716), cim1 promoter (see U.S. Pat. No. 6,225,529) maize

end1 and end2 promoters (See U.S. Pat. Nos. 6,528,704 and 6,903,205); nuc1 promoter (U.S. Pat. No. 6,407,315); Zm40 promoter (U.S. Pat. No. 6,403,862); eep1 and eep2; lec1 (U.S. Pat. No. 7,122,658); thioredoxin H promoter; mlip15 promoter (U.S. Pat. No. 6,479,734); PCNA2 promoter; and the shrunken-2 promoter. (Shaw et al., Plant Phys 98:1214-1216, 1992; Zhong Chen et al., PNAS USA 100:3525-3530, 2003).

[0051] Examples of inducible promoters include promoters responsive to pathogen attack, anaerobic conditions, elevated temperature, light, drought, cold temperature, or high salt concentration. Pathogen-inducible promoters include those from pathogenesis-related proteins (PR proteins), which are induced following infection by a pathogen (e.g., PR proteins, SAR proteins, beta-1,3-glucanase, chitinase). See, for example, Redolfi et al. (1983) Neth. J. Plant Pathol. 89:245-254; Uknes et al. (1992) Plant Cell 4:645-656; and Van Loon (1985) Plant Mol. Virol. 4:111-116. See also inducible maize promoters described in U.S. Pat. No. 6,429,362.

[0052] In addition to plant promoters, other suitable promoters may be derived from bacterial origin (e.g., the octopine synthase promoter, the nopaline synthase promoter and other promoters derived from Ti plasmids), or may be derived from viral promoters (e.g., 35S and 19S RNA promoters of cauliflower mosaic virus (CaMV), constitutive promoters of tobacco mosaic virus, cauliflower mosaic virus (CaMV) 19S and 35S promoters, or figwort mosaic virus 35S promoter).

[0053] The term "enhancer" refers to a nucleic acid molecule, or a nucleic acid sequence, that can recruit transcriptional regulatory proteins such as transcriptional activators, to enhance transcriptional activation by increasing promoter activity. Suitable enhancers can be derived from regions proximate to a native promoter of interest (homologous sources) or can be derived from non-native contexts (heterologous sources) and operably-linked to any promoter of interest within NtHMA RNAi expression vectors to enhance the activity and/or the tissue-specificity of a promoter. Some enhancers can operate in any orientation with respect to the orientation of a transcription unit. For example, enhancers may be positioned upstream or downstream of a transcriptional unit comprising a promoter and a NtHMA RNAi construct. Persons skilled in the art are capable of operably-linking enhancers and promoters to optimize the transcription levels of NtHMA RNAi constructs.

B. RNAi EXPRESSION VECTORS COMPRISING NtHMA RNAi CONSTRUCTS ENCODING NtHMA RNAi POLYNUCLEOTIDES

[0054] RNA Interference ("RNAi") or RNA silencing is an evolutionarily conserved process by which specific mRNAs can be targeted for enzymatic degradation. A double-stranded RNA (dsRNA) must be introduced or produced by a cell (e.g., dsRNA virus, or NtHMA RNAi polynucleotides) to initiate the RNAi pathway. The dsRNA can be converted into multiple siRNA duplexes of 21-23 bp length ("siRNAs") by RNases III, which are dsRNA-specific endonucleases ("Dicer"). The siRNAs can be subsequently recognized by RNA-induced silencing complexes ("RISC") that promote the unwinding of siRNA through an ATP-dependent process. The unwound antisense strand of the siRNA guides the activated RISC to the targeted mRNA (e.g., NtHMA RNA variants) comprising a sequence complementary to the siRNA anti-sense strand. The targeted mRNA and the anti-sense strand can form an A-form helix, and the major groove of the

A-form helix can be recognized by the activated RISC. The target mRNA can be cleaved by activated RISC at a single site defined by the binding site of the 5'-end of the siRNA strand. The activated RISC can be recycled to catalyze another cleavage event.

[0055] FIG. 2A illustrates the construction of an exemplary NtHMA RNAi expression vector. NtHMA RNAi expression vectors comprising NtHMA RNAi constructs encoding NtHMA RNAi polynucleotides exhibit RNA interference activity by reducing the expression level of NtHMA mRNAs, NtHMA pre-mRNAs, and related NtHMA RNA variants. The expression vectors comprise a promoter positioned upstream and operably-linked to a NtHMA RNAi construct, as further defined below. NtHMA RNAi expression vectors comprise a suitable minimal core promoter, a NtHMA RNAi construct of interest, an upstream (5') regulatory region, a downstream (3') regulatory region, including transcription termination and polyadenylation signals, and other sequences known to persons skilled in the art, such as various selection markers.

[0056] The NtHMA polynucleotides can be produced in various forms, including as double-stranded hairpin-like structures ("dsRNAi"). The NtHMA dsRNAi can be enzymatically converted to double-stranded NtHMA siRNAs. One of the strands of the NtHMA siRNA duplex can anneal to a complementary sequence within the target NtHMA mRNA and related NtHMA RNA variants. The siRNA/mRNA duplexes are recognized by RISC that can cleave NtHMA RNAs at multiple sites in a sequence-dependent manner, resulting in the degradation of the target NtHMA mRNA and related NtHMA RNA variants.

[0057] FIG. 2B illustrates the formation of a hypothetical double-stranded RNA duplex formed (as "stem-loop-stem" structure) as a product transcribed from an exemplary NtHMA RNAi construct. In FIG. 2B, a hypothetical NtHMA RNAi construct 10 is shown, comprising 3 double-stranded DNA fragments, such as fragments 1-3. Fragment 1 is positioned upstream and operably-linked to fragment 2, which is positioned upstream and operably-linked to fragment 3, for which DNA strands/sequences 4, 6, and 8 are linked together in tandem to form strand 11, as shown. Alternatively, a NtHMA RNAi construct comprises "a sense sequence" 5, which is positioned upstream and operably-linked to "a spacer sequence" 7, which is positioned upstream and operably-linked to "a reverse complementary sequence" 9. The strands/sequences 5, 7, and 9 can be linked together in tandem to form strand/sequence 12. Alternatively, a NtHMA RNAi construct comprises "a sense sequence" 8, which is positioned upstream and operably-linked to "a spacer sequence" 6, which is positioned upstream and operably-linked to "a reverse complementary sequence" 4. The strands/sequences 8, 6, and 4 can be linked together in tandem to form strand/sequence 11. Strand 12 is complementary to strand 11. Strand 11 is a template strand that can be transcribed into a NtHMA RNAi polynucleotide 13. The NtHMA RNAi polynucleotide 13 forms a hair-pin ("stem-loop-stem") structure, in which the stem 16 is a complementary region resulting from intramolecular base-pair interactions of the NtHMA RNAi polynucleotide 15 and the loop 17 represents a non-complementary region encoded by a spacer sequence, such as strands/sequences 6 or 7.

[0058] Any NtHMA RNA polynucleotide of interest can be produced by selecting a suitable sequence composition, loop size, and stem length for producing the NtHMA hairpin duplex. A suitable range for designing stem lengths of a

hairpin duplex, includes stem lengths of 20-30 nucleotides, 30-50 nucleotides, 50-100 nucleotides, 100-150 nucleotides, 150-200 nucleotides, 200-300 nucleotides, 300-400 nucleotides, 400-500 nucleotides, 500-600 nucleotides, and 600-700 nucleotides. A suitable range for designing loop lengths of a hairpin duplex, includes loop lengths of 4-25 nucleotides, 25-50 nucleotides, or longer if the stem length of the hair duplex is substantial. In certain context, hairpin structures with duplexed regions longer than 21 nucleotides may promote effective siRNA-directed silencing, regardless of loop sequence and length.

[0059] Exemplary NtHMA RNAi constructs for down-regulating the expression level of the NtHMA gene (SEQ ID NO:1) and other NtHMA-related genes include the following:

[0060] Various embodiments are directed to NtHMA RNAi expression vectors comprising a NtHMA RNAi construct that comprises one or more of: intron 1 (SEQ ID NO:4), exon 1 (SEQ ID NO:5), intron 2 (SEQ ID NO:6), exon 2 (SEQ ID NO:7), intron 3 (SEQ ID NO:8), exon 3 (SEQ ID NO:9), intron 4 (SEQ ID NO:10), exon 4 (SEQ ID NO:11), intron 5 (SEQ ID NO:12), exon 5 (SEQ ID NO:13), intron 6 (SEQ ID NO:14), exon 6 (SEQ ID NO:15), intron 7 (SEQ ID NO:16), exon 7 (SEQ ID NO:17), intron 8 (SEQ ID NO:18), exon 8 (SEQ ID NO:19), intron 9 (SEQ ID NO:20), exon 9 (SEQ ID NO:21), intron 10 (SEQ ID NO:22), exon 10 (SEQ ID NO:23), intron 11 (SEQ ID NO:24), exon 11 (SEQ ID NO:25), and intron 12 (SEQ ID NO:26), fragments thereof, and variants thereof.

[0061] Various embodiments are directed to NtHMA RNAi expression vectors comprising: a NtHMA RNAi construct having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to a sequence selected from the group consisting of: exon 1 (SEQ ID NO:5), a fragment of exon 1 (SEQ ID NO:5), exon 2 (SEQ ID NO:7), a fragment of exon 2 (SEQ ID NO:7), exon 3 (SEQ ID NO:9), a fragment of exon 3 (SEQ ID NO:9), exon 4 (SEQ ID NO:11), a fragment of exon 4 (SEQ ID NO:11), exon 5 (SEQ ID NO:13), a fragment of exon 5 (SEQ ID NO:13), exon 6 (SEQ ID NO:15), a fragment of exon 6 (SEQ ID NO:15), exon 7 (SEQ ID NO:17), a fragment of exon 7 (SEQ ID NO:17), exon 8 (SEQ ID NO:19), a fragment of exon 8 (SEQ ID NO:19), exon 9 (SEQ ID NO:21), a fragment of exon 9 (SEQ ID NO:21), exon 10 (SEQ ID NO:23), a fragment of exon 10 (SEQ ID NO:23), exon 11 (SEQ ID NO:25), and a fragment of exon 11 (SEQ ID NO:25).

[0062] Various embodiments are directed to NtHMA RNAi expression vectors comprising: a NtHMA RNAi construct encoding NtHMA RNAi polynucleotides capable of self-annealing to form a hairpin structure, in which the RNAi construct comprises (a) a first sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:3 or SEQ ID NO:47; (b) a second sequence encoding a spacer element of the NtHMA RNAi polynucleotide that forms a loop of the hairpin structure; and (c) a third sequence comprising a reverse complementary sequence of the first sequence, positioned in the same orientation as the first sequence, wherein the second sequence is positioned between the first sequence and the third sequence, and the second sequence is operably-linked to the first sequence and to the third sequence.

[0063] Various embodiments are directed to NtHMA RNAi expression vectors comprising: a NtHMA RNAi construct encoding NtHMA RNAi polynucleotides capable of self-

annealing to form a hairpin structure, in which the RNAi construct comprises (a) a first sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:3 or SEQ ID NO:47; (b) a second sequence encoding a spacer element of the NtHMA RNAi polynucleotide that forms a loop of the hairpin structure; and (c) a third sequence comprising a reverse complementary sequence of the first sequence (SEQ ID NO:46 or SEQ ID NO:48), positioned in the same orientation as the first sequence, wherein the second sequence is positioned between the first sequence and the third sequence, and the second sequence is operably-linked to the first sequence and to the third sequence.

[0064] Various embodiments are directed to NtHMA RNAi expression vectors comprising: a NtHMA RNAi construct that comprises a first sequence having “substantial similarity,” or having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:3, or portions of SEQ ID NO:3. Various embodiments are directed to NtHMA RNAi expression vectors comprising a NtHMA RNAi construct that comprises a first sequence having “substantial similarity,” or having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:47, or portions of SEQ ID NO:47.

[0065] Various embodiments are directed to a NtHMA RNAi expression vectors comprising: a NtHMA RNAi construct that comprises a second sequence having “substantial similarity,” or having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to a sequence selected from the group consisting of: intron 1 (SEQ ID NO:4), a fragment of intron 1 (SEQ ID NO:4), intron 2 (SEQ ID NO:6), a fragment of intron 2 (SEQ ID NO:6), intron 3 (SEQ ID NO:8), a fragment of intron 3 (SEQ ID NO:8), intron 4 (SEQ ID NO:10), a fragment of intron 4 (SEQ ID NO:10), intron 5 (SEQ ID NO:12), a fragment of intron 5 (SEQ ID NO:12), intron 6 (SEQ ID NO:14), a fragment of intron 6 (SEQ ID NO:14), intron 7 (SEQ ID NO:16), a fragment of intron 7 (SEQ ID NO:16), intron 8 (SEQ ID NO:18), a fragment of intron 8 (SEQ ID NO:18), intron 9 (SEQ ID NO:20), a fragment of intron 9 (SEQ ID NO:20), intron 10 (SEQ ID NO:22), a fragment of intron 10 (SEQ ID NO:22), intron 11 (SEQ ID NO:24), a fragment of intron 11 (SEQ ID NO:24), intron 12 (SEQ ID NO:26), and a fragment of intron 12 (SEQ ID NO:26). Alternatively, the second sequence of the NtHMA RNAi construct can be randomly generated without utilizing an intron sequence derived from the NtHMA gene (SEQ ID NO:1).

[0066] Various embodiments are directed to NtHMA RNAi expression vectors comprising: a NtHMA RNAi construct that comprises a third sequence having “substantial similarity,” or having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:46, or portions of SEQ ID NO:46. Various embodiments are directed to NtHMA RNAi expression vectors comprising a NtHMA RNAi construct that comprises a third sequence having “substantial similarity,” or having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NO:48, or portions of SEQ ID NO:48.

[0067] Various embodiments are directed to NtHMA RNAi expression vectors comprising: a NtHMA RNAi construct that comprises a third sequence having “substantial similarity,” or having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to a reverse complementary sequence selected from the group consisting of: SEQ

ID NO:27 (exon 1), a fragment of SEQ ID NO:27 (exon 1), SEQ ID NO:28 (exon 2), a fragment of SEQ ID NO:28 (exon 2), SEQ ID NO:29 (exon 3), a fragment of SEQ ID NO:29 (exon 3), SEQ ID NO:30 (exon 4), a fragment of SEQ ID NO:30 (exon 4), SEQ ID NO:31 (exon 5), a fragment of SEQ ID NO:31 (exon 5), SEQ ID NO:32 (exon 6), a fragment of SEQ ID NO:32 (exon 6), SEQ ID NO:33 (exon 7), a fragment of SEQ ID NO:33 (exon 7), SEQ ID NO:34 (exon 8), a fragment of SEQ ID NO:34 (exon 8), SEQ ID NO:35 (exon 9), a fragment of SEQ ID NO:35 (exon 9), SEQ ID NO:36 (exon 10), a fragment of SEQ ID NO:36 (exon 10), SEQ ID NO:37 (exon 11), and a fragment of SEQ ID NO:37 (exon 11).

[0068] Various embodiments are directed to NtHMA RNAi expression vectors comprising a NtHMA RNAi construct that comprises: SEQ ID NO:38 (“sense sequence/fragment”), the second sequence comprises SEQ ID NO:39 (“spacer sequence/fragment”) and the third sequence comprises SEQ ID NO:40 (“anti-sense sequence/fragment”).

[0069] Various embodiments are directed to NtHMA RNAi expression vectors comprising a NtHMA RNAi construct that comprises: SEQ ID NO:42 (“sense sequence/fragment”), the second sequence comprises SEQ ID NO:43 (“spacer sequence/fragment”), and the third sequence comprises SEQ ID NO:44 (“anti-sense sequence/fragment”).

[0070] Alternatively, the disclosed sequences can be utilized for constructing various NtHMA polynucleotides that do not form hairpin structures. For example, a NtHMA long double-stranded RNA can be formed by (1) transcribing a first strand of the NtHMA cDNA by operably-linking to a first promoter, and (2) transcribing the reverse complementary sequence of the first strand of the NtHMA cDNA fragment by operably-linking to a second promoter. Each strand of the NtHMA polynucleotide can be transcribed from the same expression vector, or from different expression vectors. The NtHMA RNA duplex having RNA interference activity can be enzymatically converted to siRNAs to reduce NtHMA RNA levels.

C. EXPRESSION VECTORS FOR REDUCING NtHMA GENE EXPRESSION BY CO-SUPPRESSION

[0071] Various compositions and methods are provided for reducing the endogenous expression levels for members of the NtHMA gene family by promoting co-suppression of NtHMA gene expression. The phenomenon of co-suppression occurs as a result of introducing multiple copies of a transgene into a plant cell host. Integration of multiple copies of a transgene can result in reduced expression of the transgene and the targeted endogenous gene. The degree of co-suppression is dependent on the degree of sequence identity between the transgene and the targeted endogenous gene. The silencing of both the endogenous gene and the transgene can occur by extensive methylation of the silenced loci (i.e., the endogenous promoter and endogenous gene of interest) that can preclude transcription. Alternatively, in some cases, co-suppression of the endogenous gene and the transgene can occur by post transcriptional gene silencing (“PTGS”), in which transcripts can be produced but enhanced rates of degradation preclude accumulation of transcripts. The mechanism for co-suppression by PTGS is thought to resemble RNA interference, in that RNA seems to be both an important initiator and a target in these processes, and may be

mediated at least in part by the same molecular machinery, possibly through RNA-guided degradation of mRNAs.

[0072] Co-suppression of members of the NtHMA gene family can be achieved by integrating multiple copies of the NtHMA cDNA or fragments thereof, as transgenes, into the genome of a plant of interest. The host plant can be transformed with an expression vector comprising a promoter operably-linked to NtHMA cDNA or fragments thereof. Various embodiments are directed to expression vectors for promoting co-suppression of endogenous genes of the NtHMA family comprising: a promoter operably-linked to NtHMA cDNA identified as Clone P6663 (SEQ ID NO:3) or a fragment thereof, or NtHMA cDNA identified as Clone P6643 (SEQ ID NO:47) or a fragment thereof. Various embodiments are directed to expression vectors for promoting co-suppression of endogenous genes of the NtHMA family comprising: a promoter operably-linked to NtHMA cDNA, or a fragment thereof, having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:3 or SEQ ID NO:47.

[0073] Various embodiments are directed to methods for reducing the expression level of endogenous genes of the NtHMA family by integrating multiple copies of NtHMA cDNA or a fragment thereof into a plant genome, comprising: transforming a plant cell host with an expression vector that comprises a promoter operably-linked to SEQ ID NO:3, or a fragment thereof; or SEQ ID NO:47, or a fragment thereof. Various embodiments are directed to methods for reducing the expression level of endogenous genes of the NtHMA family by integrating multiple copies of NtHMA cDNA, or a fragment thereof, into a plant genome, comprising: transforming a plant cell host with an expression vector that comprises a promoter operably-linked to NtHMA cDNA, or a fragment thereof, having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:3 or SEQ ID NO:47.

D. EXPRESSION VECTORS FOR REDUCING NtHMA GENE EXPRESSION BY INHIBITION OF TRANSLATION BY ANTI-SENSE AGENTS

[0074] Various compositions and methods are provided for reducing the endogenous expression level of the NtHMA gene family by inhibiting the translation of NtHMA mRNA. A host plant cell can be transformed with an expression vector comprising: a promoter operably-linked to NtHMA cDNA or a fragment thereof, positioned in anti-sense orientation with respect to the promoter to enable the expression of RNA polynucleotides having a sequence complementary to a portion of NtHMA mRNA. Various expression vectors for inhibiting the translation of HMA mRNA comprise: a promoter operably-linked to NtHMA cDNA identified as Clone P6663 (SEQ ID NO:3) or a fragment thereof; or NtHMA cDNA identified as Clone P6643 (SEQ ID NO:47) or a fragment thereof, in which the NtHMA cDNA, or the fragment thereof, is positioned in anti-sense orientation with respect to the promoter. Various expression vectors for inhibiting the translation of HMA mRNA comprise: a promoter operably-linked to a NtHMA cDNA, or a fragment thereof, having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:3 or SEQ ID NO:47, in which the NtHMA cDNA, or the fragment thereof, is positioned in anti-sense orientation with respect to the promoter. The lengths of anti-sense NtHMA RNA polynucleotides can vary, including 15-20 nucleotides, 20-30 nucleotides, 30-50

nucleotides, 50-75 nucleotides, 75-100 nucleotides, 100-150 nucleotides, 150-200 nucleotides, and 200-300 nucleotides.

[0075] Various embodiments are directed to methods for reducing the expression level of endogenous genes of the NtHMA family by inhibiting NtHMA mRNA translation, comprising: transforming a plant cell host with an expression vector that comprises a promoter operably-linked to NtHMA cDNA identified as Clone P6663 (SEQ ID NO:3) or a fragment thereof; or NtHMA cDNA identified as Clone P6643 (SEQ ID NO:47) or a fragment thereof, in which the NtHMA cDNA, or the fragment thereof, is positioned in anti-sense orientation with respect to the promoter. Various embodiments are directed to methods for reducing the expression level of endogenous genes of the NtHMA family by inhibiting NtHMA mRNA translation, comprising: transforming a plant cell host with an expression vector that comprises a promoter operably-linked to a NtHMA cDNA, or a fragment thereof, having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:3 or SEQ ID NO:47, in which the NtHMA cDNA, or the fragment thereof, is positioned in anti-sense orientation with respect to the promoter.

E. OTHER COMPOSITIONS AND METHODS FOR REDUCING NtHMA GENE EXPRESSION

[0076] Methods for obtaining conservative variants and more divergent variants of NtHMA polynucleotides and polypeptides are known to persons skilled in the art. Any plant of interest can be genetically modified by various methods known to induce mutagenesis, including site-directed mutagenesis, oligonucleotide-directed mutagenesis, chemically-induced mutagenesis, irradiation-induced mutagenesis, and other equivalent methods. For example, site-directed mutagenesis is described in, e.g., Smith (1985) "In vitro mutagenesis," *Ann. Rev. Genet.* 19:423-462, and references therein, such as Botstein & Shortle (1985) "Strategies and Applications of in vitro Mutagenesis," *Science* 229:1193-1201; and in Carter (1986) "Site-directed mutagenesis," *Biochem. J.* 237:1-7. Oligonucleotide-directed mutagenesis is described in, e.g., Zoller & Smith (1982) "Oligonucleotide-directed Mutagenesis using M13-derived Vectors: an Efficient and General Procedure for the Production of Point mutations in any DNA Fragment," *Nucleic Acids Res.* 10:6487-6500. Mutagenesis utilizing modified bases is described in, e.g., Kunkel (1985) "Rapid and Efficient Site-specific Mutagenesis without Phenotypic Selection," *Proc. Natl. Acad. Sci. USA* 82:488-492, and in Taylor et al. (1985) "The Rapid Generation of Oligonucleotide-directed Mutations at High Frequency using Phosphorothioate-modified DNA," *Nucl. Acids Res.* 13: 8765-8787. Mutagenesis utilizing gapped duplex DNA is described in, e.g., Kramer et al. (1984) "The Gapped Duplex DNA Approach to Oligonucleotide-directed Mutation Construction," *Nucl. Acids Res.* 12: 9441-9460. Point-mismatch mutagenesis is described in, e.g., Kramer et al. (1984) "Point Mismatch Repair," *Cell* 38:879-887. Double-strand break mutagenesis is described in, e.g., Mandecki (1986) "Oligonucleotide-directed Double-strand Break Repair in Plasmids of *Escherichia coli*: A Method for Site-specific Mutagenesis," *Proc. Natl. Acad. Sci. USA*, 83:7177-7181, and in Arnold (1993) "Protein Engineering for Unusual Environments," *Current Opinion in Biotechnology* 4:450-455. Mutagenesis utilizing repair-deficient host strains is described in, e.g., Carter et al. (1985) "Improved Oligonucleotide Site-directed Mutagenesis using

M13 Vectors,” Nucl. Acids Res. 13: 4431-4443. Mutagenesis by total gene synthesis is described in, e.g., Nambiar et al. (1984) “Total Synthesis and Cloning of a Gene Coding for the Ribonuclease S Protein,” Science 223: 1299-1301. DNA shuffling is described in, e.g., Stemmer (1994) “Rapid Evolution of a Protein in vitro by DNA Shuffling,” Nature 370: 389-391, and in Stemmer (1994) “DNA shuffling by random fragmentation and reassembly: In Vitro Recombination for Molecular Evolution,” Proc. Natl. Acad. Sci. USA 91:10747-10751.

[0077] Alternatively, NtHMA genes can be targeted for inactivation by introducing transposons (and IS elements) into the genomes of plants of interest. These mobile genetic elements can be introduced by sexual cross-fertilization and insertion mutants can be screened for loss in NtHMA activity, such as reduced Cd transport. The disrupted NtHMA gene in a parent plant can be introduced into other plants by crossing the parent plant with plant not subjected to transposon-induced mutagenesis by, e.g., sexual cross-fertilization. Any standard breeding techniques known to persons skilled in the art can be utilized. In one embodiment, one or more NtHMA-related genes can be inactivated by the insertion of one or more transposons. Mutations can result in homozygous disruption of one or more NtHMA genes, in heterozygous disruption of one or more NtHMA genes, or a combination of both homozygous and heterozygous disruptions if more than one NtHMA gene is disrupted. Suitable transposable elements can be selected from two broad classes, designated as Class I and Class II. Suitable Class I transposable elements include retrotransposons, retroposons, and SINE-like elements. Such methods are known to persons skilled in the art as described in Kumar and Bennetzen (1999), Plant Retrotransposons in Annual Review of Genetics 33:479.

[0078] Alternatively, NtHMA genes can be targeted for inactivation by a method referred to as Targeting Induced Local Lesions IN Genomics (“TILLING”), which combines high-density point mutations with rapid sensitive detection of mutations. Typically, plant seeds are exposed to mutagens, such as ethylmethanesulfonate (EMS) or EMS alkylates guanine, which typically leads to mispairing. Suitable agents and methods are known to persons skilled in the art as described in McCallum et al., (2000), “Targeting Induced Local Lesions IN Genomics (TILLING) for Plant Functional Genomics,” Plant Physiology 123:439-442; McCallum et al., (2000) “Targeted screening for induced mutations,” Nature Biotechnology 18:455-457; and Colbert et al., (2001) “High-Throughput Screening for Induced Point Mutations,” Plant Physiology 126:480-484.

[0079] Alternatively, NtHMA genes can be targeted for inactivation by introducing ribozymes derived from a number of small circular RNAs that are capable of self-cleavage and replication in plants. These RNAs can replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples of suitable RNAs include those derived from avocado sunblotch viroid and satellite RNAs derived from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus, and subterranean clover mottle virus. Various target RNA-specific ribozymes are known to persons skilled in the art as described in Haseloff et al. (1988) Nature, 334:585-591.

III. Transgenic Plants, Cell Lines, and Seeds Comprising NtHMA RNAi Polynucleotides and Related Methods

[0080] Various embodiments are directed to transgenic plants genetically modified to reduce the NtHMA gene

expression level by various methods that can be utilized for silencing NtHMA gene expression, and thereby, producing transgenic plants in which the expression level of NtHMA transporters can be reduced within plant tissues of interest. Rates of heavy metal transport and distribution patterns of heavy metal transport, in particular, cadmium transport, can be altered in transgenic plants produced according to the disclosed methods and compositions. Plants suitable for genetic modification include monocots and dicots.

[0081] Various embodiments are directed to transgenic tobacco plants genetically modified to reduce the NtHMA gene expression level by various methods that can be utilized for down-regulating NtHMA gene expression, and thereby, producing transgenic tobacco plants in which the expression level of NtHMA transporters can be reduced within plant tissues of interest. Various expression vectors have been provided to produce transgenic lines of tobacco of any variety exhibiting reduced levels of NtHMA gene expression. The disclosed compositions and methods can be applied to any plant species of interest, including plants of the genus *Nicotiana*, various species of *Nicotiana*, including *N. rustica* and *N. tabacum* (e.g., LA B21, LN KY171, TI 1406, Basma, Galpao, Perique, Beinhart 1000-1, and Petico). Other species include *N. acaulis*, *N. acuminata*, *N. acuminata* var. *multiflora*, *N. africana*, *N. alata*, *N. amplexicaulis*, *N. arentsii*, *N. attenuata*, *N. benavidesii*, *N. benthamiana*, *N. bigelovii*, *N. bonariensis*, *N. cavicola*, *N. clevelandii*, *N. cordifolia*, *N. corymbosa*, *N. debneyi*, *N. excelsior*, *N. forgetiana*, *N. fragrans*, *N. glauca*, *N. glutinosa*, *N. goodspeedii*, *N. gossei*, *N. hybrid*, *N. ingulba*, *N. kawakamii*, *N. knightiana*, *N. langsdorffii*, *N. linearis*, *N. longiflora*, *N. maritima*, *N. megalosiphon*, *N. miersii*, *N. noctiflora*, *N. nudicaulis*, *N. obtusifolia*, *N. occidentalis*, *N. occidentalis* subsp. *hesperis*, *N. otophora*, *N. paniculata*, *N. pauciflora*, *N. petunioides*, *N. plumbaginifolia*, *N. quadrivalvis*, *N. raimondii*, *N. repanda*, *N. rosulata*, *N. rosulata* subsp. *ingulba*, *N. rotundifolia*, *N. setchellii*, *N. simulans*, *N. solanifolia*, *N. spegazzinii*, *N. stocktonii*, *N. suaveolens*, *N. sylvestris*, *N. thyrsoflora*, *N. tomentosa*, *N. tomentosiformis*, *N. trigonophylla*, *N. umbratica*, *N. undulata*, *N. velutina*, *N. wigandiioides*, and *N. x sanderae*. Suitable plants for transformation include any plant tissue capable of transformation by various methods of transforming plants known by persons skilled in the art, including electroporation, micro-projectile bombardment, and *Agrobacterium*-mediated transfer as described, for example, in U.S. Pat. No. 4,459,355 that discloses a method for transforming susceptible plants, including dicots, with an *Agrobacterium* strain containing a Ti plasmid; U.S. Pat. No. 4,795,855 that discloses transformation of woody plants with an *Agrobacterium* vector; U.S. Pat. No. 4,940,838 that discloses a binary *Agrobacterium* vector; U.S. Pat. No. 4,945,050; and U.S. Pat. No. 5,015,580.

[0082] Various embodiments are directed to transgenic tobacco plants genetically modified to exogenously express a RNAi construct encoding NtHMA RNAi polynucleotides that facilitate the degradation of NtHMA RNA transcripts, and consequently, that reduce the number of RNA transcripts available for translation into NtHMA transporters. Various embodiments are directed to transgenic plants comprising an expression vector that enable the expression of NtHMA polynucleotides produced according to the disclosed methods. Various embodiments are directed to cell lines derived from transgenic plants produced according to the disclosed meth-

ods. Various embodiments are directed to transgenic seeds derived from transgenic plants produced according to the disclosed methods.

[0083] Various embodiments are directed to methods for reducing the NtHMA gene expression levels in plants, the method comprising reducing the expression level of a NtHMA gene, which can be accomplished by various methods known to persons skilled in the art. As examples, this includes: (1) RNA interference method for reducing steady-state level of endogenous NtHMA RNA variants available for translation by expression of NtHMA RNAi polynucleotides; (2) co-suppression method for reducing transcription of NtHMA gene(s) by integrating multiple copies of the NtHMA cDNA or fragments thereof, as transgenes, into a plant genome; (3) anti-sense method for reducing the NtHMA translation by the expression of anti-sense polynucleotides that can target NtHMA RNA; and (4) various methods for inducing mutagenesis.

[0084] Various embodiments are directed to transgenic tobacco plants genetically modified to reduce the NtHMA gene expression level by various methods, known to persons skilled in the art, and further modified either to reduce the expression of a second endogenous gene of interest (i.e., not NtHMA-related) or to enhance the expression of an exogenous gene of interest (i.e., not NtHMA-related). For example, the down-regulation of a second endogenous gene of interest encoding an enzyme involved in the biosynthesis of alkylloids may be desirable. In other situations, the enhancement in the expression level of a transgene encoding a recombinant protein of interest, such as a human hormone for therapeutic use, may be desirable. Persons skilled in the art are capable of producing various transgenic plants that can be modified, for example, to exogenously express NtHMA RNAi polynucleotides and at least one recombinant gene product of interest, such as a recombinant human growth factor or RNAi polynucleotides that can target a second gene of interest not related to the NtHMA family.

[0085] Producing transgenic plants according to the disclosed methods provides a number of advantages. Transgenic plants, including transgenic tobacco plants, can be grown in soils containing variable Cd concentrations, or in soils containing less than desirable Cd concentrations. These transgenic plants and derivative seeds can provide more options for cultivating them in a broader range of soil environments, which may increase the amount of cultivatable soils available to practitioners (e.g., farmers). Furthermore, these transgenic plants, exhibiting reduced Cd content, compared to non-transgenic counterparts can be consumed directly as edible products. The consumption of edible portions of these transgenic plants can be a healthier option compared to the consumption of non-transgenic counterparts. Suitable plants that can be genetically modified according to the disclosed methods, include plants cultivatable for agricultural use, including rice, corn, squash, soybeans, lettuce, potatoes, beans, wheat, barley, carrots, etc. The % Cd reduction in these transgenic plants, including the leaf lamina portion, can be approximately at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99%, when compared to non-transgenic counterparts. The Cd content of these transgenic plants, including the leaf lamina portion, is a value from a range from about 0.01 to about 0.05 ppm, from about 0.01

to about 0.1 ppm, from about 0.01 to about 0.5 ppm, from about 0.01 to about 1.0 ppm, and from about 0.01 to about 5 ppm.

IV. Consumable Products Incorporating Tobacco Leaves Genetically Modified to Contain Reduced Cd Content

[0086] Various embodiments provide transgenic plants, in which the expression level of members of the NtHMA gene family is substantially reduced to curtail or impede Cd transport into the leaf lamina. The leaf lamina derived from transgenic tobacco plants, produced according to the disclosed methods, can be incorporated into various consumable products containing Cd at a level substantially below that of consumable products made by incorporating tobacco leaves derived from plants of the same genotype that were grown under identical conditions, but not genetically modified with respect to the reduced expression level of members of the NtHMA gene family ("non-transgenic counterparts").

[0087] In some embodiments, these transgenic plants exhibiting reduced Cd content compared to non-transgenic counterparts can be incorporated into consumable products, including various smokable articles, such as cigars, cigarettes, and smokeless tobacco products (i.e., non-combustible). Smokable articles and smokeless tobacco products, produced by incorporating tobacco leaves derived from tobacco plants genetically modified to contain reduced Cd levels according to the disclosed methods, can provide healthier options compared to non-transgenic counterparts.

[0088] Smokeless tobacco products incorporating tobacco plants genetically modified according to the disclosed methods can be manufactured in any format suitable for comfort in a consumer's oral cavity. Smokeless tobacco products contain tobacco in any form, including as dried particles, shreds, granules, powders, or a slurry (i.e., tobacco extract), deposited on, mixed in, surrounded by, or otherwise combined with other ingredients in any format, such as flakes, films, tabs, foams, or beads. Smokeless tobacco products may be wrapped with a material, which may be edible (i.e., orally disintegrable) or nonedible. Liquid contents of smokeless tobacco products can be enclosed in a form, such as beads, to preclude interaction with a water-soluble wrapper. The wrapper may be shaped as a pouch to partially or completely enclose tobacco-incorporating compositions, or to function as an adhesive to hold together a plurality of tabs, beads, or flakes of tobacco. A wrapper may also enclose a moldable tobacco composition that conforms to the shape of a consumer's mouth. An orally disintegrable wrapper may enclose smokeless tobacco, e.g., as dry snuff or soluble tobacco, and may be formed on continuous thermoforming or horizontal form/fill/seal equipment or other suitable packaging equipment using edible films (which may or may not contain tobacco). Exemplary materials for constructing a wrapper include film compositions comprising HPMC, CMC, pectin, alginates, pullulan, and other commercially viable, edible film-forming polymers. Other wrapping materials may include pre-formed capsules produced from gelatin, HPMC, starch/carrageenan, or other commercially available materials. Such wrapping materials may include tobacco as an ingredient. Wrappers that are not orally disintegrable may be composed of woven or nonwoven fabrics, of coated or uncoated paper, or of perforated or otherwise porous plastic films. Wrappers may incorporate flavoring and/or coloring agents. Smokeless products can be assembled together with a

wrapper utilizing any method known to persons skilled in the art of commercial packaging, including methods such as blister packing and stick-packing, in which a small package can be formed by a vertical form/fill/seal packaging machine.

[0089] The % Cd reduction in these smokable articles and smokeless products, produced by incorporating tobacco leaves derived from tobacco plants genetically modified to contain reduced Cd levels, is a value of at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, and 100%, when compared to consumable products derived from non-transgenic counterparts. In some embodiments, the Cd content of these smokable articles and smokeless products, produced by incorporating tobacco leaves derived from tobacco plants genetically modified to contain reduced Cd levels, is a value from a range from about 0.01 to about 0.05 ppm, from about 0.01 to about 0.1 ppm, from about 0.01 to about 0.5 ppm, from about 0.01 to about 1.0 ppm, and from about 0.01 to about 5 ppm.

[0090] The degree of Cd accumulation in plants can be substantially variable depending on several parameters attributed to the complexity of the genotype and the growth environment. For example, Cd concentrations in field-grown tobacco leaves can be extremely variable depending on factors such as the agro-climate, soil parameters, and cultivars. Furthermore, the relative Cd distribution patterns within different portions of a tobacco plant can vary according to the species, the organ/tissue, and growth conditions (i.e., field-grown vs. hydroponically-grown). On average, the Cd concentrations measured in field-grown tobacco leaves (including midribs and veins) can be in the range from approximately 0.5 to 5 ppm (parts per million, or ug/g of dry weight of tobacco leaves). However, many published Cd levels typically do not define the tobacco maturity stage, the tobacco variety, or the particular leaf portions (i.e., removal from leaf stalk position) harvested for analysis. In some varieties, the lower leaves may accumulate higher Cd levels than the medium and upper leaves. At the intracellular level, Cd can be found in various cell components of a plant cell, including the cell wall, cytoplasm, chloroplast, nucleus, and vacuoles.

[0091] Furthermore, Cd content measured in tobacco leaves can vary substantially depending on the Cd levels in the soil environment where the tobacco plants were grown. The leaves of tobacco grown in Cd-contaminated areas can accumulate Cd from about 35 ppm or higher, compared to the leaves of genetically identical counterparts grown in non-contaminated areas, which can accumulate Cd at a range from approximately 0.4 to approximately 8 ppm. The vacuoles within the leaves of plants grown in Cd-contaminated areas can accumulate very high Cd concentrations. Methods for applying the disclosed compositions to be suitable for a given plant species of interest are known to persons skilled in the art.

EXAMPLES

Example 1

Cloning and Exon Mapping of a Full-Length *Nicotiana* NtHMA Genomic Clone

[0092] Two partial genomic clones representing different portions of an endogenous NtHMA gene were independently identified, referred to as “CHO_OF96xf01.ab1” and “CHO_OF261xo09c1.ab1.” Based on sequence information obtained from the partial genomic clones, a full-length

genomic clone (_HO-18-2) and 4 full-length NtHMA cDNAs were subsequently identified, including clone P6663 (SEQ ID NO:3) and clone P6643 (SEQ ID NO:47). The exon and intron subregions of full-length genomic clone (_HO-18-2) (17,921 bp) were mapped. As shown in FIG. 1A, the full-length, endogenous NtHMA gene cloned from *Nicotiana* comprises 11 exons consisting of 3392 nucleotides in total.

Example 2

Construction of NtHMA RNAi Expression Vector PBI121-NtHMA (660-915) Encoding RNAi Polynucleotides

[0093] FIG. 1B provides a list of nucleotide positions mapped to each exon within the isolated NtHMA genomic clone (SEQ ID NO: 1) (“Table 1”). The partial genomic clone CHO_OF96xf01.ab1 includes a part of intron 4, exon 4, intron 5, exon 5, intron 6, and a part of exon 6, as shown in FIG. 1A, and listed under Table 1 of FIG. 1B. The partial genomic clone CHO_OF261xo09c1 includes a part of intron 7, exon 7, intron 8, exon 8, and a part of exon 9, as shown in FIG. 1A. To produce transgenic plants that can stably produce recombinant NtHMA RNAi polynucleotides of interest that can facilitate the degradation of endogenous RNA transcripts encoding NtHMA polypeptides, two sets of NtHMA RNAi expression vectors, the PBI121-NtHMA (660-915) RNAi expression vector as further described below, and the PBI121-NtHMA (1382-1584) RNAi expression vector as further described in Example 3.

[0094] FIG. 2 illustrates an exemplary subcloning strategy for constructing a NtHMA RNAi expression vector that enables the constitutive expression of NtHMA RNAi polynucleotides of interest. Based on exon mapping and sequence analysis of genomic clone CHO_OF96xf01.ab1, RNAi constructs were designed.

[0095] FIG. 3A shows an exemplary RNAi sequence, NtHMA (660-915), for producing NtHMA RNAi polynucleotides of interest. In FIG. 3A, NtHMA RNAi RNAi construct comprises a sense fragment (272 bp) (SEQ ID NO:38) composed of exon 4 (272 bp), which is positioned upstream and operably-linked to a spacer fragment (80 bp) (SEQ ID NO:39) composed of intron 5, which is positioned upstream and operably-linked to a reverse complementary fragment (272 bp) (SEQ ID NO:40) composed of exon 4 positioned in anti-sense orientation. RNAi constructs encoding NtHMA RNAi polynucleotides of interest were inserted into the PBKCMV cloning vector, and were placed downstream and operably-linked to a cytomegalovirus (CMV) promoter. XbaI and HindIII sites were incorporated into the 5' and 3' ends of the 352 bp NtHMA sense fragment, which included the 80 bp intron fragment by utilizing PCR primers modified to incorporate these restriction enzyme sites (PMG783F: ATTCTAGACTGCTGCTATGTCATCACTGG and PMG783R: ATAAGCTTAGCCTGAAGAATTGAGCAAA). Similarly, SpeI and SacI sites were incorporated into the 5' and 3' ends of the corresponding NtHMA reverse complementary fragment by utilizing PCR primers (PMG 785F: ATGAGCTCTGGT-TATGTAGGCTACTGCTGCT and PMG 786R: ATACTAG-TATTTGTAGTGCCAGCCCAGA) to produce the PBKCMV-NtHMA RNAi plasmid. The PBI121-NtHMA RNAi expression vectors were constructed by (a) excising the β -glucuronidase ORF from the binary expression vector (“pBI121” from CLONTECH), and (b) substituting the NtHMA RNAi construct, excised from the PBKCMV-

NtHMA RNAi plasmid, into XbaI/SacI sites of the PBI121 plasmid in place of the removed β -glucuronidase ORF. The PBI121-NtHMA RNAi expression vectors comprise: (i) 352 bp XbaI-HindIII NtHMA sense fragment that includes (ii) 80 bp intron fragment, operably-linked to the (iii) 272 bp Spell-SacI NtHMA reverse complementary fragment.

Example 3

Construction of NtHMA RNAi Expression Vector PBI121-NtHMA (1382-1584) Encoding RNAi Polynucleotides

[0096] FIG. 4A shows an exemplary RNAi sequence, NtHMA (1382-1584), for producing NtHMA RNAi polynucleotides of interest. Based on exon mapping and sequence analysis of genomic clone CHO_OF261xo09c1, a RNAi construct was designed that includes a sense fragment (191 bp) (SEQ ID NO:42) comprising sequences of exon 7, which is positioned upstream and operably-linked to a spacer DNA fragment (139 bp) (SEQ ID NO:43) comprising sequences of intron 8, which is positioned upstream and operably-linked to a reverse complementary fragment (196 bp) (SEQ ID NO:44) comprising sequences of exon 7 positioned in anti-sense orientation. These RNAi constructs encoding NtHMA RNAi polynucleotides of interest were inserted into the PBKCMV cloning vector, and were placed downstream and operably-linked to a cytomegalovirus (CMV) promoter. XbaI and HindIII sites were incorporated into the 5' and 3' ends of the 330 bp NtHMA sense fragment, which included the 139 bp intron fragment by utilizing PCR primers modified to incorporate these restriction enzyme sites (PMG754F: ATTCTAGATGAGAGCAAGTCAGGTCATCC and PMG754R: ATAAGCTTTTCAAACATCCACCGCATTA). Similarly, PstI and SacI sites were incorporated into the 5' and 3' ends of the corresponding NtHMA reverse complementary fragment by utilizing PCR primers PMG757F: ATGAGCTCGCATTGAGAGCAAGTCAGGTC and PM 757R: ATCTGCAGCCTGTGGTACATCCAGCTCTT) to produce the PBKCMV-NtHMA RNAi expression vector.

[0097] The PBI121-NtHMA RNAi expression vectors were constructed by (a) excising the β -glucuronidase ORF from the binary expression vector ("pBI121" from CLONTECH), and (b) substituting the NtHMA RNAi construct, excised from the PBKCMV-NtHMA RNAi plasmid, into XbaI/SacI sites of the PBI121 plasmid in place of the removed β -glucuronidase ORF. The PBI121-NtHMA RNAi expression vectors comprise: (i) 330 bp XbaI-HindIII NtHMA sense fragment that includes (ii) 139 bp intron fragment, operably-linked to the (iii) 196 bp Spell-SacI NtHMA reverse complementary fragment. The PBI121-NtHMA RNAi expression vectors, such as those described in Examples 2 and 3, can be introduced into any host plant cell of interest by various methods known to persons skilled in the art.

Example 4

Transformation of Burley (TN90), Flue-Cured (K326), and Dark (VA359) Tobacco Varieties with NtHMA RNAi Expression Vectors

[0098] Tobacco seeds from three different varieties, Burley (TN90), Flue-cured (K326), and Dark (VA359), were sterilized and germinated in a petridish containing MS basal media supplemented with 5 ml/L plant preservative mixture (PPM).

Seedlings, at approximately 7 to 10 days post-germination, were selected for transformation with various NtHMA RNAi expression vectors. A single colony of *Agrobacterium tumefaciens* LBA4404 was inoculated into a liquid LB medium containing 50 mg l⁻¹ kanamycin (kanamycin mono sulphate), and were incubated for 48 h at 28° C. with reciprocal shaking (150 cycles min⁻¹). Cultured bacterial cells were collected by centrifugation (6000×g, 10 min), and were suspended to a final density of 0.4-0.7 OD₆₀₀, with 20 ml liquid MS medium containing 20 g⁻¹ sucrose. The 7-10 day seedling explants were immersed into a bacterial suspension for 5 mins, and were blotted on sterile filter papers. Fifty explants were placed onto 40 ml aliquots of REG agar medium (MS basal medium supplemented with 0.1 mg l⁻¹ NAA and 1 mg l⁻¹ BAP) in 100 mm×20 mm petri dishes. The explants were co-cultivated with *Agrobacterium* at 25° C. After 3 days of co-cultivation, the explants were washed and transferred to RCPK medium (REG medium with 100 mg⁻¹ kanamycin, 500 mg l⁻¹ carbenicillin, and 5 ml PPM) to select for transformants. The explants were subcultured every 2 weeks. After 8-12 weeks of growth under selective conditions, the surviving plants, representing transformants that have integrated the NtHMA RNAi expression constructs into their genomes, were transferred to a rooting medium (MS basal medium supplemented with 100 mg l⁻¹-Kanamycin). Rooted plants were transferred to pots to promote further growth.

Example 5

Cd Reduction in Leaf Lamina of First Generation Transgenics Genetically Modified to Express NtHMA RNAi Polynucleotides

[0099] To determine the effect of NtHMA RNAi polynucleotide expression on Cd transport from the root to aerial portions of transgenic plants, the Cd levels were determined for several transgenic lines that have been genetically modified to express either the NtHMA (660-915) or the (1382-1584) RNAi polynucleotides.

[0100] Approximately 40 independent transgenic plants, representing three tobacco varieties, were transformed with various PBI121-NtHMA RNAi expression vectors. Initially, transformants were grown in floating trays containing Hoaglands medium for 4 weeks. PCR positive plants for NPT II were selected and potted in 10" pots with a hydroponic system containing Hoaglands medium containing 5 μ M CdCl₂. After 4-8 weeks, two middle leaves samples were harvested and freeze-dried for metal analysis, or were frozen in liquid nitrogen for gene expression analysis. Approximately 500 mg of tobacco was weighed and digested in 10 ml of concentrated HNO₃ by microwave-accelerated, reaction system 5 digestion system (CEM corporation, Mathews, N.C.). Heavy metal concentrations were analyzed utilizing inductively coupled plasma-mass spectrophotometry ("ICP-MS," Agilent 7500A; Agilent Technologies, Palo Alto, Calif.). As non-transgenic tobacco control, a sample consisting of polish-certified, Virginia tobacco leaves, CTA-VTL-2, was prepared under comparable conditions (Dybczynski et al., 1997).

[0101] FIGS. 3B-3D show Cd reduction in leaf lamina of multiple first generation (T0) transgenic lines, representing three varieties, that have been genetically modified to express NtHMA RNAi polynucleotides (660-915).

[0102] FIGS. 4B-4D show Cd reduction in leaf lamina of multiple first generation (T0) transgenic lines, representing

three varieties, that have been genetically modified to express NtHMA RNAi polynucleotides (1382-1584).

Example 6

Reduction in NtHMA RNA Transcripts in Transgenic Tobacco Leaf by the Expression of NtHMA RNAi Polynucleotides

[0103] To determine the effect of NtHMA RNAi polynucleotide expression on the steady-state levels of endogenous NtHMA RNA transcripts, the relative change in NtHMA RNA transcripts was measured by isolating total cellular RNA from leaf lamina portions of various transgenic lines, representing three tobacco varieties.

[0104] Total RNA was isolated from middle leaves of T0 plants using TRI® Reagent (Sigma-Aldrich, St. Louis, Mo.). To remove DNA impurities, purified RNA was treated with RNase-free DNase (TURBO DNA-free, Ambion, Austin Tex.). To synthesize the first cDNA strand, approximately 10 µg of total RNA was reverse transcribed utilizing the High Capacity cDNA Archive Kit (Applied Biosystems, Foster City, Calif.). To measure the level of NtHMA transcripts in the samples, a quantitative 2-step RT-PCR was performed according to the Taqman MGB probe-based chemistry. The RT mixture contained 4 µM dNTP mix, 1× random primers, 1×RT Buffer, 10 g cDNA, 50U Multiscribe Reverse transcriptase (Applied Biosystems), 2U Superase-In RNase Inhibitor (Ambion), and nuclease-free water. The PCR mixture contained 1× Taqman Universal PCR Master Mix (Applied Biosystems, Foster City, Calif.), 400 nM forward primer, 400 nM reverse primer, 250 nM Taqman MGB probe, 2 ng of cDNA, and nuclease-free water. RT-PCR was performed utilizing an ABI 7500 Real-Time System (Applied Biosystems, Foster City, Calif.) and under amplification conditions: 50° C. for 2 min.; 95° C. for 10 min.; 40 cycles of 95° C. for 15 sec.; and 60° C. for 1 min. For normalizing the measured NtHMA RNA transcript levels, the Glyceraldehyde-3-Phosphate Dehydrogenase (G3PDH) was selected as a control endogenous RNA transcript, whose expression level is not responsive to the sequence-specific RNA interference activity of the NtHMA RNAi polynucleotides under analysis. The fold change in NtHMA RNA transcript level caused by NtHMA-RNAi-polynucleotide expression was calculated by determining the ratio of (a)/(b), in which (a) represents the normalized value of NtHMA RNA transcript level determined for samples derived from transgenic plants transformed with a NtHMA RNAi expression vector, and (b) represents the normalized value of NtHMA RNA transcript level determined for samples derived from transgenic plants transformed with a control expression vector deficient in the NtHMA RNAi RNAi construct.

[0105] FIGS. 5A-C show normalized NtHMA RNA transcript levels in various first generation (T0) transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest, as determined by quantitative realtime PCR analysis of leaf lamina extracts. FIG. 5A shows that for multiple independently derived K326 transgenic lines, the RNA transcript levels were reduced by the RNA interference activity of NtHMA (660-915) RNAi polynucleotides. FIG. 5B shows that for multiple, independently derived TN90 transgenic lines, the RNA transcript levels were reduced by the RNA interference activity of NtHMA (660-915) RNAi polynucleotides. FIG. 5C shows that for multiple independently derived VA359 transgenic lines, the

RNA transcript levels were reduced by the RNA interference activity of NtHMA (660-915) RNAi polynucleotides. The reduction in NtHMA RNA transcript levels is consistent with the reduction in Cd content measured in the middle leaves for the same transgenic lines tested. “PBI121” represents an expression vector deficient in the RNAi construct encoding NtHMA (660-915) RNAi polynucleotides.

Example 7

Distribution of Cd and Zn in Transgenic Lines Genetically Modified to Express NtHMA RNAi Polynucleotides

[0106] To determine the effect of NtHMA (660-915) RNAi polynucleotide expression on the distribution of Cd and Zn within the leaf lamina and the root, the metal content of transgenic plants of three varieties were analyzed. Five transgenic lines of each variety, i.e., Flue-cured (K326), Burley (TN90), and Dark (VA359), were selected for exhibiting Cd content at the lowest range in the leaf lamina. The middle leaves and roots of these transgenic plants and control plants were harvested for metal analysis by ICP_MS. For 8 weeks, all plants were grown in Hoaglands medium supplemented with 5 µM CdCl₂ prior to harvesting.

[0107] Table 2 lists Cd and Zn levels measured in the leaf lamina and the root of several transgenic lines, representing three tobacco varieties, as provided below. In Table 2, the Cd distribution between the leaf lamina and the root were substantially modified by the expression of NtHMA (660-915) RNAi polynucleotides for all three varieties, Flue-cured (K326), Burley (TN90), and Dark (VA359). For the K326 transgenic lines, the % Cd reduction ranged from 97.16-98.54% when compared to Cd levels observed in K326 Control plants. For TN90 transgenic lines, the % Cd reduction ranged from 85.12-90.96% when compared to Cd levels observed in the TN90 Control. For VA359 transgenic lines, the % Cd reduction ranged from 93.24-99.07% when compared to Cd levels observed in the VA359 Control. The VA359 NtHMA-11 transgenic line exhibited the lowest Cd level (1.62 µg/g) and the highest % Cd reduction (99.07%), when compared against two NtHMA RNAi transgene-deficient control lines (“VA359 PBI121”) that exhibited Cd levels at 158.3-205.96 µg/g. Comparable root analysis of the transgenic lines showed, that a substantial amount of Cd can accumulate in the root, resulting in fold increase in root Cd levels ranging from 6.90-15.38, relative to the Cd levels observed in the respective controls.

[0108] In contrast to the significant Cd reduction in the leaf lamina of transgenic lines, the Zn content of the leaf lamina was not substantially reduced, although some reduction was observed in most transgenic lines, caused by the expression of NtHMA (660-915) RNAi polynucleotides. The Zn content within the root (last column of Table 3) increased in all transgenic lines, resulting in a 4-6 fold increase in the transgenic lines of the K326 and VA359 varieties, and a 3-5 fold increase in the TN90 variety.

[0109] FIG. 6 shows the distribution of Cd and Zn between the leaf lamina and the root of various first generation transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest, as presented in Table 2.

TABLE 2

Transgenic Variety	Leaf		Root	
	Cd $\mu\text{g/g}$	Zn $\mu\text{g/g}$	Cd $\mu\text{g/g}$	Zn $\mu\text{g/g}$
K326 06T458	7.09	22.2	703	201
K326 06T459	4.97	24.1	696	225
K326 06T473	3.7	34	929	215
K326 06T480	3.93	38.6	989	224
K326 06T482	2.55	36.3	520	126
K326 Control	174.7	36.3	64.3	35.7
TN90 06T428	26.3	48.6	626	184
TN90 06T430	16.08	37.2	684	213
TN90 06T444	15.98	28.1	738	234
TN90 06T445	20.72	32.6	618	186
TN90 06T455	17.87	24.4	582	157
TN90 PBI121	181.2	35.5	62.6	44.3
TN90 Control	172.4	32.3	72.9	46.6
VA359 06T493	7.59	23.1	543	148
VA359 06T498	1.62	26.2	706	175
VA359 06T506	5.72	28.8	351	109
VA359 06T542	7.03	27.1	738	136
VA359 06T543	11.78	29.3	547	106
VA359 PBI121	206	47.5	35.3	27.6
VA359 Control	158.5	32.6	37.6	26.2

Example 8

Cd Distribution in Various Tissues of Transgenic Lines Genetically Modified to Express NtHMA RNAi Polynucleotides

[0110] To determine the effect of NtHMA (1382-1584) RNAi polynucleotide expression on Cd distribution within various tissues (i.e., the bark, lamina, pith, and root), the metal content of several transgenic lines representing two varieties, Burley (TN90) and Flue-cured (K326), were analyzed. Fully matured transgenic plants and control plants were harvested for metal analysis by ICP_MS. For 8 weeks, all plants were grown in 5 μM CdCl₂ in Hoaglands medium prior to harvesting.

[0111] Table 3 lists Cd content in the bark, lamina, pith, and root tissues of several transgenic lines, as provided below. In Table 3, Cd levels were substantially reduced in the bark, lamina, and pith tissues of all transgenic lines tested when compared that of control plants. The "Control" represents non-transgenic plants. The "PBI121" represents transgenic plants transformed with an expression vector deficient in NtHMA RNAi RNAi construct. The extent of Cd reduction in the bark, pith, and leaf lamina of K326 transgenic lines was significantly greater than that observed in TN90 transgenic lines. The expression of RNAi (1382-1584) polynucleotides in K326 transgenic plants resulted in a 9-11 fold Cd reduction in the bark, a 6-13 fold Cd reduction in the pith, and a 31-32 fold Cd reduction in the leaf lamina. The expression of RNAi (1382-1584) polynucleotides in TN90 transgenic plants resulted in a 4-7 fold Cd reduction in the bark, a 5-8 fold Cd reduction in the pith, and a 6-20 fold Cd reduction in the leaf lamina. In contrast, more modest increases (5-6 fold) in Cd content in the root of these transgenic lines were observed when compared to that of control plants.

[0112] FIG. 7 shows Cd distribution among the bark, leaf lamina, pith, and the root of various first generation transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest, as presented in Table 3.

TABLE 3

Transgenic Seed Variety	Bark Cd	Lamina Cd	Pith Cd	Root Cd
TN90 06T619	7.36	31.1	4.67	557
TN90 06T658	3.76	8.89	2.89	727
TN90 Control	30.9	151	25.3	115
TN90 PBI121	23.1	201	20	124
K326 06T682	2.02	4.32	1.97	1020
K326 06T696	2.53	4.48	4.25	1030
K326 Control	19.5	133	25.3	145
K326 PBI121	25.5	143	26.2	253

Example 9

Cd Reduction in leaf lamina of Second Generation Transgenic Lines Genetically Modified to Express NtHMA RNAi Polynucleotides

[0113] To determine the effect of NtHMA (660-915) RNAi polynucleotide expression on Cd content in leaf lamina, the metal content of two (T1) transgenic lines of VA359 variety were grown in soil containing variable Cd concentrations for 4 weeks. Two transgenic lines, 06T498 and 06T506, selected as kanamycin positives were screened by PCR. Several 10" Pots filled with sand:soil mixture were saturated with either 0, 0.1, 0.5, or 5 μM CdCl₂. Three plants per treatment per transgenic line were grown for 4 weeks by adding Hoaglands medium to the saucer. Total number of leaves, leaf area index, leaf weight, stalk weight, and root weight were observed. Two middle leaves and root samples were freeze-dried and were subjected to heavy metal analysis.

[0114] FIG. 8 shows Cd distribution between the leaf lamina and the root of various second generation (T1) transgenic lines that have been genetically modified to express NtHMA RNAi polynucleotides of interest. In FIG. 8, the Cd content of the transgenic plants was consistently lower than that of control plants at all Cd concentrations tested (0, 0.1, 0.5, and 5 μM). A reduction in Cd content of the leaf lamina (2-4.7 fold) was observed in various transgenic lines tested. The Cd level for the line 06T498 was only ~20% of control plants at 5 μM CdCl₂. An increase in root Cd content (4-16 fold) was observed in various transgenic lines tested. The highest root Cd content (a 16 fold increase) was observed for line 06T498 at 5 μM CdCl₂. Thus, the reduced heavy metal content in the leaf lamina/shoots in transgenic lines, expressing NtHMA (660-915) RNAi polynucleotide, suggested that the translocation of a substantial amount of heavy metals from the root to the leaf lamina/shoots can be interrupted by RNAi interference. The results are consistent with Cd reduction observed in the leaf lamina of first generation transgenic lines, in that the second generation transgenic lines also demonstrated (a) reduced Cd levels in the leaf lamina, and (b) increased Cd in the roots. The transgenic lines did not demonstrate phenotypical differences in general appearance, growth, and development relative to that of control plants.

Example 10

NtHMA Polynucleotides

[0115] A NtHMA polynucleotide will generally contain phosphodiester bonds, although in some cases, nucleic acid analogs are included that may have alternate backbones, comprising, e.g., phosphoramidate, phosphorothioate, phospho-

rodithioate, or O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press); and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, and non-ribose backbones. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, e.g. to increase the stability and half-life of such molecules in physiological environments or as probes on a biochip. Mixtures of naturally occurring nucleic acids and analogs can be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

[0116] A variety of references disclose such nucleic acid analogs, including, for example, phosphoramidate (Beaucage et al., *Tetrahedron* 49(10): 1925 (1993) and references therein; Letsinger, *J. Org. Chem.* 35:3800 (1970); Sprinzl et al., *Eur. J. Biochem.* 81:579 (1977); Letsinger et al., *Nucl. Acids Res.* 14:3487 (1986); Sawai et al., *Chem. Lett.* 805 (1984); Letsinger et al., *J. Am. Chem. Soc.* 110:4470 (1988); and Pauwels et al., *Chemica Scripta* 26:141 (1986)), phosphorothioate (Mag et al., *Nucleic Acids Res.* 19:1437 (1991); and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al., *J. Am. Chem. Soc.* 111:2321 (1989), O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, *J. Am. Chem. Soc.* 114:1895 (1992); Meier et al., *Chem. Int. Ed. Engl.* 31:1008 (1992); Nielsen, *Nature*, 365:566 (1993); Carlsson et al., *Nature* 380:207 (1996), all of which are incorporated by reference). Other analog nucleic acids include those with positive backbones (Denpcey et al., *Proc. Natl. Acad. Sci. USA* 92:6097 (1995); non-ionic backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowski et al., *Angew. Chem. Intl. Ed. English* 30:423 (1991); Letsinger et al., *J. Am. Chem. Soc.* 110:4470 (1988); Letsinger et al., *Nucleoside & Nucleotide* 13:1597 (1994); Chapters 2 and 3, *ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research"*, Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker et al., *Bioorganic & Medicinal Chem. Lett.* 4:395 (1994); Jeffs et al., *J. Biomolecular NMR* 34:17 (1994); *Tetrahedron Lett.* 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, *ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research"*, Ed. Y. S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids (see Jenkins et al., *Chem. Soc. Rev.* (1995) pp 169-176). Several nucleic acid analogs are described in Rawls, *C & E News* Jun. 2, 1997 page 35. These references are hereby expressly incorporated by reference.

[0117] Other analogs include peptide nucleic acids (PNA) which are peptide nucleic acid analogs. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This results in two advantages. First, the PNA backbone exhibits improved hybridization kinetics. PNAs have larger changes in the melting temperature (T_m) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4° C. drop in T_m for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9° C. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is

relatively insensitive to salt concentration. In addition, PNAs are not degraded by cellular enzymes, and thus can be more stable.

[0118] Among the uses of the disclosed NtHMA polynucleotides, and combinations of fragments thereof, is the use of fragments as probes or primers or in the development of RNAi molecules. Such fragments generally comprise at least about 17 contiguous nucleotides of a DNA sequence. In other embodiments, a DNA fragment comprises at least 30, or at least 60 contiguous nucleotides of a DNA sequence. The basic parameters affecting the choice of hybridization conditions and guidance for devising suitable conditions are set forth by Sambrook et al., 1989 and are described in detail above. Using knowledge of the genetic code in combination with the amino acid sequences set forth above, sets of degenerate oligonucleotides can be prepared. Such oligonucleotides are useful as primers, e.g., in polymerase chain reactions (PCR), whereby DNA fragments are isolated and amplified. In certain embodiments, degenerate primers can be used as probes for non-human genetic libraries. Such libraries would include but are not limited to cDNA libraries, genomic libraries, and even electronic EST (express sequence tag) or DNA libraries. Homologous sequences identified by this method would then be used as probes to identify non-human homologues of the NtHMA sequence identified herein.

[0119] Also of potential use are polynucleotides and oligonucleotides that hybridize under reduced stringency conditions, typically moderately stringent conditions, and commonly highly stringent conditions, to an NtHMA polynucleotide described herein. The basic parameters affecting the choice of hybridization conditions and guidance for devising suitable conditions are set forth by Sambrook, J., E. F. Fritsch, and T. Maniatis (1989, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., chapters 9 and 11; and *Current Protocols in Molecular Biology*, 1995, F. M. Ausubel et al., eds., John Wiley & Sons, Inc., sections 2.10 and 6.3-6.4, incorporated herein by reference), and can be readily determined by those having ordinary skill in the art based on, for example, the length and/or base composition of the polynucleotide. One way of achieving moderately stringent conditions involves the use of a prewashing solution containing 5×SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0), hybridization buffer of about 50% formamide, 6×SSC, and a hybridization temperature of about 55° C. (or other similar hybridization solutions, such as one containing about 50% formamide, with a hybridization temperature of about 42° C.), and washing conditions of about 60° C., in 0.5×SSC, 0.1% SDS. Generally, highly stringent conditions are defined as hybridization conditions as above, but with washing at approximately 68° C., 0.2×SSC, 0.1% SDS. SSPE (1×SSPE is 0.15M NaCl, 10 mM NaH₂PO₄, and 1.25 mM EDTA, pH 7.4) can be substituted for SSC (1×SSC is 0.15M NaCl and 15 mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes after hybridization is complete. It should be understood that the wash temperature and wash salt concentration can be adjusted as necessary to achieve a desired degree of stringency by applying the basic principles that govern hybridization reactions and duplex stability, as known to those skilled in the art and described further below (see, e.g., Sambrook et al., 1989). When hybridizing a nucleic acid to a target polynucleotide of unknown sequence, the hybrid length is assumed to be that of the hybridizing nucleic acid. When nucleic acids of known sequence are hybridized,

the hybrid length can be determined by aligning the sequences of the nucleic acids and identifying the region or regions of optimal sequence complementarity. The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5 to 10° C. less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, T_m (° C.) = 2(# of A+T bases) + 4(# of G+C bases). For hybrids above 18 base pairs in length, T_m (° C.) = 81.5 + 16.6(log₁₀ [Na+]) + 0.41 (% G+C) - (600/N), where N is the number of bases in the hybrid, and [Na+] is the concentration of sodium ions in the hybridization buffer ([Na+] for 1×SSC=0.165M). Typically, each such hybridizing nucleic acid has a length that is at least 25% (commonly at least 50%, 60%, or 70%, and most commonly at least 80%) of the length of a polynucleotide to which it hybridizes, and has at least 60% sequence identity (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 97.5%, or at least 99%) with a polynucleotide to which it hybridizes.

Example 11

NtHMA Polypeptides

[0120] A polypeptide may be prepared by culturing transformed or recombinant host cells under culture conditions suitable to express a polypeptide. The resulting expressed polypeptide may then be purified from such culture using known purification processes. The purification of the polypeptide may also include an affinity column containing agents which will bind to the polypeptide; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-Toyopearl® or Cibacrom blue 3GA Sepharose®; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography. Alternatively, the polypeptide may also be expressed in a form that will facilitate purification. For example, it may be expressed as a fusion polypeptide, such as those of maltose binding polypeptide (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX). Kits for expression and purification of such fusion polypeptides are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.), and InVitrogen, respectively. The polypeptide can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, e.g., silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the polypeptide. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous recombinant polypeptide. The polypeptide thus purified is substantially free of other mammalian polypeptides and is defined herein as an “substantially purified polypeptide”; such purified polypeptides include NtHMA polypeptide, fragment, variant, and the like. Expression, isolation, and purification of the polypeptides and fragments can be accomplished by any suitable technique, including but not limited to the methods described herein.

[0121] It is also possible to utilize an affinity column such as a monoclonal antibody generated against polypeptides, to affinity-purify expressed polypeptides. These polypeptides can be removed from an affinity column using conventional

techniques, e.g., in a high salt elution buffer and then dialyzed into a lower salt buffer for use or by changing pH or other components depending on the affinity matrix utilized, or be competitively removed using the naturally occurring substrate of the affinity moiety, such as a polypeptide derived from the disclosure.

[0122] A polypeptide may also be produced by known conventional chemical synthesis. Methods for constructing the polypeptides or fragments thereof by synthetic means are known to those skilled in the art. The synthetically-constructed polypeptide sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with a native polypeptides may possess biological properties in common therewith, including biological activity.

Example 12

Anti-NtHMA Antibodies

[0123] In another embodiment, antibodies that are immunoreactive with the polypeptides are provided herein. The NtHMA polypeptides, fragments, variants, fusion polypeptides, and the like, as set forth herein, can be employed as “immunogens” in producing antibodies immunoreactive therewith. Such antibodies specifically bind to the polypeptides via the antigen-binding sites of the antibody. Specifically binding antibodies are those that will specifically recognize and bind with NtHMA family polypeptides, homologues, and variants, but not with other molecules. In one embodiment, the antibodies are specific for polypeptides having an NtHMA amino acid sequence as set forth in SEQ ID NO:2 and do not cross-react with other polypeptides.

[0124] More specifically, the polypeptides, fragment, variants, fusion polypeptides, and the like contain antigenic determinants or epitopes that elicit the formation of antibodies. These antigenic determinants or epitopes can be either linear or conformational (discontinuous). Linear epitopes are composed of a single section of amino acids of the polypeptide, while conformational or discontinuous epitopes are composed of amino acids sections from different regions of the polypeptide chain that are brought into close proximity upon polypeptide folding. Epitopes can be identified by any of the methods known in the art. Additionally, epitopes from the polypeptides can be used as research reagents, in assays, and to purify specific binding antibodies from substances such as polyclonal sera or supernatants from cultured hybridomas. Such epitopes or variants thereof can be produced using techniques known in the art such as solid-phase synthesis, chemical or enzymatic cleavage of a polypeptide, or using recombinant DNA technology.

[0125] Both polyclonal and monoclonal antibodies to the polypeptides can be prepared by conventional techniques. See, for example, *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Kennet et al. (eds.), Plenum Press, New York (1980); and *Antibodies: A Laboratory Manual*, Harlow and Land (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1988); Kohler and Milstein, (U.S. Pat. No. 4,376,110); the human B-cell hybridoma technique (Kosbor et al., *Immunology Today* 4:72, 1983; Cole et al., *Proc. Natl. Acad. Sci. USA* 80:2026, 1983); and the EBV-hybridoma technique (Cole et al., 1985, *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Hybridoma cell lines that produce monoclonal antibodies specific for the polypeptides are also con-

templated herein. Such hybridomas can be produced and identified by conventional techniques. For the production of antibodies, various host animals may be immunized by injection with an NtHMA polypeptide, fragment, variant, or mutants thereof. Such host animals may include, but are not limited to, rabbits, mice, and rats, to name a few. Various adjuvants may be used to increase the immunological response. Depending on the host species, such adjuvants include, but are not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. The monoclonal antibodies can be recovered by conventional techniques. Such monoclonal antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD, and any subclass thereof.

[0126] The antibodies can also be used in assays to detect the presence of the polypeptides or fragments, either in vitro or in vivo. The antibodies also can be employed in purifying polypeptides or fragments by immunoaffinity chromatography.

Example 13

Double-Stranded RNAs

[0127] In one embodiment, double-stranded ribonucleic acid (dsRNA) molecules are provided for inhibiting the expression of the NtHMA gene in a cell (e.g., a plant cell), wherein the dsRNA comprises an antisense strand comprising a region of complementarity which is complementary to at least a part of an mRNA formed in the expression of the NtHMA gene, and wherein the region of complementarity is less than 30 nucleotides in length and wherein said dsRNA, upon contact with a cell expressing said NtHMA gene, inhibits the expression of said NtHMA gene by at least 20%. The dsRNA comprises two RNA strands that are sufficiently complementary to hybridize to form a duplex structure. One strand of the dsRNA (the antisense strand) comprises a region of complementarity that is substantially complementary, and typically fully complementary, to a target sequence, derived from the sequence of an mRNA formed during the expression of the NtHMA gene, the other strand (the sense strand) comprises a region which is complementary to the antisense strand, such that the two strands hybridize and form a duplex structure when combined under suitable conditions. The duplex structure is between about 15 and 30 (e.g., between about 18 and 25), typically between about 19 and 24 (e.g., between 21 and 23) base pairs in length. Similarly, the region of complementarity to the target sequence is between 15 and 30 (e.g., between about 18 and 25), typically between about 19 and 24 (e.g., between 21 and 23) base pairs in length. The dsRNA may further comprise one or more single-stranded nucleotide overhang(s). The dsRNA can be synthesized by standard methods known in the art as further discussed below, e.g., by use of an automated DNA synthesizer, such as are commercially available from, for example, Biosearch, Applied Biosystems, Inc. In another aspect, an expression vector can be used to express an RNAi molecule in vivo.

[0128] The dsRNA can contain one or more mismatches to the target sequence. In one embodiment, the dsRNA contains more than 3 mismatches. If the antisense strand of the dsRNA contains mismatches to a target sequence, it is typical that the

area of mismatch not be located in the center of the region of complementarity. If the antisense strand of the dsRNA contains mismatches to the target sequence, it is typical that the mismatch be restricted to 5 nucleotides from either end, for example 5, 4, 3, 2, or 1 nucleotide from either the 5' or 3' end of the region of complementarity. For example, for a 23 nucleotide dsRNA strand which is complementary to a region of the NtHMA gene, the dsRNA preferably does not contain any mismatch within the central 13 nucleotides. The methods described herein can be used to determine whether a dsRNA containing a mismatch to a target sequence is effective in inhibiting the expression of the NtHMA gene.

[0129] In one embodiment, at least one end of the dsRNA has a single-stranded nucleotide overhang of 1 to 4 (e.g., 1 or 2 nucleotides). dsRNAs having at least one nucleotide overhang have inhibitory properties. The dsRNA may also have a blunt end, typically located at the 5'-end of the antisense strand.

[0130] In yet another embodiment, the dsRNA is chemically modified to enhance stability. The nucleic acids may be synthesized and/or modified by methods well established in the art, such as those described in "Current protocols in nucleic acid chemistry", Beaucage, S. L. et al. (Eds.), John Wiley & Sons, Inc., New York, N.Y., USA, which is hereby incorporated herein by reference. Chemical modifications may include, but are not limited to 2' modifications, introduction of non-natural bases, covalent attachment to a ligand, and replacement of phosphate linkages with thiophosphate linkages. In this embodiment, the integrity of the duplex structure is strengthened by at least one, and typically two, chemical linkages. Chemical linking may be achieved by any of a variety of well-known techniques, for example by introducing covalent, ionic or hydrogen bonds; hydrophobic interactions, van der Waals or stacking interactions; by means of metal-ion coordination, or through use of purine analogues.

[0131] In yet another embodiment, the nucleotides at one or both of the two single strands may be modified to prevent or inhibit the activation of cellular enzymes, such as, for example, without limitation, certain nucleases. Techniques for inhibiting the activation of cellular enzymes are known in the art including, but not limited to, 2'-amino modifications, 2'-fluoro modifications, 2'-alkyl modifications, uncharged backbone modifications, morpholino modifications, 2'-O-methyl modifications, and phosphoramidate (see, e.g., Wagner, Nat. Med. (1995) 1:1116-8). Thus, at least one 2'-hydroxyl group of the nucleotides on a dsRNA is replaced by a chemical group. Also, at least one nucleotide may be modified to form a locked nucleotide. Such locked nucleotide contains a methylene or ethylene bridge that connects the 2'-oxygen of ribose with the 4'-carbon of ribose. Oligonucleotides containing the locked nucleotide are described in Koshkin, A. A., et al., Tetrahedron (1998), 54: 3607-3630) and Obika, S. et al., Tetrahedron Lett. (1998), 39: 5401-5404). Introduction of a locked nucleotide into an oligonucleotide improves the affinity for complementary sequences and increases the melting temperature by several degrees (Brasch, D. A. and D. R. Corey, Chem. Biol. (2001), 8:1-7).

[0132] Conjugating a ligand to a dsRNA can enhance its cellular absorption. In certain instances, a hydrophobic ligand is conjugated to the dsRNA to facilitate direct permeation of the cellular membrane. Alternatively, a ligand conjugated to the dsRNA is a substrate for receptor-mediated endocytosis. These approaches have been used to facilitate cell permeation of antisense oligonucleotides. In certain instances, conjugation

tion of a cationic ligand to oligonucleotides often results in improved resistance to nucleases. Representative examples of cationic ligands are propylammonium and dimethylpropylammonium. Anti-sense oligonucleotides can retain their high binding affinity to mRNA when the cationic ligand is dispersed throughout the oligonucleotide. See M. Manoharan *Antisense & Nucleic Acid Drug Development* 2002, 12, 103 and references therein.

Example 15

Methods For Identifying NtHMA Modulatory Agents

[0133] Methods for identifying agents that can modulate NtHMA expression level and/or activity are disclosed below. Candidates ("a test agent") that may be screened to identify NtHMA-specific modulatory activity include small molecules, chemicals, peptidomimetics, antibodies, peptides, polynucleotides (e.g., RNAi, siRNA, antisense or ribozyme molecules), and agents developed by computer-based design. Modulation of NtHMA includes an increase or decrease in activity or expression. For example, a method for identifying candidates that can modulate NtHMA expression and/or activity, comprises: contacting a sample containing an NtHMA polypeptide or polynucleotide with a test agent under conditions that allow the test agent and the NtHMA polypeptide or polynucleotide to interact, and measuring the expression and/or activity of the NtHMA polypeptide in the presence or absence of the test agent.

[0134] In one embodiment, a cell containing an NtHMA polynucleotide is contacted with a test agent under conditions such that the cell and test agent are allowed to interact. Such conditions typically include normal cell culture conditions consistent with the particular cell type being utilized, known in the art. It may be desirable to allow the test agent and the cell to interact under conditions associated with increased temperature or in the presence of reagents that facilitate the

uptake of the test agent by the cell. A control is treated similarly but in the absence of the test agent. Alternatively, the NtHMA activity or expression may be measured prior to contact with the test agent (e.g., the standard or control measurement) and then again following contact with the test agent. The treated cell is then compared to the control and a difference in the expression or activity of NtHMA compared to the control is indicative of an agent that modulates NtHMA activity or expression.

[0135] When NtHMA expression is being measured, detecting the amount of mRNA encoding an NtHMA polypeptide in the cell can be quantified by, for example, PCR or Northern blot. Where a change in the amount of NtHMA polypeptide in the sample is being measured, detecting NtHMA by use of anti-NtHMA antibodies can be used to quantify the amount of NtHMA polypeptide in the cell using known techniques. Alternatively the biological activity (e.g., heavy metal transport) can be measured before and after contact with the test agent.

[0136] It will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without departing from the spirit and the scope of the invention. Accordingly, the invention is not limited except as by the appended claims. Unless defined otherwise, all technical and scientific terms have standard meaning as commonly understood to persons skilled in the art. Although exemplary methods, devices, and materials have been described with particularity, alternative methods and materials, that may be similar or equivalent to those described herein, are applicable for making the disclosed compositions and for practicing the disclosed methods.

[0137] Any publication cited or described herein provides relevant information disclosed prior to the filing date of the present application. Statements herein are not to be construed as an admission that the inventors are not entitled to antedate such disclosures.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 49

<210> SEQ ID NO 1

<211> LENGTH: 17921

<212> TYPE: DNA

<213> ORGANISM: *Nicotiana tabacum*

<400> SEQUENCE: 1

```

gcgtagagcta tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca ggtatccggt      60
aagcggcagg gtcggaacag gagagcgac gagggagctt ccagggggaa acgcctggta      120
tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgctgatttt tgtgatgctc      180
gtcagggggg cggagcctat ggaaaaacgc cagcaacgcg gcctttttac ggttcctggc      240
cttttgctgg ccttttgctc acatgttctt tctgcggtta tccctgatt ctgtggataa      300
ccgtattacc gcctttgagt gagctgatac cgctcgccgc agccgaacga ccgagcgag      360
cgagtcagtg agcagggaag cggaagagcg cccaatacgc aaaccgcctc tcccgcgcg      420
ttggccgatt cattaatgca gctggcacga caggtttccc gactggaaag cgggcagtga      480
gcgcaacgca attaatgtga gttagctcac tcattaggca cccagggctt tacactttat      540

```

-continued

gcttccggct	cgatatgtgt	gtggaattgt	gagcggataa	caatttcaca	caggaaacag	600
ctatgaccat	gattacgcc	agctatttag	gtgacgcgtt	agaatactca	agctatgcat	660
caagcttggt	accgagctcg	gatccactag	taacggccgc	cagtgtgctg	gaattcgccc	720
ttatcacctc	tttccaaaac	aaaagtatat	ccaatatatt	ccagaactag	aaattccttt	780
ttcatctatt	atcttctcct	cctccttaga	gaaggagaaa	aatggtggaa	agtgaaaaaa	840
tgaatgaaac	aaagaagttg	agcaagagct	atcttgatgt	tttgggaatt	tgctgtactt	900
cagaagttgt	tctagttgaa	aaaattctca	agaatcttga	aggggttaaa	gagggtttcag	960
taattgtcac	aacaaagact	gtcattgtta	ttcatgattc	tcttctcatt	tctccgcaac	1020
aaattggtaa	gaaagatagt	tacacacett	tattactctc	tcagttctat	tttttacgtg	1080
atactatttt	ctttttaata	tgttctgaaa	agaacgttac	ctttttatat	attgatgtaa	1140
tttcaactta	aacttcatat	tttttttctt	aaatagtatt	gttttatagt	cataaaaaata	1200
ttattataag	tctttcttaa	gttttgtggc	ttgttaaata	aattcacata	aatgaaata	1260
aagcggtagg	agtaccattc	ttcatctttt	cttaattaac	tgacatttcc	tttctttttt	1320
gtaagtttat	atatattaag	taaagtgatt	tttctcttaa	gaaatcgcca	aaaaaaaaaa	1380
ggaaagagaa	gagaagagac	aagaaggggg	agacaaatgg	aatcagaaac	ggttttgatt	1440
tgttcaagtg	atatttgggt	gtagttgttt	caagcactgt	tctttctggt	tggtggcatt	1500
tgctagaatt	aagtgtatat	attaatttgg	gaaaatcttt	aaagtggtaa	tttttagttt	1560
tattagattg	agtaaccaag	acggaaaaaa	catgaactat	tttcttttg	aaatttctag	1620
tcaagaacat	gcataaaaat	tctcttttaa	aacgactctc	ataaaaattc	atgtggtcga	1680
gtttacgcag	cacatggacc	catagtctcc	gcctaactaa	gtattttaag	tatgtatttt	1740
ctaaaattca	tctaataatt	tctgttggcg	cacatgctcc	acaaaagatg	aattgtgcat	1800
ttgtttgaat	attgagttat	tactcaagg	aatatggatc	aattccactt	tttttctctt	1860
ttctttaata	ttgtaccat	atcttaaaaa	ttctagctcc	gcctccgact	tgcccatcgt	1920
atccgcccc	gccctcagtg	ggatagagta	ggtagaggag	atagtttata	aatgtttttt	1980
ctctaacaat	taggttttgg	aataaattct	gagattcaat	ggtctatttg	aaacctgtaa	2040
ttactactcc	ctcgtttca	tattagatga	ttactttcct	tttagtcta	ttccaaaaca	2100
aataacacat	ttctaatttt	ggaaataatt	caattttaaa	atctttcatt	ttaccatttt	2160
accttaatga	atgtttttat	agccacacaa	atgtcatgac	cccacaaatt	ttttaccctt	2220
taaaacttta	agaccacaaa	ttttaaaagt	ttcttctttt	ttcttaaaact	atatgccaa	2280
taaaactaac	tcatcttgaa	acggaggagt	actataagaa	tagttatata	tgatttggcc	2340
ccacaaatta	cataattgtg	gcaagaaatc	caagtttcta	gttggtgata	atttagtgat	2400
ggaggcgatc	tctgttgaac	cgttagagaa	aatattttga	ctcatgtcgc	tttgcattca	2460
ctgtgttgaa	gagtagttgc	taggaactaa	caaagtaatt	gtagtctcct	tgtttttttt	2520
ttttctcttc	ttgatgaatt	actgtacttt	atcttcccat	ttttttaacg	tctgttactt	2580
atggtcagtt	ttcaagtgtg	aaaaattagc	cagaagaagt	aattcaacaa	acttaacctt	2640
ttcttttctt	cattcttgcg	atctttccta	aatatattta	gtagagtctc	ttttctactt	2700
cagtcttatt	gcatttcttg	cacacctaac	agtagtggtg	cataaattgg	acctagctaa	2760
caaaaggtaa	ctcttatctg	caagttaaat	tggtacagta	acatgaaact	tggtataatc	2820

-continued

ttttaaatg cagccgataa tatatgtatg ttttacattg gttgtgtgta tataatgtaa	2880
attctttaca gcatgtttga ctaacctgca aaaattagtt tttttttttt caaaagtatt	2940
tttggtgaga agcagtttgt gtttggctaa gtaatttgaa aaatacttct gagcaacaat	3000
tagtgtttgt ccaagctttt aaaaactgct ttttaattgta tttttgtcaa aagagctttt	3060
taaaaaagta ttttttagag agaaactact tttttctgct tctccaaaat tgtttctgct	3120
tctcctcaaa aacacttttt ttccttctaa aagcttgtag aaacacttca actaaaaaa	3180
atatattagc acttatatta tccttataat tattgaagtt accatctatc ttttgtggat	3240
gtagttaaag cattgaatca agcaagatta gaagcaagca taagagtgaaggagagaaa	3300
aactacaaa agaaatggcc aagtccattt gcaattggca gtgaaatatt gcttgactc	3360
tcacttttga agtacttttt tgcaccttcc caatgggttag cacttgagc tgttgagtt	3420
gggattcctc caattatttt tagagggtgtg gctgccgtgc gaaacctcac tcttgacatc	3480
gacattcttg ttttaatagc aggtactttg cttttccttt tctttctttt taactttttg	3540
aaaagcaaaa agagatactc ccttttgtcc caatttatat ggcggtat tt aattaacaa	3600
aaaaatctaa gaaaaatag aatactttaa aaatttatgg tttgaaataa atcttaggca	3660
tctaaataca aggttaaaat aataatttta aagttaaaaa attcaagaaa gaaaggaaga	3720
cttctaaaaa ttgtggcaag aaataaatct tagagatttg tgttataaat catctcacta	3780
aggctaaaaa aataatttta aagttaaatt acttttaatt atgaaaaggt gatagatttt	3840
tgggaaaaga tagaaaagaa aactgtgcta tataaattga gacagagaga gtaatttacc	3900
ataagagtat ttgaagttaga ttggttata gaattaacat gtgctcctac tttagctttg	3960
tgattgaaat ttggctattg ctttaatttc ttagctaatt ggatgcttct aatgtgatgc	4020
ttgaatcctt attcttgata atgctttgtg attcattatt tcattaaaaa gcatgcacca	4080
tagggcatgc aattaatat tgtatttaag aatgcacttt taactaacc acaagatttg	4140
agtgggaggg atgattcttg ggtgttagat gctaatttg gaccacccta gtgactttaa	4200
ataacgaaag catgaaaaac aattattgga tgcttgatat actttatggt ataatatttt	4260
caggtgacct agaactacac aaataaactt ctttttccat attggaatag gatgatttag	4320
atttcaagat ggaattagtg attctatcac agattatctt ttatttaatg atatcaata	4380
tgaaaagaga aaaaaaaag gtgtctagga aatagtcaga aagatagtat gacatatatt	4440
attgatacaa attaatcagc tcaacacaaa ggctagcata tatttaacaa atagttatag	4500
acttgtagtg ttgtaccttt agttaagaga ctaaacacta ccaaaagggt gcgatggagc	4560
agtaggtact cctctattgt catccaggcg ccttgatata aattaattag atcccaaac	4620
gaggatcaga taccaaaata aaaacaaaaa gaaactatct ttagtgtttg ttgatttaca	4680
tggcgtagaa aaataaagga caataataa aaaaccttgg aatatcagtt aagatgcctt	4740
aatttcaaat cagtggagta tattccgact tccgtattaa ttcgctcta atcaagtctt	4800
ttaaaagatt aaataatgaa aagtttttat agtcacgaaa acattatgac caaaaattt	4860
tgctctttaa gcttttttaga ccataaattt caaaaaaaa aatttgactg aaatatttaa	4920
taccacatat ttttaaaaaa ttattttttt cttacattta gtgtcaagta aaattagatg	4980
tgataggaaa aaagaatttc gatttcgtct agtaaggaaa gtacaaaatt gcttatgtaa	5040
ctcgtaatat atcagaaatg gtttccagg tgtaaagcac aaacaacgga gccacatca	5100

-continued

gtaatgtgta aatgtgtaac gtttactctc tttttttcct caaataattg caaagaaaat	5160
gacttttcctg caagtctttc ttgccccctt tatagaaggt gacgttgact cgtaccagat	5220
tttcttgta tagtattaat cagtagtaat atcagtcacg gcgagggtta aaatttagtc	5280
tattagtaac gtaaataatt tttaaattgt catatttagt taatctaaaa tatcatgcaa	5340
ttttttatat caagtttgat atggtgacgt aaataataga acaataatct ataaacttga	5400
ttaaattgca ttgataataa atcaaaatat tatttgaacg attaaaatct cacagttaat	5460
tttgcatctt gcataaaaa atatccatca tatttccttg attatcattt gatgccgtct	5520
tagtttcttt taaaaaatt tagaatttat atcaaatatg attttttaat tactcgaact	5580
tacaagcaga ctaagtttga tattttccta attcaacgat atacgggtac tacggaaggc	5640
atttactaga aatactctga gatgttactg caattattat tattattatt ttaaaaagag	5700
aaaaaaataa cttttaaacg tccatgtgaa attatgtata ttttattata gcatgaagtg	5760
accccatctt ttatctcata aataacattg atcccatatt tttctactgt atcatcacta	5820
tcatgaaaa tacatctaga ttactgagat gtttattggc agtagtatct acagatacaa	5880
cagatgcttc tagttctatt gatgttttca ttttgacaaa aatttaattg agaaagcaaa	5940
agattttgca agaattctta gggttttatt caaaaaaaa aaagaattct tagaaatata	6000
agttttggca attaaataat tccagtaatt gggaaaaaac acttgaatag gctatagaag	6060
taaaagaaac ttctatatat taccaggcag cagagtttcc caaaatcctt ttttctaaaa	6120
aaaaaatagt agaaaatgag caatgtaatt tctttaagta acattctcta tttagtaaaa	6180
tgccatctt tctaatagag taaaagcaat agcaaataaa agaaagtta ttctttttt	6240
tcgaggggtg tgccgaccaa ggcttaata gataggaata atcaccta atagaaaaac	6300
tacctatcaa tttttgtctc ggttggattt caacatgata cctcatagtt ctacgtacgc	6360
ccacttcaat taccactagg aacaccgttt ggtgcaggaa aattttggtc ataaacttca	6420
attttaagcc ttcataataa caataaaaga gctaattagc agttaacagt cgagttaata	6480
tagctgaata atgcagttca actaaatatg tatggaaatg gtgaaaagca caaagggtgac	6540
ttatccttag gtactttata gctttatctt ttaaaaagta gtagcaagat atgatatgat	6600
tccttgaaga agaaaaggtc actgtgactc tttatttcta tcagtacctg tttgaataaa	6660
attggctaag aaagtgtgaa aatggactag caggacaaga atctcaattt ggtgttgctt	6720
taccatctt agagtgcag caatcaaaaa cccaaaaatt aaataataa taagaaaaga	6780
aaaagatgag ttattggaaa gggtaaaatg aatgaagata agaatactcc attcagtcca	6840
aaatattaag ctagcatttg gacatagatt tgggtgaagc ttgaagaaaa gaaaaagag	6900
tttttgaaga ttatatgaaa aataattttt gaaagttaaa attgtgttag ggaagttttg	6960
tgattcaaaa actactctag aactgttttt gggatttaaa tattttcttt tcaacatgtg	7020
ccaaaaaatg attaaatct ataaacgaac acataattga aaaagctct caaattttat	7080
gacaaaacag gagcttattg ttttggcaaa ttgaatttgt ttcgaaatac ttgatgatta	7140
aggagagaca aaggatagta tagacaaaag gctattagta tttgtaattg aggcattata	7200
atgtcttttt aaagcgatgt gcaaaaacct taaaaagac gaattatatg gattatatta	7260
gaagtagtat taaatttttaa tggaaacttag tgcaatcatt ttacaagggc atagtgttaa	7320
agctagaaat gctgattctt atagctggct ttgtcatgtg cagtggctgg atcaattggt	7380

-continued

ttacacgatt	attggaagc	tggtactatt	gtcttcttat	tcgccattgc	agaatggcta	7440
gagtcagg	caagtcacaa	ggtttgtgaa	tttctgtccc	tgtttaat	cttgcacaa	7500
gaaatggcta	gtgtacatct	actttcctag	gatgacagta	atgtgtgttt	ttttattta	7560
ttgtgatacc	aaaagtatga	acacctaatt	tttctcagga	tacatat	ccactttggc	7620
caaaactgact	ttaacaagtt	gtctcttgga	ttacgccaca	tcgaacctaa	aagtgttggc	7680
actaccatgt	caacttgtgc	tgtcttatag	ggatttgaga	cattttgtct	aggggtgcat	7740
atggacccct	tcttcgatgag	ctcgatagca	caaaagcttg	ctagtcattc	tacttgacct	7800
gcctcattag	cccgccctac	gcatgggcat	cacaaaccag	gccacacgta	ggaattgagt	7860
ctctccctg	cgggtcacc	acagacctta	cagttttgtc	gcttgggtaa	tgccacaact	7920
agcccgataa	cgtgcttacc	atgtgaactt	gtaccatata	ttacctagca	cttcaccttc	7980
ctctcccttt	gcgacgcgcc	taaaatgtca	tatgcaccgt	ttttccgaa	gtttgctagt	8040
ctagaagcct	attagttctt	aggcttaacc	tgccctatta	accactctcc	atcataggcg	8100
tcaccttaat	ttaaaagaat	ttgttcttag	aaggctctaa	cttaaacaccag	aaatcagttc	8160
agctgtcttc	tgttctttta	cctcaataac	attgtataat	ggtaaggact	aaactgcagc	8220
tctcttgg	atgagtggt	taaatttcat	tctgaaaatt	aatttacttc	acagctctat	8280
ttggagaaaa	taaagattaa	atattgtgag	aatgcacggg	agaaaaatat	tatattgatt	8340
aaagtgtgt	acaaccttat	ttatatacag	taattatata	ataatatgta	tctacttccc	8400
gatgtgggac	actaaatatg	actaactact	taacacttcc	tctcaagccg	gtgcatataa	8460
atcatacgta	ccgagcttgt	tacagatgta	accaatacga	gaaccagtaa	gagacttagt	8520
gaaaatat	gctagcta	cattcgactt	tacaaacttt	gtaacaatat	ctcctgagag	8580
tatcttttct	ctgacaaagt	gacagtcgat	ctcaatgtgt	ttagtcctct	catggaatac	8640
cggatttgac	gcaatatgaa	gagcagcttg	attatcacac	accagttcca	tcttattgat	8700
ttctccgaat	ttcaactcct	tgagcaactg	cttgatccac	actagctcac	acgttgccat	8760
agccatggcc	cgatattcgg	cttcgggtgct	agatcgagca	actacattct	gtttcttgct	8820
cttccaagag	accaaattac	ctcctactag	aacacaatat	ccagacgtag	aacgtttatc	8880
agaaggtgat	cctgcccatt	cagcatttgt	gtaccctgta	atctgctcgt	agcctcgatc	8940
ctcgaatagt	aaccctttgc	ctggagctga	ctttatatat	cgaagaatgc	gaacaactgc	9000
atccagtgta	ctatcacagg	gagaatccat	aaactgactt	acaacactca	cgggaaaaga	9060
aatgtctggt	ttagtcaccg	tgaggtaatt	caatttgcca	accaacctcc	gatattctcat	9120
aggatctcta	agaggtctcc	cctgtccatt	gcagaagctt	agcattcgga	tccataggag	9180
tgtcaacagg	tctacatccc	atcattccag	tcttctcaag	aatgtctaag	gcatacttcc	9240
gctgtgaaat	aacaatacct	gagctagact	gagcgacctc	aatacctgga	aaatacttca	9300
atctgcctag	atccttagtc	tggaagtgc	aaaagagatg	ttgcttcaga	ttagtaatac	9360
catcctgac	attgtcagta	ataacaatat	cgtctacata	aaccaccata	gtaagggttat	9420
gaacaaaaca	cgtgcttcta	aagacacggg	gtggaagaga	gaacaaaggt	aagtggggaa	9480
acaggacaga	gaatggaact	tgattctgga	tagctgaaga	tgacatacga	ttaataagat	9540
agcaagatgt	aagaactgca	tcaccccaaa	aatgcaacgg	agcatgagat	tgtatgagta	9600
gagtagcagt	agtttcaata	agatgtctac	tctttcttcc	agctacccca	ttttgttgag	9660

-continued

atgtgtacgg	acaagatggt	tgatgaataa	tcccatgaga	gttcataaac	ttctgaaatg	9720
gggaagacaa	atactctagg	gcattatcac	tacgaaatgt	gcggatagaa	accccaaatt	9780
gattttgaat	ttcagcgtgg	aaggtctgaa	aaatagaaaa	tagctcagac	cgatttttca	9840
tcaaaaaat	ccaagtgcac	ctggaataat	catcaatgaa	actgacaaag	tagcgggaatc	9900
ccaaggtaga	actgaccgga	ctaggacccc	aaacatctga	atggaccaaa	gtaaaagggtg	9960
acgctgaggg	aaatgggagt	gggtatactt	accgagctgt	catgactcat	actctagagt	10020
ggacaagtga	gataaaccag	gtatcatttt	ctaaagtttt	gacaaactgg	gatgtcccaa	10080
ccgtttatgt	aataaatctg	gtgaatcggt	aacaagacaa	gttattaaag	gaagacaaga	10140
tgcgagtcca	tatgattttg	caaggataag	gtaataaaat	ccatctaatt	catgtctgat	10200
accaatgatc	cgccccgtac	tatgtttctc	tataaaaaata	aggtcaccaa	gaaataaaac	10260
agcgcattta	agtgatttgg	ctaagcgact	aacggctatg	agattaaaag	gactattggg	10320
aacataaagg	actgaatcta	aaggtgagga	aggaagtggg	cctgcttgac	ctattgcagt	10380
tgccatggtt	tgagactcat	tggccattgt	gactgttaga	agagattgag	aatatgaata	10440
ctggtgaaaa	gagatttgtt	accagaaata	tgatcagatg	catttgaaac	aatgacccaa	10500
gactcaaagg	tttaagatcc	ggagacacaa	gtcacgctac	tatctgtttg	aacaatggaa	10560
gctatccctg	aagatgtctg	tttacatgct	ttgtactgaa	ggaactcagt	ataatccgat	10620
aaagaaacca	tctggatttg	attcaacgaa	ttggatccaa	cagattgtga	gttaaaggct	10680
tctgatacga	atgtcattat	aatgttgtgc	cgaaaaaaca	aaaaattcct	ggaaattact	10740
attcacgcgc	gaaaaatata	aaagtgatct	gaatttgatt	taaattggat	gggtatgctc	10800
gtatttgcaa	ggagaagaca	ctgccctgaa	ggaattttac	caaattcttg	cgggaaattg	10860
cctcatgtgc	ggcgtgtggg	cgtcagaact	tcgtcggaaa	aattcttcgc	gcggcgcgtg	10920
agggcgcgtg	tagccttttt	tgccagagat	tttttaatag	gttggtcgct	gagctctgaa	10980
ctacttcccc	gtggtgttac	cttttgcaca	acactgacag	atagtatgat	tcttgcggac	11040
agacctattt	ttgccggaaa	agagcttccg	gttgactggt	ttcttttccc	ggagtcgctg	11100
gaatttatgc	actacgataa	atttctcacg	gttgctctga	taccatgtga	gaatgcacgg	11160
gagaaaaatta	ttatatgtat	taaagtattg	tacaacccta	tttatataca	gtaattacat	11220
aataataggt	atctacttcc	cgatgtggga	cactaaacat	gactaactac	ttaacaaata	11280
tggtatttga	atttagtctc	tttgacataa	acgacataag	cctatgctta	tcttttctta	11340
cttttttagc	aatgctaaat	agtaggtcct	aactacaaac	tttatagcac	actgaaaatt	11400
acaaaaatat	agagatggcc	aatgaagggt	ttgtctgcta	acataactct	gtgtctttat	11460
tttctcactg	atattgtata	tggataaagc	attctgataa	atgaaaacct	ttatggttat	11520
gtaggctacc	gccgctatgt	catcactggt	caatatagtc	cctccaacag	cagtttttagc	11580
ggaaagcgga	gaagtcgtaa	atggtgatga	agtcaagggtg	aatagcattc	ttgctgtgaa	11640
agctgggtgaa	actataccta	ttgatggagt	tgtagtggaa	ggggaatgtg	acgtggacga	11700
gaaaacactg	acaggcgagt	cgtttccagt	ttctaagcaa	agagattcaa	cgtgtctgggc	11760
tggcgctaca	aatctaaatg	gtagtatagt	atttcttcat	gcttcattta	tttagtgctg	11820
aaacttcaag	tattgtttgt	taatgttatt	tgctcaattc	ttcaggctat	atcagtgtta	11880
agactacggc	tttggtgtaa	gattgtgcgg	tggctaggat	ggcacagcct	gtcgaagatg	11940

-continued

ctcagaacaa gaaatcaaaa acccaaagat acatcgacaa gtgtgctaaa tattatacac	12000
caggctagtg aatcttatgt tgtgccacat caagtcaaaa atgcacgta cegtgtgaac	12060
ttgttctttg tcttatgaat cacgtcacta tcctctccct ttcgatatg agatttcctt	12120
aagggtgcat atgaatccct tcttcggaag cttgccagca taggagtcta tcagtccttt	12180
cacttgaccc gccctctcag cctgcctgca gtcatggcgcg tcgcactact atattgctct	12240
ttcgtttaaa actttttatt tctaatactt ccctgctctt tgtgtatgtc taatttcgac	12300
tgggtgatgt ttgcagcaat tgtggctata ccagcttctt tggcaattgt tcctactgca	12360
ttaagagtgc acaatcgaaa tgaatggtat cgcttggett tggtcacatt ggtgagtgc	12420
tgcccggtg cactcgttct atctacacca gttgccatgt gttgcgcact ttcaaaagca	12480
gcaacgtccg gtcttctgtt taaaggagca gagtaccttg agactctagc taaaatcaaa	12540
atcatggctt ttgacaaaac agggactata actaaaggag aatttatggt gaccgagttc	12600
aagtctctga ttgatggttt tagtctcaat aactgcttt actggtaaag gttaccactc	12660
atacatattc ttttatgttg ccaaagagaa ttcaaatct taactggta tctttcacgg	12720
cacattgata gcgatataac atgattgatt tatatcatat attcataaaa gatgaaatag	12780
ggagtgccac attcacattc tcatattgaa gtttctgaaa tggtctaat ggttcaccat	12840
agagcaaaa taacatatag acacaacgtc agccgtctga tattcaggaa cttagatgga	12900
atagttggat cttatacatt gaggacacat aaaagtactt ggtcatataa attttagaaa	12960
tataatcaat gtattataat ctaaaattct tcaaatattc ttgatactgc aatacaaaag	13020
cacatggcac actgaataga agccttgctt ggtggtctaa aacattcgtt tagagtaa	13080
actgagttgt ctagtgaata ttttcagggt ttcaagcatt gagagcaagt caggctatcc	13140
gatggcaacc gctctggtgg actatgcaca atcaaattcc gttgagcaa agcctgatag	13200
agttgagcgg tttcaaaatt ttccctggtg agggatattt ggaagaattg atggaatgga	13260
aatctatgtc gggaatagga aaatttcttc aagagctgga tgtaccacag gtaaatggtt	13320
gaatcatttc ttatgctcat agtagagata aaacatcaga gttataatta taagtatatg	13380
atttctccag ttaattttgc tgtagattt tctttgacct gtttagcact aatgcggtgg	13440
atgtttgaat ttcagtacca gaaatagagg gtgatagttt caaaggaaag tctgttggt	13500
acatatttt gggaatcctt ccagctggaa ttttcagctt tccgatggt tgcgaattg	13560
gtgtaaaaga agcaatgaga gaactgaagc agatgggtat caaaaccgcg atgcttactg	13620
gtgatcgta tgcagctgcc aacctgtgc aggatcaggt atattaataa ttctgcatta	13680
cgctgaaatg attataaaac cctttggatt attgtttagt cttagaatt ttactgaac	13740
tcttattggt tccttctctt atcatcaaca ttggttaaac atttcatcta aatttagaga	13800
acgtatcacc aagtaagtgc ttaccttta cagggtcata taaaatactt aagacagtgt	13860
gatgtgaaga tgaaggttta atgttgatct ggataaacca agttattatc acaactaata	13920
taagatatgc tattgttctc caataattgg acgattttcg gacgtacgac gtacaattct	13980
tcacatatga aacctacatc agacgtacat gacacgtat gtttagcata aagagtcaag	14040
attagcatga tgatttaagc tgaatctgaa tttcaagtat ctattcttgt attgtacca	14100
ggggcggaac tagtgtgtg cttagaggtc tcaaacattg tatttgtgtt aaaaaattca	14160
cttcatatgt atttaataa tttatccaga gcagtgagcc atatttttta gaatccagaa	14220

-continued

```

cccataaaact caaaatcata gatccacctc tgattgtaag tcggaacaat tatgcagtta 14280
gggtggagctt tggatgaatt tcaagcagaa ctccctaccag aggacaaggc aacaatcatc 14340
aagggttttc agaaggaagc tccaacagcg atgataggcg acggccttaa tgatgctcct 14400
gcattagcaa cagctgacat tggcatctca atgggcatct ctgggtcagc tctcgctaaa 14460
gaaacaggcc atgctatact aatgacaaat gacatcggaa gaataccgaa agctgcacgt 14520
cttgctagaa gagttcgaag gaagattgtt gagaatatga ttatatcagc cgttaciaag 14580
gctgccatag ttgcattggc aatagcaggt tatccattgg tttgggctgc tgtcctcgca 14640
gatactggga catgcttgct agtgattttg aacagcatgc tacttctacg aggaggcaca 14700
cgagacatg ggaaaaaatg ttggagatct tctactcctt cgcagtctcc ccaccacaaa 14760
gacaaaagctt catgttgcaa gtcggaaaat gctccccagc tgtgttgctc tgatattgag 14820
tcacaaaaga aatgtacaag tcaatcatgc tcgtccgagg tgtgtgttcc aagatgtcaa 14880
cctgtctcct cgggatcaaa gtcattgtga aataatcagt gcccagactc cattgaaaat 14940
agtgggttttc attctcatcg ccgtcctcaa tgctgctcgt cgaagatggc tgctaaagca 15000
tgccaatctg cagtttcaga atcaaatgca tgcggaaata atcagtgcgc agactccgtt 15060
gaaaatagtg gttttcatte tcatccccgt cctgaatgct gctcgtcgaa gatggctgct 15120
aaagcgtgcc aatctgcagt ttcagaatca aagtcattgt gaaataatca gtgccagac 15180
tccgttgga atagtggttt tcattctcat ccccgctctc aatgctgttc atcgaagatg 15240
gctgctaaag caggccaatc tgcactttca gaatcaaagt catgtggaaa taacaattgc 15300
tcagactcca ttcacaagag taattgtcat tctttaacta actctctagt atgttcttcc 15360
aagatgtctg ctccacaatg tcattctgct acttcaagca acaaatcatg tggaagtacc 15420
aagtgtccg acttcagtga taaaaaatgt tgtcaatccg acaaaattcc tcaagcgtgc 15480
tctaccaaga agtctgctcc aggggtgtcaa tctgcagttt ctgggtctaa atcatgtgga 15540
aatagcaagt gttcagactc aaaagacaat agtagccatc cttcacatcc cgatcatcaa 15600
acatgcatgt ctaagtgtg tgctccacaa agccaatctg caacttcaag ctccaggaca 15660
tgtggaaata caaagtgtc ggacaccaat agcaagaatt cttgttattc acaaccaaac 15720
tctgaatcat gctcttcaaa gatgtctggt ccatcatgca aaactgctaa ttcaggtacg 15780
accagattcc tcttcttgtt aatacccccc gaaccaaact ataggatcta aagtattatt 15840
tggccctctg tcggaggata taatgggttag ttaaacctga aatcatgtag tctatgaatt 15900
gcaaatctct agcatcgtga caaaattctt agatcatata caactttgag aatataggct 15960
gtcacagacc cttcttcata ctgcattaga ggagagcagc caatttttta ttcattgatt 16020
gaaacaaata aagttcttct tgaggtgtat ggagaggcta tgagaatcat ttgctgagta 16080
ggtttgagat tttcaggttc aaggtcatgc agaaataaga agtgccagga ccttgcaacc 16140
gagaacagtt ttcattcacc acttactaat ccactcagtg gggaaaagct ttcggagcag 16200
aaaagcttg atttagtccg aaaagataag gaatcaagtc atgatcttcg tcatggctgc 16260
tctgacgagg aacatgatca tacaaattta gacaaggcat atgacagttg tgccttacia 16320
gaatgttgtt attcggttca aggcaataaa actgatgtat cagaaactgg aatccaggaa 16380
actgctcatt gtgacagcac caatcaaaac tgccaaaactg caagttcagg taggcactac 16440
caaatcatat gatccaaagt gtcctccac cttcactcct acaataaatg ttcgatcaaa 16500

```

-continued

```

cttcataaga agatagcata tgcacgcaa atctctaaaa aaatgatgga taatgttact 16560
caccaactag tttgagatag aagtttaact gattgctata tatcggttaac tataaaaaac 16620
tacttggttaa ttagagctga gattttcagg atcgatgaca tgcggaaatg ataagatcct 16680
ggactctcta agcatccatg gttgtcattc gcatgataat ccactccacg aggagaacaa 16740
cctggagcag aaaatcttgg atgttggttg agaaggtata aaatcacctc atgctgtcgg 16800
tcatggctgt tcggacaagg aacacgatca ctacatcca gaaaaggcat atgacagtgtg 16860
tgcaacagat gattgttgtt tttcagttca agtccatggc attgacgacg tatcaaaaag 16920
tgaaattcaa gaaactgtct attgtgacag cacaagcag agcatggtca tctccagcag 16980
ctgcaaacat gaacaaaaag atcaggtaaa tcaactgtgga cttactcta aaactactcc 17040
aactgatgaa gaactagcca agctgggttag aagatgctgc gaatacaaac catgccacga 17100
cgtccgttct ggctgcagga agcatgctgc agaattgtgtt ccaaccgttc gatcaactat 17160
caatatctta cgggacaacc atcatcatta cctagactgc agtggtcgta aggtttgttc 17220
gctgttgtag aagagacaca tcggtggatg ctgtgacagc ttcagaaaag aatgttgtgc 17280
caagaagaac caccttgtag caagtctcgg aggaggttta tcagaaattg tcatagagta 17340
gatgcaatcc gaagtgtaca tatgttgtaa acttctacc tattttatct tcaagaagtt 17400
gagctgctaa tttgaacaaa gcaagggcga attctgcaga tatccatcac actggcggcc 17460
gctcgagcat gcatctagag ggcccaattc gccctatagt gagtcgtatt acaattcact 17520
ggccgctcgt ttacaacgct gtgactggga aaaccctggc gttaccaac ttaatcgct 17580
tgcagcacat cccctcttcg ccagctggcg taatagcgaa gaggccgca ccgatcgccc 17640
ttcccaacag ttgcgcagcc tatacgtacg gcagtttaag gtttacacct ataaaagaga 17700
gagccgttat cgtctgtttg tggatgtaca gagtgatatt attgacacgc cggggcgacg 17760
gatggtgatc cccctggcca gtgcacgtct gctgtcagat aaagtctccc gtgaacttta 17820
cccgtgtgtg catatcgggg atgaaagctg gcgcatgatg accaccgata tggccagtgt 17880
gccggtctcc gttatcgggg aagaagtggc tgatctcagc c 17921

```

<210> SEQ ID NO 2

<211> LENGTH: 1397

<212> TYPE: PRT

<213> ORGANISM: *Nicotiana tabacum*

<400> SEQUENCE: 2

```

Met Asn Glu Thr Lys Lys Leu Ser Lys Ser Tyr Phe Asp Val Leu Gly
1           5           10           15

Ile Cys Cys Thr Ser Glu Val Val Leu Val Glu Lys Val Leu Lys Asn
20          25          30

Leu Glu Gly Val Lys Glu Val Ser Val Ile Val Thr Thr Lys Thr Val
35          40          45

Ile Val Ile His Asp Ser Leu Leu Ile Ser Pro Gln Gln Ile Val Lys
50          55          60

Ala Leu Asn Gln Ala Arg Leu Glu Ala Ser Ile Arg Val Lys Gly Glu
65          70          75          80

Lys Asn Tyr Gln Lys Lys Trp Pro Ser Pro Phe Ala Ile Gly Ser Gly
85          90          95

Ile Leu Leu Gly Leu Ser Phe Leu Lys Tyr Phe Phe Ala Pro Phe Gln

```

-continued

100	105	110
Trp Leu Ala Leu Ala	Ala Val Ala Val Gly	Ile Pro Pro Ile Ile Phe
115	120	125
Arg Gly Val Ala Ala	Val Arg Asn Leu Thr	Leu Asp Ile Asn Ile Leu
130	135	140
Val Leu Ile Ala Val	Ala Gly Ser Ile Val	Leu His Asp Tyr Trp Glu
145	150	155 160
Ala Gly Thr Ile Val	Phe Leu Phe Ala Ile	Ala Glu Trp Leu Glu Ser
165	170	175
Arg Ala Ser His Lys	Ala Thr Ala Ala Met	Ser Ser Leu Val Asn Ile
180	185	190
Val Pro Pro Thr Ala	Val Leu Ala Glu Ser	Gly Glu Val Val Asn Val
195	200	205
Asp Glu Val Lys Val	Asn Ser Ile Leu Ala	Val Lys Ala Gly Glu Thr
210	215	220
Ile Pro Ile Asp Gly	Val Val Val Glu Gly	Glu Cys Asp Val Asp Glu
225	230	235 240
Lys Thr Leu Thr Gly	Glu Ser Phe Pro Val	Ser Lys Gln Arg Asp Ser
245	250	255
Thr Val Trp Ala Gly	Thr Thr Asn Leu Asn	Gly Tyr Ile Ser Val Lys
260	265	270
Thr Thr Ala Leu Ala	Glu Asp Cys Ala Val	Ala Arg Met Ala Gln Leu
275	280	285
Val Glu Asp Ala Gln	Asn Lys Lys Ser Lys	Thr Gln Arg Tyr Ile Asp
290	295	300
Lys Cys Ala Lys Tyr	Tyr Thr Pro Ala Ile	Val Ala Ile Ser Ala Ser
305	310	315 320
Leu Ala Ile Val Pro	Thr Ala Leu Arg Val	His Asn Arg Asn Glu Trp
325	330	335
Tyr Arg Leu Ala Leu	Val Thr Leu Val Ser	Ala Cys Pro Cys Ala Leu
340	345	350
Val Leu Ser Thr Pro	Val Ala Met Cys Cys	Ala Leu Ser Lys Ala Ala
355	360	365
Thr Ser Gly Leu Leu	Phe Lys Gly Ala Glu	Tyr Leu Glu Thr Leu Ala
370	375	380
Lys Ile Lys Ile Met	Ala Phe Asp Lys Thr	Gly Thr Ile Thr Lys Gly
385	390	395 400
Glu Phe Met Val Thr	Glu Phe Lys Ser Leu	Ile Asp Gly Phe Ser Leu
405	410	415
Asn Thr Leu Leu Tyr	Trp Val Ser Ser Ile	Glu Ser Lys Ser Gly His
420	425	430
Pro Met Ala Ala Ala	Leu Val Asp Tyr Ala	Gln Ser Asn Ser Val Glu
435	440	445
Pro Lys Pro Asp Arg	Val Glu Gln Phe Gln	Asn Phe Pro Gly Glu Gly
450	455	460
Ile Phe Gly Arg Ile	Asp Gly Met Glu Ile	Tyr Val Gly Asn Arg Lys
465	470	475 480
Ile Ser Ser Arg Ala	Gly Cys Thr Thr Val	Pro Glu Ile Glu Gly Asp
485	490	495
Ser Phe Lys Gly Lys	Ser Val Gly Tyr Ile	Phe Leu Gly Ser Ser Pro
500	505	510

-continued

Ala Gly Ile Phe Ser	Leu Ser Asp Val Cys	Arg Ile Gly Val Lys Glu
515	520	525
Ala Met Arg Glu Leu	Lys Gln Met Gly Ile	Lys Thr Ala Met Leu Thr
530	535	540
Gly Asp Cys Tyr Ala	Ala Ala Asn His Val	Gln Asp Gln Leu Gly Gly
545	550	555 560
Ala Leu Asp Glu Phe	Gln Ala Glu Leu Leu	Pro Glu Asp Lys Ala Thr
565	570	575
Ile Ile Lys Gly Phe	Gln Lys Glu Ala Pro	Thr Ala Met Ile Gly Asp
580	585	590
Gly Leu Asn Asp Ala	Pro Ala Leu Ala Thr	Ala Asp Ile Gly Ile Ser
595	600	605
Met Gly Ile Ser Gly	Ser Ala Leu Ala Lys	Glu Thr Gly His Val Ile
610	615	620
Leu Met Thr Asn Asp	Ile Gly Arg Ile Pro	Lys Ala Ala Arg Leu Ala
625	630	635 640
Arg Arg Val Arg Arg	Lys Ile Val Glu Asn	Met Ile Ile Ser Val Val
645	650	655
Thr Lys Ala Ala Ile	Val Ala Leu Ala Ile	Ala Gly Tyr Pro Leu Val
660	665	670
Trp Ala Ala Val Leu	Ala Asp Thr Gly Thr	Cys Leu Leu Val Ile Leu
675	680	685
Asn Ser Met Leu Leu	Leu Arg Gly Gly Thr	Arg Arg His Gly Lys Lys
690	695	700
Cys Trp Arg Ser Ser	Thr Pro Ser His Ala	Pro Thr Thr Lys Thr Lys
705	710	715 720
Leu His Val Ala Ser	Arg Lys Met Leu Pro	Ser Cys Val Ala Leu Ile
725	730	735
Leu Ser His Lys Arg	Asn Val Gln Val Asn	His Ala Arg Pro Arg Cys
740	745	750
Val Phe Gln Asp Val	Asn Leu Ser Pro Gln	Asp Gln Ser His Val Glu
755	760	765
Ile Ile Ser Ala Gln	Thr Pro Leu Lys Ile	Val Val Phe Ile Leu Ile
770	775	780
Ala Val Leu Asn Ala	Ala Arg Arg Arg Trp	Leu Leu Lys His Ala Asn
785	790	795 800
Leu Gln Phe Gln Asn	Gln Ser Arg Ala Glu	Ile Ile Ser Ala Gln Thr
805	810	815
Pro Leu Lys Ile Val	Val Phe Ile Leu Ile	Pro Val Leu Asn Ala Ala
820	825	830
Arg Arg Arg Trp Leu	Leu Lys Arg Ala Asn	Leu Gln Phe Gln Asn Gln
835	840	845
Ser His Val Glu Ile	Ile Ser Ala Gln Thr	Pro Leu Lys Asn Ser Gly
850	855	860
Phe His Ser His Pro	Arg Pro Gln Cys Cys	Ser Ser Lys Met Ala Ala
865	870	875 880
Lys Ala Gly Gln Ser	Ala Leu Ser Glu Ser	Lys Ser Cys Gly Asn Asn
885	890	895
Asn Cys Ser Asp Ser	Ile His Lys Ser Asn	Cys His Ser Leu Thr Asn
900	905	910

-continued

Ser Leu Val Cys Ser	Ser Lys Met Ser Ala	Pro Gln Cys His Ser Ala
915	920	925
Thr Ser Ser Asn Lys	Ser Cys Gly Ser Thr	Lys Cys Ser Asp Phe Ser
930	935	940
Asp Lys Lys Cys Cys	Gln Ser Asp Lys Ile	Pro Gln Thr Cys Ser Thr
945	950	955 960
Lys Lys Ser Ala Pro	Gly Cys Gln Ser Ala	Val Ser Gly Ser Lys Ser
965	970	975
Cys Gly Asn Ser Lys	Cys Ser Asp Ser Lys	Asp Asn Ser Ser His Pro
980	985	990
Ser His Pro Asp His	Gln Thr Cys Met Ser Lys	Leu Cys Ala Pro Gln
995	1000	1005
Ser Gln Ser Ala Thr	Ser Ser Ser Arg Thr	Cys Gly Asn Thr Lys
1010	1015	1020
Cys Ser Asp Thr Asn	Ser Lys Asn Ser Cys	Tyr Ser Gln Thr Asn
1025	1030	1035
Ser Glu Ser Cys Ser	Ser Lys Met Ser Gly	Pro Ser Cys Lys Thr
1040	1045	1050
Ala Asn Ser Gly Ser	Arg Ser Cys Arg Asn	Lys Lys Cys Gln Asp
1055	1060	1065
Ser Ala Thr Glu Asn	Ser Phe His Ser Pro	Leu Thr Asn Pro Leu
1070	1075	1080
Ser Gly Glu Lys Leu	Ser Glu Gln Lys Ser	Leu Asp Leu Val Arg
1085	1090	1095
Lys Asp Lys Glu Ser	Ser His Asp Leu Arg	His Gly Cys Ser Asp
1100	1105	1110
Glu Gly His Asp His	Thr Asn Leu Asp Lys	Ala Tyr Asp Ser Cys
1115	1120	1125
Ala Leu Gln Glu Cys	Cys Tyr Ser Val Gln	Gly Asn Lys Thr Asp
1130	1135	1140
Val Ser Glu Thr Gly	Ile Gln Glu Thr Ala	His Cys Asp Ser Thr
1145	1150	1155
Asn Gln Thr Cys Gln	Thr Ala Ser Ser Gly	Ser Met Thr Cys Gly
1160	1165	1170
Asn Asp Lys Ile Leu	Asp Ser Leu Ser Ile	His Gly Cys His Ser
1175	1180	1185
His Asp Asn Pro Leu	His Glu Glu Asn Asn	Leu Glu Gln Lys Ile
1190	1195	1200
Leu Asp Val Val Gly	Glu Gly Ile Lys Ser	Pro His Ala Val Gly
1205	1210	1215
His Gly Cys Ser Asp	Lys Glu His Asp His	Ser His Pro Glu Lys
1220	1225	1230
Ala Tyr Asp Ser Cys	Ala Thr Asp Asp Cys	Cys Phe Ser Val Gln
1235	1240	1245
Val His Gly Ile Asp	Asp Val Ser Lys Ser	Glu Ile Gln Glu Thr
1250	1255	1260
Ala His Cys Asp Ser	Thr Lys Gln Ser Met	Val Ile Ser Ser Ser
1265	1270	1275
Cys Lys His Glu Pro	Lys Asp Gln Val Asn	His Cys Gly Leu His
1280	1285	1290
Ser Lys Thr Thr Pro	Thr Asp Glu Glu Leu	Ala Lys Leu Val Arg

-continued

1295	1300	1305	
Arg Cys Cys Lys Tyr Lys Pro Cys His Asp Val Arg Ser Gly Cys			
1310	1315	1320	
Arg Lys His Ala Ala Glu Cys Gly Pro Thr Val Arg Ser Thr Ile			
1325	1330	1335	
Asn Ile Leu Arg Asp Asn His His His Tyr Leu Asp Cys Ser Gly			
1340	1345	1350	
Arg Lys Val Cys Ser Leu Leu Glu Lys Arg His Ile Gly Gly Cys			
1355	1360	1365	
Cys Asp Ser Phe Arg Lys Glu Cys Cys Ala Lys Lys Lys His Leu			
1370	1375	1380	
Gly Ala Ser Phe Gly Gly Gly Leu Ser Glu Ile Val Ile Glu			
1385	1390	1395	
 <210> SEQ ID NO 3			
<211> LENGTH: 4392			
<212> TYPE: DNA			
<213> ORGANISM: Nicotiana tabacum			
 <400> SEQUENCE: 3			
tcacctcttt ccaaaacaaa agtatatcca atatattcca gaactagaaa ttcctttttc	60		
atctattatc ttctcctcct ccttagagaa ggagaaaaat ggtggaaagt gaaaaaatga	120		
atgaaacaaa gaagttgagc aagagctatt ttgatgtttt gggaatttgc tgtacttcag	180		
aagttgttct agttgaaaaa gttctcaaga atcttgaagg ggttaaagag gtttcagtaa	240		
ttgtcacaac aaagactgtc attgttatcc atgattctct tctcatttct ccgcaacaaa	300		
ttgttaaagc attgaatcaa gcaagattag aagcaagcat aagagtgaag ggagagaaaa	360		
actaccaaaa gaaatggcca agtccatttg caattggcag tggaatattg cttggactct	420		
catttttgaa gtactttttt gcacctttcc aatgggttagc acttgcagct gttgcagttg	480		
ggattcctcc aattattttt agaggtgtgg ctgccgtgcg aaacctcact cttgacatca	540		
acattcttgt ttaatatgca gtggctggat caattgtttt acacgattat tgggaagctg	600		
gtactattgt cttcttattc gccattgcag aatggctaga gtcaagggca agtcacaagg	660		
ctaccgtgc tatgtcatca ctggtcaata tagtcctcc aacagcagtt ttagctgaaa	720		
gcggagaagt cgtaaagtgt gatgaagtca aggtgaatag cattcttgct gtgaaagctg	780		
gtgaaactat acctattgat ggagttgtag tggaagggga atgtgacgtg gacgagaaaa	840		
cactgacagg cgagtcgttt ccagtttcta agcaaagaga ttcaacggtc tgggctggca	900		
ctacaaatct aaatggctat atcagtggtt agactacggc tttggctgaa gattgtgctg	960		
tggctaggat ggcacagctt gtcgaagatg ctgagaacaa gaaatcaaaa acccaaagat	1020		
acatcgacaa gtgtgctaaa tattatacac cagcaattgt ggctatatca gcttcttttg	1080		
caattgttcc tactgcatta agagttcaca atcgaaatga atggtatcgc ttggcttttg	1140		
tcacattggt gagtgcattg ccgtgtgcac ttgttctatc tacaccagtt gccatgtgtt	1200		
gcgcactttc aaaagcagca acgtccggtc ttctgtttta aggagcagag taccttgaga	1260		
ctctagctaa aatcaaaatc atggcttttg acaaaacagg gactataact aaaggagaat	1320		
ttatggtgac cgagttcaag tctctgattg atggttttag tctcaatata ctgctttact	1380		
gggtttcaag cattgagagc aagtcaggtc atccgatggc agccgctctg gtggactatg	1440		

-continued

cacaatcaaa	ttccgttgag	ccaaagcctg	atagagttga	gcagtttcaa	aattttcctg	1500
gtgaagggat	atttggaaga	attgatggaa	tggaaatcta	tgtcgggaat	aggaaaattt	1560
cttcaagagc	tggatgtacc	acagtaccag	aaatagaggg	tgatagtttc	aaaggaaagt	1620
ctgttgata	catatttttg	ggatcatctc	cagctggaat	tttcagtcctt	tccgatgttt	1680
gtcgaattgg	tgtaaaagaa	gcaatgagag	aactgaagca	gatgggtatc	aaaaccgcga	1740
tgcttactgg	tgattgttat	gcagctgcca	accatgtgca	ggatcagtta	ggtggagctt	1800
tggatgaatt	tcaagcagaa	ctcctaccag	aggacaaggc	aacaatcatc	aagggttttc	1860
agaaggaagc	tccaacagcg	atgataggcg	acggccttaa	tgatgctcct	gcattagcaa	1920
cagctgacat	tggcatctca	atgggcatct	ctgggtcagc	tctcgctaaa	gaaacaggcc	1980
atgttatact	aatgacaaat	gacatcggaa	gaataccgaa	agctgcacgt	cttgctagaa	2040
gagttcgaag	gaagattgtt	gagaatatga	ttatatcagt	cgttacaaag	gctgccatag	2100
ttgcattggc	aatagcaggt	tatccattgg	tttgggctgc	tgctctcgca	gatactggga	2160
catgcttgct	agtgattttg	aacagcatgc	tacttctacg	aggaggcaca	cgcagacatg	2220
ggaaaaaatg	ttggagatct	tctactcctt	cgcagtctcc	caccacaaag	acaaagcttc	2280
atgttgcaag	tcggaataatg	ctccccagct	gtgttgctct	gatattgagt	cacaaaagaa	2340
atgtacaagt	caatcatgct	cgtccgaggt	gtgtgttcca	agatgtcaac	ctgtctcttc	2400
aggatcaaaag	tcatgtggaa	ataatcagtg	cccagactcc	attgaaaata	gtggttttca	2460
ttctcatcgc	cgtccccaat	gctgctcgtc	gaagatggct	gctaaagcat	gccaatctgc	2520
agtttcagaa	tcaaagtcgt	gcggaaataa	tcagtgccca	gactccgttg	aaaatagtgg	2580
ttttcattct	catccccgtc	ctgaatgctg	ctcgtcgaag	atggctgcta	aagcgtgcca	2640
atctgcagtt	tcagaatcaa	agtcattgtg	aaataatcag	tgcccagact	ccgttgaaaa	2700
atagtgggtt	tcattctcat	ccccgtcttc	aatgctgttc	atcgaagatg	gctgctaaag	2760
caggccaatc	tgcactttca	gaatcaaagt	catgtggaaa	taacaattgc	tcagactcca	2820
ttcacaagag	taattgtcat	tctttaacta	actctctagt	atgttcttcc	aagatgtctg	2880
ctccacaatg	tcattctgct	acttcaagca	acaaatcatg	tggaaagtacc	aagtgtctcg	2940
acttcagtga	caaaaaatgt	tgtcaatccg	acaaaattcc	tcaaacgtgc	tctaccaaga	3000
agtctgctcc	aggatgtcaa	tctgcagttt	ctgggtctaa	atcatgtgga	aatagcaagt	3060
gttcagactc	aaaagacaat	agtagccatc	cttcacatcc	cgatcatcaa	acatgcatgt	3120
ctaagtgtg	tgctccacaa	agccaatctg	caacttcaag	ctccaggaca	tgtggaaata	3180
caaaagtctc	ggacaccaat	agcaagaatt	cttggttatc	acaaaccaac	tctgaatcat	3240
gctcttcaaa	gatgtctggt	ccatcatgca	aaactgctaa	ttcaggttca	aggatcatgca	3300
gaaataagaa	gtgccaggac	tctgcaaccg	agaacagttt	tcattcacca	cttactaatc	3360
cactcagtg	ggaaaagctt	tgggagcaga	aaagcttgga	tttagtccga	aaagataagg	3420
aatcaagtca	tgatcttctg	catggctgct	ctgacgaggg	acatgatcat	acaaatttag	3480
acaaggcata	tgacagttgt	gccttacaag	aatgttggtt	ttcgggtcaa	ggcaataaaa	3540
ctgatgtatc	agaaactgga	atccaggaaa	ctgctcattg	tgacagcacc	aatcaaacat	3600
gccaaactgc	aagttcagga	tcgatgacat	gcggaaatga	taagatcctg	gactctctaa	3660
gcatccatgg	ttgtcattcg	catgataatc	cactccacga	ggagaacaac	ttggagcaga	3720

-continued

```

aaatcttgga tgttggtgga gaaggtataa aatcacctca tgctgtcggt catggctggt 3780
cggacaagga acacgatcac tcacatccag aaaaggcata tgacagtgtg gcaacagatg 3840
attgttggtt ttcagttcaa gtccatggca ttgacgacgt atcaaaaagt gaaattcaag 3900
aaactgctca ttgtgacagc acaaagcaga gcatgggtcat ctccagcagc tgcaaacatg 3960
aaccaaaaga tcaggtaaat cactgtggac ttcactctaa aactactcca actgatgaag 4020
aactagccaa gctggttaga agatgctgca aatacaaac atgccacgac gtccgttctg 4080
gctgcaggaa gcatgctgca gaatgtggtc caaccgttcg atcaaccatc aatatcttac 4140
gggacaacca tcatcattac ctgactgca gtggtcgtaa gggttggtcg ctgttgagaga 4200
agagacacat cgggtggatgc tgtgacagct tcagaaaaga atgttggtgcc aagaaaaaac 4260
accttgagc aagttttgga ggaggtttat cagaaattgt catagagtag atgcaatccg 4320
aagtgtacat atgttgtaaa cttcctacct attttatctt caagaagttg agctgctaata 4380
ttgaacaaag ca 4392

```

```

<210> SEQ ID NO 4
<211> LENGTH: 723
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(723)

```

```

<400> SEQUENCE: 4

```

```

gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca ggtatccggt 60
aagcggcagg gtcggaacag gagagcgac gagggagctt ccagggggaa acgcctggta 120
tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc 180
gtcagggggg cggagcctat ggaaaaacgc cagcaacgcg gcctttttac ggttcctggc 240
cttttgctgg ccttttgctc acatgttctt tctgctgta tccctgatt ctgtggataa 300
ccgtattacc gcctttgagt gagctgatac cgctcgccgc agccgaacga ccgagcgcag 360
cgagtcagtg agcgaggaag cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg 420
ttggccgatt cattaatgca gctggcacga caggtttccc gactggaaag cgggcagtga 480
gcgcaacgca attaatgtga gtagctcac tcattaggea cccaggett tacactttat 540
gcttcgggct cgtatgttgt gtggaattgt gagcgataa caatttcaca caggaaacag 600
ctatgaccat gattacgcca agctatttag gtgacgcgtt agaatactca agctatgcat 660
caagcttggt accgagctcg gatccactag taacggccgc cagtgtgctg gaattcgccc 720
tta 723

```

```

<210> SEQ ID NO 5
<211> LENGTH: 303
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(303)

```

```

<400> SEQUENCE: 5

```

```

tca cct ctt tcc aaa aca aaa gta tat cca ata tat tcc aga act aga 48
Ser Pro Leu Ser Lys Thr Lys Val Tyr Pro Ile Tyr Ser Arg Thr Arg
1          5          10          15

```

-continued

```

aat tcc ttt ttc atc tat tat ctt ctc ctc ctc ctt aga gaa gga gaa    96
Asn Ser Phe Phe Ile Tyr Tyr Leu Leu Leu Leu Leu Arg Glu Gly Glu
20                25                30

aaa tgg tgg aaa gtg aaa aaa tga atg aaa caa aga agt tga gca aga    144
Lys Trp Trp Lys Val Lys Lys      Met Lys Gln Arg Ser      Ala Arg
35                40                45

gct att ttg atg ttt tgg gaa ttt gct gta ctt cag aag ttg ttc tag    192
Ala Ile Leu Met Phe Trp Glu Phe Ala Val Leu Gln Lys Leu Phe
50                55                60

ttg aaa aaa ttc tca aga atc ttg aag ggg tta aag agg ttt cag taa    240
Leu Lys Lys Phe Ser Arg Ile Leu Lys Gly Leu Lys Arg Phe Gln
65                70                75

ttg tca caa caa aga ctg tca ttg tta ttc atg att ctc ttc tca ttt    288
Leu Ser Gln Gln Arg Leu Ser Leu Leu Phe Met Ile Leu Phe Ser Phe
80                85                90

ctc cgc aac aaa ttg    303
Leu Arg Asn Lys Leu
95

```

```

<210> SEQ ID NO 6
<211> LENGTH: 2218
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(2218)

```

```

<400> SEQUENCE: 6

```

```

gtaagaaaga tagttacaca cctttattac tctctcagtt ctatttttta cgtgatacta    60
ttttcttttt aatatgttct gaaaagaacg ttaccttttt atattattgat gtaatttcac    120
tttaaaacttc atattttttt tcttaaatag tattgtttta tagtcataaa aatattatta    180
taagtctttc ttaagttttg tggcttggtt aataaattca cataaaatga aataaagcgg    240
taggagtacc attcttcacg ttttcttaat taactgacat ttcctttctt ttttgtaagt    300
ttatatatat taagtaaagt gatttttctc ttaagaaatc gccaaaaaaa aaaaggaaag    360
agaagagaag agacaagaag ggggagacaa atggaatcag aaacgggttt gatttgttca    420
agtgatattt ggtggtagtt gtttcaagca ctgttctttc tgtttgttgg catttgctag    480
aattaagtgt atattattaat ttgggaaaaa ctttaaagtg gttattttta gttttattag    540
attgagtaac caagacggaa aaaacatgaa ctatttttct tttgaaattt ctagtcaaga    600
acatgcataa aaattctctt ttaaaacgac tctcataaaa attcatgtgg tcgagtttac    660
gcagcacatg gacctatagt ctccgcctaa ctaagtattt taagtatgta ttttctaaaa    720
ttcatctaatt ttttctgttt ggccgcacatg ctccacaaaa gatgaattgt gcatttgttt    780
gaatattgag ttattactca aaggaatatg gatcaattcc actttttttc tcttttcttt    840
aatattgtac ccatacttta aaaattctag ctccgcctcc gacttgccca tcttatccgc    900
cccagccctc agtgggatag agtaggtgag gagtatagtt tataaatggt ttttctctaa    960
caattaggtt ttggaataaa ttctgagatt caatgggtcta tttgaaacct gtaattacta    1020
ctccctccgt ttcataattg atgattactt tcttttttag tctattccaa aacaaataac    1080
acatttctaa atttggaagt aattcaattt taaaatcttt cattttaccc atttacctta    1140
atgaatgttt ttatagccac acaaatgtca tgaccccaca aattttttac ccttaaaact    1200
tttaagacca caaattttta aagtttcttc ttttttctta aactatatgc caagtaaaac    1260

```

-continued

```

taactcatct tgaacggag gagtactata agaatagtta tatatgattt ggccccacaa 1320
attacataat tgtggcaaga aatccaagtt tctagttggt gataatttag tgatggaggc 1380
gatctctggt gaaccgtag agaaaatatt ttgactcatg tcgctttgca tctactgtgt 1440
tgaagagtag ttgctaggaa ctaacaaagt aattgtagtt tccttgtttt ttttttttcc 1500
tttcttgatg aattactgta ctttattttc ccattttttt aacgtctggt acttatgggc 1560
agttttcaag tgtgaaaaat tagccagaaa gaataattca acaaacttaa ctttttcttt 1620
tcttcattct tgcgatcttt cctaaatata tttagtagag tctcttttct acttcagtct 1680
tattgcattt ctgacacacc taacagtagt ggtacataaa ttggacctag ctaacaaaag 1740
gtaactctta tctgcaagtt taattggtac agtaacatga aacttggtat aatcttttaa 1800
attgcagccg ataatatatg tatgttttac attggttggtg tgtatataat gtaaattctt 1860
tacagcatgt ttgactaacc tgcaaaaatt agtttttttt ttttcaaaag tatttttggg 1920
gagaagcagt ttgtgttttg ctaagtaatt tgaaaaatac ttctgagcaa caattagtgt 1980
ttgtccaagc ttttaaaaac tgcttttaat tgtatttttg tcaaaagagc tttttaaaaa 2040
agtatttttt agagagaaac tacttttttc tgcttctcca aaattgtttc tgcttctcct 2100
caaaaacact ttttttcctt ctaaaagctt gtacaaacac ttcaactaaa aaaaatatat 2160
tagcacttat attatcctta taattattga agttaccatc tatcttttgt ggatgtag 2218

```

```

<210> SEQ ID NO 7
<211> LENGTH: 258
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(258)

```

```

<400> SEQUENCE: 7

```

```

tta aag cat tga atc aag caa gat tag aag caa gca taa gag tga aag      48
Leu Lys His   Ile Lys Gln Asp   Lys Gln Ala   Glu   Lys
1             5             10
gag aga aaa act acc aaa aga aat ggc caa gtc cat ttg caa ttg gca      96
Glu Arg Lys Thr Thr Lys Arg Asn Gly Gln Val His Leu Gln Leu Ala
15          20          25
gtg aaa tat tgc ttg gac tct cat ctt tga agt act ttt ttg cac ctt     144
Val Lys Tyr Cys Leu Asp Ser His Leu   Ser Thr Phe Leu His Leu
30          35          40
tcc aat ggt tag cac ttg cag ctg ttg cag ttg gga ttc ctc caa tta     192
Ser Asn Gly   His Leu Gln Leu Leu Gln Leu Gly Phe Leu Gln Leu
45          50          55
ttt tta gag gtg tgg ctg ccg tgc gaa acc tca ctc ttg aca tcg aca     240
Phe Leu Glu Val Trp Leu Pro Cys Glu Thr Ser Leu Leu Thr Ser Thr
60          65          70
ttc ttg ttt taa tag cag                                           258
Phe Leu Phe   Gln
75

```

```

<210> SEQ ID NO 8
<211> LENGTH: 3861
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(3861)

```

-continued

<400> SEQUENCE: 8

gtacttttgc	tttccttttc	tttcttttta	actttttgaa	aagcaaaaag	agatactccc	60
ttttgtccca	atttatatgg	cggattttta	ttaaacaaaa	aatctaaga	aaaatatgaa	120
tactttaaaa	atttatgggt	tgaaataaat	cttaggcata	taaatacaag	gttaaaataa	180
taatttttaa	gttaaaaaat	tcaagaaaga	aaggaagact	tctaaaattt	gtggcaagaa	240
ataaatctta	gagatttggt	ttataaatca	tctcactaag	gctaaaataa	taatttttaa	300
gttaaattac	ttttaattat	gaaaagggtg	tagatttttg	ggaaaagata	gaaaagaaaa	360
ctgtgctata	taaattgaga	cagagagagt	aatttaccat	aagagtattt	gaagttgatt	420
tggttataga	attaacatgt	gctcctaact	tagctttgtg	attgaaattt	ggctattgct	480
ttaatcttct	agctaattgg	atgcttctaa	tgtgatgctt	gaatccttat	tcttgataat	540
gctttgtgat	tcattatttc	attaaaaagc	atgcaccata	gggcatgcaa	ttaaatattg	600
tatttaagaa	tgcaacttta	actaaccac	aagattggag	tgggagggat	gattcttggg	660
tgtagtagtc	taatatggga	ccaccctagt	gactttaaat	aacgaaagca	tgaaaaacaa	720
ttattggatg	cttgatatac	tttatgttat	aatattttca	ggtagcctag	aactacacaa	780
ataaacttct	ttttccatat	tggaatagga	tgatttagat	ttcaagatgg	aattagtgat	840
tctatcacag	attatctttt	atttaatgat	atcaaatatg	aaaagagaaa	aaaaaaaggt	900
gtctaggaaa	tagtcagaaa	gatagtatga	catatattat	tgatacaaat	taatcagctc	960
aacacaaaag	ctagcatata	tttaacaaat	agttatagac	ttgtagtgtt	gtacctttag	1020
ttaagagact	aaacactacc	aaaagggtgc	gatggagcag	taggtactcc	tctattgtca	1080
tccaggcgcc	ttgatataaa	ttaattagat	cccaaaacga	ggatcagata	ccaaataaaa	1140
aacaaaaaga	aactatcttt	agtgtttgtt	gatttacatg	gcgtagaaaa	ataaaggaca	1200
atataataaa	aaccttgtaa	tatcagttaa	gatgccttaa	tttcaaatca	gtggagtata	1260
ttccgacttc	cgtattaatt	tcgctctaata	caagtctttt	aaaagattaa	ataatgaaaa	1320
gtttttatag	tcacgaaaac	attatgacca	caaaattttg	tcctttaagc	tttttagacc	1380
ataaatttca	aaaaaaaaaa	tttgactgaa	atatttaata	ccacatatatt	ttaaaaaaatt	1440
atttttttct	tacatttagt	gtcaagtaaa	attagatgtg	ataggaaaaa	agaatttcga	1500
tttcgtctag	taaggaaagt	acaaaattgc	ttatgtaact	cgtaatatat	cagaaatggg	1560
ttcccagggt	taaagcacia	acaacggagc	ccacatcagt	aatgtgtaaa	tgtgtaacgt	1620
ttactctctt	tttttctcca	aataattgca	aagaaaatga	ctttcctgca	agtcctttctt	1680
gcccctttta	tagaagggtga	cgttgactcg	taccagattt	tcttgtcata	gtattaatca	1740
gtagtaatat	cagtcacggc	gagggttttaa	atttagtcta	ttagtaacgt	aaataatttt	1800
taaattgtca	tatttagtta	atctaaaata	tcatgcaatt	ttttatatca	agtttgatat	1860
gttgacgtaa	ataatagaac	aataatttat	aaacttgatt	aaattgcatt	gataataaat	1920
caaaatatta	tttgaacgat	taaaatctca	cagttaattt	tgcatcttgc	ataaaaaaat	1980
atccatcata	tttccttgat	tatcatttga	tgcctgtcta	gtttctttta	aaaaaattta	2040
gaatttatat	caaatatgat	tttttaatta	ctcgaactta	caagcagact	aagtttgata	2100
ttttcctaata	tcaacgatata	acggttacta	cgggaaggcat	ttactagaaa	tactctgaga	2160
tgttactgca	attattatta	ttattatttt	aaaaagagaa	aaaaataaact	tttaaagctc	2220

-continued

```

catgtgaaat tatgtatatt ttattatagc atgaagtgac cccatttttt atctcataaa 2280
taacattgat cccatatttt tctactgtat catcactatc atgaaaaata catctagatt 2340
actgagatgt ttattggcag tagtatctac agatacaaca gatgcttcta gttctattga 2400
tgttttcatt ttgacaaaaa ttaattgag aaagcaaaag attttgcaag aattcttagg 2460
gttttattca aaaaaaaaaa agaattctta gaaatataag ttttggcaat taaataattc 2520
cagtaattgg gaaaaaacac ttgaataggc tatagaagta aaagaaactt ctatatatta 2580
ccaggcagca gagtttccca aaatcctttt ttctaaaaaa aaaatagtag aaaatgagca 2640
atgtaatttc ttaagtaac attctctatt tagtaaaatg tccatttttc taatgaggta 2700
aaagcaatag caaataaaag aaagtttatt cctttttttc gaggggtgtg ccgaccaagg 2760
cttaaataga taggaataat cacctaatta agaaaaacta cctatcaatt tttgtctcgg 2820
ttggatttca acatgatacc tcatagtctt acgtacgccc acttcaatta ccactaggaa 2880
caccgtttgg tgcaggaaaa ttttggctat aaacttcaat ttaagcctt catataaaca 2940
ataaaagagc taattagcag ttaacagtcg agttaatata gctgaataat gcagttcaac 3000
taaatatgta tggaaatggt gaaaagcaca aagggtgactt atccttaggt actttatagc 3060
tttatctttt aaaaagtagt agcaagatat gatatgattc cttgaagaag aaaaggtcac 3120
tgtgactcct tatttctatc agtacctgtt tgaataaaat tggctaagaa agttgtaaaa 3180
tggactagca ggacaagaat ctcaatttgg tgttgcttta ccactcttag agtgacagca 3240
atcaaaaacc caaaaattaa ataaataata agaaaagaaa aagatgagtt attggaaagg 3300
gtaaaatgaa tgaagataag aatactccat tcagtcctaaa atattaagct agcatttgga 3360
catagatttg gttgaagcct gaagaaaaga aaaaagagtt tttgaagatt atatgaaaaa 3420
taatttttga aagttaaaaat tgtgttaggg aagttttgtg attcaaaaac tactctagaa 3480
ctgttttttg gatttaataa ttttcttttc aacatgtgcc aaaaaatgat taaaatctat 3540
aaacgaacac ataattgaaa aaagctctca aattttatga ccaaacagga gcttattgtt 3600
ttggcaaatt gaatttgttt cgaaatactt gatgattaag gagagacaaa ggatagtata 3660
gacaaaaggc tattagtatt tgtaattgag gctattaaat gtctttttta agcgatgtgc 3720
aaaaacctta aaaaagacga atttatatga ttatattaga agtagtatta aattttaatg 3780
gaacttagtg caatcatttt acaagggcat agtggttaag ctagaaatgc tgattcttat 3840
agctggcttt gtcattgtgca g 3861

```

```

<210> SEQ ID NO 9
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(98)

```

```

<400> SEQUENCE: 9

```

```

tgg ctg gat caa ttg ttt tac acg att att ggg aag ctg gta cta ttg 48
Trp Leu Asp Gln Leu Phe Tyr Thr Ile Ile Gly Lys Leu Val Leu Leu
1 5 10 15

tct tct tat tcg cca ttg cag aat ggc tag agt caa ggg caa gtc aca 96
Ser Ser Tyr Ser Pro Leu Gln Asn Gly Ser Gln Gly Gln Val Thr
20 25 30

```

-continued

ag

98

<210> SEQ ID NO 10
 <211> LENGTH: 4063
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: Intron
 <222> LOCATION: (1)..(4063)

<400> SEQUENCE: 10

```

gtttgtgaat ttctgtccct gtttaatttc cttgcacaag aaatggctag tgtacatcta    60
ctttcctagg atgacagtaa tgtgtgtttt ttttatttat tgtgatacca aaagtatgaa    120
cacctaattt ttctcaggat acatatatttc cactttggcc aaactgactt taacaagttg    180
tctcttggat tacgccacat cgaacctaaa agtggttgca ctaccatgtc aacttgtgct    240
gtcttatagg gatttgagac attttgtcta ggggtgcata tggacccctt cttcatgagc    300
tcgatatgac aaaagcttgc tagtcattct acttgacctg cctcattagc ccgccctacg    360
catgggcatc acaaaccagg ccacacgtag gaattgagtc tcctccctgc gggttcacca    420
cagaccttac agttttgtcg cttgggtaat gccacaacta gcccgataac gtgcttacca    480
tgtgaacttg taccatatac tacctagcac ttcaccttcc tctccctttg cgacgcgcct    540
aaaaatgcat atgcaccgtt ttttccgaag tttgctagtc tagaagccta ttagttctta    600
ggcttaacct gcctcattaa ccactctcca tcataggcgt caccttaatt taaaagaatt    660
tgttcttaga aggccttaac ttaaaccaga aatcagttca gctgtcttct gttcttttac    720
ctcaataaca ttgtataatg gtaaggacta aactgcagct ctcttgtgga tgagtggatt    780
aaatttcatt ctgaaaatta atttacttca cagctctatt tggagaaaat aaagattaaa    840
tattgtgaga atgcacggga gaaaaatatt atattgatta aagtgttgta caaccctatt    900
tatatacagt aattatataa taatatgtat ctacttcccg atgtgggaca ctaaatatga    960
ctaactactt aacacttctc ctcaagccgg tgcataataa tcatacgtac cgagcttgtt   1020
acagatgtaa ccaatacgag aaccagtaag agacttagtg aaaatatattg ctagctaatac   1080
attcgacttt acaaactttg taacaatata tcctgagagt atcttttctc tgacaaagtg   1140
acagtcgac tcaatgtgtt tagtcctctc atggaatacc ggatttgacg caatatgaag   1200
agcagcttga ttatcacaca ccagttccat cttattgatt tctccgaatt tcaactcctt   1260
gagcaactgc ttgateccaca ctagctcaca cgttgccata gccatggccc gatattcggc   1320
ttcggtgcta gatcgagcaa ctacattctg tttcttgctc ttccaagaga ccaaattacc   1380
tcctactaga acacaatata cagacgtaga acgtttatca gaagggtgac ctgcccaatc   1440
agcatttggt taccctgtaa tctgctcgta gcctcgatcc tcgaatagta accctttgcc   1500
tggagctgac tttatatata gaagaatgcg aacaactgca tcccagtac taccacaggg   1560
agaatccata aactgactta caaactcac cggaaaagaa atgtctgggt tagtcaccgt   1620
gaggtaattc aatttgccaa ccaacotccg atatctcata ggatctctaa gaggtcctcc   1680
ctgtccattg cagaagctta gcattcggat ccataggagt gtcaacaggt ctacatccca   1740
tcattccagt cttctcaaga atgtctaagg catacttccg ctgtgaaata acaataactg   1800
agctagactg agcgacctca atacctggaa aatacttcaa tctgcctaga tccttagtct   1860
ggaagtgcc aagagatgt tgcttcagat tagtaatacc atcctgatca ttgtcagtaa   1920
  
```

-continued

taacaatatc	gtctacataa	accaccatag	taaggttatg	aacaaaacac	gtgcttctaa	1980
agacacgggg	tggaagagag	aacaaaggta	agtggggaaa	caggacagag	aatggaactt	2040
gattctggat	agctgaagat	gacatacgat	taataagata	gcaagatgta	agaactgcat	2100
cacccccaaa	atgcaacgga	gcatgagatt	gtatgagtag	agtacgagta	gtttcaataa	2160
gatgtctact	ctttctttca	gtaccccat	tttgttgaga	tgtgtacgga	caagatgttt	2220
gatgaataat	cccatgagag	ttcataaact	tctgaaatgg	ggaagacaaa	tactctaggg	2280
cattatcact	acgaaatgtg	cggatagaaa	ccccaaattg	attttgaatt	tcagcgtgga	2340
aggtctgaaa	aatagaaaat	agctcagacc	gatttttcat	caaaaatc	caagtgcacc	2400
tggaataatc	atcaatgaaa	ctgacaaagt	agcggaatcc	caaggtagaa	ctgacccgac	2460
taggacccca	aacatctgaa	tggaaccaag	taaaagggtga	cgtgagggga	aatgggagtg	2520
ggtatactta	ccgagctgtc	atgactcata	ctctagagtg	gacaagtgag	ataaaccagg	2580
tatcattttc	taaagttttg	acaaactggg	atgtcccaac	cgtttatgta	ataaatctgg	2640
tgaatcggta	acaagacaag	ttattaaagg	aagacaagat	gcgagtccat	atgattttgc	2700
aaggataaag	taataaaatc	catctaattc	atgtctgata	ccaatgatcc	gccccgtact	2760
atgtttctct	ataaaaaata	ggtcacaaag	aaataaaaca	gcgcatttaa	gtgatttggc	2820
taagcgacta	acggctatga	gattaaaagg	actattggga	acataaagga	ctgaatctaa	2880
aggtgaggaa	ggaagtgggc	ctgcttgacc	tattgcagtt	gccatggttt	gagactcatt	2940
ggccattgtg	actgttagaa	gagattgaga	atatgaatac	tggtgaaaag	agatttgtaa	3000
ccagaaatat	gatcagatgc	atttgaatca	atgacccaag	actcaaaggt	ttaagattcg	3060
gagacacaag	tcacgtact	atctgtttga	acaatggaag	ctatccctga	agatgtctgt	3120
ttacatgcct	tgtactgaag	gaactcagta	taatccgata	aagaaacat	ctggattgga	3180
ttcaacgaat	tgatccaac	agattgtgag	ttaaaggctt	ctgatacgaa	tgtcattata	3240
atgttgtgcc	gaaaaacaa	aaaattcctg	gaaattacta	ttcacgccgg	aaaaatataa	3300
aagtgatctg	aatttgattt	aaattggatg	ggtagtctcg	tatttgcaag	gagaagacac	3360
tgccctgaag	gaattttacc	aaattctggc	cggaaattgc	ctcatgtgcg	gcgtgtgggc	3420
gtcagaactt	cgtcgaaaa	attcttccgg	cggcgcgtga	ggcgcgtgt	agcctttttt	3480
gccagagatt	ttttaatagg	ttggtcgctg	agctctgaac	tacttcccg	tggtgttacc	3540
ttttgcacaa	cactgacaga	tagtatgatt	cttgccgaca	gacctatttt	tgccggaaaa	3600
gagcttccgg	ttgactgttt	tcttttcccg	gagtcgctgg	aatttatgca	ctacgataaa	3660
tttctcacgg	ttgctctgat	accatgtgag	aatgcacggg	agaaaattat	tatattgatt	3720
aaagtattgt	acaaccttat	ttatatacag	taattacata	ataataggta	tctacttccc	3780
gatgtgggac	actaaacatg	actaactact	taacaaatat	ggatttgga	tttagtctct	3840
ttgacataaa	cgacataaag	ctatgcttat	cttttcttac	ttttttagca	atgctaaata	3900
gtaggtccta	actacaaact	ttatagcaca	ctgaaaatta	ccaaaatata	gagatggcca	3960
atgaagggtt	tgtctgctaa	cataactctg	tgtctttatt	ttctcactga	tattgtatat	4020
ggataaagca	ttctgataaa	tgaaaacctt	tatggttatg	tag		4063

<210> SEQ ID NO 11

<211> LENGTH: 256

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(256)

<400> SEQUENCE: 11

gct acc gcc gct atg tca tca ctg gtc aat ata gtc cct cca aca gca      48
Ala Thr Ala Ala Met Ser Ser Leu Val Asn Ile Val Pro Pro Thr Ala
1          5          10          15

ggt tta gcg gaa agc gga gaa gtc gta aat gtt gat gaa gtc aag gtg      96
Val Leu Ala Glu Ser Gly Glu Val Val Asn Val Asp Glu Val Lys Val
20         25         30

aat agc att ctt gct gtg aaa gct ggt gaa act ata cct att gat gga     144
Asn Ser Ile Leu Ala Val Lys Ala Gly Glu Thr Ile Pro Ile Asp Gly
35         40         45

ggt gta gtg gaa ggg gaa tgt gac gtg gac gag aaa aca ctg aca ggc     192
Val Val Val Glu Gly Glu Cys Asp Val Asp Glu Lys Thr Leu Thr Gly
50         55         60

gag tcg ttt cca gtt tct aag caa aga gat tca acg gtc tgg gct ggc     240
Glu Ser Phe Pro Val Ser Lys Gln Arg Asp Ser Thr Val Trp Ala Gly
65         70         75         80

gct aca aat cta aat g
Ala Thr Asn Leu Asn
85

```

```

<210> SEQ ID NO 12
<211> LENGTH: 85
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(85)

<400> SEQUENCE: 12

gtagtatagt attttcttcat gcttcattta tttagtgtcg aaacttcaag tattgtttgt      60
taatgttatt tgctcaattc ttcag
85

```

```

<210> SEQ ID NO 13
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(142)

<400> SEQUENCE: 13

gct ata tca gtg tta aga cta cgg ctt tgg ctg aag att gtg cgg tgg      48
Ala Ile Ser Val Leu Arg Leu Arg Leu Trp Leu Lys Ile Val Arg Trp
1          5          10          15

cta gga tgg cac agc ctg tcg aag atg ctc aga aca aga aat caa aaa      96
Leu Gly Trp His Ser Leu Ser Lys Met Leu Arg Thr Arg Asn Gln Lys
20         25         30

ccc aaa gat aca tcg aca agt gtg cta aat att ata cac cag gct a     142
Pro Lys Asp Thr Ser Thr Ser Val Leu Asn Ile Ile His Gln Ala
35         40         45

```

```

<210> SEQ ID NO 14
<211> LENGTH: 309
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:

```

-continued

```

<221> NAME/KEY: Intron
<222> LOCATION: (1)..(309)

<400> SEQUENCE: 14

gtgaatctta tgttggtgcca catcaagtca aaaaatgcac gtaccgtgtg aacttgttct    60
ttgtcttatg aatcacgtca ctatcctctc ccttttcgat atgagatttc cctaagggtg    120
catatgaatc ccttcttcgg aagcttgcca gcataggagt ctatcagtcc ttctacttga    180
cccgccctct cagcctgcct gcagtcattg gcgtcgact actatattgc tctttcgttt    240
aaaaactttt atttctaata cttccctgct ctttgtgtat gtctaatttc gactgggtgat    300
gttttgcag                                         309

```

```

<210> SEQ ID NO 15
<211> LENGTH: 328
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(328)

<400> SEQUENCE: 15

caa ttg tgg cta tac cag ctt ctt tgg caa ttg ttc cta ctg cat taa    48
Gln Leu Trp Leu Tyr Gln Leu Leu Trp Gln Leu Phe Leu Leu His
1      5      10      15

gag ttc aca atc gaa atg aat ggt atc gct tgg ctt tgg tca cat tgg    96
Glu Phe Thr Ile Glu Met Asn Gly Ile Ala Trp Leu Trp Ser His Trp
20     25     30

tga gtg cat gtc cgt gtg cac tcg ttc tat cta cac cag ttg cca tgt    144
Val His Val Arg Val His Ser Phe Tyr Leu His Gln Leu Pro Cys
35     40     45

gtt gcg cac ttt caa aag cag caa cgt ccg gtc ttc tgt tta aag gag    192
Val Ala His Phe Gln Lys Gln Gln Arg Pro Val Phe Cys Leu Lys Glu
50     55     60

cag agt acc ttg aga ctc tag cta aaa tca aaa tca tgg ctt ttg aca    240
Gln Ser Thr Leu Arg Leu Leu Lys Ser Lys Ser Trp Leu Leu Thr
65     70     75

aaa cag gga cta taa cta aag gag aat tta tgg tga ccg agt tca agt    288
Lys Gln Gly Leu Leu Lys Glu Asn Leu Trp Pro Ser Ser Ser
80     85     90

ctc tga ttg atg gtt tta gtc tca ata cac tgc ttt act g    328
Leu Leu Met Val Leu Val Ser Ile His Cys Phe Thr
95     100

```

```

<210> SEQ ID NO 16
<211> LENGTH: 463
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(463)

<400> SEQUENCE: 16

gtaaagggtta ccaactcatac atattctttt atgttgccaa agagaattca aaatcttaac    60
tggttatctt tcacggcaca ttgatagcga tataacatga ttgatttata tcatatattc    120
ataaaagatg aaataggagg tgccacattc acattctcat attgaagttt ctgaaatggc    180
tctaattggtt caccatagag ccaaaataac atatagacac aacgtcagcc gtctgatatt    240
caggaactta gatggaatag ttggatctta tacattgagg acacataaaa gtacttggtc    300

```

-continued

```

atataaattt tagaaatata atcaatgtat tataatctaa aattcttcaa atattcttga    360
tactgcaata caaaagcaca tggcacactg aatagaagcc ttgttcggtg gtctaaaaca    420
ttcgtttaga gtaaatactg agttgtctag tgaatatttt cag                        463

```

```

<210> SEQ ID NO 17
<211> LENGTH: 203
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(203)

```

```

<400> SEQUENCE: 17

```

```

ggt ttc aag cat tga gag caa gtc agg tca tcc gat ggc aac cgc tct    48
Gly Phe Lys His      Glu Gln Val Arg Ser Ser Asp Gly Asn Arg Ser
1          5          10          15

ggt gga cta tgc aca atc aaa ttc cgt tga gcc aaa gcc tga tag agt    96
Gly Gly Leu Cys Thr Ile Lys Phe Arg      Ala Lys Ala      Ser
20          25

tga gcg gtt tca aaa ttt tcc tgg tga agg gat att tgg aag aat tga    144
Ala Val Ser Lys Phe Ser Trp      Arg Asp Ile Trp Lys Asn
30          35          40

tgg aat gga aat cta tgt cgg gaa tag gaa aat ttc ttc aag agc tgg    192
Trp Asn Gly Asn Leu Cys Arg Glu      Glu Asn Phe Phe Lys Ser Trp
45          50          55

atg tac cac ag                        203
Met Tyr His

```

```

<210> SEQ ID NO 18
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(145)

```

```

<400> SEQUENCE: 18

```

```

gtaaattggtt gaatcatttc ttatgctcat agtagagata aaacatcaga gttataatta    60
taagtatatg atttctccag ttaattttgc tgtagatttt tctttgacct gtttagcact    120
aatgcggtgg atgtttgaat ttcag                        145

```

```

<210> SEQ ID NO 19
<211> LENGTH: 203
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(203)

```

```

<400> SEQUENCE: 19

```

```

tac cag aaa tag agg gtg ata gtt tca aag gaa agt ctg ttg gat aca    48
Tyr Gln Lys      Arg Val Ile Val Ser Lys Glu Ser Leu Leu Asp Thr
1          5          10          15

tat ttt tgg gat cat ctc cag ctg gaa ttt tca gtc ttt ccg atg ttt    96
Tyr Phe Trp Asp His Leu Gln Leu Glu Phe Ser Val Phe Pro Met Phe
20          25          30

gtc gaa ttg gtg taa aag aag caa tga gag aac tga agc aga tgg gta    144
Val Glu Leu Val      Lys Lys Gln      Glu Asn      Ser Arg Trp Val
35          40

```

tgc agg atc ag 203
Cys Arg Ile

```
<210> SEQ ID NO 20
<211> LENGTH: 619
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(619)
```

<400> SEQUENCE: 20

gtatattaat aattctgcat tacgctgaaa tgattataaa accctttgga ttattgttta	60
gtcttaagaa ttttactga actcttattg ttctcttctt ctatcatcaa cattgggttaa	120
acatttcatc taaatttaga gaacgtatca ccaagtaagt gctttacett tacagggtca	180
tataaaatac ttaagacagt gtgatgtgaa gatgaaggtt aaatgttgat ctggataaac	240
caagttatta tcacaactaa tataagatat gctattgttc tccaataatt ggacgatttt	300
cggacgtacg acgtacaatt cttcacatat gaaacctaca tcagacgtac atgacacgct	360
atgtttagca taaagagtca agattagcat gatgatttaa gctgaatctg aatttcaagt	420
atctattctt gtattgtacc cagggggcga actagtgttg tgcttagagg tctcaaacat	480
tgtatttgtg ttaaaaaatt cacttcatat gtattttaat aatttatcca gagcagtgag	540
ccatattttt tagaatccag aacccataaa ctcaaaatca tagatccacc tctgattgta	600
agtcggaaca attatcga	619

```
<210> SEQ ID NO 21
<211> LENGTH: 1498
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(1498)
```

<400> SEQUENCE: 21

tta ggt gga gct ttg gat gaa ttt caa gca gaa ctc cta cca gag gac	48
Leu Gly Gly Ala Leu Asp Glu Phe Gln Ala Glu Leu Leu Pro Glu Asp	
1 5 10 15	
aag gca aca atc atc aag ggt ttt cag aag gaa gct cca aca gcg atg	96
Lys Ala Thr Ile Ile Lys Gly Phe Gln Lys Glu Ala Pro Thr Ala Met	
20 25 30	
ata ggc gac ggc ctt aat gat gct cct gca tta gca aca gct gac att	144
Ile Gly Asp Gly Leu Asn Asp Ala Pro Ala Leu Ala Thr Ala Asp Ile	
35 40 45	
ggc atc tca atg ggc atc tct ggg tca gct ctc gct aaa gaa aca ggc	192
Gly Ile Ser Met Gly Ile Ser Gly Ser Ala Leu Ala Lys Glu Thr Gly	
50 55 60	
cat gct ata cta atg aca aat gac atc gga aga ata ccg aaa gct gca	240
His Ala Ile Leu Met Thr Asn Asp Ile Gly Arg Ile Pro Lys Ala Ala	
65 70 75 80	
cgt ctt gct aga aga gtt cga agg aag att gtt gag aat atg att ata	288
Arg Leu Ala Arg Arg Val Arg Arg Lys Ile Val Glu Asn Met Ile Ile	
85 90 95	

-continued

tca gcc gtt aca aag gct gcc ata gtt gca ttg gca ata gca ggt tat	336
Ser Ala Val Thr Lys Ala Ala Ile Val Ala Leu Ala Ile Ala Gly Tyr	
100 105 110	
cca ttg gtt tgg gct gct gtc ctc gca gat act ggg aca tgc ttg cta	384
Pro Leu Val Trp Ala Ala Val Leu Ala Asp Thr Gly Thr Cys Leu Leu	
115 120 125	
gtg att ttg aac agc atg cta ctt cta cga gga ggc aca cgc aga cat	432
Val Ile Leu Asn Ser Met Leu Leu Leu Arg Gly Gly Thr Arg Arg His	
130 135 140	
ggg aaa aaa tgt tgg aga tct tct act cct tcg cat gct ccc cac cac	480
Gly Lys Lys Cys Trp Arg Ser Ser Thr Pro Ser His Ala Pro His His	
145 150 155 160	
aaa gac aaa gct tca tgt tgc aag tcg gaa aat gct ccc cag ctg tgt	528
Lys Asp Lys Ala Ser Cys Cys Lys Ser Glu Asn Ala Pro Gln Leu Cys	
165 170 175	
tgc tct gat att gag tca caa aag aaa tgt aca agt caa tca tgc tcg	576
Cys Ser Asp Ile Glu Ser Gln Lys Lys Cys Thr Ser Gln Ser Cys Ser	
180 185 190	
tcc gag gtg tgt gtt cca aga tgt caa cct gtc tcc tcg gga tca aag	624
Ser Glu Val Cys Val Pro Arg Cys Gln Pro Val Ser Ser Gly Ser Lys	
195 200 205	
tca tgt gga aat aat cag tgc cca gac tcc att gaa aat agt ggt ttt	672
Ser Cys Gly Asn Asn Gln Cys Pro Asp Ser Ile Glu Asn Ser Gly Phe	
210 215 220	
cat tct cat cgc cgt cct caa tgc tgc tcg tcg aag atg gct gct aaa	720
His Ser His Arg Arg Pro Gln Cys Cys Ser Ser Lys Met Ala Ala Lys	
225 230 235 240	
gca tgc caa tct gca gtt tca gaa tca aag tca tgc gga aat aat cag	768
Ala Cys Gln Ser Ala Val Ser Glu Ser Lys Ser Cys Gly Asn Asn Gln	
245 250 255	
tgc cca gac tcc gtt gaa aat agt ggt ttt cat tct cat ccc cgt cct	816
Cys Pro Asp Ser Val Glu Asn Ser Gly Phe His Ser His Pro Arg Pro	
260 265 270	
gaa tgc tgc tcg tcg aag atg gct gct aaa gcg tgc caa tct gca gtt	864
Glu Cys Cys Ser Ser Lys Met Ala Ala Lys Ala Cys Gln Ser Ala Val	
275 280 285	
tca gaa tca aag tca tgt gga aat aat cag tgc cca gac tcc gtt gga	912
Ser Glu Ser Lys Ser Cys Gly Asn Asn Gln Cys Pro Asp Ser Val Gly	
290 295 300	
aat agt ggt ttt cat tct cat ccc cgt cct caa tgc tgt tca tcg aag	960
Asn Ser Gly Phe His Ser His Pro Arg Pro Gln Cys Cys Ser Ser Lys	
305 310 315 320	
atg gct gct aaa gca ggc caa tct gca ctt tca gaa tca aag tca tgt	1008
Met Ala Ala Lys Ala Gly Gln Ser Ala Leu Ser Glu Ser Lys Ser Cys	
325 330 335	
gga aat aac aat tgc tca gac tcc att cac aag agt aat tgt cat tct	1056
Gly Asn Asn Asn Cys Ser Asp Ser Ile His Lys Ser Asn Cys His Ser	
340 345 350	
tta act aac tct cta gta tgt tct tcc aag atg tct gct cca caa tgt	1104
Leu Thr Asn Ser Leu Val Cys Ser Ser Lys Met Ser Ala Pro Gln Cys	
355 360 365	
cat tct gct act tca agc aac aaa tca tgt gga agt acc aag tgc tcc	1152
His Ser Ala Thr Ser Ser Asn Lys Ser Cys Gly Ser Thr Lys Cys Ser	
370 375 380	
gac ttc agt gat aaa aaa tgt tgt caa tcc gac aaa att cct caa gcg	1200
Asp Phe Ser Asp Lys Lys Cys Cys Gln Ser Asp Lys Ile Pro Gln Ala	
385 390 395 400	

-continued

tgc tct acc aag aag tct gct cca ggg tgt caa tct gca gtt tct ggg	1248
Cys Ser Thr Lys Lys Ser Ala Pro Gly Cys Gln Ser Ala Val Ser Gly	
405 410 415	
tct aaa tca tgt gga aat agc aag tgt tca gac tca aaa gac aat agt	1296
Ser Lys Ser Cys Gly Asn Ser Lys Cys Ser Asp Ser Lys Asp Asn Ser	
420 425 430	
agc cat cct tca cat ccc gat cat caa aca tgc atg tct aag ttg tgt	1344
Ser His Pro Ser His Pro Asp His Gln Thr Cys Met Ser Lys Leu Cys	
435 440 445	
gct cca caa agc caa tct gca act tca agc tcc agg aca tgt gga aat	1392
Ala Pro Gln Ser Gln Ser Ala Thr Ser Ser Ser Arg Thr Cys Gly Asn	
450 455 460	
aca aag tgc tcg gac acc aat agc aag aat tct tgt tat tca caa acc	1440
Thr Lys Cys Ser Asp Thr Asn Ser Lys Asn Ser Cys Tyr Ser Gln Thr	
465 470 475 480	
aac tct gaa tca tgc tct tca aag atg tct ggt cca tca tgc aaa act	1488
Asn Ser Glu Ser Cys Ser Ser Lys Met Ser Gly Pro Ser Cys Lys Thr	
485 490 495	
gct aat tca g	1498
Ala Asn Ser	

<210> SEQ ID NO 22
 <211> LENGTH: 321
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: Intron
 <222> LOCATION: (1)..(321)

<400> SEQUENCE: 22

gtacgaccag attcctcttc ttgttaatac cccccgaacc aaactatagg atctaaagta	60
ttatttggcc ctctgtcgga ggatataatg gttagttaaa cctgaaatca tgtagtctat	120
gaattgcaaa tctctagcat cgtgacaaaa ttcttagatc atatacaact ttgagaatat	180
aggctgtcac agacccttct tcatactgca ttagaggaga gcagccaatt ttttattcat	240
gatttgaaac aaataaagtt cttcttgagg tgtatggaga ggctatgaga atcatttgct	300
gagtaggttt gagattttca g	321

<210> SEQ ID NO 23
 <211> LENGTH: 333
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: exon
 <222> LOCATION: (1)..(333)

<400> SEQUENCE: 23

gtt caa ggt cat gca gaa ata aga agt gcc agg act ctg caa ccg aga	48
Val Gln Gly His Ala Glu Ile Arg Ser Ala Arg Thr Leu Gln Pro Arg	
1 5 10 15	
aca gtt ttc att cac cac tta cta atc cac tca gtg ggg aaa agc ttt	96
Thr Val Phe Ile His His Leu Leu Ile His Ser Val Gly Lys Ser Phe	
20 25 30	
cgg agc aga aaa gct tgg att tag tcc gaa aag ata agg aat caa gtc	144
Arg Ser Arg Lys Ala Trp Ile Ser Glu Lys Ile Arg Asn Gln Val	
35 40 45	
atg atc ttc gtc atg gct gct ctg acg agg aac atg atc ata caa att	192
Met Ile Phe Val Met Ala Ala Leu Thr Arg Asn Met Ile Ile Gln Ile	
50 55 60	

-continued

```

tag aca agg cat atg aca gtt gtg cct tac aag aat gtt gtt att cgg      240
Thr Arg His Met Thr Val Val Pro Tyr Lys Asn Val Val Ile Arg
65                      70                      75

```

```

ttc aag gca ata aaa ctg atg tat cag aaa ctg gaa tcc agg aaa ctg      288
Phe Lys Ala Ile Lys Leu Met Tyr Gln Lys Leu Glu Ser Arg Lys Leu
80                      85                      90

```

```

ctc att gtg aca gca cca atc aaa cat gcc aaa ctg caa gtt cag      333
Leu Ile Val Thr Ala Pro Ile Lys His Ala Lys Leu Gln Val Gln
95                      100                     105

```

```

<210> SEQ ID NO 24
<211> LENGTH: 220
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: Intron
<222> LOCATION: (1)..(220)

```

```

<400> SEQUENCE: 24

```

```

gtaggcacta ccaaatcata tgatccaaag tgctcctcca ccttcactcc tacaataaat      60
gttcgatcaa acttcataag aagatagcat atgcatcgca aatctctaaa aaaatgatgg      120
ataatgttac tcaccaacta gtttgagata gaagtttaac tgattgctat atacgttaa      180
ctataaaaaa ctacttgta attagagctg agattttcag      220

```

```

<210> SEQ ID NO 25
<211> LENGTH: 774
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<221> NAME/KEY: exon
<222> LOCATION: (1)..(774)

```

```

<400> SEQUENCE: 25

```

```

gat cga tga cat gcg gaa atg ata aga tcc tgg act ctc taa gca tcc      48
Asp Arg      His Ala Glu Met Ile Arg Ser Trp Thr Leu      Ala Ser
1                      5                      10

```

```

atg gtt gtc att cgc atg ata atc cac tcc acg agg aga aca acc tgg      96
Met Val Val Ile Arg Met Ile Ile His Ser Thr Arg Arg Thr Thr Trp
15                      20                      25                      30

```

```

agc aga aaa tct tgg atg ttg ttg gag aag gta taa aat cac ctc atg      144
Ser Arg Lys Ser Trp Met Leu Leu Glu Lys Val      Asn His Leu Met
35                      40                      45

```

```

ctg tcg gtc atg gct gtt cgg aca agg aac acg atc act cac atc cag      192
Leu Ser Val Met Ala Val Arg Thr Arg Asn Thr Ile Thr His Ile Gln
50                      55                      60

```

```

aaa agg cat atg aca gtt gtg caa cag atg att gtt gtt ttt cag ttc      240
Lys Arg His Met Thr Val Val Gln Gln Met Ile Val Val Phe Gln Phe
65                      70                      75

```

```

aag tcc atg gca ttg acg acg tat caa aaa gtg aaa ttc aag aaa ctg      288
Lys Ser Met Ala Leu Thr Thr Tyr Gln Lys Val Lys Phe Lys Lys Leu
80                      85                      90

```

```

ctc att gtg aca gca caa agc aga gca tgg tca tct cca gca gct gca      336
Leu Ile Val Thr Ala Gln Ser Arg Ala Trp Ser Ser Pro Ala Ala Ala
95                      100                     105

```

```

aac atg aac caa aag atc agg taa atc act gtg gac ttc act cta aaa      384
Asn Met Asn Gln Lys Ile Arg      Ile Thr Val Asp Phe Thr Leu Lys
110                     115                     120

```

```

cta ctc caa ctg atg aag aac tag cca agc tgg tta gaa gat gct gcg      432

```

-continued

Leu	Leu	Gln	Leu	Met	Lys	Asn	Pro	Ser	Trp	Leu	Glu	Asp	Ala	Ala	
125					130					135					
aat	aca	aac	cat	gcc	acg	acg	tcc	gtt	ctg	gct	gca	gga	agc	atg	ctg
Asn	Thr	Asn	His	Ala	Thr	Thr	Ser	Val	Leu	Ala	Ala	Gly	Ser	Met	Leu
140					145					150				155	
cag	aat	gtg	gtc	caa	ccg	ttc	gat	caa	cta	tca	ata	tct	tac	ggg	aca
Gln	Asn	Val	Val	Gln	Pro	Phe	Asp	Gln	Leu	Ser	Ile	Ser	Tyr	Gly	Thr
160					165					170					
acc	atc	atc	att	acc	tag	act	gca	gtg	gtc	gta	agg	ttt	gtt	cgc	tgt
Thr	Ile	Ile	Ile	Thr		Thr	Ala	Val	Val	Val	Arg	Phe	Val	Arg	Cys
175						180					185				
tgg	aga	aga	gac	aca	tcg	gtg	gat	gct	gtg	aca	gct	tca	gaa	aag	aat
Trp	Arg	Arg	Asp	Thr	Ser	Val	Asp	Ala	Val	Thr	Ala	Ser	Glu	Lys	Asn
190					195					200					
gtt	gtg	cca	aga	aga	acc	acc	ttg	gag	caa	gtt	tcg	gag	gag	gtt	tat
Val	Val	Pro	Arg	Arg	Thr	Thr	Leu	Glu	Gln	Val	Ser	Glu	Glu	Val	Tyr
205					210					215					
cag	aaa	ttg	tca	tag	agt	aga	tgc	aat	ccg	aag	tgt	aca	tat	gtt	gta
Gln	Lys	Leu	Ser		Ser	Arg	Cys	Asn	Pro	Lys	Cys	Thr	Tyr	Val	Val
220					225						230				
aac	ttc	cta	cct	att	tta	tct	tca	aga	agt	tga	gct	gct	aat	ttg	aac
Asn	Phe	Leu	Pro	Ile	Leu	Ser	Ser	Arg	Ser		Ala	Ala	Asn	Leu	Asn
235					240						245				
aaa	gca														
Lys	Ala														
250															

<210> SEQ ID NO 26
 <211> LENGTH: 498
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: Intron
 <222> LOCATION: (1)..(498)

<400> SEQUENCE: 26

agggcggaatt	ctgcagatat	ccatcacact	ggcggccgct	cgagcatgca	tctagagggc	60
ccaattcgcc	ctatagttag	tcgtattaca	attcactggc	cgctgtttta	caacgtcgtg	120
actgggaaaa	ccctggcggt	acccaactta	atcgccctgc	agcacatccc	cctttcgcca	180
gctggcgtaa	tagcgaagag	gccccacccg	atcgcccttc	ccaacagttg	cgcagcctat	240
acgtacggca	gtttaaggtt	tacacctata	aaagagagag	ccgttatcgt	ctgtttgtgg	300
atgtacagag	tgatattatt	gacacgccgg	ggcgacggat	ggtgatcccc	ctggccagt	360
cacgtctgct	gtcagataaa	gtctcccggt	aactttaccc	ggtggtgcat	atcggggatg	420
aaagctggcg	catgatgacc	accgatatgg	ccagtgtgcc	ggtctccggt	atcggggaag	480
aagtggctga	tctcagcc					498

<210> SEQ ID NO 27
 <211> LENGTH: 303
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 1

<400> SEQUENCE: 27

caatttgttg	cggagaaatg	agaagagaat	catgaataac	aatgacagtc	ttgtttgtga	60
------------	------------	------------	------------	------------	------------	----

-continued

caattactga aacctcttta accccttcaa gattcttgag aattttttca actagaacaa	120
cttctgaagt acagcaaatt cccaaaacat caaaatagct cttgctcaac ttctttgttt	180
cattcatttt ttcactttcc accatttttc tccttctcta aggaggagga gaagataata	240
gatgaaaaag gaatttctag ttctggaata tattggatat acttttggtt tggaaagagg	300
tga	303

<210> SEQ ID NO 28
 <211> LENGTH: 258
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 2

<400> SEQUENCE: 28	
ctgctattaa aacaagaatg tcgatgtcaa gagtgagggt tcgcacggca gccacacctc	60
taaaaataat tggaggaatc ccaactgcaa cagctgcaag tgctaaccat tggaaagggtg	120
caaaaaagta cttcaaagat gagagtccaa gcaatatttc actgccaatt gcaaatggac	180
ttggccattt cttttggtag tttttctctc ctttactct tatgcttgct tctaattctg	240
cttgattcaa tgctttaa	258

<210> SEQ ID NO 29
 <211> LENGTH: 98
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 3

<400> SEQUENCE: 29	
cttgtgactt gcccttgact ctagccattc tgcaatggcg aataagaaga caatagtacc	60
agcttcccaa taatcgtgta aaacaattga tccagcca	98

<210> SEQ ID NO 30
 <211> LENGTH: 256
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 4

<400> SEQUENCE: 30	
catttagatt tgtagcgcca gccagaccg ttgaatctct ttgcttagaa actggaaacg	60
actcgctgt cagtgttttc tcgtccacgt cacattcccc ttccactaca actccatcaa	120
taggtatagt ttcaccagct ttcacagcaa gaatgctatt caccttgact tcatcaacat	180
ttacgacttc tccgctttcc gctaaaactg ctgttgaggg gactatatattg accagtgatg	240
acatagcggc ggtagc	256

<210> SEQ ID NO 31
 <211> LENGTH: 142
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 5

<400> SEQUENCE: 31	
tagcctgggtg tataatattt agcacacttg tcgatgtatc tttgggtttt tgatttcttg	60

-continued

```

ttctgagcat cttcgacagg ctgtgccatc ctagccaccg cacaatcttc agccaaagcc 120
gtagtcttaa cactgatata gc 142

```

```

<210> SEQ ID NO 32
<211> LENGTH: 328
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 6

```

```

<400> SEQUENCE: 32
cagtaaagca gtgtattgag actaaaacca tcaatcagag acttgaactc ggtcaccata 60
aattctcctt tagttatagt ccctgttttg tcaaaagcca tgattttgat tttagctaga 120
gtctcaaggt actctgctcc tttaaacaga agaccggacg ttgctgcttt tgaaagtgcg 180
caacacatgg caactggtgt agatagaacg agtgcacacg gacatgcact caccaatgtg 240
accaaagcca agcgatacca ttcatttcga ttgtgaactc ttaatgcagt aggaacaatt 300
gccaaagaag ctggtatagc cacaattg 328

```

```

<210> SEQ ID NO 33
<211> LENGTH: 203
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 7

```

```

<400> SEQUENCE: 33
ctgtggtaca tccagctctt gaagaaattt tcctattccc gacatagatt tccattccat 60
caattcttcc aaatatccct tcaccaggaa aattttgaaa ccgctcaact ctatcaggct 120
ttggctcaac ggaatttgat tgtgcatagt ccaccagagc ggttgccatc ggatgacctg 180
acttgccttc aatgcttgaa acc 203

```

```

<210> SEQ ID NO 34
<211> LENGTH: 203
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 8

```

```

<400> SEQUENCE: 34
ctgatcctgc acatgggttg cagctgcata acgataacca gtaagcatcg cggttttgat 60
acccatctgc ttcagttctc tcattgcttc ttttacacca attcgacaaa catcggaag 120
actgaaaatt ccagctggag atgatoccaa aaatatgtat ccaacagact ttcctttgaa 180
actatcacc cctattttctg gta 203

```

```

<210> SEQ ID NO 35
<211> LENGTH: 1498
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 9

```

```

<400> SEQUENCE: 35
ctgaattagc agttttgcat gatggaccag acatctttga agagcatgat tcagagttgg 60
tttgtgaata acaagaattc ttgctattgg tgtccgagca ctttgtattt ccacatgtcc 120

```

-continued

tggagcttga agttgcagat tggctttgtg gagcacacaa cttagacatg catgtttgat	180
gatcgggatg tgaaggatgg ctactattgt cttttgagtc tgaacacttg ctatttccac	240
atgatttaga cccagaaact gcagattgac accctggagc agacttcttg gtagagcacg	300
cttgaggaat tttgtcggat tgacaacatt ttttatcact gaagtcggag cacttggtac	360
ttccacatga tttgttgctt gaagtagcag aatgacattg tggagcagac atcttggaag	420
aacatactag agagttagtt aaagaatgac aattactctt gtgaatggag tctgagcaat	480
tgttattttc acatgacttt gattctgaaa gtgcagattg gcttgcttta gcagccatct	540
tcgatgaaca gcattgagga cggggatgag aatgaaaacc actatttcca acggagtctg	600
ggcactgatt atttccacat gactttgatt ctgaaactgc agattggcac gctttagcag	660
ccatcttcga cgagcagcat tcaggacggg gatgagaatg aaaaccacta ttttcaacgg	720
agtctgggca ctgattatth ccgcatgact ttgattctga aactgcagat tggcatgctt	780
tagcagccat cttcgacgag cagcattgag gacggcagtg agaatgaaaa ccactattht	840
caatggagtc tgggcactga ttattttccac atgactttga tcccaggagg acaggttgac	900
atcttggaac acacacctcg gacgagcatg attgacttgt acatttcttt tgtgactcaa	960
tatcagagca acacagctgg ggagcatttt ccgacttgca acatgaagct ttgtctttgt	1020
ggtgggggagc atgcgaagga gtagaagatc tccaacattt tttcccatgt ctgcgtgtgc	1080
ctctcgtag aagtagcatg ctgttcaaaa tccactagcaa gcattgtcca gtatctgcga	1140
ggacagcagc ccaaaccaat ggataacctg ctattgcaa tgcaactatg gcagcctttg	1200
taacggctga tataatcata ttctcaacaa tcttccttcg aactcttcta gcaagacgtg	1260
cagctttcgg tattcttccg atgtcatttg tcattagtat agcatggcct gtttctttag	1320
cgagagctga cccagagatg cccattgaga tgccaatgtc agctgttgct aatgcaggag	1380
catcattaag gccgtgcct atcatcgctg ttggagcttc cttctgaaaa ccttgatga	1440
ttgttgccct gtcctctggt aggagtctg cttgaaatc atccaaagct ccacctaa	1498

<210> SEQ ID NO 36
 <211> LENGTH: 333
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 10

<400> SEQUENCE: 36

ctgaacttgc agtttgcat gtttgattgg tgcgtcaca atgagcagtt tcctggattc	60
cagtttctga tacatcagtt ttattgcctt gaaccgaata acaacattct tgtaaggcac	120
aactgtcata tgcttctgt aaatttgtat gatcatgttc ctgctcagag cagccatgac	180
gaagatcatg acttgattcc ttatcttttc ggactaaatc caagcttttc tgctccgaaa	240
gcttttcccc actgagtggg ttagtaagtg gtgaatgaaa actgttctcg gttgcagagt	300
cctggcactt cttatttctg catgacctg aac	333

<210> SEQ ID NO 37
 <211> LENGTH: 774
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Reverse Complementary Sequence of Exon 11

-continued

<400> SEQUENCE: 37

```

tgctttgttc aaattagcag ctcaacttct tgaagataaa ataggttagga agtttacaac    60
atatgtacac ttcggttgc atctactcta tgacaatttc tgataaacct cctccgaaac    120
ttgctccaag gtggttcttc ttggcacaac attcttttct gaagctgtca cagcatccac    180
cgatgtgtct cttctccaac agcgaacaaa ccttacgacc actgcagtct aggtaatgat    240
gatggttgtc ccgtaagata ttgatagttg atcgaacggg tggaccacat tctgcagcat    300
gcttctctga gccagaacgg acgtcgtggc atggtttgta ttcgcagcat cttctaacca    360
gcttggtctag ttcttcatca gttggagtag ttttagagtg aagtcacag tgatttacct    420
gatcttttgg ttcattgttg cagctgctgg agatgacat gctctgcttt gtgctgtcac    480
aatgagcagt ttcttgaatt tcacttttgg atacgtcgtc aatgccatgg acttgaactg    540
aaaaacaaca atcatctgtt gcacaactgt catatgcctt ttctggatgt gagtgcgt    600
gttccttgtc cgaacagcca tgaccgacag catgaggtga ttttatacct tctccaacaa    660
catccaagat tttctgtctc aggttggtct cctcgtggag tggattatca tgcgaatgac    720
aaccatggat gcttagagag tccaggatct tatcatttcc gcatgtcatc gate      774

```

<210> SEQ ID NO 38

<211> LENGTH: 260

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Sense Sequence Transgene Construct

<400> SEQUENCE: 38

```

attttagtag ccagcccaga ccgttgaatc tatttgctta gaaactggaa acgactcgcc    60
tgtcagtggt ttctcgtcca cgtcacattc ccttccatt acaactccat caataggtat    120
agtttcacca gctttaacag caagaatgct attcaacttg acttcatcaa catttacgac    180
ttctccactt tcagctaaaa ctgctgttgg agggactata ttgaccagtg atgacatagc    240
agcagtagcc tacataacca                                260

```

<210> SEQ ID NO 39

<211> LENGTH: 96

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Spacer Sequence Transgene Construct

<400> SEQUENCE: 39

```

agcctgaaga attgagcaaa taacattaac aaacaatact tgaagtttca gcactaaata    60
aatgaagcat gaaggaatac tacactacca tttaga                                96

```

<210> SEQ ID NO 40

<211> LENGTH: 260

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Reverse Complement Sequence Transgene Construct

<400> SEQUENCE: 40

```

tggttatgta ggctactgct gctatgtcat cactggtcaa tatagtcctt ccaacagcag    60
ttttagctga aagtggagaa gtcgtaaatg ttgatgaagt caagttgaat agcattcttg    120

```

-continued

ctgttaaagc tgggtgaaact atacctattg atggagttgt aatggaaggg gaatgtgacg	180
tggacgagaa aacctgaca ggcgagtcgt ttccagtttc taagcaaata gattcaacgg	240
tctgggctgg cactacaaat	260

<210> SEQ ID NO 41
 <211> LENGTH: 616
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: RNAi sequence

<400> SEQUENCE: 41

atattgtagtg ccagcccaga ccgttgaatc tatttgctta gaaactggaa acgactcgcc	60
tgtcagtgtt ttctcgtcca cgtcacattc cccttcattt acaactccat caataggtat	120
agtttcacca gctttaacag caagaatgct attcaacttg acttcatcaa catttacgac	180
ttctccactt tcagctaaaa ctgctgttgg agggactata ttgaccagtg atgacatagc	240
agcagtagcc tacataacca agcctgaaga attgagcaaa taacattaac aaacaatact	300
tgaagtttca gcactaaata aatgaagcat gaaggaatac tacactacca tttagatggt	360
tatgtaggct actgctgcta tgtcatcact ggtcaatata gtccctccaa cagcagtttt	420
agctgaaagt ggagaagtcg taaatgttga tgaagtcaag ttgaatagca ttcttgctgt	480
taaagctggt gaaactatac ctattgatgg agttgtaatg gaaggggaat gtgacgtgga	540
cgagaaaaca ctgacaggcg agtcgtttcc agtttctaag caaatagatt caacggtctg	600
ggctggcact acaaat	616

<210> SEQ ID NO 42
 <211> LENGTH: 192
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Sense Sequence Transgene Construct

<400> SEQUENCE: 42

tgagagcaag tcaggtcatc cgatggcagc cgctctggtg gactatgcac aatcaaattc	60
cgttgagcca aagcctgata gagttgagca gtttcaaaat tttctggtg aagggatatt	120
tggaagaatt gatggaatgg aaatctatgt cgggaatagg aaaatttctt caagagctgg	180
atgtaccaca gg	192

<210> SEQ ID NO 43
 <211> LENGTH: 138
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Space Sequence Transgene Construct

<400> SEQUENCE: 43

taaatggttg aatcatttct tatgctcata gtagagataa aacatcagag ttataattat	60
aagtatatga ttctccagt taattttgct gttagatttt ctttgacctg tttagcacta	120
atgcggtgga tgtttgaa	138

<210> SEQ ID NO 44
 <211> LENGTH: 196
 <212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Reverse Complementary Sequence Transgene Construct

<400> SEQUENCE: 44

```
cctgtggtac atccagctct tgaagaaatt ttcctattcc cgacatagat ttccattcca    60
tcaattcttc caaatatccc ttcaccagga aaattttgaa actgctcaac tctatcaggc    120
tttggtcaa cggaatttga ttgtgcatag tccaccagag cggctgccat cggatgacct    180
gacttgctct caatgc                                          196
```

<210> SEQ ID NO 45

<211> LENGTH: 526

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: RNAi Sequence

<400> SEQUENCE: 45

```
tgagagcaag tcaggtcac cgatggcagc cgctctggtg gactatgcac aatcaaattc    60
cggttagcca aagcctgata gagttgagca gtttcaaaat ttcctggtg aagggatatt    120
tggaagaatt gatggaatgg aaatctatgt cgggaatagg aaaatttctt caagagctgg    180
atgtaccaca ggtaaatggt tgaatcattt cttatgctca tagtagagat aaaacatcag    240
agttataatt ataagtatat gattttctca gttaattttg ctgtagatt ttctttgacc    300
tgtttagcac taatgcgggt gatgtttgaa cctgtggtac atccagctct tgaagaaatt    360
ttcctattcc cgacatagat ttccattcca tcaattcttc caaatatccc ttcaccagga    420
aaattttgaa actgctcaac tctatcaggc tttggtcaa cggaatttga ttgtgcatag    480
tccaccagag cggctgccat cggatgacct gacttgctct caatgc                    526
```

<210> SEQ ID NO 46

<211> LENGTH: 4392

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Reverse Complement Sequence

<400> SEQUENCE: 46

```
acgaaacaag tttaatcgtc gagttgaaga acttctattt tatccatcct tcaaatgttg    60
tatacatgtg aagcctaacg tagatgagat actgttaaag actatttggg ggagggtttg    120
aacgagggtc cacaaaaaag aaccgtgttg taagaaaaga cttcgacagt gtcgtaggtg    180
gctacacaga gaagagggtg tcgcttggtt ggaatgctgg tgacgtcaga tccattacta    240
ctaccaacag ggcatctcat aactaccaac tagcttgcca acctggtgta agacgtcgta    300
cgaaggacgt cggctctgcc tgcagcaccg taccaaacat aaacgtcgta gaagattggt    360
cgaaccgata aagaagtagt caacctcaco aaaatctcac ttcaggtgct actaaatgga    420
ctagaaaacc aagtacaaac gtcgacgacc tctactggta cgagacgaaa cagcagctg    480
ttactcgta aagaacttaa agtgaaaaac tatgcagcag ttacggtacc tgaacttgac    540
ttttgttgt tagtagacaa cgtgttgaca gtatacggaa aagacctaca ctactagca    600
caaggaacag gcttgctggt actggtgctc gtactccact aaaatatgga agagggtgtt    660
gtaggttcta aaagacgagg ttcaacaaga ggagcacctc acctaatagt acgcttactg    720
```

-continued

ttggtaccta	cgaatctctc	aggctcctaga	atagtaaagg	cgtacagtag	ctaggacttg	780
aacgtcaaac	cgtacaaact	aaccacgaca	gtgttactcg	tcaaaggacc	taagggtcaaa	840
gactatgtag	tcaaaataac	ggaacttggc	ttattgttgt	aagaacattc	cgtgttgaca	900
gtatacggaa	cagatttaaa	catactagta	cagggagcag	tctcgtcggg	actgcttcta	960
gtactgaact	aaggaataga	aaagcctgat	ttaggttcga	aaagacgagg	ctttcgaaaa	1020
ggggtgactc	acctaatacat	tcaccactta	cttttgacaa	gagccaacgt	ctcaggaccg	1080
tgaagaataa	agacgtactg	gaacttgac	ttaatcgta	aaacgtacta	cctgggtctgt	1140
agaaacttct	cgtactaagt	ctcaacaaaa	cacttattgt	tcttaagaac	gataaccaca	1200
ggctcgtgaa	acataaagg	gtacaggacc	tcgaacttca	acgtctaacc	gaaacacctc	1260
gtgtgttgaa	tctgtacgta	caaactacta	gccctacact	tcctaccgat	gataacagaa	1320
aactcagact	tgtgaacgat	aaaggtgtac	taaatctggg	tctttgacgt	ctaactgtag	1380
gacctcgtct	gaagaacat	ctcgtgcaaa	ctccttaaaa	cagcctaact	gttgtaaaaa	1440
acagtgactt	cagcctcgtg	aacctgaag	gtgtactaaa	caacgaactt	catcgtctta	1500
ctgtaacacc	tcgtctgtag	aaccttcttg	tatgatctct	caatcaattt	cttactgtta	1560
atgagaacac	ttacctcaga	ctcgttaaca	ataaagggtg	actgaaacta	agactttcac	1620
gtctaaccgg	acgaaatcgt	cggtagaagc	tacttgctgt	aactcctgcc	cctactctta	1680
cttttggtga	taaaaagttg	cctcagaccc	gtgactaata	aaggtgtact	gaaactaaga	1740
ctttgacgtc	taaccgtgcg	aaatcgtcgg	tagaagctgc	tcgtcgtaag	tcctgcccct	1800
actcttactt	ttgggtgataa	aagttgcctc	agacccgtga	ctaataaagg	cgtgctgaaa	1860
ctaagacttt	gacgtctaac	cgtacgaaat	cgtcggtaga	agctgctcgt	cgtaactcct	1920
gccgctactc	ttacttttgg	tgataaaaagt	tacctcagac	ccgtgactaa	taaagggtgta	1980
ctgaaactag	gactcctctg	tccaactgta	gaaccttgtg	tgtggagcct	gctcgtacta	2040
actgaacatg	taaagaaaac	actgagttat	agtctcgttg	tgtcgacccc	tcgtaaaagg	2100
ctgaacgttg	tacttcgaaa	cagaaacacc	accctcgtac	gcttcctcat	cttctagagg	2160
ttgtaaaaaa	gggtacagac	gcacacggag	gagcatcttc	atcgtacgac	aagttttagt	2220
gategttcgt	acagggtcat	agacgctcct	gtcgtcgggt	ttggttacct	attggacgat	2280
aacgggtacg	ttgataccgt	cggaaacatt	gctgactata	ttagtataag	agttgttaga	2340
aggaagcttg	agaagatcgt	tctgcacgtc	gaaagccata	agaaggctac	agtaaacagt	2400
aatcatattg	taccggacaa	agaaatcgct	ctcagctggg	tctctacggg	taactctacg	2460
gttacagtcg	acaacgatta	cgtcctcgta	gtaattccgg	cagcggatag	tagcgacaac	2520
ctcgaaggaa	gacttttggg	aactactaac	aacggaacag	gagaccatcc	tcaagacgaa	2580
ctttaagtag	gtttcgagg	ggattgacta	ggacgtgtac	caaccgtcga	cgtattgtta	2640
gtggtcattc	gtagcgccaa	aactatgggt	agacgaagtc	aagagagtaa	cgaagaaaat	2700
gtggttaagc	tgtttgtagc	ctttctgact	tttaaggctg	acctctacta	gggtttttat	2760
acatagggtg	tctgaaagga	aactttgata	gtgggagata	aagaccatga	caccatgtag	2820
gtcgagaact	tctttaaaaa	gataagggtc	gtatctaaag	gtaaggtagt	taagaagggt	2880
tatagggaag	tggctccttt	aaaactttga	cgagttgaga	tagtccgaaa	ccgagttgcc	2940
ttaaactaac	acgtatcagg	tggctctcgc	gacggtagcc	tactggactg	aacgagagtt	3000

-continued

```

acgaactttg ggtcatttcg tcacataact ctgattttgg tagttagtct ctgaacttga 3060
gccagtggtg tttaagagga aatcaatata agggacaaaa cagttttcgg tactaaaaact 3120
aaaatcgatc tcagagttcc atgagacgag gaaatttgtc ttctggcctg caacgacgaa 3180
aactttcacg cgttgtgtac cgttgaccac atctatcttg ttcacgtgtg cctgtacgtg 3240
agtggttaca ctggtttcgg ttctgtatgg taagtaaagc taacacttga gaattacgtc 3300
atccttggtt acggtttctt cgactataatc ggtgttaacg accacatatt ataaatcgtg 3360
tgaacagcta catagaaacc caaaaactaa agaacaagac tcgtagaagc tgttcgacac 3420
ggtaggacgt gtggcgtggt agaagtcggt ttccggcatca gaattgtgac tatatcggtg 3480
aatctaaaca tcacggtcgg gtctggcaac ttagagaaac gaatcttga cctttgctga 3540
gcggacagtc acaaaagagc aggtgcagtg taaggggaag gtgatgttga ggtagtattc 3600
catatcaaa ggttcgaaag tgtcgttctt acgataagtg gaactgaagt agttgtaaat 3660
gctgaagagg cgaaagtcga ttttgacgac aacctccctg atataactgg tcactactgt 3720
atcgtcgcca tcggaacact gaacgggaac tgagatcggt aagacgttac cgcttattct 3780
tctgttatca tggtcgaagg gttattagca cattttggtt actaggtcgg tgacgataat 3840
tttgttctta caactacagt tctactcca aagcgtgccg tcggtgtgga gatttttatt 3900
aacctcctta gggttgacgt tgtcgacgtt cagcattggt aacctttcca cgttttttca 3960
tgaagttttt actctcaggt tcggtataag gtgacgggta acgtttacct gaaccggtaa 4020
agaaaaccat caaaaagaga ggaaagttag aatacgaacg aagattagaa cgaactaagt 4080
tacgaaattg ttaaacacg cctctttact cttctcttag tacttattgt tactgtcaga 4140
aacaacactg ttaatgactt tggagaaatt ggggaagttc taagaactct tgaaaaagtt 4200
gatcttggtt aagacttcat gtcgtttaag ggtttttag ttttatcgag aacgagttga 4260
agaaacaaag taagtaaaaa agtgaaaggt ggtaaaaaga ggaagagatt cctcctctc 4320
ttctattatc tactttttcc ttaaagatca agaccttata taacctatat gaaaacaaaa 4380
cctttctcca ct 4392

```

```

<210> SEQ ID NO 47
<211> LENGTH: 4011
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sense Sequence

```

```

<400> SEQUENCE: 47

```

```

acgacggcca gtgaattgta atacgactca ctatagggcg aattgggccc tctagatgca 60
tgctcgagcg gccgccagtg tgatggatat ctgcagaatt cgcccttgag gaaacataga 120
aagaagagaa tgggtgaaag tgagaaaatg aatgacacaa agaactgag caagagctat 180
tttgatgttt tgggaatttg ctgtacttca gaagttgttc ttgttgaaaa aattctcaag 240
aatcttgaag ggggttaaaga ggtttcagta attgtcacia caaagactgt cattgttatt 300
catgattctc tcctcatctc tcagcaacaa attgttaaag cattgaatca agcaagatta 360
gaagcaagta taagagttaa aggagagaaa aactaccaa agaaatggcc aagtcattt 420
gcaattggca gtggaatatt gcttggactc tcatttttga agtacttttt tgcacctttc 480
caatggttag cacttgcagc tgttgcagtt gggattcctc caattatttt taggggtgtg 540

```

-continued

gctgccgtgc gaaacctcac tcttgacatc aacattcttg ttttaatagc agtgacggga	600
tcaattgttt tacacgatta ttgggaagct gatactattg tcttcttatt caccattgca	660
gaatggctag agtcaagggc aagtcacaag gctactgctg ctatgtcatc actggccaat	720
atagtccttc caacagcagt tttagctgaa agtggagaag tcgtaaatgt tgatgaagtc	780
aagttgaata gcattcttgc tgtaaagct ggtgaaacta tacctattga tggagtgtga	840
atggaagggg aatgtgacgt ggacgagaaa aactgacag gcgagtcgtt tccagtttct	900
aagcaaatag attcaacggt ctgggctggc actacaaatc taaatggcta tatcagtgtt	960
aagactacgg ctttggtgta agattgtgcg gtggctagga tggcgagct tgtcgaagat	1020
gctcagaaca agaaatcaaa aacccaaaga tacattgaca agtgtgctaa atattatata	1080
ccagcaattg tggtatatac agcttctttg gcaatagtgc ctactgcatt aagagttcac	1140
aatcgaaatg agtggatcgc cttggctttg gtcacgttgg tgagtgcagc tccgtgtgca	1200
cttgtgctat ctacaccagt tgccatgtgt tgtgcacttt ctaaagcagc aacgtccggt	1260
cttctgttta aaggagcaga gtaccttgag actcttgcta aaatcaaaat catggctttt	1320
gacaaaaacag ggactataac tagaggagaa tttatggtga ccgagttcaa gtctctggtt	1380
gatggctctg gtctcaatac actgctttac tgggtttcaa gtattgagag caagtcaggt	1440
catccgatgg cagccgctct ggttgactat gcacaatcaa attccgttga gccaaagcct	1500
gatagagttg agcagtttca aaattttcct ggtgaagggg tatttggaag aattgatgga	1560
atggaaatct atgtcgggaa taggaaaatt tcttcaagag ctggatgtac tacagtacca	1620
gaaatagagg gtgatagttt ccaaggaaag tctgttgat acatatTTTT gggatcatct	1680
cccgtggaa ttttcggtct ttccgatgtt tgtcgaattg gtgtaaaaga agcaatgaga	1740
gagctgaagc agatgggtat caaaaccgag atgcttactg gtgattgtta tgcagctgcc	1800
aaccatgtgc aggatcagtt aggtggagct atggatgaat ttcaagcgga actcttacca	1860
gaggacaagg caacaatcat caagggtttt cagaaggaag ctccaacagc gatgataggc	1920
gacggcctta atgatgctcc tgcattagca acagctgaca ttggcatctc aatgggcac	1980
tctgggtcag ctctcggcaa agaaacaggc catgttatac taatgacaaa tgacatcgga	2040
agaataccaa aagctgcacg tcttgctaga agagttcgaa ggaagattgt tgagaatatg	2100
attatatcag tcgttacaaa ggccgccata gttgcattgg caatagcagg ttatccattg	2160
gtttgggctg ctgtcctcgc ggatactggg acatgcttgc tagtgatctt gaacagcatg	2220
ctacttctac gagtaggcac acacagacat gggaaaaaat gttgtagatc tgctactcct	2280
tcgcatgctc ccaaccacaa agacaaaagt tcttgttgca agtcggaaaa tgctccgcag	2340
ctgtgttget ctgatattga gtcacaaaag aaatgtacga gtcaatcatg ctgctccgag	2400
gtgtgtgttc caagatgcca acctgtctcc tcgggatcaa agtcatgtgg aaataatcag	2460
tgccagact ccgttgaaaa tagtggtttt cattctcatc cccgtcctct agtatgttct	2520
tccaagatgt ctgtccaca atgtcattct gccacttcaa gctccaaatc atgtggaagt	2580
accaagtgtc ccaacttcag tgacaaaaaa tgttgccaat atgacaaaat tctcaaacg	2640
tgctctacca agaagtctgc tccaggatgt caatctgcag tttctgggtc taaatcatgt	2700
ggagatagca agtgttcaga ctcgaaagac aatagtagcc atccttcaca tccgatcat	2760
caaatatgca cgtctaagtt gtgtgtctca caaagccaat ctgcaacttc aagctccagg	2820

-continued

```

acatgtggaa atatgaagtg ctcgacaccc aatagcaaga attcttgta ttcacatacc 2880
aactctgaat catgctcttc aaagatgtct ggtccagcat gcaaaactgc taattcaggt 2940
tcaaggttat gcggaaataa gaagtgccta gactctgcaa acgagaacag ttttcattca 3000
cttactaatc cactctgtga ggaaaagctt ttggagaagg aaagcttga tttagcccca 3060
aaagataggg aatcaaatca tgatcttagt catggttact ctgacgagga acatgatcat 3120
ctaaatttag acaaggcaca tgacagtgtg gccttacaag aatgttgta ttctgttcaa 3180
ggcaataaaa ctgatgtatc agaaactgga atccaggaag ctgctcattg tgacagcatc 3240
aatcaaacat gccaaactgc aatttcagga tcaatgacat gcggaaataa taagagtctg 3300
gactctctaa gcatccatgg ttgtcattca catgatagtc cactccacaa ggagagcaac 3360
ttggagcaga aaagcttga tggtgctgga gaaggataa aatcacctca tgcgtcgggt 3420
caaggctgtt cggacaagga gcacaatcac tcgcatccag aaaaggcgta tgacagtgt 3480
gcaacagacg attgtgtgtt ttcagttcaa gtccatggca ttgacgacgt atcaagaagt 3540
gaaattcaag aaactgctca ttgtgacagc acaaaacaga gcacgggtcat cccagcagc 3600
tgcaacatg aacaaaaaga tcaggtaaat cactgtggat ctactctaa aagtattcca 3660
actgatgaag aactagccaa gctgggtaga agatgctgca aatacaaac atgccacgat 3720
gtccgctctg gctgcaggaa gcatgctgca gaatgtggtc caaccgttcg atcaaccatc 3780
aatatcttac gggacaacca tcatcatcat ctactgca gtggtcgtaa ggtttgttcg 3840
ctgttgaga agagacacat tgggtgatgc tgtgacagct tcagaaaaga atgttgtgcc 3900
aagaacaatc accttgagc aagttttgga ggaggtttat cagaaattgt caaagggcga 3960
attccagcac actggcgccc gttactagt gatccgagct cggtaaccaag c 4011

```

<210> SEQ ID NO 48

<211> LENGTH: 4011

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Reverse Complement Sequence

<400> SEQUENCE: 48

```

gcttggtacc gagctcggat ccactagtaa cggccgccag tgtgctggaa ttgcgccttt 60
gacaatttct gataaacctc ctccaaaact tgcctcaagg tgattgttct tggcacaaca 120
ttcttttctg aagctgtcac agcatocacc aatgtgtctc ttctccaaca gcgaacaaac 180
cttacgacca ctgcagtcta gatgatgatg atggttgtcc cgtaagatat tgatggttga 240
tcgaacgggt ggaccacatt ctgcagcatg cttcctgcag ccagagcgga catcgtggca 300
tggtttgtat ttgcagcatc ttctaaccag cttggctagt tcttcatcag ttggaatact 360
ttagagtga gatccacagt gatttacctg atcttttggg tcatgttcgc agctgctggg 420
gatgaccgtg ctctgttttg tgctgtcaca atgagcagtt tcttgaattt cacttcttga 480
tacgtcgta atgccatgga cttgaactga aaaacaacaa tcgtctgttg cacaactgtc 540
atacgccctt tctggatgcg agtgattgtg ctcttgtcc gaacagcctt gaccgacagc 600
atgaggtgat tttatacctt ctccagcaac atccaagctt ttctgctcca agttgctctc 660
cttgtggagt ggactatcat gtgaatgaca accatggatg cttagagagt ccagactctt 720
attatttccg catgtcattg atcctgaaat tgcagtttgg catgtttgat tgatgctgtc 780

```

-continued

acaatgagca gcttctctgga ttccagtttc tgatacatca gttttattgc cttgaacaga	840
ataacaacat tcttgtaagg cacaactgtc atgtgccttg tctaaattta gatgatcatg	900
ttctctgtca gagtaacat gactaagatc atgatttgat tccctatctt ttcgggctaa	960
atccaagctt tcttcttcca aaagcttttc ctcacagagt ggattagtaa gtgaatgaaa	1020
actgttctcg ttgacagagt ctaggcactt cttatttccg cataaccttg aacctgaatt	1080
agcagttttg catgctggac cagacatctt tgaagagcat gattcagagt tggatatgtga	1140
ataacaagaa ttcttgctat tgggtgtccg gcacttcata tttccacatg tctggagct	1200
tgaagttgca gattggcttt gtggagcaca caacttagac gtgcataatt gatgatcggg	1260
atgtgaagga tggctactat tgtctttcga gtctgaacac ttgctatctc cacatgattt	1320
agaccagaa actgcagatt gacatcctgg agcagacttc ttggtagagc acgtttgagg	1380
aattttgtca tattggcaac attttttgtc actgaagttg gagcacttgg tacttccaca	1440
tgatttggag cttgaagtgg cagaatgaca ttgtggagca gacatcttgg aagaacatac	1500
tagaggacgg ggatgagaat gaaaaccact attttcaacg gagtctgggc actgattatt	1560
tccacatgac ttgtatcccg aggagacagg ttggcatctt ggaacacaca cctcggacga	1620
gcatgattga ctctacatt tcttttgtga ctcaatatca gagcaacaca gctgcggagc	1680
attttccgac ttgcaacaag aagctttgtc tttgtggttg ggagcatgcg aaggagtagc	1740
agatctacaa catttttttc catgtctgtg tgtgcctact cgtagaagta gcatgctgtt	1800
caagatcact agcaagcatg tcccagtatc cgcgaggaca gcagcccaa ccaatggata	1860
acctgctatt gccaatgcaa ctatggcggc ctttgtaacg actgatataa tcatattctc	1920
aacaatcttc cttcgaacte ttctagcaag acgtgcagct tttggtattc ttcagatgtc	1980
atltgtcatt agtataacat ggcctgttct tttcgcgaga gctgacccag agatgcccac	2040
tgagtgcca atgtcagctg ttgctaagtc aggagcatca ttaaggccgt cgcctatcat	2100
cgtgttggga gcttctctct gaaaaccctt gatgattgtt gccttgcctc ctggtgaagag	2160
ttcgcgttga aattcatcca tagctccacc taactgatcc tgcacatggt tggcagctgc	2220
ataacaatca ccagtaagca tcgcgggttt gatacccatc tgcttcagct ctctcattgc	2280
ttcttttaca ccaattgcac aaacatcgga aagaccgaaa attccagcgg gagatgatcc	2340
caaaaatatg tatccaacag actttccttg gaaactatca cctctatctt ctggtactgt	2400
agtacatcca gctcttgaag aaattttcct attcccgaca tagatttcca tttcatcaat	2460
tcttccaaat atcccttcac caggaaaatt ttgaaactgc tcaactctat caggctttgg	2520
ctcaacggaa ttgattgtg catagtcaac cagagcggct gccatcggat gacctgactt	2580
gctctcaata cttgaaaccc agtaaagcag tgtattgaga ccaagaccat caaccagaga	2640
cttgaactcg gtcaccataa attctcctct agttatagtc cctgttttgt caaaagccat	2700
gattttgatt ttagcaagag tctcaaggta ctctgctcct ttaaacagaa gaccggacgt	2760
tgctgcttta gaaagtgcac aacacatggc aactgggtga gatagcacia gtgcacacgg	2820
acatgcactc accaactgta ccaaagccaa gcgataccac tcatttcgat tgtgaactct	2880
taatgcagta ggaactattg ccaaagaagc tgatatagcc acaattgctg gtgtataata	2940
tttagcacac ttgtcaatgt atctttgggt ttttgatttc ttgttctgag catcttcgac	3000
aagctgcgcc atcttagcca cgcacaatc ttcagccaaa gccgtagtct taacactgat	3060

-continued

```

atagccattt agattttag tgccagccca gaccgttgaa tctatttgct tagaaactgg 3120
aaacgactcg cctgtcagtg ttttctcgtc caggtcacat tccccctcca ttacaactcc 3180
atcaataggt atagtttcac cagctttaac agcaagaatg ctattcaact tgacttcac 3240
aacatttacg acttctccac tttcagctaa aactgctggt ggagggacta tattgaccag 3300
tgatgacata gcagcagtag ccttgtgact tgcccttgac tctagccatt ctgcaatggt 3360
gaataagaag acaatagtat cagcttccca ataatcgtgt aaaacaattg atcccgtcac 3420
tgctattaaa acaagaatgt tgatgtcaag agtgaggttt cgcacggcag ccacaccct 3480
aaaaataatt ggaggaatcc caactgcaac agctgcaagt gctaaccatt ggaaagggtc 3540
aaaaaagtac ttcaaaaatg agagtccaag caatattcca ctgccaatg caaatggact 3600
tggccatttc ttttggtagt ttttctctcc tttcactctt atacttgctt ctaactctgc 3660
ttgattcaat gctttaacaa tttgtgtctg agaaatgagg agagaatcat gaataacaat 3720
gacagtcttt gttgtgacaa ttactgaaac ctctttaacc ccttcaagat tcttgagaat 3780
tttttcaaca agaacaactt ctgaagtaca gcaaatccc aaaacatcaa aatagctctt 3840
gctcagattc tttgtgtcat tcattttctc actttccacc attctcttct ttctatgttt 3900
cctcaagggc gaattctgca gatatecatc aactggcgg ccgctcgagc atgcatctag 3960
agggcccaat tcgccctata gtgagtcgta ttacaattca ctggccgtcg t 4011

```

<210> SEQ ID NO 49

<211> LENGTH: 1325

<212> TYPE: PRT

<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 49

```

Gly Glu Leu Gly Pro Leu Asp Ala Cys Ser Ser Gly Arg Gln Cys Asp
1           5           10           15

```

```

Gly Tyr Leu Gln Asn Ser Pro Leu Arg Lys His Arg Lys Lys Arg Met
20           25           30

```

```

Val Glu Ser Glu Lys Met Asn Asp Thr Lys Asn Leu Ser Lys Ser Tyr
35           40           45

```

```

Phe Asp Val Leu Gly Ile Cys Cys Thr Ser Glu Val Val Leu Val Glu
50           55           60

```

```

Lys Ile Leu Lys Asn Leu Glu Gly Val Lys Glu Val Ser Val Ile Val
65           70           75           80

```

```

Thr Thr Lys Thr Val Ile Val Ile His Asp Ser Leu Leu Ile Ser Gln
85           90           95

```

```

Gln Gln Ile Val Lys Ala Leu Asn Gln Ala Arg Leu Glu Ala Ser Ile
100          105          110

```

```

Arg Val Lys Gly Glu Lys Asn Tyr Gln Lys Lys Trp Pro Ser Pro Phe
115          120          125

```

```

Ala Ile Gly Ser Gly Ile Leu Leu Gly Leu Ser Phe Leu Lys Tyr Phe
130          135          140

```

```

Phe Ala Pro Phe Gln Trp Leu Ala Leu Ala Ala Val Ala Val Gly Ile
145          150          155          160

```

```

Pro Pro Ile Ile Phe Arg Gly Val Ala Ala Val Arg Asn Leu Thr Leu
165          170          175

```

```

Asp Ile Asn Ile Leu Val Leu Ile Ala Val Thr Gly Ser Ile Val Leu
180          185          190

```

-continued

His Asp Tyr Trp Glu	Ala Asp Thr Ile Val	Phe Leu Phe Thr Ile Ala
195	200	205
Glu Trp Leu Glu Ser	Arg Ala Ser His Lys	Ala Thr Ala Ala Met Ser
210	215	220
Ser Leu Val Asn Ile	Val Pro Pro Thr Ala	Val Leu Ala Glu Ser Gly
225	230	235 240
Glu Val Val Asn Val	Asp Glu Val Lys Leu	Asn Ser Ile Leu Ala Val
245	250	255
Lys Ala Gly Glu Thr	Ile Pro Ile Asp Gly	Val Val Met Glu Gly Glu
260	265	270
Cys Asp Val Asp Glu	Lys Thr Leu Thr Gly	Glu Ser Phe Pro Val Ser
275	280	285
Lys Gln Ile Asp Ser	Thr Val Trp Ala Gly	Thr Thr Asn Leu Asn Gly
290	295	300
Tyr Ile Ser Val Lys	Thr Thr Ala Leu Ala	Glu Asp Cys Ala Val Ala
305	310	315 320
Arg Met Ala Gln Leu	Val Glu Asp Ala Gln	Asn Lys Lys Ser Lys Thr
325	330	335
Gln Arg Tyr Ile Asp	Lys Cys Ala Lys Tyr	Tyr Thr Pro Ala Ile Val
340	345	350
Ala Ile Ser Ala Ser	Leu Ala Ile Val Pro	Thr Ala Leu Arg Val His
355	360	365
Asn Arg Asn Glu Trp	Tyr Arg Leu Ala Leu	Val Thr Leu Val Ser Ala
370	375	380
Cys Pro Cys Ala Leu	Val Leu Ser Thr Pro	Val Ala Met Cys Cys Ala
385	390	395 400
Leu Ser Lys Ala Ala	Thr Ser Gly Leu Leu	Phe Lys Gly Ala Glu Tyr
405	410	415
Leu Glu Thr Leu Ala	Lys Ile Lys Ile Met	Ala Phe Asp Lys Thr Gly
420	425	430
Thr Ile Thr Arg Gly	Glu Phe Met Val Thr	Glu Phe Lys Ser Leu Val
435	440	445
Asp Gly Leu Gly Leu	Asn Thr Leu Leu Tyr	Trp Val Ser Ser Ile Glu
450	455	460
Ser Lys Ser Gly His	Pro Met Ala Ala Ala	Leu Val Asp Tyr Ala Gln
465	470	475 480
Ser Asn Ser Val Glu	Pro Lys Pro Asp Arg	Val Glu Gln Phe Gln Asn
485	490	495
Phe Pro Gly Glu Gly	Ile Phe Gly Arg Ile	Asp Gly Met Glu Ile Tyr
500	505	510
Val Gly Asn Arg Lys	Ile Ser Ser Arg Ala	Gly Cys Thr Thr Val Pro
515	520	525
Glu Ile Glu Gly Asp	Ser Phe Gln Gly Lys	Ser Val Gly Tyr Ile Phe
530	535	540
Leu Gly Ser Ser Pro	Ala Gly Ile Phe Gly	Leu Ser Asp Val Cys Arg
545	550	555 560
Ile Gly Val Lys Glu	Ala Met Arg Glu Leu	Lys Gln Met Gly Ile Lys
565	570	575
Thr Ala Met Leu Thr	Gly Asp Cys Tyr Ala	Ala Ala Asn His Val Gln
580	585	590

-continued

Asp	Gln	Leu	Gly	Gly	Ala	Met	Asp	Glu	Phe	Gln	Ala	Glu	Leu	Leu	Pro
595					600					605					
Glu	Asp	Lys	Ala	Thr	Ile	Ile	Lys	Gly	Phe	Gln	Lys	Glu	Ala	Pro	Thr
610					615					620					
Ala	Met	Ile	Gly	Asp	Gly	Leu	Asn	Asp	Ala	Pro	Ala	Leu	Ala	Thr	Ala
625					630					635					640
Asp	Ile	Gly	Ile	Ser	Met	Gly	Ile	Ser	Gly	Ser	Ala	Leu	Ala	Lys	Glu
645					650					655					
Thr	Gly	His	Val	Ile	Leu	Met	Thr	Asn	Asp	Ile	Gly	Arg	Ile	Pro	Lys
660					665					670					
Ala	Ala	Arg	Leu	Ala	Arg	Arg	Val	Arg	Arg	Lys	Ile	Val	Glu	Asn	Met
675					680					685					
Ile	Ile	Ser	Val	Val	Thr	Lys	Ala	Ala	Ile	Val	Ala	Leu	Ala	Ile	Ala
690					695					700					
Gly	Tyr	Pro	Leu	Val	Trp	Ala	Ala	Val	Leu	Ala	Asp	Thr	Gly	Thr	Cys
705					710					715					720
Leu	Leu	Val	Ile	Leu	Asn	Ser	Met	Leu	Leu	Leu	Arg	Val	Gly	Thr	His
725					730					735					
Arg	His	Gly	Lys	Lys	Cys	Cys	Arg	Ser	Ala	Thr	Pro	Ser	His	Ala	Pro
740					745					750					
Asn	His	Lys	Asp	Lys	Ala	Ser	Cys	Cys	Lys	Ser	Glu	Asn	Ala	Pro	Gln
755					760					765					
Leu	Cys	Cys	Ser	Asp	Ile	Glu	Ser	Gln	Lys	Lys	Cys	Thr	Ser	Gln	Ser
770					775					780					
Cys	Ser	Ser	Glu	Val	Cys	Val	Pro	Arg	Cys	Gln	Pro	Val	Ser	Ser	Gly
785					790					795					800
Ser	Lys	Ser	Cys	Gly	Asn	Asn	Gln	Cys	Pro	Asp	Ser	Val	Glu	Asn	Ser
805					810					815					
Gly	Phe	His	Ser	His	Pro	Arg	Pro	Leu	Val	Cys	Ser	Ser	Lys	Met	Ser
820					825					830					
Ala	Pro	Gln	Cys	His	Ser	Ala	Thr	Ser	Ser	Ser	Lys	Ser	Cys	Gly	Ser
835					840					845					
Thr	Lys	Cys	Ser	Asn	Phe	Ser	Asp	Lys	Lys	Cys	Cys	Gln	Tyr	Asp	Lys
850					855					860					
Ile	Pro	Gln	Thr	Cys	Ser	Thr	Lys	Lys	Ser	Ala	Pro	Gly	Cys	Gln	Ser
865					870					875					880
Ala	Val	Ser	Gly	Ser	Lys	Ser	Cys	Gly	Asp	Ser	Lys	Cys	Ser	Asp	Ser
885					890					895					
Lys	Asp	Asn	Ser	Ser	His	Pro	Ser	His	Pro	Asp	His	Gln	Ile	Cys	Thr
900					905					910					
Ser	Lys	Leu	Cys	Ala	Pro	Gln	Ser	Gln	Ser	Ala	Thr	Ser	Ser	Ser	Arg
915					920					925					
Thr	Cys	Gly	Asn	Met	Lys	Cys	Ser	Asp	Thr	Asn	Ser	Lys	Asn	Ser	Cys
930					935					940					
Tyr	Ser	His	Thr	Asn	Ser	Glu	Ser	Cys	Ser	Ser	Lys	Met	Ser	Gly	Pro
945					950					955					960
Ala	Cys	Lys	Thr	Ala	Asn	Ser	Gly	Ser	Arg	Leu	Cys	Gly	Asn	Lys	Lys
965					970					975					
Cys	Leu	Asp	Ser	Ala	Asn	Glu	Asn	Ser	Phe	His	Ser	Leu	Thr	Asn	Pro
980					985					990					
Leu	Cys	Glu	Glu	Lys	Leu	Leu	Glu	Lys	Glu	Ser	Leu	Asp	Leu	Ala	Arg

-continued

995				1000				1005				
Lys Asp	Arg	Glu	Ser	Asn His	Asp	Leu	Ser	His Gly	Tyr	Ser	Asp	
1010				1015				1020				
Glu Glu	His	Asp	His	Leu Asn	Leu	Asp	Lys	Ala His	Asp	Ser	Cys	
1025				1030				1035				
Ala Leu	Gln	Glu	Cys	Cys Tyr	Ser	Val	Gln	Gly Asn	Lys	Thr	Asp	
1040				1045				1050				
Val Ser	Glu	Thr	Gly	Ile Gln	Glu	Ala	Ala	His Cys	Asp	Ser	Ile	
1055				1060				1065				
Asn Gln	Thr	Cys	Gln	Thr Ala	Ile	Ser	Gly	Ser Met	Thr	Cys	Gly	
1070				1075				1080				
Asn Asn	Lys	Ser	Leu	Asp Ser	Leu	Ser	Ile	His Gly	Cys	His	Ser	
1085				1090				1095				
His Asp	Ser	Pro	Leu	His Lys	Glu	Ser	Asn	Leu Glu	Gln	Lys	Ser	
1100				1105				1110				
Leu Asp	Val	Ala	Gly	Glu Gly	Ile	Lys	Ser	Pro His	Ala	Val	Gly	
1115				1120				1125				
Gln Gly	Cys	Ser	Asp	Lys Glu	His	Asn	His	Ser His	Pro	Glu	Lys	
1130				1135				1140				
Ala Tyr	Asp	Ser	Cys	Ala Thr	Asp	Asp	Cys	Cys Phe	Ser	Val	Gln	
1145				1150				1155				
Val His	Gly	Ile	Asp	Asp Val	Ser	Arg	Ser	Glu Ile	Gln	Glu	Thr	
1160				1165				1170				
Ala His	Cys	Asp	Ser	Thr Lys	Gln	Ser	Thr	Val Ile	Pro	Ser	Ser	
1175				1180				1185				
Cys Glu	His	Glu	Pro	Lys Asp	Gln	Val	Asn	His Cys	Gly	Ser	His	
1190				1195				1200				
Ser Lys	Ser	Ile	Pro	Thr Asp	Glu	Glu	Leu	Ala Lys	Leu	Val	Arg	
1205				1210				1215				
Arg Cys	Cys	Lys	Tyr	Lys Pro	Cys	His	Asp	Val Arg	Ser	Gly	Cys	
1220				1225				1230				
Arg Lys	His	Ala	Ala	Glu Cys	Gly	Pro	Thr	Val Arg	Ser	Thr	Ile	
1235				1240				1245				
Asn Ile	Leu	Arg	Asp	Asn His	His	His	His	Leu Asp	Cys	Ser	Gly	
1250				1255				1260				
Arg Lys	Val	Cys	Ser	Leu Leu	Glu	Lys	Arg	His Ile	Gly	Gly	Cys	
1265				1270				1275				
Cys Asp	Ser	Phe	Arg	Lys Glu	Cys	Cys	Ala	Lys Asn	Asn	His	Leu	
1280				1285				1290				
Gly Ala	Ser	Phe	Gly	Gly Gly	Leu	Ser	Glu	Ile Val	Lys	Gly	Arg	
1295				1300				1305				
Ile Pro	Ala	His	Trp	Arg Pro	Leu	Leu	Val	Asp Pro	Ser	Ser	Val	
1310				1315				1320				
Pro Ser												
1325												

We claim:

1. An isolated polynucleotide selected from the group consisting of:
 - a polynucleotide comprising a sequence having at least 70% sequence identity to SEQ ID NO:1, and encoding a NtHMA transporter having P1B-type ATPase activity;

- a polynucleotide comprising a sequence having at least 70% sequence identity to SEQ ID NO: 3;
- a polynucleotide comprising a sequence having at least 70% sequence identity to SEQ ID NO: 47;
- a polynucleotide comprising a sequence encoding a polypeptide having at least 70% sequence identity to

SEQ ID NO:2, wherein the polypeptide is a NtHMA transporter having P1B-type ATPase activity; and
 a polynucleotide comprising a sequence encoding a polypeptide having at least 70% sequence identity to SEQ ID NO: 49, wherein the polypeptide is a NtHMA transporter having P1B-type ATPase activity.

2. An expression vector comprising the isolated polynucleotide of claim 1.

3. A transgenic plant made by a process comprising introducing the isolated polynucleotide of claim 1.

4. A cell line made by a process comprising introducing the isolated polynucleotide of claim 1.

5. A consumable tobacco product incorporating leaves harvested from the transgenic plant of claim 3.

6. A NtHMA RNAi construct capable of inhibiting the expression of a NtHMA messenger RNA to which it corresponds, the construct comprising:

a first sequence having at least 70% sequence identity to a portion of SEQ ID NO:3 or 47;

a second sequence; and

a third sequence having a reverse complementary sequence of the first sequence, positioned in the same orientation as the first sequence,

wherein the second sequence is positioned between the first sequence and the third sequence, and the second sequence is operably-linked to the first sequence and to the third sequence.

7. The NtHMA RNAi construct of claim 6, wherein:

(a) the first sequence has at least 95% sequence identity to a sequence selected from the group consisting of: exon 1 (SEQ ID NO:5), a fragment of exon 1 (SEQ ID NO:5), exon 2 (SEQ ID NO:7), a fragment of exon 2 (SEQ ID NO:7), exon 3 (SEQ ID NO:9), a fragment of exon 3 (SEQ ID NO:9), exon 4 (SEQ ID NO:11), a fragment of exon 4 (SEQ ID NO:11), exon 5 (SEQ ID NO:13), a fragment of exon 5 (SEQ ID NO:13), exon 6 (SEQ ID NO:15), a fragment of exon 6 (SEQ ID NO:15), exon 7 (SEQ ID NO:17), a fragment of exon 7 (SEQ ID NO:17), exon 8 (SEQ ID NO:19), a fragment of exon 8 (SEQ ID NO:19), exon 9 (SEQ ID NO:21), a fragment of exon 9 (SEQ ID NO:21), exon 10 (SEQ ID NO:23), a fragment of exon 10 (SEQ ID NO:23), exon 11 (SEQ ID NO:25), and a fragment of exon 11 (SEQ ID NO:25);

(b) the second sequence has at least 95% sequence identity to a sequence selected from the group consisting of: intron 1 (SEQ ID NO:4), a fragment of intron 1 (SEQ ID NO:4), intron 2 (SEQ ID NO:6), a fragment of intron 2 (SEQ ID NO:6), intron 3 (SEQ ID NO:8), a fragment of intron 3 (SEQ ID NO:8), intron 4 (SEQ ID NO:10), a fragment of intron 4 (SEQ ID NO:10), intron 5 (SEQ ID NO:12), a fragment of intron 5 (SEQ ID NO:12), intron 6 (SEQ ID NO:14), a fragment of intron 6 (SEQ ID NO:14), intron 7 (SEQ ID NO:16), a fragment of intron 7 (SEQ ID NO:16), intron 8 (SEQ ID NO:18), a fragment of intron 8 (SEQ ID NO:18), intron 9 (SEQ ID NO:20), a fragment of intron 9 (SEQ ID NO:20), intron 10 (SEQ ID NO:22), a fragment of intron 10 (SEQ ID NO:22), intron 11 (SEQ ID NO:24), a fragment of intron 11 (SEQ ID NO:24), intron 12 (SEQ ID NO:26), and a fragment of intron 12 (SEQ ID NO:26);

(c) the third sequence has at least 95% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:27, a fragment of SEQ ID NO:27, SEQ ID NO:28, a fragment of SEQ ID NO:28, SEQ ID NO:29, a

fragment of SEQ ID NO:29, SEQ ID NO:30, a fragment of SEQ ID NO:30, SEQ ID NO:31, a fragment of SEQ ID NO:31, SEQ ID NO:32, a fragment of SEQ ID NO:32, SEQ ID NO:33, a fragment of SEQ ID NO:33, SEQ ID NO:34, a fragment of SEQ ID NO:34, SEQ ID NO:35, a fragment of SEQ ID NO:35, SEQ ID NO:36, a fragment of SEQ ID NO:36, SEQ ID NO:37, and a fragment of SEQ ID NO:37;

(d) the first sequence comprises SEQ ID NO:38, the second sequence comprises SEQ ID NO:39, and the third sequence comprises SEQ ID NO:40;

(e) the first sequence comprises SEQ ID NO:42, the second sequence comprises SEQ ID NO:43, and the third sequence comprises SEQ ID NO:44; or

(f) two or three of (a), (b), (c), and (d).

8. The NtHMA RNAi construct of claim 6, wherein the first and the second sequence each have a length selected from the group consisting of 20-30 nucleotides, 30-50 nucleotides, 50-100 nucleotides, 100-150 nucleotides, 150-200 nucleotides, 200-300 nucleotides, 300-400 nucleotides, 400-500 nucleotides, 500-600 nucleotides, and 600-700 nucleotides.

9. A transgenic plant comprising the RNAi of claim 6 having a reduced Cd level in at least a part of the plant compared to the part in a non-transgenic counterpart.

10. A consumable tobacco product incorporating leaves harvested from the transgenic plant of claim 9.

11. The product claim 10, wherein a % Cd reduction is a value of at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%.

12. The product article of claim 10, wherein the Cd content is a value ranging from about 0.01 to about 0.05 ppm, from about 0.01 to about 0.1 ppm, from about 0.01 to about 0.5 ppm, from about 0.01 to about 1.0 ppm, or from about 0.01 to about 5 ppm.

13. A method for reducing Cd levels in at least a part of a plant, comprising:

reducing levels of an NtHMA mRNA in the plant by causing expression of an RNAi construct.

14. The method of claim 13, wherein the RNAi construct comprises:

a first sequence having at least 90% sequence identity to a portion of SEQ ID NO:3 or 47;

a second sequence; and

a third sequence having a reverse complementary sequence of the first sequence, positioned in the same orientation as the first sequence,

wherein the second sequence is positioned between the first sequence and the third sequence, and the second sequence is operably-linked to the first sequence and to the third sequence.

15. The method of claim 13, wherein following expression of the RNAi construct, the part of the plant has a Cd content reduced by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%.

16. The method of claim 13, wherein following expression of the RNAi construct, the part of the plant has a Cd content ranging from about 0.01 to about 0.05 ppm, from about 0.01 to about 0.1 ppm, from about 0.01 to about 0.5 ppm, from

about 0.01 to about 1.0 ppm, or from about 0.01 to about 5 ppm.

17. A substantially purified polypeptide having at least 70% sequence identity to SEQ ID NO:2 or 49, wherein the polypeptide is a NtHMA transporter having P1B-type ATPase activity.

18. The substantially purified polypeptide of claim **17**, wherein the polypeptide has at least 95% sequence identity to SEQ ID NO:2 or 49.

19. An antibody that specifically binds to a polypeptide comprising SEQ ID NO:2 or 49.

* * * * *

Transcript levels of *AtMRPs* after cadmium treatment: induction of *AtMRP3*

L. BOVET^{1,4}, T. EGGMANN¹, M. MEYLAN-BETTEX¹, J. POLIER¹, P. KAMMER², E. MARIN³, U. FELLER⁴ & E. MARTINOIA⁵

¹Plant Physiology Laboratory, University of Neuchâtel, Emile Argand 13, 2007 Neuchâtel, Switzerland, ²BES, University of Bern, Gertrudwokerstrasse 5, 3012 Bern, Switzerland, ³CEA-Cadarache, St-Paul Lez Durance, France, ⁴Institute of Plant Sciences, University of Bern, 31 Altenbergrain, 3013 Bern, Switzerland and ⁵Laboratory for Molecular Plant Physiology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

ABSTRACT

In yeasts, the ABC-type transporters are involved in vacuolar sequestration of cadmium. In plants, transport experiments with isolated vacuoles indicate that this is also true. In order to know more about the response of *AtMRPs*, a subclass of *Arabidopsis* ABC transporters, to cadmium, their expression pattern was analysed using the microchip technology and semi-quantitative reverse transcriptase-polymerase chain reaction. From 15 putative sequences coding for *AtMRPs*, transcript levels were detected for 14. All were expressed in the roots as well as in the shoots, although at a different level. In 4-week-old *Arabidopsis*, transcript levels of four *AtMRPs* were up-regulated after cadmium treatment. In all cases up-regulation was exclusively observed in the roots. The increase of transcript levels was most pronounced for *AtMRP3*. A more detailed analysis revealed that induction of *AtMRP3* could also be observed in the shoot when leaves were cut and cadmium allowed to be taken up in the shoot. In young plantlets, a far higher portion of Cd^{2+} was translocated in the aerial part compared with adult plants. Consequently, *AtMRP3* transcript levels increased in both root and shoot of young plants. This suggests that 7-day-old seedlings do not exhibit such a strict root-shoot barrier as 4-week-old plants. Expression analysis with mutant plants for glutathione and phytochelatin synthesis as well as with compounds producing oxidative stress indicate that induction of *AtMRP3* is likely due to the heavy metal itself.

Key-words: *Arabidopsis thaliana*; ABC transporters; *AtMRP*; cadmium; glutathione; heavy metal transport; phytochelatins.

INTRODUCTION

Due to industrialization Cd^{2+} , one of the most toxic heavy metals, can now be considered to be one of the major environmental pollutant (Alloway & Steinnes 1999). As a non-essential heavy metal, cadmium affects general metab-

olism in plants such as photosynthesis and respiration, thereby reducing growth and development (Prasad & Hagemeyer 1999). Cd^{2+} is suggested to be transported in the roots via divalent cation transporters for Fe^{2+} and Zn^{2+} and possibly Ca^{2+} channels (Korshunova *et al.* 1999; Rogers, Eide & Guerinot 2000; Pence *et al.* 2000). Within the cell, it has been shown that the vacuolar $\text{Ca}^{2+}/\text{H}^{+}$ antiporter CAX2 also exchanges cadmium with protons (Hirschi *et al.* 2000). In most plants, such as pea (Prasad & Hagemeyer 1999), wheat (Keltjens & Beusichem 1998) or *Sinapis alba* (Fargašová 2001), Cd^{2+} is mostly retained in the roots. Contrarily, plants known to accumulate large amounts of heavy metals, so-called hyperaccumulators, translocate a large part of heavy metals from the root to the shoot. The translocation mechanism remains to be elucidated, but it is known that transport of cadmium to the aerial part of the plant is influenced by factors such as transpiration, root uptake and xylem loading (Salt *et al.* 1995; Raskin, Smith & Salt 1997; Prasad & Hagemeyer 1999).

All plants described so far have evolved mechanisms to cope at least with low amounts of heavy metals. It is generally accepted that phytochelatins (PCs), also classified as class III of metallothioneins (MTs) and constituted by $(\gamma\text{glu-cys})_n\text{gly}$, play a major role in heavy metal detoxification. Indeed, PCs form complexes with thio-reactive toxic metals such as Cd(II) or Hg(II) , which are then likely to be transported into vacuoles (Rauser 1995; Salt & Rauser 1995; Cobbett 1999; Cobbett 2000a). Until very recently, PCs had been described only in plants and some fungi. The observation that *Caenorhabditis elegans* contains a gene for PC-synthesis and that a knockout of this gene strongly affects the heavy-metal tolerance of *C. elegans* indicates that PCs may be found in more organisms than originally postulated (Vatamaniuk *et al.* 2001; Clemens, Schroeder & Degenkolb 2001). On the other hand, class I and II MTs, which in contrast to PC are proteins, have been mainly shown to play a role in the detoxification of heavy metals in fungi and animals, but their role as detoxifiers in plants remains controversial. Indeed, although MTI and II complemented an MT-deficient (*cup1 delta*) mutant of yeast (Zhou & Goldsbrough 1994), their transcripts have not been reported to be markedly induced after heavy metal

treatments (Schäfer *et al.* 1997; García-Hernández, Murphy & Taiz 1998). In comparison, light stress (Dunaeva & Adamska 2001) and natural senescence (Miller, Arteca & Pell 1999) have been shown to stimulate the expression of *MTI* and *MTII*.

Glutathione, which is the major source of non-protein thiols in plants and the precursor for PC synthesis, could also itself play a role in protection against cadmium by forming Cd-GS₂ complexes (Li *et al.* 1997; Xiang & Oliver 1998). Altering glutathione levels by antisensing γ -ECS results in cadmium hypersensitive *Arabidopsis* plants (Xiang *et al.* 2001).

The export of Cd²⁺ from the cytoplasm can occur either into internal (vacuoles) or apoplastic compartments. For this purpose, specific transporters are involved. Data obtained with yeast and fission yeast suggest that ABC transporters are involved in vacuolar transport of Cd-complexes. In *Saccharomyces cerevisiae*, which is not able to synthesize PCs, the MRP homologue YCF1 transports glutathione complexes (Li *et al.* 1997). In contrast, HMT1 mediates apo-phytochelatin and phytochelatin-Cd²⁺ transport in *Schizosaccharomyces pombe* (Ortiz *et al.* 1995). Other ATP-dependent transporters, found in procaryotes and eucaryotes, such as P-type ATPase, have been demonstrated to confer resistance to toxic concentrations of divalent metals in *Escherichia coli* (Sharma *et al.* 2000). In plants, the ABC transporter(s) involved in cadmium transport has not been identified, but AtMRP3 is able to partially complement the cadmium sensitivity of the Δ YCF1 mutant (Tommasini *et al.* 1998). However, their mode of action remains to be elucidated. Recent data support the idea that other classes of transporters, maintaining metal (cation?) homeostasis in plants (Lombi *et al.* 2002), and which can also transport Cd²⁺ such as CAX2 (Hirschi *et al.* 2000), AtNramps (Thomine *et al.* 2000), IRT1 (Rogers *et al.* 2000), and ZNT1 (Pence *et al.* 2000), are involved in Cd²⁺ distribution within plants.

In this study, cDNA-microarray analyses indicate that transcript levels of some members of the *AtMRP* subfamily are increased after Cd²⁺ treatment. Reverse transcriptase (RT)-polymerase chain reaction (PCR) data show that *AtMRP3* exhibits the strongest induction and therefore we analysed the transcript pattern of this ABC transporter in more detail.

MATERIALS AND METHODS

Plant material

Plants used in this study were: *Arabidopsis thaliana* (columbia) as wild-type, two EMS-mutants of *Arabidopsis thaliana*, *cad2* and *cad1-3* (Howden *et al.* 1995a,b) and mutants overexpressing sense and antisense γ -ECS (Xiang *et al.* 2001). All mutants have *columbia* as background.

Hydroponic culture

Seeds (*Arabidopsis thaliana*, columbia) were placed on small plastic pots containing sterilized clay beads (Migros,

Marin, Switzerland). The pots were placed at room temperature under 100% humidity and soaked for 7 d in water to allow seed germination. The bottom of the pots were then soaked in the following culture medium: 1.5 mM Ca(NO₃)₂ 4H₂O, 1 mM KNO₃, 1 mM KH₂PO₄, 0.75 mM MgSO₄ 7H₂O, 30 μ M Fe³⁺ethylenediaminetetraacetic acid, 1 μ M MnCl₂, 5 μ M H₃BO₃, 0.175 μ M ZnSO₄ 7H₂O, 0.125 μ M CuSO₄ and 0.2 μ M MoO₃. Plants were transferred weekly to fresh medium to prevent the development of micro-organisms or other contaminants. After 4 weeks, plants were subjected to stress (see figure legends and text). Roots were carefully removed from beads using distilled water and finally separated from leaves. Both organs were rapidly immersed in liquid nitrogen. When RNA extraction was not carried out immediately, plant material was stored at -80 °C for not more than 1 week to prevent any RNA degradation. The cadmium concentrations (from 10 to 60 μ M CdCl₂) that were used in our study had no visible toxic effect during the experiments. However, prolonged exposure affected plant growth at different levels according to the concentrations applied.

Bactoagar culture

Seeds (*Arabidopsis thaliana*, columbia) were sterilized by immersion in a saturated water solution of calcium hypochlorite (CaCl₂O₂) for 10 min. After a quick centrifugation at 3000 $\times$ g, seeds were first washed with 70% ethanol and then four times with sterile distilled water. The seeds were maintained overnight at 4 °C, before laying on 0.4–0.6% (w/v) bactoagar plates. The bactoagar was solubilized in the following nutrient solution (1 L): myo-inositol 0.099 g, 2.5 mM MES-KOH pH 5.7, 10 g sucrose, 0.132 g (NH₄)₂HPO₄ and 0.163 g Hoaglands (H-2395; Sigma, Buchs, Switzerland), before sterilization. After 7 d in the phytotron (25 °C, 16 h light and 70% humidity), whole plants or roots and leaves were collected and immersed in liquid nitrogen before RNA extraction.

Polymerase chain reaction

The specific sense (S) and antisense (AS) primers (5'-3') were designed using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_http://www.cgi) and used to amplify the different AtMRPs by RT-PCR. The nomenclature of AtMRPs was according to Martinoia *et al.* (2002): (AC025295, T4K22.12, AtMRP1) AtMRP1-S, ccgc agaaatcctcttggtcttgatg and AtMRP1-AS, gtgaatcatcacctg tagcttctctg; (AC003096, T29F13, AtMRP2) AtMRP2-S, ccgcagaaatcctcttggtcttgatg and AtMRP2-AS, ccttgtaagtgtg tagtcatcttctg; (AP000375, MJG19.3, AtMRP3) AtMRP3-S, ccactgctctgtgtgactg and AtMRP3-AS, gaggtgtactcagcc acaagc; (AC005309, F17A22.19, AtMRP4) AtMRP4-S, ctgg aacggtgtcaactcaaggatgttg and AtMRP4-AS, attccggcagatcg gagagcgtactct; (AC002411, F20D22.11, AtMRP5) AtMRP5-S, cacttgagcagcattactga and AtMRP5-AS, tcttctaatagccgt gcagga; (AP000375, MJG19.4, AtMRP6) AtMRP6-S, ggcca gagacaattggtgtgc and AtMRP6-AS, accttggtctaggagcaggac;

(AP000375, MJG19.5, AtMRP7) AtMRP-7S, aactggtgtgtct tggacgag and AtMRP7-AS, tcttgaatccgaactgtgtg; (AB023045, MXL8.11, AtMRP8) AtMRP-8S, ctctgcaagg gagctgattagatca and AtMRP8-AS, tggagctatgtagcccttg gaataa; (AL138658, T209.140, AtMRP9) AtMRP9-S, gccactgtctctgttattct and AtMRP9-AS, gagccggcaagttag tagat; (AL163527, F17J16.190, AtMRP10) AtMRP10-S, ccacggcatcgatagataatg, and AtMRP10-AS, gccgattgttaag gagacaa; (AC025295, T4K22.1, AtMRP11) AtMRP-11S, tcgctcacagattgaatacca and AtMRP11-AS, ccaccttgactcattc cattc; (AC025295, T4K22.13, AtMRP12) AtMRP12-S, attcgcgaggaaattcaagtct and AtMRP12-AS, caccacactcattc cattct; (AC006225, T5E7.1, AtMRP13) AtMRP13-S, tca gagccctttctgtttca and AtMRP13-AS, tcgtgtgtggagtagg gaag; (AL162651, F26K9.130, AtMRP14) AtMRP14-S, cgagcgatgtcaacttaagga and AtMRP14-AS, gcaacaacgact gtctctcc. For this study, the selected housekeeping genes were S16 (At5g18380, this 40S ribosomal protein of *Arabidopsis thaliana* belongs to the S9P family of ribosomal proteins) and actin2 (At3g18780), the primers were S16-S: ggcgactcaaccagctactga; S16-AS: cggtactcttctgttaacga; actin2-S, tggatccacgagacaaccta; actin2-AS, ttctgtgaacgattc ctggac, respectively. The primers for PR1 (M90508) were, respectively, PR1-S: ggccttacggggaaaacttag; PR1-AS: cgttcacataatcccacgag. The PCR reactions were performed in a final volume of 25 μ L containing the following mixture: PCR buffer, 0.2 mM dNTPs, 1 μ M of both 5'- and 3'- primers, 1 U *Taq* DNA polymerase (Promega, Catalys, Wallisellen, Switzerland) and adjusted amounts of cDNA. Total RNA was purified from plants using the RNeasy Plant Mini Kit (Qiagen, Basel, Switzerland) and stored at -80°C after quantification by spectrophotometry. After DNase treatment (DNase, RQ1, RNase free, Promega, Catalys), cDNAs were prepared using M-MLV reverse transcriptase, RNase H minus, point mutant (Promega, Catalys) as indicated by the manufacturer and stored at -20°C . The cDNAs used in the PCR reaction were between a 1/10 and 1/100 dilution. After 2 min denaturation at 95°C , 35 PCR cycles (94°C for 45 s, 58°C for 45 s and 72°C for 1 min) were run. The PCR products for all 14 AtMRPs were sequenced to confirm that they corresponded to the right amplified fragments.

Semi-quantitative RT-PCR

The RT reactions were conducted exactly as indicated above. However, PCR amplification were slightly modified, 0.5 MBq [$\gamma^{32}\text{P}$]ATP (110 TBq mmol^{-1}) were added to the PCR mixture using the following set of dNTPs (0.07 mM dATP, 0.2 mM dGTP, 0.2 mM dTTP and 0.2 mM dCTP). The number of PCR cycles was 20, to be in the exponential phase of PCR for both the housekeeping gene (*S16*, *Actin2*) and the genes of interest (*AtMRP3*, *PR1*). This point was determined by sampling aliquots during the PCR reaction after 16, 18 and 20 cycles (data not shown). The radioactive PCR products were then separated by gel electrophoresis in 2% agarose gels. After incubation in 0.25 M HCl for 10 min and two washes in $2\times$ SSC, the gels were transferred

onto nitrocellulose membranes (Porablot NY plus; Macherey Nagel, Oensingen, Switzerland) under vacuum for 2 h. The membranes were then subjected to overnight autoradiography (BIOMAX-MR, Kodak, Integra Biosciences, Wallisellen, Switzerland). The labelling intensities were finally quantified using phosphorimager (GS-250 Molecular Imager; Bio-Rad, Heinach, Switzerland). The expression of the gene of interest was the mean of phosphorimager data $\pm$ SD ($n = 3$), as the ratio between the gene of interest and the housekeeping gene. The addition of an artificial RNA within the total RNA preparation makes it possible to control the quality of the reverse transcription and to confirm the use of this technique as quantitative.

CDNA-microarray

In this experiment, plants were grown for 4 weeks in aerated hydroponic culture (8 h light) at 22°C . Total RNA was prepared according to Verwoerd, Dekker & Hoekema (1989). The amplification of cDNA clones or the 3'- ends of the predicted transcribed sequences, selection of housekeeping genes, microarray preparation, mRNA isolation, preparation of fluorescent probes, hybridization reaction and microarray analysis were performed as described by Reymond *et al.* (2000). For more details, see <http://www.unil.ch/ibpv>. For our purpose, a fluorescent-labelled cDNA probe was prepared from mRNA isolated from control *Arabidopsis* roots by reverse transcription in the presence of Cy3-dCTP. A second probe, labelled with Cy5-dCTP, was prepared from roots of plants treated with $50\text{ }\mu\text{M}$ CdCl_2 . After hybridization of both probes on a cDNA-microarray spotted with 50 ESTs containing gene fragments for ABC-transporters, The 3'- end PCR fragments for *AtMRPs* and miscellaneous genes, in duplicate, a computer image was generated following scanning.

Cadmium measurement by atomic absorption spectrometry

Dried plant samples were weighed and subsequently transferred into glass tubes. After adding 100 μL 1.8 M H_2SO_4 to each tube, the samples were kept for 8 h at 95°C and afterwards for additional 8 h at 550°C . After cooling, 100 μL 10 M HCl and, after mixing, 2 mL H_2O were added to the samples. Cadmium was quantified after appropriate dilution with 0.1 M HCl by atomic absorption spectrometry (228.8 nm) in an air/acetylene flame (SpectraAA 220 FS, Varian, Mulgrave, Australia).

Uptake of $^{109}\text{Cd}^{2+}$ and root : shoot ratio

Four-week-old plants were grown on clay beads. To prevent plant touching during the removal of the roots from the beads using deionized water, a nylon net was placed between beads and leaf materials before germination. A similar approach was used to avoid touching 7-day-old seedlings. In this case a rigid plastic net (1 cm^2) was placed on the bactoagar solution, before sowing the sterilized

seeds. A solution of 0.4% bactoagar allowed the removal of seedlings without damaging the roots. In both case the roots of plants were then soaked in 1/8 MS, 10 mM MES-KOH pH 6.0. After 30 min, the medium was replaced by 1/8 MS, 10 mM MES-KOH pH 6.0, 0.44 MBq carrier free $^{109}\text{CdCl}_2$ (31.3 MBq mL $^{-1}$) for 1 h. After four washings of 30 min in 1/8 MS, 10 mM MES-KOH pH 6.0, 1 mM CaCl_2 , the plants were removed from the net, gently sponged on a filter paper and the leaves were separated from the roots using small scissors. The plant material was finally immersed in 3 mL of scintillation cocktail (Irga safe plus) and subjected to liquid scintillation counting.

Root length measurements

After sterilization, seeds (approximately 20) were placed on 0.8% bactoagar plates. The plates were stored at 4 °C for 12 h (overnight) for synchronization of seed germination and then placed vertically in the phytotron (25 °C, 16 h light and 70% humidity). After 4 d, the position of the apex of the roots were pointed with a marker at the back of the plates and turned right at 90°, if necessary. Finally, the lengths of the roots were measured after another 24 h growth in the phytotron.

RESULTS

Up-regulation of *AtMRPs* after cadmium treatment

The analysis of the transcript levels by RT-PCR using specific primers revealed that 14 out of the 15 putative genes for *AtMRPs* are constitutively expressed in leaves and roots. The amplified PCR products were sequenced and confirmed to correspond to each respective *AtMRP* (data not shown). In order to identify specific *AtMRPs* induced or repressed by cadmium, total RNA from roots and leaves of plants previously treated with 20 μM CdCl_2 for 24 h were isolated. A first, qualitative RT-PCR (35 PCR cycles) showed that cadmium treatment did not modify the transcript level in leaves of *Arabidopsis thaliana*. In contrast, these results suggested that *AtMRP3*, and at a lower level, *AtMRP6*, 7 and 14 were possibly induced in roots (Fig. 1). In parallel with these first RT-PCR analyses, the 14 *AtMRPs* were spotted together with other 50 stress-related or control genes with a robot onto glass slides. The hybridization of the slides was performed using Cy3 (control plants) and Cy5 (50 μM CdCl_2 for 24 h) labelled cDNAs prepared from root mRNA isolated from 4-week-old plants (Fig. 2a). The data revealed that in the set of *AtMRPs*, the highest induction could be observed for *AtMRP3* (2.4–2.5) (Fig. 2b). Some induction could also be observed for *AtMRP4*, *AtMRP9*, *AtMRP12*, *AtMRP13* and *AtMRP14*, but with weaker signals. The results expressed in the first quadrant were similar to those found in the second, although the data of the first quadrant were slightly higher. Cy3 and Cy5 signals for housekeeping genes (actin-2, ubiquitin and ubiquitin-4) were superior to 20 000, whereas

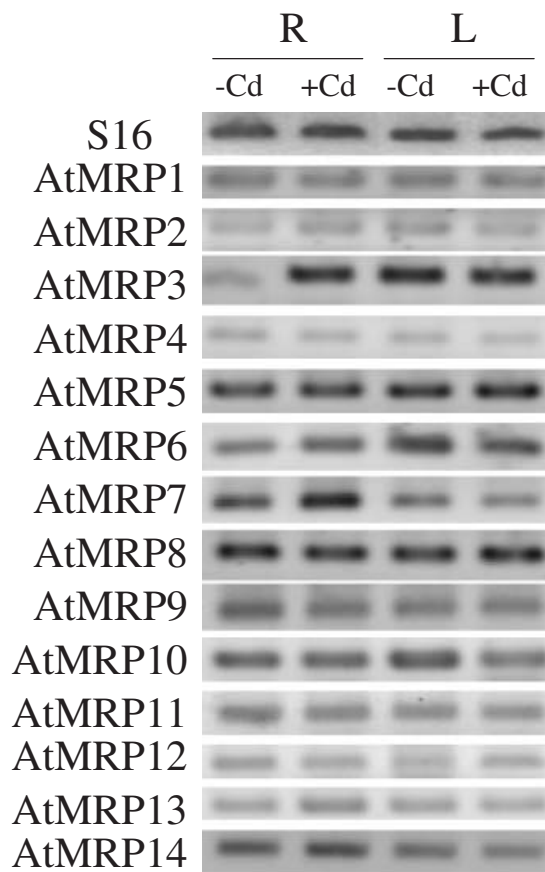


Figure 1. Transcript levels of *AtMRPs* in roots and leaves of 4-week-old *A. thaliana* plants after treatment with 20 μM CdCl_2 for 24 h. Non-quantitative two step RT-PCR (35 cycles) were performed as indicated in material and methods. Amplified DNA fragments were separated using a 1.5% agarose gel and stained with ethidium bromide. The nomenclature of the corresponding genes is according to Martinoia *et al.* (2002).

those for the *AtMRPs* did not exceed 10 000, demonstrating the relatively low expression levels of *AtMRPs*. The microarray data were confirmed by hybridization of a second microarray and by radioactive RT-PCR analysis for *AtMRP3* (Fig. 2c). As observed in qualitative RT-PCR, no cadmium-dependent induction of the 14 *AtMRPs* was detected in leaf tissue using microarray analysis (data not shown).

AtMRP3 induction and ^{109}Cd root : shoot ratio in adult plants

In order to learn more about the mechanisms of cadmium-dependent gene induction and why such a difference between roots and leaves can be observed, we investigated in more detail under which conditions the transcript levels of *AtMRP3* increased. To perform appropriate experiments, the plants were grown hydroponically on clay beads. Raising Cd^{2+} concentrations from 20 to 60 μM in the culture medium shows that *AtMRP3* induction in roots is saturated at 20 μM . In contrast, no induction could be observed in

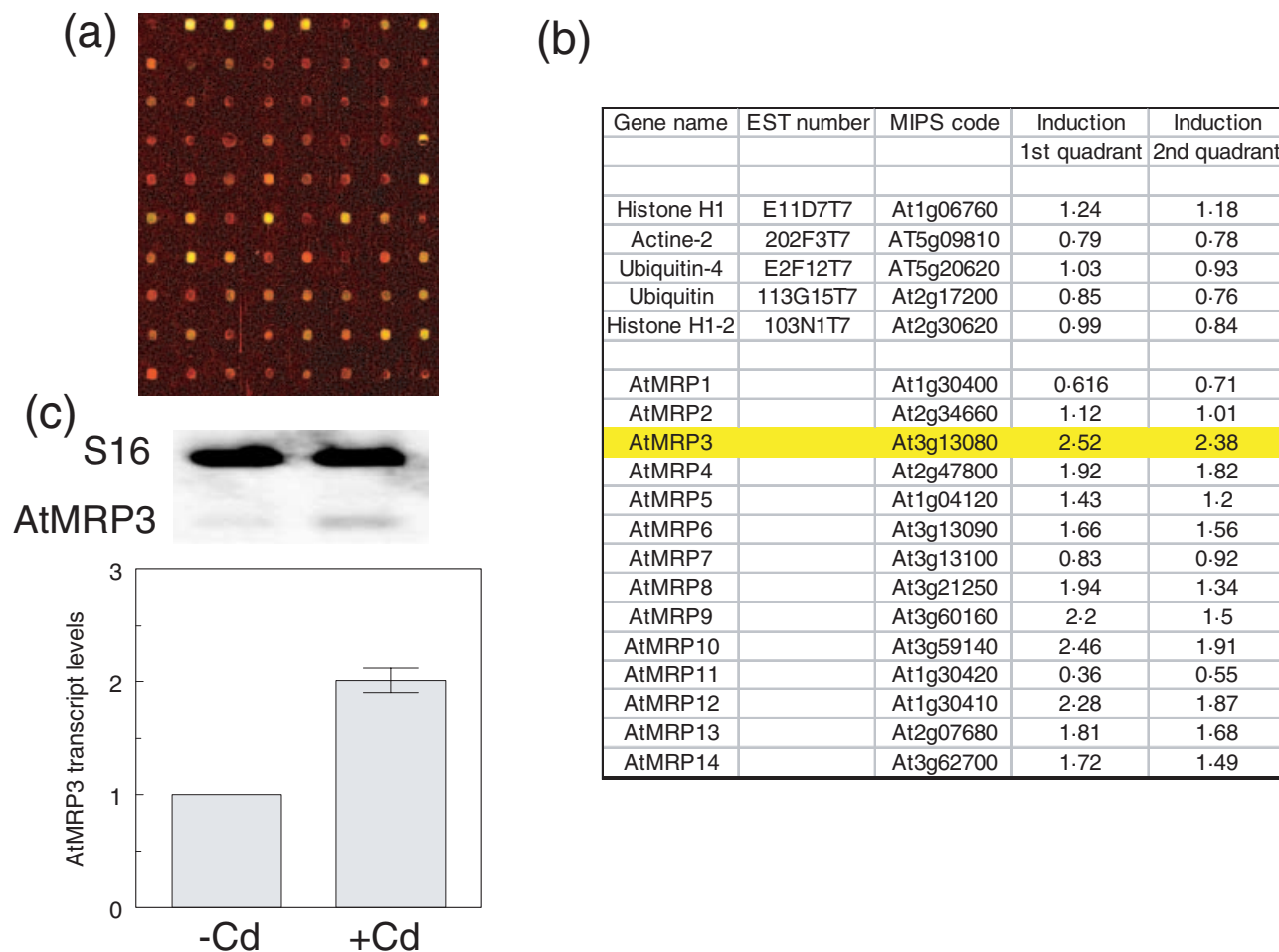


Figure 2. Transcript levels of *AtMRPs* in roots of plants treated with 50 μM CdCl_2 for 24 h using cDNA-microarray. Genes induced or repressed after cadmium treatment are represented as red or green signals, respectively. Genes expressed at approximately equal levels between treatments appear as yellow spots. Control genes which were used for normalization of fluorescent signals in both channels are in the first row of the quadrant [panel (a) shows the first quadrant which contains only one copy of the genes]. Panel (b) shows the expression levels (duplicates on first and second quadrants) of five housekeeping genes (used for normalization of the signals) and the 14 *AtMRP* genes (3'-end amplified fragments by PCR using whole plant cDNAs and amplified EST inserts, as templates). EST numbers and MIPS code are also indicated. Panel (c) shows a two step semi-quantitative RT-PCR with the RNA used as for slide hybridization. The autoradiography of the nitrocellulose membrane (after agarose gel transfer) and histograms resulting from quantification of the phosphorimager data ($n = 3$) are presented.

leaves up to 60 μM (Fig. 3a). A 48 h treatment with 20 μM CdCl_2 resulted in similar expression levels as a 24 h treatment (data not shown). The fact that *AtMRP3* induction is only observed in roots but not in leaves might suggest that Cd is efficiently retained in the root, and hence cannot affect *AtMRP3* expression in leaves. To test this hypothesis, 30 individual 4-week-old plants were incubated with $^{109}\text{CdCl}_2$ for 1 h. The data showed a high variability of radioactive cadmium distribution within the different plants (data for seven plants are shown here). Indeed, the root : shoot ratio varied from about 6 to 25 (Fig. 3b). However, values inferior to 4 were never found in the 30 plants tested. These results demonstrate that in 4-week-old plants Cd is mainly retained in the roots.

To determine whether the presence of Cd^{2+} within the leaves resulted in an increase of the *AtMRP3* transcript levels, we fed cadmium directly to leaves. A net was placed

between the plants and the beads before germination to prevent any contact between the plants and the research worker carrying out the experiment. Indeed, if the plants were touched by hand, increases in the transcript levels of stress-induced genes, which do not commonly respond to cadmium, were found to occur (data not shown). Thus, after carefully removing the beads from the roots with distilled water, the roots of the plants were cut, leaving only the first 0.5 cm of the primary root. These plants, as well as plants with intact roots, were exposed to 10 mM MES-KOH pH 6.0, 10 mM KNO_3 in the presence or absence of 20 μM CdCl_2 for 24 h. Under these conditions a 2.4-fold induction of *AtMRP3* could be observed in leaves of plants, where the roots have been removed, suggesting that under those conditions cadmium concentration in the leaves was increased and resulted in *AtMRP3* up-regulation (Fig. 4a). Indeed, measurement of cadmium in leaves of plants with-

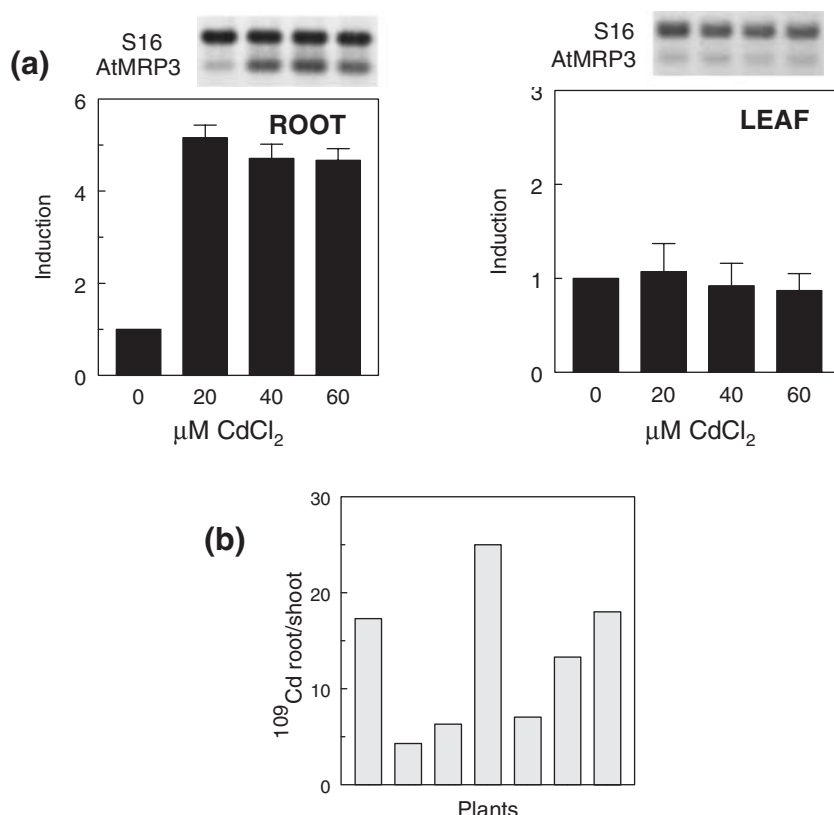


Figure 3. Effect of increasing Cd concentration on *AtMRP3* transcript levels and ^{109}Cd root : shoot ratio of 4-week-old plants. (a) Expression of *AtMRP3* (semi-quantitative RT-PCR) in leaves (right) and roots (left) after addition of 0, 20, 40, 60 $\mu\text{M CdCl}_2$ to the culture medium for 24 h. The 4-week-old plants were maintained in clay beads during the 24 h stress using freshly prepared nutrient solution. (b) ^{109}Cd root/shoot distribution in seven individual plants, after 1 h exposure to $^{109}\text{CdCl}_2$, as described in material and methods (the values obtained were comparable with two other independent experiments)

out roots and treated with CdCl_2 showed that the cadmium content was nearly double in comparison with control plants (Fig. 4b).

AtMRP3 induction and ^{109}Cd root : shoot ratio in young seedlings

The situation was analysed for young seedlings treated with cadmium. When seedlings were cultivated for 7 d

under sterile conditions on a 0.6% bactoagar medium supplemented with 10 $\mu\text{M CdCl}_2$, roots exhibited a strongly reduced growth rate (Fig. 5a), whereas lowering the concentrations reduced the effects. By contrast, under higher CdCl_2 concentration (>10 $\mu\text{M CdCl}_2$) chlorophyll content strongly diminished in addition to reduced growth (data not shown). The *AtMRP3* expression levels were analysed in 7-day-old seedlings treated either with 20 $\mu\text{M CdCl}_2$ for 24 h (Fig. 5b) or from the outset of germination on bacto-

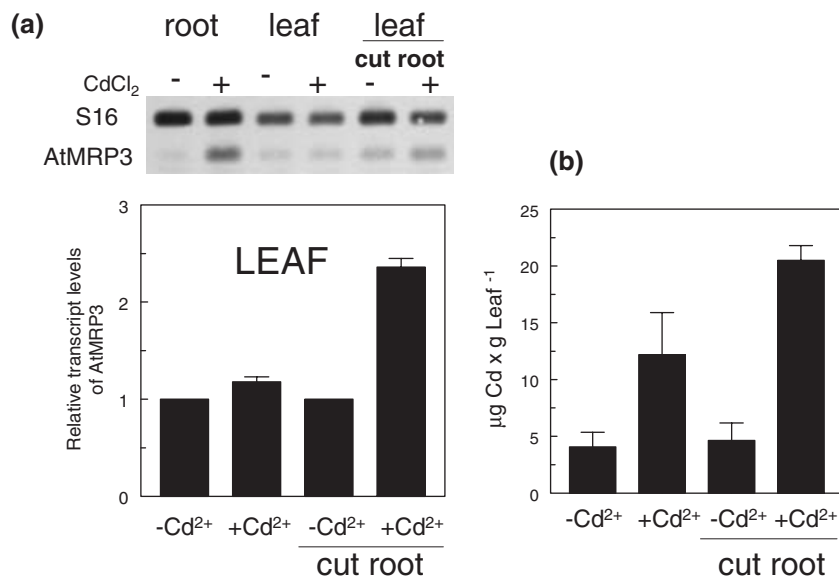


Figure 4. Comparison of *AtMRP3* transcript levels and cadmium contents in leaves of 4-week-old plants with or without cut roots. (a), Transcript levels in leaves of intact plants and plants with cut roots, after CdCl_2 treatment for 24 h. As control, data obtained with plants which were not subjected to CdCl_2 treatment are shown. Two-step semi-quantitative RT-PCR were performed as indicated in materials and methods. (b), Cadmium contents in plants corresponding to those shown in (a) are presented.

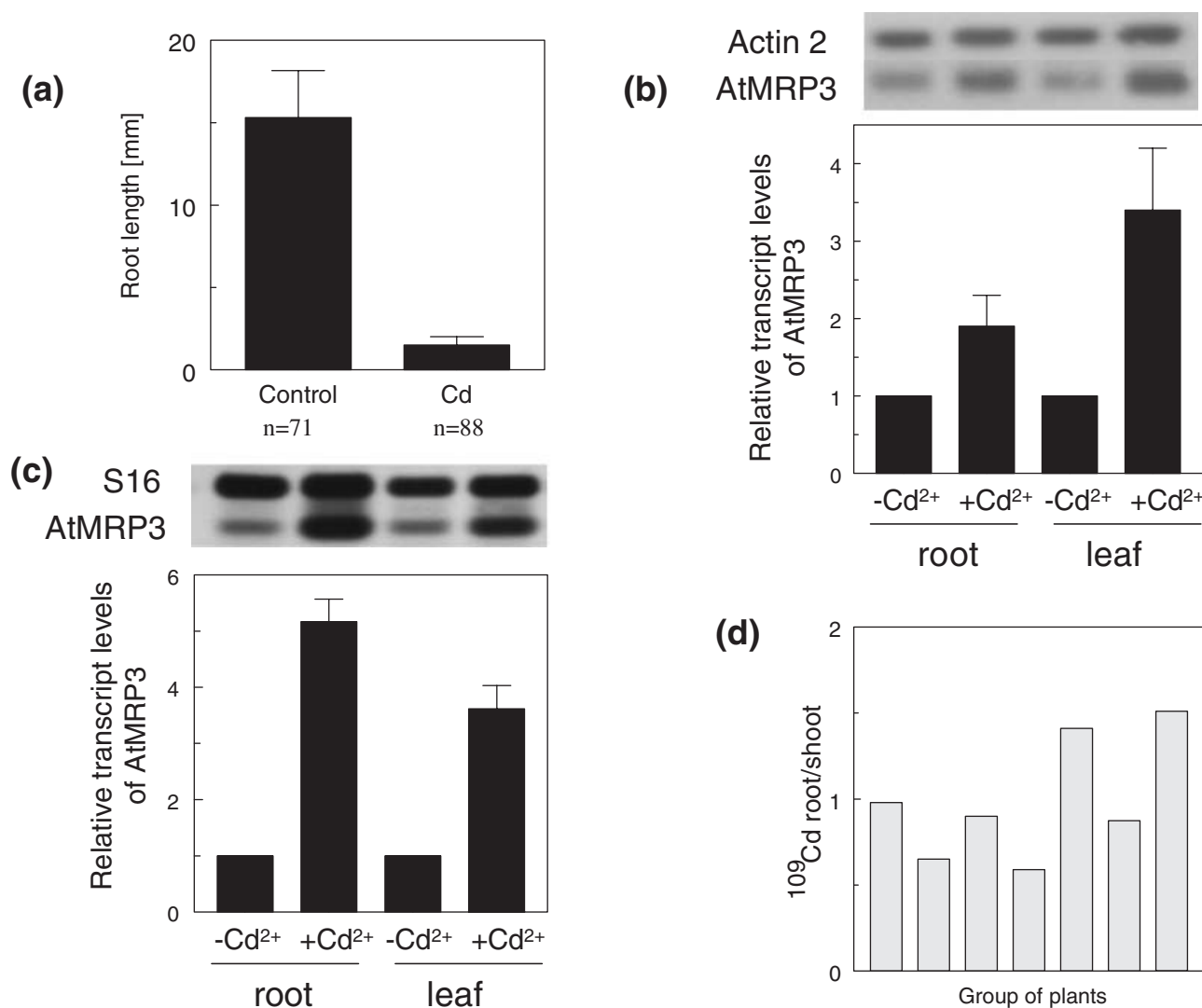
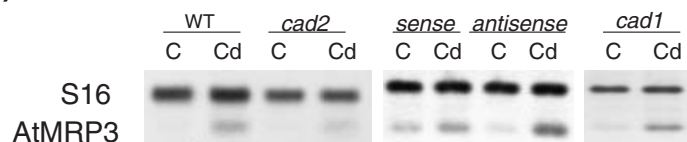


Figure 5. Transcript level of *AtMRP3* in roots and leaves of 7-day-old seedlings after Cd²⁺ treatment. The effect of 10 μ M CdCl₂ on roots grown from the outset of germination was analysed by measuring the root length (a). Semi-quantitative RT-PCRs were performed to analyse the transcript level of *AtMRP3* in 7-day-old plants (6–10 seeds per plastic net, as described in the text). Plantlets were transferred from bactoagar to 10 mM MES-KOH, 10 mM KNO₃ in the presence or absence of 20 μ M CdCl₂ for 24 h (b). Semi-quantitative RT-PCR were performed to analyse the transcript level of *AtMRP3* in 7-day-old seedlings grown onto bactoagar medium supplemented with 10 μ M CdCl₂ from the outset of germination (c). ¹⁰⁹Cd root : shoot distribution within seven groups of plantlets (approximately 6–10 seedlings per plastic net) after 1 h exposure to ¹⁰⁹CdCl₂, as described in material and methods (d).

agar medium supplemented with 10 μ M CdCl₂ (Fig. 5c). In both cases, *AtMRP3* was up-regulated in the root by a factor higher than 2 (Fig. 5b & c), as for 4-week-old plants (Figs 2 & 3). However, in contrast to the results presented above for mature plants, *AtMRP3* was also strongly induced in the leaves of seedlings (about three times more, compare Fig. 3a with Fig. 5b & c). It should be noticed that a longer exposure to cadmium during germination (from 0 to 7 d), compared to a 24 h treatment, did not affect the *AtMRP3* expression level in the leaves (about 3), but seemed to stimulate *AtMRP3* induction in the roots (compare Fig. 5c with Fig. 5b). To demonstrate a possible correlation between Cd²⁺ distribution and *AtMRP3* induction, 7-day-old plantlets were incubated

with ¹⁰⁹CdCl₂. For this purpose, sterilized seeds were placed on 0.4% bactoagar medium covered by 1 cm² plastic nets. After 7 d, the nets containing the seedlings were transferred into liquid medium. Uptake of radioactive cadmium was performed exactly as for 4-week-old plants (see above, Fig. 3b). Determination of the distribution of ¹⁰⁹Cd between roots and shoots resulted in a much lower retention of Cd²⁺ in the roots (0.5–1.5) in comparison with the 4-week-old plants (compare Fig. 5d with Fig. 3b), indicating that the root–shoot barrier for cadmium is either not or only weakly established in young plantlets. Therefore higher cadmium concentrations are present in shoots and likely to be responsible for inducing *AtMRP3* transcript levels.

(a)



(b)

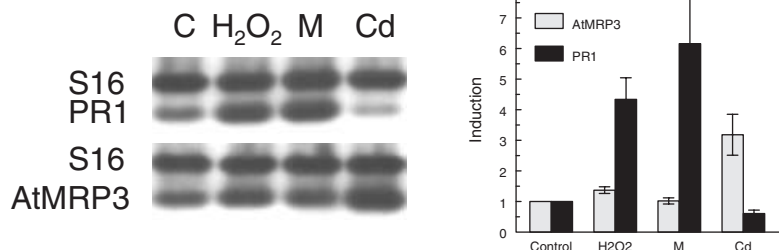


Figure 6. Induction of *AtMRP3* in mutants affected in GSH or phytochelatin levels (*cad2*, γ -ECS sense and antisense plants, *cad1*) (a) and effect of oxidative stress on the expression of *AtMRP3* (b). WT, *cad1* (impaired phytochelatin synthesis), *cad2* (reduced glutathione synthesis) and γ -ECS sense and antisense plants were grown for 7 d on 0.8% bactoagar supplemented with or without 10 μ M CdCl₂. Transcription levels were determined using semi-quantitative RT-PCR (a). To measure the effect of oxidative stress on the expression of *AtMRP3*, total RNA was prepared from 7-day-old seedlings, grown on bactoagar, which were sprayed with 1 mM H₂O₂ or 100 μ M menadione and collected after 24 h. The expression levels of *AtMRP3* and *PR1* were estimated by semi-quantitative RT-PCR (b).

AtMRP3 expression level in *cad1*, *cad2* and γ -ECS transgenic plants

It has been reported that Cd²⁺ detoxification involves glutathione (GSH) and phytochelatin (PC), which form complexes with many heavy metals (Grill, Winnacker & Zenk 1985; Li *et al.* 1997; Cobbett 2000a). *In vivo*, Cd(GSH)₂ and/or Cd-PCs are presumably transported into the vacuole via directly energized transporters (Salt & Rauser 1995; Li *et al.* 1996; Tommasini *et al.* 1996; Rea 1999; Cobbett 2000b). Two EMS-mutants of *Arabidopsis thaliana*, *cad2* and *cad1-3* (Howden *et al.* 1995a,b), knocked out in the first step of glutathione synthesis, γ -glutamyl-cys synthetase (γ -ECS) and phytochelatin synthetase (PCs) activities, respectively, as well as mutant plants overexpressing sense and antisense γ -ECS, were used to investigate whether the complexes play a role in *AtMRP3* induction. All these plants, with the exception of those overexpressing the sense γ -ECS (Xiang *et al.* 2001), have been demonstrated to be less tolerant to Cd in comparison with wild-type plants. These mutants were used to determine whether a disruption in the glutathione or PC biosynthesis pathway affect the induction of *AtMRP3*. Both *cad* mutants, γ -ECS antisense and sense, as well as wild-type plants (*columbia* ecotype) were grown on bactoagar supplemented with 10 μ M CdCl₂ for 7 d. Root growth was strongly affected by Cd²⁺ in all mutants (data not shown). Interestingly, Cd²⁺ induced comparable *AtMRP3* expression in wild-type and mutant plants (Fig. 6a). Treatment of wild-type plants with BSO (L-buthionine sulfoxime), an inhibitor of glutathione biosynthesis, gave similar results. Therefore, our results suggest that the induction of the *AtMRP3* gene is not related to the biosynthesis or levels of GSH or PC and hence not mediated by a heavy metal-GS or heavy metal-PC complexes in *Arabidopsis thaliana*.

Effect of oxidative stress on the *AtMRP3* expression level

The induction of *AtMRP3* could therefore result either from oxidative stress generated by Cd²⁺ (Prasad & Hagemeyer 1999, references herein) or to the heavy metal itself. To answer this question, 7-day-old seedlings were subjected to two different oxidative stress, H₂O₂ (1 mM) and menadione (100 μ M). The transcript levels of *AtMRP3* and *PR1*, a gene known to be induced by oxidative stress (Chamnongpol *et al.* 1998), were compared for their capacity to be induced by Cd²⁺ (similar experimental set up as in Fig. 5). The data show that *AtMRP3* is not induced by menadione and only very slightly by H₂O₂ compared to Cd (Fig. 6b). On the other hand, *PR1* [as well as other oxidative stress-dependent genes like *GST6* and *GST11* (Xiang & Oliver 1998, data not shown)], were not induced by Cd²⁺, but strongly expressed after menadione or H₂O₂ exposure (Fig. 6b). These results strongly suggest that neither oxidative stress (generated by H₂O₂ or menadione), nor GSH and PC levels or complexes are responsible for inducing *AtMRP3*. Interestingly, copper appears to induce *AtMRP3* to a similar level as cadmium in agar plates. In contrast only a slight effect was observed for zinc (data not shown).

DISCUSSION

In *Saccharomyces cerevisiae* (Li *et al.* 1997), *Schizosaccharomyces pombe* (Ortiz *et al.* 1995; Cobbett 2000b) and *Leishmania tarentolae* (Legare *et al.* 2001) it has been demonstrated that ABC transporters play a central role in transport of heavy metals into the vacuole and hence in the final step of heavy metal detoxification. In *Saccharomyces*, YCF1 (yeast cadmium factor 1), the ABC transporter responsible for heavy metal tolerance, belongs to the sub-

family of the MRPs. MRPs are full length ABC transporters containing two cytosolic, two transmembrane regions and a N-terminal extension. YCF1 is able to transport glutathione conjugates (Li *et al.* 1996, Tommasini *et al.* 1996), as well as bis glutathione-cadmium complexes (Li *et al.* 1997). In contrast, HMT1, the ABC transporter responsible for cadmium tolerance in fission yeast, is a half size transporter consisted of only one cytosolic and one transmembrane region. In this case, the substrates transported are phytochelatins (PC), both as apo-PC and as PC-Cd complexes (Ortiz *et al.* 1995). In plants, transport experiments using isolated vacuoles indicate that an ABC-type transporter is involved in vacuolar uptake of PC (Rausser 1995; Salt & Rausser 1995). However no plant transporter for GS₂Cd or PC has been presently identified. In *A. thaliana* it has been shown that AtMRP3 can partially restore the heavy metal hypersensitive phenotype *ycf1*, however, the mechanisms involved have not yet been elucidated.

In *Arabidopsis* grown in phytotron for 4 weeks, 14 AtMRPs are expressed. To analyse whether one of these AtMRPs responds to Cd²⁺ treatment and might be involved in Cd detoxification, RT-PCR and cDNA microarray were performed. Similar results were obtained with the two methods. Indeed, in both cases, AtMRP3 was the most strongly induced MRP. A few other AtMRPs (see Fig. 2) also exhibited increased transcript levels after cadmium treatment, but to a lower extent. In every case AtMRP3 induction was detected only in the roots. This result can be explained by the observations made in different plant species (pea, sugar beet) showing that the bulk of heavy metals is stored in the root, whereas only a minor portion is transferred into the shoot (Prasad & Hagemeyer 1999; Cakmak *et al.* 2000; Fargašová 2001; Vitória, Lea & Azevedo 2001). The expression level of AtMRP3 in the roots of 4-week-old plants was saturated at 20 µM Cd²⁺ and did not more increase when the cadmium concentrations raised up to 60 µM CdCl₂ (Fig. 3). It is worth mentioning that, under our conditions, induction of AtMRP3 was negligible at 10 µM Cd²⁺ for adult plants. This is probably due to a partial adsorption to the clay beads, depletion in the solution or interaction with compounds present in the culture medium, which both are likely to reduce the free cadmium concentration. In a similar experiment using cell cultures of *A. thaliana* treated with 20 µM CdSO₄ for 1 to 4 h, Sanchez-Fernandez *et al.* (1998) did not find any induction of AtMRP3. Three reasons may explain this observation: (i) cell cultures are not suitable to follow transcript levels of AtMRP3 exposed to Cd²⁺ (ii) the time of exposure (4 h) was too short; and (iii) the MS medium efficiently interacts with Cd²⁺, thereby reducing the free heavy metal concentration.

To verify our hypothesis that the absence of AtMRP3 induction in leaves is due to the low cadmium concentrations accumulated within the leaves and not to a root-specific promoter element, we cut the roots of *Arabidopsis* and exposed the base of the root to a cadmium solution. Under these conditions an increase of transcript levels was observed and was correlated with the uptake of cadmium to the aerial part of the plants (Fig. 4). The fact that

AtMRP3 induction can also occur in leaves of 7-day-old seedlings (Fig. 5b & c) confirms this hypothesis, because at this stage Cd²⁺ translocation from the root to the shoot is much higher compared to 4-week-old plants (Figs 3b & 5d). Apparently, the establishment of the root-shoot barrier for cadmium and hence the transport of heavy metals to the shoot depends on the developmental stage of the plant. Furthermore, our results demonstrate that it is not always possible to extrapolate data obtained from seedlings to mature plants.

In addition to heavy metal sequestration, GSH may also play a role in protecting plants against oxidative stress, which can originate from heavy metal uptake (Noctor *et al.* 1998; Dixit, Pandey & Shyam 2001). However, under our experimental conditions, AtMRP3 was not induced by oxidative stress (100 µM menadione or 1 mM H₂O₂), an observation also made by Sanchez-Fernandez *et al.* (1998), except for aminotriazole which induces mild oxidative stress through inhibition of catalases, thus indicating that redox status does probably not influence AtMRP3 expression levels. Results with mutants having reduced levels of GSH and PCs, as well as with plants having reduced or elevated level of GSH after antisensing or overexpressing γ-ECS (Xiang *et al.* 2001), show that it is likely not PC-Cd and also not GS₂Cd which influences AtMRP3 induction. These results strongly indicate that the promoter of AtMRP3 is either inducible directly by free Cd²⁺ or via a cadmium-inducible transcription factor. However, it cannot be excluded that very low levels of Cd-complexes (Cd-GSH₂) may also play a role in gene expression. Recent data obtained with *S. cerevisiae* showed that Cu and Zn homeostasis are controlled by specific genes, Ace1, Mac1 and Zap1 coding for transcription factors, respectively (Bird *et al.* 2000; Gross *et al.* 2000). Although only Zap1 has a homologous gene in *A. thaliana*, which likely play a role in zinc homeostasis (De Pater *et al.* 1996), it is tempting to speculate that transcriptional activators can be activated by other heavy metal divalent cations, including Cd²⁺.

In conclusion, our results show that transcript levels of genes coding for ABC-type transporters might be modulated directly by heavy metals and not through oxidative stress. A direct correlation between heavy metals and the plant ABC transporter transcripts suggest that in plant, as in yeasts, ABC transporters are involved in heavy metal fluxes. However, to date we have no results demonstrating that AtMRP3 is really involved in cadmium transport.

ACKNOWLEDGMENTS

This project was supported by the Swiss National Foundation within the project 31-63708.00 and the NCCR Plant Survival as well as by the Bundesamt für Bildung und Wissenschaft (BBW C990060 and BBW 00-0413/EU proposal Metallophytes, QLRT-2001-02894). The authors are grateful to Dr Xiang for providing us with γ-ECS transgenic plants.

REFERENCES

- Alloway B.J. & Steinnes E. (1999) Anthropogenic additions of cadmium to soils. In *Cadmium in Soils and Plants* (eds M.J. McLaughlin & B.R. Singh), pp. 97–123. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Bird A.J., Zhao H., Luo H., Jensen L.T., Srinivasan C., Evans-Galea M., Winge D.R. & Eide D.J. (2000) A dual role for zinc fingers in both DNA binding and zinc sensing by the Zap1 transcriptional activator. *EMBO Journal* **19**, 3704–3713.
- Cakmak I., Welch R.M., Hart J., Norvell W.A., Ozturk L. & Kochian L.V. (2000) Uptake and retranslocation of leaf-applied cadmium (109Cd) in diploid, tetraploid and hexaploid wheats. *Journal of Experimental Botany* **51**, 221–226.
- Chamnongpol S., Willekens H., Moeder W., Langebartels C., Sandermann H., Van Montagu M., Inze D. & Van Camp W. (1998) Defense activation and enhanced pathogen tolerance induced by H₂O₂ in transgenic tobacco. *Proceedings of the National Academy of Science of the United States of America* **95**, 5818–5823.
- Clemens S., Schroeder J.I. & Degenkolb T. (2001) *Caenorhabditis elegans* expresses a functional phytochelatin synthase. *European Journal of Biochemistry* **268**, 3640–3643.
- Cobbett C.S. (1999) A family of phytochelatin synthase genes from plant, fungal and animal species. *Trends in Plant Science* **9**, 335–337.
- Cobbett C.S. (2000a) Phytochelatins and their roles in heavy metal detoxification. *Plant Physiology* **123**, 825–832.
- Cobbett C.S. (2000b) Phytochelatin biosynthesis and function in heavy-metal detoxification. *Current Opinion in Plant Biology* **3**, 211–216.
- De Pater S., Greco V., Pham K., Memelink J. & Kijne J. (1996) Characterization of a zinc-dependent transcriptional activator from *Arabidopsis*. *Nucleic Acids Research* **24**, 4624–4631.
- Dixit V., Pandey V. & Shyam R. (2001) Differential antioxidative responses to cadmium in roots and leaves of pea (*Pisum sativum* L. cv. Azad). *Journal of Experimental Botany* **52**, 1101–1109.
- Dunaeva M. & Adamska I. (2001) Identification of genes expressed in response to light stress in leaves of *Arabidopsis thaliana* using RNA differential display and natural senescence. *European Journal of Biochemistry* **268**, 5521–5528.
- Fargašová A. (2001) Phytotoxic effects of Cd, Zn, Pb, Cu and Fe on *Sinapis alba* L. seedlings and their accumulation in roots and shoots. *Biologia Plantarum* **44**, 471–473.
- García-Hernández M., Murphy A. & Taiz L. (1998) Metallothioneins 1 and 2 have distinct but overlapping expression patterns in *Arabidopsis*. *Plant Physiology* **118**, 387–397.
- Grill E., Winnacker E.L. & Zenk M.H. (1985) Phytochelatins: the principal heavy-metal complexing peptides of higher plants. *Science* **230**, 674–676.
- Gross C., Kelleher M., Lyer V.R., Brown P.O. & Winge D.R. (2000) Identification of the copper regulon in *Saccharomyces cerevisiae* by DNA microarrays. *Journal of Biological Chemistry* **275**, 32310–32316.
- Hirschi K.D., Korenkov V.D., Wilganowski N.L. & Wagner G.J. (2000) Expression of *Arabidopsis* CAX2 in tobacco. Altered metal accumulation and increased manganese tolerance. *Plant Physiology* **124**, 125–133.
- Howden R., Andersen C.R., Goldsbrough P.B. & Cobbett C.S. (1995b) A cadmium-sensitive, glutathione-deficient mutant of *Arabidopsis thaliana*. *Plant Physiology* **107**, 1067–1073.
- Howden R., Goldsbrough P.B., Andersen C.R. & Cobbett C.S. (1995a) Cadmium-sensitive, cad1 mutants of *Arabidopsis thaliana* are phytochelatin deficient. *Plant Physiology* **107**, 1059–1066.
- Keltjens W.G. & van Beusichem M.L. (1998) Phytochelatins as biomarkers for heavy metal stress in maize (*Zea mays* L.) and wheat (*Triticum aestivum* L.): combined effects of copper and cadmium. *Plant and Soil* **203**, 119–126.
- Korshunova Y.O., Eide D., Clark W.G., Gueriot M.L. & Pakrasi H.B. (1999) The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range. *Plant Molecular Biology* **40**, 37–44.
- Legare D., Richard D., Mukhopadhyay R., Stierhof Y.D., Rosen B.P., Haimeur A., Papadopoulos B. & Ouellette M. (2001) The Leishmania ATP-binding cassette protein PGPA is an intracellular metal-thiol transporter ATPase. *Journal of Biological Chemistry* **276**, 26301–26307.
- Li Z.S., Lu Y.P., Zhen R.G., Szczypka M., Thiele D.J. & Rea P.A. (1997) A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae*: YCF1-catalyzed transport of bis (glutathionato) cadmium. *Proceedings of the National Academy of Science of the United States of America* **94**, 42–47.
- Li Z.S., Szczypka M., Lu Y.P., Thiele D.J. & Rea P.A. (1996) The Yeast Cadmium Factor Protein (YCF1) is a vacuolar glutathione S-conjugate pump. *Journal of Biological Chemistry* **271**, 6509–6517.
- Lombi E., Tearall K.L., Howarth J.R., Zhao F.J., Hawkesford M.J. & McGrath S.P. (2002) Influence of iron status on cadmium and zinc uptake by different ecotypes of the hyperaccumulator *Thlaspi caerulescens*. *Plant Physiology* **128**, 1359–1367.
- Miller J.D., Arteca R.N. & Pell E.J. (1999) Senescence-associated gene expression during ozone-induced leaf senescence in *Arabidopsis*. *Plant Physiology* **120**, 1015–1024.
- Martinoia E., Klein M., Geisler M., Bovet L., Forestier C., Kolukisaoglu U., Muller-Rober B. & Schulz B. (2002) Multifunctionality of plant ABC transporters – more than just detoxifiers. *Planta* **214**, 345–355.
- Noctor G., Arisi A., Jouanin J., Kunert K., Rennenberg H. & Foyer C. (1998) Glutathione: biosynthesis, metabolism and relationship to stress tolerance explored in transformed plants. *Journal of Experimental Botany* **49**, 623–647.
- Ortiz D.F., Ruscini T., McCue K.F. & Ow D.W. (1995) Transport of metal-binding peptides by HMT1, a fission yeast ABC-type vacuolar membrane protein. *Journal of Biological Chemistry* **270**, 4721–4728.
- Pence N.S., Larsen P.B., Ebbs S.D., Letham D.L., Lasat M.M., Garvin D.F., Eide D. & Kochian L.V. (2000) The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator *Thlaspi caerulescens*. *Proceedings of the National Academy of Science of the United States of America* **97**, 4956–4960.
- Prasad M.N. & Hagemeyer J. (1999) *Heavy Metal Stress in Plants* (eds M.N. Prasad & J. Hagemeyer). Springer, Berlin, Germany.
- Raskin I., Smith R.D. & Salt D.E. (1997) Phytoremediation of metals: using plants to remove pollutants from the environment. *Current Opinion in Biotechnology* **8**, 221–226.
- Rauser W.E. (1995) Phytochelatins and related peptides. structure, biosynthesis, and function. *Plant Physiology* **109**, 1141–1149.
- Rea P. (1999) MRP subfamily ABC transporters from plants and yeast. *Journal of Experimental Botany* **50**, 895–913.
- Reymond P., Weber H., Damond M. & Farmer E. (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* **12**, 707–720.
- Rogers E.E., Eide D.J. & Gueriot M.L. (2000) Altered selectivity in an *Arabidopsis* metal transporter. *Proceedings of the National Academy of Science of the United States of America* **97**, 12356–12360.
- Salt D.E., Blaylock M., Kumar N.P., Dushenkov V., Ensley B.D., Chet I. & Raskin I. (1995) Phytoremediation: a novel strategy for the removal of toxic metals from the environment using plants. *Biotechnology (N Y)* **13**, 468–474.
- Salt D.E. & Rauser W.E. (1995) MgATP-dependent transport of

- phytochelatins across the tonoplast of the oat roots. *Plant Physiology* **107**, 1293–1301.
- Sanchez-Fernandez R., Ardiles-Diaz W., Van Montagu M., Inze D. & May M.J. (1998) Cloning and expression analyses of AtMRP4, a novel MRP-like gene from *Arabidopsis thaliana*. *Molecular and General Genetics* **258**, 655–662.
- Schäfer H.J., Greiner S., Rausch T. & Haag-Kerwer A. (1997) In seedlings of the heavy metal accumulator *Brassica juncea* Cu²⁺-differentially affects transcript amounts for -glutamylcysteine synthetase (-ECS) and metallothionein (MT2). *FEBS Letters* **404**, 216–220.
- Sharma R., Rensing C., Rosen B.P. & Mitra B. (2000) The ATP hydrolytic activity of purified ZntA, a Pb (II) /Cd (II) /Zn (II) -translocating ATPase from *Escherichia coli*. *Journal of Biological Chemistry* **275**, 3873–3878.
- Thomine S., Wang R., Ward J.M., Crawford N.M. & Schroeder J.I. (2000) Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to Nramp genes. *Proceedings of the National Academy of Science of the United States of America* **97**, 4991–4996.
- Tommasini R., Evers R., Vogt E., Mornet C., Zaman G.J., Schinkel A.H., Borst P. & Martinoia E. (1996) The human multidrug resistance-associated protein functionally complements the yeast cadmium resistance factor 1. *Proceedings of the National Academy of Science of the United States of America* **93**, 6743–6748.
- Tommasini R., Vogt E., Fromenteau M., Hortensteiner S., Matile P., Amrhein N. & Martinoia E. (1998) An ABC-transporter of *Arabidopsis thaliana* has both glutathione-conjugate and chlorophyll catabolite transport activity. *Plant Journal* **13**, 773–780.
- Vatamaniuk O.K., Bucher E.A., Ward J.T. & Rea P.A. (2001) A new pathway for heavy metal detoxification in animals. phytochelatin synthase is required for cadmium tolerance in *Caenorhabditis elegans*. *Journal of Biological Chemistry* **276**, 20817–20820.
- Verwoerd T.C., Dekker B.M. & Hoekema A. (1989) A small-scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Research* **17**, 2362.
- Vitória A.P., Lea P.J. & Azevedo R.A. (2001) Antioxidant enzymes responses to cadmium in radish tissues. *Phytochemistry* **57**, 701–710.
- Xiang C. & Oliver D.J. (1998) Glutathione metabolic genes coordinately respond to heavy metals and jasmonic acid in *Arabidopsis*. *Plant Cell* **10**, 1539–1550.
- Xiang C., Werner B.L., Christensen E.M. & Oliver D.J. (2001) The biological functions of glutathione revisited in *Arabidopsis* transgenic plants with altered glutathione levels. *Plant Physiology* **126**, 564–574.
- Zhou J. & Goldsbrough P.B. (1994) Functional homologs of fungal metallothionein genes from *Arabidopsis*. *Plant Cell* **6**, 875–884.

Received 21 May 2002; received in revised form 22 August 2002; accepted for publication 26 August 2002



US 20050091709A1

(19) **United States**(12) **Patent Application Publication** (10) **Pub. No.: US 2005/0091709 A1****Song et al.** (43) **Pub. Date: Apr. 28, 2005**(54) **TRANSGENIC ORGANISM EXPRESSING
FUNGAL MRP-LIKE ABC TRANSPORTERS**

Oct. 15, 2002 (KR) 2002/62984

Publication Classification(76) Inventors: **Won Yong Song**, Kyungsangbuk-do
(KR); **Young Yeul Yang**, Kyungki-do
(KR); **Youngsook Lee**,
Kyungsangbuk-do (KR); **InWhan**
Hwang, Pohang-City (KR); **Eun Woon**
Noh, Kyungki-do (KR); **Young Im**
Choi, Kyungki-do (KR); **Eun Hwa**
Jeong, Kyungsangbuk-do (KR); **Enrico**
Martinoia, Zurich (CH)(51) **Int. Cl.⁷** **A01H 1/00**; C12N 15/82;
C07H 21/04; C12N 5/04;
C07K 14/415
(52) **U.S. Cl.** **800/278**; 435/69.1; 435/419;
435/468; 530/370; 536/23.6(57) **ABSTRACT**Correspondence Address:
BAKER & BOTTS
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112

The present invention relates to an isolated DNA encoding fungal MRP (multidrug resistance associated protein) subfamily of ATP-binding cassette (ABC) transporter protein which confers organisms with resistance to and/or accumulation of toxic materials such as heavy metals and herbicides, vectors containing the isolated DNA, and organisms transformed with the isolated DNA. Organisms transformed with the fungal MRP subfamily of ABC transporters of the present invention can be used for remediation of environment polluted with toxic materials. For example, the transformed plants can be used for cleaning polluted soil or water, and thereby providing an environmentally-friendly way to remediate polluted resources with low cost.

(21) Appl. No.: **10/492,880**(22) PCT Filed: **Oct. 16, 2002**(86) PCT No.: **PCT/KR02/01934**(30) **Foreign Application Priority Data**

Oct. 16, 2001 (KR) 2001/68302

FIG.1

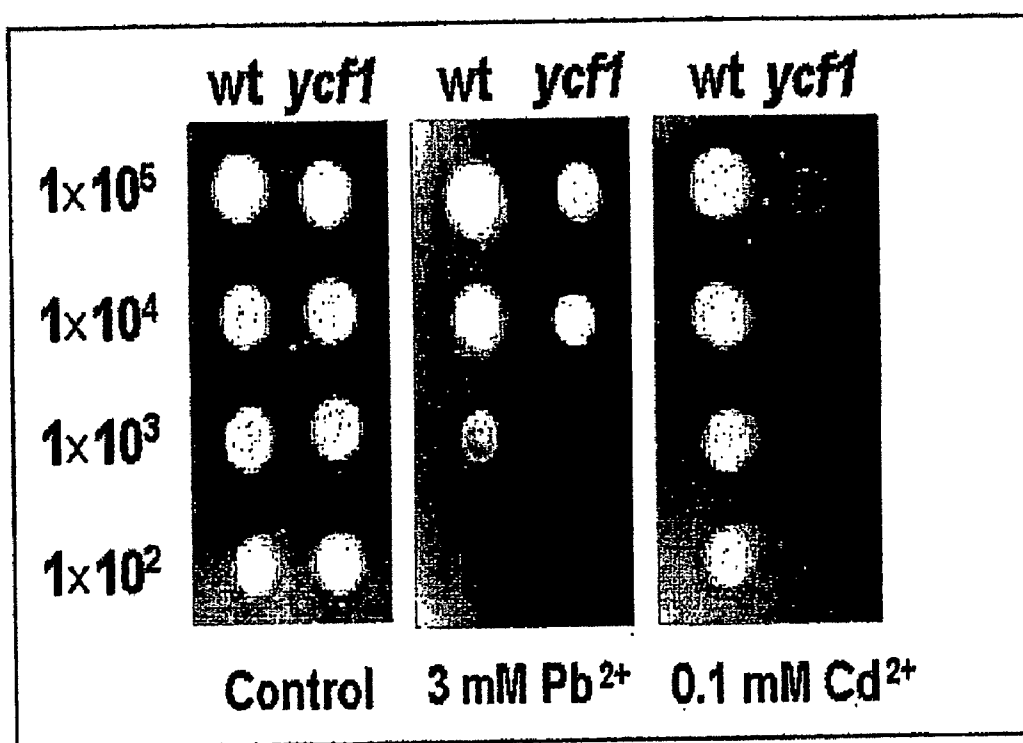


FIG.2

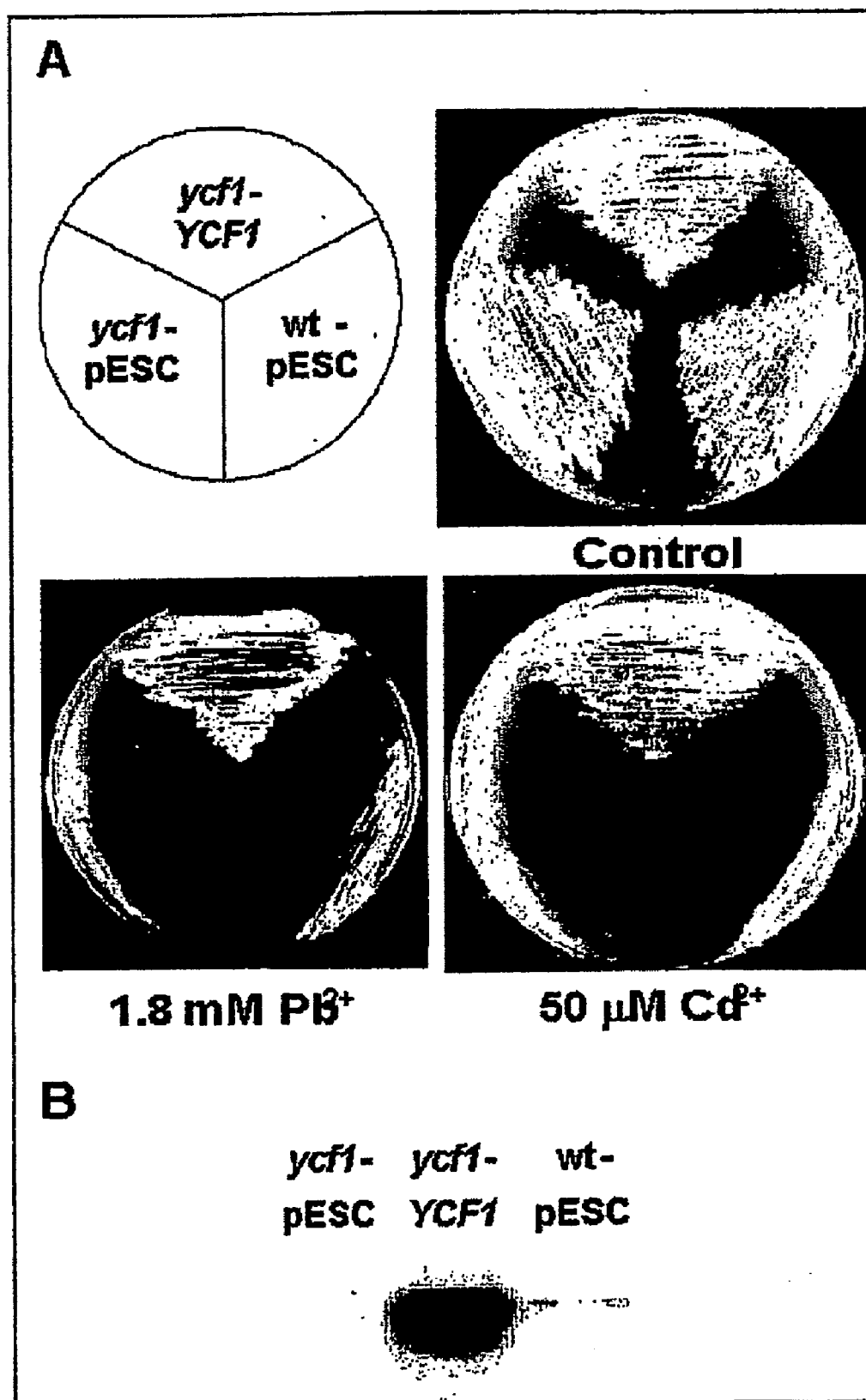


FIG.3

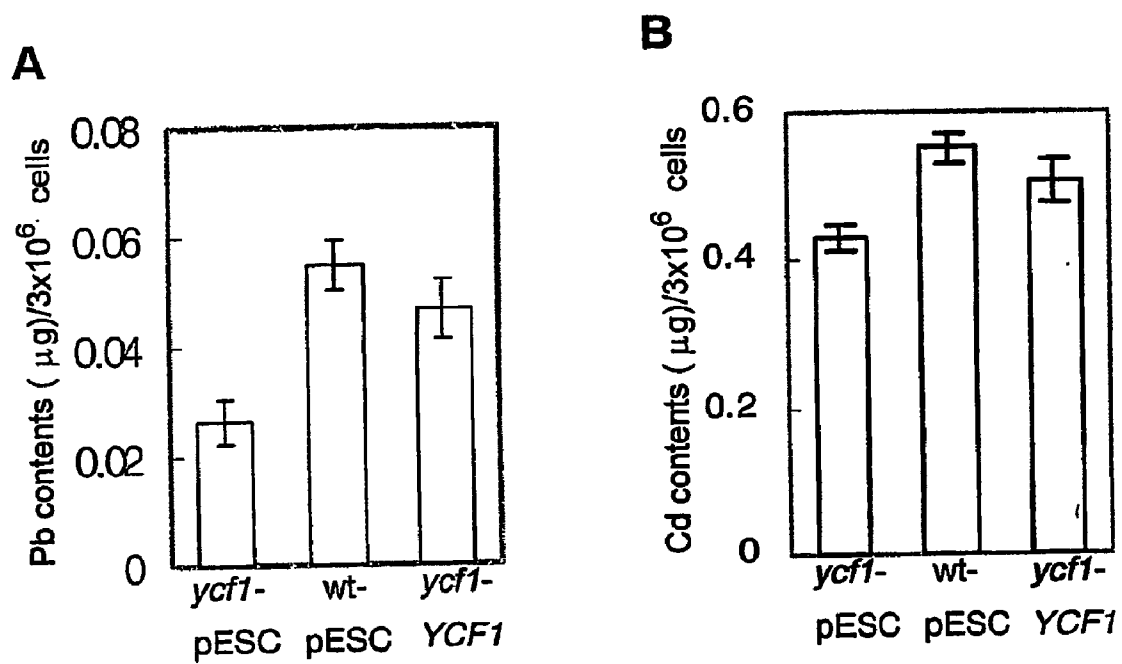


FIG.4

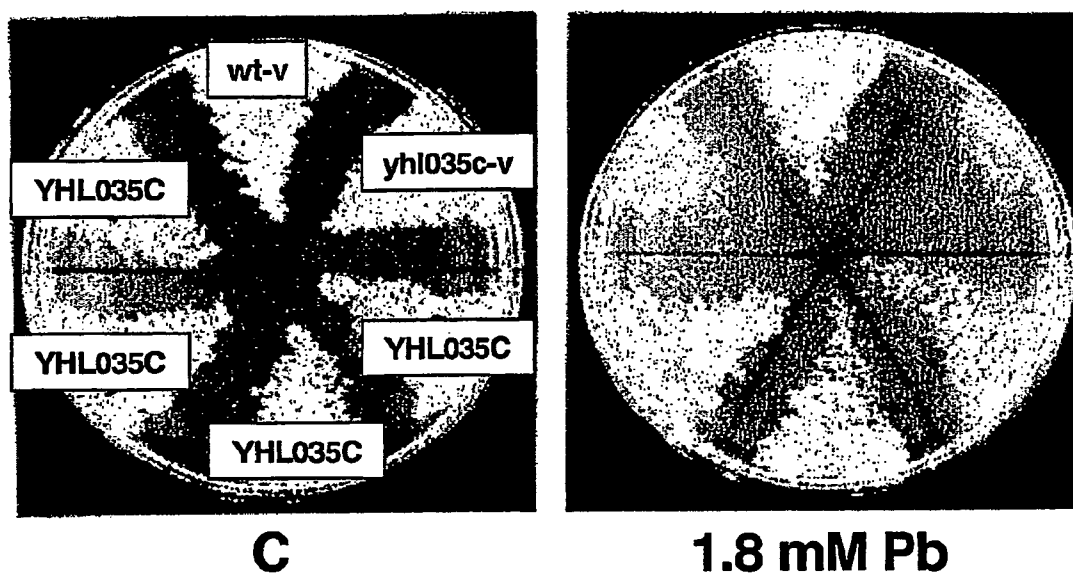


FIG.5

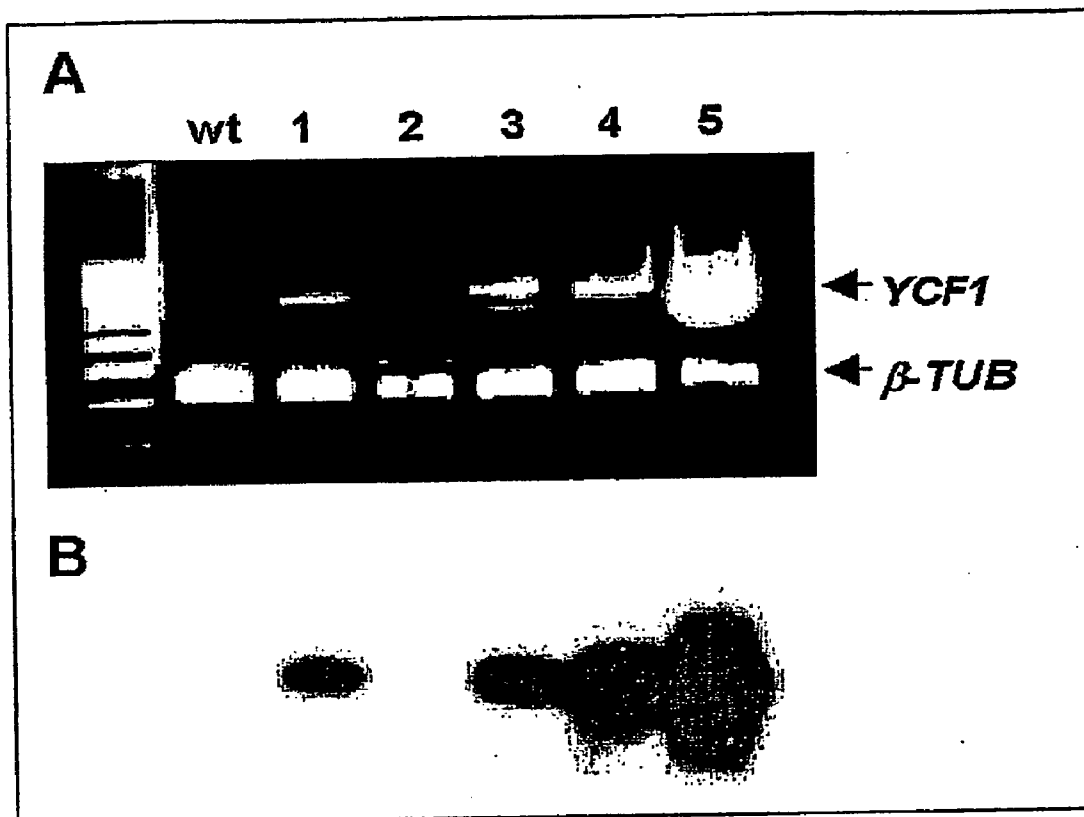


FIG.6

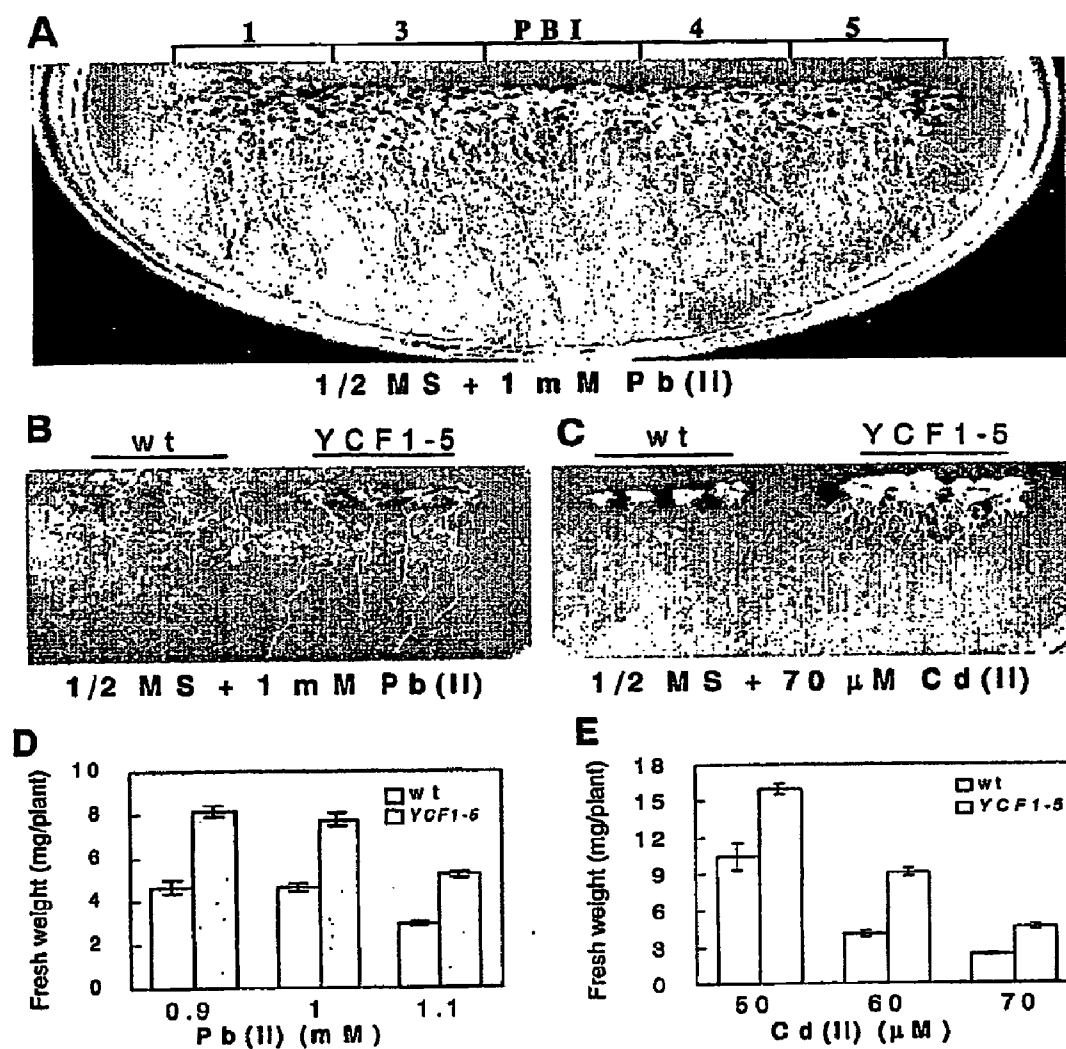


FIG.7

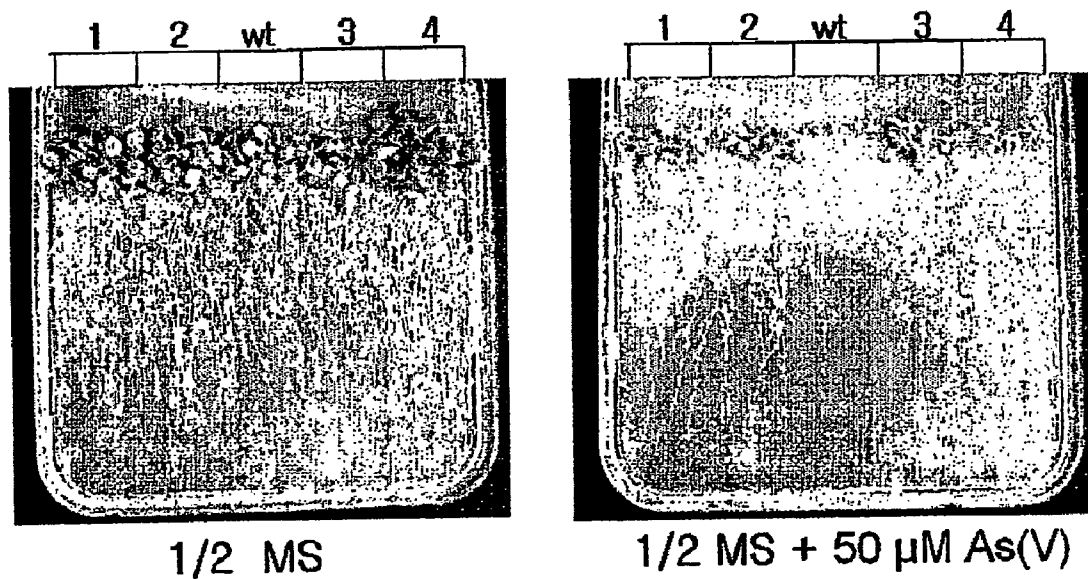


FIG.8

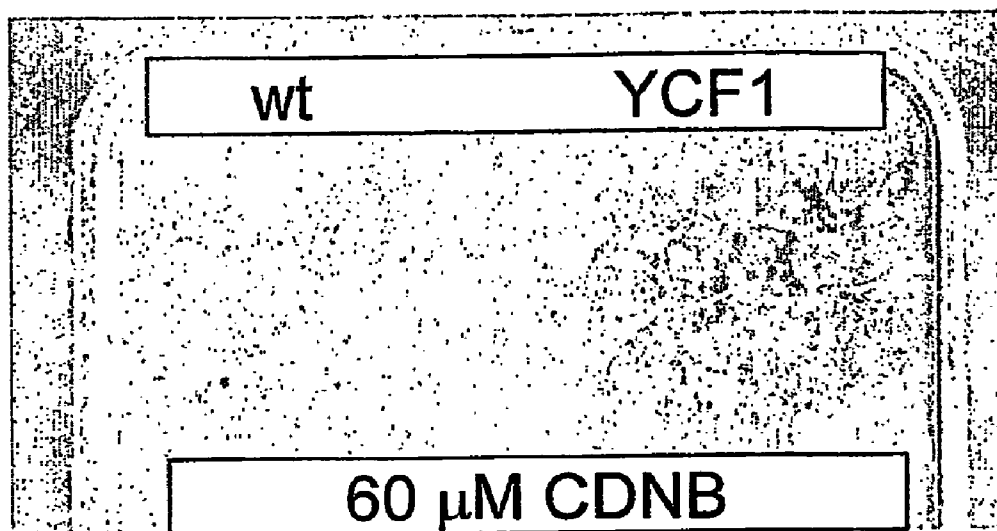
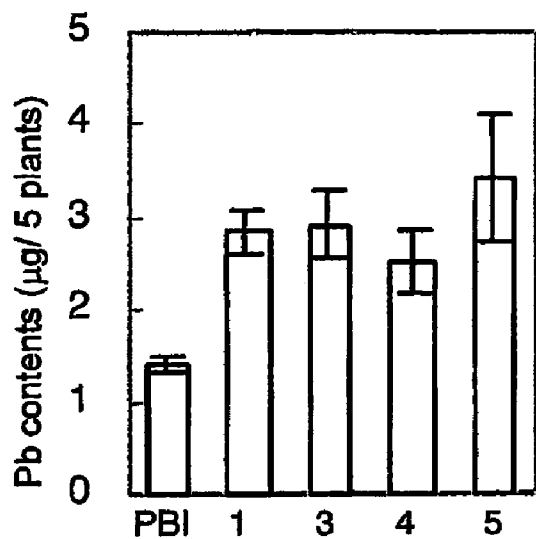


FIG.9

A



B

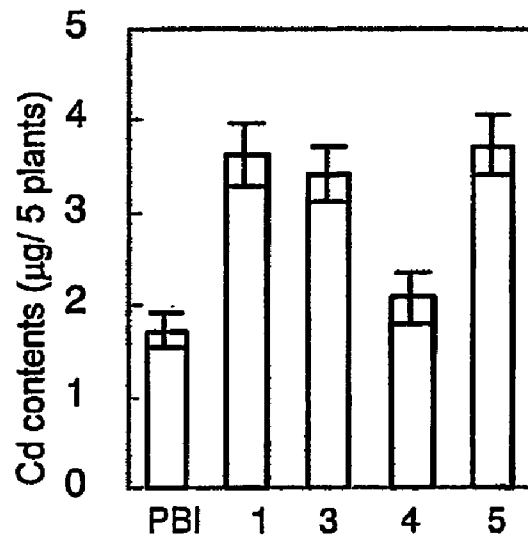
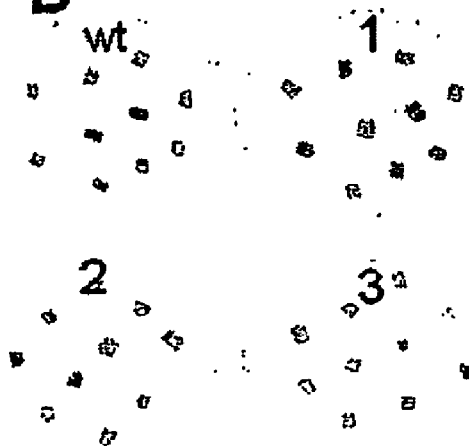


FIG.10

A



B



C

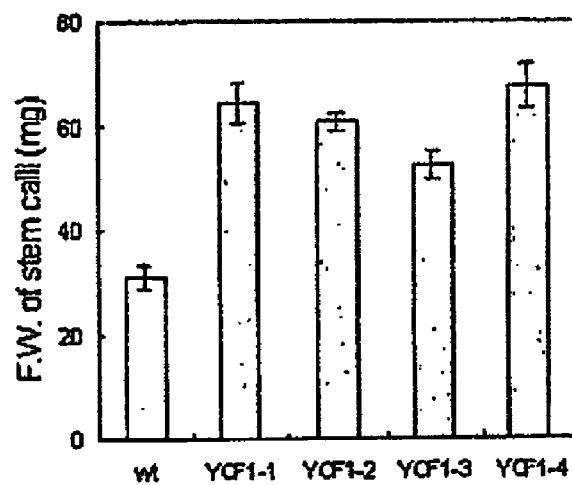
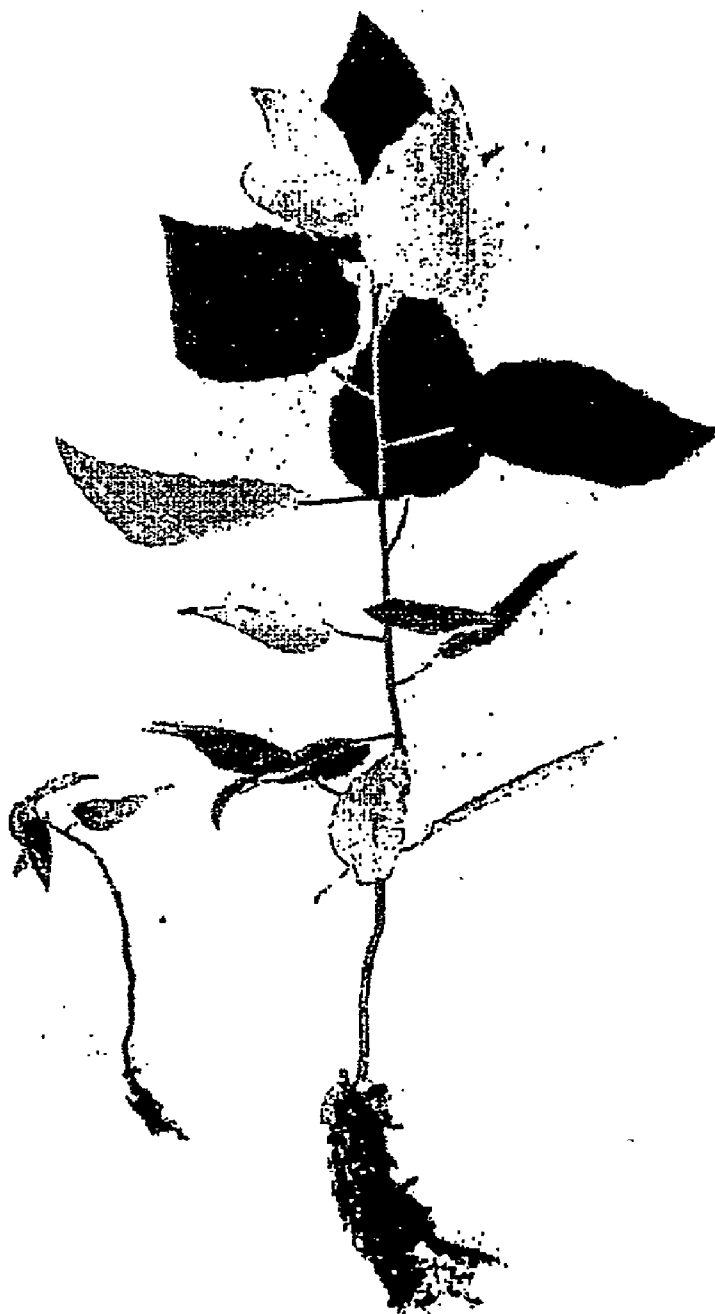


FIG.11



wt

PBI-YHL035C

TRANSGENIC ORGANISM EXPRESSING FUNGAL MRP-LIKE ABC TRANSPORTERS

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application is based on Korean Patent Application Nos. 2001-0063802 and 2002-0062984 filed with Korean Intellectual Property Office on Oct. 16, 2001 and Oct. 15, 2002, respectively.

BACKGROUND OF THE INVENTION

[0002] (a) Field of the Invention

[0003] The present invention relates to a fungal MRP-like ABC transporter gene and organisms transformed with the gene, and more particularly, to transformed organisms expressing fungal MRP-like ABC transporter genes including YCF1 or YHL035C, and thereby having improved resistance to and accumulation of toxic materials such as lead, cadmium, arsenic, and herbicides.

[0004] (b) Description of the Related Art

[0005] Heavy metals such as lead, cadmium, mercury and so on accumulate in the human body through nature's food chain and cause chronic damage to the brain, nerves, bones, etc., and the polluted environment and damage continues from generation to generation. Typical examples of problems caused by heavy metal toxicity are Minamata disease and itaiitai disease, which have occurred in Japan. As lead is a pollutant that causes the most damage among the heavy metals (Salt, D. E., Smith, R. D., and Raskin, I. Phytoremediation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 643-668 (1998)), it is very important to rid lead from the environment. The United States government expends approximately 500 million dollars each year in order to remove lead from environments where children are raised (Lanphear, B. P. The paradox of lead poisoning prevention. *Science*. 281; 1617-1618 (1998)). Additionally, as arsenic pollutes drinking water and causes skin diseases and cancer, it is becoming a serious problem.

[0006] As known genes associated with resistance to or accumulation of toxic materials such as heavy metals and agricultural chemicals, bacterial P-type ATPase has a role in pumping lead to outside of cells at bacterial cell membranes (Rensing, C., Sun, Y., Mitra, B., and Rosen, B. P. Pb(II)-translocating P-type ATPases. *J. Biol. Chem.* 273: 32614-32617 (1998)); genes associated with cadmium resistance include the YCF1 gene of yeast; and genes associated with resistance to arsenic include the YCF1 and ACR genes of yeast, ArsAB of bacteria, etc.

[0007] Living organisms have a mechanism for mitigating toxicity of materials using transporter proteins or biological materials having affinity for noxious materials that invade the body. Use of genes contributing to living organism's resistance against noxious materials would provide an environmentally-friendly way to remediate environments polluted with noxious materials at a very low cost as compared with the physical and/or chemical remediation that is currently widely being employed (Mejare and Bulow, Trends in Biotechnology; 2001, Raskin I. and Ensley B. D. Phytoremediation of Toxic Metals., John Wiley & Sons, New York; 2000). In particular, as plants have many advantages such as their ability to express foreign genes readily and thus

exhibit new phenotypes, they can be produced and maintained at a low cost, they are aesthetically pleasing, etc., research on improvement of plants by inserting useful genes thereinto for use in environmental remediation is being actively conducted. This technique, the use of plants for cleaning up environment, is called "Phytoremediation."

[0008] Under the circumstance where environmental pollution in soil, etc. due to toxic materials such as lead, cadmium, arsenic, and agricultural chemicals is serious, there is a great need for organisms that are transformed by genes that confer resistance to and/or accumulation of these toxic materials.

[0009] Transformed plants that can be used to remove cadmium from the 30 environment have been disclosed in several papers (Zhu et al., (1999) *Plant Physiol.* 119: 73-79, Hirschi et al., (2000) *Plant Physiol.* 124:125-33, Dominguez-Solis et al., (2001) *J. Biol. Chem.* 276: 9297-9302), but there have been no report of transgenic plants that are enhanced in the capacity to remove lead or arsenic from the environment. Further, attempts to develop organisms transformed with YCF1 to improve resistance to not only lead but also to cadmium, arsenic, and herbicides for removal of these toxic materials have not yet been disclosed.

SUMMARY OF THE INVENTION

[0010] The present invention relates to a DNA molecule exhibiting resistance to and accumulation of lead, and encoding fungal MRP-like ABC transporter protein (multi-drug resistance-associated protein ATP-binding cassette transporter protein, MRP-like ABC transporter protein).

[0011] Further, the invention relates to a recombinant vector comprising said DNA molecule encoding MRP-like ABC transporter protein.

[0012] Still further, the invention relates to transformed organisms with improved resistance to and/or accumulation of toxic materials, which are transformed with said DNA molecule encoding fungal MRP-like ABC transporter protein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 shows photographs of wild type and YCF1 null (ycf1) yeasts growing on control, lead or cadmium-containing media, and shows that the ycf1 mutant yeast is more sensitive to lead and cadmium than wild type yeast.

[0014] FIGS. 2A and 2B show photographs of wild type yeast transformed with empty vector (wt-pESC), ycf1 yeast transformed with empty vector (ycf1-pESC), and YCF1-transformed ycf1 (ycf1-YCF1) yeast growing on control, lead or cadmium-containing media (A), and the result from Northern blot of YCF1 mRNA from the yeast lines (B).

[0015] FIGS. 3A and 3B show graphs of lead (A) or cadmium (B) content in ycf1-pESC yeast, wt-pESC yeast, and ycf1-YCF1 yeast.

[0016] FIG. 4 shows photographs of wild type, YHL035C mutant (yhl035C-v), and YHL035-transformed yhl035C-v yeasts growing on control or lead-containing media, and shows that yhl035C-v yeast is more sensitive to lead than wild type or yhl035C-v yeast transformed with YHL035C.

[0017] FIGS. 5A and 5B show photographs of RT-PCR results, which show the expression of YCF1 in YCF1-transformed *Arabidopsis thaliana*.

[0018] FIG. 6 shows photographs (A,B,C) and graphs (D,E) showing that YCF1-transformed *Arabidopsis thaliana* are enhanced in resistance to lead and cadmium.

[0019] FIG. 7 shows photographs showing that YCF1-transformed *Arabidopsis thaliana* are enhanced in resistance to arsenic.

[0020] FIG. 8 shows a photograph showing that YCF1-transformed *Arabidopsis thaliana* are enhanced in resistance to herbicide CDNB.

[0021] FIGS. 9A and 9B show graphs of lead (A) and cadmium (B) content in YCF1-transformed *Arabidopsis thaliana*.

[0022] FIGS. 10A-10C show a series of photographs (A,B) and a graph (C) showing that stem calli and leaf segments from YCF1-transformed poplar are enhanced in resistance to lead compared to those from wild type poplar.

[0023] FIG. 11 shows a photograph showing that YHL035C-transformed poplar plant is more resistant to lead than wild type poplar plant.

DETAILED DESCRIPTION AND THE PREFERRED EMBODIMENTS

[0024] The present invention relates to a gene exhibiting resistance to and/or accumulation of toxic materials, a vector comprising the gene, and cells and organisms transformed therewith.

[0025] The gene exhibiting resistance to and/or accumulation of toxic materials in the present invention is a gene encoding fungal multidrug resistance associated protein (MRP)-like ATP-binding cassette transporter protein (hereinafter referred to as "MRP-like ABC transporter protein").

[0026] The fungal MRP-like ABC transporter protein, which is a kind of ABC transporter protein, has a role in transporting several organic materials present in cytoplasm to the outside of the cytoplasm. ABC transporter proteins exist in several organisms ranging from prokaryotic organisms to human liver cells, and they transport a large variety of materials. Organic materials transported by the ABC transporter proteins include heavy metals conjugated with glutathione, bile acids, etc. YCF1 belonging to the fungal MRP-like ABC transporter proteins is known to transport cadmium, arsenic, and agricultural chemicals into vacuoles and thereby confer resistance to these noxious materials (Li et al., (1997) *Proc. Natl. Acad. Sci. USA* 94: 42-47, Ghosh et al., (1999) *Proc. Natl. Acad. Sci. USA* 96: 5001-5006). However, an attempt to develop organisms with improved resistance to all of cadmium, arsenic, and agricultural chemicals using YCF1 for removal of noxious materials has not yet been disclosed. Furthermore, the contribution of YCF1 to resistance to lead has not yet been reported, and there has been no disclosure regarding the function of YHL035C protein and transformants expressing it.

[0027] The fungal MRP-like ABC transporter genes include, as examples, YCF1 (Yeast cadmium factor 1) and YHL035C genes, and MRP-like ABC transporter genes having at least 28% sequence homology in amino acid sequence with YCF1 protein and YHL035C protein. YCF1 and YHL035C genes, or their proteins, have 28% sequence homology with each other. Also, this invention is intended to include DNA molecules encoding fungal MRP-like ABC

transporter proteins and having at least 28% sequence homology, preferably at least 40% homology, and more preferably at least 50% homology in amino acid sequence with YCF1 or YHL035C protein. For example, BPT1, YBT1, and YOR1 genes, which have at least 28% homology in amino acid sequence with YCF1 or YHL035C proteins, can be included from a comparison conducted in the amino acid database from the GenBank using the CLUSTALW program. The BPT1 protein has 40% homology in amino acid sequence with YCF1, YBT1 protein has 51% homology in amino acid sequence with YHL035C, and YOR1 protein has 28% homology in amino acid sequence with YCF1. YBT1 and BPT1 are fungal MRP-like ABC transporter proteins that are known to transport bile acids (Ortiz et al., (1997) *Journal of Biological Chemistry* 272: 15358-15365, Petrovic et al., (2000) *Yeast* 16: 561-571), and YOR1 protein, also a fungal MRP-like ABC transporter protein, has been known to transport several drugs (Decottignies et al., (1998) *Journal of Biological Chemistry* 273: 12612-12622).

[0028] The MRP-like ABC transporter proteins according to the present invention have been known to have a common domain structure, comprising an N-terminal extension domain, which is a considerably lengthy and hydrophobic domain at the N-terminal; a first transmembrane spanning domain; a first nucleotide binding fold domain, which is a cytoplasmic domain; a second transmembrane spanning domain; and a second nucleotide binding fold domain, which is a cytoplasmic domain located at the C-terminal.

[0029] In a preferred embodiment of the invention, the fungal MRP-like ABC transporter proteins have a sequence homology of at least 28%, preferably at least 40%, and more preferably at least 50%, with the amino acid sequence of YCF1 protein of SEQ ID NO:2 or YHL035C protein of SEQ ID NO:4, each of which comprises an N-terminal extension domain, a first transmembrane spanning domain, a first nucleotide binding fold domain, a second transmembrane spanning domain, and a second nucleotide binding fold domain, and each domain of the fungal MRP-like ABC transporter proteins may have at least 28% sequence homology with the amino acid sequence of each corresponding domain of YCF1 protein of SEQ ID NO:2 or YHL035C protein of SEQ ID NO:4.

[0030] The gene conferring both resistance to and accumulation of toxic materials in the present invention is a YCF1 gene comprising a sequence encoding the polypeptide of YCF1 protein, for example, a nucleotide sequence encoding the polypeptide of SEQ ID NO:2 exhibiting resistance to noxious materials and accumulation of noxious materials; preferably a YCF1 gene exhibiting resistance to or accumulation of one or more noxious materials selected from the group consisting of cadmium, arsenic, and herbicides as well as resistance to and accumulation of lead and comprising a nucleotide sequence encoding YCF1 polypeptide of SEQ ID NO:2; and more preferably a YCF1 gene having the nucleotide sequence of SEQ ID NO:1. The YCF1 gene exists at the sixth chromosome in *Saccharomyces cerevisiae*. The YCF1 gene exhibits its function when expressed, and therefore the present invention is intended to encompass proteins having at least 28% homology, preferably at least 40% homology, and more preferably 50% homology with the amino acid sequence of the YCF1 protein and contributing to resistance to and accumulation of noxious materials, and DNA molecules encoding them.

[0031] YCF1 protein is one of the ABC transporter proteins, it exists at the vacuolar membrane in yeast, and it is known to mitigate the toxicity of cadmium by transporting cadmium conjugated with glutathione present within cytoplasm into the inside of vacuoles using MgATP as the energy source (Li, Z. S. et al. A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae*: YCF1-catalyzed transport of bis(glutathionato)cadmium. *Proc. Natl. Acad. Sci. USA* 94, 4247 (1997)).

[0032] Another example of a gene exhibiting resistance to and/or accumulation of noxious materials is the YHL035C gene, and a YHL035C gene exhibiting resistance to lead and comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:4, and more preferably the nucleotide sequence of SEQ ID NO:3, is provided. The YHL035C gene exists at the eighth chromosome in *Saccharomyces cerevisiae* and is one of the MRP-like ABC transporter proteins. The YHL035C gene exhibits its function when expressed, and therefore the present invention is intended to encompass proteins having at least 28% homology, preferably at least 40% homology, and more preferably 50% homology with the amino acid sequence of YHL035C protein and contributing to resistance to and accumulation of noxious materials, and DNA molecules encoding them.

[0033] In addition, the present invention provides a recombinant vector comprising said DNA molecule encoding the fungal MRP-like ABC transporter protein, and preferably it provides recombinant vectors comprising the YCF1 gene or YHL035C gene. Specific examples of the recombinant vectors include pESC-YCF1, ENpCambia-YCF1, or PBI121-YCF1 recombinant vectors, or the pESC-YHL035C recombinant vector, and pPBI121-YHL035C. The construction of these recombinant vectors can be conducted according to known processes by a person having ordinary knowledge in the art to which the invention pertains.

[0034] In the present invention, noxious materials can include heavy metals including lead, cadmium, arsenic, etc., or agricultural chemicals and herbicides. The herbicides, which are generally lipophilic compounds having a low molecular weight, have been known to readily pass through plant cell walls and hinder plant-specific processes, for example, photosynthetic electron transport or biosynthetic metabolism of essential amino acids, etc., and for example, chlorosulfuron, axidofluorpen, norflurazon, and chlorodinitrobenzene (CDNB) may be included.

[0035] Hence, transgenic organisms capable of exhibiting resistance to noxious materials as well as accumulating noxious materials can be prepared using the DNA molecules comprising nucleotide sequences encoding the polypeptide of YCF1 protein and YHL035C protein of the present invention, or DNA molecules having at least 28% homology therewith, and the transgenic organisms thus prepared can be employed to remediate sites polluted by noxious materials with ease and low cost.

[0036] In addition, the present invention relates to organisms transformed with said DNA molecules encoding the fungal MRP-like ABC transporter proteins. Also, the invention is directed to transgenic cells, preferably plant cells, which are transformed with said DNA molecules encoding the fungal MRP-like ABC transporter proteins. The ABC

genes include all of the foregoing genes, and as examples, the YCF1 gene and YHL035C gene are preferably employed.

[0037] The transgenic organisms are preferably prokaryotic or eukaryotic organisms, and as examples, plants, animals, yeast, *E. coli*, and fungus may be employed. Transgenic plants comprise heterogeneous DNA sequences according to genetic engineering methods, which are constructed to be properly expressed in plant cells, plant tissues, or plant bodies. Plant transformants can be prepared according to known techniques, and *Agrobacterium tumefaciens*-mediated DNA transfer is typically employed. More preferably, recombinant *agrobacterium* constructed by a method selected from the group consisting of electroporation, micro-particle injection, and use of a gene gun is introduced into plants by a dipping method. In an embodiment of the present invention, transgenic plants can be prepared by constructing an expression cassette comprising the MRP-like ABC transporter protein coding sequence which is operably linked to permit its transcription and translation, constructing a recombinant vector comprising said expression cassette, and introducing said recombinant vector into plant cells or plant tissues.

[0038] The above plants include herbaceous plants such as *Arabidopsis*, rapeseeds, leaf mustards, tobaccos, onions, carrots, cucumbers, sweet potatoes, potatoes, napa cabbages, radishes, lettuces, broccoli, petunias, sunflowers, grass, etc., and trees such as olive, willow, white birch, poplar, and birch, and preferably poplar and *Arabidopsis* are employed.

[0039] In a preferred embodiment of the present invention, the transgenic organisms may include YCF1 *Arabidopsis thaliana* (accession number KCTC10064BP), YCF1 poplar, or YHL035C poplar. YCF1 *Arabidopsis thaliana* of the present invention was deposited with the Korean Collection for Type Cultures at the Korea Research Institute of Bioscience and Biotechnology located at 52, Eoun-dong, Yuseong-gu, Daejeon, Korea on Sep. 5, 2001, and assigned Accession Number KCTC10064BP. YCF1 *Arabidopsis thaliana* can be asexually reproduced by tissue culture and grown into a plant according to conventional plant cell culturing methods and differentiation methods. YCF1 *Arabidopsis thaliana* is excellent in resistance to heavy metals and other noxious materials (FIGS. 6, 7, and 8) and accumulation thereof (FIG. 9), as compared with its wild-type counterpart, and in particular, it exhibits resistance to and accumulation of lead, cadmium, arsenic, and agricultural chemicals.

[0040] YCF1 poplar and YHL035C poplar plants of the present invention can be asexually reproduced by tissue culture and grown into plants according to conventional plant cell culturing methods and differentiation methods. YCF1 poplar and YHL035C poplar plants are excellent in resistance to lead as compared with their wild-type counterparts (FIG. 10 and FIG. 11).

[0041] The preferred examples are hereinafter presented for better understanding of the invention. The following examples, however, are provided solely in order for better understanding of the present invention: the present invention should not be construed to be limited thereto.

EXAMPLE 1

Sensitivity to Lead and Cadmium in YCF1 Mutant Yeast

[0042] Wild-type yeast (DTY 165) and *yef1* mutant yeast (DTY 167, MATa *ura3 leu2 his3 trp3 lys2 suc2 yef::hisG*) were cultured in YPD liquid media (1% yeast extract, 2% peptone, 2% dextrose) at 30° C. until OD600 reached 1-2, and then yeast cells of the same number (1×10^2 , 1×10^3 , 1×10^4 or 1×10^5) were cultured in half-diluted YPD solid media containing 3 mM lead at 30° C. for three days. Likewise, they were also cultured in half-diluted YPD solid media containing 0.1 mM cadmium. The experimental results are shown in **FIG. 1**.

[0043] As shown in **FIG. 1**, the growth of *ycf1* mutant yeast was decreased in 3 mM lead-containing media as compared with wild-type yeast, and *ycf1* mutant yeast grew little in 0.1 mM cadmium-containing media. Therefore, it was verified that YCF1 is a gene conferring resistance against lead and cadmium.

EXAMPLE 2

Construction of Cloning Vector

[0044] 2-1: Cloning Vector (PESC-YCF1)

[0045] For the isolation of the YCF1 gene, wild-type yeast was incubated in 3 ml YPD liquid media at 30° C. for 12 hours, and then centrifuged (12,000 rpm, 20-60 sec.). Resultant pellets were suspended in 400 μ l of yeast lysis buffer (1 M sobitol, 0.1 M EDTA, 50 mM DTT (pH 7.5)), and after 40 μ l of zymolase (5 mg/1 ml, 0.9 M sobitol) was added thereto, they were maintained at 37° C. for 15 to 30 minutes. Subsequently, they were mixed with 400 μ l of a urea buffer solution (7 M urea, 0.3125 M NaCl, 0.05 M Tris-HCl (pH 8.0), 0.02 M EDTA (pH 8.0), 1 % Sarcosine) and then mixed with phenol/chloroform to isolate supernatants. The supernatants were mixed with 1 ml of 100% ethanol and centrifuged (12,000 rpm, 10 min.), and then the precipitated DNA was suspended in a TE buffer solution (10 mM Tris-HCl, 1 mM EDTA pH 8.0). PCR was performed using the isolated DNA as a template, YCfa primer (SEQ ID NO:3), YCFb primer (SEQ ID NO:4), and an LA taq polymerase kit (Takara) to isolate the YCF1 gene.

[0046] To prove that the YCF1 gene in yeast is associated with resistance to lead, the YCF1 gene isolated above (1) was cloned into a pESC-URA (yeast shuttle vector, Stratagene) vector. That is, YCF1 PCR products were cleaved with restriction enzymes Xho I and Sca I to prepare YCF1 (Xho I/ Sca I), and a pESC-URA vector was digested with Hind III, treated with Klenow fragments and dNTPs to create a vector with blunt ends, and cleaved with restriction enzyme Xho I. The above YCF1 (Xho I/Sca I) and digested pESC-URA vector (Xho I/blunt-end) were ligated using T4 DNA ligase to construct the recombinant vector pESC-YCF1.

[0047] 2-2: Cloning Vector (PESC-YHL035C)

[0048] The procedures were carried out in a manner substantially identical to Example 2-1 except that PCR was performed using yeast genomic DNA used in Example 2-1 as a template, YHL035Ca primer (SEQ ID NO:9),

YHL035Cb primer (SEQ ID NO:10), and an LA taq polymerase kit (Takara) to isolate the YHL035C gene from yeast

YHL035Ca:
5'-cgacgcgcccgcacatgggaacggatccccttattatc-3'

YHL035Cb:
5'-cgacgcgqccgocatcatcttacttgattgcttgg-3'

[0049] To express the YHL035C gene In yeast, a YHL035C gene was cloned into pESC-URA (yeast shuttle vector, stratagene). YHL035C PCR products were cleaved with Not I and ligated into pESC-URA using T4 DNA ligase to construct the recombinant vector pESC-YHL035C.

EXAMPLE 3

Resistance to Lead and Cadmium in YCF1 Recombinant Yeast

[0050] Recombinant yeast (ycf1-pESC yeast) where an empty vector was introduced into ycf1 mutant yeast, recombinant yeast (wt-pESC yeast) where an empty vector was introduced into wild-type yeast, and a recombinant yeast (ycf1-YCF1 yeast) where YCF1 was overexpressed in ycf1 mutant yeast were each constructed and tested for resistance to lead or cadmium.

[0051] 3-1: Construction of YCF1 Recombinant Yeast

[0052] Yeast was inoculated into a 3 ml liquid YPC media and cultured at 30° C. for 12 hours, and then 0.5 ml of the culture was put Into 10 ml liquid YPD media and cultured at 30° C. for 6 to 8 hours until OD600 reached 0.5 to 0.8. The resultant culture were centrifuged (1,500 rpm, 5 min.) to collect yeasts, which was re-suspended in 5 ml of buffer (0.1 M LiOAc, TE, pH 7.5) and centrifuged. The centrifuged yeast was resuspended in buffer (0.1 M LiOAc, TE (pH 7.5)) and cultured in an agitating incubator at 30° C. for 1 hour, and after plasmid pESC, pESC-YCF1 or pESC-YCF1, and salmon testis DNA were added thereto, it was cultured at 30° C. for 30 minutes. The cultured yeast was mixed with 0.7 ml of buffer (40% PEG 3300, 0.1M LiOAc, TE (pH 7.5)) and incubated while shaking at 30° C. for 1 hour. Thereafter, the above mixture was placed at 42 to 45° C. for 5 minutes to be subjected to heat shock, and centrifuged (2,500 rpm, 5 min.) to collect yeast. The yeast was washed with 1 ml of TE buffer solution (pH 7.5), suspended In 0.2 ml of water or TE buffer solution (pH 7.5), and then cultured on selectable media (CM Ura7-) for 2 to 3 days to select transformed yeasts (ycf1-pESC yeast, wt-pESC yeast, ycf1-YCF1 yeast).

[0053] 3-2: Resistance to Lead and Cadmium in Recombinant Yeast

[0054] The ycf1-pESC yeast, wt-pESC yeast, and ycf1-YCF1 yeast 25 constructed above were each cultured in galactose media (2% galactose, 1% of 0.17% YNB, 0.13% dropout powder, 0.5% ammonium sulfate) to which 1.8 mM lead was added, or in galactose media to which 50 uM of cadmium was added. Control group was cultured in galactose media without heavy metals. Growth degree of each of the recombinant ycf1-pESC yeast, wt-pESC yeast, and ycf1-YCF1 yeast in lead or cadmium-containing media is shown in the photographs of **FIG. 2**.

[0055] It can be seen from **FIG. 2** that ycf1-YCF1 yeast where YCF1 was overexpressed in ycf1 mutant yeast grew

better than ycf1-pESC yeast or wt-pESC yeast in media containing lead or cadmium, which thus supported the previous results that the YCF1 gene has an important role in conferring cadmium resistance. Additionally, it newly revealed that this gene is important to lead resistance.

[0056] 3-3: YCF1 Expression in Recombinant Yeast

[0057] To investigate the expression of YCF1 in ycf1-pESC yeast, wt-pESC yeast, and ycf1-YCF1 yeast constructed in Example 3, Northern Blotting was performed.

[0058] Each recombinant yeast was ground with liquid nitrogen and total RNA extraction buffer solution (0.25 M Tris HCl pH 9.0, 0.25 M NaCl, 0.05 M EDTA, 0.345 M p-Aminosalicylic acid, 0.027 M triisopropyl naphthalene sulfonic acid, 0.02% β -mercaptoethanol, 0.024% phenol) was mixed with phenol/chloroform in a 1:1 ratio. The supernatants obtained from centrifugation at 12,000 rpm for 10 minutes were transferred into a new tube, to which 400 μ l of isopropanol was added. Centrifugation was performed again at 12,000 rpm for 10 minutes to precipitate RNA, which was then lysed in DEPC-treated water and stored in a freezer.

[0059] To perform Northern Blot, 30 μ g of RNA was electrophoresed on agarose gel for RNA and transferred onto a nylon membrane. The nylon membrane was incubated while stirring in a hybridization reaction solution (6 $\times$ SSPE, 0.5% SDS, 10% PEG, 1% nonfat milk, 50% formamide) at 42° C. for 2 hours. Then, YCF1 labeled with ^{32} P dCTP was added thereto and the reaction was performed at 42° C. for 12 hours. After the hybridization reaction, the nylon membrane was washed twice with a buffer (2 $\times$ SSPE and 0.5% SDS), washed with another buffer (0.2 $\times$ SSPE, 0.5% SDS), dried, and then autoradiographed on X-ray film. The experimental results are shown in FIG. 2B.

[0060] From FIG. 2B, Northern Blot photographs of three kinds of recombinant yeasts, it can be seen that ycf1-YCF1 yeast overexpressed YCF1 mRNA. That is, it was proven that lead and cadmium resistance in ycf1-YCF1 yeast, which was exhibited in Example 3-2, is due to the overexpression of YCF1.

[0061] 3-4: Lead and Cadmium Resistance Mechanism

[0062] To investigate whether lead and cadmium resistance conferred by the YCF1 gene is due to the intracellular accumulation of heavy metals or due to extracellular discharge, experiments were conducted.

[0063] Three kinds of yeasts (ycf1-pESC yeast, wt-pESC yeast, and ycf1-YCF1 yeast) were each cultured in 1/2 galactose solid media containing 1.5 mM lead or 15 μ M cadmium for one day, and the cultured yeasts were scraped for harvest. The harvested yeasts were put into 1 ml of concentrated nitric acid, digested for 200° C. for about 6 hours, and then diluted with 10 ml of 0.5 N nitric acid, and the amount of heavy metals contained in the yeasts was measured using an atomic absorption spectrometer (AAS). The measurement results for ycf1-pESC yeast, wt-pESC yeast, and ycf1-YCF1 yeast are represented by graph in FIG. 3.

[0064] Consequently, wt-pESC yeast and ycf1-YCF1 yeast showed a high accumulation of lead and cadmium as compared with ycf1-pESC yeast, and in particular, wt-pESC yeast and ycf1-YCF1 yeast exhibited about a 2-fold higher

accumulation of lead than ycf1-pESC yeast. Therefore, it was verified that lead and cadmium resistance of the YCF1 gene is due to Intracellular accumulation of these heavy metals.

EXAMPLE 4

Lead Resistance in YHL035C Recombinant Yeast

[0065] 4-1: Construction of Recombinant Yeast

[0066] Recombinant yeast (yhl035c-v yeast) where an empty vector was introduced into yhl035c mutant yeast, recombinant yeast (wt-v yeast) where an empty vector was introduced into wild-type yeast, and recombinant yeast (YHL035C yeast) where YHL035C was overexpressed in yhl035c mutant yeast were each constructed according to methods substantially similar to the above Example 3-1, and transformed yeasts (wt-v yeast, yhl035c-v yeast, YHL035C yeast) were selected. Resistance to lead was tested using these recombinant yeasts.

[0067] 4-2: Lead Resistance Test in Recombinant Yeast

[0068] Resistance to lead in wt-v yeast, yhl035c-v yeast, and YHL035C yeast was tested according to methods substantially similar to the above Example 3-2. The growth degree of each of the recombinant wt-v yeast, yhl035c-v yeast, and YHL035C yeast in lead-containing media is exhibited by photographs in FIG. 4.

[0069] As can be seen in FIG. 4, yhl035c-v, which is a yhl035c mutant yeast, was more sensitive than wt-v yeast in media containing 1.8 mM lead. However, YHL035C yeast, where YHL035C was expressed in yhl035c mutant yeast, again recovered resistance to lead and exhibited growth similar to that of wt-v yeast. In media containing 1.8 mM lead. Hence, it newly revealed that the YHL035C gene has an important role in conferring lead resistance.

EXAMPLE 5

Preparation of Transgenic Plants

[0070] 5-1: Construction of YCF1 Vector and YHL035C Vector for Plant Transformation

[0071] To introduce YCF1 gene into plants, 4.6 kb of YCF1, which is a BamHI/SnaB I fragment of pESC-YCF1 plasmid, was inserted into PBI121 (BamHI/Sma I) whereby a PBI121-YCF1 vector was constructed. To improve the expression of the YCF1 gene, an EnPCAMBIA1302-YCF1 vector was constructed. A PCAMBIA1302 vector was digested with restriction enzyme SalI and treated with Klenow fragments and dNTPs to 25 create a vector with blunt ends, and 35S enhancer (BamHI/blunt-end) was inserted into the vector, which was digested with BamHI, to construct the EnPCAMBIA1302 vector. The EnPCAMBIA1302-YCF1 vector was constructed by the insertion of a YCF1 gene (BamHI/blunt-end) at site BglII/PmlI of the EnPCAMBIA1302 vector.

[0072] PBI121-YCF1 and EnPCAMBIA1302-YCF1 vectors were introduced into *E. coli* using an electroporator (BIO-RAD) and cultured on LB solid media. A single colony was inoculated in 3 ml LB (Amp) liquid media, cultured for 12 to 16 hours, and centrifuged to harvest the transformed *E. coli*. Then, 100 μ l of solution I (50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA) was added to the har-

vested *E. coli* to re-suspend it, 200 μ l of solution II (1% SDS, 0.2 N NaOH) was added thereto, the mixture was gently mixed, and then incubated on ice water for 5 minutes. Then, 150 μ l of Solution III (5 M Potassium acetate) was added to the above mixture, which was then slowly mixed 3 to 5 times and centrifuged (12,000 rpm, 10 min.) to collect the supernatant. The supernatant was mixed with 100% ethanol to precipitate DNA, which was then isolated and dried. The PBI121-YCF1 plasmid DNA thus obtained was lysed in TE buffer solution and cut with restriction enzymes BamHI and EcoRI, and the insertion of the YCF1 gene in the correct orientation was confirmed.

[0073] To introduce the YHL035C gene into plants, a pHI121-YHL035C vector was constructed according to methods substantially similar to those used in the construction of the above YCF1 vector for plant transformation, except that the YHL035C gene was cut out from the pESC-YHL035C vector constructed in Example 2 and restriction sites SacI and EcoICR1 were created at both termini thereof, which were then put into a pBI121 vector digested with SacI and SmaI to construct the pBI121-YHL035C vector.

[0074] 5-2: Preparation of Transgenic *Arabidopsis thaliana*

[0075] The vectors PBI121-YCF1 and EnPCAMBIA1302-YCF1 constructed in the above Example 5-1 were introduced into *Agrobacterium* (LBA4404). The transformed *Agrobacterium* was screened on MS (Murashige-Skoog) media containing kanamycin, and transfected into the flower of *Arabidopsis thaliana* by a dipping method (Clough, S. J., and Bent, A. F., Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735-743 (1988)) to introduce the YCF1 gene into the plants. After 4 to 5 weeks, the seeds of the *Arabidopsis* plants were harvested and selected for YCF1 *Arabidopsis* plants using kanamycin for plants transformed with the PBI121-YCF1 vector and hygromycin for plants transformed with the EnPCAMBIA1302-YCF1 vector.

[0076] A total of five lines of YCF1 *Arabidopsis* plants were obtained, and they were tested for expression levels of YCF1 mRNA, and whether the expressed mRNA included the C-terminal 700 bp of YCF1 mRNA to ensure that the full length mRNA was expressed. First, mRNA was extracted from the five lines of YCF1 *Arabidopsis* plants and subjected to RT-PCR using the primers of SEQ ID NO:5/SEQ ID NO:6, and the RT-PCR results are exhibited in FIG. 5A. As shown in FIG. 5A, 700 bp of YCF1 are expressed in plants 1, 3, 4, and 5, and in plants 4 and 5, YCF1 was highly expressed.

[0077] Southern Blotting was carried out for the above RT-PCR products, and the results are exhibited in FIG. 5B. It can be seen from FIG. 5B that the gene expressed in plants 1, 3, 4, and 5 in FIG. 5A is YCF1, and the complete transcription of wild-type YCF1 mRNA occurred, confirming the conclusion from FIG. 5A. This YCF1 *Arabidopsis thaliana* showing high YCF1 expression was deposited to the Korean Collection for Type Cultures at the Korea Research Institute of Bioscience and Biotechnology located at 52, Eoun-dong, Yusung-gu, Daejeon, Korea on Sep. 5, 2001, and assigned Accession Number KCTC10064BP.

[0078] 5-3: Preparation of Transgenic Poplar

[0079] As plants for transformation, Bong-hwa 1, a clone of hybrid poplar (*Populus alba* x *P. glandulosa*), which does

not bloom outdoors, was used after proliferation. To ensure in vitro aseptic materials of poplar, a stalk was obtained from clone bank which was being conserved in nursery at the Korea Forest Research Institute, and surface-sterilized with ethanol (5 minutes) and 2% NaCl (20 min.), and then the stalk, which was developed after 4-weeks of cultivation on MS media, was employed as a specimen for transformation.

[0080] For transformation using *Agrobacterium* containing a gene of Interest, Induction of a callus, and induction of the stalk, etc., the method by Noh et al. (Genetic Engineering of Poplar (2002), written by Noh, Eun-un et al. ISBN#89-8176-098-5 93520) was used. *Agrobacterium tumefaciens* constructed in Example 5-2, into which PBI121-YCF1 and EnPCAMBIA1302-YCF1 vectors were introduced, was inoculated in LB media and cultured overnight at 30° C., and after centrifugation at 1,000 g for 10 minutes, the media were discarded and the pellet was re-suspended in a 0.85% NaCl solution. The suspension was poured into a petri dish, and then the internode tissue of poplar cultivated in vitro was dipped thereinto for 20 minutes. After that, it was placed between two disinfected absorption filtering papers and gently pressed to eliminate excessive *Agrobacterium*, and then they were co-cultivated in callus-inducible media containing no antibiotic (MS+2.4-D 1.0 mg/L, BA 0.1mg/L, NM 0.01 mg/L) (Murashige and Skoog, 1962) for 2 days followed by the selection of transformed cells by use of selectable media containing 50 mg/L kanamycin and 500 mg/L cefotaxime.

[0081] The callus thus formed was grown in stalk-inducible media (WPM+zeatin 1.0 mg/L, BA 0.1 mg/L, NM 0.01 mg/L) (Lloyd and McCown, 1981) containing 50 mg/L of kanamycin to induce stalks. Once the stalks were induced, they were induced to elongate through subcultures in MS basic media containing 50 mg/L kanamycin, and roots were induced by addition of 0.2 mg/L of IBA to the same media, thus obtaining the whole plant body.

EXAMPLE 6

Growth of Transformed *Arabidopsis* Plants in the Presence of Noxious Materials

[0082] YCF1 *Arabidopsis* plants prepared in Example 5 were cultivated in 1/2 MS media containing lead (0.9, 1, 1.1 mM), cadmium (50, 60, 70 μ M), herbicide chloro-dinitrobenzene (CDNB) (60 μ M), and pentavalent arsenic (50 μ M) to investigate their growth degree. As a control, *Arabidopsis* plants transformed with an empty vector (PBI) and wild-type plants were used. The experimental results are shown in FIG. 6 to FIG. 8.

[0083] FIG. 6 shows photographs and graphs exhibiting growth of YCF1-transformed *Arabidopsis thaliana*. As shown in FIG. 6, when YCF1 *Arabidopsis* and control plants were grown in media containing lead for three weeks, YCF1 *Arabidopsis* (1, 3, 4, and 5) showed less chlorosis in leaves and better growth of roots than PBI empty vector transformed plants and wild-type plants (A, B, and D). Also, when YCF1 and wild-type *Arabidopsis* plants were grown in cadmium media at various concentrations, wild-type *Arabidopsis* plants showed more chlorosis in leaves and their roots grew poorly and were shorter than the YCF1 transformants (C and E).

[0084] FIG. 7 shows photographs exhibiting resistance to arsenic in YCF1-transformed *Arabidopsis thaliana*. When grown in media containing 60 μ M pentavalent arsenic for three weeks, wild-type *Arabidopsis* showed a little growth,

whereas YCF1 transformants (1, 2, 3, and 4) showed much better growth as compared with wild-type.

[0085] FIG. 8 shows a photograph exhibiting resistance to CDNB in YCF1-transformed *Arabidopsis thaliana*. When YCF1 *Arabidopsis* and control plants were grown in media containing 60 μ M CDNB for two months, wild-type plants showed poor germination and almost died, whereas YCF1 *Arabidopsis* plants grew well, almost like *Arabidopsis* plants grown in normal condition.

[0086] Hence, it was verified that YCF1-transformed *Arabidopsis thaliana* prepared in Example 5 has resistance to lead, cadmium, arsenic, and herbicides.

EXAMPLE 7

Accumulation of Heavy Metals in Transformed *Arabidopsis* Plants

[0087] YCF1 *Arabidopsis* plants obtained in Example 5 and wild-type *Arabidopsis* transformed with empty vector, PBI, were cultivated in 1/2 MS media containing lead (0.75 mM) and cadmium (70 μ M) for three weeks to investigate accumulation of heavy metals. The amount of accumulation of lead or cadmium was investigated by experiments substantially similar to Example 6, and the results are shown in FIG. 9, in which FIG. 9A shows the lead content and FIG. 9B shows the cadmium content of the plants.

[0088] As shown in FIG. 9A, YCF1 *Arabidopsis* plants showed 2-fold or 1.4-fold higher accumulation of lead than PBI. Also, as shown in FIG. 9B, YCF1 *Arabidopsis* plants showed 2-fold or 3-fold higher accumulation of cadmium than PBI.

[0089] Hence, it was verified that YCF1 *Arabidopsis* plants have higher accumulation of lead and cadmium than wild-type *Arabidopsis*.

EXAMPLE 8

Mechanism of Resistance to Noxious Materials in YCF1-Transformed Plants

[0090] YCF1-transformed plants showed resistance to lead, cadmium, arsenic, and herbicides and showed accumulation of lead and cadmium. To investigate that these phenomena occurred due to the fact that YCF1 proteins transport heavy metals into vacuoles, the vacuoles were separated from YCF1 transformed plants and wild-type plants, and experiments of transporting cadmium and herbicides were carried out. The transport experiment results of cadmium (Cd+GSH) and herbicides (DNB-GS) in the vacuoles of the YCF transformed plants and wild-type plants are shown in Table 1 below.

TABLE 1

Accumulation of GS-associated Compounds in the Vacuoles Separated from <i>Arabidopsis</i> (unit: pmol/1 μ l vacuole/20 min)				
Compounds	Wild-type Plants		YCF1 Transformed Plants	
	-MgATP	+MgATP	-MgATP	+MgATP
DNB - GS	21 $\pm$ 1.1	28 $\pm$ 1.3	24 $\pm$ 0.9	33 $\pm$ 1.2
Cd + GSH	75 $\pm$ 3.3	104 $\pm$ 4.9	71 $\pm$ 3.4	177 $\pm$ 9.8
GSH	—	70 $\pm$ 2.2	—	69 $\pm$ 1.1

[0091] As shown in Table 1 above, as YCF1 employs MgATP as its energy source when transporting materials, there was, in the -MgATP group, no difference in cadmium (Cd+GSH) and herbicide (DNB-GS) contents between YCF1 transformants and wild-type plants. However, when MgATP was added, YCF1 transformants' vacuoles showed about a 1.7-fold higher accumulation of cadmium (Cd-GSH) than the wild-type vacuoles. In case of the herbicide (DNB-GS), YCF1 transformants showed a little higher accumulation than wild-type.

[0092] It was verified from FIGS. 6, 7, 8, and 9 that YCF1 transformed *Arabidopsis* plants have higher resistance to lead, cadmium, arsenic, and herbicides than wild-type, and they accumulate more lead and cadmium. It was supported from the results of Table 1 that such phenomenon is due to the fact that YCF1 proteins expressed in the vacuoles of YCF1 transformed *Arabidopsis* plants transport cadmium and herbicides into the vacuoles and stabilize them, thereby conferring resistance to and accumulation of those noxious materials.

EXAMPLE 9

Growth of Transgenic Poplar Plants

[0093] 9-1: YCF1 Transgenic Poplar Plants

[0094] The leaf fragments and stalks of YCF1 poplar plants prepared in Example 5 were mounted on callus-inducible media containing 500 ppm of lead, and cultivated for two weeks to investigate the difference in their growth. As a control, wild-type poplar plants, which were not transformed, were treated in the same manner, and the results of the experiments are shown in FIG. 10. FIG. 10 shows photographs and a graph showing the growth of YCF1 transformed and control poplar plants.

[0095] As shown in FIG. 10, in media containing lead, YCF1 poplar (1, 2, 3, and 4) showed about a 2-fold higher growth in stalk pieces than wild-type (wt) (A and C), and the leaf fragments of YCF1 poplar (1, 2, and 3) showed less browning than those of wild-type (wt) and kept the color of chlorophyll green (B).

[0096] 9-2: YHL035C Transformed Poplar Plants

[0097] YHL035C transformed poplar plants prepared in Example 5 were transferred to a pot and allowed to grow into pot seedlings. These pot seedlings were transferred to soil that was dipped into a solution containing 500 ppm of Pb(NO)₃. The results of the experiment are shown in FIG. 11.

[0098] FIG. 11 shows a photograph showing that wild-type poplar grew poorly when cultivated in soil containing lead for four weeks, whereas the YHL035C transformant poplar grew much better than the wild-type. Hence, it was verified that YCF1 transformed poplar plants and YHL035C poplar plants prepared in Example 5 have resistance to lead.

[0099] As described above, the YCF1 gene of the present invention improves resistance to and accumulation of heavy metals and other noxious materials, and the YHL035C gene improves resistance to lead, and accordingly, the transformants capable of expressing these genes can be used for the purpose of remediating environments polluted with noxious materials. Thus, the transformants of the invention provide an environmentally-friendly way to remediate the environment at a low cost.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 10

<210> SEQ ID NO 1

<211> LENGTH: 4548

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<220> FEATURE:

<221> NAME/KEY: gene

<222> LOCATION: (0)...(0)

<223> OTHER INFORMATION: *ycf1* gene

<400> SEQUENCE: 1

atggctggta atcttgtttc atgggcctgc aagctctgta gatctcctga agggtttgga	60
cctatatcct ttacagtgta ctttactcaa tgcttcatcg acggtgtgat cctaaatcta	120
tcagcaattt tcatgataac cttcggatc agagatttag ttaacctttg caagaaaaaa	180
cactctggca tcaaatatag gcggaattgg attattgtct ctaggatggc actagtctctg	240
ttggagatag cgtttgtttc acttgcgtct ttaaatattt ctaaagaaga agcggaaaac	300
tttaccattg taagtcaata tgcttctaca atgttatctt tatttgttgc tttagcetta	360
cactggatag aatacgatag atcagttgta gccaatcagg tacttttatt ctattggcctt	420
tttgaacat tcggtaatth tgctaaacta ataaatattc taattagaca cacctacgaa	480
ggcatttggt attccggaca aacgggtttc atactaacgt tattccaagt aataacatgt	540
gccagtatcc tgttacttga agctcttcca aagaagccgc taatgccaca tcaacacata	600
catcaaaact taacaagaag aaaacccaaat ccatacagata gcgcaaacat attttccagg	660
attaccttct ctgggatgac aggtttgatg aaaactggct atgaaaaata cttagtggaa	720
gcagatttat ataaattacc gaggaacttt agtagtgaag aactctctca aaaattggag	780
aaaaactggg aaaatgagtt gaagcaaaaa tcaaatcctt cattatcatg ggctatatgc	840
agaacttttg gatctaaaat gcttttagcc gcattcttta aagcaattca tgatgttcta	900
gcatttactc aaccacaact actaaggatt ttaatcaagt tcgtcacaga ctataacagt	960
gagagacagg atgaccatcc ttctcttcaa gggtttgaaa ataaccaccc acaaaaaatta	1020
cccattgtaa gaggggtttt gattgcgttt gctatgtttc tgggtgggctt tactcagaca	1080
tctgtcctgc atcaatattt cctgaatgac ttcaacacag gcattgtatat taagagcgcc	1140
ctaacggcct taatatatca aaaatcctta gtgctatcta atgaggcttc tggactttcc	1200
tctaccggtg acattgtcaa tctcatgagt gtggatgttc aaaaattaca agatttaaca	1260
caatggctaa atttaatatg gtcagggcct tttcaaatca ttatttgctt atattctctg	1320
tataagttgt tgggaaatcc catgtgggtt ggcgtgatta tactagtatt tatgatgcca	1380
ttgaactcat ttttgatgag gatacaaaag aagttgcaa aatccagat gaagtacaaa	1440
gatgaaagga cccgtgttat aagtgaata ctaacaata ttaaatcttt gaagttatat	1500
gcatgggaga agccttatag ggaaaagcta gaagaagtaa gaaataacaa agagttaaaa	1560
aatcttcaaa aactaggatg ttatatggct gtgacaagtt ttcagttcaa tatagtacca	1620
ttccttggtt catgttgtac ctttgctgta tttgtttata ctgaggatag agcattgact	1680
actgacttag ttttccctgc ttgactctg ttcaacctgc tctcattccc actaatgatt	1740
attcctatgg ttttaaatcc ttttatcgaa gcttctgttt ctattggtag attatttaca	1800

-continued

ttctttacca	atgaagagct	acaaccagat	tcggttcagc	gtttaccaa	agtaaaaaat	1860
attggcgatg	tagccattaa	cattggagat	gatgctacct	ttttatggca	acggaaaccg	1920
gaatacaaa	tagccttaaa	gaatattaat	ttccaagcta	aaaaggaaa	tttgacctgt	1980
attgttggtg	aagttggcag	tggtaaaaca	gctctattgt	catgcatggt	aggtgatcta	2040
ttcaggggta	aaggtttcgc	caccgttcat	ggttctgttg	cttatgtttc	acaagttcca	2100
tgataaatga	atgggtactgt	aaaggaaaac	attttatttg	ggcatagata	cgacgcggaa	2160
ttttacgaaa	aaacgatcaa	ggcctgtgcg	ttaactattg	atcttgcaat	tttgatggat	2220
ggagataaga	cattagtgtg	cgagaaagg	atctccttat	ctggaggaca	aaaagctcgt	2280
ttgtctttag	caagagcagt	ttatgcgaga	gctgacactt	atttacttga	tgatcctttg	2340
gcagctgttg	atgaacacgt	tgccaggcac	ttgatcgaac	atgtgttggg	tccaaatggt	2400
ttattacata	caaaaacgaa	ggtattagcc	actaataagg	tgagcgcgtt	atccatcgca	2460
gattctattg	cattattaga	taatggagaa	atcacacagc	aggttacata	tgatgagatt	2520
acgaaggacg	ctgattcgcc	attatggaaa	ttgctcaaca	actatggtaa	aaaaaataac	2580
ggtaagtcga	atgaattcgg	tgactcctct	gaaagctcag	ttcgagaaag	tagtatacct	2640
gtagaaggag	agctggaaca	actgcagaaa	ttaaatgatt	tggtttttgg	caactctgac	2700
gccataagtt	taaggaggcg	cagtgatgca	actttgggaa	gcacgatttt	tggtgacgat	2760
gaaaatattg	ctaaaagaga	gcacgtgtaa	cagggaagaa	taaagtggaa	catttaccta	2820
gagtagcgta	aagcttgcaa	cccgaaaagc	gtttgtgtat	tcatattggt	tattgttata	2880
tcgatgttcc	tctctgttat	gggtaacgtt	tggttgaaac	attggtctga	agttaatagc	2940
cgctatggat	ctaatacaaa	tgccgcgcgt	tacttgacca	tttattttgc	acttggtatt	3000
ggttcagcac	tggaacatt	aatccagaca	atcgttctct	gggttttttg	taccattcat	3060
gcctccaaat	atttacacaa	cttgatgaca	aactctgtgt	tgagagcccc	aatgacgttt	3120
tttgaacaaa	caccaatcgg	tagaattcta	aacagattct	caaagacat	atacaaatg	3180
gatgctttat	taggaagaac	attttctcag	ttttctgca	atgcagtga	agtcacattc	3240
actattacgg	ttatctgtgc	gacgacatgg	caatttatct	tcattatcat	tccactaagt	3300
gtgttttaca	tctactacca	gcagtattac	ctgagaacat	caaggaggtt	gcgtcgttta	3360
gactctatta	ctaggctctc	aatctactct	catttccaag	agactttggg	tggccttgca	3420
acggttagag	gttattctca	acagaaaagg	ttttcccaca	ttaatcaatg	ccgcattgat	3480
aataacatga	gtgcgttcta	tccctctatc	aatgctaacc	gttggttagc	atataggttg	3540
gaacttattg	gttcaattat	cattctaggt	gctgcaactt	tatccgtttt	tagactaaaa	3600
caaggcacat	taacggcagg	tatggtgggt	ttatcattaa	gctatgcttt	acaaatcact	3660
caaacgttaa	attggattgt	tagaatgact	gtggaagttg	aaacgaatat	tgtttcagtg	3720
gaaagaataa	aggaatatgc	tgatttgaag	agcgaggcac	ctttaatagt	tgaaggccac	3780
agaccaccca	aagaatggcc	gagccagggt	gatataaagt	ttaataatta	ttccactcgt	3840
tataggccgg	agcttgatct	tgttctgaag	cacattaata	tacacattaa	accaaagtga	3900
aaagttggta	tcgtgggtag	aacgggtgcg	ggaaaatcct	cattaacgct	agcattattc	3960
aggatgattg	aggctagcga	gggaaacatc	gtaatcgaca	acattgccat	caacgagatt	4020
gggttatatg	atttgagaca	taaattgtca	atcatacctc	aggattctca	agtttttgag	4080

-continued

```

ggcactgttc gtgagaacat tgatccatt aaccaatata ctgatgaagc tatttggagg 4140
gcattggaac tttctcattt gaaagaacac gtgctatcaa tgagcaatga cggattagat 4200
gcccactaa ccgaagggtg tggcaactta agtgttgac aaagacaatt attatgtctt 4260
gcaagagcaa tgttggttcc atcaaagatt ttggtgcttg atgaagccac ggccgcagtc 4320
gacgtggaga cagataaagt cgtccaagag acgattcgta ctgctttcaa ggacagaact 4380
atcttgacca tcgcgcatag actgaacacg ataatggaca gtgatagaat catagtgttg 4440
gacaatggta aagtagccga gtttgactct ccgggccagt tattaagtga taacaaatca 4500
ttgttctatt cactgtgcat ggaggctggt ttggtcaatg aaaattaa 4548

```

<210> SEQ ID NO 2

<211> LENGTH: 1515

<212> TYPE: PRT

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 2

```

Met Ala Gly Asn Leu Val Ser Trp Ala Cys Lys Leu Cys Arg Ser Pro
  1           5           10          15
Glu Gly Phe Gly Pro Ile Ser Phe Tyr Gly Asp Phe Thr Gln Cys Phe
          20          25          30
Ile Asp Gly Val Ile Leu Asn Leu Ser Ala Ile Phe Met Ile Thr Phe
          35          40          45
Gly Ile Arg Asp Leu Val Asn Leu Cys Lys Lys Lys His Ser Gly Ile
          50          55          60
Lys Tyr Arg Arg Asn Trp Ile Ile Val Ser Arg Met Ala Leu Val Leu
          65          70          75          80
Leu Glu Ile Ala Phe Val Ser Leu Ala Ser Leu Asn Ile Ser Lys Glu
          85          90          95
Glu Ala Glu Asn Phe Thr Ile Val Ser Gln Tyr Ala Ser Thr Met Leu
          100         105         110
Ser Leu Phe Val Ala Leu Ala Leu His Trp Ile Glu Tyr Asp Arg Ser
          115         120         125
Val Val Ala Asn Thr Val Leu Leu Phe Tyr Trp Leu Phe Glu Thr Phe
          130         135         140
Gly Asn Phe Ala Lys Leu Ile Asn Ile Leu Ile Arg His Thr Tyr Glu
          145         150         155         160
Gly Ile Trp Tyr Ser Gly Gln Thr Gly Phe Ile Leu Thr Leu Phe Gln
          165         170         175
Val Ile Thr Cys Ala Ser Ile Leu Leu Leu Glu Ala Leu Pro Lys Lys
          180         185         190
Pro Leu Met Pro His Gln His Ile His Gln Thr Leu Thr Arg Arg Lys
          195         200         205
Pro Asn Pro Tyr Asp Ser Ala Asn Ile Phe Ser Arg Ile Thr Phe Ser
          210         215         220
Trp Met Ser Gly Leu Met Lys Thr Gly Tyr Glu Lys Tyr Leu Val Glu
          225         230         235         240
Ala Asp Leu Tyr Lys Leu Pro Arg Asn Phe Ser Ser Glu Glu Leu Ser
          245         250         255
Gln Lys Leu Glu Lys Asn Trp Glu Asn Glu Leu Lys Gln Lys Ser Asn
          260         265         270
Pro Ser Leu Ser Trp Ala Ile Cys Arg Thr Phe Gly Ser Lys Met Leu

```

-continued

275	280	285
Leu Ala Ala Phe Phe Lys Ala Ile His Asp Val Leu Ala Phe Thr Gln 290 295 300		
Pro Gln Leu Leu Arg Ile Leu Ile Lys Phe Val Thr Asp Tyr Asn Ser 305 310 315 320		
Glu Arg Gln Asp Asp His Ser Ser Leu Gln Gly Phe Glu Asn Asn His 325 330 335		
Pro Gln Lys Leu Pro Ile Val Arg Gly Phe Leu Ile Ala Phe Ala Met 340 345 350		
Phe Leu Val Gly Phe Thr Gln Thr Ser Val Leu His Gln Tyr Phe Leu 355 360 365		
Asn Val Phe Asn Thr Gly Met Tyr Ile Lys Ser Ala Leu Thr Ala Leu 370 375 380		
Ile Tyr Gln Lys Ser Leu Val Leu Ser Asn Glu Ala Ser Gly Leu Ser 385 390 395 400		
Ser Thr Gly Asp Ile Val Asn Leu Met Ser Val Asp Val Gln Lys Leu 405 410 415		
Gln Asp Leu Thr Gln Trp Leu Asn Leu Ile Trp Ser Gly Pro Phe Gln 420 425 430		
Ile Ile Ile Cys Leu Tyr Ser Leu Tyr Lys Leu Leu Gly Asn Ser Met 435 440 445		
Trp Val Gly Val Ile Ile Leu Val Ile Met Met Pro Leu Asn Ser Phe 450 455 460		
Leu Met Arg Ile Gln Lys Lys Leu Gln Lys Ser Gln Met Lys Tyr Lys 465 470 475 480		
Asp Glu Arg Thr Arg Val Ile Ser Glu Ile Leu Asn Asn Ile Lys Ser 485 490 495		
Leu Lys Leu Tyr Ala Trp Glu Lys Pro Tyr Arg Glu Lys Leu Glu Glu 500 505 510		
Val Arg Asn Asn Lys Glu Leu Lys Asn Leu Thr Lys Leu Gly Cys Tyr 515 520 525		
Met Ala Val Thr Ser Phe Gln Phe Asn Ile Val Pro Phe Leu Val Ser 530 535 540		
Cys Cys Thr Phe Ala Val Phe Val Tyr Thr Glu Asp Arg Ala Leu Thr 545 550 555 560		
Thr Asp Leu Val Phe Pro Ala Leu Thr Leu Phe Asn Leu Leu Ser Phe 565 570 575		
Pro Leu Met Ile Ile Pro Met Val Leu Asn Ser Phe Ile Glu Ala Ser 580 585 590		
Val Ser Ile Gly Arg Leu Phe Thr Phe Phe Thr Asn Glu Glu Leu Gln 595 600 605		
Pro Asp Ser Val Gln Arg Leu Pro Lys Val Lys Asn Ile Gly Asp Val 610 615 620		
Ala Ile Asn Ile Gly Asp Asp Ala Thr Phe Leu Trp Gln Arg Lys Pro 625 630 635 640		
Glu Tyr Lys Val Ala Leu Lys Asn Ile Asn Phe Gln Ala Lys Lys Gly 645 650 655		
Asn Leu Thr Cys Ile Val Gly Lys Val Gly Ser Gly Lys Thr Ala Leu 660 665 670		
Leu Ser Cys Met Leu Gly Asp Leu Phe Arg Val Lys Gly Phe Ala Thr 675 680 685		

-continued

Val	His	Gly	Ser	Val	Ala	Tyr	Val	Ser	Gln	Val	Pro	Trp	Ile	Met	Asn
690						695					700				
Gly	Thr	Val	Lys	Glu	Asn	Ile	Leu	Phe	Gly	His	Arg	Tyr	Asp	Ala	Glu
705					710					715					720
Phe	Tyr	Glu	Lys	Thr	Ile	Lys	Ala	Cys	Ala	Leu	Thr	Ile	Asp	Leu	Ala
				725					730					735	
Ile	Leu	Met	Asp	Gly	Asp	Lys	Thr	Leu	Val	Gly	Glu	Lys	Gly	Ile	Ser
			740					745					750		
Leu	Ser	Gly	Gly	Gln	Lys	Ala	Arg	Leu	Ser	Leu	Ala	Arg	Ala	Val	Tyr
	755						760					765			
Ala	Arg	Ala	Asp	Thr	Tyr	Leu	Leu	Asp	Asp	Pro	Leu	Ala	Ala	Val	Asp
770						775					780				
Glu	His	Val	Ala	Arg	His	Leu	Ile	Glu	His	Val	Leu	Gly	Pro	Asn	Gly
785					790					795					800
Leu	Leu	His	Thr	Lys	Thr	Lys	Val	Leu	Ala	Thr	Asn	Lys	Val	Ser	Ala
				805					810					815	
Leu	Ser	Ile	Ala	Asp	Ser	Ile	Ala	Leu	Leu	Asp	Asn	Gly	Glu	Ile	Thr
			820					825					830		
Gln	Gln	Gly	Thr	Tyr	Asp	Glu	Ile	Thr	Lys	Asp	Ala	Asp	Ser	Pro	Leu
		835					840					845			
Trp	Lys	Leu	Leu	Asn	Asn	Tyr	Gly	Lys	Lys	Asn	Asn	Gly	Lys	Ser	Asn
850					855						860				
Glu	Phe	Gly	Asp	Ser	Ser	Glu	Ser	Ser	Val	Arg	Glu	Ser	Ser	Ile	Pro
865					870					875					880
Val	Glu	Gly	Glu	Leu	Glu	Gln	Leu	Gln	Lys	Leu	Asn	Asp	Leu	Asp	Phe
				885					890					895	
Gly	Asn	Ser	Asp	Ala	Ile	Ser	Leu	Arg	Arg	Ala	Ser	Asp	Ala	Thr	Leu
			900					905					910		
Gly	Ser	Ile	Asp	Phe	Gly	Asp	Asp	Glu	Asn	Ile	Ala	Lys	Arg	Glu	His
		915					920					925			
Arg	Glu	Gln	Gly	Lys	Val	Lys	Trp	Asn	Ile	Tyr	Leu	Glu	Tyr	Ala	Lys
930						935					940				
Ala	Cys	Asn	Pro	Lys	Ser	Val	Cys	Val	Phe	Ile	Leu	Phe	Ile	Val	Ile
945					950					955					960
Ser	Met	Phe	Leu	Ser	Val	Met	Gly	Asn	Val	Trp	Leu	Lys	His	Trp	Ser
				965					970					975	
Glu	Val	Asn	Ser	Arg	Tyr	Gly	Ser	Asn	Pro	Asn	Ala	Ala	Arg	Tyr	Leu
			980					985					990		
Ala	Ile	Tyr	Phe	Ala	Leu	Gly	Ile	Gly	Ser	Ala	Leu	Ala	Thr	Leu	Ile
		995					1000					1005			
Gln	Thr	Ile	Val	Leu	Trp	Val	Phe	Cys	Thr	Ile	His	Ala	Ser	Lys	Tyr
1010						1015					1020				
Leu	His	Asn	Leu	Met	Thr	Asn	Ser	Val	Leu	Arg	Ala	Pro	Met	Thr	Phe
1025					1030					1035					1040
Phe	Glu	Thr	Thr	Pro	Ile	Gly	Arg	Ile	Leu	Asn	Arg	Phe	Ser	Asn	Asp
				1045					1050					1055	
Ile	Tyr	Lys	Val	Asp	Ala	Leu	Leu	Gly	Arg	Thr	Phe	Ser	Gln	Phe	Phe
		1060						1065					1070		
Val	Asn	Ala	Val	Lys	Val	Thr	Phe	Thr	Ile	Thr	Val	Ile	Cys	Ala	Thr
		1075					1080						1085		

-continued

Thr	Trp	Gln	Phe	Ile	Phe	Ile	Ile	Ile	Pro	Leu	Ser	Val	Phe	Tyr	Ile
1090						1095					1100				
Tyr	Tyr	Gln	Gln	Tyr	Tyr	Leu	Arg	Thr	Ser	Arg	Glu	Leu	Arg	Arg	Leu
1105					1110					1115					1120
Asp	Ser	Ile	Thr	Arg	Ser	Pro	Ile	Tyr	Ser	His	Phe	Gln	Glu	Thr	Leu
				1125					1130					1135	
Gly	Gly	Leu	Ala	Thr	Val	Arg	Gly	Tyr	Ser	Gln	Gln	Lys	Arg	Phe	Ser
			1140					1145					1150		
His	Ile	Asn	Gln	Cys	Arg	Ile	Asp	Asn	Asn	Met	Ser	Ala	Phe	Tyr	Pro
		1155					1160					1165			
Ser	Ile	Asn	Ala	Asn	Arg	Trp	Leu	Ala	Tyr	Arg	Leu	Glu	Leu	Ile	Gly
	1170					1175					1180				
Ser	Ile	Ile	Ile	Leu	Gly	Ala	Ala	Thr	Leu	Ser	Val	Phe	Arg	Leu	Lys
1185				1190						1195					1200
Gln	Gly	Thr	Leu	Thr	Ala	Gly	Met	Val	Gly	Leu	Ser	Leu	Ser	Tyr	Ala
			1205						1210					1215	
Leu	Gln	Ile	Thr	Gln	Thr	Leu	Asn	Trp	Ile	Val	Arg	Met	Thr	Val	Glu
			1220					1225					1230		
Val	Glu	Thr	Asn	Ile	Val	Ser	Val	Glu	Arg	Ile	Lys	Glu	Tyr	Ala	Asp
	1235						1240					1245			
Leu	Lys	Ser	Glu	Ala	Pro	Leu	Ile	Val	Glu	Gly	His	Arg	Pro	Pro	Lys
	1250					1255					1260				
Glu	Trp	Pro	Ser	Gln	Gly	Asp	Ile	Lys	Phe	Asn	Asn	Tyr	Ser	Thr	Arg
1265				1270						1275					1280
Tyr	Arg	Pro	Glu	Leu	Asp	Leu	Val	Leu	Lys	His	Ile	Asn	Ile	His	Ile
				1285					1290					1295	
Lys	Pro	Asn	Glu	Lys	Val	Gly	Ile	Val	Gly	Arg	Thr	Gly	Ala	Gly	Lys
		1300					1305						1310		
Ser	Ser	Leu	Thr	Leu	Ala	Leu	Phe	Arg	Met	Ile	Glu	Ala	Ser	Glu	Gly
		1315					1320					1325			
Asn	Ile	Val	Ile	Asp	Asn	Ile	Ala	Ile	Asn	Glu	Ile	Gly	Leu	Tyr	Asp
	1330					1335						1340			
Leu	Arg	His	Lys	Leu	Ser	Ile	Ile	Pro	Gln	Asp	Ser	Gln	Val	Phe	Glu
1345				1350						1355					1360
Gly	Thr	Val	Arg	Glu	Asn	Ile	Asp	Pro	Ile	Asn	Gln	Tyr	Thr	Asp	Glu
			1365					1370						1375	
Ala	Ile	Trp	Arg	Ala	Leu	Glu	Leu	Ser	His	Leu	Lys	Glu	His	Val	Leu
		1380						1385					1390		
Ser	Met	Ser	Asn	Asp	Gly	Leu	Asp	Ala	Gln	Leu	Thr	Glu	Gly	Gly	Gly
	1395					1400						1405			
Asn	Leu	Ser	Val	Gly	Gln	Arg	Gln	Leu	Leu	Cys	Leu	Ala	Arg	Ala	Met
	1410					1415					1420				
Leu	Val	Pro	Ser	Lys	Ile	Leu	Val	Leu	Asp	Glu	Ala	Thr	Ala	Ala	Val
1425				1430						1435					1440
Asp	Val	Glu	Thr	Asp	Lys	Val	Val	Gln	Glu	Thr	Ile	Arg	Thr	Ala	Phe
			1445						1450					1455	
Lys	Asp	Arg	Thr	Ile	Leu	Thr	Ile	Ala	His	Arg	Leu	Asn	Thr	Ile	Met
	1460							1465					1470		
Asp	Ser	Asp	Arg	Ile	Ile	Val	Leu	Asp	Asn	Gly	Lys	Val	Ala	Glu	Phe
	1475						1480					1485			
Asp	Ser	Pro	Gly	Gln	Leu	Leu	Ser	Asp	Asn	Lys	Ser	Leu	Phe	Tyr	Ser

-continued

1490	1495	1500	
Leu Cys Met Glu Ala Gly Leu Val Asn Glu Asn			
1505	1510	1515	
 <210> SEQ ID NO 3			
<211> LENGTH: 4776			
<212> TYPE: DNA			
<213> ORGANISM: Saccharomyces cerevisiae			
<220> FEATURE:			
<221> NAME/KEY: gene			
<222> LOCATION: (0)...(0)			
<223> OTHER INFORMATION: yhl035c gene			
 <400> SEQUENCE: 3			
atgggaacgg atcccccttat tatccgaaat aatgggttcat tttgggaagt tgatgatttt		60	
actcgtttag gaagaactca gctattgagc tactattttac cattggctat catagcctca		120	
attggcattt tcgcactttg tcgcagtgga ttatctcgtt atgtaagatc tgccgagtgc		180	
gatttagtga acgaatatct atttggcgca caagaagaga gaaaagaaga taatagtata		240	
gaaagacttc tacggaactc aaatacccaa gcccaattacg tcaacgtcaa aaagcaagga		300	
aggattttga aacttagaca ttttgatata acaactatag atgtcaagca aatcgatgct		360	
aaaaatcatg gtggactaac gtttagtaga ccgtctacta gtgaccactt aagaaatca		420	
tctgaaattg tattaatgtc tttacaaata attggccttt cctttttaag agtaacaaaa		480	
atcaatattg aattaacgaa cagagatggt acaactttac tattattttg gttaatacta		540	
ctttccctaa gtatcttaag agtttacaag cgttcaacga atctttgggc catctgtttt		600	
actgccata caactatttg gatttcaacc tggattccaa ttcgttcggt ctatattggt		660	
aatatcgatg atgtaccctc acagatattt tacatctttg aattcgtaat tacttcaacc		720	
ttacagccaa taaagctcac ttcaccgatt aaagacaact catccatcat ctacgttaga		780	
gacgaccata cgtctccttc gagggaaacac atatcctcaa ttttaagttg cattacttg		840	
agctggatta ccaattttat atgggaggcc caaaagaaca ctattaagtt aaaggatatt		900	
tggggcttat caatggaaga ctatagcatt ttcattctaa aagggtttac caggagaaac		960	
aagcacatta ataatttgac gctagcactg tttgaatctt tcaaaacata tttactcata		1020	
ggaatgttat ggggttcgtt gaacagtatt gtaaaccttc tccccacaat tttaatgaaa		1080	
agatttttag aaattgtgga taacccaaac cgttcctcat catgcatgaa tttggcgttg		1140	
ctttatatta ttggtatggt catttgtaga ttgacattag caatttgtaa ttcccaaggt		1200	
caatttggtt ctgataagat ttgtttaaga ataagagcca tactcatagg agaaatttat		1260	
gcaaaaggct tacgtaggag gctgtttaca tctccaaaaa ccagctctga ttcagatagt		1320	
atctccgcaa acotttggtac cataattaat ctcatctcta ttgactcatt taaggatcgt		1380	
gaactagcaa actaccctta tgtgacagtt caggcagtaa ttatgataat agttgttgta		1440	
ggactacttt tcaacttttt aggtgtttca gcttttgtag gaatttcaat tatcttagtg		1500	
atgttcccat tgaatttctt gtttagcgaat ttgttaggta agtttcaaaa gcaaacactg		1560	
aaatgtactg accaaagaat ctcaaaattg aacgagtgtt tacagaacat aagaattgtc		1620	
aaatatattg cttgggaaaag gaatattata aatgaaatca aatcaataag gcaaaaggaa		1680	
ttaagatcct tattaataaa atcttttggtg tggtcgtaa cttcttttct ttggttcgtg		1740	
acaccgacct tgggtgacag tgtcactttc gccatctgta catttgttca acatgaagat		1800	

-continued

ttgaatgccc	cgcttgcttt	cactactttg	tcactcttca	ctttgttaaa	gacaccctg	1860
gatcaattat	caaatatgct	aagtttcata	aatcaatcaa	aagtctctct	aaaaagaata	1920
agcgattttt	taaggatgga	cgatacagaa	aaatataatc	aactaacat	atctccagac	1980
aaaaataaaa	ttgaatttaa	aaatgcgact	ttaacctgga	atgaaaatga	cagcgatatg	2040
aatgcattca	aattatgttg	tttgaatatt	aaatttcaaa	ttggttaagt	aaatttgatt	2100
ttgggttcta	caggatctgg	taaaagtgca	ttgctgctgg	gtttactggg	tgaactaaat	2160
ctaattagtg	gctctatcat	tgttccgagc	ttagaaccaa	agcatgattt	aattcccgac	2220
tgcaaggtt	taaccaattc	cttcgcata	tgttcacaaa	gtgcgtggct	attaaatgac	2280
acggtaaaaa	acaatattat	ctttgataac	ttctataacg	aggataggta	caacaaagta	2340
attgatgcat	gtgggctgaa	aagagacctg	gagattttac	cagcagggtg	cctaacagaa	2400
attggtgaaa	aggggtataac	tttatcagga	gggcagaaac	agagaatttc	cttggcgaga	2460
gctgtttatt	cgagtgtcaa	gcatgtctta	ctagatgatt	gtttgagcgc	tgtcgattca	2520
catactgctg	tatggatcta	tgaaaattgc	atcacaggtc	cactaatgaa	aaatagaacc	2580
tgcattttag	ttacgcacaa	tgtttcatta	acacttagaa	atgccatttt	cgcgattgtg	2640
ttggaaaatg	gcaaagtga	gaatcaagga	actattacag	aattacaaag	caaagggctt	2700
tttaaggaaa	aatatgttca	actttcttct	cgagatagca	ttaatgaaa	gaacgcta	2760
agattaaaag	ctcccagaaa	aaatgactct	cagaaaatcg	aacctgtcac	cgagaacata	2820
aattttgatg	caaattttgt	caatgatggc	cagctaatag	aagaggaaga	aaaatcaaac	2880
ggtgccataa	gccccgatgt	ttataaatgg	tacctgaaat	tttttgagg	cttcaaagct	2940
ttaacagccc	tgctcgtctt	ttatatcaca	gctcaaattt	tgctcatcag	tcagtcttgg	3000
tggtacgac	attgggtcaa	cgataccaat	gtacgaataa	atgctccagg	ttttgcgatg	3060
gacacgctgc	cattaaaagg	gatgaccgac	tcttcgaaaa	ataaacataa	tgcattttat	3120
tacttaaccg	tatattttct	tattgggtatc	attcaggcaa	tgctaggtgg	ttttaaaca	3180
atgatgacgt	ttttatccgg	tatgcgagcc	tccagaaaga	tctttaataa	tctgctagat	3240
ctagttctac	atgcccaaat	acgatttttc	gacgtgacgc	cggttggtag	aatcatgaat	3300
cgcttttcaa	aggacatcga	aggtgttgat	caagaattga	ttccatactt	agaagtaact	3360
atattttgcc	taattcaatg	cgcatacaatt	atattttctca	ttaccgtaat	aactcctcgc	3420
tttttgacag	tcgccgttat	cgtttttgtt	ttatatctct	ttgtggggaa	atggtagtta	3480
acggcaagta	gagaattgaa	aaggttagat	tcaataacca	aatcacccat	ttttcaacat	3540
ttctcagaga	ccttggttag	cgtttgcaca	attcgtgcat	ttggcgacga	gaggagattc	3600
attttagaaa	atatgaacaa	aattgaccaa	aataacagag	cattctttta	tttatcagtt	3660
actgtcaaat	ggttttcttt	tagagtgcac	atgattggcg	cattcattgt	tttagcatca	3720
ggttctttta	ttctgctcaa	tattgcaaat	attgactcgg	gtcttgccgg	catttctttg	3780
acatatgcc	ttttgtttac	agatgggtgt	ttatgggttag	ttagactgta	ctcaacattt	3840
gaaatgaaca	tgaactctgt	tgaagacta	aaagaatatt	ctagcattga	acaagagAAC	3900
tatcttggcc	atgatgaagg	ccgcattcta	cttctaaacg	aaccatcgtg	gccaaaagat	3960
ggagaaattg	aaattgaaaa	cttatcttta	cgttacgcgc	caaatttgcc	tcctgtcata	4020
agaaatgtta	gtttcaaagt	ggatcctcaa	agtaagattg	ggattgtcgg	aagaactggc	4080

-continued

```

gcaggcaaat ctaccataat aacggcatta ttcagattac tagaaccaat aaccggatgt 4140
atcaaaaatag atgggcagga tataagtaaa attgatctcg ttacattacg tcgttccatt 4200
actatcatcc ctcaggaccc tattctatct gcagggtacaa tcaaaagtaa tgttgatcca 4260
tatgatgaat atgatgaaaa aaaaatattc aaagcacttt cacaagtaaa tctaatttct 4320
tcacatgaat ttgaagaagt gcttaactcg gaggaacgct ttaacagcac tcataataaa 4380
tttttaaadc ttcacacaga aatagctgag ggcggcttaa atctgtccca aggtgaaagg 4440
caattgcttt ttattgcacg atcattgtta cgcgagccaa agataatact tttggacgag 4500
gctacttctt ctattgatta cgattctgac catttaattc aggggtattat aagaagtgag 4560
tttaataaaa gcacaattct tactattgca catcgtttga gatctgttat cgattacgac 4620
aggataattg tgatggatgc cgggtaggtg aaagaatatg atcgccctag tgaactgttg 4680
aaagatgaac gcggtatatt ttatagtatg tgcgtgaca gtgggggcct agagcttttg 4740
aagcaaatag ccaagcaatc aagtaagatg atgaaa 4776

```

<210> SEQ ID NO 4

<211> LENGTH: 1592

<212> TYPE: PRT

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 4

```

Met Gly Thr Asp Pro Leu Ile Ile Arg Asn Asn Gly Ser Phe Trp Glu
  1             5             10             15
Val Asp Asp Phe Thr Arg Leu Gly Arg Thr Gln Leu Leu Ser Tyr Tyr
          20             25             30
Leu Pro Leu Ala Ile Ile Ala Ser Ile Gly Ile Phe Ala Leu Cys Arg
          35             40             45
Ser Gly Leu Ser Arg Tyr Val Arg Ser Ala Glu Cys Asp Leu Val Asn
          50             55             60
Glu Tyr Leu Phe Gly Ala Gln Glu Glu Arg Lys Glu Asp Asn Ser Ile
          65             70             75             80
Glu Arg Leu Leu Arg Asn Ser Asn Thr Gln Ala Asn Tyr Val Asn Val
          85             90             95
Lys Lys Gln Gly Arg Ile Leu Lys Leu Arg His Phe Asp Ile Thr Thr
          100            105            110
Ile Asp Val Lys Gln Ile Asp Ala Lys Asn His Gly Gly Leu Thr Phe
          115            120            125
Ser Arg Pro Ser Thr Ser Asp His Leu Arg Lys Ser Ser Glu Ile Val
          130            135            140
Leu Met Ser Leu Gln Ile Ile Gly Leu Ser Phe Leu Arg Val Thr Lys
          145            150            155            160
Ile Asn Ile Glu Leu Thr Asn Arg Asp Val Thr Thr Leu Leu Leu Phe
          165            170            175
Trp Leu Ile Leu Leu Ser Leu Ser Ile Leu Arg Val Tyr Lys Arg Ser
          180            185            190
Thr Asn Leu Trp Ala Ile Cys Phe Thr Ala His Thr Thr Ile Trp Ile
          195            200            205
Ser Thr Trp Ile Pro Ile Arg Ser Val Tyr Ile Gly Asn Ile Asp Asp
          210            215            220
Val Pro Ser Gln Ile Phe Tyr Ile Phe Glu Phe Val Ile Thr Ser Thr

```

-continued

225	230	235	240
Leu Gln Pro Ile Lys 245	Leu Thr Ser Pro Ile Lys 250	Asp Asn Ser Ser Ile 255	
Ile Tyr Val Arg Asp 260	Asp His Thr Ser Pro Ser Arg 265	Glu His Ile Ser 270	
Ser Ile Leu Ser Cys Ile Thr 275	Trp Ser Trp Ile Thr Asn 280	Phe Ile Trp 285	
Glu Ala Gln Lys Asn Thr 290	Ile Lys Leu Lys Asp 295	Ile Trp Gly Leu Ser 300	
Met Glu Asp Tyr Ser 305	Ile Phe Ile Leu Lys Gly 310	Phe Thr Arg Arg Asn 315	320
Lys His Ile Asn Asn 325	Leu Thr Leu Ala Leu Phe 330	Glu Ser Phe Lys Thr 335	
Tyr Leu Leu Ile Gly 340	Met Leu Trp Val Leu Val 345	Asn Ser Ile Val Asn 350	
Leu Leu Pro Thr Ile 355	Leu Met Lys Arg Phe Leu 360	Glu Ile Val Asp Asn 365	
Pro Asn Arg Ser Ser 370	Ser Cys Met Asn Leu Ala 375	Trp Leu Tyr Ile Ile 380	
Gly Met Phe Ile Cys 385	Arg Leu Thr Leu Ala 390	Ile Cys Asn Ser Gln Gly 395	400
Gln Phe Val Ser Asp 405	Lys Ile Cys Leu Arg 410	Ile Arg Ala Ile Leu Ile 415	
Gly Glu Ile Tyr Ala 420	Lys Gly Leu Arg Arg Arg 425	Leu Phe Thr Ser Pro 430	
Lys Thr Ser Ser Asp 435	Ser Asp Ser Ile Ser Ala 440	Asn Leu Gly Thr Ile 445	
Ile Asn Leu Ile Ser 450	Ile Asp Ser Phe Lys Val 455	Ser Glu Leu Ala Asn 460	
Tyr Leu Tyr Val Thr 465	Val Gln Ala Val Ile 470	Met Ile Ile Val Val Val 475	480
Gly Leu Leu Phe Asn 485	Phe Leu Gly Val Ser Ala 490	Phe Ala Gly Ile Ser 495	
Ile Ile Leu Val Met 500	Phe Pro Leu Asn Phe 505	Leu Leu Ala Asn Leu Leu 510	
Gly Lys Phe Gln Lys 515	Gln Thr Leu Lys Cys Thr 520	Asp Gln Arg Ile Ser 525	
Lys Leu Asn Glu Cys 530	Leu Gln Asn Ile Arg 535	Ile Val Lys Tyr Phe Ala 540	
Trp Glu Arg Asn Ile 545	Ile Asn Glu Ile Lys 550	Ser Ile Arg Gln Lys Glu 555	560
Leu Arg Ser Leu Leu 565	Lys Lys Ser Leu Val 570	Trp Ser Val Thr Ser Phe 575	
Leu Trp Phe Val Thr 580	Pro Thr Leu Val Thr 585	Gly Val Thr Phe Ala Ile 590	
Cys Thr Phe Val Gln 595	His Glu Asp Leu Asn 600	Ala Pro Leu Ala Phe Thr 605	
Thr Leu Ser Leu Phe 610	Thr Leu Leu Lys Thr 615	Pro Leu Asp Gln Leu Ser 620	
Asn Met Leu Ser Phe 625	Ile Asn Gln Ser Lys 630	Val Ser Leu Lys Arg Ile 635	640

-continued

Ser	Asp	Phe	Leu	Arg	Met	Asp	Asp	Thr	Glu	Lys	Tyr	Asn	Gln	Leu	Thr	645	650	655
Ile	Ser	Pro	Asp	Lys	Asn	Lys	Ile	Glu	Phe	Lys	Asn	Ala	Thr	Leu	Thr	660	665	670
Trp	Asn	Glu	Asn	Asp	Ser	Asp	Met	Asn	Ala	Phe	Lys	Leu	Cys	Gly	Leu	675	680	685
Asn	Ile	Lys	Phe	Gln	Ile	Gly	Lys	Leu	Asn	Leu	Ile	Leu	Gly	Ser	Thr	690	695	700
Gly	Ser	Gly	Lys	Ser	Ala	Leu	Leu	Leu	Gly	Leu	Leu	Gly	Glu	Leu	Asn	705	710	715
Leu	Ile	Ser	Gly	Ser	Ile	Ile	Val	Pro	Ser	Leu	Glu	Pro	Lys	His	Asp	725	730	735
Leu	Ile	Pro	Asp	Cys	Glu	Gly	Leu	Thr	Asn	Ser	Phe	Ala	Tyr	Cys	Ser	740	745	750
Gln	Ser	Ala	Trp	Leu	Leu	Asn	Asp	Thr	Val	Lys	Asn	Asn	Ile	Ile	Phe	755	760	765
Asp	Asn	Phe	Tyr	Asn	Glu	Asp	Arg	Tyr	Asn	Lys	Val	Ile	Asp	Ala	Cys	770	775	780
Gly	Leu	Lys	Arg	Asp	Leu	Glu	Ile	Leu	Pro	Ala	Gly	Asp	Leu	Thr	Glu	785	790	795
Ile	Gly	Glu	Lys	Gly	Ile	Thr	Leu	Ser	Gly	Gly	Gln	Lys	Gln	Arg	Ile	805	810	815
Ser	Leu	Ala	Arg	Ala	Val	Tyr	Ser	Ser	Ala	Lys	His	Val	Leu	Leu	Asp	820	825	830
Asp	Cys	Leu	Ser	Ala	Val	Asp	Ser	His	Thr	Ala	Val	Trp	Ile	Tyr	Glu	835	840	845
Asn	Cys	Ile	Thr	Gly	Pro	Leu	Met	Lys	Asn	Arg	Thr	Cys	Ile	Leu	Val	850	855	860
Thr	His	Asn	Val	Ser	Leu	Thr	Leu	Arg	Asn	Ala	His	Phe	Ala	Ile	Val	865	870	875
Leu	Glu	Asn	Gly	Lys	Val	Lys	Asn	Gln	Gly	Thr	Ile	Thr	Glu	Leu	Gln	885	890	895
Ser	Lys	Gly	Leu	Phe	Lys	Glu	Lys	Tyr	Val	Gln	Leu	Ser	Ser	Arg	Asp	900	905	910
Ser	Ile	Asn	Glu	Lys	Asn	Ala	Asn	Arg	Leu	Lys	Ala	Pro	Arg	Lys	Asn	915	920	925
Asp	Ser	Gln	Lys	Ile	Glu	Pro	Val	Thr	Glu	Asn	Ile	Asn	Phe	Asp	Ala	930	935	940
Asn	Phe	Val	Asn	Asp	Gly	Gln	Leu	Ile	Glu	Glu	Glu	Glu	Lys	Ser	Asn	945	950	955
Gly	Ala	Ile	Ser	Pro	Asp	Val	Tyr	Lys	Trp	Tyr	Leu	Lys	Phe	Phe	Gly	965	970	975
Gly	Phe	Lys	Ala	Leu	Thr	Ala	Leu	Phe	Ala	Leu	Tyr	Ile	Thr	Ala	Gln	980	985	990
Ile	Leu	Phe	Ile	Ser	Gln	Ser	Trp	Trp	Ile	Arg	His	Trp	Val	Asn	Asp	995	1000	1005
Thr	Asn	Val	Arg	Ile	Asn	Ala	Pro	Gly	Phe	Ala	Met	Asp	Thr	Leu	Pro	1010	1015	1020
Leu	Lys	Gly	Met	Thr	Asp	Ser	Ser	Lys	Asn	Lys	His	Asn	Ala	Phe	Tyr	1025	1030	1035

-continued

Tyr	Leu	Thr	Val	Tyr	Phe	Leu	Ile	Gly	Ile	Ile	Gln	Ala	Met	Leu	Gly	
				1045					1050					1055		
Gly	Phe	Lys	Thr	Met	Met	Thr	Phe	Leu	Ser	Gly	Met	Arg	Ala	Ser	Arg	
			1060					1065					1070			
Lys	Ile	Phe	Asn	Asn	Leu	Leu	Asp	Leu	Val	Leu	His	Ala	Gln	Ile	Arg	
		1075					1080					1085				
Phe	Phe	Asp	Val	Thr	Pro	Val	Gly	Arg	Ile	Met	Asn	Arg	Phe	Ser	Lys	
	1090					1095					1100					
Asp	Ile	Glu	Gly	Val	Asp	Gln	Glu	Leu	Ile	Pro	Tyr	Leu	Glu	Val	Thr	
1105					1110					1115				1120		
Ile	Phe	Cys	Leu	Ile	Gln	Cys	Ala	Ser	Ile	Ile	Phe	Leu	Ile	Thr	Val	
			1125						1130					1135		
Ile	Thr	Pro	Arg	Phe	Leu	Thr	Val	Ala	Val	Ile	Val	Phe	Val	Leu	Tyr	
			1140					1145					1150			
Phe	Phe	Val	Gly	Lys	Trp	Tyr	Leu	Thr	Ala	Ser	Arg	Glu	Leu	Lys	Arg	
	1155						1160					1165				
Leu	Asp	Ser	Ile	Thr	Lys	Ser	Pro	Ile	Phe	Gln	His	Phe	Ser	Glu	Thr	
	1170					1175					1180					
Leu	Val	Gly	Val	Cys	Thr	Ile	Arg	Ala	Phe	Gly	Asp	Glu	Arg	Arg	Phe	
1185					1190					1195					1200	
Ile	Leu	Glu	Asn	Met	Asn	Lys	Ile	Asp	Gln	Asn	Asn	Arg	Ala	Phe	Phe	
			1205						1210					1215		
Tyr	Leu	Ser	Val	Thr	Val	Lys	Trp	Phe	Ser	Phe	Arg	Val	Asp	Met	Ile	
			1220					1225					1230			
Gly	Ala	Phe	Ile	Val	Leu	Ala	Ser	Gly	Ser	Phe	Ile	Leu	Leu	Asn	Ile	
		1235					1240					1245				
Ala	Asn	Ile	Asp	Ser	Gly	Leu	Ala	Gly	Ile	Ser	Leu	Thr	Tyr	Ala	Ile	
	1250					1255					1260					
Leu	Phe	Thr	Asp	Gly	Ala	Leu	Trp	Leu	Val	Arg	Leu	Tyr	Ser	Thr	Phe	
1265					1270					1275				1280		
Glu	Met	Asn	Met	Asn	Ser	Val	Glu	Arg	Leu	Lys	Glu	Tyr	Ser	Ser	Ile	
			1285						1290				1295			
Glu	Gln	Glu	Asn	Tyr	Leu	Gly	His	Asp	Glu	Gly	Arg	Ile	Leu	Leu	Leu	
		1300					1305						1310			
Asn	Glu	Pro	Ser	Trp	Pro	Lys	Asp	Gly	Glu	Ile	Glu	Ile	Glu	Asn	Leu	
		1315				1320						1325				
Ser	Leu	Arg	Tyr	Ala	Pro	Asn	Leu	Pro	Pro	Val	Ile	Arg	Asn	Val	Ser	
	1330					1335					1340					
Phe	Lys	Val	Asp	Pro	Gln	Ser	Lys	Ile	Gly	Ile	Val	Gly	Arg	Thr	Gly	
1345					1350					1355				1360		
Ala	Gly	Lys	Ser	Thr	Ile	Ile	Thr	Ala	Leu	Phe	Arg	Leu	Leu	Glu	Pro	
			1365					1370					1375			
Ile	Thr	Gly	Cys	Ile	Lys	Ile	Asp	Gly	Gln	Asp	Ile	Ser	Lys	Ile	Asp	
		1380						1385					1390			
Leu	Val	Thr	Leu	Arg	Arg	Ser	Ile	Thr	Ile	Ile	Pro	Gln	Asp	Pro	Ile	
	1395						1400					1405				
Leu	Phe	Ala	Gly	Thr	Ile	Lys	Ser	Asn	Val	Asp	Pro	Tyr	Asp	Glu	Tyr	
	1410					1415					1420					
Asp	Glu	Lys	Lys	Ile	Phe	Lys	Ala	Leu	Ser	Gln	Val	Asn	Leu	Ile	Ser	
1425					1430					1435				1440		
Ser	His	Glu	Phe	Glu	Glu	Val	Leu	Asn	Ser	Glu	Glu	Arg	Phe	Asn	Ser	

-continued

1445	1450	1455
Thr His Asn Lys Phe Leu Asn Leu His Thr Glu Ile Ala Glu Gly Gly		
1460	1465	1470
Leu Asn Leu Ser Gln Gly Glu Arg Gln Leu Leu Phe Ile Ala Arg Ser		
1475	1480	1485
Leu Leu Arg Glu Pro Lys Ile Ile Leu Leu Asp Glu Ala Thr Ser Ser		
1490	1495	1500
Ile Asp Tyr Asp Ser Asp His Leu Ile Gln Gly Ile Ile Arg Ser Glu		
1505	1510	1515
Phe Asn Lys Ser Thr Ile Leu Thr Ile Ala His Arg Leu Arg Ser Val		
1525	1530	1535
Ile Asp Tyr Asp Arg Ile Ile Val Met Asp Ala Gly Glu Val Lys Glu		
1540	1545	1550
Tyr Asp Arg Pro Ser Glu Leu Leu Lys Asp Glu Arg Gly Ile Phe Tyr		
1555	1560	1565
Ser Met Cys Arg Asp Ser Gly Gly Leu Glu Leu Leu Lys Gln Ile Ala		
1570	1575	1580
Lys Gln Ser Ser Lys Met Met Lys		
1585	1590	
 <210> SEQ ID NO 5		
<211> LENGTH: 33		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: YCFa primer		
 <400> SEQUENCE: 5		
actaccgtaa agctcgagaa aatggctggt aat		33
 <210> SEQ ID NO 6		
<211> LENGTH: 27		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: YCFb primer		
 <400> SEQUENCE: 6		
cttgcctaag tgacgtgacg tctcctt		27
 <210> SEQ ID NO 7		
<211> LENGTH: 24		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: RT-YCF1A primer		
 <400> SEQUENCE: 7		
catgagtgcg ttctatccct ctat		24
 <210> SEQ ID NO 8		
<211> LENGTH: 24		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: RT-YCF1B primer		
 <400> SEQUENCE: 8		
ccaccttcgg ttagttgggc atct		24

-continued

<210> SEQ ID NO 9
 <211> LENGTH: 36
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: forward primer: YHL035Ca

<400> SEQUENCE: 9

cgacgcggcc gcatgggaac ggatccctt attatc

36

<210> SEQ ID NO 10
 <211> LENGTH: 35
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: reverse primer: YHL035Cb

<400> SEQUENCE: 10

cgacgcggcc gccatcatct tacttgattg cttgg

35

1. A DNA molecule exhibiting resistance to and enhanced accumulation of lead and encoding a fungal multidrug resistance-associated protein (MRP)-like ATP-binding cassette transporter protein (MRP-like ABC transporter protein).

2. The DNA molecule according to claim 1, wherein said fungal MRP-like ABC transporter protein is selected from the group consisting of YCF1 protein, YHL035C protein, BPT1 protein, YBT1 protein, and YOR1 protein.

3. The DNA molecule according to claim 2, wherein said YCF1 protein has the amino acid sequence of SEQ ID NO:2 and the DNA molecule has a nucleotide sequence encoding said amino acid sequence.

4. The DNA molecule according to claim 3, wherein said DNA molecule encoding YCF1 protein has the nucleotide sequence of SEQ ID NO:1.

5. The DNA molecule according to claim 2, wherein said YCF1 protein further exhibits resistance to and enhanced accumulation of one or more noxious materials selected from the group of cadmium, arsenic, and herbicide, in addition to resistance to and accumulation of lead.

6. The DNA molecule according to claim 5, wherein said herbicide is chloro-dinitrobenzene.

7. The DNA molecule according to claim 2, wherein said YHL035C protein has the amino acid sequence of SEQ ID NO:4 and the DNA molecule has a nucleotide sequence encoding said amino acid sequence.

8. The DNA molecule according to claim 7, wherein said DNA molecule encoding YHL035C protein has the nucleotide sequence of SEQ ID NO:3.

9. The DNA molecule according to claim 1, wherein said fungal MRP-like ABC transporter protein is a fungal MRP-like transporter protein having at least 28% amino acid sequence homology with the amino acid sequence of YCF1 protein of SEQ ID NO:2 or the amino acid sequence of YHL035C protein of SEQ ID NO:4.

10. The DNA molecule according to claim 9, wherein said fungal MRP-like ABC transporter protein has at least 28%

sequence homology with the amino acid sequence of YCF1 protein of SEQ ID NO:2 or YHL035C protein of SEQ ID NO:4, and

wherein said YCF1 protein and YHL035C protein each comprise an N-terminal extension domain, a first transmembrane spanning domain, a first nucleotide binding fold domain, a second transmembrane spanning domain, and a second nucleotide binding fold domain, and each domain of said fungal MRP-like ABC transporter proteins has at least 28% sequence homology with the amino acid sequence of each corresponding domain of YCF1 protein of SEQ ID NO:2 or YHL035C protein of SEQ ID NO:4.

11. A recombinant vector comprising the DNA molecule encoding the fungal MRP-like ABC transporter protein according to claim 1.

12. The recombinant vector according to claim 11, wherein said recombinant vector is pESC-YCF1, ENpCambia-YCF1, or PBI121-YCF1.

13. The recombinant vector according to claim 11, wherein said recombinant vector is a pESC-YHL035C recombinant vector, or pPBI121-YHL035C recombinant vector.

14. A transgenic organism transformed with the DNA molecule encoding the fungal MRP-like ABC transporter protein according to claim 1.

15. The transgenic organism according to claim 14, wherein said transgenic organism is a prokaryotic organism or a eukaryotic organism.

16. The transgenic organism according to claim 15, wherein said eukaryotic organism is animal, plant, or yeast.

17. The transgenic organism according to claim 16, wherein said plant is selected from the group of *Arabidopsis*, rapeseed, leaf mustards, tobaccos, onions, carrots, cucumbers, sweet potatoes, potatoes, napa cabbages, radishes, lettuces, broccoli, petunias, sunflowers, grass, white birch, poplar, olive, willow, and birch.

18. The transgenic organism according to claim 17, wherein said transgenic organism is YCF1 *Arabidopsis thaliana* (accession number KCTC10064BP) or YCF1 poplar.

19. The transgenic organism according to claim 17, wherein said transgenic organism is YHL035C poplar.

20. A plant cell transformed with the DNA molecule encoding the fungal MRP-like ABC transporter protein according to claim 1.

21. The plant cell according to claim 20 wherein said plant is selected from the group of *Arabidopsis*, rapes, leaf mustards, tobaccos, onions, carrots, cucumbers, sweet potatoes, potatoes, napa cabbages, radishes, lettuces, broccoli, petunias, sunflowers, grass, white birch, poplar, olive, willow, and birch.

* * * * *