



No. 11-165857-00000-0000

Fecha: 2011-12-01 16:35:55

Dep. 2020 DIR.NUEVASCR

Tra. 11 PATENTEIYII

Eve: 378 FASENACIONALI

Act. 411 PRESENTACION

Folios: 294

Act. 411 PRESENTACION

Folios: 300



Industria y Comercio

SUPERINTENDENCIA

DELEGATURA DE PROPIEDAD INDUSTRIAL

División de Nuevas Creaciones

SOLICITUD

PATENTE DE INVENCION

21. EXPEDIENTE No. _____

VARIANTES DE CELULASA CON EXPRESIÓN,

54. TÍTULO _____

ACTIVIDAD Y/O ESTABILIDAD MEJORADAS,

Y USO DE LAS MISMAS

51. CLASIFICACIÓN INTERNACIONAL _____

DANISCO US INC

71. SOLICITANTE _____

925 Page Mill Road, Palo Alto, California 94304, E.U.A

DOMICILIO _____

JERÓNIMO CASTILLO BONILLA

74. APODERADO _____

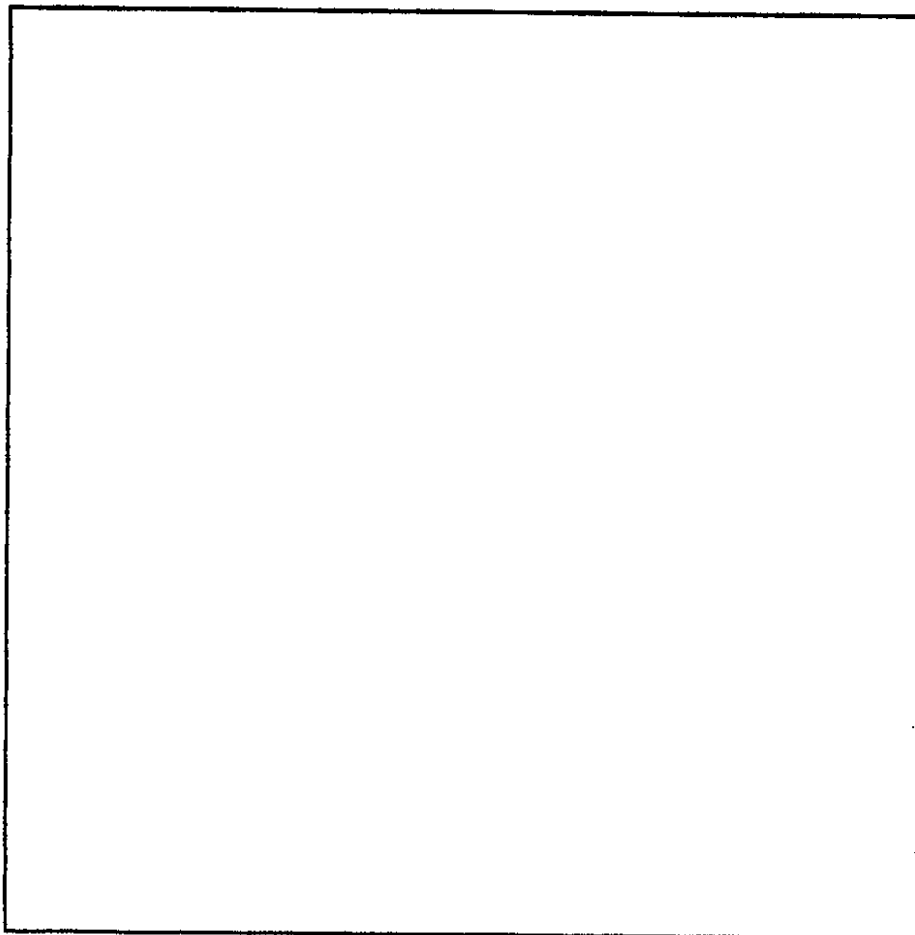
22. BOGOTÁ, D. C., _____

 Industria y Comercio SUPERINTENDENCIA		PETI		SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO  No. 11-165857-00000-0000 Fecha: 2011-12-01 16:35:55 Dep. 2020 DIR.NUEVASCR Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI Act. 411 PRESENTACION Folios: 294 Act. 411 PRESENTACION Folios: 300	
FORMULARIO ÚNICO DE SOLICITUD DE PATENTES PCT ENTRADA FASE NACIONAL					
SOLICITUD DE REGISTRO DE: ☒ Patente de Invención ☐ Patente Modelo de Utilidad					
☒ Capítulo I ☐ Capítulo II					
SOLICITANTE		Nombre: DANISCO US INC. Dirección: 925 Page Mill Road Domicilio: Palo alto, California 94304, Estados Unidos de América Lugar de Constitución: Delaware Teléfono: Fax: E - mail:		IDENTIFICACIÓN ☐ C.C. ☐ NIT ☐ C.E. ☐ Otro Cual Número:	
REPRESENTANTE O APODERADO		Nombre: JERÓNIMO CASTILLO BONILLA Dirección: Calle 93A N° 14-17, Ofc 311 Teléfono: 2363493 Fax: 2182351 E-mail: info@pcastillopineda.com		IDENTIFICACIÓN ☒ C.C. ☐ NIT ☐ C.E. ☐ Otro Número: 79462057 TP: 104167	
INVENTORES		Nombre: NIKOLAEV, Igor Dirección: 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América Nacionalidad o Domicilio: Rusa			
PCT Solicitud Internacional No.: PCT/US2010/037328 Fecha: Junio 3, 2010 Publicación Internacional No.: WO2010/141779 A1 Fecha: Diciembre 9, 2010					
Título: VARIANTES DE CELULASA CON EXPRESIÓN, ACTIVIDAD Y/O ESTABILIDAD MEJORADAS, Y USO DE LAS MISMAS					
Clasificación Internacional:					
Prioridad ☒ SI ☐ NO		País de Origen US	Fecha Junio 3, 2009	Número de Solicitud 61/183.959	
Comprobante de pago No:		Fecha:			
Instrucciones para el diligenciamiento del presente formulario, ver reverso de esta página					

Anexos

- ☒ Comprobante de pago de la tasa de presentación de la solicitud.
- ☒ Documento que acredita la existencia y representación legal de la persona jurídica peticionaria, en cuanto no estén inscritas en alguna Cámara de Comercio Colombiana.
- ☒ Poderes, si fuere el caso.
- ☐ Documento de cesión del inventor al solicitante o a su causante.
- ☒ Traducción de la solicitud internacional al castellano (Resumen, descripción, reivindicaciones, dibujos)
- ☐ Copia del examen preliminar internacional efectuado por la Autoridad Internacional de Examen preliminar (IPEA)
- ☐ Traducción de examen preliminar internacional efectuado por la IPEA
- ☐ De ser el caso, copia del contrato de acceso.
- ☐ De ser el caso, documento que acredite la licencia o autorización de uso de conocimientos tradicionales de las comunidades indígenas.
- ☐ De ser el caso, certificado de depósito del material biológico
- ☐ Arte final 12x12.
- ☐ De ser el caso, información sobre otras solicitudes de patente o títulos obtenidos en el extranjero por el mismo titular o su causante, relacionadas parcial o totalmente con la invención de esta solicitud.
- ☐ De ser el caso, modificaciones efectuadas a la solicitud internacional.
- ☒ Otros. Modificación nombres inventores e inclusión solicitante para E.U.A.

FIGURA CARACTERÍSTICA:



Solicito la concesión de la patente

NOMBRE: JERÓNIMO CASTILLO B.

FIRMA:

C.C. 79462057

T.P. 104167

Instrucciones para el diligenciamiento del presente formulario, ver reverso de esta página

(Continuación Inventores)

Nombre : AEHLE, Wolfgang
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : alemana

Nombre : BOTT, Richard R.
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : estadounidense

Nombre : BOWER, Benjamin
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : estadounidense

Nombre : CASPI, Jonathan
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : estadounidense

Nombre : ESTELL, David A.
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : estadounidense

Nombre : GOEDEGEBUUR, Frits
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : Holandesa

Nombre : HOMMES, Ronaldus W., J.
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : Holandesa

Nombre : KAPER, Thijs
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : Holandesa

Nombre : KELEMEN, Bradley
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : estadounidense

Nombre : KRALJ, Slavko
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : Holandesa

Nombre : VAN LIESHOUT, Johan
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América

Nombre : VAN STIGT THANS, Sander
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : Holandesa

Nombre : WALLACE, Louise
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : sudafricana

Nombre : VOGTENTANZ, Gudrun
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América
Nacionalidad : alemana

Nombre : SANDGREN, Mats
Dirección : Microbiology Program, Biomedical Center, Box 590, Husarg 3, S-751 Upsala, Suecia
Nacionalidad : sueca

EXTRACTO

La presente invención se relaciona con enzimas y en particular con variantes de celulasa con expresión, actividad y/o estabilidad mejoradas. También describe los ácidos nucleicos que codifican las variantes de celulasa, los métodos para identificar las variantes de celulasa útiles y los métodos de uso de los mismos.

State of California

SECRETARY OF STATE

Requested for use in Colombia.

Not for use within the United States of America.

The purpose of the Apostille is to certify the authenticity of the signature of the official signing the document, the capacity in which the official signing the document has acted, and, where appropriate, the identity of the seal or stamp.

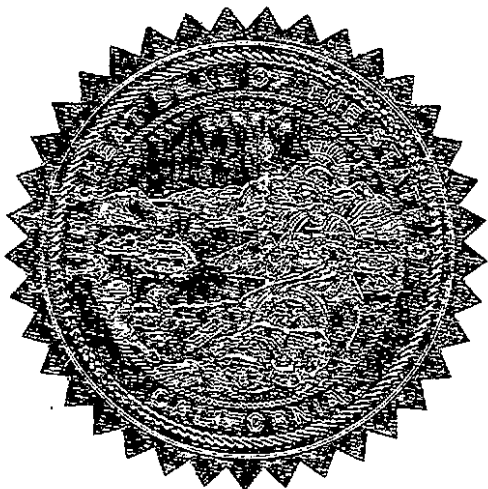
APOSTILLE

(Convention de La Haye du 5 octobre 1961)

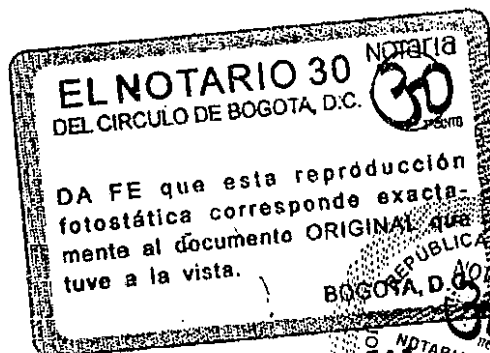
1. Country: *United States of America*
This public document
2. *has been signed by Carol A. See*
3. *acting in the capacity of Notary Public, State of California*
4. *bears the seal/stamp of Carol A. See, Notary Public,*
State of California

CERTIFIED

5. *At Sacramento, California*
6. *the 1st day of September 2009*
7. *by Deputy Secretary of State, State of California*
8. *No. 723031*
9. *Seal/Stamp:*



10. Signature



30 NOV 2011

Jana Bowen
Secretary of State

BY *[Signature]*



6

LAW OFFICES
PEDRO CASTILLO PINEDA &
ASSOCIATES
P.O. BOX 48362
Bogotá, Colombia / South America

IMPORTANT NOTE

The lodging of this GENERAL POWER with the Patent Office obviates the need of presenting special powers with each individual trade mark application.

FOR PATENTS AND TRADE MARKS IN
COLOMBIA

1 Applicant's name	(1)	Poder General	General Power of Attorney
			(1) Danisco US Inc.

2 Applicant's address	domiciliado en (2)
-----------------------	--------------------

domiciled in (2)
925 Page Mill Road, Palo Alto
California 94304
U.S.A.

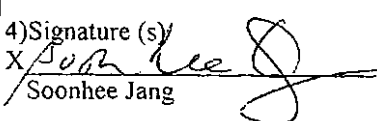
3 Date
4 Applicant's signature

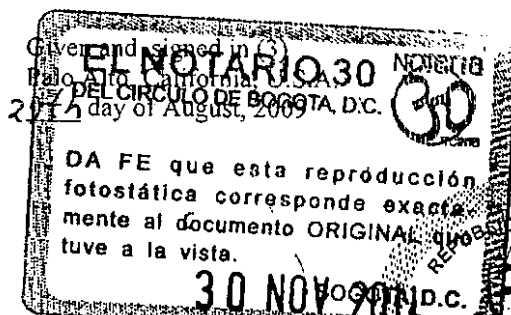
nombra al Dr. JERONIMO CASTILLO BONILLA y/o Dra. XIMENA BONILLA PEREIRA y/o Dr. ALVARO CASTILLO GRAU Calle 93A N° 14-17 Oficina 311 en Bogotá, Colombia, como sus representantes con poder especial, amplio y suficiente, para recabar de las oficinas y autoridades Colombianas, administrativas, judiciales o de policía, en primera o segunda instancia, la obtención de Patentes de Invención, Registro de Marcas, de Modelos y de Dibujos Industriales, de Nombres Comerciales y Enseñas, Lemas Comerciales y Denominaciones de Origen, lo mismo que traspasos, extensiones, renovaciones, cambios de nombre, registro de licencias de explotación y registro de licencias de uso referentes a tales actuaciones, cancelación de Registros de marca, a cuyo efecto les faculta para dar ante dichas autoridades todos los pasos necesarios al objeto indicado, elevar solicitudes, formular descripciones, enmiendas, protestas, declaraciones, apelaciones y reclamos, solicitar la aplicación de medidas cautelares; solicitar la nulidad de Patentes de Invención, modelos de utilidad, diseños industriales, así como la de registros de marcas ante la autoridad nacional competente; aceptar en nuestro nombre y representación la Cesión de derechos de invención y de patente de invención que se nos haga por cualquier persona, abonar todos los impuestos, cuotas y pagos determinados por la ley, recibir todos los documentos y valores dando el descargo respectivo, llenar cualesquiera otros requisitos, desistir y tomar en fin todas las medidas que creyeren conducentes al resguardo de nuestros intereses y en caso de presentarse oposiciones y/o observaciones para que intervengan como demandantes o como demandados, con todas las demás facultades necesarias, así como iniciar y tramitar procesos jurisdiccionales y/o administrativos relativos a competencia desleal. Este poder comprende además las facultades de recibir, desistir, transigir, comprometer, sustituir, revocar sustituciones, conciliar, renunciar, recibir notificaciones y nombrar apoderados judiciales o extrajudiciales.
Dado y firmado en (3)

hereby appoint, JERONIMO CASTILLO BONILLA and/or XIMENA BONILLA PEREIRA and/or ALVARO CASTILLO GRAU, Calle 93A N° 14-17 Office 311 en Bogotá, Colombia, as our representatives with special, ample and sufficient power to obtain from the offices and Colombian authorities, administrative, judicial and police, in first or in second instance, Privileged Patents of invention, Registration of Trade Marks, of Models and Industrial Drawings, of Commercial Names and ensigns, Commercial slogans and Origin Denominations as well as assignments, extensions, renewals, changes of name, registration of license of exploitation and of license of use referred to said actions, Cancellation Actions against trademark Registrations, who are empowered to take before said authorities all of the necessary steps for the purposes indicated, to file applications, to formulate descriptions, corrections, protests, declarations, appeals, and claims, apply for precautionary measures; apply for the nullity of Patent Inventions, Utility Models, Industrial Designs, as well as for Trademark Registrations before the competent national authority; to accept on our behalf and representation the assignment of rights of inventions and titles of invention that may be done to us by any person, to pay all taxes, quotas and payments prescribed by law, to receive all the documents and proceeds giving the respective acquittance or receipt, to fulfill any other requisites to desist and take in fact any other measures that may be deemed conducive to the safeguard of our interests, and in case that oppositions and/or observations are presented to, intervene as complainant and defendant, with all necessary legal faculties, as well as to file jurisdictional processes and/or administrative related to disloyal competition.

This power of attorney includes the faculties to receive, desist, compromise, transact substitute it in all or in part, to revoke substitutions, to conciliate, to renounce, to receive notifications and to appoint extra-judicial or judicial attorneys.

4) Signature (s)

X 
Soonhee Jang



DECLARACION NOTARIAL
(Individual)

A los días del mes de
 compareció y/o comparecieron

personalmente ante mi, Notario Público

quien (es) doy fe de conocer y me consta que es
(son) la(s) persona(s) descrita(s) y firmante(s) del
documento que precede, declarando ante mi que
otorga(n) el mismo para los fines y propósitos que
en él se expresan.

Notario Público,

.....

(Sociedades)

El suscrito Notario Certifica: que la firma que
antecede y dice

es auténtica; que el signatario desempeña
actualmente el cargo de

de,

que tiene facultad para otorgar este poder; y que
dicha compañía esta domiciliada en

y actualmente existe y esta organizada de acuerdo
a las leyes de

Todos los hechos anteriores le constan por haber
examinado los documentos respectivos y de todo
ello da fe.

Dado y firmado en

El Notario Público,

.....

(Con legalización Consular)

NOTARIAL CERTIFICATE
(Individual)

On this day of of
 personally appeared before me

a Notary Public

to me known, and known to me to be the
person(s) described in and who executed the
foregoing document, and he (they) duly
acknowledged to me that he (they) executed
same for the uses and purposes therein expresed.

.....

Notary Public.

(Corporations)

The undersigned Notary Public certifies: that the
preceding signature reading
Soonhee Jang

is authentic; that the signer at present fills the post
of Vice President, IP Strategy & Chief IP Counsel

of Danisco US Inc.

and that she is empowered to grant this power; and
that said company is domiciled in
Palo Alto, California, U.S.A.

and actually exists and is organized in accordance
with the laws of Delaware, USA

All the foregoing facts are known to the Notary
due to her having examined the respective
documents to which she, the Notary attests..

Given and signed in Palo Alto, California, USA
Carol A. See, Notary Public

Carol A. See

Notary Public.

(With Consular legalization)



From the INTERNATIONAL BUREAU

PCT**NOTIFICATION OF THE RECORDING
OF A CHANGE**(PCT Rule 92bis.1 and
Administrative Instructions, Section 422)

To:

KAPLAN, Jonathan, M.
DANISCO US INC.
925 Page Mill Road
Palo Alto, California 94304
ETATS-UNIS D'AMERIQUE

Date of mailing (day/month/year) 03 February 2011 (03.02.2011)	IMPORTANT NOTIFICATION
Applicant's or agent's file reference 31414WO-2	
International application No. PCT/US2010/037328	International filing date (day/month/year) 03 June 2010 (03.06.2010)

1. The following indications appeared on record concerning:		
☒ the applicant	☒ the inventor	☐ the agent
☐ the common representative		
Name and Address NIKOLAEV, Igor DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 United States of America	State of Nationality RU	State of Residence US
	Telephone No.	
	Facsimile No.	
	E-mail address	
2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:		
☐ the person	☐ the name	☐ the address
☐ the nationality	☐ the residence	
Name and Address	State of Nationality	State of Residence
	Telephone No.	
	Facsimile No.	
	E-mail address ☐ Notifications by e-mail authorized	
3. Further observations, if necessary: The person identified in Box 1 is now recorded as applicant for the United States of America only and inventor for all designated States.		
4. A copy of this notification has been sent to:		
☒ the receiving Office	☐ the International Preliminary Examining Authority	
☐ the International Searching Authority	☒ the designated Offices concerned	
☐ the Authority(ies) specified for supplementary search	☐ the elected Offices concerned	
	☐ other:	

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Ratsimba Landry-Andosoa e-mail pt03.pct@wipo.int Telephone No. +41 22 338 74 03
---	---

Facsimile No. +41 22 338 87 20

Form PCT/IB/306 (January 2009)

1/EQ3PTGJ90

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING
OF A CHANGE

(PCT Rule 92bis.1 and
Administrative Instructions, Section 422)

To:	RECEIVED
KAPLAN, Jonathan, M. DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 ETATS-UNIS D'AMERIQUE	SEP 30 2011 PATENT DEPARTMENT

Date of mailing (day/month/year) 02 September 2011 (02.09.2011)	IMPORTANT NOTIFICATION
Applicant's or agent's file reference 31414WO-2	
International application No. PCT/US2010/037328	International filing date (day/month/year) 03 June 2010 (03.06.2010)

1. The following indications appeared on record concerning:

☒ the applicant ☒ the inventor ☐ the agent ☐ the common representative

Name and Address KELEMAN, Bradley DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 United States of America	State of Nationality US	State of Residence US
	Telephone No.	
	Facsimile No.	
	E-mail address	

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

☐ the person ☒ the name ☐ the address ☐ the nationality ☐ the residence

Name and Address KELEMEN, Bradley DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 United States of America	State of Nationality US	State of Residence US
	Telephone No.	
	Facsimile No.	
	E-mail address ☐ Notifications by e-mail authorized	

3. Further observations, if necessary:

4. A copy of this notification has been sent to:

☒ the receiving Office	☐ the International Preliminary Examining Authority
☐ the International Searching Authority	☒ the designated Offices concerned
☐ the Authority(ies) specified for supplementary search	☐ the elected Offices concerned
	☐ other:

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Marquis Catherine
Facsimile No. +41 22 338 87 20	e-mail pt03.pct@wipo.int Telephone No. +41 22 338 74 03

PCT REQUEST

Original (for SUBMISSION)

III-10	Applicant and/or inventor	Applicant and inventor US only KRALJ, Slavko DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 United States of America
III-10-1	This person is	
III-10-2	Applicant for	
III-10-4	Name (LAST, First)	
III-10-5	Address	
III-10-6	State of nationality	NL
III-10-7	State of residence	US
III-11	Applicant and/or inventor	Applicant only All designated States except US NIKOLAEV, Igor DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 United States of America
III-11-1	This person is	
III-11-2	Applicant for	
III-11-4	Name (LAST, First)	
III-11-5	Address	
III-11-6	State of nationality	RU
III-11-7	State of residence	US
III-12	Applicant and/or inventor	Applicant and inventor US only VAN LIESHOUT, Johan DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 United States of America
III-12-1	This person is	
III-12-2	Applicant for	
III-12-4	Name (LAST, First)	
III-12-5	Address	
III-12-6	State of nationality	NL
III-12-7	State of residence	US
III-13	Applicant and/or inventor	Applicant and inventor US only VAN STIGT TRANS, Sander DANISCO US INC. 925 Page Mill Road Palo Alto, California 94304 United States of America
III-13-1	This person is	
III-13-2	Applicant for	
III-13-4	Name (LAST, First)	
III-13-5	Address	
III-13-6	State of nationality	NL
III-13-7	State of residence	US

**VARIANTES DE CELULOSA CON EXPRESIÓN, ACTIVIDAD Y/O
ESTABILIDAD MEJORADAS, Y USO DE LAS MISMAS**

5

REFERENCIA A SOLICITUDES RELACIONADAS

[01] Esta solicitud reivindica el beneficio de prioridad de la Solicitud Provisional de los Estados Unidos No. 61/183.959, presentada el 3 de junio de 2009, la cual se incorpora acá en su totalidad por referencia.

- 10 [02] Esta invención fue hecha con el soporte del Gobierno bajo el premio condicional No. DE-FC36-08GO 18708 otorgado por el Departamento de Energía. El gobierno tiene ciertos derechos en esta invención.

I. CAMPO

- 15 [03] La presente divulgación se relaciona con enzimas y en particular con variantes de celulasa. También se describen los ácidos nucleicos que codifican las variantes de celulasa, composiciones que comprenden las variantes de celulasa, los métodos para identificar las variantes de celulasa útiles y los métodos de uso de los mismos.

20 II. ANTECEDENTES

- [04] La celulosa y la hemicelulosa son los materiales de planta más abundantes producidos por fotosíntesis. Estas pueden ser degradadas y utilizadas como fuente de energía por numerosos microorganismo (por ejemplo, bacterias, levadura y hongos) que producen enzimas extracelulares capaces de hacer la hidrólisis de los sustratos poliméricos a azúcares monoméricos (Aro y otros, J Biol Chem, 276: 24309 a 24314, 2001). A medida que se alcanzan los límites de los recursos no renovables, el potencial de que la celulosa se convierta en una fuente de energía renovable mayor es enorme (Krishna y otros, Bioresource Tech, 77: 193 a 196, 2011). La utilización efectiva de la celulosa a través de

- 25 [05] procesos biológicos es un acercamiento a superar la escasez de alimentos y combustibles (Ohmiya y otros, Biotechnol Gen Engineer Rev, 14: 365 a 414, 1997).

[06] Las celulasas son enzimas que hidrolizan celulosa (uniones beta-1,4-glucan o beta D-glucodídico) lo que resulta en la formación de glucosa, celobiosa, celooligosacáridos,

- y similares. Las celulasas han sido tradicionalmente divididas en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) ("CBH") y beta-glucosidasas (glucohidrolasa [beta]-D-glucosida; EC 3.2.1.21) ("BG") (Knowles y otros, TIBTECH 5: 255 a 261, 1987; y Schelein, Métodos de Enzimol, 160: 234 a 243, 1988). Las endoglucanasas actúan principalmente en las partes amorfas de la fibra celulosa, mientras que las celobiohidrolasas son también capaces de degradar celulosa cristalina (Nevalainen y Penttilä, Mycota, 303 a 319, 1995). Así, la presencia de una celobiohidrolasa en un sistema de celulasa es requerida para la solubilización eficiente de la celulosa cristalina (Suurnakki y otros, Celulosa 7: 189 a 209, 2000). La
- 5 beta-glucosidasa actúa para liberar unidades de D-glucosa a partir de celobiosa, celooligosacáridos, y otros glucósidos (Freer, J Biol Chem, 268: 9337 a 9342, 1993).
- [07] Las celulasas son conocidas por ser producidas por un amplio número de bacterias, levadura y hongos. Ciertos hongos producen un sistema de celulasa completo capaz de degradar formas cristalinas de celulosa, de modo tal que las celulasas son producidas
- 15 fácilmente en grandes cantidades por medio de fermentación. Los hongos filamentosos juegan un papel especial dado que varias levaduras, tal como *Saccharomyces cerevisiae*, carecen de la habilidad para hidrolizar celulosa (Ver, por ejemplo, Wood y otros, Métodos en Enzimología, 60: 87 a 116, 1988).
- [08] Las clasificaciones de celulasa por hongo de CBH, EG y BG pueden ser además
- 20 expandidas para incluir múltiples componentes dentro de cada clasificación. Por ejemplo, múltiples CBH, EG y BG han sido aislados de una variedad de fuentes de hongos incluyendo *Trichoderma reesei* (también conocido como *Hypocrea jecorina*), el cual contiene genes conocidos para dos CBH, es decir, CBH I ("CBH1") y CBH II ("CBH2"), por lo menos 8 EG, es decir, EG I, EG II, EG III, EG IV, EG V, EG VI, EG
- 25 VII y EG VIII, y por lo menos 5 BG, es decir, BG1, BG2, BG3, BG4 y BG5. Los EGIV, EGVI y EGVIII también tienen actividad xiloglucanasa.
- [09] Con el fin de convertir eficientemente la celulosa cristalina en glucosa, se requiere el sistema de celulasa completo comprende componentes de cada uno de las clasificaciones de CBH, EG y BG, con componentes aislados menos efectivos en la
- 30 hidrolización de celulosa cristalina (Filho y otros, Can J Microbiol, 42: 1 a 5, 1996). Una relación sinérgica se ha observado entre los componentes de celulasa de diferentes clasificaciones. En particular, las celulasas de tipo EG y las celulasas de tipo CBH interactúan sinérgicamente para degradar celulosa más eficientemente.

- [10] Las celulasas son conocidas en el estado de la técnica por ser útiles en el tratamiento de textiles para propósitos de mejorar la habilidad de limpieza de las composiciones detergentes, para su uso como un agente suavizante para mejorar la textura y la apariencia de telas de algodón; y similares (Kumar y otros, Química Textil y de Colores, 29: 37 a 42, 1997). Las composiciones detergentes que contienen celulasa con desempeño de limpieza mejorado (Patente de los Estados Unidos No. 4.435.307; Solicitudes de Reino Unido No. 2.095.275 y 2.094.826) y para uso en el tratamiento de telas para mejorar la textura y la apariencia del textil (Patentes de los Estados Unidos No. 5.648.263, 5.691.178, y 5.776.757; y Solicitud de Reino Unido No. 1.358.59), han sido descritas. Así, las celulasas producidas en hongos y bacterias han recibido atención significativa. En particular, la fermentación de *Trichoderma* spp. (por ejemplo, *Trichoderma longibrachiatum* o *Trichoderma reesei*) se ha mostrado para producir un sistema de celulasa completo capaz de degradar las formas cristalinas de celulosa.
- [11] Aunque las composiciones de celulasa han sido descritas anteriormente, aún permanece una necesidad por nuevas y mejoradas composiciones de celulasa. Las composiciones de celulosa mejoradas son utilizadas en detergentes caseros, tratamientos de textiles, conversión de biomasa y fabricación de papel. Las celulasas que presentan expresión, actividad y estabilidad mejoradas son de interés particular.

20 III. RESUMEN

- [12] La presente divulgación se relaciona con enzimas y en particular con variantes de celulasa. También se describen ácidos nucleicos que codifican las variantes de celulasa, composiciones que comprenden las variantes de celulasa, métodos para identificar variantes de celulasa adicionales útiles y métodos de uso de las mismas.
- 25 [13] La presente divulgación proporciona variantes de celulasa, en donde las variantes son formas maduras que tienen actividad de celulasa y que comprenden una sustitución en una o más posiciones seleccionadas a partir del grupo que consiste de: 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 30 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 y 447, en donde las posiciones son

enumeradas por correspondencia con la secuencia de aminoácido de una celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3. En algunas incorporaciones, la sustitución en una o más posiciones resulta en una variante de celulasa con expresión, actividad y/o estabilidad mejoradas en comparación con la CBH2 de referencia. En algunas incorporaciones, la variante comprende una sustitución adicional en una o más posiciones adicionales seleccionadas a partir de: (i) un primer grupo que consiste de 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 y 405; o (ii) un segundo grupo que consiste de 94, 98, 107, 120, 134, 144, 147, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, y 429, en donde las posiciones adicionales son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3. En algunas incorporaciones, la sustitución en una o más posiciones se selecciona a partir del grupo que consiste de 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10 posiciones.

[14] En otro aspecto, la divulgación suministra variantes de celulasa, en donde las variantes son formas maduras que tienen actividad de celulasa y que comprenden una sustitución en una o más posiciones seleccionadas a partir del grupo que consiste de: 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 y 405, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de una celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3 y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una sustitución K/E, R/Q, N/D, Q/E, D/N o E/Q. En algunas incorporaciones, la sustitución en una o más posiciones resulta en una variante de celulasa con expresión, actividad y/o estabilidad mejoradas en comparación con la CBH2 de referencia. En algunas incorporaciones, la variante de celulasa aislada comprende una sustitución adicional en una o más posiciones adicionales seleccionadas a partir de: (i) un primer grupo que consiste de 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362,

- 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 y 447; o (ii) un segundo grupo que consiste de 94, 98, 107, 120, 134, 144, 147, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, y 429, en donde las posiciones adicionales
- 5 son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3. En algunas incorporaciones, la sustitución en una o más posiciones se selecciona a partir del grupo que consiste de 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10 posiciones.
- [15] En otro aspecto, la divulgación suministra variantes de celulasa, en donde las
- 10 variantes son formas maduras que tienen actividad de celulasa y que comprenden una sustitución en una o más posiciones seleccionadas a partir del grupo que consiste de: 94, 98, 107, 120, 134, 144, 147, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, y 429, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de una
- 15 celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una del grupo que consiste de V94E, P98L, E107Q, M120L, M134G, M134L, M134V, L144G, L144R, L144S, T154A, A178V, L179C, V206L, S210L, S210R, T214M, T214Y, G231P, G231I, G231A, G231N, G231S, G231T, T232V, H266S,
- 20 H266A, W272A, W272D, W272Y, N275L, S316P, V323L, V323N, V323Y, G360R, S380T, A381T, E399N, E399D, R413Y, R413P, y A416G. En algunas incorporaciones, la sustitución en una o más posiciones resulta en una variante de celulasa con expresión, actividad y/o estabilidad mejoradas en comparación con la CB H2 de referencia. En algunas incorporaciones, la variante comprende una sustitución adicional en una o más
- 25 posiciones adicionales seleccionadas a partir de: (i) un primer grupo que consiste de 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322,
- 30 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 y 447; o (ii) un segundo grupo que consiste de 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344,

356, 378, 382 y 405, en donde las posiciones adicionales son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3. En algunas incorporaciones, la sustitución en una o más posiciones se selecciona a partir del grupo que consiste de 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10 posiciones.

[16] La presente divulgación además suministra variantes de celulasa, en donde las variantes son formas maduras que tienen actividad de celulasa y que comprenden una sustitución en de una a seis posiciones seleccionadas a partir del grupo que consiste de 111, 144, 154, 162, 410 y 413, en donde la variante de celulasa comprende: una leucina o una serina en la posición 111; una leucina, una glutamina o un triptófano en la posición 144; una treonina, una cisteína o una valina en la posición 154; una tirosina o una asparagina en la posición 172; una arginina o una serina en la posición 410; y una serina, un triptófano o una tirosina en la posición 413; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una sustitución S413Y.

[17] También se suministran por medio de la presente divulgación variantes de celulasa, en donde las variantes son formas maduras que tienen actividad de celulasa y que comprenden una sustitución en de una a seis posiciones seleccionadas a partir del grupo que consiste de 98, 194, 313, 316, 384 y 443, en donde la variante de celulasa comprende: una prolina o una leucina en la posición 98; una lisina, una cisteína o un ácido glutámico en la posición 194; una serina o una treonina en la posición 313; una serina o una prolina en la posición 316; una glicina, una cisteína o una glutamina en la posición 384; y una asparagina o una isoleucina en la posición 443; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una del grupo seleccionado de una sustitución P98L, una sustitución K194E, y una sustitución S316P.

[18] Adicionalmente, la presente divulgación suministra variantes de celulasa, en donde las variantes son formas maduras que tienen actividad de celulasa y que comprenden una sustitución en de una a seis posiciones seleccionadas a partir del grupo que consiste

de 153, 161, 203, 233, 422 y 444, en donde la variante de celulasa comprende: una arginina o una glutamina en la posición 153; una asparagina, una alanina o un triptófano en la posición 161; una arginina o una histidina en la posición 203; una prolina o un ácido aspártico en la posición 233; una glutamina o una valina en la posición 422; y una prolina o una glutamina en la posición 443; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una sustitución R153Q.

[19] Además, la presente divulgación suministra variantes de celulasa, en donde las variantes son formas maduras que tienen actividad de celulasa y que comprenden una sustitución en de una a siete posiciones seleccionadas a partir del grupo que consiste de 98, 111, 144, 313, 316, 413 y 422, en donde la variante de celulasa comprende: una prolina o una leucina en la posición 98; una leucina o una serina en la posición 111; una leucina o un triptófano en la posición 144; una serina o una treonina en la posición 313; una serina o una prolina en la posición 316; una serina o un triptófano en la posición 413; y una glutamina o una valina en la posición 422; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una del grupo seleccionado de una sustitución P98L, y una sustitución S316P.

[20] En aspectos adicionales de la divulgación se suministran variantes de celulasa, en donde las variantes comprenden una glutamina en la posición 98, y una sustitución seleccionada del grupo que consiste de un T138C, un S316P, un S343Q, un Q362I, un S386C, un C400S y un S406P, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3.

[21] Adicionalmente, se suministran variantes de celulasa las cuales comprenden una cisteína en la posición 138, y una sustitución seleccionada a partir del grupo que consiste de un S316P, un S343Q, un Q362I, un S386C, un C400S y un S406P, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3.

[22] En otro aspecto, se suministran variantes de celulasa las cuales comprenden una prolina en la posición 316, y una sustitución seleccionada del grupo que consiste de un

S343Q, un Q362I, un S386C, un C400S y un S406P, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3.

[23] Además; la presente divulgación suministra variantes de celulasa que comprenden una glutamina en la posición 343, y una sustitución seleccionada del grupo que consiste de un Q362I, un S386C, un C400S y un S406P, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3.

[24] La presente divulgación también suministra variantes de celulasa que comprenden una isoleucina en la posición 362, y una sustitución seleccionada del grupo que consiste de un S386C, un C400S y un S406P, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3.

[25] En un aspecto, se suministran variantes de celulasa que comprenden una cisteína en la posición 386, y una sustitución seleccionada del grupo que consiste de un C400S y un S406P, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3. En algunas incorporaciones, las variantes comprenden una serina en la posición 400 y una prolina en la posición 406.

[26] Adicionalmente, la presente divulgación suministra variantes de celulasa, en donde la variante es una forma madura que tiene actividad de celulasa y que comprende una sustitución en de una a siete posiciones seleccionadas del grupo que consiste de 98, 111, 182, 291, 316, 362 y 400, en donde la variante de celulasa comprende: prolina, leucina o glutamina en la posición 98; leucina o serina en la posición 111; aspargina o triptófano en la posición 182; serina o un ácido glutámico en la posición 291; serina o prolina en la posición 316; glutamina, isoleucina o leucina en la posición 362; y cisteína o serina en la posición 400; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una sustitución P98L o una sustitución S316P. En algunas incorporaciones, la variante además comprende una sustitución L439P o una sustitución T74S.

- [27] En algunas incorporaciones preferidas de la divulgación, las variantes de celulasa de cualquiera de los párrafos anteriores del sumario se derivan de una celulasa madre seleccionada del grupo que consiste de CBH2 *Hypocrea jecorina*, CBH2 *Hypocrea koningii*, CBH2 *Humicola insolens*, CBH2 *Acremonium cellulolyticus*, CBH2 *Agaricus bisporus*, EG *Fusarium osysporum*, CBH2 *Phanerochaete chrysosporium*, CBH2 *Talaromyces emersonii*, 6B/E3 CBH2 *Thermobifida fusca*, 6A/E2 EG *Thermobifida fusca*, y EG *Cellulomonas fimi* CenA. En algunas incorporaciones, la variante de celulasa se deriva de una celulasa madre cuya secuencia de aminoácido es por lo menos 75% 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% idéntica a un miembro del grupo que consiste de SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, y SEQ ID NO: 13. En algunas incorporaciones, las variantes son aisladas. También se suministran composiciones que comprenden una variante de celulasa. En algunas incorporaciones preferidas, la composición es enriquecida en la variante de celulasa.
- 15 [28] En otros aspectos, la presente divulgación suministra un ácido nucleico asilado que codifica la variante de celulasa de cualquiera de los párrafos anteriores del resumen. También se suministra un vector de expresión que comprende el ácido nucleico asilado vinculado operablemente con una secuencia regulatoria, tal como una célula huésped que comprende el vector de expresión. En algunas incorporaciones, se suministran
- 20 métodos para producir una variante de celulasa, que comprende cultivar las células huésped en un medio de cultivo bajo condiciones adecuadas para producir la variante de celulasa.
- [29] También se suministran composiciones que comprenden la variante de celulasa de cualquiera de los párrafos anteriores del resumen. En algunas incorporaciones
- 25 preferidas, las composiciones además comprenden por lo menos una enzima adicional seleccionada del grupo que consiste de una subtilisina, una metaloproteasa neutral, una lipasa, una cutinasa, una amilasa, una carbohidrasa, una pectinasa, una mananasa, un arabinasa, un galactanasa, una xilanas, una oxidasa, y una peroxidasa. La divulgación además suministra métodos para limpiar o cuidar telas que comprende poner en
- 30 contacto una superficie o un artículo que comprende una tela con la composición. Adicionalmente, se suministran métodos para limpiar y terminar la superficie de una tela que comprende poner en contacto una superficie o un artículo que comprende una tela con la composición.

- [30] En algunas incorporaciones preferidas, se suministran métodos para convertir biomasa en azúcares que comprende poner en contacto la biomasa con la variante de celulasa de cualquiera de los párrafos anteriores del resumen. En algunas incorporaciones preferidas se suministran métodos para producir un combustible que
- 5 comprende: poner en contacto una composición de biomasa con una composición enzimática que comprende la variante de celulasa de cualquiera de los párrafos anteriores para producir una solución de azúcar; y cultivar la solución de azúcar con un microorganismo de fermentación bajo condiciones suficientes para producir un combustible.
- 10 [31] Otros objetos, características y ventajas de la presente divulgación serán evidentes a partir de la siguiente descripción detallada. Se deberá entender, sin embargo, que la descripción detallada y los ejemplos específicos, mientras que se indican las incorporaciones preferidas de la divulgación, están dados como ilustración únicamente, dado que varios cambios y modificaciones dentro del alcance de la divulgación serán
- 15 evidentes para un experto en el estado de la técnica a partir de la lectura de esta descripción detallada.

IV. BREVE DESCRIPCIÓN DE LAS FIGURAS

- [32] La figura 1 suministra una alineación de las secuencias de aminoácidos de la forma
- 20 madura de carias celulasas: CBH2 *Hypocrea jecorina* (también conocida como *T. reesei*) (SEQ ID NO:3), CBH2 *Hypocrea koningii* (SEQ ID NO:4), CBH2 *Humicola insolens* (SEQ ID NO:5), CBH2 *Acremonium cellulolyticus* (SEQ ID NO:6), CBH2 *Agaricus bisporus* (SEQ ID NO:7), EG *Fusarium osysporum* (SEQ ID NO:8), CBH2 *Phanerochaete chrysosporium* (SEQ ID NO:9), CBH2 *Talaromyces emersonii* (SEQ ID
- 25 NO: 10), 6B/E3 CBH2 *Thermobifida fusca* (SEQ ID NO: 11), 6A/E2 EG *Thermobifida fusca* (SEQ ID NO: 12), y EG *Cellulomonas fimi* *CenA* (SEQ ID NO: 13).

- [33] La figura 2A suministra una gráfica de la frecuencia relativa de los ganadores del ensayo de rastreo de maíz pretratado (PCS) observado sobre el esperado a medida que un producto de carga cambia del CBH2 SEL. La reducción en carga de CBH2 resulta en
- 30 una frecuencia significativamente más alta de los ganadores del ensayo de PCS. La figura 2B suministra una gráfica de los ganadores de ensayo de celulosa hinchada de ácido fosfórico (PASC) observado sobre el esperado a medida que un producto de carga

cambia del CBH2 SEL. Se observó una frecuencia significativamente más alta de ganadores de PASC para variantes sin un cambio de carga.

[34] La figura 3 suministra un mapa plásmido para pTTTpyrG-cbh2.

[35] La figura 4 suministra un esquemático del módulo de eliminación *T. reesei* ku80.

- 5 [36] La figura 5 ilustra la inactivación del gen ku80 nativo como consecuencia de su reemplazo con un marcador seleccionable de als.

[37] La figura 6 suministra un esquemático del módulo de eliminación *T. reesei* pyr2.

[38] La figura 7 ilustra la inactivación del gen pyr2 nativo como consecuencia de su reemplazo con un marcador seleccionable de hygR y un fragmento amdS.

- 10 [39] La figura 8 suministra un esquemático del módulo de eliminación *T. reesei* hygR.

[40] La figura 9 ilustra la inactivación del gen hygR heterólogo como consecuencia de su reemplazo con un marcador seleccionable de pyr2.

[41] La figura 10 suministra un esquemático del módulo de eliminación *T. reesei* bgl1.

[42] La figura 11 suministra un esquemático del módulo de eliminación *T. reesei* egl3.

- 15 [43] La figura 12 suministra un esquemático del vector plásmido telomérico *T. reesei* utilizado para la expresión de recombinasa ere.

[44] La figura 13 ilustra la inactivación del marcador seleccionable pyr2 y la activación del marcador seleccionable amdS como consecuencia de introducción de un polinucleótido de un gen de interés (GOI). En incorporaciones ejemplares, el GOI

- 20 codifica una variante de CB H2.

[45] La figura 14 suministra un esquemático que ilustra la derivación de la cepa huésped MAD6, a partir de la cepa quad-eliminada.

V. DESCRIPCIÓN DETALLADA DE VARIAS INCORPORACIONES

- 25 [46] La presente divulgación se relaciona con enzimas y en particular con variantes de celulasa. También se describen ácidos nucleicos que codifican las variantes de celulasa, composiciones que comprenden las variantes de celulasa, métodos para identificar las variantes de celulasa útiles adicionales y métodos para utilizar las composiciones.

- [47] Se entiende que la anterior descripción general y la siguiente descripción detallada
30 son ejemplares y explicativas únicamente y no son restrictivas de las composiciones y los métodos acá descritos. A menos que se defina de otra forma acá, todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado como se entiende comúnmente por un experto en el estado de la técnica a quien pertenezca esta

divulgación. El uso de "o" significa "y/o" a menos que se indique lo contrario. Del mismo modo, los términos "comprende", "que comprende", "incluye" y "que incluye" no pretenden ser limitantes. Todas las patentes y publicaciones, incluyendo todas las secuencias de aminoácidos y nucleótidos divulgadas dentro de dichas patentes y publicaciones, referenciadas acá son incorporadas expresamente por referencia. Los encabezados suministrados acá no son limitaciones de los varios aspectos o incorporaciones de la divulgación, los cuales se pueden tener por referencia a la descripción completa.

[48] A menos que se defina acá lo contrario, todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado como se entienden comúnmente por un experto en el estado de la técnica al cual pertenezca esta divulgación. Singleton y otros, DICCIONARIO DE MICROBIOLOGIA Y BIOLOGIA MOLECULAR, 2D ED., John Wiley e Hijos, New York (1994), y Hale & Marham, EL DICCIONARIO HARPER COLLINS DE BIOLOGIA, Harper Perennial, NY (1991) suministran a un experto con un diccionario general de varios de los términos utilizados en esta divulgación. Aunque se pueden utilizar cualquier método y materiales similares o equivalentes a aquellos acá descritos en la práctica o la prueba de la presente divulgación, se describen los métodos y los materiales preferidos. Los rangos numéricos son inclusivos de los números que definen el rango. A menos que se indique lo contrario, los ácidos nucleicos están escritos de izquierda a derecha en orientación 5' a 3'; las secuencias de aminoácidos están escritas de izquierda a derecha en orientación amino a carboxil, respectivamente. Los expertos son dirigidos particularmente a Sambrook y otros, CLONACIÓN MOLECULAR: UN MANUAL DE LABORATORIO (Segunda Edición), Cold Spring Harbos Press, Plainview, NY, 1989, y Ausubel FM y otros, Protocolos Actuales en Biología Molecular, John Wiley e Hijos, New York, NY, 1993, para definiciones y términos del estado de la técnica. Se entenderá que esta divulgación no está limitada a la metodología, protocolos y reagentes particulares descritos, dado que estos pueden variar.

30 I. Definiciones

[49] Los términos a continuación son definidos más completamente por referencia a la descripción en su totalidad.

[50] El término "polipéptido" como se usa acá se refiere a un compuesto hecho de una cadena simple de residuos de aminoácidos vinculados por uniones de péptido. El término "proteína" como se utiliza acá puede ser sinónimo con el término "polipéptido".

- 5 [51] "Variante" significa una proteína la cual se deriva a partir de una proteína precursora (por ejemplo, la proteína nativa) por adición de uno o más aminoácidos a alguno o ambos del extremo de terminal C- y N-, sustituciones de uno o más aminoácidos en uno o un numero de diferentes sitios en la secuencia de aminoácido, o la eliminación de uno o más aminoácidos en alguno o ambos extremos de la proteína o
- 10 en uno o más sitios en la secuencia de aminoácido. La preparación de una variante de celulasa puede ser llevada a cabo por cualquier medio conocido en el estado de la técnica. En incorporaciones preferidas, una variante de celulasa se prepara al modificar una secuencia de ADN la cual codifica para la proteína nativa, transformación de la secuencia de ADN modificada en un huésped adecuado, y la expresión de la secuencia
- 15 de ADN modificada para formar la enzima de variante. La variante de celulasa de la divulgación incluye péptidos que comprenden secuencias alteradas de aminoácidos en comparación con una secuencia de aminoácido de enzima precursora en donde la variante de celulasa retiene la característica de naturaleza celulolítica de la enzima precursora pero la cual puede haber alterado las propiedades en algún aspecto
- 20 específico. Por ejemplo, una variante de celulasa puede tener un pH aumentado óptimo o temperatura aumentada o estabilidad oxidativa o afinidad reducida o unión a materiales no celulósicos pero retiene sus características de actividad celulolítica. Se contempla que las variantes de acuerdo con la presente divulgación se pueden derivar a partir de un fragmento de ADN que codifica una variante de celulasa en donde la
- 25 actividad de funcionalidad de la variante de celulasa expresada se retiene. Por ejemplo, un fragmento de ADN que codifica una celulasa puede además incluir una secuencia de ADN o porción del mismo que codifica una bisagra o enlazador unido a la secuencia de ADN de celulasa en cualquiera de los extremos 5' y 3' en donde la actividad funcional del dominio de celulasa codificado se retiene. Los términos variante y derivado pueden
- 30 ser utilizados intercambiabilmente en este documento.

[52] Los "residuos equivalentes" pueden también ser definidos al determinar la homología en el nivel de estructura terciaria para una celulasa precursora cuya estructura terciaria ha sido determinada por cristalografía de rayos x. Los residuos

- equivalentes se definen como aquellos para los cuales las coordenadas atómicas de dos o más de los átomos de cadena principal de un residuo de aminoácido particular de una celulasa y CBH2 *Hypocrea jecorina* (N sobre N, CA sobre CA, C sobre C, y O sobre O) se encuentran dentro de 0.13 nm y preferiblemente 0.1 nm después de la alineación. La
- 5 alineación se logra después de que el mejor modelo ha sido orientado y posicionado para dar la superposición máxima de coordenadas atómicas de átomos de proteínas no hidrógeno de la celulasa en cuestión al CB H2 *H. jecorina*. El mejor modelo es el modelo cristalográfico que da el factor R más bajo para los datos de difracción experimental en la resolución más alta disponible, ver por ejemplo US 2006/0205042.
- 10 [53] Los residuos equivalentes los cuales son funcionalmente análogos a un residuo específico de CB H2 *H. jecorina* se definen como aquellos aminoácidos de una celulasa la cual puede adoptar una conformación tal que ellos se alterna, modifican o contribuyen a la estructura de proteína, la unión de sustrato o la catálisis en una forma definida y atribuida a un residuo específico del CB H2 *H. jecorina*. Además, ellos son
- 15 aquellos residuos de la celulasa (para los cuales se ha obtenido una estructura terciaria por cristalografía de rayos x) la cual ocupa una posición análoga en la medida en que, aunque los átomos de cadena principal del residuo dado pueden no satisfacer el criterio de equivalencia con base en ocupar una posición homóloga, las coordenadas atómicas de por lo menos dos de los átomos de cadena lateral del residuo se encuentran con 0.13
- 20 nm de los átomos de cadena lateral correspondientes de CBH2 *H. jecorina*. La estructura de cristal del CB H2 *H. jecorina* se en Zou y otros (1999) (Ref. 5, supra).
- [54] El término "molécula de ácido nucleico" incluye moléculas de ARN, ADN y cADN. Se entenderá que, como resultado de la degeneración del código genético, se pueden producir una multitud de secuencias de nucleótido que codifican una proteína
- 25 dada tal como CBH2 y/o variantes de la misma. La presente divulgación contempla todas las posibilidades de secuencia variante de nucleótido, que codifican variante de celulasa tal como CB H2, todas las cuales son posiblemente dadas por la degeneración del código genético.
- [55] Una construcción o secuencia de ácido nucleico "heterólogo" tiene una porción de
- 30 la secuencia la cual no es nativa a la célula en la cual se expresa. Heterólogo, con respecto a una secuencia de control se refiere a una secuencia de control (es decir, promotor o mejorador) que no funciona en la naturaleza para regular el mismo gen la expresión de la cual actualmente se regula. Generalmente, las secuencias heterólogas de

ácido nucleico no son endógenas a la célula o parte del genoma en el cual ellas están presenten, y han sido adicionadas a la célula, por infección, transfección, transformación, microinyección, electroporación, o similar. Una construcción de ácido nucleico "heterólogo" puede contener una secuencia o ADN de control que codifica la

5 combinación de secuencia encontrado en la célula nativa.

[56] Como se utiliza acá, el término "vector" se refiere a una construcción de ácido nucleico diseñada para transferirse entre las diferentes células huésped. Un "vector de expresión" se refiere a un vector que tiene la habilidad de incorporar y expresar fragmentos heterólogos de ADN en una célula extraña. La selección de los vectores de
10 expresión apropiados se encuentra dentro del conocimiento de aquellos expertos en el estado de la técnica.

[57] De acuerdo con esto, un "módulo de expresión" o "vector de expresión" es una construcción de ácido nucleico generada de forma recombinante o sintética, con una serie de elementos específicos de ácido nucleico que permiten la transcripción de un
15 ácido nucleico particular en una célula objetivo. El módulo de expresión recombinante se puede incorporar en un plásmido, cromosoma, ADN de mitocondria, ADN plastidio, virus, o fragmento de ácido nucleico. Típicamente, la porción de módulo de expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico para ser transcrito y un promotor.

20 [58] Como se utiliza acá, el término "plásmido" se refiere a una construcción de ADN circular de doble cadena (ds) utilizado como un vector de clonación, y el cual forma un elemento genético extracromosómico de auto-replicación en varias bacterias y algunos eucariotas.

[59] Como se utiliza acá, el término "secuencia de nucleótido que codifica el marcador
25 seleccionable" se refiere a una secuencia de nucleótido la cual es capaz de la expresión en células y donde la expresión del marcador seleccionable confiere a las células que contienen el gen expresado la habilidad de crecer en la presencia de un agente selectivo correspondiente, o bajo las condiciones de crecimiento selectivas correspondientes.

[60] Como se usa aquí, el término "promotor" se refiere a una secuencia de ácido
30 nucleico que funciona para dirigir la transcripción de un gen hacia abajo. El promotor generalmente será apropiado a la célula huésped en la que se expresa el gen objetivo. El promotor junto con otras secuencias de ácido nucleico de transcripción y regulatorias traslacionales (también llamado "secuencias de control") son necesarios para expresar

un determinado gen. En general, las secuencias de aminoácidos reguladoras de transcripción y traslación incluyen, pero no se limitan a, secuencias del promotor, sitios de unión ribosomal, secuencias de inicio y detención de transcripción, secuencias de inicio y detención traslacional, y secuencias potenciadoras o activadoras.

- 5 [61] "Gen quimérico" o "construcción de ácido nucleico heterólogo", como se define acá, se refiere a un gen no nativo (es decir, uno que ha sido introducido en un huésped) que puede estar compuesto de partes de diferentes genes, incluyendo elementos regulatorios. Una construcción de gen quimérico para transformación de una célula huésped está típicamente compuesta de una región regulatoria de transcripción
- 10 (promotor) vinculado operablemente con una secuencia que codifica proteína heteróloga, o, en un gen quimérico marcador seleccionable, a un gen marcador seleccionable que codifica una proteína que confiere, por ejemplo, resistencia antibiótica a células transformadas. Un gen quimérico típico de la presente divulgación, para transformación en una célula huésped, incluye una región regulatoria de
- 15 transcripción que es constitutiva o inducible, una secuencia que codifica proteína, y una secuencia terminadora. Una construcción de gen quimérico puede también incluir una segunda secuencia de ADN que codifica un péptido de señal si se desea la secreción de la proteína objetivo.

- [62] Un ácido nucleico es "vinculado operativamente" cuando es ubicado en una
- 20 relación funcional con otra secuencia de ácido nucleico. Por ejemplo, el ADN que codifica un líder secretor es vinculado operativamente con el ADN para un polipéptido si éste se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o mejorador es vinculado operativamente con una secuencia codificadora si ésta afecta la transcripción de la secuencia; o un sitio de unión de ribosoma es vinculado
- 25 operativamente con una secuencia codificadora si éste es ubicado con el fin de facilitar la traslación. Generalmente, "vinculado operativamente" significa que las secuencias de ADN que son vinculadas están contiguas, y, en el caso de un líder secretor, contiguo y en marco de lectura. Sin embargo, los mejoradores no tienen que estar contiguos. La vinculación se realiza por ligación en los sitios de restricción convenientes. Si dichos
- 30 sitios no existen, los adaptadores, vinculadores o cebadores sintéticos de oligonucleótido para PCR son utilizados de acuerdo con la práctica convencional.

[63] Como se usa acá, el término "gen" significa el segmento de ADN involucrado en la producción de una cadena de polipéptido, que puede o no incluir regiones que preceden

y siguientes a la región de codificación, por ejemplo, secuencias sin trasladar 5' (5' UTR) o "líder" y secuencias 3' UTR o "posteriores", así como las secuencias que intervienen (intrones) entre los segmentos codificadores individuales (exones).

[64] En general, las moléculas de ácido nucleico que codifican la celulasa variante tal como CBH2 se hibridarán, bajo condiciones de rigor moderadas a altas a la secuencia de tipo salvaje tal como se suministra acá como SEQ ID NO:1. Sin embargo, en algunos casos una secuencia de nucleótido que codifica CBH2 se emplea que posee un uso de codones sustancialmente diferente, mientras que la proteína codificada por la secuencia de nucleótido que codifica CBH2 tiene la misma o sustancialmente la misma secuencia de aminoácido que la proteína nativa. Por ejemplo, la secuencia codificadora puede ser modificada para facilitar la expresión más rápida de CBH2 en un sistema de expresión procariótico o eucariótico, de acuerdo con la frecuencia con la cual se utiliza un codón particular por el huésped (Te'o y otros, Cartas de Microbiología FEMS, 190: 13 a 19, 2000, por ejemplo describe la optimización de genes para expresión en hongos filamentosos).

[65] Una secuencia de ácido nucleico se considera que es "hibrizable selectivamente" con una secuencia de ácido nucleico de referencia si las dos secuencias específicamente se hibridan con otra bajo condiciones de hibridación y lavado de rigor moderado a alto. Las condiciones de hibridación se basan en la temperatura de fusión (T_m) del complejo o sonda de unión de ácido nucleico. Por ejemplo, el "rigor máximo" típicamente se presenta a aproximadamente $T_m - 5^\circ\text{C}$ (5°C por debajo de la T_m de la sonda); el "rigor alto" a aproximadamente 5 a 10°C por debajo de la T_m ; "rigor moderado" o "intermedio" a aproximadamente 10 a 20°C por debajo de la T_m de la sonda; y "rigor bajo" a aproximadamente 20 a 25°C por debajo de la T_m . Funcionalmente, las condiciones de rigor máximo pueden ser utilizadas para identificar secuencias que tienen identidad estricta o identidad casi estricta con la sonda de hibridación; mientras que las condiciones de rigor alto son utilizadas para identificar las secuencias que tienen aproximadamente 80% o más identidad de secuencia con la sonda.

[66] Las condiciones de hibridación de rigor moderado y alto son bien conocidas en el estado de la técnica (ver, por ejemplo, Sambrook y otros, 1989, Capítulos 9 y 11, y en Ausubel F. M. y otros, 1993, incorporados expresamente acá por referencia). Un ejemplo de condiciones de rigor alto incluye la hibridación a aproximadamente 42°C en 50% de formamida, 5xSSC, solución de Denhardt de 5 veces, 0.5% de SDS y 100 ug/ml

de ADN portador desnaturalizado seguido por el lavado dos veces en 2 veces SSC y 0.5% SDS a temperatura ambiente y dos veces adicionales en 0.1x SSC y 0.5% SDS a 42°C.

[67] El término “recombinante” cuando se utiliza con la referencia, es decir, a una
 5 célula o ácido nucleico, proteína, o vector, indica que la célula, ácido nucleico, proteína
 o vector, ha sido modificado por la introducción de un ácido nucleico heterólogo o
 proteína o la alteración de un ácido nucleico nativo o proteína, o que la célula se deriva
 de una célula modificada de esta forma. Así, por ejemplo, las células recombinantes
 expresan genes que no se encuentran dentro de la forma nativa (no recombinante) de la
 10 célula o expresa genes nativos que son expresados de forma anormal, sub-expresados o
 no expresados del todo.

[68] Como se utiliza acá, los términos “transformado”, “establemente transformado” o
 “transgénico” con referencia a una célula significa que la célula tiene una secuencia de
 ácido nucleico no nativo (heterólogo) integrada con su genoma o como un plásmido
 15 episomal que se mantiene a través de múltiples generaciones.

[69] Como se utiliza acá, el término “expresión” se refiere al proceso por medio del cual
 se produce un polipéptido con base en la secuencia de ácido nucleico de un gen. El
 proceso incluye tanto la transcripción como la traslación.

[70] El término “introducido” en el contexto de insertar una secuencia de ácido nucleico
 20 en una célula, significa “transfección”, o “transformación” o “transducción” e incluye la
 referencia a la incorporación de una secuencia de ácido nucleico en una célula
 eucariótica o procariótica donde se puede incorporar la secuencia de ácido nucleico en
 el genoma de la célula (por ejemplo, ADN de cromosoma, plásmido, plástido, o
 mitocondrial), convertido en un replicón autónomo, o expresado transitoriamente (por
 25 ejemplo, mRNA transfectado).

[71] Sigue que el término “expresión de CBH2” se refiere a la transcripción y traslación
 del gen *cbh2* o variantes del mismo, los productos que incluyen ARN precursor, mRNA,
 polipéptido, polipéptidos procesados pos-traslacionalmente, y derivados de los mismos,
 incluyendo CBH2 de especies relacionadas tal como *Trichoderma koningii*, *Hypocrea*
 30 *jecorina* (también conocida como *Trichoderma longibrachiatum*, *Trichoderma reesei* o
Trichoderma viride) y *Hypocrea schweinitzii*. A modo de ejemplo, los ensayos para la
 expresión de CBH2 incluyen mancha Occidental para la proteína CBH2, análisis de
 mancha Norte y ensayos de reacción de cadena de polimerasa transcriptasa (RT-PCR)

para mRNA *cbh2*, y Celulosa Hinchada de Ácido Fosfórico (PASC) y ensayos de hidrazida de ácido p-hidroxibenzoico (PAHBAH) como se describen los siguientes: (a) PASC: (Karlsson, J. y otros (2001), Eur. J. Biochem, 268, 6498 a 6507, Wood, T. (1988) en Métodos en Enzimología, Vol. 160. Parte de Biomasa una Celulosa y Hemicelulosa
 5 (Wood, W. & Kellog, S. Eds.), pp.19 a 25, Prensa Académica, San Diego, CA, USA) y (b) PAHBAH: (Lever, M. (1972) Bioquímica Analítica, 47, 273, Blakeney, A. B. & Mutton, L. L. (1980) Diario de Ciencia de Comida y Agricultura, 31, 889, Henry, R. J. (1984) Diario del Instituto de Fabricación de Cerveza, 90, 37).

[72] El término "empalme alternativo" se refiere al proceso mediante el cual se generar
 10 múltiples isoformas de polipéptido a partir de un gen simple, e involucra el empalme de exones no consecutivos durante el procesamiento de algunas, pero no todas, las transcripciones del gen. Así, un exón particular se puede conectar con cualquiera de varios exones alternativos para formar ARN mensajeros. Los ARN alternativamente empalmados producen polipéptidos ("variantes de empalme") en los cuales algunas
 15 partes son comunes mientras que otras partes son diferentes.

[73] El término "secuencia de señal" se refiere a una secuencia de aminoácidos en la porción de terminal N de una proteína que facilita la secreción de la forma madura de la proteína por fuera de la célula. La forma madura de la proteína extracelular carece de la secuencia de señal que se escinde durante el proceso de secreción.

[74] El término "célula huésped" significa una célula que contiene un vector y soporta la replicación, y/o transcripción o transcripción y traslación (expresión) de la construcción de expresión. Las células huésped para su uso en la presente divulgación pueden ser células procarióticas, tal como *E. coli*, o células eucarióticas tal como células de levadura, planta, insecto, anfibio, o mamífero. En general, las células huésped son
 25 hongos filamentosos.

[75] El término "hongos filamentosos" significa cualquier y todos los hongos filamentosos reconocidos por los expertos en el estado de la técnica. Un hongo preferido se selecciona a partir del grupo que consiste de *Aspergillus*, *Trichoderma*, *Fusarium*, *Chrysosporium*, *Penicillium*, *Humicola*, *Neurospora*, o formas sexuales
 30 alternativas de los mismos tal como *Emericella*, *Hypocrea*. Se ha demostrado ahora que el hongo industrial asexual *Trichoderma reesei* es un clon derivado del ascomiceto *Hypocrea jecorina* (Ver, Kuhls y otros, PNAS, 93:7755 a 7760, 1996).

[76] El término “celooligosacárido” se refiere a grupos de oligosacáridos que contienen de 2 a 8 unidades de glucosa y que tienen uniones beta-1.4, por ejemplo, celobiosa.

[77] Los términos “celulasa”, “enzimas celulolíticas” o “enzimas de celulasa” se refieren a una categoría de enzimas capaces de hidrolizar polímeros de celulasa en oligómeros celooligosacáridos más cortos, celobiosa y/o glucosa. Varios ejemplos de celulasas, tal como exoglucanasas, exocelobiohidrolasas, endoglucanasas, y glucósidos se han obtenido de organismos celulolíticos, particularmente que incluyen hongos, plantas y bacterias. Las enzimas hechas por estos microbios son mezclas de proteínas con tres tipos de acciones útiles en la conversión de celulosa en glucosa:

5 endoglucanasas (EG), celobiohidrolasas (CBH), y beta-glucósidos. Estos tres tipos diferentes de enzimas de celulasa actúan sinérgicamente para convertir celulosa y sus derivados en glucosa.

[78] Varios microbios hacen enzimas que hidrolizan celulosa, incluyendo el hongo de madera podrida *Trichoderma*, la bacteria de compost *Thermomonospora*, *Bacillus*, y

15 *Cellulomonas*, *Streptomyces*; y el hongo *Humicola*, *Espergillus* y *Fusarium*.

[79] El CBH2 de *Hypocrea jecorina* es un miembro de la Familia de Hidrolasa Glicosil 6 (por lo tanto Cel6) y, específicamente, fue el primer miembro de esa familia identificado en *Hypocrea jecorina* (por lo tanto Cel6A). La Familia de Hidrolasa Glicosil 6 contiene tanto Endoglucanasas como Celobiohidrolasas/exoglucanasas, y ese

20 CBH2 es el último. Así, las frases CBH2, proteína de tipo CBH2 y celobiohidrolasas Cel6 pueden ser utilizadas acá de forma intercambiable.

[80] El término “dominio de unión de celulosa” como se utiliza acá se refiere a la porción de la secuencia de aminoácido de una celulasa o una región de la enzima que está involucrada en la actividad de unión de celulosa de una celulasa o derivado de la

25 misma. Los dominios de unión de celulosa generalmente funcionan al unir de forma no covalente la celulasa con celulosa, un derivado de celulosa u otro polisacárido equivalente de los mismos. Los dominios de unión de celulosa permiten o facilitan la hidrólisis de las fibras de celulosa por la región de núcleo catalítica estructuralmente distinta, y típicamente funcionan independiente del núcleo catalítico. Así, un dominio de

30 unión de celulosa no poseerá la actividad hidrolítica significativa atribuible a un núcleo catalítico. En otras palabras, un dominio de unión de celulosa es un elemento estructural de la estructura terciaria de proteína de enzima celulasa que es distinto del elemento

estructural que posee actividad catalítica. El dominio de unión de celulosa y el módulo de unión de celulosa pueden ser utilizados acá de forma intercambiable.

[81] Como se usa acá, el término “surfactante” se refiere a cualquier compuesto generalmente reconocido en el estado de la técnica como que tiene cualidades de superficie activa. Así, por ejemplo, los surfactantes comprenden surfactantes aniónicos, catiónicos y no iónicos tales como aquellos comúnmente encontrados en detergentes. Los surfactantes aniónicos incluyen alquilbenzenosulfonatos lineales o ramificados; sulfatos de alquil o alquenil éter que tienen grupos alquilo lineales o ramificados o grupos alquenil; sulfatos de alquil o alquenil; olefinsulfonatos; y alcanosulfonatos. Los surfactantes anfóliticos incluyen sulfonatos de sal de amonio cuaternarios, y surfactantes anfóliticos de tipo betaína. Dichos surfactantes anfóliticos tienen los grupos cargados positivos y negativos en la misma molécula. Los surfactantes no iónicos pueden comprender éteres de polioxialquilenos, así como alcanolamidas de ácido graso mayor u óxido de alquilenos aducto del mismo, monoésteres de glicerina de ácido graso, y similares.

[82] Como se usa acá, el término “tela que contiene celulosa” se refiere a cualquier tela cosida o sin coser, hilos o fibras hechas de celulosa que contiene algodón o no o mezclas de celulosa que contiene o no algodón que incluye celulósicas naturales y celulósicas artificiales (tal como yute, lino, ramio, rayón, y lyocell).

[83] Como se usa acá, el término “tela que contiene algodón” se refiere a telas cosidas o no cosidas, hilos o fibras hechas de algodón puro o mezclas de algodón que incluyen telas tejidas de algodón, tejidos de algodón, pantalones de algodón, hilos de algodón, algodón en bruto y similares.

[84] Como se usa acá, el término “composición de lavado de piedra” se refiere a una formación para su uso en telas que contienen celulosa de lavado de piedra. Las composiciones de lavado de piedra son utilizadas para modificar las telas que contienen celulosa antes de su venta, es decir, durante el proceso de fabricación. En contraste, las composiciones de detergente están destinadas para la limpieza de telas sucias y no son utilizadas durante el proceso de fabricación.

[85] Como se usa acá, el término “composición detergente” se refiere a una mezcla que está destinada para su uso en un medio de lavado para el lavado de telas sucias que contienen celulosa. En el contexto de la presente divulgación, dichas composiciones pueden incluir, adicional a las celulasas y surfactantes, enzimas hidrolíticas adicionales,

constructores, agentes blanqueadores, agentes azulantes y colorantes fluorescentes, inhibidores de aglomeración, agentes enmascarantes, activadores de celulasa, antioxidantes, y solubilizantes.

5 [86] Como se usa acá, el término “reducción o eliminación en la expresión del gen cbh2” significa que el gen cbh2 ha sido eliminado del genoma y por lo tanto no puede ser expresado por el microorganismo huésped recombinante; o que el gen cbh2 o transcripto ha sido modificado de modo que una enzima CBH2 funcional no se produce por el microorganismo huésped o en niveles que son significativamente menores que el gen cbh2 no modificado o transcrito.

10 [87] El término “gen cbh2 variante” significa que la secuencia de ácido nucleico del gen cbh2 de *H. jecorina* ha sido alterada al remover, adicionar, y/o manipular la secuencia de codificación.

15 [88] Como se usa acá, el término “purificador” generalmente se refiere a ácido nucleico transcrito sujetador o proteína que contiene células para purificación bioquímica y/o cromatografía de columna.

[89] Como se usa acá, los términos “activo” y “biológicamente activo” se refieren a una actividad biológica asociada con una proteína particular y son usados acá de forma intercambiable. Por ejemplo, la actividad enzimática asociada con una proteasa es proteólisis y, así, una proteasa activa tiene actividad proteolítica. Sigue que la actividad
20 biológica de una proteína dada se refiere a cualquier actividad biológica típicamente atribuida a esa proteína por aquellos expertos en el estado de la técnica.

[90] Como se usa acá, el término “enriquecido” significa que la celulasa tal como CBH2 se encuentra en una concentración que es mayor con relación a la concentración de CBH2 encontrada en una composición de celulasa de hongo de tipo salvaje, o que
25 ocurre naturalmente. Los términos enriquecido, elevado y mejorado pueden ser utilizados acá de forma intercambiable.

[91] Una composición de celulasa de hongo de tipo salvaje es una producida por una fuente de hongo que ocurre naturalmente y la cual comprende uno o más componentes de BGL, CBH y EG en donde cada uno de estos componentes se encuentra en la
30 proporción producida por la fuente de hongo. Así, una composición de CBH enriquecido tendría CBH a una proporción alterada en donde la proporción de CBH a otros componentes de celulasa (es decir, EG, betaglucósidos y otras endoglucanasas) es elevada. Esta proporción se puede aumentar por el aumento en CBH o la reducción (o

eliminación) de por lo menos otro componente por cualquier medio conocido en el estado de la técnica.

[92] El término "aislado" o "purificado" como se usa acá se refiere a un ácido nucleico o un aminoácido que es removido de por lo menos un componente con el cual está asociado naturalmente.

[93] Así, para ilustrar, un sistema de celulasa que ocurre naturalmente puede ser purificado en componentes sustancialmente puros por medio de técnicas reconocidas de separación publicadas en la literatura, incluyendo cromatografía de intercambio a un pH adecuado, cromatografía de afinidad, exclusión de tamaño y similares. Por ejemplo, en la cromatografía de intercambio de iones (usualmente cromatografía de intercambio de aniones), es posible separar los componentes de celulasa al eludir con un gradiente de pH, o un gradiente de sal, o ambos un gradiente de pH y de sal. El CBH purificado puede entonces ser adicionado a la solución enzimática que resulta en una solución CBH enriquecida. También es posible elevar la cantidad de CBH producido por un microbio utilizando métodos de genética molecular para sobreexpresar el CBH que codifica el gen, posiblemente en conjunto con la eliminación de uno o más genes que codifican otras celulasas.

[94] Las celulasas de hongos pueden contener más de un componente de CBH. Los diferentes componentes generalmente tienen diferentes puntos isoeléctricos que permiten su separación por medio de cromatografía de intercambio de iones y similares. Cualquiera de un componente CBH simple o una combinación de componentes CBH se pueden emplear en una solución enzimática.

[95] Cuando se emplea en soluciones enzimáticas, el componente CBH2 homólogo o variante es generalmente adicionado en una cantidad suficiente para permitir la tasa más alta de liberación de azúcares solubles a partir de la biomasa. La cantidad de componente CBH2 homólogo o variante adicionado depende, con el tipo de biomasa a ser sacarificada, lo cual puede ser determinado fácilmente por el experto en la materia cuando se emplea, el porcentaje de peso del componente CBH2 homólogo o variante presente en la composición de celulasa está preferiblemente entre 1 y 100 con ejemplos ilustrativos que son aproximadamente 1, preferiblemente aproximadamente 5, preferiblemente aproximadamente 10, preferiblemente aproximadamente 15, o preferiblemente aproximadamente 20 por ciento en peso hasta preferiblemente aproximadamente 25, preferiblemente aproximadamente 30, preferiblemente

aproximadamente 35, preferiblemente aproximadamente 40, preferiblemente aproximadamente 45 o preferiblemente aproximadamente 50 por ciento en peso. Adicionalmente, los rangos preferidos pueden ser aproximadamente 0.5 a aproximadamente 15 por ciento en peso, aproximadamente 0.5 a aproximadamente 20 por ciento en peso, de aproximadamente 1 a aproximadamente 10 por ciento en peso, de aproximadamente 1 a aproximadamente 15 por ciento en peso, de aproximadamente 1 a aproximadamente 20 por ciento en peso, de aproximadamente 1 a aproximadamente 25 por ciento en peso, de aproximadamente 5 a aproximadamente 20 por ciento en peso, de aproximadamente 5 a aproximadamente 25 por ciento en peso, de aproximadamente 5 a aproximadamente 30 por ciento en peso, de aproximadamente 5 a aproximadamente 35 por ciento en peso, de aproximadamente 5 a aproximadamente 40 por ciento en peso, de aproximadamente 5 a aproximadamente 45 por ciento en peso, de aproximadamente 5 a aproximadamente 50 por ciento en peso, de aproximadamente 10 a aproximadamente 20 por ciento en peso, de aproximadamente 10 a aproximadamente 25 por ciento en peso, de aproximadamente 10 a aproximadamente 30 por ciento en peso, de aproximadamente 10 a aproximadamente 35 por ciento en peso, de aproximadamente 10 a aproximadamente 40 por ciento en peso, de aproximadamente 10 a aproximadamente 45 por ciento en peso, de aproximadamente 10 a aproximadamente 50 por ciento en peso, de aproximadamente 15 a aproximadamente 60 por ciento en peso, de aproximadamente 15 a aproximadamente 65 por ciento en peso, de aproximadamente 15 a aproximadamente 70 por ciento en peso, de aproximadamente 15 a aproximadamente 75 por ciento en peso, de aproximadamente 15 a aproximadamente 80 por ciento en peso, de aproximadamente 15 a aproximadamente 85 por ciento en peso, de aproximadamente 15 a aproximadamente 95 por ciento en peso. Sin embargo, cuando se emplea, el porcentaje en peso del componente CBH2 homólogo o variante con relaciona a cualquier componente de tipo EG presente en la composición de celulasa es de preferiblemente aproximadamente 1, preferiblemente aproximadamente 5, preferiblemente aproximadamente 10, preferiblemente aproximadamente 15, o preferiblemente aproximadamente 20 por ciento en peso hasta preferiblemente aproximadamente 25, preferiblemente aproximadamente 30, preferiblemente aproximadamente 35, preferiblemente aproximadamente 40, preferiblemente aproximadamente 45 o preferiblemente aproximadamente 50 por ciento en peso. Adicionalmente, los rangos preferidos pueden ser aproximadamente 0.5 a

- aproximadamente 15 por ciento en peso, aproximadamente 0.5 a aproximadamente 20 por ciento en peso, de aproximadamente 1 a aproximadamente 10 por ciento en peso, de aproximadamente 1 a aproximadamente 15 por ciento en peso, de aproximadamente 1 a aproximadamente 20 por ciento en peso, de aproximadamente 1 a aproximadamente 25 por ciento en peso, de aproximadamente 5 a aproximadamente 20 por ciento en peso, de aproximadamente 5 a aproximadamente 25 por ciento en peso, de aproximadamente 5 a aproximadamente 30 por ciento en peso, de aproximadamente 5 a aproximadamente 35 por ciento en peso, de aproximadamente 5 a aproximadamente 40 por ciento en peso, de aproximadamente 5 a aproximadamente 45 por ciento en peso, de aproximadamente 5 a aproximadamente 50 por ciento en peso, de aproximadamente 10 a aproximadamente 20 por ciento en peso, de aproximadamente 10 a aproximadamente 25 por ciento en peso, de aproximadamente 10 a aproximadamente 30 por ciento en peso, de aproximadamente 10 a aproximadamente 35 por ciento en peso, de aproximadamente 10 a aproximadamente 40 por ciento en peso, de aproximadamente 10 a aproximadamente 45 por ciento en peso, de aproximadamente 10 a aproximadamente 50 por ciento en peso, de aproximadamente 15 a aproximadamente 20 por ciento en peso, de aproximadamente 15 a aproximadamente 25 por ciento en peso, de aproximadamente 15 a aproximadamente 30 por ciento en peso, de aproximadamente 15 a aproximadamente 35 por ciento en peso, de aproximadamente 15 a aproximadamente 40 por ciento en peso, de aproximadamente 15 a aproximadamente 45 por ciento en peso, de aproximadamente 15 a aproximadamente 50 por ciento en peso.

II. Celulasas

- [96] Las celulasas son conocidas en el estado de la técnica como enzimas que hidrolizan celulosa (vínculos beta-1,4-glucano o beta D-glucosídico) que resultan en la formación de glucosa, celobiosa, celooligosacáridos, y similares. Como se indicó anteriormente, las celulasas se han dividido tradicionalmente en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) ("CBH") y beta-glucósidos (EC 3.2.1.21) ("BG").
- [97] Ciertos hongos producen sistemas completos de celulasa los cuales incluyen exocelobiohidrolasas o celulasa de tipo CBH, endoglucanasas o celulasas tipo EG y beta-glucósidos o celulasas tipo BG. Sin embargo, algunas veces estos sistemas carecen de

- celulasas tipo CBH y las celulasas de bacterias típicamente también incluyen poco o ninguna celulasa tipo CBH. Adicionalmente, se ha mostrado que los componentes EG y los componentes CBH interactúan sinérgicamente para degradar celulosa más eficientemente. Los diferentes componentes, es decir, las varias endogluconasas y
- 5 exocelobiohidrolasas en un componente múltiple o sistema de celulasa completo, generalmente tienen diferentes propiedades, tal como punto isoelectrico, peso molecular, grado de glicosilación, especificidad de sustrato y patrones de acción enzimática.
- [98] Se cree que las celulasas de tipo endoglucanasa hidrolizan las uniones glucosídicas
- 10 internas beta-1,4- en regiones de baja cristalinidad de la celulosa y las celulasas de tipo exocelobiohidrolasa hidrolizan celobiosa a partir del extremo reductor o no reductor de celulosa. Sigue que la acción de los componentes de endoglucanasa puede facilitar ampliamente la acción de las exocelobiohidrolasas al crear nuevos extremos de cadena los cuales son reconocidos por componentes de exo-celobiohidrolasa. Además, las
- 15 celulasas tipo beta-glucósido se han mostrado para catalizar la hidrólisis de beta-D-glucósidos alquilo y/o arilo, tal como beta-D-glucósido metilo y glucósido p-nitrofenil así como glucósidos que contienen únicamente residuos de carbohidrato, tal como celobiosa. Esto produce glucosa como el producto solo para el microorganismo y reduce o elimina la celobiosa que inhibe celobiohidrolasas y endoglucanasas.
- 20 [99] Las celulasas también encuentran un número de usos en composiciones detergentes incluyendo para mejorar la habilidad de limpieza, como un agente suavizante y para mejorar la textura de telas de algodón (Hemmpel, ITB Tintura/Impresión/Terminado 3:5 a 14, 1991; Tyndall, Química de Textiles y Colores 24:23 a 26, 1992; y Kumar y otros, Química Textil y Colores, 29:37 a 42, 1997). Dado que el mecanismo no es parte de la
- 25 divulgación, las propiedades suavizantes y de restauración de color de la celulasa han sido atribuidas a los componentes de endoglucanasa alcalina en composiciones de celulasa como se ejemplifican por las Patentes de los Estados Unidos No. 5.648.263, 5.691.178, y 5.776.757, las cuales divulgan que las composiciones detergentes que contienen una composición de celulasa enriquecida en un componente específico de
- 30 endoglucanasa alcalina imparten restauración de color y suavizante mejorado a prendas tratadas en comparación con composiciones de celulasa no enriquecidas en dicho componente. Adicionalmente, el uso de dichos componentes de endoglucanasa alcalina en composiciones detergentes se ha mostrado para complementar los requerimientos de

pH de la composición detergente (por ejemplo, al exhibir actividad máxima en un pH alcalino de 7.5 a 10, como se describe en la Patente de los Estados Unidos No. 5.648.263, 5.691.178, y 5.776.757).

[100] Las composiciones de celulasa también se han mostrado para degradar telas que contienen algodón, lo que resulta en pérdida de fuerza reducida en la tela (Patente de los Estados Unidos No. 4.822.516), lo que contribuye a la renuencia para usar composiciones de celulasa en aplicaciones comerciales de detergente. Las composiciones de celulasa que comprenden componentes de endoglucanasa han sido sugeridas para exhibir pérdida de fuerza reducida para telas que contienen algodón en comparación con composiciones que comprenden un sistema de celulasa completo.

[101] Las celulasa también han sido mostradas para ser útiles en la degradación de biomasa de celulasa en etanol (en donde la celulasa degrada celulosa en glucosa y levadura y otros microbios que además fermentan la glucosa en etanol), en el tratamiento de pulpa mecánica (Pere y otros, In Proco. Tappi Pulping Conf. Nashville, TN, 27 a 31, pp. 693 a 696, 1996), para su uso como un aditivo de alimentación (WO 91/04673) y en la molienda de grano húmedo.

[102] La mayoría de los CBH y EG tienen una estructura multidominio que consiste de un dominio de núcleo separado de un dominio de unión de celulosa (CBD) por un péptido vinculador (Suurnakki y otros, 2000). El dominio de núcleo contiene el sitio activo mediante el cual el CBD interactúa con la celulosa al unir la enzima a éste (van Tilbeurgh y otros, FEBS Lett. 204:23 a 227, 1986; Tomme y otros, Eur. J. Biochem. 170:575 a 581, 1988). Los CBD son particularmente importantes en la hidrólisis de celulosa cristalina. Se ha mostrado que la habilidad de las celobiohidrolasas para degradar celulosa cristalina claramente se reduce cuando el CBD está ausente (Linder y Teeri, J. Biotechnol. 57:15 a 28, 1997). Sin embargo, el papel exacto y el mecanismo de acción de los CBD todavía es una materia de especulación. Se ha sugerido que el CBD mejora la actividad enzimática simplemente al aumentar la concentración efectiva de enzima en la superficie de celulosa (Stahlberg y otros, Bio/Technol. 9:286 a 290, 1991), y/o al soltar las cadenas de celulosa simple de la superficie de celulosa (Torino y otros, EMBO J. vol. 15, no. 21, pp. 5739 a 5751, 1996). La mayoría de los estudios sobre los efectos de los dominios de celulasa en diferentes sustratos se han llevado a cabo con proteínas de núcleo de celobiohidrolasas, a medida que sus proteínas de núcleo pueden ser fácilmente producidas por proteólisis limitada con papaína (Tomme y otros, 1988).

- Varias celulasas se han descrito en la literatura científica, cuyos ejemplos incluyen: desde *Trichoderma reesei*: Shoemaker, S. y otros, Bio/Tecnología, 1:691 a 696, 1983, la cual divulga CBH1; Teeri, T. y otros, Gen, 51:43 a 52, 1987, la cual divulga CBH2. Las celulasas de especies diferentes a *Trichoderma* también han sido descritas por ejemplo,
- 5 Ooi y otros, Investigación de Ácidos Nucleicos, vol. 18, no. 19, 1990, la cual divulga la secuencia de cADN que codifica para la endogluconasa F1-CMC producida por *Aspergillus aculeatus*; Kawaguchi T y otros, Gen 173(2):287 a 8, 1996, la cual divulga la clonación y la secuencia del cADN que codifica beta-glucósido 1 de *Aspergillus aculeatus*; Sakamoto y otros, Curr. Genet. 27:435 a 439, 1995, la cual divulga la
 - 10 secuencia de cADN que codifica la endoglucanasa CMCase-1 de *Aspergillus kawachii* IFO 4308; Saarilahti y otros, Gen 90:9 a 14, 1990, la cual divulga una endoglucanasa de *Erwinia carotovora*; Spilliaert R. y otros, Eur. J. Biochem. 24(3):923 a 30, 1994, la cual divulga la clonación y secuencia de bglA, que codifica para una beta-glucanasa termoestable de *Rhodothermus marinus*; y Halldorsdottir S y otros, Appl. Microbiol.
 - 15 Biotechnol. 49(3):277 a 84, 1998, la cual divulga la clonación y secuencia y sobreexpresión de un gen de *Rhodothermus marinus* que codifica una celulasa termoestable de glicosil hidrolasa familia 12. Sin embargo, aún existe una necesidad por identificación y caracterización de celulasas novedosas, con propiedades mejoradas, tal como desempeño mejorado bajo condiciones de estrés térmico o en la presencia de
 - 20 surfactantes, actividad específica aumentada, patrón de escisión de sustrato alterado, y/o alto nivel de expresión in vitro.

[103] El desarrollo de nuevas y mejoradas composiciones de celulasa que comprenden cantidades variantes de celulasas tipo CBH, tipo EG y tipo BG es de interés para uso:

- (1) en composiciones detergentes que exhiben habilidad de limpieza mejorada, funciona
- 25 como un agente suavizante y/o mejora la textura de las telas de algodón (por ejemplo, “lavado por piedra” o “biopulido”); (2) en composiciones para degradar pulpa de madera u otra biomasa en azúcares (por ejemplo, para la producción de bio-combustible); y/o (3) en composiciones de alimentación.

[104] También se suministra acá la preparación completa de celulasa que comprende

- 30 variantes de celulasa. Como se usa acá, la frase “preparación completa de celulasa” se refiere a composiciones que contienen celulasa que ocurre naturalmente y no naturalmente. Una composición “que ocurre naturalmente” es una producida por una fuente que ocurre naturalmente y la cual comprende uno o más componentes tipo

celobiohidrolasa, uno o más tipo endoglucanasa, y uno o más beta-glucosidasa en donde cada uno de estos componentes se encuentra en la proporción producida por la fuente. Una composición que ocurre naturalmente es una que es producida por un organismo no modificado con respecto a las enzimas celulolíticas de modo tal que la proporción de las

5 enzimas componentes no se altera de aquella producida por el organismo nativo. Una composición "que no ocurre naturalmente" abarca aquellas composiciones producidas por: (1) combinar enzimas celulolíticas componentes en una proporción que ocurre naturalmente o que no ocurre naturalmente, es decir, proporción alterada; o (2) modificar un organismo para sobreexpresar o subexpresar una o más enzimas

10 celulolíticas; o (3) modificar un organismo de modo tal que por lo menos una enzima celulolítica se elimina. De acuerdo con esto, en algunas incorporaciones, la preparación completa de celulasa puede tener uno o más de los varios EG y/o CBH, y/o beta-glucosidasas eliminadas. Por ejemplo, el EG1 puede ser eliminado solo o en combinación con otros EG y/o CBH.

15 [105] En general, la preparación completa de celulasa incluye enzimas que incluyen, pero no se limitan a: (i) endoglucanasas (EG) o 1,4- β -d-glucano-4-glucanohidrolasas (EC 3.2.1.4), (ii) exoglucanasas, que incluyen 1,4- β -d-glucano glucanohidrolasas (también conocidas como celodextrinasas) (EC 3.2.1.74) y 1,4- β -d-glucano celobiohidrolasas (exo-celobiohidrolasas, CBH) (EC 3.2.1.91), y (iii) β -glucosidasa

20 (BG) o β -glucosida glucohidrolasas (EC 3.2.1.21).

[106] En la presente divulgación, la preparación completa de celulosas puede ser a partir de cualquier microorganismo que es útil para la hidrólisis de un material celulósico. En algunas incorporaciones, la preparación completa de celulosas es una celulasa completa de hongo filamentoso. Los "hongos filamentosos" incluyen todas las formas

25 filamentosas de la subdivisión Eumycota y Oomycota. En algunas incorporaciones, la preparación completa de celulasa es una especie *Acremonium*, *Aspergillus*, *Emericella*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Scytalidium*, *Thielavia*, *Tolypocladium*, o *Trichoderma*, de celulasa completa. En algunas incorporaciones, la preparación completa de celulasa es una celulasa completa *Aspergillus aculeatus*,

30 *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, o *Aspergillus oryzae*. En otro aspecto, la preparación completa de celulasa es una celulasa completa *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*,

- Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, o *Fusarium venenatum*. En otro aspecto, la preparación completa de
- 5 celulasa es una celulasa completa *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Scytalidium thermophilum*, o *Thielavia terrestris*. En otro aspecto, la preparación completa de celulasa es una celulasa completa *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, por ejemplo, RL-P37 (Sheir-Neiss y otros, Appl. Microbiol.
- 10 Biotecnología, 20 (1984) pp. 46 a 53; Montenecourt B.S., Can., 1 a 20, 1987), QM9414 (ATCC No. 26921), NRRL 15709, ATCC 13631, 56764, 56466, 56767, o *Trichoderma viride*, por ejemplo, ATCC 32098 and 32086. En algunas incorporaciones, la preparación completa de celulasa es una celulasa completa *Trichoderma reesei* RutC30, la cual está disponible de la Colección de Cultivo Tipo Americano como *Trichoderma*
- 15 *reesei* ATCC 5675.
- [107] Los ejemplos de preparaciones comerciales de celulasa adecuadas para su uso en la presente divulgación incluyen, por ejemplo CELLUCLAST™ (disponible de Novozymes A/S) y LAMINEX™, IndiAge™ y enzima Primafast™ LAMINEX BG (disponible de Genencor Division, Danisco US. Inc.).
- 20 [108] En la presente divulgación, la preparación completa de celulasa puede ser partir de cualquier método de cultivación de microorganismo conocido en el estado de la técnica y que resulta en la expresión de enzimas capaces de hidrolizar un material celulósico. La fermentación puede incluir cultivación en frasco para agitar, fermentación a pequeña o grande escala, tal como fermentaciones continuas, de lote, de
- 25 lote alimentado, o de estado sólido en fermentadores de laboratorio o industriales llevado a cabo en un medio adecuado y bajo las condiciones que permiten que la celulasa sea expresada o aislada.
- [109] Generalmente, el microorganismo es cultivado en un medio de cultivo de célula adecuado para la producción de enzimas capaces de hidrolizar un material celulósico.
- 30 La cultivación se lleva a cabo en un medio de nutriente adecuado que comprende fuentes de carbono y nitrógeno y sales orgánicas, utilizando procedimientos conocidos en el estado de la técnica. El medio de cultivo adecuado, los rangos de temperatura y otras condiciones adecuadas para el crecimiento y la producción de celulasa son

conocidos en el estado de la técnica. Como un ejemplo no limitante, el rango de temperatura normal para la producción de celulasas por *Trichoderma reesei* es 24°C a 28°C.

- [110] Generalmente, la preparación completa de celulasa es utilizada a medida que es producida por fermentación sin o con mínima recuperación y/o purificación. Por ejemplo, una vez las celulasas son secretadas por una célula en el medio de cultivo de célula, el medio de cultivo de célula que contiene las celulasas puede ser utilizado. En algunas incorporaciones, la preparación completa de celulasa comprende los contenidos no fraccionados de material de fermentación, incluyendo el medio de cultivo celular, enzimas extracelulares y células. Alternativamente, la preparación completa de celulasa puede ser procesada por cualquier método conveniente, por ejemplo, por precipitación, centrifugación, afinidad, filtración o cualquier otro método conocido en el estado de la técnica. En algunas incorporaciones, la preparación completa de celulasa puede ser concentrada, por ejemplo, y luego utilizada sin purificación adicional. En algunas incorporaciones, la preparación completa de celulasa comprende agentes químicos que reducen la viabilidad celular o que mata las células. En algunas incorporaciones, las células se lisan o permeabilizan utilizando métodos conocidos en el estado de la técnica.

20 **III. Biología Molecular**

- [111] En una incorporación, esta divulgación suministra la expresión de genes de *cbh2* variante bajo el control de un promotor funcional en un hongo filamentoso. Por lo tanto, esta divulgación se basa en técnicas de rutina en el campo de genética recombinante (ver, por ejemplo, Sambrook y otros, *Clonación Molecular, Un Manual de Laboratorio*, 2ª Ed., 1989; Kriegler, *Transferencia y Expresión de Genes: Un Manual de Laboratorio*, 1990; y Ausubel y otros, eds., **PROTOCOLOS ACTUALES EN BIOLOGÍA MOLECULAR**, Greene Publishing and Wiley-Interscience, New York, 1994).

Métodos para Mutar Secuencias de Ácido Nucleico *cbh2*

- 30 [112] Cualquier método conocido en el estado de la técnica que pueda introducir mutaciones, se contempla por parte de la presente divulgación.
- [113] La presente divulgación se relaciona con la expresión, purificación y/o asilamiento y uso del CBH2 variante. Estas enzimas son preferiblemente preparadas por

métodos recombinantes que utilizan el gen *cbh2* a partir de *H. jecorina*. El caldo de fermentación puede ser utilizado con o sin purificación.

- [114] Después del aislamiento y la clonación del gen *cbh2* de *H. jecorina*, otros métodos conocidos en el estado de la técnica, tal como mutagenesis dirigida en el sitio, son utilizados para hacer las sustituciones, adiciones o eliminaciones que corresponden a los aminoácidos sustituidos en la variante de CBH2 expresada. De nuevo, la mutagenesis dirigida en el sitio y otros métodos para incorporar cambios de aminoácidos en proteínas expresadas en el nivel de ADN son conocidos en el estado de la técnica (Sambrook y otros, supra; y Ausubel y otros, supra).
- 10 [115] El ADN que codifica una variante de secuencia de aminoácido del CBH2 de *H. jecorina* se prepara por medio de una variedad de métodos conocidos en el estado de la técnica. Estos métodos incluyen, pero no se limitan a, preparación por mutagenesis dirigida en el sitio (o mediada de oligonucleótido), mutagenesis PCR, y mutagenesis de módulo de un ADN preparado anteriormente que codifica el CBH2 de *H. jecorina*.
- 15 [116] La mutagenesis dirigida en el sitio es un método preferido para preparar las variantes de sustituciones. Esta técnica es bien conocida en el estado de la técnica (ver, por ejemplo, Carter y otros, Res. Ácidos Nucleicos, 13:4431 a 4443 (1985) y Kunkel y otros, Proc. Natl. Acad. Sci. USA 82:488 (1987)). Brevemente, al llevar a cabo la mutagenesis dirigida en el sitio de ADN, el ADN de partida es alterado al primero
- 20 hibridar un oligonucleótido que codifica la mutación deseada a una rama simple de dicho ADN de partida. Después de la hibridación, una polimerasa de ADN es utilizada para sintetizar una segunda rama completa, utilizando el oligonucleótido hibridado como un cebador, y utilizando la rama simple del ADN de inicio como una plantilla. Así, el oligonucleótido que codifica la mutación deseada se incorpora en el ADN de
- 25 doble rama resultante.
- [117] La mutagenesis PCR es también adecuada para hacer variantes de secuencia de aminoácido del polipéptido de inicio, es decir, el CBH2 de *H. jecorina*. Ver Higuchi, en Protocolos PCR, pp. 177 a 183 (Prensa Académica, 1990); y Valletta y otros, Res. Ácidos Nuc. 17:723 a 733 (1989). Ver, también, por ejemplo Cadwell y otros, Métodos y Aplicaciones PCR, Vol. 2, 28 a 33 (1992). Brevemente, cuando se utilizan cantidades
- 30 pequeñas de ADN de plantilla como material de inicio en un PCR, los cebadores que difieren un poco en la secuencia de la región correspondiente en un ADN de plantilla pueden ser utilizados para generar cantidades relativamente grandes de un fragmento de

ADN específico que difiere de la secuencia de plantilla únicamente en las posiciones donde los cebadores difieren de la plantilla.

- [118] Otro método para preparar variantes, mutagenesis de módulo, se basa en la técnica descrita por Wells y otros, Gen 34:315 a 323 (1985). El material de inicio es el plásmido (u otro vector) que comprende el ADN polipéptido de inicio a ser mutado. Se identifica el codón en el ADN de inicio a ser mutado. Debe haber un sitio de endonucleasa de restricción único en cada lado del sitio de mutación identificado. Si dichos sitios de restricción no existen, ellos pueden ser generados utilizando el método de mutagenesis mediada por oligonucleótido descrito anteriormente para introducirlos en ubicaciones apropiadas en el ADN polipéptido de inicio. Al ADN plásmido se corta en estos sitios para alinearlo. Un oligonucleótido de doble rama que codifica la secuencia del ADN entre los sitios de restricción pero que contiene las mutaciones deseadas, es sintetizado utilizando procedimientos estándar, en donde las dos ramas del oligonucleótido son sintetizadas de forma separada y luego hibridadas juntas usando técnicas estándar. Este oligonucleótido de doble rama se refiere como el módulo. Este módulo se diseña para tener extremos 5' y 3' que son compatibles con los extremos del plásmido alineado, de modo tal que puede ser ligado directamente con el plásmido. Este plásmido ahora contiene la secuencia de ADN mutado.

- [119] Alternativamente, o adicionalmente, la secuencia de aminoácido deseada que codifica un CBH2 variante puede ser determinada, y una secuencia de ácido nucleico que codifica dicha variante de secuencia de aminoácido se puede generar de forma sintética.

- [120] El CBH2 variante preparado de esta forma puede ser sujeto a modificaciones adicionales, a menudo dependiendo del uso previsto de la celulasa. Dichas modificaciones pueden involucrar alteración adicional de la secuencia de aminoácido, fusión a polipéptidos heterólogos y/o modificaciones covalentes.

IV. Ácidos Nucleicos de cbh2 y Polipéptidos de CBH2

30 A. Ácidos Nucleicos de Tipo cbh2 Variante

- [121] La secuencia de ácido nucleico para el cbh2 de tipo salvaje se muestra en la SEQ ID NO:1. La divulgación abarca una molécula de ácido nucleico que codifica las celulasas variantes acá descritas. El ácido nucleico puede ser una molécula de ADN.

[122] Después que las secuencias de ADN que codifican las variantes de CBH2 han sido clonadas en construcciones de ADN, el ADN es utilizado para transformar microorganismos. El microorganismo a ser transformado para el propósito de expresar un CBH2 variante de acuerdo con la presente divulgación puede ventajosamente comprender una cepa derivada de *Trichoderma sp.* Así, un modo preferido para preparar celulasas de CBH2 variante de acuerdo con la presente divulgación comprende transformar una célula huésped de *Trichoderma sp.* con una construcción de ADN que comprende por lo menos un fragmento de ADN que codifica una porción o todo el CBH2 variante. La construcción de ADN generalmente estará unida funcionalmente con un promotor. La célula huésped transformada luego crece bajo condiciones de modo que expresa la proteína deseada. Subsecuentemente, el producto de proteína deseada puede ser purificado a homogeneidad sustancial.

[123] Sin embargo, esto puede de hecho ser que el mejor vehículo de expresión para un ADN dado que codifica un CBH2 variante puede diferir de *H. jecorina*. Así, puede ser que será más ventajoso expresar una proteína en un huésped de transformación que soporta similitud filogenética con el organismo fuente para el CBH2 variante. En una incorporación alternativa, el *Aspergillus niger* puede ser utilizado como un vehículo de expresión. Para una descripción de técnicas de transformación con *A. niger*, ver WO 98/31821, cuya divulgación es incorporada por referencia en su totalidad.

[124] De acuerdo con lo anterior, la presente descripción de un sistema de expresión de *Aspergillus spp.* se suministra para propósitos ilustrativos únicamente y como una opción para expresar el CBH2 variante de la divulgación. Un experto en el estado de la técnica, sin embargo, puede verse inclinado o motivado a expresar el ADN que codifica el CBH2 variante en una célula huésped diferente si lo considera apropiado y se deberá entender que la fuente del CBH2 variante debe ser considerada en la determinación del huésped de expresión óptimo. Adicionalmente, una persona experta en el campo será capaz de seleccionar el mejor sistema de expresión para un gen particular a través de técnicas de rutina utilizando las herramientas disponibles en el estado de la técnica.

30 B. Polipéptidos de CBH2 Variante

[125] Los CBH2 variantes de esta divulgación tienen secuencias de aminoácido que se derivan de la secuencia de aminoácido de un CBH2 precursor. La secuencia de aminoácido de la variante de CBH2 difiere de la secuencia de aminoácido de CBH2

precursor por la sustitución, eliminación o inserción de uno o más aminoácidos de la secuencia de aminoácido precursor. En una incorporación preferida, el CBH2 precursor es CBH2 de *Hypocrea hecorina*. La secuencia de aminoácido madura del CBH2 de *H. jecorina* se muestra en el Ejemplo 2 (SEQ ID NO:3). Así, esta divulgación se dirige a

5 variantes de CBH2 las cuales contienen residuos de aminoácido en las posiciones en las cuales son equivalentes con el residuo identificado particular en CBH2 de *H. jecorina*. Un residuo (aminoácido) de un CBH2 homólogo es equivalente a un residuo de CBH2 de *Hypocrea jecorina* si éste es homólogo (es decir, que corresponde en posición en la estructura primaria o terciaria) o es funcionalmente análogo a un residuo específico o

10 porción del residuo en el CBH2 de *Hypocrea jecorina* (es decir, que tiene la misma o similar capacidad funcional para combinar, reaccionar, o interactuar químicamente o estructuralmente). Como se usa acá, la numeración pretende corresponder a la de la secuencia de aminoácido de CBH2 maduro (SEQ ID NO:3).

[126] La alineación de las secuencias de aminoácido para determinar la homología es preferiblemente determinada al usar un "algoritmo de comparación de secuencia". La alineación óptima de las secuencias para la comparación se puede llevar a cabo, por ejemplo, por medio del algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2:482 (1981), por medio del algoritmo de alineación de homología de Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), por el método de búsqueda de similitud de

20 Pearson & Lipman, Proc. Natl Acad. Sci. USA 85:2444 (1988), por implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA en el Paquete de Software de Genética de Wisconsin, Grupo de Computador de Genética, 575 Ciencia Dr., Madison, Wis), o por inspección visual, la inspección visual puede utilizar paquetes gráficos tales como, por ejemplo, MOE por Grupo de Computación Química, Montreal,

25 Canadá.

[127] Un ejemplo de un algoritmo que es adecuado para determinar la similitud de secuencia es el algoritmo BLAST, el cual se describe en Altschul y otros, J. Mol. Biol. 215:403 a 410 (1990). El software para llevar a cabo los análisis BLAST se encuentra disponible públicamente a través del Centro Nacional para Información de

30 Biotecnología (www.ncbi.nlm.nih.gov). Este algoritmo involucra primero identificar los pares de frecuencia de alta puntuación (HSP) al identificar palabras cortas de longitud W en la secuencia de pregunta que iguala o satisface algún puntaje de umbral de valor positivo T cuando se alinea con una palabra de la misma longitud en una secuencia de

base de datos. La palabra circundante inicial actúa como puntos de inicio para encontrar HSP más largos que los contienen. Las palabras se expanden en ambas direcciones a lo largo de cada una de las dos secuencias que son comparadas por hasta que el puntaje de alineación acumulativa pueda aumentarse. La extensión de las palabras se detiene cuando: la puntuación de alineación acumulativa cae por la cantidad X de un valor máximo alcanzado; la puntuación acumulativa va a cero o por debajo; o el extremo de cualquier secuencia es alcanzado. Los parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y la velocidad de la alineación. El programa BLAST utiliza por defecto una longitud de palabra (W) de 11, las alineaciones de matriz de puntuación (B) BLOSUM62 (ver Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)) de 50, expectativa (E) de 10, M²5, N²-4, y una comparación de ambos filamentos.

[128] El algoritmo BLAST luego lleva a cabo un análisis estadístico de la similitud entre las dos secuencias (ver, por ejemplo, Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873 a 5787 (1993)). Una medida de similitud provista por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), la cual suministra una indicación de la probabilidad por la cual una coincidencia entre dos nucleótidos o secuencias de aminoácido se presentaría por probabilidad. Por ejemplo, una secuencia de aminoácido es considerada similar a una proteasa su la probabilidad de suma más pequeña en una comparación de la secuencia de aminoácido de prueba a una secuencia de aminoácido de proteasa es menor a aproximadamente 0.1, más preferiblemente menor a aproximadamente 0.01, y más preferiblemente menor a aproximadamente 0.001.

[129] Para propósitos de la presente divulgación, el grado de identidad puede ser determinado adecuadamente por medio de programas de computador conocidos en el estado de la técnica, tal como GAP suministrado en el paquete de programa GCG (Manual de Programa para el Paquete Wisconsin, Versión 8, Agosto 1994, Grupo de Computación de Genética, 575 Science Drive, Madison, Wis., USA 53711) (Needleman, S. B. y Wunsch, C. D., (1970), Diario de Biología Molecular, 48, 443 a 45), usando GAP con las siguientes configuraciones para la comparación de secuencia de polinucleótido: penalidad de creación GAP de 5.0 y penalidad de extensión GAP de 0.3.

[130] Una alineación estructural entre un CBH2 de *T. reesei* y otras celulasas puede ser utilizada para identificar las posiciones equivalentes o correspondientes en otras celulasas que tienen un grado de homología moderado a alto, por ejemplo, aproximadamente 50%, 55%, 60 %, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o hasta 99%, con CBH2 de *T. reesei* (SEQ ID NO:3). Un método para obtener la alineación estructural es utilizar el programa Pile Up del paquete GCG que utiliza valores por defecto de penalidades GAP, es decir, una penalidad de creación GAP de 3.0 y una penalidad de extensión GAP de 0.1. Otros métodos de alineación estructural incluyen el análisis de grupo hidrofóbico (Gaboriaud y otros, Cartas FEBS, 224:149 a 155, 1987) y roscado contrario (Huber y Torda, Ciencia de Proteína, 7:142 a 149, 1998).

[131] Una alineación ejemplar de la forma madura de varias celulasas de referencia se suministra como Figura 1. Las celulasas de referencia incluyen *Hypocrea jecorina* (también conocida como *T. reesei*) CBH2 (SEQ ID NO:3), *Hypocrea koningii* CBH2 (SEQ ID NO:4), *Humicola insolens* CBH2 (SEQ ID NO:5), *Acremonium cellulolyticus* CBH2 (SEQ ID NO:6), *Agaricus bisporus* CBH2 (SEQ ID NO:7), *Fusarium osysporum* EG (SEQ ID NO: 8), *Phanerochaete chrysosporium* CBH2 (SEQ ID NO:9), *Talaromyces emersonii* CBH2 (SEQ ID NO: 10), *Thermobifida fusca* 6B/E3 CBH2 (SEQ ID NO: 11), *Thermobifida fusca* 6A/E2 EG (SEQ ID NO: 12), y *Cellulomonas fimi* *CenA* EG (SEQ ID NO: 13). Las secuencias fueron alineadas utilizando los algoritmos de alineación de secuencia múltiple ClustalW y MUSCLE. Una matriz que muestra el porcentaje de identidad de celulasas de la alineación de secuencia de la Figura 1 se suministra en la Tabla 1.

Tabla 1. Matriz de Porcentaje de Identidad de Celulasa*

*Los números en la columna superior y la columna de la izquierda corresponden a las SEQ ID NOS de las secuencias alineadas de la Figura 1.

[132] Las búsquedas de secuencias son típicamente llevadas a cabo usando el programa BLASTN cuando se evalúa una secuencia de ácido nucleico dada con relación a las secuencias de ácido nucleico en las Secuencias de ADN GenBank y otras bases de datos públicas. La programa BLASTX es preferida para buscar las secuencias de ácido

nucleico que han sido trasladadas en todos los marcos de lectura contra las secuencias de aminoácido en las Secuencias de Proteína del GenBank y otras bases de datos públicas. Tanto el BLASTN como el BLASTX se ejecutan usando parámetros por defecto de una penalidad de brecha abierta de 11.0, y una penalidad de brecha extendida de 1.0, y utilizan la matriz BLOSUM-62 (Ver, por ejemplo, Altschul y otros, 1997).

V. Expresión de Variantes de CBH2 Recombinante

[133] Los métodos de la divulgación confían en el uso de células para expresar CBH2 variante, sin ningún método particular de expresión de CBH2 requerido. EL CBH2 variante es preferiblemente secretado a partir de las células. La divulgación suministra células huésped las cuales han sido transducidas, transformadas o transfectadas con un vector de expresión que comprende una secuencia de ácido nucleico que codifica CBH2 variante. Las condiciones de cultivo, tales como temperatura, pH y similares, son aquellos utilizados previamente para la célula huésped madre antes de la transducción, transformación o transfección y serán evidentes para los expertos en el estado de la técnica.

[134] En un acercamiento, una célula de hongo filamentoso o célula de levadura es transfectada con un vector de expresión que tiene un promotor o un fragmento de promotor biológicamente activo o uno o más (por ejemplo, una serie) de mejoradores los cuales funcionan en la línea de célula huésped, vinculada operablemente con un segmento de ADN que codifica el CBH2 variante, de modo tal que el CBH2 variante se expresa en la línea celular.

A. Construcciones de Ácido Nucleico/Vectores de Expresión

[135] Los fragmentos de polinucleótido natural o sintético que codifica el CBH2 variante ("secuencias de ácido nucleico que codifica CBH2") pueden ser incorporados en construcciones o vectores de ácido nucleico heterólogo, capaces de introducción en, y replicación en, una célula de hongo filamentoso o de levadura. Los vectores y métodos acá divulgados son adecuados para su uso en las células huésped para la expresión de CBH2 variante. Cualquier vector puede ser utilizado siempre y cuando sea replicable y viable en las células en las cuales se introducen. Los números grandes de vectores y promotores adecuados son conocidos para los expertos en el estado de la técnica, y están comercialmente disponibles. Los vectores de clonación y de expresión

también se describen en Sambrook y otros, 1989, Ausubel F M y otros, 1989, y Strathern y otros, La Biología Molecular de la Levadura *Saccharomyces*, 1981, cada una de las cuales se incorpora expresamente acá por referencia. Los vectores de expresión apropiados para hongos se describen en van den-Hondel, C. A. M. J. J. y otros

5 (1991) En: Bennet, J. W. y Lasure, L. L. (eds.) *Más Manipulaciones de Gen en Hongos*. Prensa Académica, pp. 396 a 428. La secuencia de ADN apropiada puede ser insertada en un plásmido o vector (colectivamente relacionado acá como "vectores") por medio de una variedad de procedimientos. En general, la secuencia de ADN se inserta en un sitio de restricción de endonucleasa apropiado por procedimientos estándar. Dichos

10 procedimientos y los procedimientos de sub-clonación relacionados son considerados por encontrarse dentro del alcance del conocimiento de los expertos en el estado de la técnica.

[136] Los hongos filamentosos recombinantes que comprenden la secuencia de codificación para el CBH2 variante pueden ser producidos al introducir una

15 construcción de ácido nucleico heterólogo que comprende la secuencia que codifica el CBH2 variante en las células de una cepa seleccionada de hongos filamentosos.

[137] Una vez que se obtiene la forma deseada de una secuencia de ácido nucleico de *cbh2* variante, ésta puede ser modificada en una variedad de formas. Donde la secuencia involucra regiones no codificadores de acompañamiento, las regiones de

20 acompañamiento pueden estar sujetas a resección, mutagenesis, etc. Así, las transiciones, transversiones, eliminaciones, e inserciones pueden ser llevadas cabo en la secuencia que ocurre naturalmente.

[138] Una secuencia que codifica *cbh2* variante seleccionada se puede insertar en un vector adecuado de acuerdo con técnicas recombinantes bien conocidas y usadas para

25 transformar hongos filamentosos capaces de la expresión de CBH2. Debido a la degeneración inherente del código genético, otras secuencias de ácido nucleico que codifican sustancialmente la o misma o una secuencia de aminoácido funcionalmente equivalente puede ser utilizada para clonar y expresar CBH2 variante. Por lo tanto, se apreciará que dichas sustituciones en la región de codificación caerán dentro de las

30 variantes de secuencia cubiertas por la presente divulgación. Cualquiera y todas de estas variantes de secuencia se pueden utilizar de la misma forma como se describe acá para una secuencia de ácido nucleico que codifica CBH2 madre.

- [139] La presente divulgación también incluye construcciones de ácido nucleico recombinante que comprende una o más de las secuencias de ácido nucleico que codifica CBH2 variante como se describieron anteriormente. Las construcciones comprenden un vector, tal como un plásmido o un vector viral, en el cual se ha insertado una secuencia de la divulgación, en una orientación hacia adelante o hacia atrás.
- [140] Las construcciones de ácido nucleico heterólogo pueden incluir la secuencia de codificaciones para cbh2 variante. (i) en aislamiento; (ii) en combinación con secuencias de codificación adicionales; tal como secuencias de codificación de proteína de fusión o de péptido de señal, donde la secuencia que codifica cbh2 es la secuencia de codificación dominante; (iii) en combinación con secuencias no codificadoras, tal como intrones y elementos de control, tal como elementos promotores y terminadores o regiones no trasladadas 5' y/o 3', efectivas para la expresión de la secuencia de codificación en un huésped adecuado; y/o (iv) en un vector o ambiente huésped en el cual la secuencia que codifica cbh2 es un gen heterólogo.
- [141] En un aspecto de la presente divulgación, una construcción de ácido nucleico heterólogo se emplea para transferir una secuencia de ácido nucleico que codifica CBH2 variante en una célula in vitro, con líneas preferidas de hongos filamentosos y levadura. Para producción a largo término de CBH2 variante, se prefiere la expresión estable. Sigue que cualquier método efectivo para generar transformantes estables puede ser utilizado en la práctica de la divulgación.
- [142] Los vectores apropiados están típicamente equipados con una secuencia de ácido nucleico que codifica el marcador, sitios de inserción, y elementos de control adecuados, tal como secuencias promotoras o de terminación. El vector puede comprender secuencias regulatorias, que incluyen, por ejemplo, secuencias no codificadoras, tal como intrones y elementos de control, es decir, elementos promotores y terminadores o regiones no trasladadas 5' y/o 3', efectivas para la expresión de la secuencia de codificación en células huésped (y/o en un vector o ambiente de célula huésped en el cual una secuencia que codifica antígeno de proteína soluble modificada no se expresa normalmente), vinculada operablemente con la secuencia de codificación.
- Números grandes de vectores y promotores adecuados son conocidos para los expertos en el estado de la técnica, varios de los cuales se encuentran disponibles comercialmente y/o se describen en Sambrook, y otros (supra).

[143] Los promotores ejemplares incluyen los promotores constitutivos y promotores inducibles, cuyos ejemplos incluyen un promotor CMV, un promotor temprano SV40, un promotor RSV, un promotor alfa EF-1, un promotor que contiene el elemento de respuesta tet (TRE) en el sistema tet-on o tet-off como se describió (ClonTech y BASF),
 5 el promotor de actina beta y el promotor de metalotionina que se pueden regular por la adición de ciertas sales de metal. Una secuencia de promotor es una secuencia de ADN la cual es reconocida por el hongo filamentoso particular para propósitos de expresión. Este es vinculado operablemente con la secuencia de ADN que codifica un polipéptido de CBH2 variante. Dicha vinculación comprende el posicionamiento del promotor con
 10 respecto al codón de iniciación de la secuencia de ADN que codifica el polipéptido de CBH2 variante en los vectores de expresión divulgados. La secuencia de promotor contiene secuencia de control de transcripción y traslación los cuales median la expresión del polipéptido de CBH2 variante. Los ejemplos incluyen los promotores de *Aspergillus niger*, *A. awamori* o *A. oryzae* glucoamilasa, alfa-amilasa, o genes de
 15 codificación de alfa-glucosidasa; los Genes de *A. nidulans* *gpdA* o *trpC*; los genes de *Neurospora crassa* *cbh1* o *trp1*; los genes que codifican la proteinasa aspártica de *A. niger* o *Rhizomucor miehei*; los genes que codifican *H. jecorina* (*T. reesei*) *cbh1*, *cbh2*, *egl1*, *cgl2*, u otra celulasa.

[144] La selección del marcador seleccionable apropiado dependerá de la célula
 20 huésped, y los marcadores apropiados para diferentes huéspedes son bien conocidos en el estado de la técnica. Los genes marcadores seleccionables típicos incluyen *argB* de *A. nidulans* o *T. reesei*, *amdS* de *A. nidulans*, *pyr4* de *Neurospora crassa* o *T. reesei*, *pyrG* de *Aspergillus niger* o *A. nidulans*. Los marcadores seleccionables ejemplares adicionales incluyen, pero no se limitan a *trpC*, *trp1*, *oliC31*, *niaD* o *leu2*, los cuales se
 25 incluyen en construcciones de ácido nucleico heterólogo utilizados para transformar una cepa mutante tal como *trp*-, *pyr*-, *leu*- y similares.

[145] Dichos marcadores seleccionables le confieren a los transformantes la habilidad de utilizar un metabolito que usualmente no es metabolizado por el hongo filamentoso. Por ejemplo, el gen *amdS* de *H. jecorina* el cual codifica la enzima acetamidasa que
 30 permite que las células transformantes crezcan en acetamida como una fuente de nitrógeno. El marcador seleccionable (por ejemplo, *pyrG*) puede restaurar la habilidad de una cepa mutante auxotrófica de crecer en un medio mínimo selectivo o el marcador

seleccionable (por ejemplo, olic31) puede conferir a los transformantes la habilidad de crecer en la presencia de un fármaco inhibidor o antibiótico.

[146] El marcador seleccionable que codifica la secuencia es clonado en cualquier plásmido adecuado usando métodos generalmente empleados en el estado de la técnica.

- 5 Los plásmidos ejemplares incluyen pUC18, pBR322, pRAX y pUC100. El plásmido pRAX contiene secuencias AMAL de *A. nidulans*, las cuales hacen posible que se replique en *A. niger*.

- [147] La práctica de la presente divulgación empleará, a menos que se indique lo contrario, técnicas convencionales de biología molecular, microbiología, ADN
10 recombinante, e inmunología, los cuales se encuentran dentro del campo de la técnica. Dichas técnicas se explican completamente en la literatura. Ver, por ejemplo, Sambrook y otros, 1989; Freshney, Cultivo de Células Animales, 1987; Ausubel y otros, 1993; y Coligan y otros, Protocolos Actuales en Inmunología, 1991.

15 **B. Células Huésped y Condiciones de Cultivo para la Producción de CBH2 de (i) Hongos Filamentosos**

- [148] Así, la presente divulgación suministra hongos filamentosos que comprenden células las cuales han sido modificadas, seleccionadas y cultivadas en una forma efectiva para resultar en producción de CBH2 variante o la expresión relativa al hongo
20 madre no transformado correspondiente.

- [149] Los ejemplos de especies de hongos filamentosos madre que pueden ser tratados y/o modificados para la expresión de CBH2 variante incluyen, pero no se limitan a Trichoderma, por ejemplo, Trichoderma reesei, Trichoderma longibrachiatum, Trichoderma viride, Trichoderma koningii; Penicillium sp., Humicola sp., including
25 Humicola insolens; Aspergillus sp., Chrysosporium sp., Fusarium sp., Hypocrea sp., y Emericella sp.

- [150] Las células que expresan CBH2 son cultivadas bajo condiciones típicamente empleadas en el cultivo de la línea de hongos madre. Generalmente, las células son cultivadas en un medio estándar que contiene sales fisiológicas y nutrientes, tal como lo
30 descritos en Pourquie, J. y otros, Bioquímica y Genética de Degradación de Celulosa, eds. Aubert, J. P. y otros, Prensa Académica, pp. 71 a 86, 1988 y Ilmen, M. y otros, Appl. Environ. Microbiol. 63:1298 a 1306, 1997. Las condiciones de cultivo son también estándar, por ejemplo, los cultivos son incubados a 28 grados C en cultivos de

agitación o en fermentadores hasta que se alcanzan los niveles deseados de expresión de CBH2.

[151] Las condiciones de cultivo preferidas para un hongo filamentoso dado se pueden encontrar en la literatura científica y/o a partir de la fuente del hongo tal como la

- 5 Colección de Cultivo de Tipo Americano (ATCC; www.atcc.org/). Después que el crecimiento del hongo se ha establecido, las células son expuestas a condiciones efectivas para causar o permitir la expresión del CBH2 variante.

- [152] En los casos donde una secuencia que codifica CBH2 se encuentra bajo el control de un promotor inducible, el agente de inducción, por ejemplo, un azúcar, una sal de
10 metal o antibiótico, se adiciona al medio en una concentración efectiva para inducir la expresión de CBH2.

- [153] En una incorporación, la cepa comprende *Aspergillus niger*, la cual es una cepa útil para obtener proteína sobreexpresada. Por ejemplo, *A. niger* var *awamori* 246 es conocida por secretar cantidades elevadas de celulasas secretadas (Goedegebuur y otros,
15 *Curr. Genet.* (2002) 41:89 a 98). Otras cepas de *Aspergillus niger* var *awamori* tales como GCD AP3, GCD AP4 y GAP3-4 son conocidas por Ward y otros (Ward, M, Wilson, L. J. y Kodama, K. H., 1993, *Appl. Microbiol. Biotechnol.* 39:738 a 743).

- [154] En otra incorporación, la cepa comprende *Trichoderma reesei*, la cual es una cepa útil para obtener proteína sobreexpresada. Por ejemplo, RL-P37, descrita por Sheir-
20 Neiss y otros, *Appl. Microbiol. Biotechnol.* 20:46 a 53 (1984) es conocida por secretar cantidades elevadas de enzimas de celulasa. Los equivalentes funcionales de RL-P37 incluyen la cepa *Trichoderma reesei* RUT-C30 (ATCC No. 56765) y la cepa QM9414 (ATCC No. 26921). Se contempla que estas cepas también serían útiles en la sobreexpresión de CBH2 variante.

- [155] Cuando se desea obtener el CBH2 variante en la ausencia de actividad celulolítica nativa potencialmente perjudicial, es útil obtener una célula huésped de *Trichoderma* la cual ha tenido uno o más genes de celulasa eliminados antes de la introducción de una construcción de ADN o plásmido que contiene el fragmento de ADN que codifica el CBH2 variante. Dichas cepas se pueden preparar por medio de los métodos divulgados
25 en la Patente de los Estados Unidos No. 5.246.853 y WO 92/06209, cuyas divulgaciones se incorporan acá por referencia. Al expresar una celulasa de CBH2 variante en un microorganismo huésped que carece de uno o más genes de celulasa, los procedimientos de identificación y siguiente purificación son simplificados. Cualquier

gen de *Trichoderma* sp. que ha sido clonado puede ser eliminado, por ejemplo, los genes *cbh1*, *cbh2*, *egl1*, y *egl2* así como aquellos que codifican proteína EG III y/o EGV (ver por ejemplo, la patente de los Estados Unidos No. 5.475.101 y WO 94/28117, respectivamente).

- 5 [156] La eliminación del gen se puede lograr al insertar una forma del gen deseado a ser eliminado o interrumpido en un plásmido mediante métodos conocidos en el estado de la técnica. El plásmido de eliminación es luego cortado en un sitio de enzima de restricción apropiado, interno a la región deseada que codifica el gen, y la secuencia que codifica el gen o parte del mismo reemplazada con un marcador seleccionable. Las
- 10 secuencias de ADN de acompañamiento del lugar del gen a ser eliminado o interrumpido, preferiblemente entre aproximadamente 0.5 a 2.0 kb, permanecen en cualquier lado del gen marcador seleccionable. Un plásmido de eliminación apropiado generalmente tendrá sitios de restricción de enzima únicos presenten allí para permitir que los fragmentos que contienen el gen eliminado, incluyendo las secuencias de ADN
- 15 de acompañamiento, y el gen marcador seleccionable será removidos como una pieza lineal simple.

- [157] Un marcador seleccionable debe ser seleccionado con el fin de permitir la eliminación del microorganismo transformado. Cualquier gen marcador seleccionable que sea expresado en el microorganismo seleccionado será adecuado. Por ejemplo, con
- 20 *Aspergillus* sp., el marcador seleccionable se selecciona de modo que la presencia del marcador seleccionable en los transformantes no afectará significativamente las propiedades del mismo. Dicho marcador seleccionable puede ser un gen que codifica un producto ensayable. Por ejemplo, una copia funcional de un gen *Aspergillus* sp. puede ser utilizada la cual si carece en la cepa huésped resulta en que la cepa huésped visualiza
- 25 un fenotipo auxotrófico. De forma similar, los marcadores seleccionables existen para *Trichoderma* sp.

- [158] En una incorporación, una cepa derivada de *pyrG*- de *Aspergillus* sp. se transforma con un gen *pyrG* funcional, el cual así suministra un marcador seleccionable para transformación. Una cepa derivada de *pyrG*- se puede obtener por medio de
- 30 selección de cepas de *Aspergillus* sp. que son resistentes al ácido fluoroorótico (FOA). El gen *pyrG* codifica decarboxilasa orotidina-5'-monofosfato, una enzima requerida para la biosíntesis de uridina. Las cepas con un gen *pyrG* intacto crecen en un medio que carece de uridina pero que son sensibles al ácido fluoroorótico. Es posible

seleccionar cepas derivadas de pyrG- que carecen de una enzima de decarboxilasa monofosfato orotidina funcional y requieren uridina para el crecimiento al seleccionar resistencia a FOA. Utilizando la técnica de selección de FOA también es posible obtener cepas que requieren uridina las cuales carecen de una transferasa pirofosforobosil orotato funcional. Es posible transformar estas células con una copia funcional del gen que codifica esta enzima (Berges & Barreau, Curr. Genet. 19:359 a 365 (1991), y van Hartingsveldt y otros (1986) Desarrollo de un sistema de transformación homólogo para *Aspergillus niger* basado en el gen pyrG. Mol. Gen. Genet. 206:71 a 75). La selección de las cepas derivadas es llevada a cabo fácilmente utilizando la técnica de resistencia a FOA relacionada anteriormente, y así, el gen pyrG es empleado preferiblemente como un marcador seleccionable.

[159] En una segunda incorporación, una cepa derivada de pyr4 de *Hypocrea* sp. (*Hypocrea* sp. (*Trichoderma* sp.)) se transforma con un gen pyr4 funcional, el cual así suministra un marcador seleccionable para transformación. Una cepa derivada de pyr4.sup se puede obtener por medio de la selección de cepas de *Hyprocera* sp. (*Trichoderma* sp.) que son resistentes al ácido fluoroorótico (FOA). El gen pyr4 codifica decarboxilasa orotidina-5'-monofosfato, una enzima requerida para la biosíntesis de uridina. Las cepas con un gen pyr4 intacto crecen en un medio que carece de uridina pero son sensibles al ácido fluoroorótico. Es posible seleccionar cepas derivadas de pyr4.sup que carecen de una enzima de decarboxilasa monofostato orotidina funcional y requieren de uridina para el crecimiento al seleccionar resistencia a FOA. Utilizando la técnica de selección de FOA también es posible obtener cepas que requieren de uridina las cuales carecen de una transferasa pirofosforobosil orotato. Es posible transformar estas células con una copia funcional del gen que codifica esta enzima (Berges & Barreau, 1991). La selección de las cepas derivadas se lleva a cabo fácilmente utilizando la técnica de resistencia a FOA relacionada anteriormente, y así, el gen pyr4 es empleado preferiblemente como un marcador seleccionable.

[160] Para transformar pyrG- *Aspergillus* sp. o pyr4- *Hyprocera* sp. (*Trichoderma* sp.) con el fin de ser carente de la habilidad de expresar uno o más genes de celulasa, un fragmento simple de ADN que comprende un gen de celulasa interrumpido o transformado es luego aislado del plásmido de eliminación y utilizado para transformar un huésped apropiado de pyr- *Aspergillus* o pyr- *Trichoderma*. Los transformantes son luego identificados y seleccionados con base en su habilidad de expresar el gen pyrG o

pyr4, respectivamente, y así adicionar la auxotrofia de uridina la cepa huésped. El análisis de blot del Sur luego se lleva a cabo en los transformantes resultantes para identificar y confirmar un evento de integración de cruzado doble que reemplaza parte o toda la región de codificación de la copia genómica del gen a ser eliminado con los marcadores seleccionables apropiados de pyr.

[161] Aunque los vectores plásmidos específicos descritos anteriormente se relacionan con la preparación de transformantes pyr-, la presente divulgación no se limita a estos vectores. Varios genes pueden ser eliminados y reemplazados en la cepa *Aspergillus* sp. o *Hyprocrea* sp. (*Trichoderma* sp.) utilizando las técnicas anteriores. De hecho, cualquier huésped, por ejemplo, el gen de *Aspergillus* sp. o *Hyprocrea* sp. que ha sido clonado, y así identificado, puede ser eliminado del genoma usando la estrategia anteriormente descrita.

[162] Como se indicó anteriormente, las cepas huésped utilizadas pueden ser derivados de *Hyprocrea* sp. (*Trichoderma* sp.) que carecen o tienen un gen no funcional o genes que corresponden a los marcadores seleccionables escogidos. Por ejemplo, si se selecciona el marcador seleccionable de pyrG para *Aspergillus* sp., entonces se utiliza una cepa específica derivada de pyrG- como un recipiente en el procedimiento de transformación. También, por ejemplo, si se selecciona el marcador seleccionable de pyr4 para un *Hyprocrea* sp., entonces se utiliza una cepa específica derivada de pyr4- en el procedimiento de transformación. De forma similar, se pueden utilizar los marcadores seleccionables que comprenden genes de *Hyprocrea* sp. (*Trichoderma* sp.) equivalentes a los genes de *Aspergillus* *nidulans* *amdS*, *argB*, *trpC*, *niaD*. La cepa recipiente correspondiente debe por lo tanto ser una cepa derivada tal como *argB*-, *trpC*-, *niaD*-, respectivamente.

[163] El ADN que codifica la variante de CBH2 es luego preparada para su inserción en un microorganismo apropiado. De acuerdo con la presente divulgación, el ADN que codifica una variante de CBH2 comprende el ADN necesario para codificar una proteína que tiene actividad celulolítica funcional. El fragmento de ADN que codifica la variante de CBH2 puede estar unido funcionalmente con una secuencia promotora de hongos, el promotor del gen *glaA* en *Aspergillus* o el promotor de los genes *cbh1* o *egl1* en *Trichoderma*.

[164] También se contempla que más de una copia de ADN que codifica una variante de CBH2 puede ser recombinada en la cepa para facilitar la sobreexpresión. El ADN que

- codifica la variante de CBH2 se puede preparar por medio de la construcción de un vector de expresión que lleva el ADN que codifica la variante. El vector de expresión que lleva el fragmento de ADN insertado que codifica la variante de CBH2 puede ser cualquier vector que sea capaz de replicarse autónomamente en un organismo huésped
- 5 dado o de integrarse en el ADN del huésped, típicamente un plásmido. En incorporaciones preferidas, se contemplan dos tipos de vectores de expresión para obtener expresión de genes. El primero contiene secuencias de ADN en las cuales el promotor, la región que codifica el gen, y la secuencia terminadora todas se originan del gen a ser expresado. La interrupción del gen se puede obtener cuando se desee al
- 10 eliminar las secuencias de ADN no deseadas (por ejemplo, codificación para dominios indeseados) para dejar que el dominio sea expresado bajo control de sus propias secuencias regulatorias de transcripción y traslación. Un marcador seleccionable puede también estar contenido en el vector que permite la selección para la integración en el huésped de múltiples copias de las secuencias de gen novedosas.
- 15 [165] El segundo tipo de vector de expresión es preensamblado y contiene las secuencias requeridas para la transcripción de alto nivel y un marcador seleccionable. Se contempla que la región de codificación para un gen o parte del mismo se puede insertar en este vector de expresión de propósito general de modo tal que se encuentra bajo control de transcripción del promotor de módulo de expresión y las secuencias
- 20 terminadoras.
- [166] Por ejemplo, en *Aspergillus*, el pRAX es dicho vector de expresión de propósito general. Los genes o parte de los mismos se pueden insertar hacia abajo del promotor fuerte glaa.
- [167] Por ejemplo, en *Hypocrea*, el pTEX es dicho vector de expresión de propósito
- 25 general. Los genes o parte de los mismos se pueden insertar hacia abajo del promotor fuerte cbhl.
- [168] En el vector, la secuencia de ADN que codifica la variante de CBH2 de la presente divulgación debe estar vinculada operablemente con las secuencias de transcripción y traslación, es decir, una secuencia de promotor adecuada y la secuencia
- 30 de señal en el marco de lectura al gen estructural. El promotor puede ser cualquier secuencia de ADN que muestra actividad de transcripción en la célula huésped y se puede derivar de genes que codifican las proteínas homólogas o heterólogas a la célula huésped. Un péptido de señal opcional suministra producción extracelular de la variante

de CBH2. EL ADN que codifica la secuencia de señal es preferiblemente aquel que se encuentra asociado naturalmente con el gen a ser expresado, sin embargo la secuencia de señal de cualquier fuente adecuada, por ejemplo una exo- celobiohidrolasa o endoglucanasa de *Trichoderma*, se contempla en la presente divulgación.

- 5 [169] Los procedimientos utilizados para ligar las secuencias de ADN que codifican para el CBH2 variante de la presente divulgación con el promotor, y la inserción en vectores adecuados son bien conocidos en el estado de la técnica.

- [170] El vector o construcción de ADN descrito anteriormente puede ser introducido en la célula huésped de acuerdo con técnicas conocidas tal como transformación,
10 transfección, microinyección, bombardeo biolístico y similares.

- [171] En la técnica de transformación preferida, se debe tener en cuenta que la permeabilidad de la pared celular a ADN en *Hyprocrea* sp. (*trichodemra* sp.) es muy baja. De acuerdo con esto, el consumo de la secuencia, gen o fragmento de gen deseado de ADN es mejor en el mínimo. Existe una pluralidad de métodos para aumentar la
15 permeabilidad de la pared celular de *Hyprocrea* sp. (*Trichoderma* sp.) en la cepa derivada (es decir, que carece de un gen funcional que corresponde al marcador seleccionable utilizado) antes del proceso de transformación.

- [172] El método preferido en la presente divulgación para preparar *Aspergillus* sp. o *Hyprocrea* sp. (*Trichoderma* sp.) para transformación involucra la preparación de
20 protoplastos a partir de hongo micelio. Ver Campbell y otros, Eficiencia de transformación mejorada de *A. niger* usando el gen homólogo de *niaD* a partir de esporas vegetales germinadas. El micelio es tratado con una enzima que digiere la pared celular resultante en los protoplastos. Los protoplastos son luego protegidos por la presencia de un estabilizador osmótico en el medio de suspensión. Estos estabilizadores
25 incluyen sorbitol, manitol, cloruro de potasio, sulfato de magnesio y similares. Usualmente, la concentración de estos estabilizadores varía entre 0.8 M y 1.2 M. Es preferible utilizar aproximadamente una solución de 1.2 M de sorbitol en el medio de suspensión.

- [173] El consumo del ADN en la cepa huésped, *Aspergillus* sp. o *Hyprocrea* sp. (*Trichoderma* sp.), depende de la concentración de iones de calcio. Generalmente se
30 utiliza entre aproximadamente 10 mM CaCl_2 y 50 mM CaCl_2 en una solución de consumo. Aparte de la necesidad por el ion de calcio en la solución de consumo, otros ítems generalmente incluidos son un sistema de búfer tal como búfer TE (10 Mm

Tris, pH 7.4; 1 mM EDTA) o búfer de 10 mM MOPS, pH 6.05 (ácido morfolinapropanosulfónico) y polietilenglicol (PEG). Se cree que el polietilenglicol actúa para fusionar las membranas celulares permitiendo así que el contenido del medio sea entregado en el citoplasma de la célula huésped, a modo de ejemplo la cepa

5 *Aspergillus* sp. o *Hyprocrea* sp., y el ADN plásmido se transfiere al núcleo. Esta fusión frecuentemente deja múltiples copias del ADN plásmido integradas en el cromosoma huésped.

[174] Usualmente, una suspensión que contiene los protoplastos de *Aspergillus* sp. o células que han sido sujetas a un tratamiento de permeabilidad a una adensidad de

10 $10^{5.5}$ a $10^{6.6}$ /mL, preferiblemente $2 \times 10^{5.5}$ /mL se utiliza en la transformación. De forma similar, una suspensión que contiene los protoplastos de *Hyprocrea* sp. (*Trichoderma* sp.) o células que han sido sujetas a un tratamiento de permeabilidad a una densidad de $10^{8.8}$ a $10^{9.9}$ /mL, preferiblemente $2 \times 10^{8.8}$ /mL se utiliza en la transformación. Un volumen de 100 μ L de estos

15 protoplastos o células en una solución apropiada (por ejemplo, 1.2M sorbitol; 50 mM CaCl₂) se mezcla con el ADN deseado. Generalmente, una concentración alta de PED es adicionada a la solución de consumo. Se puede adicionar de 0.1 a 1 de volumen de 25% PEG 4000 a la suspensión de protoplastos. Sin embargo, es preferible adicionar aproximadamente 0.25 volúmenes a la suspensión de protoplastos. Los aditivos tal

20 como dimetilsulfóxido, heparina, espermidina, cloruro de potasio y similares pueden también ser adicionados a la solución de consumo y ayudar en la transformación.

[175] Generalmente, la mezcla es luego incubada a aproximadamente 0 grados C por un periodo de entre 10 a 30 minutos. El PEG adicional es luego adicionando a la mezcla para además mejorar el consumo del gen deseado o la secuencia de ADN. El 25% de

25 PEG 4000 es preferiblemente aproximadamente 10 veces el volumen de la mezcla de transformación. Después que el PEG es adicionado, la mezcla de transformación es luego incubada bien sea a temperatura ambiente o sobre hielo antes de la adición de una solución de sorbitol y CaCl₂. La suspensión de protoplasto es luego adicionada a alícuotas fundidos de un medio de crecimiento. Este medio de crecimiento permite el

30 crecimiento de los transformantes únicamente. Cualquier medio de crecimiento puede ser utilizado en la presente divulgación que sea adecuado para crecer los transformantes deseados. Sin embargo, si los transformantes de *Pyr*.sup.+ son seleccionados es preferible utilizar un medio de crecimiento que no contenga uridina. Las colonias

subsecuentes son transferidas y purificas en un medio de crecimiento agotado de uridina.

[176] En esta etapa, los transformantes estables se pueden distinguir de los transformantes inestables por su tasa de crecimiento más rápida y, en *Trichoderma*, por ejemplo, la formación de colonias circulares con un contorno suave en lugar de imperfecto en el medio de cultivo sólido que carece de uridina. Adicionalmente, en algunos casos una prueba adicional de estabilidad se puede hacer al crecer los transformantes en el medio sólido no selectivo (es decir, que contiene uridina), cosechar las esporas de este medio de cultivo y determinar el porcentaje de estas esporas las cuales subsecuentemente germinarán y crecerán en el medio selectivo que carece de uridina.

[177] En una incorporación particular del método anterior, las variantes de CBH2 se recuperan en la forma activa a partir de la célula huésped después del crecimiento en el medio líquido como resultado del proceso apropiado de translación posterior de la variante de CBH2.

(ii) Levadura

[178] La presente divulgación también contempla el uso de levadura como una célula huésped para la producción de CBH2. Otros varios genes que codifican las enzimas hidrolíticas se han expresado en varias cepas de la levadura *S. cerevisiae*. Estos incluyen secuencias que codifican dos endoglucanasas (Penttilä y otros, *Levadura* vol. 3, pp 175 a 185, 1987), dos celobiohidrolasas (Penttilä y otros, *Gen*, 63: 103 a 112, 1988) y una beta-glucosidasa de *Trichoderma reesei* (Cummings y Fowler, *Curr. Genet.* 29:227 a 233, 1996), una xilanasa de *Aureobasidium pullulans* (Li y Ljungdahl, *Appl. Environ. Microbiol.* 62, no. 1, pp. 209 a 213, 1996), una alfa-amilasa de trigo (Rothstein y otros, *Gen*, 55:353 a 356, 1987), etc. Adicionalmente, un módulo de gen de celulasa que codifica endo-[beta]-1,4-glucanasa de *Butyrivibrio fibrisolvens* (END1), celobiohidrolasa de *Phanerochaete chrysosporium* (CBH1), celodextrinasa de *Ruminococcus flavefaciens* (CEL1) y la celobiasa de *Endomyces fibrilizer* (Bgl1) fue exitosamente expresado en una cepa de laboratorio de *S. cerevisiae* (Van Rensburg y otros, *Levadura*, vol. 14, pp. 67 a 76, 1998).

C. Introducción de una Secuencia de Ácido Nucleico que Codifica CBH2 en Células Huésped

- [179] La divulgación además suministra células y composiciones celulares que han sido modificadas genéticamente para comprender una secuencia de ácido nucleico que
- 5 codifica CBH2 variante suministrada exógenamente. Una célula madre o línea de célula puede ser modificada genéticamente (es decir, transducida, transformada o transfectada) con un vector de clonación o un vector de expresión. El vector puede estar, por ejemplo, en la forma de un plásmido, una partícula viral, un fago, etc., como se describirá más adelante.
- 10 [180] Los métodos para la transformación de la presente divulgación pueden resultar en la integración estable de todo o parte del vector de transformación en el genoma del hongo filamentoso. Sin embargo, también se contempla la transformación que resulta en el mantenimiento de un vector de transformación extra-cromosomal auto-replicante.
- [181] Varios métodos estándar de transfección se pueden utilizar para producir líneas
- 15 celulares de *Trichoderma reesei* que expresan grandes cantidades de la proteína heteróloga. Algunos de los métodos publicados para la introducción de construcciones de ADN en cepas que producen celulasa de *Trichoderma* incluyen Lorito, Hayes, DiPietro y Harman, 1993, *Curr. Genet.* 24: 349 a 356; Goldman, VanMontagu y Herrera-Estrella, 1990, *Curr. Genet.* 17:169 a 174; Penttilä, Nevalainen, Ratto, Salminen
- 20 y Knowles, 1987, *Gen* 6: 155 a 164, para *Aspergillus* Yelton, Hamer y Timberlake, 1984, *Proc. Natl. Acad. Sci. USA* 81: 1470 a 1474, para *Fusarium* Bajar, Podila y Kolattukudy, 1991, *Proc. Natl. Acad. Sci. USA* 88: 8202 a 8212, para *Streptomyces* Hopwood y otros, 1985, La Fundación John Innes, Norwich, UK y para *Bacillus* Brigidi, DeRossi, Bertarini, Riccardi y Matteuzzi, 1990, *FEMS Microbiol. Lett.* 55: 135
- 25 a 138).
- [182] Otros métodos para introducir una construcción de ácido nucleico heterólogo (vector de expresión) en hongos filamentosos (por ejemplo, *H. jecorina*) incluyen, pero no se limitan al uso de un arma partícula o de gen, permeabilización de paredes celulares de hongos filamentosos antes del proceso de transformación (por ejemplo,
- 30 mediante el uso de altas concentraciones de álcali, por ejemplo, 0.05 M a 0.4 M CaCl₂ o acetato de litio), fusión de protoplasto o transformación mediada de *Agrobacterium*. Un método ejemplar para la transformación de hongos filamentosos por medio del tratamiento de protoplastos o esferoplastos con polietilenglicol y CaCl₂

se describe en Campbell, E. I. y otros, Curr. Genet. 16:53 a 56, 1989 y Penttila, M. y otros, Gen, 63:11 a 22, 1988.

[183] Se puede utilizar cualquiera de los procedimientos bien conocidos para introducir secuencias externas de nucleótido en células huésped. Estos incluyen el uso de

5 transfección de fosfato de calcio, polibreno, fusión de protoplasto, electroporación, biolística, liposomas, microinyección, vectores plasma, vectores virales y cualquiera de los otros métodos bien conocidos para introducir ADN genómico clonado, cADN, ADN sintético u otro material genético externo en una célula huésped (ver, por ejemplo, Sambrook y otros, supra). También se utiliza el método de transfección mediado con

10 Agrobacterium descrito en la Patente de los Estados Unidos No. 6.255.115. Sólo es necesario que el procedimiento de ingeniería de genética particular utilizado sea capaz de introducir satisfactoriamente por lo menos un gen en la célula huésped capaz de expresar el gen heterólogo.

[184] Adicionalmente, las construcciones de ácido nucleico heterólogo que comprende

15 una secuencia de ácido nucleico que codifica CBH2 variante se puede transcribir in vitro, y el ARN resultante introducido en la célula huésped por medio de métodos bien conocidos, por ejemplo, por inyección.

[185] La divulgación además incluye transformantes novedosos y útiles de hongos filamentosos tal como H. jecorina y A. niger para su uso en la producción de

20 composiciones de celulasa de hongo. La divulgación incluye transformantes de hongos filamentosos que comprenden la secuencia que codifica el CBH2 variante, o la eliminación de la secuencia que codifica cbh endógeno.

[186] Seguido a la introducción de una construcción de ácido nucleico heterólogo que comprende la secuencia de codificación para un cbh2 variante, las células modificadas

25 genéticamente pueden ser cultivadas en medios de nutrientes convencionales como sea apropiado para activar promotores, seleccionar transformantes o amplificar la expresión de una secuencia de ácido nucleico que codifica el CBH2 variante. Las condiciones de cultivo, tales como temperatura, pH y similares, son aquellas previamente utilizadas para la célula huésped seleccionada para expresión, y serán evidentes para los expertos

30 en el estado de la técnica.

[187] La progenie de células en las cuales dichas construcciones de ácido nucleico heterólogo han sido introducidas son generalmente consideradas por comprender la

secuencia de ácido nucleico que codifica CBH2 variante encontrada en la construcción de ácido nucleico heterólogo.

[188] La divulgación además incluye transformantes novedosos y útiles de hongos filamentosos tal como *H. jecorina* para su uso en la producción de composiciones de celulasa de hongo. La *Aspergillus niger* puede también ser utilizada en la producción del CBH2 variante. La divulgación incluye transformantes de hongos filamentosos, especialmente hongos que comprenden la secuencia que codifica el *cbh2* variante, o la eliminación de la secuencia que codifica el *cbh2* endógeno.

[189] Los transformantes estables de hongos filamentosos pueden generalmente distinguirse de los transformantes inestables por su tasa de crecimiento más rápida y, en *Trichoderma*, por ejemplo, la formación de colonias circulares con un contorno más suave en lugar de imperfecto en el medio de cultivo sólido. Adicionalmente, en algunos casos, una prueba adicional de estabilidad se puede realizar al crecer los transformantes en un medio sólido no selectivo, cosechar las esporas de este medio de cultivo y determinar el porcentaje de estas esporas las cuales subsecuentemente germinarán y crecerán en el medio selectivo.

VI. Aislamiento y Purificación de Proteína de CBH2 Recombinante

[190] En general, una proteína de CBH2 variante producida en cultivo celular es secretada en el medio y puede ser purificada o aislada, por ejemplo, al remover los componentes indeseados del medio de cultivo celular. Sin embargo, en algunos casos, una proteína de CBH2 variante puede ser producida en una forma celular que necesita recuperación de un lisado celular. En dichos casos, la proteína de CBH2 variante es purificada de las células en las cuales se produjo utilizando técnicas empleadas rutinariamente por los expertos en el estado de la técnica. Los ejemplos incluyen, pero no se limitan a, cromatografía de afinidad (Goyal y otros, *Tecnol. Biorecursos* 36:37 a 50, 1991; Fliess y otros, *Eur. J. Appl. Microbiol. Biotechnol.* 17:314 a 318, 1983; Bhikhabhai y otros, *J. Appl. Biochem.* 6:336 a 345, 1984; Ellouz y otros, *J. Chromatografía* 396:307 a 317, 1987), incluyendo intercambio de iones que utiliza materiales con alto poder de resolución (Medve y otros, *J. Chromatografía A* 808: 153 a 165, 1998), cromatografía de interacción hidrofóbica (Tomaz y Queiroz, *J. Chromatografía A* 865: 123 a 128, 1999); y particionamiento de dos fases (Brumbauer y otros, *Bioseparación* 7:287 a 295, 1999).

- [191] Típicamente, la proteína de CBH2 variante es fraccionada para segregar las proteínas que tienen propiedades seleccionadas, tal como afinidad de unión a agentes de unión particulares, por ejemplo, anticuerpos o receptores; o los cuales tienen un rango de peso molecular seleccionado; o rango de puntos isoelectrónicos.
- 5 [192] Una vez se logra la expresión de una proteína de CBH2 variante dada, la proteína de CBH2 producida de esta forma se purifica de las células o el cultivo celular. Los procedimientos ejemplares adecuados para dicha purificación incluyen los siguientes: cromatografía de columna de afinidad a anticuerpo, cromatografía de intercambio de iones; precipitación a etanol; HPLC de fase reversa; cromatografía en silicio o en una 10 resina de intercambio de cationes tal como DEAE; cromatofocado; SD-PAGE; precipitación de sulfato de amonio; y filtración de gel usando, por ejemplo, Sephadex G-75. Varios métodos para la purificación de proteína pueden ser empleados y dichos métodos son conocidos en el estado de la técnica y descritos por ejemplo en Deutscher, Métodos en Enzimología, vol. 182, No. 57, pp. 779, 1990; Scopes, Métodos de 15 Enzimol. 90: 479 a 91, 1982. Los pasos de purificación seleccionados dependerán de, por ejemplo, la naturaleza del proceso de producción utilizado y la proteína particular producida.

VII. Utilidad de cbh2 y CBH2

- 20 [193] Se puede apreciar que los ácidos nucleicos de cbh variante, la proteína de CBH2 variante y las composiciones que comprenden actividad de proteína de CBH2 variante encuentran utilidad en una amplia variedad de aplicaciones, algunas de las cuales se describen más adelante.
- [194] Las nuevas y mejoradas composiciones de celulasa que comprenden diferentes 25 cantidades de celulasas tipo BG, tipo EG y tipo CBH variante encuentran utilidad en composiciones detergentes que exhiben habilidad de limpieza mejorada, funcionan como un agente suavizante y/o mejoran la textura de las telas de algodón (por ejemplo, "lavado por piedra" o "biopulido"), en composiciones para degradar pulpa de madera en azúcares (por ejemplo, por producción de bioetanol), y/o en composiciones alimenticias.
- 30 El aislamiento y la caracterización de la celulasa de cada tipo suministran la habilidad de controlar los aspectos de dichas composiciones.

[195] En un acercamiento, la celulasa de la divulgación encuentra utilidad en composiciones detergentes o en el tratamiento de telas para mejorar la textura y la apariencia.

[196] Dado que la tasa de hidrólisis de los productos celulósicos puede aumentar al usar
5 un transformante que tiene por lo menos una copia adicional del gen cbh insertado en el genoma, los productos que contienen celulosa o heteroglicanos pueden ser degradados a una tasa más rápida y a un grado mayor. Los productos hechos de celulosa tal como papel, algodón, pañales de celulosa y similares, pueden ser degradados más eficientemente en un relleno sanitario. Así, el producto de fermentación obtenible a
10 partir de los transformantes o los transformantes solos puede ser utilizado en composiciones para ayudar a degradar por licuefacción una variedad de productos celulósicos que se adicionan a los rellenos sanitarios superpoblados.

[197] La sacarificación y fermentación separadas es un proceso por el cual la celulosa presente en biomasa, por ejemplo, rastrojo de maíz, se convierte en glucosa y
15 subsecuentemente las cepas de levadura convierten la glucosa en etanol. La sacarificación y fermentación simultáneas es un proceso por el cual la celulosa presente en biomasa, por ejemplo, rastrojo de maíz, se convierte en glucosa y, al mismo tiempo y en el mismo reactor, las cepas de levadura convierten la glucosa en etanol. Así, en otro acercamiento, la celulasa de tipo CBH variante de la divulgación encuentra utilidad en
20 la degradación de biomasa en etanol. La producción de etanol a partir de fuentes fácilmente disponibles de celulosa suministra una fuente de combustible estable y renovable.

[198] Las materias primas a base de celulosa se componen de desechos agrícolas, hierbas y maderas y otra biomasa de bajo valor tal como desechos municipales (por
25 ejemplo, papel reciclado, recortes de jardín, etc.). El etanol se puede producir a partir de la fermentación de cualquiera de estas materias primas celulósicas. Sin embargo, la celulosa debe primero convertirse en azúcares antes que pueda haber conversión a etanol.

[199] Una amplia variedad de materias primas se pueden utilizar con el CBH variante
30 inventivo y el seleccionado para su uso puede depender de la región donde se realiza la conversión. Por ejemplo, en el Medio Oeste de los Estados Unidos los desechos agrícolas tal como la paja de trigo, rastrojo de maíz y bagasa pueden predominar mientras que en California la paja de arroz puede predominar. Sin embargo, se debe

entender que cualquier biomasa celulósica disponible puede ser utilizada en cualquier región.

[200] Los métodos de la presente divulgación pueden ser utilizados en la producción de monosacáridos, disacáridos, y polisacáridos como materias primas químicas o de fermentación para microorganismos para la producción de productos orgánicos, químicos y combustibles, plásticos, y otros productos o intermedios. En particular, el valor de procesar residuos (granos secos de destilería, granos de elaboración de cerveza, bagazo de caña, etc.) puede aumentar por medio de solubilización parcial o completa de celulosa o hemicelulosa. Adicional a etanol, algunos químicos que se pueden producir de celulosa y hemicelulosa incluyen, acetona, acetato, glicina, lisina, ácidos orgánicos (por ejemplo, ácido láctico), 1,3-propanediol, butanediol, glicerol, etilenglicol, furfural, polihidroxicanoatos, cis, ácido mucónico cis, alimento de animales y xilosa.

[201] Una composición de celulasa que contiene una cantidad mejorada de celobiohidrolasa encuentra utilidad en la producción de etanol. El etanol a partir de este proceso puede ser utilizado adicionalmente como un mejorador de octanos o directamente como un combustible en lugar de gasolina lo cual es ventajoso ya que el etanol como fuente de combustible es más amigable con el medio ambiente que los productos derivados de petróleo. Es conocido que el uso de etanol mejorará la calidad del aire y posiblemente reduce los niveles locales de ozono y contaminación. Además, la utilización de etanol en lugar de gasolina puede ser de importancia estratégica en la amortiguación del impacto de cambios repentinos en energía no renovable y suministros petroquímicos.

[202] El etanol puede ser producido por medio de procesos de sacarificación y fermentación de biomasa celulósica tal como árboles, plantas herbáceas, desechos municipales sólidos y residuos agrícolas y forestales. Sin embargo, la proporción de enzimas individuales de celulasa dentro de una mezcla de celulasa que ocurre naturalmente producida por un microbio puede no ser la más eficiente para conversión rápida de celulosa en biomasa a glucosa. Es conocido que las endoglucanasas actúan para producir nuevos extremos de cadena de celulosa los cuales por si solos son sustratos para la acción de celobiohidrolasas y de esta forma mejoran la eficiencia de la hidrólisis de todo el sistema de celulasa. Por lo tanto, el uso de actividad mejorada u optimizada de celobiohidrolasa puede mejorar ampliamente la producción de etanol.

[203] Así, la celobiohidrolasa inventiva encuentra uso en la hidrólisis de celulosa a sus componentes de azúcares. En una incorporación, una celobiohidrolasa variante es adicionada a la biomasa antes de la adición de un organismo de fermentación. En una segunda incorporación, una celobiohidrolasa variante es adicionada a la biomasa al mismo tiempo que un organismo de fermentación. Opcionalmente, puede haber otros componentes de celulosa presentes en cualquier incorporación.

[204] En otra incorporación, la materia prima celulósica puede ser pretratada. El pretratamiento puede ser por medio de temperatura elevada y la adición de ácido diluido, ácido concentrado o solución álcali diluida. La solución de pretratamiento se adiciona por un tiempo suficiente para por lo menos hidrolizar parcialmente los componentes de hemicelulosa y luego neutralizados.

[205] El principal producto de la acción de CBH2 en celulosa es celobiosa la cual está disponible para conversión a glucosa por actividad BG (por ejemplo en un producto de celulosa de hongo). Bien sea por el pretratamiento de la biomasa celulósica o por la acción enzimática en la biomasa, otros azúcares, adicionales a glucosa y celobiosa, se pueden hacer a partir de biomasa. El contenido de hemicelulosa de la biomasa se puede convertir (por hemicelulasas) en azúcares tal como xilosa, galatosa, manosa y arabinosa. Así, en un proceso de conversión de biomasa, la sacarificación enzimática puede producir azúcares que son disponibles para conversiones biológicas o químicas a otros intermedios o productos finales. Por lo tanto, los azúcares generados a partir de biomasa encuentran uso en una variedad de procesos adicionales a la generación de etanol. Los ejemplos de dichas conversiones son fermentación de glucosa a etanol (como revisado por M. E. Himmel y otros, pp 2 a 45, en "Combustibles y Químicos a partir de Biomasa", Simposio ACS Serie 666, ed B. C. Saha y J. Woodward, 1997) y otras conversiones biológicas de glucosa a 2,5-diceto-D-gluconato (Patente de los Estados Unidos No. 6.599.722), ácido láctico (R. Datta y S-P, Tsai pp 224 a 236, *ibid*), succinato (R. R. Gokarn, M. A. Eiteman y J. Sridhar pp 237 a 263, *ibid*), 1,3-propanediol (A- P. Zheng, H. Biebl y W-D. Deckwer pp 264 a 279, *ibid*), 2,3-butanediol (C.S. Gong, N. Cao y G. T. Tsao pp 280 a 293, *ibid*), y las conversiones químicas y biológicas de xilosa a xilitol (B. C. Saha y R. J. Bothast pp 307 a 319, *ibid*). Ver también, por ejemplo, WO 98/21339.

[206] Las composiciones de detergente de esta divulgación pueden emplear aparte de la composición de celulasa (independientemente del contenido de celobiohidrolasa, es

decir, libre de celobiohidrolasa, sustancialmente libre de celobiohidrolasa, o mejorada con celobiohidrolasa), un surfactante, incluyendo surfactantes aniónicos, no iónicos y anfólicos, una hidrolasa, agentes de construcción, agentes blanqueadores, agentes azulantes y colorantes fluorescentes, inhibidores de aglomeración, solubilizantes, surfactantes catiónicos y similares. Todos estos componentes son conocidos en el estado de la técnica de los detergentes. La composición de celulasa como se describió anteriormente puede ser adicionada a la composición detergente en un diluyente líquido, en gránulos, en emulsiones, en gel, en pastas, y similares. Dichas formas son bien conocidas en el campo del experto. Cuando una composición sólida de detergente es empleada, la composición de celulasa es preferiblemente formulada como gránulos. Preferiblemente, los gránulos pueden ser formados con el fin de contener un agente protector de celulasa. Para una discusión más completa, ver la Patente de los Estados Unidos No. 6.162.782 titulada "Composiciones detergentes que contienen composiciones de celulasa deficientes en componentes de tipo CBH2", la cual se incorpora acá por referencia.

[207] Preferiblemente, las composiciones de celulasa se emplean de aproximadamente 0.0005 por ciento en peso a aproximadamente 5 por ciento en peso con relación a la composición detergente total. Más preferiblemente, las composiciones de celulasa son empleadas de aproximadamente 0.0002 por ciento en peso a aproximadamente 2 por ciento en peso con relación a la composición detergente total.

[208] Adicional, la secuencia de ácido nucleico de CBH2 variante encuentra utilidad en la identificación y caracterización de secuencias de ácido nucleico relacionadas. Una cantidad de técnicas útiles para determinar (predecir o confirmar) la función de los genes relacionados o los productos de genes incluyen, pero no se limitan a, (A) análisis de ADN/ARN, tal como (1) sobreexpresión, expresión ectópica, y expresión en otras especies; (2) eliminación de genes (genética inversa, eliminación objetivo, silenciamiento de gen inducido por virus (VIGS, ver Baulcombe, 100 Años de Virología, Calisher y Horzinek eds., Springer-Verlag, New York, NY 15:189 a 201, 1999); (3) análisis del estado de metilación del gen, especialmente regiones regulatorias de acompañamiento; y (4) hibridación in situ; (B) análisis de producto de gen tal como (1) expresión de proteína recombinante; (2) producción de antisueros, (3) inmunolocalización; (4) ensayos biológicos para catálisis u otra actividad; (5) estado de fosforilación; y (6) interacción con otras proteínas por medio de análisis doble híbrido

- de levadura; (C) análisis de camino, tal como ubicar un gen o producto de gen dentro de un camino bioquímico o de señalización particular con base en su fenotipo de sobreexpresión o por homología de secuencia con genes relacionados; y (D) otros análisis los cuales también pueden ser llevados a cabo para determinar o confirmar la participación del gen aislado y su producto en un camino particular metabólico o de señalización, y ayudar a determinar la función del gen.

EXPERIMENTAL

- [209] La presente divulgación se describe en mayor detalle en los siguientes ejemplos, los cuales no pretenden de ninguna forma limitar el alcance de la divulgación como se reivindica. Las figuras adjuntas deben ser consideradas como partes integrales de la descripción de la divulgación. Los siguientes ejemplos son ilustrativos, pero no limitan la divulgación reivindicada.

- [210] En la divulgación experimental que sigue, se aplican las siguientes abreviaciones:
- M (molar); mM (milimolar); μ M (micromolar); nM (nanomolar); mol (moles); mmol (milimoles); μ mol (micromoles); nmol (nanomoles); gm (gramos); mg (miligramos); μ g (microgramos); pg (picogramos); L (litros); ml y mL (mililitros); μ l y μ L (microlitros); cm (centímetros); mm (milímetros); μ m (micrómetros); nm (nanómetros); U (unidades); V (voltios); MW (peso molecular); sec (segundos); min(s) (minuto/minutos); h(s) y hr(s) (hora/horas); $^{\circ}$ C (grados Centígrados); QS (cantidad suficiente); ND (no hecho); NA (no aplica); rpm (revoluciones por minuto); H₂O (agua); dH₂O (agua desionizada); HCl (ácido clorhídrico); aa (aminoácido); bp (base par); kb (kilobase par); kD (kilodaltones); cADN (ADN copia o complementario); ADN (ácido desoxirribonucleico); ssADN (ADN de filamento simple); dsADN (ADN de filamento doble); dNTP (trifosfato desoxirribonucleotido); ARN (ácido ribonucleico); MgCl₂ (cloruro de magnesio); NaCl (cloruro de sodio); w/v (peso a volumen); v/v (volumen a volumen); g (gravedad); OD (densidad óptica); CNPG (cloro-nitro- fenil-beta-D-glucósido); CNP (2-cloro-4-nitrofenol); APB (bagazo de ácido pretratado); PASC (celulosa hinchada de ácido fosfórico); PCS (rastroy de maíz de ácido pretratado); Pi or PI (índice de desempeño); HPLC (cromatografía líquida de alta presión); PAGE (electroforesis de gel de poliacrilamida); PCR (reacción en cadena de polimerasa); RT-PCR (PCR de transcripción inversa); y SEL (librería de evaluación de sitio).

EJEMPLO 1

Ensayos

[211] Los siguientes ensayos fueron utilizados en los ejemplos descritos más adelante. Cualquier desviación de los protocolos provistos a continuación se indica en los
5 ejemplos. En estos experimentos, se utilizó un espectrofotómetro para medir la absorbencia de los productos formados después de la terminación de las reacciones.

A. Ensayo de Hexoquinasa para la Medición de Glucosa Residual

[212] La glucosa residual de sobrenadantes de cultivo *H.jecorina* que expresan variantes
10 de CBH2 fue medida usando un ensayo de hexoquinasa. Un volumen de 5 µl de sobrenadante se adicionó a 195 µl glucosa-hexoquinasa en un ensayo (Laboratorio de Instrumentación, Breda, Holanda) en una placa de microtitulación 96 (Costar Flat Bottom PS). Las placas fueron incubadas a temperatura ambiente por 15 min. Seguido a la incubación, la absorbencia se midió a 340 nm OD. Los sobrenadantes de los cultivos
15 que expresan glucosa residual fueron excluidos de la puesta en común para estudios adicionales.

B. Ensayo HPLC para la Determinación de Contenido de Proteína

[213] La concentración de proteínas variantes de CBH2 de los sobrenadantes agrupados
20 de cultivo se determinó utilizando un Agilent 1100 (Hewlett Packard) HPLC equipado con una columna Proswift RB 2H (Dionex). Diez microlitros de muestra, mezclados con 50 µl de 10% de acetonitrilo en agua desmineralizada filtrada se inyectó seguido al equilibrio de la columna de HPLC con 10% de acetonitrilo que contiene 0,01% de ácido trifluoroacético. Los compuestos se eluyeron utilizando un gradiente de 10% a 30% de
25 acetonitrilo de 0,3 min a 1 min, seguido de un gradiente de 30% a 65% de 1 min a 4 min. Las concentraciones de proteína de variantes de CBH2 se determinó a partir de una curva de calibración generada utilizando CBH2 de tipo salvaje purificado (6.25, 12.5, 25, 50 µg/ml). Para calcular el índice de rendimiento (P_i o PI), la proporción de la proteína total (promedio) producida por una variante y la proteína total (promedio)
30 producida por la de tipo salvaje a la misma dosis fueron promediadas.

C. Determinación de la Actividad Específica por medio de Ensayo de Hidrólisis de Celulosa Hinchada de Ácido Fosfórico (PASC)

[214] Hidrólisis de celulasa: La celulosa hinchada de ácido fosfórico (PASC) fue preparada de Avicel de acuerdo con un método publicado (Walseth, Tappi 35:228, 1971; y Wood, Biochem J, 121:353 a 362, 1971). Este material se diluyó con búfer y agua para lograr una mezcla de 1% w/v de tal manera que la concentración final de acetato de sodio fue 50 mM, pH 5.0. Cien microlitros de una suspensión al 1% de PASC en 50 mM de búfer de acetato de sodio (pH 5.0) se entregó en una placa microtitulación de 96 agujeros (Costar Flat Bottom PS). Diez microlitros de 5 mg/ml de sobrenadante de cultivo de una cepa eliminada de CBH2 fueron introducidos en el PASC, y 5, 10, 15, o 20 μ l de sobrenadantes de cultivo de células *H. jecorina* que expresan CBH2 de tipo salvaje o variantes de CBH2 se añadieron a la misma. La eliminación del gen de CBH2 de *Hypocrea jecorina* (también conocida como *Tricoderma reesei*) se describe en las Patentes de los Estados Unidos No. 5.861.271 y 5.650.322. Volúmenes de compensación del búfer de acetato fueron adicionados para compensar las diferencias en el volumen total. La placa de microtitulación fue sellada e incubada en un incubador termostizado a 50°C bajo condiciones de agitación a 900 rpm. Después de dos horas, la reacción de hidrólisis se detuvo por medio de la adición de 100 μ l de búfer de glicina, pH 10 en cada agujero o pozo. Para calcular el índice de rendimiento (P_i o PI), la proporción del azúcar total (promedio) producido por una variante y el azúcar total (promedio) producido por la de tipo salvaje a la misma dosis fueron promediadas. Los productos de reacción de hidrólisis se analizaron con el ensayo PAHBAH. Ensayo PAHBAH: Alícuotas de 150 μ l de PAHBAH que reducen el reactivo de azúcar (5% w/v de hidrazida de ácido p-hidroxibenzoico (PAHBAH, Sigma # H9882), disuelto en 0.5 N HCl), (Lever, Anal. Biochem, 47:273 a 279, 1972) fueron adicionados a todos los agujeros o pozos de una placa de microtitulación vacía. Diez microlitros de los sobrenadantes de reacción de hidrólisis fueron adicionados a la placa de reacción PABA. Todas las placas fueron selladas e incubadas a 69°C bajo agitación continua de 900 rpm. Después de una hora las placas se ubicaron en hielo por cinco minutos y se centrifugaron a 720 x g a temperatura ambiente por cinco minutos. Las muestras de 80 μ L de las mezclas de reacción PAHBAH desarrolladas se transfirieron a una placa fresca y se midió la absorbencia a 410 nm en un espectrofotómetro. Se incluyó una celobiosa estándar como control. Una curva de respuesta a dosis se generó para la proteína CBH2 tipo salvaje.

D. Determinación de la Actividad Específica por medio de Hidrólisis de Rastrojo de Maíz Pretratado con Ácido Diluido

[215] Rastrojo de Maíz Pretratado (PCS): El rastrojo de maíz fue pretratado con 2% w/w de H_2SO_4 como se describió (Schell y otros, J. Appl Biochem Biotechnol, 105:69 a 86, 2003) y seguido por múltiples lavados con agua desionizada para obtener una pasta que tiene un pH de 4.5. El búfer de acetato de sodio (pH 5.0) fue luego adicionado a una concentración final de 50 mM de acetato de sodio y, si es necesario, esta mezcla fue luego titulada a pH de 5.0 utilizando 1N NaOH. La concentración de celulasa en la mezcla de reacción fue de aproximadamente 7%. Sesenta y cinco microlitros de esta suspensión de celulosa fueron adicionados por agujero a una placa de microtitulación de 96 agujeros (Nunc Flat Bottom PS). Diez microlitros de un sobrenadante de cultivo de 5 mg/ml de una cepa eliminada de CBH2 se adicionaron al PCS, y 5, 10, 15, o 20 μ l de los sobrenadantes de cultivo de células de *H. jecorina* que expresan CBH2 de tipo salvaje o variantes de CBH2 fueron adicionadas a la misma. Después de sellar la placa, las placas se ubicaron en una incubadora termostizada a 50°C bajo condiciones de agitación a 1300 rpm por 5 minutos. Las placas fueron luego incubadas a 50°C mientras que se agitaba a 200 rpm bajo 80% de humedad para evitar el secado. Después de 7 días las placas se pusieron en hielo por 5 min y la reacción de hidrólisis se detuvo por la adición de 100 μ l de búfer de glicina, pH 10 a cada agujero. Los productos de reacción de hidrólisis fueron analizados con el ensayo PAHBAH. Para calcular el índice de rendimiento (P_i o PI), la proporción del azúcar total (promedio) producido por una variante y el azúcar total (promedio) producido por la de tipo salvaje a la misma dosis fueron promediadas. Ensayo PAHBAH: Alícuotas de 150 μ l de PAHBAH que reducen el reactivo de azúcar (5% w/v de hidrazida de ácido p-hidroxibenzoico (PAHBAH, Sigma # H9882), disuelto en 0.5 N HCl), (Lever, Anal. Biochem, 47:273 a 279, 1972) fueron adicionados a todos los agujeros o pozos de una placa de microtitulación vacía. Diez microlitros de los sobrenadantes de reacción de hidrólisis fueron adicionados a la placa de reacción PAHBAH. Todas las placas fueron selladas e incubadas a 69°C bajo agitación continua de 900 rpm. Después de una hora las placas se ubicaron en hielo por cinco minutos y se centrifugaron a 720 x g a temperatura ambiente por cinco minutos. Las muestras de 80 μ L de las mezclas de reacción PAHBAH desarrolladas se transfirieron a una placa fresca y se midió la absorbencia a 410 nm en un

espectrofotómetro. Se incluyó una celobiosa estándar como control. Una curva de respuesta a dosis se generó para la proteína CBH2 tipo salvaje.

E. Determinación de la Actividad Específica por medio de Hidrólisis de Mazorca de Maíz Pretratada (CC) con Amoniaco

[216] La mazorca de maíz fue molida para pasar a través de un filtro de 0.9 mm, luego fue pretratada como se describió en el Ejemplo 4 de la solicitud de patente WO2006110901 (incorporada acá por referencia para este método). La mazorca de maíz pretratada fue utilizada como una suspensión de 7% de celulosa en 50 mM de acetato de sodio pH 5.0. Sesenta y cinco microlitros de la suspensión se adicionaron por agujero a una placa de microtitulación de 96 agujeros (Nunc Flat Bottom PS). A cada agujero, se adicionaron 20 µl de 3.4 mg/ml de sobrenadante de una cepa (Δ egl1, Δ egl2, Δ cbh1, Δ cbh2) complementado con 0.23 mg/ml de glucosidasa 1 purificada de *T. reesei* (Cel3 A). Se adicionaron veinte o cuarenta microlitros de sobrenadantes de cultivo de células *H. jecorina* que expresan CBH2 tipo salvaje o variantes de CBH2. Los volúmenes de compensación del búfer de acetato fueron adicionados para compensar las diferencias en el volumen total. La placa fue incubada a 50°C mientras que se agitaba a 220 rpm bajo 80% de humedad para evitar el secado. Después de 3 días la placa fue puesta en hielo por 5 min y se adicionaron 100 µl de 100 mM de glicina pH 10.0. Después de mezclar, la placa fue centrifugada a 3000 rpm por 5 min. Un volumen de 10 µl de sobrenadante fue diluido en 190 µl de agua. Veinte µl de la solución diluida se transfirieron a una nueva placa de microtitulación de 96 agujeros (Costar Flat Bottom PS) que contiene 100 µl de mezcla de ensayo de glucosa ABTS (2.74 mg/ml de 2,2'-azino-bis(3-etilbenzo-tiazolina-6-ácido sulfónico, 0.333 U/ml de peroxidasa de rábano picante tipo VI, 1 U/ml de oxidasa de glucosa) y un aumento en A_{420} fue registrado en un espectrofotómetro de placa de microtitulación (Spectramax Plus 384, Dispositivos Moleculares). Un rango de concentraciones de glucosa se incluyó como un estándar en cada placa (6.25, 12.5, 25, 50, 100, 200, 400 mM). Los ensayos fueron hechos en duplicado. Una curva de respuesta de dosis se generó para el CBH2 tipo salvaje al ajustar los datos con la ecuación de Langmuir ($y = (x*a)/(x+b)$) y las actividades de las variantes CBH2 se dividieron por una actividad calculada de CBH2 tipo salvaje de la misma placa para producir un índice de rendimiento.

F. Estabilidad de Variantes de CBH2 en Presencia de Etanol

[217] La estabilidad del CBH2 de tipo salvaje y las variantes de CBH2 se probó en presencia de 4.5% de etanol (EtOH) a 49°C. Los sobrenadantes de cultivo agrupados (80 µL) de células de *H. jecorina* que expresan variantes de CBH2 se adicionaron a una
 5 placa de 96 agujeros (Greiner V-bottom PS) que contiene 10 µl de 40.5% de EtOH por agujero. Las placas se sellaron y se incubaron en una incubadora con termostato a 49°C por 16 horas con una agitación de 900 rpm. Seguido a la incubación, las placas se ubicaron en hielo por 5 minutos. La actividad de CBH2 residual se determinó usando el ensayo de hidrólisis de celulosa hinchada de ácido fosfórico (PASC) como se describió
 10 anteriormente.

[218] Para calcular la actividad residual, el valor del producto formado por la adición de 5, 10, 15 y 20 µl de CBH2 incubado en EtOH al ensayo PASC de actividad residual se dividió por el valor del producto formado por la adición de 5, 10, 15 y 20 µl de CBH2 libre de EtOH al ensayo PASC. Los valores individuales de estas cuatro proporciones
 15 fueron luego promediados para dar la actividad residual promedio. Para determinar el valor PI para la variante, el valor de la actividad residual promedio para las variantes fue luego dividido por el promedio de los valores de actividad residual de los controles de CBH2 tipo salvaje.

20 G. Termoestabilidad de Variantes de CBH2

[219] La termoestabilidad del CBH2 tipo salvaje y los variantes de CBH2 se probó a 53°C. Los sobrenadantes de cultivo agrupados (80 µL) de células de *H. jecorina* que expresan variantes de CBH2 se adicionaron a una placa de 96 agujeros (Greiner V-bottom PS). Las placas se sellaron y se incubaron en una incubadora con termostato a
 25 53°C por 16 horas con una agitación de 900 rpm. Seguido a la incubación, las placas se ubicaron en hielo por 5 minutos. La actividad de CBH2 residual se determinó usando el ensayo de hidrólisis de celulosa hinchada de ácido fosfórico (PASC) como se describió anteriormente.

[220] Para calcular la actividad residual, el valor del producto formado por la adición de
 30 5, 10, 15 y 20 µl de CBH2 tratado con calor al ensayo PASC de actividad residual se dividió por el valor del producto formado por la adición de 5, 10, 15 y 20 µl de CBH2 sin calentar al ensayo PASC. Los valores individuales de estas cuatro proporciones fueron luego promediados para dar la actividad residual promedio. Para determinar el

valor PI para la variante, el valor de la actividad residual promedio para las variantes fue luego dividido por el promedio de los valores de actividad residual de los controles de CBH2 tipo salvaje.

5 **EJEMPLO 2**

Generación de librerías de evaluación de sitios y de combinaciones para la *Hipocrea jecorina* CBH2

[221] El plásmido pTTTPyr-cbh2 que contiene la secuencia que codifica la proteína Hipocrea jecorina CBH2 (SEQ ID NO: 1) se envió a BASECclear (Leiden, Países
10 Bajos), Gene Art AG (Regensburg, Alemania), y Sloning BioTechnology GmbH (Puchheim, Alemania) para la generación de librerías de evaluación de sitios (SELS). Se proporciona un mapa de la pTTTPyr-cbh2 en la figura 3. Se le solicitó a los proveedores que generaran librerías de posiciones en cada uno de los sitios de la proteína madura de Hipocrea jecorina CBH2 (SEQ ID NO: 3) que se muestra en la tabla 2-1. La secuencia
15 de aminoácidos de la proteína CBH2 de cadena completa se muestra en la SEQ ID NO:2.

[222] La SEQ ID NO: 1 establece la secuencia de referencia del ADN que codifica la Hipocrea jecorina CBH2:

atgattgtcggcattctcaccacgctggctacgctggccacactcgagctagtgtgcct
20 ctagaggagcggcaagcttgctcaagcgt
ctggggccaatgtgtgggccagaattggctcgggtccgacttgctgtgcttcggaagcac
atgcgtctactccaacgactattactccc
agtgtcttccggcgctgcaagctcaagctcgtccacgcgcgccgcgtcgacgactctc
gagtatccccacacatcccggtcga
25 gtcgcgcgacgcctccacctggttctactactaccagagtacctccagtcggatcgggaa
ccgctacgtattcaggcaaccctttgttg
gggtcactccttgggccaatgcatattacgcctctgaagttagcagcctcgctattccta
gcttgactggagccatggccactgctgcag
cagctgtgcgaaaggttccctctttatgtggctagatactcttgacaagacccctctca
30 tggagcaaaccttggccgacatccgcaccg
ccaacaagaatggcggttaactatgccggacagtttggtgtatgacttgccggatcgcg
attgcgctgcccttgctcgaatggcgaa
tactctattgccgatgggtggcgtcgccaaatataagaactatcgcacaccattcgtcaa

attgtcgtggaatatccgatatccggaccct
 cctggttattgagcctgactctcttgccaacctggtagcaaacctcggtagctccaaagtg
 tgccaatgctcagtcagcctaccttgagtgc
 atcaactacgccgtcacacagctgaaccttccaaatgttgcgatgtattggacgctggc
 5 catgcaggatggcttggtggccggcaaa
 ccaagacccggccgctcagctatttgcaaatgtttacaagaatgcatcgtctccgagagc
 tcttcgggattggcaaccaatgtcgcca
 actacaacgggtggaacattaccagccccccatcgtacacgcaaggcaacgctgtctaca
 acgagaagctgtacatccacgctattg
 10 gacctctcttgccaatcacggctgggtccaacgccttcttcactgatcaaggtegat
 cgggaaagcagcctaccggacagcaaca
 gtggggagactgggtgcaatgtgatcggcaccggatttggtattcgccatccgcaaacac
 tggggactcgttgctggattcgtttgtctg
 ggtcaagccaggcggcgagtgtgacggcaccagcagcagcagtgccgacgatttgactc
 15 ccactgtgcgctcccagatgccttgc
 aaccggcgccctcaagctgggtgcttggtccaagcctactttgtgcagcttctcacaacg caaacccatcgttcctgtaa.
 [223] La SEQ ID NO:2 establece la secuencia completa de la proteína de *Hipocrea jecorina* CBH2:
 MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYS
 20 NDY YSQCL
 PGAASSSSSTRAASTTSRVSP TTSRSSSATPPPGSTTTRVPPVSGTATYSGNPFVG
 VTP WA
 NAYYASEVSSLAIPSLTGAMATAAAVAKVPSFMWLDTLDKTPLMEQTLADIRT
 ANKNNG NY
 25 AGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLL
 VIEPDS LA
 NLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGW PANQ
 DPAAQLFA NV
 YKNASSPRALRGLATNVANYNGWNITSPPSYTQGNVYNEKLYIHAIGPLL ANH
 30 GWSNAF FI
 TDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGEC
 DGTSDSSA PR FDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL.

[224] la SEQ ID NO: 3 establece la secuencia de la proteína madura de Hipocrea jecorina CBH2:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAAST
TSRVSPPTS
5 RSSSATPPPGSTTTTRVPPVGSGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTG
AMATAA
AAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDC
AALASNGEYS
IADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLNLVTNLGTPKCANAQSAY
10 LECINY
AVTQLNLPNVAMYLDAGHAGWLGW PANQDPAAQLFANVYKNASSPRALRGLA
TNVANYNGWN
ITSPPSYTQGNVYNEKLYIHAIGPLL ANHGWSNAFFITDQGRSGKQPTGQQQW
GDWCNVIG
15 TGFGIRPSANTGDSLLDSFVWVKPGGECDGTSDSSAPRFD SHCALPDALQPAPQ
AGAWFQAYFVQLLTNANPSFL.

[225] Para cada uno de los sitios listados en la tabla 2-1, se generaron un promedio de 18 variantes de sustitución. Las librerías se recibieron como plásmidos puros que codifican variantes de proteínas CBH2.

Tabla 2-1: Posiciones seleccionadas de la proteína CBH2 madura

[226] Cuatro librerías de combinaciones sintéticas de CBH2 también fueron producidas por Sloning Biotechnology GmbH (Puchheim, Alemania) y BaseClear (Leiden, Países Bajos). Las tablas 2-2 a 2-5 listan las sustituciones que pueden estar presentes en miembros de las librerías de combinaciones sintéticas de CBH2 (enumeradas de acuerdo con la secuencia de aminoácidos de la CBH2 madura).

Producción de proteínas variantes de CBH2

30 [227] Los plásmidos purificados de pTTTpyr-cbh2 (p c b h i, Amp^R, Acetamida^R) que contienen marcos de lectura abiertos que codifican secuencias variantes de CBH2 fueron obtenidos de los proveedores especificados arriba. Los protoplastos de la cadena de H. jecorina cuatri-borrada (Aegl1, Aegl2, Acbh1, Acbh2) fueron transformados con

los constructos de pTTTpyrG y criados en agar selectivo que contiene acetamida a 28°C por 7 días como se describe (WO 2009/048488). Los genes que codifican celobiohidrolasa I (CBHI, Cel7a), celobiohidrolasa II (CBHII, Cel6a), endogluconasa I (EGI, Cel7b), y endogluconasa II (EGII, Cel5a) han sido inactivados en la cadena cuatri-borrada. Las esporas se cosecharon, replantadas en otro plato de agar de acetamida, e incubadas a 28°C por 7 días. Además, las esporas se cosecharon en glicerol al 15% y se almacenaron a -20°C para uso futuro. Para la producción de proteína variante de CBH2, se agregó un volumen de 10 ml de suspensión de esporas a 200 ml de medio mínimo de glicina suplementado con una mezcla de glucosa/sofrosa al 2% en un plato filtro de PVDF: 6.0 g/L de glicina, 4.7 g/L de $(\text{NH}_4)_2\text{SO}_4$; 5.0 g/L de KH_2PO_4 ; 1.0 g/L de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 33.0 g/L de PIPPS; pH 5.5; con la adición posterior a la esterilización de la mezcla de glucosa/sofrosa al 2% como fuente de carbono, 10 ml/L de 100 g/L de CaCl_2 , 2.5 ml/L de elementos traza de T. reesei (400X): 175 g/L de ácido cítrico anhidro; 200 g/L de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$; 16 g/L de ZnSO_4 ; 3.2 g/L de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$; 1.4 g/L de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$; 0.8 g/L de H_3BO_3 . Cada variante de CBH2 se crió por cuádruplicado. Después de sellar el plato con una membrana permeable al oxígeno, los platos se incubaron a 28°C por 6 días, mientras se agitaban a 220 rpm. El sobrenadante fue cosechado transfiriendo el medio de cultivo a un plato de micropipetado a baja presión y probado en busca de glucosa residual usando la prueba de hexoquinasa como se describe en el ejemplo 1.

EJEMPLO 3

Expresión, Actividad y Estabilidad de Variantes de CBH2

[228] La H. jecorina CBH2 SEL y variantes de proteínas combinatorias fueron probadas por varias propiedades de interés. En particular, las variantes de celulosa se probaron para la expresión de proteínas usando una prueba de HPLC (HPLC), actividad específica usando la prueba PASC de hidrólisis (Act. PASC) y la prueba PCS de hidrólisis (Act. PCS), estabilidad en presencia de etanol (proporción de EtOH) y estabilidad térmica (proporción de calor) como se describe en el ejemplo 1. Las variantes de combinación también se probaron para actividad específica por hidrólisis de comcon pretratado con amoníaco (Sp. Act. CC) como se describe en el ejemplo 1. Los datos de desempeño para las variantes de CBH2 SEL se muestran en la tabla 3-1, y los datos de desempeño para las variantes de combinación de CBH2 se muestran en la

Tabla 3-2. Las filas de la tabla 3-2 que no tienen datos de desempeño corresponden a variantes de combinación de CBH2 que no fueron expresadas en las pruebas iniciales.

[229] El índice de desempeño (Pi o PI) es la proporción de desempeño de la variante a la de la celulasa madre o de referencia. Varios de los términos que se establecen debajo

5 se emplean para describir la mutación: las mutaciones up tienen un $P_i > 1$; las mutaciones neutrales tienen un $P_i > 0.5$; las mutaciones no perjudiciales tienen un $P_i > 0.05$; las mutaciones perjudiciales tienen un $P_i = 0.05$; las mutaciones combinables son aquellas mutaciones para las cuales la variante tiene un valor $P_i = 0.5$ para al menos una propiedad, y > 0.05 para todas las propiedades. Las mutaciones combinables son
10 mutaciones que pueden ser combinadas para entregar proteínas con Pis apropiados para una o más propiedades deseadas. Las posiciones en las que ocurren las mutaciones se clasifican de la siguiente manera: las posiciones no restrictivas tienen $> 20\%$ de mutaciones neutras para al menos una propiedad; y las posiciones restrictivas tienen $< 20\%$ de mutaciones neutras para actividad y estabilidad.

15 [230] Estos datos pueden usarse para diseñar cualquier CBH. Aún si el CBH que se va a crear tiene un aminoácido diferente de los de la CBH2 en una posición particular, estos datos pueden usarse para encontrar una sustitución que altere las propiedades deseadas identificando la mejor elección de sustitución, incluyendo la sustitución al aminoácido del tipo salvaje de CBH2.

20 [231] La Tabla 3-1 muestra los valores de los índices de desempeño (Pi o PI) para 2.282 variantes de Hypocrea jecorina CBH2 en 162 posiciones. Los índices de desempeño menores o iguales a 0.05 se fijaron en 0.05 y se indican en **negrita itálica** en la tabla.

[232] Las variantes no combinables son aquellas para las que todos los valores de $P_i < 0.05$. Para la Hypocrea Jecorina CBH2, de las 2.282 variantes, 6 son no combinables.

25 Estas son la A121Q, L243H, A322Q, L328C, L328Y, V463Y. Cualquier CBH2 que tiene una de las sustituciones de arriba relativa a la Hypocrea jecorina CBH2 puede mejorarse mutando ese aminoácido a una de las sustituciones combinables en esa posición, o al aminoácido presente en la Hypocrea jecorina CBH2 en la posición 121, 243, 322, 328 y 463.

30

EJEMPLO 4

Clasificación de las posiciones de CBH2

- [233] Como se ha descrito hasta ahora, la funcionalidad de las variantes de celulasa se cuantificó como un índice de desempeño (PI), el cual es la relación del desempeño de una variante a la celulasa madre o de referencia. Varios de los términos que se establecen debajo se emplean para describir la mutación: las mutaciones up tienen un PI > 1; las mutaciones neutrales tienen un PI > 0.5; las mutaciones no perjudiciales tienen un PI > 0.05; las mutaciones perjudiciales tienen un PI = 0.05; las mutaciones combinables son aquellas mutaciones para las cuales la variante tiene un valor PI = 0.5 para al menos una propiedad, y > 0.05 para todas las propiedades. Las mutaciones combinables son mutaciones que pueden ser combinadas para entregar proteínas con Pis apropiados para una o más propiedades deseadas. Las posiciones en las que ocurren las mutaciones se clasifican de la siguiente manera: las posiciones no restrictivas tienen > 20% de mutaciones neutras para al menos una propiedad; y las posiciones restrictivas tienen < 20% de mutaciones neutras para actividad y estabilidad.
- [234] Como se muestra en la tabla 4-1, para las 162 posiciones probadas en la CBH2, todas las 162 son no restrictivas. Las posiciones no restrictivas son las más apropiadas para el uso en la construcción de librerías de combinaciones, ya que tienen un gran número de mutaciones combinables.

EJEMPLO 5

20 Efecto del cambio de carga sobre la actividad de las variantes de CBH2

- [236] En este ejemplo, se evalúa el efecto del cambio de carga sobre la actividad de la CBH2 en pruebas PCS y PASC. En resumen, los números de ganadores de PCS y PASC en las SELs de CBH2 se determinó como una propiedad del cambio neto de carga. En las tablas 5-1 y 5-2, se determinó la tasa de ganadores observados a los esperados (o/e) en las pruebas PCS y PASC respectivamente. Los valores en negrita itálica son significativamente diferentes del promedio de 10 distribuciones aleatorias más o menos el número de desviaciones estándar (sd) listados en las columnas respectivas.

- [237] Como se muestra en la tabla 5-1 y en la Figura 2A, reducir la carga (por ejemplo -1, -2) resulta en una frecuencia significativamente más alta de CBH2 ganadores en la prueba PCS, mientras que incrementar la carga (por ejemplo +1) resulta en una frecuencia significativamente más baja de CBH2 ganadores en la prueba PCS. En contraste, como se muestra en la tabla 5-2 y en la Figura 2B, se observa una frecuencia significativamente más alta de CBH2 ganadores para variantes sin un cambio de carga

(por ejemplo, 0), mientras que se observa una frecuencia significativamente más baja de CBH2 ganadores para cambios de carga (por ejemplo, -2, -2, +2).

[238] En conclusión, la actividad de la CBH2 en PCS se correlaciona con una disminución en la carga. La actividad de la CBH2 en la PASC por otro lado, muestra
5 que esta tendencia no se correlaciona con la actividad hidrolizante de la celulosa, pero es específica para el material del PCS.

EJEMPLO 6

Creación de cepas de tamizado de *Trichoderma reesei*

10 [239] Las cepas de tamizado mejoradas fueron creadas para incrementar la consistencia de la expresión de variantes de CBH2 en presencia de factores no relacionados con las secuencias de aminoácidos de las variantes de la enzima. En particular, las cepas de tamizado de *T. reesei* se desarrollaron en combinación con un vector objetivo para
15 forzar la integración de los genes de la variante *cbh2* (por ejemplo, la región de codificación en combinación operable con una secuencia reguladora). Las nuevas cepas preparadas durante el desarrollo de la presente divulgación, combinan diferentes mutaciones que son ventajosas para las librerías de variantes de tamizado. En la figura 4 se muestra un esquema de los pasos de ingeniería genética.

[240] El borrado de la *ku80* de la cepa derivada de *T. reesei* cuatri-borrada. Se identificó
20 un solo ortólogo de MUS52, el ortólogo *N. crassa* de la KU80 humana, mediante la búsqueda TBLASTN en el genoma de la secuencia de *H. jecorina* QM6a (*Trichoderma reesei*) y por consiguiente se denominó *T. reesei ku80*. Identificación de proteína 58213; <http://genome.jgi-psf.org/Trire2/Trire2.home.html>. La secuencia de nucleótidos del gen de *T. reesei ku80* se proporciona como la SEQ ID NO:23;

25 ATGGCGGACAAGGAAGCAACCGTCTTCATCATCGACCTCGGCGCGTCCATGG
CA
GCTGTCAATGGGGGTCGAGAAGAATCCGACCTTGATTGGAGCATGAGCTACG
TC
TGGGACAAGATCAGCAACGTCGTGGCCTCGAATCGCAAGACGCTGTGCGTT
30 GGC
GTCGTGGGGTTTCAGAACCGACGAGACAAACCACACGCTGAGCGAGGATGGG
TAC
GAGAACATCTCCATATTGCAGCCCCTGGGGCCGATGAGCATGTCCAGCCTCA

AG
GCTCTTCAGCCCAAGGTGAAGCCGAGCAGGACGGTGGAAGGCGATGCCATC
TCG
GCGATTGTCATTGCCGTCGACATGATTGACAAGTACACGAAGAAGAACAAT
5 GG
AAGCGGCAGATTGTTCTCATTACCGACGGCCAAGGCGAGATTGATCCAGATG
AT
ATTGCCGACATTGCTAGAAAGATGCGCGACTCGAATATTGAATTGACAGTCTT
GT
10 GAGTTGGCGAGACCGTTTGGCGGACGGTAATGGTGCTGACGGTGATGCAAG
GGG
CGTCGACTTTGATGCTCCCGATTACGGCTTCAAAGAGGAGGACAAACCTTCA
GTC
AAGGTA CTCCATATGTTCACTTCTTTTCTTTTCTTCTTTATTTTCTTTTCTTTT
15 GAA
GCTTTTCATTAACCTCTTCGTTAGAAGCAAAACGAAGAGACCCTAAAAAAGCT
CGT
GGATGGCTGTGGCGACGACTCAAGGTTTCGCCTCCATGGTCGAGGCCATTGAC
GA
20 CTTGAATGAGCCACGAGCAAAGTCGGTCAAGCCTTACAAAACGTACGAAGG
TCT
CTTGACCTTGGGAGATCCGAAAAACGCTCCCGCAGTGGTGGAATCCGCGTC
GA
GAGATACTTCAAGACCCATCTAGCCAGGCCACCTGCCGCCAGCACCGTGGTG
25 GTC
AAGGAGGAGCAAGCTGGGCCGTCTCAGGCAGACGAGGACGAACAGATGGA
CGG
AGCGGA ACTTACAGCTGTGAGGCAGGCCAGGACATACAAGGTCAATGATCC
AGA
30 TGCCCCTGGCGGTAAGCGTGACGTTGAGTTTGAGTCTCTGGCCAAAGGGTAC
GAG
TACGGCAGGACGGCAGTCCACATCAGCGAGTCTGATCAAAACGTCACCAAG
CTC

GCGACAGAAAAGAGCTTCAAGATCATCGGCTTCGTCCAGAAAGAAAAGGTA
TTG
GCTTGGCTCTCAGCATTGACCCGTTGCTCTTGGCTAACCCTTGTTTAGTATG
AAA
5 TGCTCCTTAATCTTGGCGAAACCTGCGTTACCGTTGCATCCAAGTACGATGAA
AA
GTCTGAGCTGGCTTTTAGCTCTCTGGTGTGGGCGCTCTCGGAGCTCGACGCC
TAC
GCCGTGGCCCGCCTAGTAACTAAGGACCAAAGGACCCCATGCTGGTGTAC
10 TG
ATGCCGTATATGGAGCCTGATTATGTTTGTCTCTATGATGTGCCTCTGCCTTTC
GC
AGAGGACATCAGGACGTACCAGTTTCCTCCCTTGGACAGAGTCGTTACCGTC
AGT
15 GGCCAAACGCTCACCAACCATCGCCTATTGCCATCCGACGAGCTCAACCAAG
CG
ATGAGCGACTACGTAGATGCCATGGACATTTCAAGTTATGGTATCGATGAAGA
TG
GGTGAGTATAGAAGATGATTGTTCAAATCTTTCACCTTCTAAGCATTGCTTCTGA
20 TC
TAGGCAACCGGCTGAATATGCCACCATCGATGAGTTATACAACCCTGCGATAC
AT
CGCATAGGCCATGCGATCAAACAACGAGCGATCCACCCAGAGAAACCCGTG
CCC
25 GAGATCCCCCAGTCTTGCTTAGATTTCGCAGCACCCCGACAGAACTCGTCG
AGA
CTGTGCAGCCTCATATCGATGCACTGATTCACGCTGCAGACGTGAAGAAAGG
TAC
TGATTCCATTACATATGCTTCTCTGCACACTGATGTTTGATTTGTGCTAACGCC
30 CC
CCTTAGTGCCGCCCAAGGCCAAGGGCAAGCGCCAAAGAGAAACAGTTAAAC
CCA
TCTCGGGACTGGATGTGGATGCCCTTCTGGGAGAAGAGCAGAAAGGTTCCAT

- TA
 GTCCGGAGAATGCCATTCCGGACTTCAAACGAGCCCTCAACTCGTCCGAAGA
 AG
 TCGAGCAGATTGCCGACGCCACAAAACAAATGGGGGCCATTGTGCGGTCTCT
 5 CA
 TTACGGACAGCTTCGGGGATAGCAAATATGCCCAGGCAATGGAAGGCATTGG
 TG
 CGATGCGTGAGGAGCTGATCAACCTGGAAGAGCCTGGCCTGTACAACGACTT
 TG
 10 TGCGCGACTTGAAGAAAAGTTTGCTATCTGGAGCCTTGGGTGGTGACAGGCG
 AG
 ATTTCTGGTTCAAGATGAGGTGGGCGAAGCTGGGCCTGATTGACAAGAAAC
 AGT
 CGGAGGTGTCTTCGGTCACTCTTGAGGAGGCGGACGAGGTGAGTGGTGACAG
 15 CAT
 GCTGTCGGATTATACGGACGTTGTTTGCTAACTTGTGGGATAGTTTACAAGT
 CG
 AGGTGAGGTATCTACGTTGACCAAGAATGGGACCATGTATATGAGCGGTGTA
 AC AACAGAATCCTGTGCTTTGAGCATTGTATGA
 20 [241] El gen de Treesei ku80 fue borrado de la cepa derivada cuatri-borrada usando
 métodos estándar en el arte (WO 2005/001036). En resumen, se empleó una platina de
 borrado de ku80 que emplea un marcador seleccionable flanqueado entre la secuencia
 1.3 kb de 5' ku80 y la secuencia 2.3 de 3' ku80, como se muestra esquemáticamente en
 la figura 5. La variante Treesei ah, la cual confiere resistencia al herbicida etil
 25 clorimuron, se empleó como marcador seleccionable como se describe en O
 2008/039370. La secuencia de nucleótidos de la platina de expulsión de ku80 tiene una
 longitud de 7685 pares de bases: las bases 1-1271 corresponden a la región homóloga
 del 5' ku80; las bases 1280-7685 corresponden a la variante resistente al etil α -s-
 clorimuron (A190D); y las bases 5381-7685 corresponden a la región homóloga de la
 30 3' ku80. La secuencia de nucleótidos de la platina de expulsión se proporciona como la
 SEQ ID NO: 14: ggccgcctcaacacccacactcgaggcacacgagttcatcggcggctccccacaagct
 ctggccaacctgctaccggctctctc
 gcgagacttccaaagcctacaaacgaggtcgacgtcaaggaggccctcgagcgccagcc

cggcagatggagcctccagggcca
gatcaaggccaacaacatgagagcccagagcgccgcactccggctcgacgacaaggaggg
caaggcgagagcctttgaggagg
ccaagcgcgagctactggcgatcaccacagcgccctcggaagccttccggcgcaagat
5 aatgagcttgatcgcaatgacgagttc acgtacgcttggccatattgttggcttttggctcctacatgtacggcgcatgg
ttgggaggatatacccacggagagtgtccgagt
ggcttctgggatttagagcgtcattagcaggatagagatggttggccaggggaatggaat
tgacttttactacaaggaacttgttactc
tgggttgattccattgcgtgactggtagtagggagggaatgcttttactttgtgccact
10 agaccgcagagaagggttgggtgcaagcgg
ggtccgtgtataccgaccaagagtgtgggcatacagcaacgtttctgaacgacttcatt
ttgtccgagtctactggatgagatgccca
gcgtgaagccgtacgccaccagggcgacgaactcgacaagggtgacgagggaggagatgc
cgtgcagcatgccaaacttctgttg agggcacgcatctcatccgactgtgcatccttgcataccactccttccgtctcgttg
15 gctgggtgggaggggtcaacaaatccatcgtc
agccatccgggggtctcaaatcaatggcgtgcatgaggagtcgggcttgaggctaacttg
tccatggcggcttctcatggcttgacag
tggcgggaagcagcacggcgagggtgacgagggcgctgacgaacatggttgcgatgggca
ccaaggagctccacttgttgggagc
20 gtcgacgagggccgcatgccgcccttgatgcccaagagggcggttccggggaacgtgag
ggcgagcagcgcggggatggccgt
ctgcatgccaaagtatggggaacagcttgccttggtgagcggaaggagggccggct
gacgggtgcggaacatgacgatgccg
ttgacgaaggactgcagtagcgtagtgtgatggttaagcagctggccggcgcgctgagac
25 aatggccggcaatggtaaaaaggacc
aagatgtactaggtagttgcaatgtggcttattacctactactacctggtaggcaccta
ctaggtacttgggtagacggacaatgaaattt
gaagtcgggggttcaggaaagcagggcgctggacacattgtgttcaggcggtagccgtc
gtcatcgtcagccaatgtcgaggccc
30 ggcagcccaggagcgcagacaaccttggccggaggagcccgcaggtacctgccaaagcgc
ggctggtacctctcaacctctcag
gcctgttggtatgccctatgacatgccctgggggatgcagctgttccccggccccgcact
ttcgggtgaccgcgaggctgctgattgg

ctgggtgccacgggctgggcgggtccctgaagttgttgcctctgaactctgtcggcgctg
 gcgtcggctgcgccaatgggaggcga
 gacaactcagggtagaatactgacagaagaagagaatcgaaagtaggtagacagcc
 aattcgttgcattggcaggcaaccgca
 5 caggagaaaaattgactaccccacaatcaggcacagtaagtagggcacagtacgtatgta
 cagacaaggcgcaagcgatactgcgc
 gacccgggtacctgcgggcttgacacgtgcgacaggctactttactagtattcgcagcgg
 cgggtcgcgcattattacatgtactgtgc
 cgccatttgatgactgggctgctgcagtattagtagatctgcccggcacgcccttccat
 10 gggcgcgacccgggactggaccctctga
 ctctacctacatgtacctaggccggggcgggcttggtagctttgtccgatcaggtcgtt
 cgcttggctacctattttctctttctctctc
 catcctgcttctggccttgcaattctcttcgccactcctcctcttccccccgcgatac
 ccttgaattcgtcagagaggaaaagacgaga
 15 aaaaaagggcagcagagacgtcgggtctggctcacgtgctgcatctctgcgcactctcat
 ttttttattgtccgacccctccctcaacctt ctccttcgttgacaggctaagccttgcttcgacgctctctctttgaattttctacttct
 accttctttcttgcgtgttaccaccatagctcgat
 tcacgatgctccgaagtcgccaagtcacagccagggccgtccgggctctggggccaggcgc
 gcgcctttacctcgacgaccaagcct
 20 gtcatgatccagagcagccagaggaaacaggccaacgccagcgctgctccgtaagtcgcc
 cattgccattgcattctctgtttgatata
 tacttctgctgcttgcgtggcgtcgtctctcggttatgcgtgtcaaggaccagggtgtgt
 tcgcatcgtggtttccagcgccgattaccgg
 gggacgaatttttggctgctcaactcgcgcgcgcgcattctgattcttcgtttcaatct
 25 tgagcgacaactggctaataatggccattg
 gcaattgcttcacacagacaagtcggccctgtaccgagccctgcttcaacgctgaagac
 aaagaccgcagccatgtgcagcctctgg
 tcaacccgtcgaagcccgcacatggatgaatcgtatgtccacgtcccctcgtcccgccta
 caaatgaacacgattacaccagaatttt
 30 gcaacaatcgacacttctataacagaccaattgagctttgttctgaccaatcatgttgc
 ctagattcattggcaaaaccggaggcgaaat
 ctccacgagatgatgctgcgacagggtgtcaagcacattttaggttccgatgccggcc
 gccacacgggctccatccttgcctcatc

tctccagctaggcaaattctcgctaaccttgagtcaccatccagtcggataccctggcggc
 gctatcctgcccgtcttcgacgccatctac
 aactcaaaacacttcgacttcatectgccccgtcatgagcaggagctggccatatggcc
 gagggctatgcccgtgcctcgggcaaa
 5 cccgggtgctcctgggtgacttcggccccgggtgctaccaatgtcatcacgcccatgcag
 gatgccctgtcggacggnaacgcccttgg
 tcgtcttctcgggccaggctcccaccacggccatcggcagc gatgacttccaagaggccg
 acgtcgtgggcatctcgcggggcctgc
 accaagtggaaacgtcatggtcaagagcgttgctgagctgccgcggagaatcaacgaggcc
 10 ttgagattgccaccagcggccgcc
 tggccccgtcctcgtcgacctgcccgaaggatgtcacggctggtatcctgaggagagccat
 ccctacggagactgctctgccgtctctg
 cccagtgccgcctcccgcgccgcatggagctgagctccaagcagctcaacgcctccatc
 aagcgtgccgccgacctcatcaacat
 15 cgccaagaagcccgtcatctacgccgggtcaggggtgcatccagtcgaggcggtga
 gctcctgaagcagctggcggacaag
 gctccatccccgtcaccaccacctccatggcctgggtgcctttgatgagctggacgag
 aagtcgctgcacatgctgggcatgcacg
 gctcggcgtatgccaacatggccatgcagcaggccgacctcatcatcgccctcggcagcc
 20 gattcgacgaccgtgttactctgaatgt
 ctccaaatttgcgctgcagccaggcaagctgctgccgagggccgcggcgatcatcca
 ctttgagatcatgccaagaacatcaa
 caaggatccaggcgaccgaggccgtcgaggcgacgtcgccaccaacctgaagcacct
 cattccccagattgccgaaaagtcca
 25 tggcggaccgaggagagtggttcggcctcatcaatgagtggagaagaagtggcccctgt
 caaactaccagcgcgcggagcgggc
 tggcctcatcaagccgcagacgggtcatggaggagattagcaacctgacggccaaccgaaa
 ggacaagacgtacattgccacgggt
 gtcggccagcaccagatgtgggttgcctcgcacttcgctggaggcacccctcgatccatg
 30 attacctctggtggtctgggcaccatgg
 gctacggtctccccgcggccattggcgccaagggtggcccagcccgacgctctcgtaatg
 acgttgatggc gatgcctcgtttaacat
 gacgctgacggagctgtcgactgctgcacagttcaacattggcgtcaagggtggttgct

caacaacgaggagcagggcatggtgac
 gcagtggcagaacctcttttacgaggaccgatatgcccacacgcaccagaagaaccccg
 ctcatgaagctggccgacgccatgg
 gcgttcagcaccagcgcgtgacggagccggagaagctggtcgatgccctgacgtggctga
 5 tcaacaccgatggccggccctgttg
 gaggtgtcacggacaagaagggtgcctgtcctgcccatgggtgcccgccggatcgccctg
 cacgagttcctcgtctttgaacctgggtg
 agtctacttcagacataattgcttgcgcatcgagatactaacactctcacagaaaaggat
 aagcagcggcgtgagctgatgaaggagag
 10 aacaaagggtgtgcactcctaaagcgatgatgtctgcgaggggttctcgttgaacctta
 gttcaggcaccatcttacctcttatttttcc
 cgtgggctttcatttgtgtcatccgagcatgacgtttaggggttgagtttcttcttt
 ttatcttgcatttactggtacctataggcgcgag
 actaggcttccatgtttgttttgcgactttcaaaaagtacttttagtggttggggcac
 15 gacgagggggggcaacctcttctgtcgaaaaa
 ggtggctggatggatgagatgagatgagatgaggggtgaagatagataacctgcagtgttt
 tgacgcgacgggatggcgatcgagca
 ccccgacagaactcgtcgagactgtgcagcctcatatcgatgcactgattcacgctgca
 gacgtgaagaaaggatgattccattac
 20 atatgcttctctgcacactgatgttgattgtgtaacgcccccttagtgccgccaa
 ggccaagggaagcgccaagagaaacag
 ttaaaccatctcgggactggatgtgatgcccttctgggagaagagcagaaagggtcca
 ttagtccggagaatgccattccggacttc
 aaacgagccctcaactcgtccgaagaagtcgagcagattgccgacgccacaaaacaaatg
 25 ggggccattgtcgggtctctcattacg
 gacagcttcggggatagcaaatatgccaggcaatggaaaggcattggtgcgatgcgtgag
 gagctgatcaacctggaagagcctgg
 cctgtacaacgactttgtgcgcgacttgaagaaaagttgctatctggagccttgggtgg
 tgacaggcgagatttctggttcaagatgag
 30 gtgggcgaagctgggcctgattgacaagaaacagtcggaggtgtcttcggtcactcttga
 ggaggcggacgaggtgagtggtgcag
 catgctgtcggattatacggacgttgttgcctaactgtgggatagtttacaagtcgag
 gtgaggtatctacgttgaccaagaatgggac

catgtatatgagcgggtgtaacaacagaatcctgtgctttgagcattgtatgatatgatta
 ttgatgaaccggacaaaaggggtagggga
 ttgatgccatcacgaccgattgaccagacctggattctgcacagcatggctgctgattt
 tgttgaccttgcgacgtaacatccctgaaga
 5 acaacctactattaacctatcatttagcagaagctctgtaaccttcttgattcttgatt
 cagcttctgagctctgcaaatgtaatcatttcgagg
 ttgtgtaattccggccaagcaggcggccgtctgccagcgcctgacctaggctgcaccgcaa
 tctgccaatcagctgcccttcagttcg ttgaccttgagctgcccctcatcctttatctgcacacaattcttttctctgctctg
 cgcattcttctctctctctctctcccttctcaagctcaa
 10 cttcacctcactcgcctcactacaagccctcccgctgctgctcgcacccatctcgcac
 tgcggccagcaaaacaagcaaagccgtga
 tcgatccctcagcatggctaccctcaacctcaccgtccgcctggagatgctcaaagaaatt
 ggaatcaccgtccaatacggcgagcatgt
 agcgaaagaagcagccagcaacgaagcagcgatggcattcgaagaagaagaagagtccc
 15 cgccgttgtgccgccaaggcaga
 acagcacgcctctgaacacgacgctggccacgatgcttgggacgcggctgccacatctc
 gacttcggcgcaagaacagcagaag
 cccaggagatggacgactcgtctatcgtgatgccgctggactactccaagttgtcgtt
 ggagagcctgcggacgaatccatcagctt
 20 ttgctcgtggaaggctcgtcgaaggcttatcctgaccagtttatcggcaaggcaaacaggcc
 tcgtgtatgtagcgattgcttctctgcatta
 tgggaatctcaagagagtatggtagaagataactgacaactgcaggccaagccgtactt
 tgacaagatttgggaagacagagtctgg
 gatttgtgaggatcttgattgatgtgcatatggcgacatgcctgctaataatcattgtagc
 25 ttctatctctacaaccccagagaagccttcaga
 gaagcctcgcgtgctgggtgccactgttcagctcgaaggctttctcaaaagcatcaacag
 agcgctcgggtacttctctcaccattccag
 gaggggcaaaccaggaccggttttatctgagggtcggccaggagacaccccaaggcctc
 gatatctacagaggctcagagagaccag
 30 aaatccctaaagattgaaacgttccccgattttcaacaggcggactacgacagcttagg
 aacgcgcatggcgccatccaggaggact
 ggttgaagaactggcagatgctggtacctcggccgagtttcgacaagaagaaaaatgcag

acaaaagagcagccaagagaaggct

cgagcgagagcgaatgcttcacaatacgcaggaattttcatttggcaggttaagggcaa aggggctgacgtgg.

- [242] La creación de la cepa Archy2 a partir de la cepa derivada de la *T.reesei* $\Delta ku80$ cuatri-borrada. El gen *pyr2* fue borrado de la cepa de expulsión de *ku80*. La platina de
- 5 borrado del *pyr2* contiene el promotor de *T. reesei* *cbhl*, un gen resistente a la higromicina y un marcador parcialmente seleccionable *amdS* flanqueado por las secuencias 5' y 3' del *pyr2*, mostrado esquemáticamente en la figura 6. El uso de este vector permite buscar resistencias a la higromicina y al ácido fluoroorotico de los transformadores de expulsión del *pyr2*. El gen *amdS* parcial contiene la porción 3' del
- 10 gen, pero carece de un promotor y la porción terminal amino de la región de codificación, y es por consiguiente no funcional. La secuencia de nucleótidos de la platina de expulsión de *pyr2* tiene una longitud de 9259 pares de bases: las bases 1-1994 corresponden a la región homologa 3' del *pyr2*; las bases 2002-3497 corresponden al promotor de *T. reesei* *cbhl*; las bases 3564-5449 corresponden al marcador seleccionable resistente a la higromicina; las bases 5450-7440 corresponden al marcador parcial 3' de
- 15 *A. nidulans* *amdS*; y las bases 7441-9259 corresponden a la región homologa 5' del *pyr2*. La secuencia de nucleótidos de la platina de expulsión del *pyr2* se proporcionan como la SEQ ID NO:15:

- atcacgccctgcataaaagaccctcaagagtccatgtgccctatctgcctgatcttctct
- 20 aacccttatttaacattggccctatcacaacc
- tagttcttctcagcctgctttgtcaacacttgtcacggttcaactcaacgtaatcagca
- ggtagcaggacgaggatagggagagaaacg
- aagagaagaagaggagagaggaagaaaaaaaagaaaagaaaagggaaggaa
- agaaggaggaaaagagaagaa
- 25 agtcagatgaagaagcaagaagacgccatggtagccaccgcttgtcagggctccttagca
- acgaacaactctagcttggggacttgtc
- gatgtgtcgtttccttaccatcagcaccaacgatgagttc gatatagacgaggacc
- tcatggaagtagagaccattgggttcgaca
- ggatctctcagtttcaacttctatgaggtctgtcgtcggatgacttttgaggagcttcc
- 30 ccttctgcttcaaccccaaactctcttctgaa
- accgcagcacgttggcacggccgtgttgctggagcagtttgcttcgagcactctcagcg
- tggtttcagcagcccactggtgagtggc
- ctcctttgacgtccacaccttgctctgtcgcgcgtatctggtgggaacgactgctcc

aaggaggattgctaacgaggtgtaggccg
 aatatgcatcagattctccgtaaccttagctacggcctctcaacatctgtgacatga
 cggagcgcaagtactgggtgtggcgacca
 agatgcgcggctggaacatcgacggctgccccgaagacgtcaggagactcatgtttgtc
 5 acatcatcgccaccctgggatgcagcc
 ccgtcgtgacggatgaagacatggactaccccaagaactgggcggcaattctccacggta
 gagacagatatccgagtgaacctgtgg
 gccaccggcctcatgggcgcaccatctgcctccactcgggtggcgtctgccctcgtctcc
 agggcttgggtctcgggtactcgactct
 10 gaagtcgtatgtgcagcgcataaacagcctcggcgccgcggaccgtgttgcctcgtttg
 ccgcaagcccgagacgagatttttgaa
 agatgcggcttcaggaacagcggccggagtagtatcaagactctggtcggcgaatactac
 aacatggtgtgtctccacatcgactt ggccagactctatacgattttcaaacctcgctatacgtcatattgacttgttcttagg
 tcttcgatttggccgggcccagactttatcga
 15 ctggaatagcattgccgacgctgccaagaagatgtgaaccatttgactgatacgtatgt
 gctacgcatgtcgaccttctttgtttgttctt
 tggcggctctttgtataccttgggacacggcagacgcatgtctatgtgaagaaaacgttc
 acggcgctgtttgcatcaggaatatgatca
 ttaaocatggagcgtaatggtattaatgatcaactagaaaaatggtatggaagggcgaga
 20 gggcgatcaacaaagcagcccggggc
 atagtctggaagcagcaggaattggaagggaaaagggaagctgcacaatgaagggatatcg
 tgagcggagtggctcacgagatc
 aacagactggcgaaagcaagcaattgccaacgccggctattaggccataagatggcctgt
 tgtgagtcacgattgcacgtatcccat
 25 atgactgctctgtcgtgacttgaaaaaaataggaggataaaggagaaagaaagtgag
 acaacccgtgagggacttggggtagta
 ggagaacacatgggcaaccgggcaatacacgcgatgtgagacgagttcaacggcgaatgg
 aaaatcttgaaaaacaaaataaata
 actgccctccatacgggtatcaaattcaagcagttgtacggaggctagctagagttgtga
 30 agtcggtaatcccgtgtatagtaatacga
 gtcgcataaatactccgaagctgctgcgaacccggagaatcgagatgtgctggaaagc
 ttctagcgagcggctaaattagcatgaa
 aggctatgagaaattctggagacggcttgtgaatcatggcgttcattcttcgacaagc

aaagcgtccgtcgcagtagcaggcactc
 atccccgaaaaaactcggagattcctaagtagc gatggaaccggaataatataaggca
 atacatigagttgcctcgacggttgcaat
 gcaggggtactgagcttggacataactgttccgtacccacctcttctcaacctttggcg
 5 tttccctgattcagcgtacccgtacaagtcg
 taatcactattaaccagactgaccggacgtgtttgcccttcatttggagaaataatgt
 cattgcgatgtgtaattgcctgcttgaccgac
 tggggctgttcgaagcccgaatgtaggattgtatccgaactctgctcgtagaggcatgt
 tgtgaatctgtgtcgggcaggacacgcctc
 10 gaagggtcacggcaagggaaccaccgatagcagtgcttagtagcaacctgtaaagccgc
 aatgcagcatcactggaaaatacaaa
 ccaatggctaaaagtacataagttaatgcctaaagaagtcataaccagcggctaataat
 tgtacaatcaagtggctaaacgtaccgtaa
 tttgccaacggcttgtgggggtgcagaagcaacggcaaagccccacttccccacgtttgt
 15 ttcttcaactcagccaatctcagctggtgat
 cccccaattgggtcgttgtttgtccgggtgaagtgaagaagacagaggtgaagaatgtc
 tgactcggagcgttttgatacaaccaag
 ggcagtgatggaagacagtgaatgttgacattcaaggagtatttagccagggatgcttg
 agtgatcgtgtaaggagggttgtctgccg
 20 atacgacgaatactgtatagtcacttctgatgaagtggtccatattgaaatgtaaagtcg
 gcactgaacaggcaaaagattgagttgaaa
 ctgcctaagatctcgggccctcgggccttcggcccttgggtgtacatgtttgtgctccgg
 gcaaattgcaaagtggttaggatcgaaca
 cactgtgcctttaccaagcagctgagggatgtgataggcaaattgtcaggggccactg
 25 catggtttcgaatagaaagagaagcttag
 ccaagaacaatagccgataaagatagcctcattaaacggaatgagctagtaggcaaagtc
 agcgaatgtgtatataaaaggttcgag
 gtccgtgcctccctcatgctctcccatctactcatcaactcagatcctccaggagactt
 gtacaccatcttttagggcacagaaacccaa
 30 tagtcaaccgggactgcgcacatcatgtatcgggaagtgccgtcatctcggccttcttg
 ccacacctcgtgctagactaggcgcgcca
 ggaagcccggaaggtgaagtgattcttcgccgtggctggagcaaccgggtgattccagcg
 tctccgacttggactgagcaattcagc

gtcacggattcacgatacacagctcagaccgctccacggctggcggcattattggttaac
 ccggaaactcagtccttggccccgtcc
 cgaagggaacccgacttaccaggctgggaaagccagggatagaatacactgtacgggcttc
 gtacgggagggttcggcgtaggggtgtt
 5 cccaagttttacacaccccccaagacagctagcgcacgaaagacgcggagggttggga
 aaaaaggggcgaaaattaagcgggag acgtatttaggtgctagggccggttcctccccattttcttcggttccctttctcct
 ggaagactttctctctctctctctctctctctctccatc
 ctacgtccatctctcttcccatcatcctcctcacctccatctcaactccatcacat
 cacaatcgatatgaaaaagcctgaactcaccg
 10 cgacgtctgtcgagaagtttctgatcgaaaagttcgacagcgtctccgacctgatgcagc
 tctcggaggggcgaaagaatctcgtgctttc
 agcttcgatgtaggaggggcgtggatatgtcctgcgggtaaatagctgcgccgatgggttc
 taaaagatcgttatgtttatcggcactttg
 catcgccgcgctccccgattccggaagtgttgacattggggaattcagcgagagcctga
 15 cctattgcatctcccgccgtgcacaggg
 tgtcacgttgcaagacctgcctgaaaccgaactgcccgtgttctgcagccggcgcgga
 ggccatggatgcgatcgtgcggccga
 tcttagccagacgagcgggttcggccattcggaccgcaaggaatcggtaatacactac
 atggcgtgatttcatatgcgcgattgctg
 20 atccccatgtgtatcactggcaactgtgatggacgacaccgtcagtcgctccgtcgcgc
 aggtctcctgatgagctgatgctttgggcc
 gaggactgccccgaagtccggcacctcgtgcacgcggattcggctccaacaatgtcctg
 acggacaatggccgcataacagcgggt
 cattgactggagcgaggcgatgttcggggattccaatacagggtcgccaacatcttctt
 25 ctggaggccgtggttggttgtatggagc
 agcagacgcgctacttcgagcggaggcatccggagcgtgcaggatcgccgcggctccggg
 cgtatatgctccgcatgggtcttgacc
 aactctatcagagcttggttgacggcaatttcgatgatgcagcttgggcgcagggtcgat
 gcgacgcaatcgtccgatccggagccgg
 30 gactgtcgggcgtacacaaatcgccgcagaagcgcggccgtctggaccgatggctgtgt
 agaagtactcggcgatagtggaacc
 gacgccccagcactcgtccgagggcaaaggaatagatagatgccgaccgggatccactt
 aacgttactgaaatcatcaaacagctt

gacgaatctggatataagatcgttggtgtcgtatgtcagctccggagttgagacaaatggt
 gttcaggatctcgataagatacgttcatttg
 tccaagcagcaaagagtgccctctagtgtttaatagctccatgtcaacaagaataaaac
 gcgttcgggtttacctctccagatacagc
 5 tcatctgcaatgcattaatgcattggacctcgcaaccctagtagcccttcaggctccgg
 cgaagcagaagaatagcttagcagagtct
 attttcattttcgggagactagcattctgtaaacgggcagcaatcgcccagcagttagta
 gggccccctctacctctcaggagatgtaa
 caacgccaccttatgggactatcaagctgacgctggcttctgtgcagacaaactgcgccc
 10 acgagttcttccctgacgccgctctcgcg
 caggcaagggaactcgatgaatactacgcaaagcacaagagacctgttggtccactccat
 ggcctccccatctctctcaaagaccag
 cttcgagtcgaaggtagaccgttgcccctaagtcgtagatgtccctttttgtcagctaac
 atatgccaccagggtctagaaacatcaatgg
 15 gctacatctcatggctaaacaagtacgacgaaggggactcggttctgacaacctgctcc
 gcaaagccgggtgccgtctttctacgtcaa
 gacctctgtcccgcagacctgatggctcgcgagacagtcaacaacatcatcgggcgcac
 cgtcaaccacgcacaagaactggtc
 gtgcggcggcagttctgggtggtgaggggtgcgacgttgggattcgtggtggcgtcatcgg
 20 ttaggaacggatatcgggtggctcgatt
 cgagtgccggccgcgttcaacttctgtacggctaaaggccgagtcattggcggctgccg
 tatgcaaagatggcgaacagcatgga
 gggtcaggagacgggtgcacagcgttgcgggccgattacgcactctgttgagggtgagtc
 ctccgctcttcttcttttctgtctata
 25 ccaggcctccactgtcctcctttctgtcttttatactatatacagagaccggcagtcact
 gatgaagtatgttagacctccgcctcttcacca
 aatccgtcctcggtcaggagccatggaaatacgaactccaagggtcatcccatgccctggc
 gccagtccgagtcggacattattgcctc
 caagatcaagaacggcgggctcaatatcggctactacaacttcgacggcaatgtccttc
 30 acaccctctatctctgcgcggcgtggaa
 accaccgtcggcgcactcgccaaagccgggtcacaccgtgaccccgaggacgccatacaag
 cacgatttcggccacgatctcatctcc
 catatctacgcgggtgacggcagcgccgacgtaatgcgcgatatcagtgcatccggcgag

ccggcgattccaaatatcaaagaccta
 ctgaacccgaacatcaaagctgttaacatgaacgagctctgggacacgcatctccagaag
 tggaattaccagatggagtaccttgaga
 aatggcgggaggctgaagaaaaggccgggaagggaactggacgccatcatcgccgatta
 5 cgcctaccgctcggttacggcatga
 ccagttccggtactatgggtatgcctctgtgatcaacctgctggattcacgagcgtggt
 tgttccggttacctttgcggataagaacatcg
 ataagaagaatgagagtttcaaggcggttagttagcttgatgccctcgtgcaggaagagt
 atgatccggaggcgtaccatggggcac
 10 cggttgcagtgacaggttatcggacggagactcagtgaaagaggacggtggcgattgcag
 aggaagtggggaagtgtctgggaaat
 gtggtgactccatagctaataagtgtagatagcaattgcacaagaaatcaataaccagc
 aactgtaataagcgtgaagtaccatg
 ccatgtctacgaaagagcagaaaaaacctgccgtagaacgaagagatatgacacgcttc
 15 catctctcaaagggaagaatcccttcag
 ggttgcgtttccagtagtgattttaccgctgatgaaatgactggactccctcctctgct
 cttatacgaaaaattgcctgactctgcaaagggt
 tgtttgtcttggagatgatgtcccccccatcgtcttatctcataccccgccatcttt
 ctgattctcatcttcaacaagaggggcaatcc
 20 atgatctgcgatccagatgtgcttctggcctcactctgccttcaggttgatgttcaact
 taattgggtgacgaattcagctgatttgctgcagt
 atgctttgtgttggttcttccaggcttgtgccagccatgagcgctttgagagcatgttg
 tcacttataaactcgagtaacggccacatattg
 ttactacttgaatcacatacctaatttgatagaattgacatgtttaagagctgaggt
 25 agctttaatgcctctgaagtattgtgacacagct
 tctcacagagtgaagaatgaaaagttggactccccctaataagtaaaagtctctctg
 aacggtgaagagcatagatccggcatcaa
 ctacctggctagactacgacgtcaattctgcggcctttgacctttatatatgtccattt
 atgcaatagattcttttttttttttttttttttttttttttttttttttttttttttttt
 30 ttttttttgcceaatttcgcagatcaaagtggacgttatagcatcataactaagctca
 gttgctgaggggaagccgtctactaccttagccc
 atccatccagctccataccttgatactttagacgtgaagcaattcacactgtacgtctcg
 cagctctccttcccgtcttcttccccactgg

ggccatggtgcgtgtatcgtcccctccacaattctatgccatggtacctccagcttatac
aatgccccgctaacaagtcgcctctttgcctt
gatagcttatcgataaaactttttccgccagaaaggctccgccacagacaagaaaaa
aaattcaccgcctagcctttggccccggc
5 atttggttaaacctcgagcctctctcccgtcttggggtatcaggaagaaaaaataat
ccatcgccaagggtgttttggcatcacca
cccgaaaacagcacttcctcgatcaaaagtgtcccgccatgaagaccacgtggaaggaca
tcctccgggtgcctacgcaccaggagt
ttctggacattgtgtgagcaggagccagcgaaactgccactcagatccgtgccgggt
10 tcaagattagcagaattcagggtacgtc
gcattgcccatcgaggtgtctcattatcggggctcttggaacgatcatgattgcat
ggcgtatgctaacacatagacagccttctac
actcgaaagggtcaagttcaccagagagcgtttccgaaaagttcgctccatcctcgac
agcttcctcgcctccaggacatccacc
15 ctccacaaggaccttctcaacacctctacgatgccgaccactcaagattgcccttgg
ccagatgtccactgccaagcacctgggtc
agaccatctcgcgcgactacgtcgtctcttgaatacgcgccagtcgctctaacagtga
agcagctcaagcggggcgcgtctcggtc
catggccacgctgggtcaagcgcctcaaggacccccctgtgtacctggaccaggtccgcca
20 gcatctcggccgtcttccctccatcgac
cccaacaccaggacctgtctatctgcgggtacccaatgttggaagtccagcttctg
cgaagtatcaccgcgcggatgtggacg
tccagccctatgctttcaccaccaagagtctgtttgtcggccacttgactacaagtacc
tgcgattccaggccattgatacccccggtatt
25 ctggaccacctcttgaggagatgaacactatcgaaatgcagaggtatgtggcgcggcta.

[243] Creación de la cepa Archy3 a partir de la cepa Archy2 de *T. reesei*. La cepa Archy2 fue transformada con un vector para integrarlo al mismo lugar del *pyr2* y reemplazar el gen de resistencia a la higromicina con la región de codificación del gen *pyr2*. La platina de expulsión de la higromicina se muestra en la figura 7. Esta reintroducción del gen de *pyr2* de vuelta al sitio del *pyr2* lo ubica entre el promotor *cbhl* de *T. reesei* y el marcador parcial seleccionable *amdS*. Esta cepa pudo seleccionarse por prototrofia de uridina y sensibilidad a la higromicina. La secuencia de nucleótidos de la platina de expulsión del *hygR* tiene una longitud de 9088 pares de bases: las bases 1-

1994 corresponden a la región homologa 3' del pyr2; las bases 1995-3497 corresponden al promotor cbhl de T. reesei; las bases 3564-5137 corresponden al marcador seleccionable de pyr2; las bases 5280-7270 corresponden al marcador parcial de A. nidulans amdS 3'; las bases 7271-9088 corresponden a la región homologa 5' del pyr2.

- 5 La secuencia de nucleótidos de la platina de expulsión de hygR se proporciona como SEQ ID NO: 16:

atcacgccctcgcataaaagaccctcaagagtccatgtgccctatctgcctgatcttct

aacccttatttaacattggccctatcacaacc

tagttcttctcagcctgctttgtcaacactgtcacgggtcaactcaacgtaatcagca

- 10 ggtagcaggacgaggatagggagagaaacg

aagagaaagagaggagagaggaaagaaaaaaaagaaaagaaaagggaaaaggaa

agaaggaggaaaagagaagaa

agtcagatgaagaaagcaagaagacgccatggtagccaccgcttgcagggtccttagca

acgaacaactctagcttggggacttgc

- 15 gatgtgtcgtttccttctacccatcagcaccaacgatgagttcgatatagacgaggacc

tcatggagtagagaccattgggttcgaca

ggatctctcagtttcaacttctatgaggtctgtcgtcggatgactttttagggagcttcc

ccttctgcttcaaccccaactcttcttctgaa

accgcagcacgttggcacggcgtgttgctggagcagtttgcttcgagcactctcagcg

- 20 tggtttcagcagcccactggtagtggc

ctcctttgacgtccacacctgtcctgtcgcgtatctgggtgggaacgactgtctcc

aaggaggattgctaacgaggtttaggccg

aatatcgcatcagattctccgtaaccttagctacggcctcttcaacatctgtgacatga

cggagcgeaagtactgggtggttggcgacca

- 25 agatgcgcggctggaacatcgacggctgccccgaagacgtcaggagactcatgtttgttc

acatcatcgccaccctgggatgcagcc

ccgtcgtgacggatgaagacatggactaccccaagaactggggggcaattctccacggta

gagacagatatccgagtgaacctgtgg

gccaccggcctcatgggcgcaccatctgcctccactcgggtggccgtctgcctcgtctcc

- 30 agggcttgggtctcgggtactgcgactct

gaagtcgtatgtgcagcgcgtgaacagcctcggcgccgcggaccgtgttgcctcgtttg

ccgcaagcccagacgagatttttgaa

agatgcggcttcaggaaacagcggccggagtagtatcaagactctggtcggcgaatactac

aacatgggtgtgctccacatcgactt ggccagactctatacgattttcaaacctcgctatacgatcatattgacttggttcttagg
 tcttcgatttgcccggggcccaagactttatcga
 ctggaatagcattgccgacgctgccaagaagatgtgaaccattgactgatacgatgtgt
 gctacgcatgtcgaccttctttgtttgttctt
 5 tggcggctctttgtataaccttgggacacggcagacgcatgtctatgtgaagaaaacgttc
 acggcgtgtttgcatcaggaatatgatca
 ttaacatggagcgtaatggtattaatgatcaactagaaaaatggtatggaagggcgaga
 gggcgatcaacaaagcagcccggggc
 atagtctggaagcagcaggaattggaagggaaaaggagctgcacaatgaagggatatcg
 10 tgagcggagtggctcaccgagagtatc
 aacagactggcgaaagcaagcaattgccaacgcccggctattaggccataagatggcctgt
 tgtgagtgccagttgcacgtatccccat
 atgactgctctgtcgctgacttgaaaaaaaataggaggataaaggagaaagaaagtgag
 acaacccgtgagggacttggggtagta
 15 ggagaacacatgggcaaccgggcaatacacgcgatgtgagacgagttcaacggcgaatgg
 aaaatcttgaaaaacaaaataaaata
 actgccctccataggggtatcaaatcaagcagttgtacggaggctagatagagttgtga
 agtcggtaatcccgtgtatagtaatacga
 gtcgcatctaaatactccgaagctgctgcgaacccggagaatcgagatgtgctggaaagc
 20 ttctagcgagcggctaaattagcatgaa
 aggctatgagaaattctggagacggcttgttgaaatcatggcgttcattcttcgacaagc
 aaagcgttccgtcgagtagcaggcactc
 attcccgaaaaaactcggagattcctaagtagcgatggaaccggaataatataataggca
 atacattgagttgcctcgacggttgcaat
 25 gcaggggtactgagcttgacataactgttccgtacccacctcttctcaacctttggcg
 ttccctgattcagcgtacccgtacaagtcg
 taatcactattaacccagactgaccggacgtgtttgcccttcatttggaagaataatgt
 cattgcgatgtgtaattgcctgcttgaccgac
 tggggctgttcgaagcccgaatgtaggattgttatccgaactctgctcgtagaggcatgt
 30 tgtgaatctgtgtcgggcaggacacgcctc
 gaagggtcacggcaagggaaccaccgatagcagtgcttagtagcaacctgtaaagccgc
 aatgcagcatcactggaaaatacaaa
 ccaatggctaaaagtacataagttaatgcctaaagaagtcataaccagcggctaataat

tgtacaatcaagtggctaaacgtaccgtaa
 ttgccaacggcttgtggggtgcagaagcaacggcaaagccccacttccccacgtttgt
 ttcttcaactcagccaatctcagctggtgat
 cccccaattgggtcgttgtttgtccggtgaagtgaaagaagacagaggtaagaatgtc
 5 tgactcggagcgtttgcatacaaccaag
 ggcaigtgatggaagacagtgaaatgttgacattcaaggagtatttagccagggatgcttg
 agtgtatcgtgtaaggagggttgtctgccg
 atacgacgaatactgtatagtcacttctgatgaagtggccatattgaaatgtaaagtcg
 gcactgaacaggcaaaagattgagttgaaa
 10 ctgcctaagatctcgggccctcgggccttcggcctttgggtgtacatgtttgtgctccgg
 gcaaatgcaaagtgtggtaggatcgaaca
 cactgctgcctttaccaagcagctgagggtatgtgataggcaaatgttcaggggccactg
 catggtttcgaatagaaagagaagcttag
 ccaagaacaatagccgataaagatagcctcattaaacggaatgagctagtaggcaaagtc
 15 agcgaatgtgtatatataaagggttcgag
 gtccgtgcctccctcatgctctccccatctactcatcaactcagatcctccaggagactt
 gtacaccatcttttgaggcacagaaacccaa
 tagtcaaccgcggactgcgcacatgtatcggaaagtggccgtcatctcggccttcttgg
 ccacacctcgtgctagactaggcgcgta
 20 atatgtggccgttactcgagttataagtgacaacatgctctcaaagcgctcatggctgg
 cacaagcctggaaagaaccaacacaaag
 catactgcagcaaatcagctgaattcgtcaccaattaagtgaacatcaacctgaaggcag
 agtatgaggccagaagcacatctggatc
 gcagatcatggattgccctcttgttgaagatgagaatctagaaagatggcggggtatga
 25 gataagagcgtatgggggggcacatcatc
 ttccaagacaaacaacctttgcagagtcaggcaattttcgtataagagcaggaggaggg
 agtccagtcatttcatcagcggtaaaatc
 actctagacaatcttcaagatgagttctgccttgggtgacttatagccatcatcatacct
 agacagaagccttgtgggatactaagaccaac
 30 gtacaagctcgcactgtacgcttgcacttccatgtgaaaactcgatacggcgcgcctcta
 aattttatagctcaaccactccaatccaacc
 tctgcatccctctcactcgtcctgatctactgttcaaatcagagaataaggacactatcc
 aaatccaacagaatggctaccacctcccagc

tgctgcctacaagcaggacttctcaaatccgccatcgacggcggcgtcctcaagtttg
gcagcttcgagctcaagtcgaagcggat
atccccctacttctcaacgcgggcgaattccacacggcgcgcctcgccggcgccatcgc
ctccgcctttgcaaagaccatcatcgag
5 gcccaggagaaggcggcctagagttcgacatcgtcttcggccggcctacaagggcac
ccgctgtgctccgccatcacatcaa
gctcggcgagctggcgcccagaacctggaccgcgtctcctactcgttgaccgcaagga
ggccaaggaccacggcgaggggcgg
caacatcgtcggcgcttcgctcaagggaagagggtcctgattgtcgacgacgtcatcac
10 cgccggcacccgccaaggaggacgcc
attgagaagatcaccaaggaggcggcatcgtcgccggcatcgtcgtggccctggaccgc
atggagaagctccccgctgcggatg
gcgacgactccaagcctggaccgagtgccattggcgagctgaggaaggagtacggcatcc
ccatctttgccatcctcactctggatg
15 acattatcgatggcatgaaggcctttgctacccctgaggatatcaagaacacggaggatt
accgtgcccaagtacaaggcgactgactg
attgagcgttcaatgtcagaaggagagaaagactgaaaagggtgaaagaaggagc
ttgttgtattattattctatctcgaat
cttctagatcttgcgtaaataaacaagcgtactagctagcctccgtacaactgcttga
20 attigatacccgatggagggcagttattttatt
ttgttttcaagattttccattcgcgttgaactcgtctcacatcgcgtgtattgcccgg
ttgcccatgtgtacgcgttcgggttacctcttcc
agatacagctcatctgcaatgcattaatgcattggacctcgcaaccctagtagcccttc
aggctccggcgaagcagaagaatagctta
25 gcagagctctattttcattttcgggagactagcattctgtaaaggcagcaatcggccag
cagttagtagggctccctctacctctcagg
gagatgtaacaacgccaccttatgggactatcaagctgacgctggcttctgtgcagacaa
actgcgccacaggttcttccctgacgcc
gctctcgcgcaggcaagggaactcgtatgaatactacgcaaagcacaagagaccggttgg
30 ccactccatggcctccccatctctctca
aagaccagcttcgagtcgaaggtagaccgttggccctaagtcgttagatgtcccttttgt
cagctaacatatgccaccagggtacgaaa
catcaatgggctacatctcatggctaacaagtacgacgaaggggactcgggtctgacaa

tcctgctcttatacgaaaaattgcctgactctg
 caaagggtgttgcttggaagatgatgtgccccccatcgctcttatctcataccccgc
 catctttctagattctcatcttcaacaagaggg
 gcaatccatgatcigcgatccagatgtgcttctggcctcatactctgccttcaggtgat
 5 gttcacttaattggtgacgaattcagctgatttg
 ctgcagtatgcttgtgttggtctttccaggcttgccagccatgagcgcttgagag
 catggtgtcacttataaactcgagtaacggcc
 acatatgttcactacttgaatcacatacctaattttgatagaattgacatgtttaaaga
 gctgaggtagctttaatgcctctgaagtattgtga
 10 cacagcttctcacagagtgaagaatgaaaagtggactccccctaataagtaaaagtctt
 gtctctgaacgggtgaagagcatagatccg
 gcatcaactacctggctagactacgacgtcaattctcgggcctttgacctttatatatg
 tccattaatgcaatagattcttttttttttttttttt
 ttttttttttttttttttgcgaattcgcagatcaaagtggacgttatagcatcat
 15 aactaagctcagttgctgagggaagccgtctactacc
 ttgcccattccatccagctccataccttgatactttagacgtgaagcaattcacactgta
 cgtctcgcagctctccttcccgtcttgcctcc
 ccactgggggtccatggtgcgtgtatcgtccctccacaattctatgccatggtacctcca
 gcttatcaatgccccgctaacaagtcgcctc
 20 tttgccttgatagcttatcgataaaacttttttccgccagaaaggctccgccacagac
 aagaaaaaaaaattcaccgcctagcctttggc
 cccggcatttggctaaacctcgagcctctctcccgtcttgggtatcaggaagaaaagaa
 aaaaatccatcgccaagggctgttttggc
 atcaccacccgaaaacagcacttctcgatcaaaagtggcccgccatgaagaccacgtgg
 25 aaggacatccctccggtgcctacgcac
 caggagtttctggacattgtgctgagcaggaccagcgaaactgccactcagatccgt
 gccggcttcaagattagcagaattcgag
 gtacgtcgcattgccatcgaggatgtctcattatcggggctcttgagaacgatcatg
 attgcatggcgaatgctaacacatagacag
 30 ccttctacactcgaaaggtaagttcacccaggagacgttttccgaaaagttcgctcca
 tctctgacagcttccctcgctccaggaca
 tccacccctccacaaggaccttctcaacacctctacgatgccgaccactcaagattg
 cccttggccagatgtccactgccaagcac

ctggtcgagaccatctcgcgcgactacgtccgtctcttgaaatacggccagtcgctctac

cagtgcgaagcagctcaagcgggcccgtc

tcggtcgcatggccacgtgggtcaagcgcctcaaggacccccctgctgtacctggaccagg

tccgccagcatctcggccgtcttcctc

5 catcgaccccaacaccaggaccctgctcatctcgggttaccccaatgttggaagtcag

cttcctgcgaagtatcacccgcgcgat

gtggacgtccagccctatgcttcaccaccaagagtctgtttgtcgccactttgactac

aagtacctgcgattccaggccattgataccc

ccgggtattctggaccaccctcttgaggagatgaacactatcgaaatgcagaggtatgtgg cgcggt.

10 [244] Creación de la cepa A5D a partir de la cepa Archy3 de *T. reesei*. El bglI de *T. reesei* nativo fue borrado de la cepa Archy 3 usando un vector de recombinación doble conocido en el arte. La resistencia a la higromicina fue empleada como marcador seleccionable para la eliminación del bglI. Además, el marcador de resistencia a la higromicina se flanqueó con sitios loxP. La platina de eliminación se muestra en la

15 figura 8. Se analizaron transformantes resistentes a la higromicina para la eliminación del bglI. Una cepa confirmada para la eliminación bglI fue luego transformada con un vector telomérico que codifica ere recombinasa y un amdS funcional para la selección de los transformantes para facilitar la remoción de la higromicina por expresión de la ere recombinasa y salirse por bucle de los sitios loxP. El vector telomérico que codifica

20 la ere recombinasa se muestra esquemáticamente en la figura 10. Los transformantes se obtuvieron primero en un medio de acetamida, luego transferidos a un agar de dextrosa de papa y transferidos a platos por réplica sobre un medio de higromicina para buscar sensibilidad a la higromicina. Las cepas sensibles a la higromicina fueron nuevamente transferidas al medio de dextrosa de papa y transferido por replicación a platos con un

25 medio medio de acetamida. Se seleccionó una cepa que perdió la habilidad de crecer en la acetamida, lo que indicaba una pérdida del vector telomérico. La secuencia de nucleótidos de la platina de expulsión del bglI se proporciona como SEQ ID NO 17:

aatggtaggaatgctgggatataggctctgtgctggcaagttgatggatcctcgaatgag

gccgccctgcaaggggaacatcagagat

30 ctaccattgcctccttgcccaatccactatcatacctacctcatgatcattcctgcgaa

ggtctaccagtaaataatttcctcgtcccggtgtt

catcatgtccagaacctcatctcgccaaattgacttggccacagtgtctggagctgggta

agcagcgtgccaaggaattgtgtcgagtc

tgtgccaggcattgtccccgacattgtgaacttcagccaggagaacttttcgatcgcacc
 tatgctgagcaccgtgggcatgcatggc
 ttcaataatgcagttcgagagggagtgtgtcatgccctaaagctcattggccacctccac
 aggctagctctacctgcatctgtatgga
 5 ctttccttgctcctccttcagaaaacctcttggtcgtcgcaggttaactgttggtgcc
 gtcattgttgacagtggatagccaaggcaaaa
 ccgtctgcttcaacggaagcattcggcggttggtgtcatcgtgtatcgategaccag
 gagaacccagacgagtgtgttcgagagaa
 tcatcgacgatgtgaagaggcgacgactagtatctagaagattataatcgaacaaatcag
 10 cgtttgtctgtcgggcgttgaggcgca
 gtgcccgccaaagcagcgtcgcaatatataggcagcgagagactgtcaacagccagccg
 ccatgtgatcgatectagccgtcttc
 cgtatctccctaaaccccttcttggggggcgggggcagcggcggttctaatttgctg
 gctgtctggataacgtgaatggtagacatg
 15 gtaatgttcggctcgcgaaacattgtacaattggagttagatcgagatggaaggaaa
 cgctccaaaactcggtagctgggtgccca
 tcaggtgtcagggcatagcgttctctgcaaatagaggnaagagaatagcactagtga
 gtgtgaatcacaatgaagaggaggtgt
 tgccggaatgcttgagcagcgtcaaagttgaacttgaagctatcacaattgcagggt
 20 aagtacatgttggtgccagttgacagcac
 agtgcgcggagcggaggatgtcgcggaagaggcgcgacgctaaccgggcttcttca
 gtgagcagaactcctgctgcaagagt
 tccttctctctgcgagatgacgtgaggcccaattgcagcttccctcgaacaaggtgatt
 gaacatctcttccctcacatttcacacac
 25 tacctcctcaattcacttctgcttcggcgttcttcattcatggtactgctctgatgc
 ctatcctgaagattgtattcctgcagtattcacgcc
 atcccaccttcggtcctcactcacagtcacaggtaaccgccttcacctcctcgcgatg
 atgtcggcaatctggtggatcaatgtcgg
 ttgagggccgccgtagtaggatgggcatggggaacgaggtcgccttcgcccacagat
 30 aacttcgtatagcatacattatacgaag
 ttatcctgggcttgtgactggtcgcgagctgccactaagtggggcagtaccattttatcg
 gacctatccagctatgggacctcgcga
 aatttttacatcattttcttttctcagtaacggccacctttgtaaagcgtaccagc

aaacaaattgcaattggccgtagcaaggtagt
 cagggttatcgtgatggaggagaaggctatatcagcctcaaaaatatgttccagctgg
 cggaagcccgaaggtaagtggattctt
 cgccgtggctggagcaaccgggtgattccagcgtctccgactggactgagcaattcagc
 5 gtcacggattcacgatagacagctcag
 accgtccacggctggcggcattattggttaaccggaaactcagtccttggccccgt
 cccgaagggaaccgacttaccaggctgg
 gaaagccagggatagaatacactgtacgggcttcgtacgggagggtcggcgtagggttgt
 tccaagttttacaccccccaagaca
 10 gctagcgcacgaaagacgcgggagggttggtaaaaaagggcgaaaattaagcgggagac
 gtatttaggtgctagggccggttctt cccatttttcttcggttcccttctctctggaagactttctctctcttcttctc
 ttcttccatcctcagtcctcttcttcccatcatccat
 ctctcaccctccatcctaactccatcacatcacaatcgatagaaaaagcctgaactcac
 cgcgacgtctgtcgagaagtttctgatcga
 15 aaagttcgacagcgtctccgacctgatgcagctctcggaggggcgaagaatctcgtgcttt
 cagcttcgatgtagggggcgtggatat.
 gtcttcgggtaaatagctgcgccgatggtttctacaaagatcgttatgtttatcggcac
 ttgcatcggccgcgctcccattccggaa
 gtgcttgacattggggaattcagcgagagcctgacctattgcatctccgccgtgcacag
 20 ggtgtcacgttgcaagacctgcctgaaac
 cgaactgcccgtgttctgcagccggtcgcggaggccatggatgcgatcgtgcggccga
 tcttagccagacgagcgggttcggcc
 cattcggaccgcaagggaatcggtaatacactacatggcgtgatttcatatgcgcgattg
 ctgatcccatgtgtatcactggcaactgt
 25 gatggacgacaccgtcagtgctcgcgtcgcgcaggctctcgtatgagctgatgcttgggc
 cgaggactgccccgaagtcggcacct
 cgtgcacgcggatttcggctccaatatgtcctgacggacaatggccgcataacagcgggt
 cattgactggagcgaggcgatgttcgg
 ggattcccaatacagggtcggcaacatcttcttctggaggccgtggttgcttgtatgga
 30 gcagcagacgcgctacttcgagcggagg
 catccggagcttgaggatcggcggtccggcggtatgtctccgattggtcttgac
 caactctatcagagcttggttgacggcaa
 ttctgatgatgcagcttgggcgcagggtcgtatgcgacgcaatcgtccgatccggagccgg

gactgtcgggcgtacacaaatgccc
 gcagaagcgcggccgtctggaccgatggctgtgtagaagtactgccgatagtgaaacc
 gacgccccagcactcgtccgagggc
 aaaggaatagagtagatgccgaccgggatccacttaacgttactgaaatcatcaaacagc
 5 ttgacgaatctggatataagatcgttgg
 gtcgatgtcagctccggagttgagacaaatgggttcaggatctcgataagatacgttca
 tttgtccaagcagcaaagagtgcccttctagt
 gatttaatagctccatgtcaacaagaataaaacgcgttcgggttacctcttccagata
 cagctcatctgcaatgcattaatgcattggac
 10 ctcgcaaccctagtagcgccttcaggctccggcgaagcagaagaatagcttagcagagtc
 tattttcatttctgggagacgagatcaag
 cagatcaacggctgtcaagagacctacgagactgaggaatccgctcttggctccacgcga
 ctatatatttgtctctaattgtactttgacat
 gtcctcttcttactctgtagacttgactatgaaaattccgtcaccagccccctgggttc
 15 gcaaagataattgcactgtttcttccctgaaactct
 caagcctacaggacacacattcatcgtagggtataaacctcgaaaatcattcctactaaga
 tgggtataacaatagtaaccatgggtgccta
 gtgaatgtccgtaacacccaatacgcggccgataactcgtatagcatacattatacg
 aagttatacttggcgcgcctagtggaaacac
 20 gagcacataagcttttacctatgggtatcgcttgcactctacgcgccgttgatgggtggagg
 atgggtggacgttcccagagaccctacgagc
 tgtggcatcgtcaaaactgtgccacagaccttgtcttgccttcataaacctcgaggagtg
 ttccagactcatcatccatacacaagcagta
 ttaatcaaagaaactcggcgcgaatggcaaaaatgggttgcaaacagaaaactatggcct
 25 ctctctattccatcattaactactctacccgt
 ttgtcataacaacatcattaaaacccttatgcgtcagggtgtagcatccttgatctgttgc
 tcctccaacggccagttctcaatcgttacctctt
 ctcccaccaactcaaactcaagcttcacagactcgtcgggtgtcaaggctagctcatact
 tgccgggggtatacaatccggttccgtgag
 30 aatcaacacgggcgagagcactgacagggatggggatgctgagcttgggaagagtgaccag
 gcttgatgtcggcaagtcggtcgaat
 ccgacgagccacttgttcgggtacggggctgggccagcgttgcttgcgaacaaacagc
 atggccgtatatggggactccgtcttgc

- ccgagttcttgatgttggcctcgaaggtgaagacgggaatctgctcgctgtaagtgtatc
 cgggggtgaggagcagagaggatcgatga
 ggtgtgaacttgaggctcttgggggtggctggcgagagtccttgaaggtggtgtagaa
 gagaccactgccaaactcgtagacgggt
- 5 ttgccggtgtaccagatgtaagtctgtccagggttgactttccatcgggtcggagggtc
 atgtcattctgggggaattggtgaacatactc
 agccgggtactgagtgggtgaccagtcggccggcaggagcacgcttgccagagagaatgtc
 gaagaggggcaacgcctcccactg
 gccgggatatccgcccagacgagggaagtacacttcttgttgccttgagcgaggatga
- 10 gtctacctgaccaccgcccatttgagg
 acgacaaggggttggcgacctcgtgagctgcttgatgagatccagctgattaccgggc
 caagcaatgtccgtgcggtcagcgcct
 cctgticaatggtgtgtcaattccaccgaggtagatgatggcatccgacttcttggcgg
 cagcaatggccttggcaaagccagtgggtg
- 15 ctgttgcggcgatctctgtgccgagttcaaagtgacgtgatagccggccttcttagca
 gcttccagagggctgatgaggtatggggc
 agggccatagtagttgccttgcatttgggtgtggcattggccoatggtcgatcagagc
 aatgctgcgcaccttcttggacagaggga
 gagtgccatcgttcttgagcaggacgatgccctcaacagcagcctcgtacgagatgtt.
- 20 [245] La secuencia del vector telomérico, pTTT-cre, se proporciona como SEQ ID NO
 18: ttgtacaaagtggatcgcgccgcgcgccagctccgtgcgaaagcctgacgcaccggta
 gattcttggtagcccgatcatgacg gcggcgggagctacatggccccgggtgattattttttgtatctacttctgaccttt
 tcaaatatacgggtcaactcatcttctacttgaga
 tgcggcctgcttggattgcgatgttgcagcttggcaaattgtggctttcgaaaacaca
- 25 aaacgattccttagtagccatgcattttaagat
 aacggaatagaagaaaggaaattaaaaaaaaaaaaaaaaacaaacatcccgttcataac
 ccgtagaatcgccgctcttctgtatcc
 cagtaccagtttatttgaatagctcgcccgctggagagcatcctgaatgcaagtaacaa
 ccgtagaggctgacacggcagggtgttgc
- 30 agggagcgtcgtgttctacaaggccagacgtcttcgcgggtgatataatgtatgtttga
 ctgcaggctgctcagcgacgacagicaagt
 tcgcccctcgtcgttgtgcaataatgcagtggggaagccacaccgtgactcccatcttt
 cagtaaagctctgttgggtgttatcagcaat

acacgtaattaaactcgttagcatggggctgatagcttaattaccgtttaccagtgcc
 tgggtctgcagctttccttggcccgtaaaattc
 ggcgaagccagccaatcaccagctaggcaccagctaaacctataattagtctcttatca
 acaccatccgctccccgggatcaatga
 5 ggagaatgagggggatgcggggctaaagaagcctacataacctcatgccaactccagt
 ttacactcgtcagccaacatcctgac
 tataagctaacacagaatgcctcaatcctgggaagaactggccgctgataagcgcgcccg
 cctcgcaaaaaccatccctgatgaatg
 gaaagtccagacgctgcctgcggaagacagcgttattgatttcccaaagaaatcggggat
 10 cctttcagaggccgaactgaagatcaca
 gaggcctccgctgcagatcttgtgtccaagctggcggccggagagttgacctcgggtgaa
 gttacgctagcattctgtaaacgggcag
 caatcgcccgagcagttagtagggctcccctctacctctcaggagatgtaacaacgccacc
 ttatgggactatcaagctgacgctggctt
 15 ctgtgcagacaaactgcgccacgagttcttccctgacgccgctctcgcgcaggcaaggg
 aactcgaatgaatactacgcaaagcaca
 agagacccggttggtccactccatggcctccccatctctctcaagaccagcttcgagtca
 aggtacaccggtgccccctaagtcgttagat
 gtccctttttgtcagctaacatatgccaccagggtacgaaacatcaatgggctacatct
 20 catggctaacaagtacgacgaaggggac
 tcggttctgacaaccatgctccgcaaagccgggtgccgttcttacgtcaagacctctgtc
 ccgcagaccctgatggtctgcgagacagt
 caacaacatcatcgggcgcaccgtcaaccacgcaacaagaactggtcgtgcggcggcag
 ttctgggtggtgaggggtgcgacgttg
 25 ggattcgtgggtggcgtcatcgggtgtaggaacggatatcgggtggctcgattcagtgccgg
 ccgcgttcaacttcctgtacgggtctaagg
 ccgagtcatgggcggctgccgtatgcaaagatggcgaacagcatggagggtcaggagacg
 gtgcacagcgttgcgggcccgtta cgcaactctgttgagggtgagtccttcgctcttcttcttcttctgctctataccaggcc
 tccactgtcctccttcttcttcttcttatactatatac
 30 gagaccggcagtcactgatgaagtatgttagacctccgctcttcaccaaatcegtctc
 ggtcaggagccatggaaatcagactcca
 aggtcatccccatgccctggcgccagtcggagtcggacattattgcctccaagatcaaga
 acggcggggtcaatatcgggtactacaa

ctfcgacggcaatgtccttccacacctctatcctgcgcggcgtggaaaccaccgtcgc
 cgcactcggcaaaagccgggtcacaccgt
 gaccccgtggacgccatacaagcacgatttcggccacgatctcatctcccatatctacgc
 ggctgacggcagcggcgacgtaatgcg
 5 cgatatcagtgcatccggcgagccggcgattccaaatatcaaagacctactgaacccgaa
 catcaaagctgttaacatgaacgagctc
 tgggacacgcctctccagaagtgggaattaccagatggagtagcttgagaaatggcgggag
 gctgaagaaaaggccgggaagggaac
 tggacgccatcatcgcgccgattacgcctaccgctgcggtacggcatgaccagtccggt
 10 actatgggtatgcctctgtgatcaacctg
 ctggattcacgagcgtggttggtccggttaccttgcggataagaacatcgataagaag
 aatgagagttcaaggcggtagtgagctt
 gatgccctcgtgcaggaagagtatgatccggaggcgtaccatggggcaccgggtgcagtg
 caggttatcggacggagactcagtga
 15 agagaggacgttggcgattgcagagggaagtggggaagttgctgggaaatgtggtgactcc
 atagctaataagtgtcagatagcaattt
 gcacaagaaatcaataccagcaactgtaaataagcgtgaagtgacctatgcatgctacg
 aaagagcagaaaaaacctgccgtag
 aaccgaagagatatgacacgcttccatctctcaaagggaagaatcccttcagggttgcgtt
 20 tccagtctagacacgtataacggcacaag
 tgtctctaccaaattgggttatatctcaaatgtgatctaaggatggaaagcccagaatat
 cgatcgcgcgcagatccatatataggcccc
 ggggtataattacctcaggaaatagctttaagtagcttattaagtattaaaattatata
 atttttaataataactataatttcttaataaataaggattt
 25 ttaagctttatataaaatataataataaaataatattatataagctttttattaataa
 ataaaaatagctaaaaatataaaaaaataagctttaaa
 atacttatttttaattagaattttatatttttaataataagatctttacttttta
 taagcttcctaccttaaaatttttacttttttactatttta
 ctatatcttaataaaggctttaaaaatataaaaaaatcttcttatataattataagcta
 30 taaggattatataatatttttttaatttttaagtaa
 gtattaaagctagaattaaagttttaatttttaaggctttatttaaaaaaggcagtaa
 tagcttataaaaagaaatttcttttctttatactaaa
 agtactttttttaataagggttaggggttaggggttaggggttaggggttaggggtta

gggttagggtaggggttagggtaggg
gttaggggttaggggttagggtaaggggttaaacaagccacgttgtgtctcaaatctctg
atgttacattgcacaagataaaaatatcat
catgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccata
5 ttcaacgggaaacgtcttgctcgaggccgc
gattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcg
ggcaatcagggtgcgacaatctatcgattgtat
gggaagcccgatgcgccagagttgtttctgaacatggcaaaggtagcgttgccaatgat
gttacagatgagatggtcagactaaact
10 ggctgacgggaatttatgcctcttcgaccatcaagcattttatccgtactcctgatgatg
catggttactcaccactgcgatccccggggaa
aacagcattccaagggtattagaagaatatcctgattcagggtgaaaatattgttgatgcgct
ggcagtgttcctgcgccgggtgcattcgattc
ctgtttgtaattgtccttttaacagcgcgcgtatttcgtctcgctcaggcgcaatcac
15 gaatgaataacggtttggttgatgcgagtgattt
tgatgacgagcgtaatggctggcctgttgacaagctcgaaagaaatgcataagctttt
gccattctcaccggattcagtcgtcactcat
ggtgatttctcacttgataaccttattttgacgaggggaaattaatagggttgattgat
gttgagcagagtcggaatcgagaccgatacca
20 ggatcttgccatcctatggaactgcctcgggtgagttttctcttcattacagaaacggct
ttttcaaaaatattggtattgataatcctgatatg
aataaattgcagtttcattigatgctcgatgagttttctaatacagaattggttaattgg
ttgtaacactggcagagcattacgctgacttgac
gggacggcggtttgttgaataaatgaacttttgctgagtgaaaggatcagatcacgca
25 tcttcccgacaacgcagaccgttccgtgg
caaagcaaaagttcaaaatcaccaactgggtccacctacaacaaagctctcatcaaccgtg
gctccctcactttctggctggatgatggg
gcgattcaggcctgggtatgagtcagcaacaccttcttcacgaggcagacctcagcgggtt
aaacctaaccctaaccctaaccctaacc
30 taaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaacc
taaccctaaccctaattggggtcgatctgaac
cgagggatgagggttctatagactaatctacaggccgtacatggtgtgattgcagatgcga
cgggcaagggtgtacagtgtccagaagga

ggagagcggcataggtattgtaatagaccagctttacataataatcgctgttgctactg
 actgatgaccttcttccctaaccagtttccta
 attaccactgcagtgaggataaccctaactcgcctctggggtattattatactgattagc
 aggtggcttatatagtgctgaagtactataag
 5 agtttctgcgggaggagggtggaaggactataaaactggacacagttaggatagagtgatg
 acaagacctgaatgttatccctccggtgt
 ggtatagcgaattggctgaccttgcatggtaatggttaggcagggttttgcagagg
 gggacgagaacgcgttctgcgatttaacg
 gctgctgccccaagctttacggttctctaattggcgggccgctcaggcgcacgtcccat
 10 ggccattcgaattcgtaatcatgggtcatag
 ctgttctctgtgtgaaattgttatccgctcacaattccacacaacatacga gccggaagc
 ataaagtgtaaagcctgggggtgcctaataga
 gtgagctaactcacattaattgcgttgcgctcactgcccgtttccagtcgggaaacctg
 tcgtgccagctgcattaatgaatcgcccaa
 15 cgcgcggggagaggcgggttgcgtattgggcgctcttccgcttccctcgtcactgactcg
 ctgcgctcggctgttcggctgcggcgag
 cgggtatcagctcactcaaaaggcggtaatacggttatccacagaatcaggggataacgcag
 gaaaagaacatgtgagcaaaaggccag
 caaaaggccaggaaccgtaaaaaggccgcgttgctggcggttttccataggctccgcccc
 20 cctgacgagcatcacaaaaatcgacgc
 tcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccctgga
 agctccctcgtgcgctctcctgttccga
 ccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctc
 atagctcacgctgtaggtatctcagttcgggt
 25 gtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccagccgctg
 cgccttatccggtaactatcgtcttgagt
 ccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagca
 gagcgagggtatgtaggcggtgctaca
 gagtcttgaagtgggtgacctaacacggctacactagaagaacagtatttggtatctgc
 30 gctctgctgaagccagttaccttcggaaaa
 agagttggtagctcttgatccggcaacaaaccaccgctggtagcgggtgtttttgtt
 tgcaagcagcagattacgcgcagaaaaaa
 aggatctcaagaagatcctttgatctttctacggggctcgtacgctcagtggaaacgaaaa

ctcacgtaagggaatttggatcatgagattat
 caaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaataaatctaaa
 gtatatatgagtaaacttggctgacagttacca
 atgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcacatagttgc
 5 ctgactccccgtcgtgtagataactacgatacg
 ggagggcttaccatctggccccagtgtgcaatgataccgcgagaccacgctcaccggc
 tccagatttatcagcaataaaccagcc
 agccggaagggccgagcgcagaagtggctctgcaactttatccgcctccatccagtctat
 taattgttgcgggaagctagagtaagta
 10 gttcggcagtaataagtttgcgcaacgttgttgcattgctacaggcacgtgggtgcac
 gctcgtcgtttggatggcttcattcagctccg
 gtcccaacgatcaaggcgagttacatgatccccatgttgtcaaaaaagcggttagct
 ccttcggctctccgatcgttgcagaagta
 agttggcgcagtggtatcactcatggttatggcagcactgcataattctcttactgtca
 15 tgccatccgtaagatgcttttctgtgactggg
 agtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcggg
 cgtcaatacgggataataccgcgccacat
 agcagaactttaaagtgtcatcattggaaaacgttcttcggggcgaaaactctcaagg
 atcttaccgctgttgagatccagttcgatgt
 20 aaccactcgtgcacccaactgatcttcagcatcttttacttcaccagcgtttctgggt
 gagcaaaaacaggaaggcaaaatgccgcaa
 aaaagggaataagggcgacacggaaatgtgaatactcatactcttcttttcaatatt
 attgaagcatttatcagggttattgtctcatga
 gcggatacatattgaatgtatttagaaaaataacaaataggggtccgcgcacatttc
 25 cccgaaaagtgcacctgacgtctaagaaa
 ccattattatcatgacattaacctataaaaaataggcgatcacgaggcccttctgtctcg
 cgcgttccggtgatgacggtgaaaacctctga
 cacatgcagctcccgagacggtcacagcttgtctgtaagcggatgccgggagcagacaa
 gcccgtcagggcgcgtcagcgggtg
 30 ttggcgggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgc
 accataaaattgtaaacgttaatatattgtt
 aaaattcgcgttaaatTTTTgttaaatcagctcatttttaaccaataggccgaaatcgg
 caaaatcccttataaatcaaaagaatagcccgga

gatagggttgagtgtgtccagtttgaacaagagtcactattaaagaacgtggactc
 caacgtcaaagggcgaaaaaccgtctatc
 agggcgatggcccactacgtgaaccatcacccaaatcaagtttttggggtcgaggtgcc
 gtaaagcactaaatcggaaccctaaagg
 5 gagccccgatttagagcttgacggggaaagccggcgacgtggcgagaaaggaaggaa
 gaaagcgaaaggagcgggcgcta
 gggcgctggcaagtgtacgggtcacgctgcgcgtaaccaccacacccgccgcgcttaatg
 cgccgctacagggcgcgctactatggt
 tgctttgacgtatgcgggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatca
 10 ggcgccattcgccattcaggtgcgcaa
 ctgttgggaagggcgatcgggtcggggcctcttcgctattacgccagctggcgaaaggggg
 atgtgctgcaaggcgattaagttgggt
 aacgccagggtttccagtcacgacgttgtaaaacgacggccagtgcccaagcttacta
 gatgcatgctcgagcggccgacgtgt
 15 gatggatatctgcagaattcgcccttgactagtgtctctatcctggtggcagcgctcaa
 gtaccagaggcagcagcgggcttagga
 gcggcctgggtgttctccgcaccctctacatgctgggctatattatagcgacaagccg
 aacggcaccggcaggtacaatggttcgct
 gtacttgcttgcgcaagcgggtctttggggattgagcgcatttggtgttgc aaaggattt
 20 gatgtaaatgtagtcgacatcttagcacaga
 ggggagagttgataaaatgtggtctgtttgaatgatagtcgggttcgtgacctatattcg
 tgatagtggagataggtcgcgcctatcttat
 cgggccggagcaaaaattccaccgcagcggggtgagttttcgttatacagccatcccact
 tccagcttcaaattgtcagtttaatccagc
 25 ccaattcaatcattggagaaccgccatcatgtcttcgaagtcacccctcccctacgcaat
 tcgcgcaaccaaccatcccaacccttaac
 atctaaactcttctccatcgccgaggagaagaaaaccaacgtcaccgtctccgcagacgt
 tactacttcggcgagctcctcgatcttgc
 tgaccgtacatcctgcaccaatgcccctccaggataacaaatagctgatgcgtagtgagt
 30 acaggcctaggccccctatatcgagttct
 gaaaaccacatcgacatctcaccgatctcacccttcgacctttcctcgctccaatc
 cctcgcgacaaagcacaacttcctcatcttt
 gaggaccgcaagttcatcgacatcggaacaccgtgcaaaagcagtagccacgggtggcgct

ctccgcatctccgaatgggcacacat
 catcaactgcgccatcctgccgggcgaagggatcgtcgaggccctcgacagacaaccaa
 gtctcctgactttaagacgcgaatca
 acgaggctcctgattcttgcgagatgacgagtaagggatctcttgcgacaggggagtc
 5 acaggcacgctcggttgagtacgcgcg
 gaagtataaggggtttgtgatgggattcgtgagtacaagggcgttgagtgggtgctgcc
 cgaacagaaaggagagcaggattt
 tgtcgtctttacgactgggggtgaatctgtcggataagggggataagctggggcagcagta
 tcagacacctgggtcggcggttgggcg
 10 aggtgcggactttatcattgcgggtaggggcatctataaggcggacgatccagtcgaggc
 gggtcagagggtaccgggaggaaggct
 ggaaaacttacgagaaaagagttggactttgagtgtgagtggaaatgtgtaacggtattg
 actaaaagggatccatatgtttattgcagc
 cagcatagtlattaccagaaagacccactgacggctctagtagtattcgaacagatatt
 15 attgtgaccagctctgaacgatatgtccct
 aatctggtagacaagcactgatctaccccttggaaacgcagcatctaggctctggctgtgc
 tctaaccctaactagacgattgatcgaga
 ccatccaatactgaaaagtctctatcagaggaaatccccaacattgtagtagtcagggtc
 ctttgtggctgggagagaattggttcgctcc
 20 actgattccagttgagaaagtgggctagaaaaagtcttgaagattggagtgggctgtg
 gttatctagttacttctcgaactctgtacatgt
 ccggctcgcgacgtacgcgtatcgatggcgccagctgcaggcggccgctgcagccacttg
 cagtcccgtggaattctcacggtgaat
 gtaggcctttttagggtaggaattgtcactcaagcaccaccaactccattacgcctcc
 25 cccatagagttcccaatcagtgagtcattg
 cactgttctcaaatagattggggagaagttgacttccgccagagctgaaggctgcacaa
 ccgcatgatatagggtcggcaacggca
 aaaaagcacgtggctcaccgaaaagcaagatgtttgcgatctaactccaggaacctgga
 tacatccatcatcacgcacgaccacttt
 30 gatctgctggtaaactcgatttcgccctaaaccgaagtgcgtggtaaactctacacgtggg
 cccctttcgggtatactgcgtgtgtcttctcta
 ggtgccattctttcccttctctagtgtgaattgtttgttggagtccgagctgtaa
 ctacctctgaatctctggagaatggtggactaac

gactaccgtgcacctgcatcatgtatataatagtatcctgagaaggggggttggagca
 atgtgggactttgatggcatcaacaaag
 aacgaagacgcctctttgcaaagtttggctacggagaactggatacttgt
 tgtgtctctgtgtattttgtggcaacaagag
 5 gccagagacaatctattcaacaccaagcttgccttttgagctacaagaacctgtgggg
 tatatatctagagttgtgaagtcggtaatcc
 cgctgtatagtaatacgaagtcgcataaatactccgaagctgctgcgaacccggagaat
 cgagatgtgctggaaagcttctagcgag
 cggctaaattagcatgaaaggctatgagaaattctggagacggcttgtgaatcatggcg
 10 ttccattcttcgacaagcaaagcgttccgt
 cgcatagcaggcactcattcccgaaaaaactcggagattcctaagtagcgatggaaccg
 gaataatataataggcaatacattgagti
 gcctcgacgggtgcaatgcaggggtactgagcttggacataactgttccgtacccacct
 cttctcaacctttggcgtttccctgattcag
 15 cgtacccgtacaagtcgtaatcactattaacccagactgaccggacgtgtttgcccctc
 atttggagaaataatgtcattgcgatgtgtaa
 tttgcctgcttgaccgactggggctgttcgaagcccgaatgtaggattgttatccgaact
 ctgctcgtagaggcatgttgtgaatctgtgtc
 gggcaggacacgcctcgaagggtcacggcaagggaaccaccgatagcagtgcttagtag
 20 caacctgtaaagccgcaatgcagcat
 cactggaaaatacaaaaccaatggctaaaagtacataagttaatgcctaaagaagtcatat
 accagcggctaataattgtacaatcaagt
 gctaaacgtaccgtaatttgccaacggcttgtggggtgcagaagcaacggcaaagcccc
 acttccccacgtttgtttcttactcagtc
 25 caatctcagctggatgaccccccaattgggtcgttgtttgtccggtgaagtgaagaag
 acagaggtaagaatgtctgactcggagcg
 tttgcatacaaccaagggcagtgatggaagacagtgaatgttgacattcaaggagtat
 tttagccagggatgcttgagtgatcgtgta
 aggaggtttgtctgccgatacgacgaatactgtatagtcacttctgatgaagtggccat
 30 attgaaatgtaaagtcggcactgaacaggc
 aaaagattgagtgaaactgcctaagatctcgggccctcgggccttcggcctttgggtgt
 acatgtttgtgctcgggcaaatgcaaagt
 gtgtaggatacgaacacactgctgcctttaccaagcagctgagggtatgtgataggcaa

- tggtcaggggccactgcatggtttcgaat
 agaaagagaagcttagccaagaacaatagccgataaagatagcctcattaaacggaatga
 gctagtaggcaaagtcagcgaatgtgt
 atatataaagggttcgaggtccgtgcctccctcatgcttccccatctactcatcaactea
 5 gatcctccaggagacttgtaaccatcttttga
 ggacagaaacccaatagtaaccatcacaaagttgtacaaaaagcaggctcaccatga
 gcaacctgctcaccgtccaccagaacc
 tcctgccctccctgtcgacgccacctctgacgaggtccgcaagaacctcatggacatgt
 tccgcgaccgccaggccttttagcgagca
 10 cacctggaagatgctcctcagegtctgccgatcttggggcgcctggtgcaagctcaacaa
 ccgcaagtggttccccgccgagccgga
 ggacgtccgcgactacctcctctacctgcaggcccaggcctggccgtcaagaccatcca
 gcagcacctcggccagctcaacatgc
 tcaccgacgctctggcctgcctcgccttagcgactctaacgccgtcagcctggtcatgc
 15 gccgcatccgcaaggagaacgtcgacg
 ctggcgagcgcgccaagcaggccctcgccttcgagcgcaccgacttcgaccaggtccgca
 gcctcatggagaacagcgaccgctg
 ccaggatatccgcaacctcgcctttctcggcattgcctacaacacctgctccgcattgc
 cgagatcggccgcatccgcgtcaaggac
 20 atctctgcaccgacggcgccgcatgctcattcacatcgccgcaccaagacctcgtg
 tctaccggcggtcgagaaggccctc
 agcctcggcgtcaccaagctcgtcgagcgctggatttctgttccggcgctgctgacgac
 cccaacaactacctcttctgccgcgtccg
 aaagaacggcgtcgccgcccttctgccacctctcagctcagcaccgagccctggaggg
 25 catctttgaggccaccaccgcctcat
 ctacggcgccaaggacgactctggccagcgctacctcgcctggtctggccactctgcccg
 agtcggcgctgcccagacatggccc
 gagccggcgctcagcatccccgagattatgcaggccggcggtggaccaacgtcaacatcg
 tcatgaactacatccgcacctcgactctgagaccggcgccatggtccgactcctcgaggacggcgactaaaccagctttc.
 30 [246] Creación de la cepa MAD6 a partir de la cepa A5D de *T. reesei*. La *egl3* fue
 borrada de la cepa A5D usando el método descrito anteriormente para la eliminación del
bgl1. Un esquema de la platina de eliminación se muestra en la figura 9. La resistencia a
 la Higromicina se empleó como el marcador selectivo para la eliminación del *egl3*. El

marcador de resistencia a la higromicina estaba flanqueado por sitios loxP. Se confirmó un transformante con la eliminación del egl3. El marcador de higromicina fue removido de la cepa como se describe en la creación de la cepa ASD. La secuencia de nucleótidos de la platina de borrado del egl3 se proporciona como SEQ ID NO: 19:

5 gggaggtaggcgcagatacgggtgcatgggacccgaacccgtaaccggaacacgacctat
 cagccctccaactcacacctctcgc
 ctatcactatcctagatagttcatcggccaaactcatgtaacctagctacctacctacgtg
 gtaagaatgcgggctatcatgtctcacggcg
 cgggtacatgtcgggtatctcgtcgttccccgcagggtgacgtcgggaatccatgcaagtac
 10 tccctgaaatcgagacgacagagagaac
 aaccaacgcgcgttaaacgcttcatgttcatctaagaggcacattcgaagaactagcttaa
 cactagacctggcttttcgacccccctcc
 gcagaaagccgttttctctcaatcctcccggttggctttgtcagtcgtaacttgc
 gcgctaacagagtcttgacgcagcgtttgc
 15 gcatcagtccttgaggcggttcacgggactaggacaacaggggatgtgacaggccggata
 gtaattatgggttatccgggtaagca
 ggggaatttacgaggccgctttacgtgggggaacagccacttgcggggggaaggagtag
 taggcgactcggtcgatgagtcga
 ggtgtctggttgacttgactgcagagcgtaggtaattgagatcgggcaacattatcgg
 20 gtccggctcgggtatggccgagttgcgact
 gcttggtcattcggcgaagctgatgtcgtggtatcctgaagcatcgatatcgaaacat
 gatggtcagtcctatctgacgtgtgcggtga
 caagcgagtcgggattttgtgacatgacgttcaacttcagtcaatgccttagggctcgat
 aagattaagattgggttctggcagcgggtcta
 25 gaacaccgccacaaattctgtccattgaggagcgtgatgtctaggcgcatacctaacacg
 gagctgtatgaccggcagctcaacgga
 cttctcttcgttcaacggcagtcctatttgcggtacacgaatggatctttctcctggctt
 tgaagtgcgcagtggcgtgcgaatgtatagat
 gtctcgctacctagaaaagctggctttctgacagggtccctccacctctctaccaac
 30 gacaaactgaacaagtatctggcggtttccc
 aacgccgaataggccagtcgccaatactccctccagccctgattggccccctcgaagtat
 cgccatgtctgtgtgtgagattatcgat
 ggacgtcactcccccaacctacaggaagagcaaatgggagcagtggttctgcaatgagct

atataatagatcgctc gatctcatacaaa
 ttgtatgctcagtcatacaacgagcgggtccaagatcccttctccaacgaccctcgaaa
 cattgcaacccgggtgcagcctgaacttgtt
 cgtatagcctagaaagcgacgccatcttcacgttttacgcgattagcctcatggctatft
 5 gtgccgaagtgggagttgtatggtagcagtg
 aggagattgtggctacgacacagcgggttctcttgagcggcttacatctccgattagg
 cctgcgtacgatccagatcatgggaaact
 ttacaatggcttactcgtttatctcaacactgagctccaaticactctatgcattgat
 taacacgtttggcatgtggttcttcagctgtaaact
 10 cttcagcttcccaagaattgcaacctcgctgattgctaatagtgttgcattgcgttgcac
 ctggtgcggcagtgcaaaggagagtgcaaag
 tagccggcagattaatttaagcttatctactcaggggtaaacagccgtaaaggaccttt
 tgatctaacatgccgatgtgtatgtatgatca
 cgcaatgcccaccatatcttggcagtcagatttgcctggcgcgccaagtataacttcg
 15 tataatgtatgctatacgaagttatcggccg
 gcgtattgggtgttacggagcattcactaggcaacctatggttactattgtatacccatct
 tagtaggaatgattttcgagggttatacctacg
 atgaatgtgtgtcctgtaggcttgagagttcaaggagaaacagtgcaattatcttgcg
 aaccaggggctggtgacggaatttcata
 20 gtcaagctatcagagtaaagaaggagcagtgcaaaagtacaattagagacaaatatata
 gtcgcgtggagccaagagcggattcctc
 agtctcgtaggtctcttgacgaccgttgatctgcttgatctcgtctcccgaatgaaaa
 tagactctgctaagctattctctgcttcgccg
 gagcctgaaggcgctactagggttgcgaggtccaatgcattaatgcattgcagatgagct
 25 gtatctggaagaggtaaaccgaaacg
 cgttttattcttgtgacatggagctattaaatcactagaaggcactctttgctgcttgg
 acaaatgaacgtatcttatcgatcctgaacac
 catttgtctcaactccggagctgacatcgacaccaacgatcttatccagattcgtcaa
 gctgtttgatgatttcagtaacgttaagtggat
 30 cccggtcggcatctactctattctttgccctcggacgagtgctggggcgctcgggttcca
 ctatcggcgagctactctacacagccatcg
 gtccagacggccgcgttctgcgggcgatttgtgtacgcccagagtcgggctccggat
 cggacgattgcgtcgcacgcacctgc

gcccaagctgcatcatcgaaattgccgtcaaccaagctctgatagagttggtcaagacca
 atgcggagcatatacgcccggagccgc
 ggcgatcctgcaagctccggatgcctccgctcgaagtagcgcgtctgctgctccatacaa
 gcccaaccacggcctccagaagaagatg
 5 ttggcgacctcgtattgggaatccccgaacatgcctcgcctcagtcfaatgaccgctgtt
 atgcggccattgtccgtcaggacattgttg
 gagccgaaatccgctgcacgaggtgccggacttcggggcagtcctcggcccaaagcatc
 agctcatcgagagcctgcgcgacgg
 acgcactgacgggtgtcgtccatcacagtttgccagtatacacatggggatcagcaatcg
 10 cgcataatgaaatcacgccatgtagtgatt
 gaccgattccttgcgggtccgaatggggccgaaccgcctcgtctggctaagatcgcccgag
 ccatcgcatccatggcctccgcgaccg
 gctgcagaacagcgggcagttcgggttcaggcaggtcttgcaacgtgacaccctgtgcac
 ggcgggagatgcaatagggtcaggctct
 15 cgctgaattcccaatgtcaagcacttccggaatcgggagcgcggccgatgcaaagtgcc
 gataaacataacgatctttgtagaacc
 atcggcgagctatttaccgcaggacatatccacgcccctctacatcgaagctgaaagc
 acgagattcttcgccctccgagagctgc
 atcaggtcggagacgctgtcgaactttcgtatcagaacttctcgacagacgtcgcggtg
 20 agttcaggcttttcatatcgattgtgatgtg
 atggagttgagatggaggtgaggagatggatgatgggaaagggaagatggactgaggatgg
 aagaagagaagaagagagagaga
 gaaagtcttcaggagagaaagggaaccgaagaaaaatggggagggaaccggccctagca
 cctaaatacgtctcccgttaattttc gccctttttcaccaaaccctccgcgttttcgtcgttagctgtcttggggggtgtgta
 25 aaacttgggaacaaccctacgccgaacctcc
 cgtacgaagcccgtacagtgtattctatccctggctttccagcctggtaagtcgggtcc
 ctccgggacggggccaaggagactgagtt
 tccgggttaaccaataatgccgccagccgtggagcggctctgagctgtctatcgtgaatcc
 gtgacgctgaattgctcagtcgaagtcgg
 30 agacgctggaatccaccggttgcctccagccacggcgagaatccacttaccttccgggct
 tccgccagctggcaacataattttgaggc
 tgatatagccttctctccatcacgataagccctgactaccttgctacgggccaattgca
 attgtttgctgggttacgctttacaaaagggtgg

ccgttactgagcaaaaagaaaatgatgtaaaaatttgcgagtgggtcccatagctggatg
gggccgataaaatggtactgccccactta
gtggcagctcgcgaccagtcacaagcccaggataacttcgtataatgtatgtatacga
gttatctgtgggcgttatgaataatagact
5 ggaaccggggcccttgattgacgactccataTTTTtagatgtagcaactcggcaagagc
attatgtgcaatacatttgttaccatacaaa
ggcagctgccagacgacttgattgcgtacaattctcacggcaagctttccagggttat
gcattatgcgcaaatgcttgatgcttaccgc
aggattaatctcggagaagcgcgtgcaagctatatgggtgtagtagatatgtagatgtac
10 caaccaatgaagaacatttatggtctagaa
cgtagtgtgaaggTTTTgagtaatttgatcaagtaagacgatattattgatataatac
caagcatataatcatgataaatacttggaaacca
cccttgcgtccggcctcacgagccttctactgccgggctcgaaggagccactggaggcc
tgtccacccttggatgcgatttctgca
15 ccttttcttgggcctgcacgtcgattagacatgattcaaatcgagatcttggaaatct
tacatgctggcgaagccaccggtgtggctgg
actgtccgcccttctgcgcaatgcttgaacctctcttggggctgtgtagaaagggtt
gttagcaacattactacaactctcaggactcg
gtggctgtaccgggtggcgaagttccgggggtatcgttgccagacattgtgtgattatt
20 tgggtgtgcaaatgtgtgctatgtgtgttgttgc
gttgggtgatgatgctgaagctgttgaaagcaggctgggtctgtgggagagacttgggata
tttatatccaaagttcggctgtgttctctg
gaagctcttctactccatacaatcatccaaagttgtcgtcattgagcgtgatcagta
gtagcctctgaggatcatcccatgatccttccg
25 gccaacagtcggcactcatcaacagcaacaatcagccgccacaaacataggtagtaag
gagttagatatcatgtagtcgtagta
ctcgacatcatgacgtacaagcttggcagtgctggtaggtgcaagtatgatgatcgtat
ccgccgttgttcgatcgaacagagtgcggt
cagattcacgggttctctcaccttgaacattggatgcaattggattgatccacaatcctg
30 gagaatggcttcaagctcactgctccagtcgc
aagcttcagagcctattactaagggtagagctacctatgtcaagagtttcaaggtaacct
aagctacatgtgatagtcggcaagccattt
gaacgcagaccgtgaacgggtgatgtaaatccgggtagacgccaagcgtgccgtgtcaa

tgacgctagatacacctcgatttacgta
 gagtgaatgccagccaatggagtcacacataacccgcttagactctgctcggggcgat
 acccgatcgagaggcagagccgctta
 aacgcgatcgcggaacctgtaacagagccagcgctcgatgaattgcatcatggaagcc
 5 attgatgtggaatgttgagcgataacaa
 cacgaattgaagacgacattgacttgcttcaagtgagtggaattgccgggcagacaag
 ataggtaggctcttggtgcgctgtcacat caatccattcctttcctctgttcaatctatgttgacattctgatagggatcattgga
 tgccaatgcaaagaacatgagagtggtgtctgc
 attcaagtatcctggctgtaagctgtggccatggggcgctgcggtcaaggtaaatcgcgat
 10 gactaatcagtcctcggtgactctggggcg
 gttagaggcagtgctgtaaccaaagcttgagccgagggcaaaaacaacggcgcatcaaac
 aatcaacgaaagcatcgtaacagt gtctcttcccagtcaattacttcgaaaaccttctcgatagaaccttcagacgatgaac
 agggcacgcaaccgtcagccgcgcccccc
 aggacagactcagcgccccgggaggcagatcgtcacaccttggtcgacgagctc.

15

EJEMPLO 7

Análisis de variantes de CBH2 de una sola sustitución en pruebas de actividad y estabilidad térmica.

[247] Se transformaron protoplastos de una cepa de *T. reesei* borrada seis veces (Aeg11,
 20 Aeg12, AegB, Acbh1, Acbh2, Abgl1) con variantes CNH⁺ SEL seleccionadas
 codificadas sobre el vector pTTTpyrG como se describe (WO 2009/048488). Las
 esporas se cosecharon, se pusieron nuevamente en platos con agar de acetamida, e
 incubaron a 28°C por 7 días. Luego las variantes CBH2 se produjeron inoculando 25 ml
 de un cultivo de YEG (5 g/L de extracto de levadura, 20 g/L de glucosa) con una cepa
 25 de *Trichoderma* esporulante sobre agar, seguido por un periodo de incubación de 2 días
 a 28°C, 200 rpm. 5 ml del cultivo de YEG se emplearon para inocular 45 ml de un
 medio de glicina que contiene una mezcla de glucosa/soforosa al 2%. El cultivo fue
 dispensado sobre una placa de microtitulación de 6 pozos a incubada estacionariamente
 por 5 días a 28°C en una atmosfera enriquecida con oxígeno. Las células fueron
 30 removidas por filtración a través de un filtro de 2 µm. Los sobrenadantes fueron
 subsiguientemente concentrados usando celdas de batido de Vivaspin con un corte de 10
 kD de peso molecular. Las variantes se probaron en su habilidad para hidrolizar celulosa
 hinchada con ácido fosfórico (PASC) a 50°C, hojas y tallos de maíz pre tratados con

ácido diluido (PCS) a 50°C, mazorcas de maíz pre tratadas con amoníaco diluido (CC) a 50°C y 57°C. Además, las variantes se probaron para estabilidad térmica.

[248] PASC a 50°C. La actividad de la enzima sobre celulosa hinchada con ácido fosfórico se examinó esencialmente como se describe en el ejemplo 1, sección C, con los siguientes cambios. A cada cubeta, se le agregaron 10 µl que contenían 4.9 µl de proteína en sobrenadante de una cepa borrada de CBH2 (Aegl1, Aegl2, Acbh1, Acbh2). La placa fue incubada a 50°C mientras se agitaba a 200 rpm. Después de 2 horas la placa se puso en hielo por 5 minutos y se agregaron 100 µl de glicina 100 mM y pH de 10.0. Después de mezclar, la placa se centrifugó a 3000 rpm por 5 min. Un volumen de 40 µl de sobrenadante se diluyeron con 160 µl de agua. 10 µl de la solución diluida fueron transferidos a una nueva placa de microtitulación de 96 cubetas (Costar Flat Bottom PS) que contenía 100 µl de una mezcla de prueba de ABTS glucosa (2.74 mg/ml de ácido 2,2'-azino-bis(3-etilbenzo)-tiazolin-6-sulfónico, 1 U/ml de peroxidasa de rabano tipo VI, 1 U/ml de oxidasa de glucosa) y se registró un incremento en el A42 en un espectrofotómetro de placas de microtitulación (Spectramax Plus 348, Molecular Devices). Se incluyó un rango de concentraciones de glucosa como estándar en cada placa (0; 0.008; 0.016; 0.031; 0.063; 0.125; 0.25; 0.5; 1 mg/ml). Las pruebas se hicieron por duplicado, se generó una curva de respuesta a la dosis para el tipo salvaje de CBH2 ajustando los datos con una ecuación de isoterma de Temkin ($y=a+b(\ln(1+c*x))$) y se dividieron las actividades de las variantes de CBH2 por una actividad calculada del tipo salvaje de CBH2 del mismo plato para obtener un índice de desempeño.

[249] PCS 50°. Se examinó la actividad de la enzima sobre hojas y tallos de maíz lavados con ácido diluido esencialmente como se describe en el ejemplo 1, sección D, con los siguientes cambios. A cada cubeta, se agregan 10 µl que contenían 49 µg de proteína en un sobrenadante de una cepa borrada de CBH2 (Agl1, Agl2, Acbh1, Acbh2). La placa se incubó a 50°C mientras se agitaba a 200 rpm. Después de 2 días la placa se puso en hielo por 5 min y se agregaron 100 µl de glicina 100mM con un pH de 10. Después del mezclado, la placa se centrifugó a 3000 rpm por 5 min. Se diluyó un volumen de 10 µl de sobrenadante en 190 µl de agua. Diez µl de la solución diluida fueron transferidos a una placa de microtitulado de 96 cubetas (Costar Flat Bottom PS) que contenían 100 µl de mezcla de ABTS glucosa de prueba y analizada como se describe en la prueba PASC.

[250] Mazorca a 50°C. Actividad de la enzima sobre la mazorca a 50°C se llevó a cabo esencialmente como se describe en el ejemplo 1, sección E, con los siguientes cambios. A cada cubeta se le agregaron 10 µl de una solución que contenía 46.55 µg de proteína de sobrenadante de una cepa CBH2 borrada (Agl1, Agl2, Acbh1, Acbh2), suplementada con 4.90 µg de CBH1 de *T. reesei*, 6.48 µg de Xyn2 Y2 de *T. reesei* (Xiong y colaboradores, *Extremophiles* 8:393-400, 2004), 2.28 µg de *Fusarium verticillioides* (Fv) 51^a, 5.32 µg de Fv3A, 0.76 µg de Fv43D, y 2.5 µg de BGL1 de *T. reesei*. Las enzimas de *Fusarium verticillioides* han sido descritas en US 61/245269 (incorporado aquí a modo de referencia para la enseñanza de este método). Se agregaron diferentes volúmenes de sobrenadante de cepa transformada de *T. reesei* que expresa CBH2 como se describe en el ejemplo 1. Las placas fueron incubadas por 2 días a 50°C con agitación a 200 rpm.

[251] Mazorca 5TC. Se molió mazorca hasta que pasara por una malla de 0.9 mm y luego se pre trató como se describe en el ejemplo 4 de WO 2006/110901 (incorporado aquí a modo de referencia para la enseñanza de este método). La mazorca pre tratada se empleó como una solución de celulosa en suspensión al 7% en acetato de sodio 50 mM con pH de 5.0. Setenta microlitros de suspensión fueron agregados a cada cubeta de una placa de microtitulación de 96 cubetas (Nunc Flat Bottom PS). A cada cubeta se le agregó una solución que contenía 46.55 µg de proteína de sobrenadante de una cepa de CBH2 borrada (Agl1, Agl2, Acbh1, Acbh2), 4.90 µg de una variante de CBH2 (S8P/T411/N89D/S92T/S113N/S196T/P227L/D249K/T255P/S278P/E95K/T296P/T332 Y/ V403D/S411F/T462I), 6.48 µg de Xyn2 Y5 de *T. reesei* (Xiong y colaboradores, *Extremophiles* 8:393-400, 2004), 2.28 µg de Fv51A, 5.32 µg de Fv3A, 0.76 µg de Fv43D, 2.45 µg de *talaromices emersonii* beta glucosidasa fueron adicionados. Hasta veinte microlitros de sobrenadante de *H. jecorina* que expresaba ya fuera el tipo salvaje de CBH2 o una variante de CBH2 fueron agregados. Se agregaron volúmenes de compensación de buffer de acetato para completar las diferencias en volumen total. La placa se incubó a 57 °C mientras se agitaba a 200 rpm. Después de 2 días la placa se puso en hielo y se centrifugó a 3000 rpm por 5 min. Un volumen de 10 µl de sobrenadante se diluyó en 190 µl de agua. Diez µl de la solución diluida fueron transferidos a una nueva placa de microtitulación de 96 cubetas (Costar Flat Bottom PS)

que contenía 100 µl de mezcla de prueba de ABTS glucosa y se probaron y analizaron como se describe arriba.

[252] Prueba 2 de estabilidad térmica de CBH2. Para probar la estabilidad térmica de las variantes de CBH2, 50 µl de sobrenadante se incubaron en una máquina PCR a lo largo de un rango de temperaturas entre 50 – 70°C. La actividad remanente se determinó incubando 20 µl de sobrenadante con 80 µl de 0.625 mg/ml de celotriosa en buffer de acetato de sodio 50 mM a un pH de 5.0 en Tween80 al 0.0025%. Después de una hora, se agregaron 40 µl de glicina pH10 100 mM. Se agregaron veinte µl a 80 µl de reactivo ABTS (ver arriba) y el desarrollo del color en OD420 se registró por 5 min. Se calculó una pseudo temperatura de fusión (T_m) ajustando la actividad restante a cada temperatura con la fórmula $y = 1/(1 + \exp(-(a*(1-b/c - \ln(c/b)) + d*((1/c) - (1/b))))/e))$ en la cual a es la capacidad calórica (kcal/mol $^{\circ}K^{-1}$), b es la temperatura de fusión (K), c es la temperatura de prueba (K), d es el cambio de entalpía (kcal/mol $^{\circ}K^{-1}$), y e es la constante de los gases (kcal/mol $^{\circ}K^{-1}$).

15

Como se determinó en la prueba 2 de estabilidad térmica. La T_m de la CBH2 de referencia de SEQ ID NO:3 fue de 57.5 °C bajo las condiciones de la prueba.

[253] Se ideó una categoría de mutaciones, o sea mutaciones altamente combinables, para abarcar las mutaciones que tienen un $PI > 0.75$ para al menos una propiedad, y < 0.05 para todas las propiedades. Como se ilustra en los ejemplos subsiguientes, algunas variantes tienen dos mutaciones altamente combinables; tienen actividad y/o estabilidad que es mayor que la enzima CBH2 de referencia. Una acumulación de mutaciones altamente combinables puede emplearse para hacer una variante beneficiosa donde cada mutación individual no es suficiente para conferir actividad, estabilidad térmica y/o niveles de expresión superiores.

EJEMPLO 8

Variantes de la librería de CBH2 y actividades de las mismas.

[254] Una librería de combinaciones de CBH2 sintética fue producida por GeneOracle (Mountain View, CA). La tabla 8-1 lista las posibles sustituciones de los miembros de la librería de combinaciones CBH2 (numerados de acuerdo con la secuencia de aminoácidos de la CBH2 madura). Esta librería fue creada combinando sustituciones

30

clasificadas como mutaciones y/o mutaciones altamente combinables en base a su Tm y su desempeño como se listan en la tabla 7-1.

La librería fue recibida del proveedor mencionado arriba como productos PCR
 5 purificados en los cuales los iniciadores GACCGGACGTGTTTTGCCCTTCAT (SEQ ID NO:20) y GTGTGACCGGCTTTGGCGAGTG (SEQ ID NO:21) fueron empelados para amplificar el gen de cbh2 flanqueado corriente arriba por alrededor de 1.1 kb de promotor cbh1 y corriente abajo por alrededor de 1.85 kb de marcador amdS para forzar la integración en el sitio pvr2 de la cepa anfitriona de *H. jecorina*. En la figura 11 se
 10 muestra un esquema de la recombinación homóloga de la platina expreson en la cepa de tamizado. La secuencia de nucleótidos del fragmento PCR (promotor cbh1 parcial, gen cbh2, y gen amdS parcial) amplificado de pTTTpyrG-CBH2 usando los iniciadores de SEQ ID NO:20 y SEQ ID NO:21, se proporciona a continuación como SEQ ID NO:22:

```

gaccggacgtgttttgccttcatttggagaaataatgcatgcatgtgtaatttgc
15 tgcttgaccgactggggctgttcgaagcccg
aatgtaggattgttatccgaactctgctcgtagaggcatgttgaatctgtgcgggca
ggacacgcctcgaaaggtcacggcaaggg
aaaccaccgatagcagtgtctagtagcaacctgtaaagccgcaatgcagcatcactggaa
aatacaaaaccaatggctaaaagtacata
20 agttaatgcctaaagaagtcataaccagcggctaataattgtacaatcaagtggctaaa
cgtaccgtaatttgccaacggcctgtgggggt
tgcaagaagcaacggcgaagccccacttccccacgtttgtttcttactcagtccaatctc
agctggatgatcccccaattgggtcgttgttt
gttccgggtgaagtgaagaagacagaggtaagaatgtctgactcggagcgtttgcatac
25 aaccaagggcagtgatggaagacagtg
aaatgttgacattcaaggagtatttagccagggatgcttgagtgtatcgtgtaaggaggt
ttgtctgccgatacgaagaatactgtatagt
cacttctgatgaagtggccatattgaaatgtaaagtcggcactgaacaggcaaaagatt
gagttgaaactgcctaagatctcgggccct
30 cgggccttcggcctttgggtgtacatgtttgtgctccgggcaaatgcaaagtgtgtagg
atcgaacacactgctgcctttaccaagca
gctgaggggtatgtgataggcaaatgttcaggggccactgcatggtttcgaatagaaagag
aagcttagccaagaacaatagccgataa

```

- agatagcctcattaaacggaatgagctagtaggcaaagtcagcgaatgtgtatatataaa
 ggcticgaggtccgtgcctccctcatgctct
 ccccatctactcatcaactcagatcctccaggagactgtacaccatctttgaggcaca
 gaaacccaatagtaaccatcacaagtttgt
- 5 acaaaaaagcaggctccgcggccgccccctcaccaccatgattgtcggcattctcacc
 acgctggctacgctggccacactcgca
 gctagtgtgcctctagaggagcggcaagcttgctcaagcgtctgtaattatgtgaaccc
 tctcaagagacccaaatactgagatatgtc
 aagggggccaatgtggtggccagaattggcgggtccgacttgctgtgcttcgggaagcac
- 10 atgctgtactccaacgactattactccc
 agtgtcttcccggcgtgcaagctcaagctcgtccacgcgcgccgctcgacgacttctc
 gagtatccccacacaatcccggtcga
 gctccgcgacgcctccacctggttctactactaccagagtacctccagtcggatcgggaa
 ccgctacgtattcaggcaaccctttgttg
- 15 gggtcacctcctgggccaatgcatattacgcctctgaagtagcagcctcgtattccta
 gcttgactggagccatggccactgctgcag
 cagctgtcgaaagggttccctcttttatgtggctgtaggctctcccggaaccaaggcaat
 ctgttactgaaggctcatcattcactgcaga
 gatactcttgacaagacccctctcatggagcaaaccttggccgacatccgcaccgccaac
- 20 aagaatggcggttaactatgccggacag
 tttgtggtgtatgacttggcggatcgcgattgcgctgcccttgccctgaatggcgaatac
 tctattgccgatgggtggcgtcgccaaatata
 agaactatatcgacaccattcgtcaaattgtcgtggaatattccgatatccggaccctcc
 tggttattggTAtgagtttaaacacctgcct
- 25 ccccccccccttcccttcttcccgcggcatcttgcgttgtgtaactattgttccc
 tcttcagagcctgactctcttgccaacctgggtg
 accaacctcggctactccaaagtgtgccaatgctcagtcagcctaccttgagtgcataaac
 tacgccgtcacacagctgaaccttccaaa
 tgttgcgatgtatttggacgctggccatgcaggatggcttggctggccggcaaccaaga
- 30 cccggccgctcagctatttgcaaatgtt
 acaagaatgcatcgtctccgagagctcttcgggattggcaaccaatgtcgccaactaca
 acgggtggaacattaccagcccccatc
 gtacacgcaaggcaacgctgtctacaacgagaagctgtacatccacgctattggacctct

tcttgccaatcacggctggccaacgcct
 tcttcatcactgatcaaggctgatcgggaaagcagcctaccggacagcaacagtggggag
 actgggtgcaatgtgatcggcaccggat
 ttgggtatcgcctatccgcaaacactggggactcgttgcctggattcgttgcctgggtca
 5 agccaggcggcgagtgtaggcaccag
 cgacagcagtgccgacgattgactccactgtgcgctccagatgccttgcaaccggc
 gcctcaagctgggtgcttgggtccaagcc
 tacttgcagcttctcacaacgcaaacccatcgttctgtaaAaggggtgggcgcgcc
 gaccagcttctgtacaaagtgggtgatc
 10 gcgccgcgcgccagctccgtgcgaaagcctgacgcaccggtagattcttggtagcccg
 atcatgacggcggcgggagctacatg gcccggggtgattatcttctgtatctacttctgaccctttcaaataacggtaac
 tcatcttctactggagatgcggcctgcttgggtattg
 cgatgtgtcagcttggcgaattgtggcttgcgaaacacaaaacgattccttagtagcc
 atgcattttaagataacgggaatagaagaaag
 15 aggaaattaaaaaaaaaaaaaaaaaacaacatcccgttcataacccgtagaatgcgcgtc
 ttcgtgtatcccagtagcagtttatttgaa
 tagctcgcgcgtggagagcatcctgaatgcaagtaacaaccgtagaggctgacacggca
 ggtgttgcctaggagcgtcgtgttcta
 caaggccagacgtcttcgcggttgatatatatgtatgttgcaggtgctcagcga
 20 cgacagtcaggctgcctcgcctgcttgc
 caataatgcagtgagggaagccacaccgtgactccatcttctagtaagctctgttgg
 gttatcagcaatacacgtaatttaactcgt
 tagcatggggctgatagcttaattaccgtttaccagtgccatggttctgcagcttctt
 ggcccgtaaaattcggcgaagccagccaatc
 25 accagctaggcaccagctaaacctataattagtctcttataacaccatccgctcccc
 gggatcaatgaggagaatgagggggatg
 cggggctaaagaagcctacataacctcatgccaaactccagttacactcgtcgagcca
 acatcctgactataagctaacacagaatg
 cctcaatcctgggaagaactggcgcgtgataagcgcggccgcctcgcaaaaaccatccct
 30 gatgaatggaaagtcagacgctgcct
 gcggaagacagcgttattgatttccaaagaatcggggatccttctcagaggccgaactg
 aagatcacagaggcctccgctgcagatc
 ttgtgtccaagctggcggccggagagttgacctgggtggaagtacgctagcattctgta

- aacgggcagcaatcgcccagcagttagt
 aggggtcccctctacctctcagggagatgtaacaacgccaccttatgggactatcaagctg
 acgctggcttctgtgcagacaaactgcgc
 ccacgagttcttcctgacgccgtctcgcgcaggcaagggaactcgaataactacgc
 5 aaagcacaagagacccgttggtccact
 ccatggcctccccatctctcctcaagaccagcttcgagtcaggtaacaccgttggcccta
 agtcgtagatgtccctttttgtcagctaaca
 tatgccaccagggtctacgaacatcaatgggctacatctcatggctaaacaagtacgacg
 aaggggactcgggtctgacaacctgct
 10 ccgcaaaagccgggtgccgtctctacgtcaagacctctgtcccgacacccgatggctg
 cgagacagtcaacaacatcatcgggcg
 caccgtcaaccacgcaacaagaactggctcgtgcggcggcagttctggtggtgaggggtgc
 gatcgttgggattcgtggtggcgtcat
 cgggttaggaacggatatcgggtggctcgattcgagtgccggccgcgttcaacttctgta
 15 cgggtctaaggccgagtcattgggcggctg
 ccgtatgcaaagatggcgaacagcatggagggtcaggagacgggtgcacagcgttgctggg
 ccgattacgcactctgttgagggtga gtccttcgcctcttcttcttcttctgctctataccaggcctccactgtcctccttctt
 gctttttatactatatacagagaccggcagtcactga
 tgaagtatgtagacctccgcctcttcaccaaattccgtcctcggtcaggagccatggaaa
 20 tacgactccaaggtcatcccatgccctg
 gcgccagtcaggagtcggacattattgcctccaagatcaagaacggcgggctcaatatcgg
 ctactacaacttcgacggcaatgtccttc
 cacaccctcctatcctgcgcggcgtggaaaccaccgtcggcgcactcgccaaagccggtc acacc.
 [255] Protoplastos de AD5 de la cepa de *H. jecorina* (Aegl1, Aegl2, Acbhl1, Acbhl2,
 25 Abgl1) descritos en el ejemplo 6 fueron transformados con la librería de ADN como se
 describe (US 2006/0094080) y cultivados en agar selectivo que contiene acetamida a
 28°C por 7 días (0.6 g/L acetamida, 1.68 g/L CsCl, 20 g/L glucosa, 6 g/L KH₂PO₄, 0.6
 g/L MgSO₄·7H₂O, 0.6 g/L CaCl₂·2H₂O, 0.5 g/L uridina, sales de elementos traza, 10
 g/L agarosa con bajo punto de fusión). Después de 24 horas el agar se cubrió con agar
 30 selectivo suplementado con 1.2 g/L de ácido fluoroorotico (FOA). Un total de 380
 colonias fueron transferidas a placas de dextrosa de papa que contenían 1.2 g/L de FOA
 e incubadas a 28°C por 4-5 días. Las esporas fueron transferidas placas de agar fresco
 de dextrosa de papa, los cuales fueron incubados a 28°C por 3 días. Alternativamente,

- los protoplastos de la cepa MAD6 descrita en el ejemplo 6 pueden emplearse en lugar de AD5 para la expresión de los miembros de la librería de variantes. Así mismo, los derivados de protoplastos de la cepa MAD6 en las cuales se han activado celulasas adicionales pueden ser usadas con este propósito. Tales derivados exhibirían aún menos actividad de celulasa en el trasfondo.
- [256] Para la producción de proteína variante de CBH2, las esporas fueron transferidas usando un replicador de 96 clavijas a un 200 µl de medio mínimo de glicina suplementado con una mezcla de glucosa/soforosa al 2% en una placa filtro de PDVF: 6.0 g/L de glicina, 4.7 g/L de $(\text{NH}_4)_2 \text{SO}_4$; 5.0 g/L de KH_2PO_4 ; 1.0 g/L de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 33.0 g/L de PIPPS; pH 5.5; con la adición esteril de una mezcla de glucosa/soforosa al 2% como fuente de carbono, 10 ml/L de 100g/L de CaCl_2 , 2.5 ml/L de elementos traza de *T. reesei* (400X): 175g/L ácido cítrico anhidro; 200g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$; 16g/L de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$; 3.2 g/L de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$; 1.4 g/L de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$; 0.8 g/L de H_3BO_3 . cada variante de CBH2 fue cultivada por cuadruplicado. Después de sellar la placa con una membrana permeable al oxígeno, las placas fueron incubadas a 28°C por 6 días, mientras se agitaban a 200 rpm. Se cosechó el sobrenadante transfiriendo el medio de cultivo a una placa de microtitulación en presencia de baja presión.
- [257] Se aislaron un total de diez variantes que mostraron actividad mejorada sobre la mazorca a 57°C. El ADN genómico de estas cepas fue aislado y se determinaron sus secuencias genéticas del *cbh2*. Las variantes de CBH2 se probaron para propiedades de interés. Las sustituciones y las actividades de los miembros de la librería de combinaciones sobre la mazorca y las hojas y tallos del maíz se muestran en la Tabla 8-2. Las actividades específicas para las hojas y tallos de maíz pre tratados con ácido diluido (PCS 50°C), para la mazorca a 50°C (CC 50°C) y para la mazorca a 57°C (CC 57°C) se determinaron como se describe en el ejemplo 7.

EJEMPLO 9

Variantes con sustitución sencilla y doble y actividad de las mismas.

- [258] GeneOracle (Mountain View, CA) variantes sintéticas de genes que codifican CBH2 en el pTTTpyrG. La Tabla 9-1 lista las variables sencillas y dobles que se contruyeron (numeradas de acuerdo con la secuencia de aminoácidos de la CBH2 madura).

[259] Los genes variantes de cbh2 se recibieron del proveedor arriba mencionado como productos PCR purificados en los cuales fueron usados los iniciadores GACCGGACGTGTTTTGCCCTTCAT (SEQ ID NO:20) y GTGTGACCGGCTTTGGCGAGTG (SEQ ID NO:21) para amplificar los genes de cbh2 flanqueados corriente arriba por 1000 bp del promotor cbhl y corriente abajo por 1000 bp del marcador de amdS para forzar la integración en el lugar de la pyr2 en la cepa de *H. jecorina* anfitriona. La secuencia de nucleótidos del fragmento PCR (promotor cbhl parcial, el gen cbh2, y el gen parcial de amdS) amplificada del pTTTpyrG-CBH2 empleando los iniciadores de la SEQ ID NO:20 y la SEQ ID NO: 21, se proporciona en el ejemplo 8 como SEQ ID NO: 2q22.

[260] Los protoplastos de la cepa MAD6 de *H. jecorina* (Aegl1, Aegl2, Aegl3, Acbh1, Acbh2, Abgl1) descritos en el ejemplo 6, fueron transformados con fragmentos lineales de ADN como se describe (US 2006/0094080) y cultivados en agar selectivo que contiene acetamida a 28°C por 7 días como se describe. Después de 24 horas el agar se cubrió con agar selectivo suplementado con 1.2 g/L de ácido fluoroortico (FOA). Las colonias fueron transferidas a placas con agar de dextrosa de papa que contienen 1.2 g/L de FOA e incubados a 28°C por 4-5 días. Las esporas fueron transferidas a placas con agar de dextrosa de papa, en donde fueron incubadas a 28°C por 3 días.

[261] Las variantes fueron cultivadas en placas de microtitulación como se describe en el ejemplo 8. Para el crecimiento en frascos de agitación, se inocularon 25 ml de cultivo YEG (5 g/l de extracto de levadura, 20 G/l de glucosa) con una cepa esporulante de trichoderma sobre agar e incubado a 28°C a 200 rpm por 2 días. Luego de la inoculación, los cultivos fueron incubados en una incubadora acústica resonante (Applikon) a 28°C por 3 días. Las células fueron removidas por filtración a través de un filtro de 0.2 µm. Los sobrenadantes se concentraron usando celdas de giro Vivaspin con un peso molecular de corte de 10 kD.

[262] Las muestras crecidas en la placa de microtitulación fueron analizadas para los niveles de producción de CBH2 por HPLC como se describe en el ejemplo 1. Tanto las muestras cultivadas en la placa de microtitulación como en el frasco de agitación fueron probadas para las propiedades de interés. La tabla 9-2 lista el desempeño de las variantes de CBH2 para múltiples propiedades de interés.

La T_m = temperatura del pseudo punto de fusión ($^{\circ}\text{C}$) se determinó mediante una prueba 2 de estabilidad térmica. IR = índice de desempeño para la estabilidad como se determina por pruebas de estabilidad térmica.

- [263] Prueba 3 de estabilidad térmica de CBH2. Para analizar la estabilidad térmica de las variantes de CBH2, 40 μl de sobrenadante fueron incubados en una máquina PCR a 58 $^{\circ}\text{C}$ por triplicado y se retiraron muestras a los 30, 60 y 120 min de incubación. La actividad restante se determinó incubando 10 μl de sobrenadante con 50 μl de 1mg/ml de celotriosa en buffer de acetato de sodio 50 mM con pH de 5.0 y 0.0025% de Tween20. Después de una hora. Se agregaron 40 μl de glicina 100 mM y pH 10. Se agregaron cincuenta μl de reactivo ABTS concentrado 3 veces (ver arriba) t se registró el desarrollo de color a OD420 por 3 minutos. La actividad residual en el tiempo fue ajustada con una fórmula de decaimiento exponencial: $y = A_0 * \exp(-k * t)$, donde A_0 es la actividad a $t=0$, t es el tiempo, y k es la constante de decaimiento. Se calcularon índices de desempeño (PI= por la siguiente fórmula: $PI = -\log(k_{\text{variant}}) / -\log(k_{\text{wt}})$.

REIVINDICACIONES

Nosotros reivindicamos:

- 5 1. Una variante de celulasa, en donde la variante es una forma madura que tiene actividad de celulasa y que comprende una sustitución en una o más posiciones seleccionadas del grupo que consiste de: 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207,
 10 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 y 447, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de una celobiohidrolasa II (CBH2) de referencia
 15 establecida como SEQ ID NO:3.

2. La variante de celulasa de la reivindicación 1, en donde la variante comprende una sustitución adicional en una o más posiciones adicionales seleccionadas a partir de:
 - (i) un primer grupo que consiste de 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194,
 20 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 y 405; o
 - (ii) un segundo grupo que consiste de 94, 98, 107, 120, 134, 144, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 291, 312, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, y 429,
- 25 en donde las posiciones adicionales son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa (CBH2) de referencia establecida como SEQ ID NO:3.

3. Una variante de celulasa, en donde dicha variante es una forma madura que tiene
 30 actividad celulasa y que comprende una sustitución en una o más posiciones seleccionadas del grupo que consiste de: 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 y 405, en donde las posiciones son enumeradas por

correspondencia con la secuencia de aminoácido de una celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una sustitución K/E, R/Q, N/D, Q/E, D/N o E/Q.

5

4. La variante de celulasa de la reivindicación 3, en donde la variante comprende una sustitución adicional en una o más posiciones adicionales seleccionada de:

(i) un primer grupo que consiste de 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 y 447; o

15 (ii) un segundo grupo que consiste de 94, 98, 107, 120, 134, 144, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 291, 312, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, y 429,

en donde las posiciones adicionales son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida

20 como SEQ ID NO:3.

5. Una variante de celulasa, en donde la variante es una forma madura que tiene actividad celulasa y que comprende una sustitución en una o más posiciones seleccionada del grupo que consiste de: 94, 98, 107, 120, 134, 144, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 291, 312, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, y 429, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de una celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una del grupo que consiste de V94E,

30 P98L, E107Q, M120L, M134G, M134L, M134V, L144G, L144R, L144S, T154A, A178V, L179C, V206L, S210L, S210R, T214M, T214Y, G231P, G231I, G231A, G231N, G231S, G231T, T232V, H266S, H266A, W272A, W272D, W272Y, N275L,

S316P, V323L, V323N, V323Y, G360R, S380T, A381T, E399N, E399D, R413Y, R413P, y A416G.

6. La variante de celulasa de la reivindicación 5, en donde la variante comprende una
 5 sustitución adicional en una o más posiciones adicionales seleccionadas de:
- (i) un primer grupo que consiste de 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292,
 10 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 y 447; o
- (ii) un segundo grupo que consiste de 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294,
 15 327, 339, 344, 356, 378, 382 and 405,
- en donde las posiciones adicionales son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3.
7. La variante de celulasa de cualquiera de las reivindicaciones 1 a 6, en donde la
 20 sustitución en una o más posiciones se selecciona del grupo que consiste de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 y 25 posiciones.
8. Una variante de celulasa, en donde la variante es una forma madura que tiene
 25 actividad celulasa y que comprende una sustitución de una a seis posiciones seleccionadas del grupo que consiste de 111, 144, 154, 162, 410 y 413, en donde la variante de celulasa comprende: leucina o serina en la posición 111; leucina, glutamina o triptófano en la posición 144; treonina, cisteína o valina en la posición 154; tirosina o asparagina en la posición 162; arginina o serina en la posición 410; y serina, triptófano o
 30 tirosina en la posición 413; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una sustitución S413Y.

9. Una variante de celulasa, en donde la variante es una forma madura que tiene actividad celulasa y que comprende una sustitución de una a seis posiciones seleccionadas del grupo que consiste de 98, 194, 313, 316, 384 y 443, en donde la variante de celulasa comprende: prolina o leucina en la posición 98; lisina, cisteína o ácido glutámico en la posición 194; serina o treonina en la posición 313; serina o prolina en la posición 316; glicina, cisteína o glutamina en la posición 384; y asparagina o isoleuceína en la posición 443; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una de grupo seleccionado de una sustitución P98L, una sustitución K194E, y una sustitución S316P.
10. Una variante de celulasa, en donde la variante es una forma madura que tiene actividad celulasa y que comprende una sustitución de una a seis posiciones seleccionadas del grupo que consiste de 153, 161, 203, 233, 422 y 444, en donde la variante de celulasa comprende: arginina o glutamina en la posición 153; asparagina, alanina o triptófano en la posición 161; arginina o histidina en la posición 203; prolina o ácido aspártico en la posición 233; glutamina o valina en la posición 422; y prolina o glutamina en la posición 443; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una sustitución R153Q.
11. Una variante de celulasa, en donde la variante es una forma madura que tiene actividad celulasa y que comprende una sustitución de una a siete posiciones seleccionadas del grupo que consiste de 98, 111, 144, 313, 316, 413 y 422, en donde la variante de celulasa comprende: prolina o leucina en la posición 98; leucina o serina en la posición 111; leucina o triptófano en la posición 144; serina o treonina en la posición 313; serina o prolina en la posición 316; serina o triptófano en la posición 413; y glutamina o valina en la posición 422; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa

consiste de una sustitución simple, la sustitución simple no es una del grupo seleccionado de una sustitución P98L, y una sustitución S316P.

12. Una variante de celulasa que comprende una glutamina en la posición 98, y una
5 sustitución seleccionada del grupo que consiste de una T138C, una S316P, una S343Q,
una Q362I, una S386C, una C400S y una S406P, en donde las posiciones son
enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa
II (CBH2) de referencia establecida como SEQ ID NO:3.
- 10 13. Una variante de celulasa que comprende una cisteína en la posición 138, y una
sustitución seleccionada del grupo que consiste de una S316P, una S343Q, una Q362I,
una S386C, una C400S y una S406P, en donde las posiciones son enumeradas por
correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de
referencia establecida como SEQ ID NO:3.
- 15 14. Una variante de celulasa que comprende una prolina en la posición 316, y una
sustitución seleccionada del grupo que consiste de una S343Q, una Q362I, una S386C,
una C400S y una S406P, en donde las posiciones son enumeradas por correspondencia
con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia
20 establecida como SEQ ID NO:3.
15. Una variante de celulasa que comprende una glutamina en la posición 343, y una
sustitución seleccionada del grupo que consiste de una Q362I, una S386C, una C400S y
una S406P, en donde las posiciones son enumeradas por correspondencia con la
25 secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida
como SEQ ID NO:3.
16. Una variante de celulasa que comprende una isoleucina en la posición 362, y una
sustitución seleccionada del grupo que consiste de una S386C, una C400S y una S406P,
30 en donde las posiciones son enumeradas por correspondencia con la secuencia de
aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID
NO:3.

17. Una variante de celulasa que comprende una cisteína en la posición 386, y una sustitución seleccionada del grupo que consiste de una C400S y una S406P, en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3.

5

18. La variante de celulasa de la reivindicación 1, que comprende una serina en la posición 400 y una prolina en la posición 406.

19. Una variante de celulasa, en donde la variante es una forma madura que tiene actividad celulasa y que comprende una sustitución de una a siete posiciones seleccionadas del grupo que consiste de 98, 111, 182, 291, 316, 362 y 400, en donde la variante de celulasa comprende: prolina, leucina o glutamina en la posición 98; leucina o serina en la posición 111; asparagina o triptófano en la posición 182; serina o un ácido glutámico en la posición 291; serina o prolina en la posición 316; glutamina, isoleucina o leucina en la posición 362; y cisteína o serina en la posición 400; en donde las posiciones son enumeradas por correspondencia con la secuencia de aminoácido de la celobiohidrolasa II (CBH2) de referencia establecida como SEQ ID NO:3, y en donde cuando la variante de celulasa consiste de una sustitución simple, la sustitución simple no es una del grupo seleccionado de una sustitución P98L, y una sustitución S316P.

20

20. La variante de celulasa de la reivindicación 19, en donde la variante además comprende una sustitución L439P o una sustitución T74S.

21. La variante de celulasa de cualquiera de las reivindicaciones 1 a 20, en donde la variante se deriva de una celulasa madre seleccionada del grupo que consiste de *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, y *Cellulomonas fimi* CenA EG.

30

22. La variante de celulasa de cualquiera de las reivindicaciones 1 a 21, en donde la variante se deriva de una celulasa madre cuya secuencia de aminoácido es por lo menos 75% 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% idéntica a un miembro del grupo

que consiste de SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, y SEQ ID NO: 13.

- 5 23. La variante de celulasa de cualquiera de las reivindicaciones 1 a 22, en donde la variante es aislada.

24. Una composición que comprende la variante de celulasa de cualquiera de las reivindicaciones 1 a 23.

10

25. La composición de la reivindicación 24, en donde la composición es enriquecida en la variante de celulasa.

- 15 26. Un ácido nucleico aislado que codifica la variante de celulasa de cualquiera de las reivindicaciones 1 a 22.

27. Un vector de expresión que comprende el ácido nucleico de la reivindicación 26 vinculado operativamente con una secuencia regulatoria.

- 20 28. Una célula huésped que comprende el vector de expresión de la reivindicación 27.

29. Un método para producir una variante de celulasa, que comprende cultivar la célula huésped de la reivindicación 28 en un medio de cultivo, bajo condiciones adecuadas para producir la variante de celulasa.

25

30. Una composición que comprende la célula huésped de la reivindicación 29 y el medio de cultivo.

- 30 31. Una composición que comprende la variante de celulasa en sobrenadante del medio de cultivo de la reivindicación 29.

32. La composición de la reivindicación 24; que además comprende por lo menos una enzima adicional seleccionada del grupo que consiste de una subtilisina, una

metaloproteasa neutral, una lipasa, una cutinasa, una amilasa, una carbohidrasa, una pectinasa, una mananasa, un arabinasa, un galactanasa, una xilanasas, una oxidasas, y una peroxidasa.

- 5 33. Un método para limpiar o cuidar telas que comprende poner en contacto una superficie o un artículo que comprende una tela con la composición de la reivindicación 24 o 32.
34. Un método para limpiar o terminar la superficie de una tela que comprende poner en
10 contacto una superficie o un artículo que comprende una tela con la composición de las reivindicaciones 24 o 32.
35. Un método para convertir biomasa en azúcares que comprende poner en contact la biomasa con la composición de las reivindicaciones 24 o 32.
- 15 36. Un método para producir un combustible que comprende: poner en contacto una composición de biomasa con una composición enzimática que comprende la variante de celulasas de cualquiera de las reivindicaciones 1 a 22 para producir una solución de azúcar; y cultivar la solución de azúcar con un microorganismo de fermentación bajo
20 condiciones suficientes para producir un combustible.

EXTRACTO

La presente invención se relaciona con enzimas y en particular con variantes de celulasa con expresión, actividad y/o estabilidad mejoradas. También describe los ácidos nucleicos que codifican las variantes de celulasa, los métodos para identificar las variantes de celulasa útiles y los métodos de uso de los mismos.

CELLULASE VARIANTS WITH IMPROVED EXPRESSION, ACTIVITY AND/OR
STABILITY, AND USE THEREOF

CROSS REFERENCE TO RELATED APPLICATION

[01] This application claims the benefit of U.S. Provisional Application No. 61/183,959, filed on June 3, 2009, which is hereby incorporated by reference in its entirety.

[02] This invention was made with Government support under conditional award no: DE- FC36-08GO 18078 awarded by the Department of Energy. The Government has certain rights in this invention.

I. FIELD

[03] The present disclosure relates to enzymes and in particular cellulase variants. Also described are nucleic acids encoding the cellulase variants, compositions comprising the cellulase variants, methods of identifying additional useful cellulase variants and methods of use thereof.

II. BACKGROUND [04] Cellulose and hemicellulose are the most abundant plant materials produced by photosynthesis. They can be degraded and used as an energy source by numerous microorganisms (e.g., bacteria, yeast and fungi) that produce extracellular enzymes capable of hydrolysis of the polymeric substrates to monomeric sugars (Aro et al., J Biol Chem, 276: 24309-24314, 2001). As the limits of non-renewable resources approach, the potential of cellulose to become a major renewable energy resource is enormous (Krishna et al., Bioresource Tech, 77: 193-196, 2001). The effective utilization of cellulose through [05] biological processes is one approach to overcoming the shortage of foods, feeds, and fuels (Ohmiya et al., Biotechnol Gen Engineer Rev, 14: 365-414, 1997). [06] Cellulases are enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. Cellulases have been traditionally

divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta- glucosidases ([beta]-D-glucoside glucohydrolase; EC 3.2.1.21) ("BG") (Knowles et al., TIBTECH 5: 255-261, 1987; and Schulein, Methods Enzymol, 160: 234-243, 1988). Endoglucanases act mainly on the amorphous parts of the cellulose fibre, whereas cellobiohydrolases are also able to degrade crystalline cellulose (Nevalainen and Penttila, Mycota, 303-319, 1995). Thus, the presence of a cellobiohydrolase in a cellulase system is required for efficient solubilization of crystalline cellulose (Suurnakki et al., Cellulose 7: 189- 209, 2000). Beta-glucosidase acts to liberate D-glucose units from cellobiose, celooligosaccharides, and other glucosides (Freer, J Biol Chem, 268: 9337-9342, 1993). [07] Cellulases are known to be produced by a large number of bacteria, yeast and fungi. Certain fungi produce a complete cellulase system capable of degrading crystalline forms of cellulose, such that the cellulases are readily produced in large quantities via fermentation. Filamentous fungi play a special role since many yeast, such as *Saccharomyces cerevisiae*, lack the ability to hydrolyze cellulose (See, e.g., Wood et al., Methods in Enzymology, 160: 87-116, 1988). [08] The fungal cellulase classifications of CBH, EG and BG can be further expanded to include multiple components within each classification. For example, multiple CBHs, EGs and BGs have been isolated from a variety of fungal sources including *Trichoderma reesei* (also referred to as *Hypocrea jecorina*), which contains known genes for two CBHs, i.e., CBH I ("CBHI") and CBH II ("CBH2"), at least 8 EGs, i.e., EG I, EG II, EG III, EGIV, EGV, EGVI, EGVII and EGVIII, and at least 5 BGs, i.e., BG1, BG2, BG3, BG4 and BG5. EGIV, EGVI and EGVIII also have xyloglucanase activity.

[09] In order to efficiently convert crystalline cellulose to glucose the complete cellulase system comprising components from each of the CBH, EG and BG classifications is

required, with isolated components less effective in hydrolyzing crystalline cellulose (Filho et al., Can J Microbiol, 42: 1-5, 1996). A synergistic relationship has been observed between cellulase components from different classifications. In particular, the EG-type cellulases and CBH- type cellulases synergistically interact to more efficiently degrade cellulose. [10] Cellulases are known in the art to be useful in the treatment of textiles for the purposes of enhancing the cleaning ability of detergent compositions, for use as a softening

agent, for improving the feel and appearance of cotton fabrics, and the like (Kumar et al., Textile Chemist and Colorist, 29:37-42, 1997). Cellulase-containing detergent compositions with improved cleaning performance (U.S. Pat. No. 4,435,307; GB App. Nos. 2,095,275 and 2,094,826) and for use in the treatment of fabric to improve the feel and appearance of the textile (U.S. Pat. Nos. 5,648,263, 5,691,178, and 5,776,757; and GB App. No. 1,358,599), have been described. Hence, cellulases produced in fungi and bacteria have received significant attention. In particular, fermentation of *Trichoderma* spp. (e.g., *Trichoderma longibrachiatum* or *Trichoderma reesei*) has been shown to produce a complete cellulase system capable of degrading crystalline forms of cellulose. [11] Although cellulase compositions have been previously described, there remains a need for new and improved cellulase compositions. Improved cellulase compositions find used in household detergents, textile treatments, biomass conversion and paper manufacturing. Cellulases that exhibit improved expression, activity and stability are of particular interest.

III. SUMMARY

[12] The present disclosure relates to enzymes and in particular cellulase variants. Also described are nucleic acids encoding the cellulase variants, compositions comprising the

cellulase variants, methods of identifying additional useful cellulase variants and methods of use thereof.

[13] The present disclosure provides cellulase variants, wherein the variants are mature forms having cellulase activity and comprising a substitution at one or more positions selected from the group consisting of: 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 and 447, wherein the positions are numbered by correspondence with the amino acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3. In some embodiments, the substitution at one or more positions results in a cellulase variant with improved expression, activity and/or stability in comparison to the reference CB H2. In some embodiments, the variant comprises a further substitution at one or more further positions selected from: (i) a first group consisting of 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 and 405; or (ii) a second group consisting of 94, 98, 107, 120, 134, 144, 147, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, and 429, wherein the further positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3. In some embodiments, the substitution at one or more positions is selected from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 positions. [14] In another aspect the disclosure provides cellulase variants, wherein the variants are mature forms having

cellulase activity and comprising a substitution at one or more positions selected from the group consisting of: 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 and 405, wherein the positions are numbered by correspondence with the amino acid sequence of a reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO: 3 and wherein when the cellulase variant consists of a single substitution, the single substitution is not a K/E, R/Q, N/D, Q/E, D/N or E/Q substitution. In some embodiments, the substitution at one or more positions results in a cellulase variant with improved expression, activity and/or stability in comparison to the reference CBH2. In some embodiments, the isolated cellulase variant comprises a further substitution at one or more further positions selected from: (i) a first group consisting of 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 and 447; or (ii) a second group consisting of 94, 98, 107, 120, 134, 144, 147, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, and 429, wherein the further positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3. In some embodiments, the substitution at one or more positions is selected from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 positions.

[15] In another aspect the disclosure provides cellulase variants, wherein the variants are mature forms having cellulase activity and comprising a substitution at one or more positions selected from the group consisting of: 94, 98, 107, 120, 134, 144, 147, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, and 429, wherein the positions are numbered by correspondence with the amino acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not one of the group consisting of V94E, P98L, E107Q, M120L, M134G, M134L, M134V, L144G, L144R, L144S, T154A, A178V, L179C, V206L, S210L, S210R, T214M, T214Y, G231P, G231I, G231A, G231N, G231S, G231T, T232V, H266S, H266A, W272A, W272D, W272Y, N275L, S316P, V323L, V323N, V323Y, G360R, S380T, A381T, E399N, E399D, R413Y, R413P, and A416G. In some embodiments, the substitution at one or more positions results in a cellulase variant with improved expression, activity and/or stability in comparison to the reference CB H2. In some embodiments, the variant comprises a further substitution at one or more further positions selected from: (i) a first group consisting of 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 and 447; or (ii) a second group consisting of 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 and 405, wherein the further positions are numbered by

correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3. In some embodiments, the substitution at one or more positions is selected from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 positions. [16] The present disclosure further provides cellulase variants, wherein the variants are mature forms having cellulase activity and comprising a substitution at from one to six positions selected from the group consisting of 111, 144, 154, 162, 410 and 413, wherein the cellulase variant comprises: a leucine or a serine at position 111; a leucine, a glutamine or a tryptophan at position 144; a threonine, a cysteine or a valine at position 154; a tyrosine or an asparagine at position 162; an arginine or a serine at position 410; and a serine, a tryptophan or a tyrosine at position 413; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not a S413Y substitution.

[17] Also provided by the present disclosure are cellulase variants, wherein the variants are mature forms having cellulase activity and comprising a substitution at from one to six positions selected from the group consisting of 98, 194, 313, 316, 384 and 443, wherein the cellulase variant comprises: a proline or a leucine at position 98; a lysine, a cysteine or a glutamic acid at position 194; a serine or a threonine at position 313; a serine or a proline at position 316; a glycine, a cysteine or a glutamine at position 384; and an asparagine or an isoleucine at position 443; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not one of the group selected from a P98L substitution, a K194E substitution, and a S316P substitution. [18] In addition the present

disclosure provides cellulase variants, wherein the variants are mature forms having cellulase activity and comprising a substitution at from one to six positions selected from the group consisting of 153, 161, 203, 233, 422 and 444, wherein the cellulase variant comprises: an arginine or a glutamine at position 153; an asparagine, an alanine or a tryptophan at position 161; an arginine or a histidine at position 203; a proline or an aspartic acid at position 233; a glutamine or a valine at position 422; and a proline or a glutamine at position 443 ; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not a R153Q substitution. [19] Moreover the present disclosure provides cellulase variants, wherein the variants are mature forms having cellulase activity and comprising a substitution at from one to seven positions selected from the group consisting of 98, 111, 144, 313, 316, 413 and 422, wherein the cellulase variant comprises: a proline or a leucine at position 98; a leucine or a serine at position 111; a leucine or a tryptophan at position 144; a serine or a threonine at position 313; a serine or a proline at position 316; a serine or a tryptophan at position 413; and a glutamine or a valine at position 422; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not one of the group selected from a P98L substitution, and a S316P substitution. [20] In further aspects of the disclosure cellulase variants are provided, wherein the variants comprise a glutamine at position 98, and a substitution selected from the group consisting of a T138C, a S316P, a S343Q, a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

[21] Additionally, cellulase variants are provided which comprise a cysteine at position 138, and a substitution selected from the group consisting of a S316P, a S343Q, a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3.

[22] In another aspect, cellulase variants are provided which comprise a proline at position 316, and a substitution selected from the group consisting of a S343Q, a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3. [23] Moreover, the present disclosure provides cellulase variants comprising a glutamine at position 343, and a substitution selected from the group consisting of a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3. [24] The present disclosure also provides cellulase variants comprising an isoleucine at position 362, and a substitution selected from the group consisting of a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3. [25] In one aspect, cellulase variants are provided comprising a cysteine at position 386, and a substitution selected from the group consisting of a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3. In some embodiments, the variants comprise a serine at position 400 and a proline at position 406.

[26] In addition, the present disclosure provides cellulase variants, wherein the variant is a mature form having cellulase activity and comprising a substitution at from one to seven positions selected from the group consisting of 98, 111, 182, 291, 316, 362 and 400, wherein the cellulase variant comprises: proline, leucine or glutamine at position 98; leucine or serine at position 111; asparagine or tryptophan at position 182; serine or a glutamic acid at position 291; serine or proline at position 316; glutamine, isoleucine or leucine at position 362; and cysteine or serine at position 400; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not a P98L substitution or a S316P substitution. In some embodiments, the variant further comprises a L439P substitution or a T74S substitution.

[27] In some preferred embodiments of the disclosure, the cellulase variants of any of the preceding paragraphs of the summary are derived from a parent cellulase selected from the group consisting of *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, and *Cellulomonas fimi* CenA EG. In some embodiments, the cellulase variant is derived from a parent cellulase whose amino acid sequence is at least 75% 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a member of the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13. In some embodiments, the variants are isolated. Also provided are compositions comprising a cellulase variant. In some preferred embodiments, the composition is

enriched in the cellulase variant. [28] In other aspects, the present disclosure provides an isolated nucleic acid encoding the cellulase variant of any of the preceding paragraphs of the summary. An expression vector comprising the isolated nucleic acid operably linked to a regulatory sequence is also provided, as is a host cell comprising the expression vector. In some embodiments, methods are provided for producing a cellulase variant, comprising culturing the host cells in a culture medium under suitable conditions to produce the cellulase variant. [29] Also provided are compositions comprising the cellulase variant of any of the preceding paragraphs of the summary. In some preferred embodiments, the compositions further comprising at least one additional enzyme selected from the group consisting of a subtilisin, a neutral metalloprotease, a lipase, a cutinase, an amylase, a carbohydrase, a pectinase, a mannanase, an arabinase, a galactanase, a xylanase, an oxidase, and a peroxidase. The disclosure further provides methods of cleaning or fabric care comprising contacting a surface or an article comprising a fabric with the composition. Additionally, methods are provided for depilling and surface finishing a fabric comprising contacting a surface or an article comprising a fabric with the composition. [30] In some preferred embodiments, methods are provided for converting biomass to sugars comprising contacting the biomass with the cellulase variant of any of the preceding paragraphs of the summary. In some preferred embodiments methods are provided for producing a fuel comprising: contacting a biomass composition with an enzymatic composition comprising the cellulase variant of any of the preceding paragraphs to yield a sugar solution; and culturing the sugar solution with a fermentative microorganism under conditions sufficient to produce a fuel.

[31] Other objects, features and advantages of the present disclosure will become apparent from the following detailed description. It should be understood, however, that

the detailed description and specific examples, while indicating preferred embodiments of the disclosure, are given by way of illustration only, since various changes and modifications within the scope and spirit of the disclosure will become apparent to one skilled in the art from reading this detailed description.

IV. BRIEF DESCRIPTION OF THE DRAWINGS

[32] FIG. 1 provides an alignment of the amino acid sequences of the mature form of various cellulases: *Hypocrea jecorina* (also known as *T. reesei*) CBH2 (SEQ ID NO:3), *Hypocrea koningii* CBH2 (SEQ ID NO:4), *Humicola insolens* CBH2 (SEQ ID NO:5), *Acremonium cellulolyticus* CBH2 (SEQ ID NO:6), *Agaricus bisporus* CBH2 (SEQ ID NO:7), *Fusarium osysporum* EG (SEQ ID NO:8), *Phanerochaete chrysosporium* CBH2 (SEQ ID NO:9), *Talaromyces emersonii* CBH2 (SEQ ID NO: 10), *Thermobifida. fusca* 6B/E3 CBH2 (SEQ ID NO: 11), *Thermobifida fusca* 6A/E2 EG (SEQ ID NO: 12), and *Cellulomonas fimi* CenA EG (SEQ ID NO: 13). [33] FIG. 2A provides a graph of the relative frequency of observed over expected pretreated corn stover (PCS) assay winners as a product of charge change of the CB H2 SELs. Decreasing CBH2 charge results in a significantly higher frequency of PCS assay winners. FIG. 2B provides a graph of the relative frequency of observed over expected phosphoric acid swollen cellulose (PASC) assay winners as a product of charge change of the CB H2 SELs. A significantly higher frequency of PASC winners is observed for variants without a charge change.

[34] FIG. 3 provides a plasmid map for pTTTpyrG-cbh2.

[35] FIG. 4 provides a schematic of the *T. reesei* ku80 deletion cassette. [36] FIG. 5 illustrates the inactivation of the native ku80 gene as a consequence of its replacement with an als selectable marker.

[37] FIG. 6 provides a schematic of the *T. reesei* pyr2 deletion cassette.

[38] FIG. 7 illustrates the inactivation of the native pyr2 gene as a consequence of its replacement with a hygR selectable marker and an amdS fragment.

[39] FIG. 8 provides a schematic of the T. reesei hygR deletion cassette

[40] FIG. 9 illustrates the inactivation of the heterologous hygR gene as a consequence of its replacement with a pyr2 selectable marker.

[41] FIG. 10 provides a schematic of the T. reesei bglI deletion cassette [42] FIG. 11 provides a schematic of the T. reesei egl3 deletion cassette

[43] FIG. 12 provides a schematic of the T. reesei telomeric plasmid vector used for expression of ere recombinase.

[44] FIG. 13 illustrates the inactivation of the pyr2 selectable marker and activation of the amdS selectable marker as a consequence of introduction of a polynucleotide of a gene of interest (GOI). In exemplary embodiments, the GOI encodes a CB H2 variant.

[45] FIG. 14 provides a schematic illustrating the derivation of the MAD6 host strain, from the quad-deleted strain.

V. DETAILED DESCRIPTION OF VARIOUS EMBODIMENTS [46] The present disclosure relates to enzymes and in particular cellulase variants. Also described are nucleic acids encoding the cellulase variants, compositions comprising the cellulase variants, methods of identifying additional useful cellulase variants and methods of using the compositions. [47] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the compositions and methods described herein. Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. In this application, the use of the singular includes the plural unless specifically stated otherwise. The use of "or" means "and/or" unless state otherwise.

Likewise, the terms "comprise," "comprising," "comprises," "include," "including" and "includes" are not intended to be limiting. All patents and publications, including all amino acid and nucleotide sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference. The headings provided herein are not limitations of the various aspects or embodiments of the disclosure, which can be had by reference to the specification as a whole. Accordingly, the terms herein are more fully defined by reference to the specification as a whole.

[48] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Singleton, et al., *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY*, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, *THE HARPER COLLINS DICTIONARY OF BIOLOGY*, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this disclosure. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, the preferred methods and materials are described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxyl orientation, respectively. Practitioners are particularly directed to Sambrook et al., *MOLECULAR CLONING: A*

LABORATORY MANUAL (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989, and Ausubel FM et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N. Y., 1993, for definitions and terms of the art. It is to be understood that this disclosure is not limited to the particular methodology, protocols, and reagents described, as these may vary.

I. Definitions

[49] The terms below are more fully defined by reference to the specification as a whole. [50] The term "polypeptide" as used herein refers to a compound made up of a single chain of amino acid residues linked by peptide bonds. The term "protein" as used herein may be synonymous with the term "polypeptide".

[51] "Variant" means a protein which is derived from a precursor protein (e.g., the native protein) by addition of one or more amino acids to either or both the C- and N-terminal end, substitution of one or more amino acids at one or a number of different sites in the amino acid sequence, or deletion of one or more amino acids at either or both ends of the protein or at one or more sites in the amino acid sequence. The preparation of a cellulase variant may be performed by any means known in the art. In preferred embodiments, a cellulase variant is prepared by modifying a DNA sequence which encodes for the native protein, transformation of the modified DNA sequence into a suitable host, and expression of the modified DNA sequence to form the variant enzyme. The variant cellulase of the disclosure includes peptides comprising altered amino acid sequences in comparison with a precursor enzyme amino acid sequence wherein the variant cellulase retains the characteristic cellulolytic nature of the precursor enzyme but which may have altered properties in some specific aspect. For example, a variant cellulase may have an increased pH optimum or increased temperature or oxidative stability or decreased affinity or binding to non-cellulosic materials but will retain its characteristic cellulolytic activity. It is contemplated that the variants according to the present disclosure may be derived from a DNA fragment encoding a cellulase variant wherein the functional activity of the expressed cellulase variant is retained. For example, a DNA fragment encoding a cellulase may further include a DNA sequence or portion thereof encoding a hinge or linker attached to the

cellulase DNA sequence at either the 5' or 3' end wherein the functional activity of the encoded cellulase domain is retained. The terms variant and derivative may be used interchangeably herein. [52] "Equivalent residues" may also be defined by determining homology at the level of tertiary structure for a precursor cellulase whose tertiary structure has been determined by x-ray crystallography. Equivalent residues are defined as those for which the atomic coordinates of two or more of the main chain atoms of a particular amino acid residue of a cellulase and *Hypocrea jecorina* CBH2 (N on N, CA on CA, C on C and O on O) are within 0.13 nm and preferably 0.1 nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the cellulase in question to the *H. jecorina* CBH2. The best model is the crystallographic model giving the lowest R factor for experimental diffraction data at the highest resolution available see for examples US 2006/0205042. [53] Equivalent residues which are functionally analogous to a specific residue of *H. jecorina* CBH2 are defined as those amino acids of a cellulase which may adopt a conformation such that they either alter, modify or contribute to protein structure, substrate binding or catalysis in a manner defined and attributed to a specific residue of the *H. jecorina* CBH2. Further, they are those residues of the cellulase (for which a tertiary structure has been obtained by x-ray crystallography) which occupy an analogous position to the extent that, although the main chain atoms of the given residue may not satisfy the criteria of equivalence on the basis of occupying a homologous position, the atomic coordinates of at least two of the side chain atoms of the residue lie within 0.13 nm of the corresponding side chain atoms of *H. jecorina* CBH2. The crystal structure of *H. jecorina* CBH2 is shown in Zou et al. (1999) (Ref. 5, supra).

[54] The term "nucleic acid molecule" includes RNA, DNA and cDNA molecules. It will be understood that, as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given protein such as CBH2 and/or variants thereof may be produced. The present disclosure contemplates every possible variant nucleotide sequence, encoding variant cellulase such as CB H2, all of which are possible given the degeneracy of the genetic code. [55] A "heterologous" nucleic acid construct or sequence has a portion of the sequence which is not native to the cell in which it is expressed. Heterologous, with respect to a control sequence refers to a control sequence (i.e. promoter or enhancer) that does not function in nature to regulate the same gene the expression of which it is currently regulating. Generally, heterologous nucleic acid sequences are not endogenous to the cell or part of the genome in which they are present, and have been added to the cell, by infection, transfection, transformation, microinjection, electroporation, or the like. A "heterologous" nucleic acid construct may contain a control sequence/DNA coding sequence combination that is the same as, or different from a control sequence/DNA coding sequence combination found in the native cell. [56] As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art. [57] Accordingly, an "expression cassette" or "expression vector" is a nucleic acid construct generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment.

Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid sequence to be transcribed and a promoter.

[58] As used herein, the term "plasmid" refers to a circular double- stranded (ds) DNA construct used as a cloning vector, and which forms an extrachromosomal self-replicating genetic element in many bacteria and some eukaryotes. [59] As used herein, the term "selectable marker-encoding nucleotide sequence" refers to a nucleotide sequence which is capable of expression in cells and where expression of the selectable marker confers to cells containing the expressed gene the ability to grow in the presence of a corresponding selective agent, or under corresponding selective growth conditions.

[60] As used herein, the term "promoter" refers to a nucleic acid sequence that functions to direct transcription of a downstream gene. The promoter will generally be appropriate to the host cell in which the target gene is being expressed. The promoter together with other transcriptional and translational regulatory nucleic acid sequences (also termed "control sequences") are necessary to express a given gene. In general, the transcriptional and translational regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. [61] "Chimeric gene" or "heterologous nucleic acid construct", as defined herein refers to a non- native gene (i.e., one that has been introduced into a host) that may be composed of parts of different genes, including regulatory elements. A chimeric gene construct for transformation of a host cell is typically composed of a transcriptional regulatory region (promoter) operably linked to a heterologous protein coding sequence, or, in a selectable marker chimeric gene, to a selectable marker gene encoding a protein conferring, for example, antibiotic resistance to transformed cells. A typical chimeric gene of the present disclosure, for transformation into a host cell, includes a transcriptional

regulatory region that is constitutive or inducible, a protein coding sequence, and a terminator sequence. A chimeric gene construct may also include a second DNA sequence encoding a signal peptide if secretion of the target protein is desired. [62] A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA encoding a secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally,

"operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading frame. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors, linkers or primers for PCR are used in accordance with conventional practice.

[63] As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain; that may or may not include regions preceding and following the coding region, e.g. 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons). [64] In general, nucleic acid molecules which encode the variant cellulase such as CBH2 will hybridize, under moderate to high stringency conditions to the wild type sequence such as provided herein as SEQ ID NO:1. However, in some cases a CBH2-encoding nucleotide sequence is employed that possesses a substantially different codon usage, while the protein encoded by the CBH2-encoding nucleotide sequence has the same or substantially the same amino acid

sequence as the native protein. For example, the coding sequence may be modified to facilitate faster expression of CBH2 in a particular prokaryotic or eukaryotic expression system, in accordance with the frequency with which a particular codon is utilized by the host (Te'o et al., FEMS Microbiology Letters, 190: 13-19, 2000, for example, describes the optimization of genes for expression in filamentous fungi). [65] A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about $T_m - 5^\circ\text{C}$. (5°C below the T_m of the probe); "high stringency" at about $5-10^\circ\text{C}$ below the T_m ; "moderate" or "intermediate stringency" at about $10-20^\circ\text{C}$. below the T_m of the probe; and "low stringency" at about $20-25^\circ\text{C}$. below the T_m . Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify sequences having about 80% or more sequence identity with the probe.

[66] Moderate and high stringency hybridization conditions are well known in the art (see, for example, Sambrook, et al, 1989, Chapters 9 and 11, and in Ausubel, F. M., et al., 1993, expressly incorporated by reference herein). An example of high stringency conditions includes hybridization at about 42°C . in 50% formamide, 5xSSC, 5.times.Denhardt's solution, 0.5% SDS and 100 ug/ml denatured carrier DNA followed by washing two times in 2. times. SSC and 0.5% SDS at room temperature and two additional times in 0.1x SSC and 0.5% SDS at 42°C .

[67] The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been

modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all. [68] As used herein, the terms "transformed", "stably transformed" or "transgenic" with reference to a cell means the cell has a non-native (heterologous) nucleic acid sequence integrated into its genome or as an episomal plasmid that is maintained through multiple generations. [69] As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

[70] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA). [71] It follows that the term "CBH2 expression" refers to transcription and translation of the *cbh2* gene or variants thereof, the products of which include precursor RNA, mRNA, polypeptide, post-translationally processed polypeptides, and derivatives thereof, including CB H2 from related species such as *Trichoderma koningii*, *Hypocrea jecorina* (also known as *Trichoderma longibrachiatum*, *Trichoderma reesei* or *Trichoderma viride*) and *Hypocrea schweinitzii*. By way of example, assays for CBH2 expression include Western blot for CB H2 protein, Northern blot analysis and reverse transcriptase polymerase chain reaction (RT-PCR) assays for *cbh2* mRNA, and Phosphoric Acid

Swollen Cellulose (PASC) and p- hydroxybenzoic acid hydrazide (PAHBAH) assays as described in the following: (a) PASC: (Karlsson, J. et al. (2001), *Eur. J. Biochem*, 268, 6498-6507, Wood, T. (1988) in *Methods in Enzymology*, Vol.160. Biomass Part a Cellulose and Hemicellulose (Wood, W. & Kellog, S. Eds.), pp.19-25, Academic Press, San Diego, Calif., USA) and (b) PAHBAH: (Lever, M. (1972) *Analytical Biochemistry*, 47, 273, Blakeney, A. B. & Mutton, L. L. (1980) *Journal of Science of Food and Agriculture*, 31, 889, Henry, R. J. (1984) *Journal of the Institute of Brewing*, 90, 37).

[72] The term "alternative splicing" refers to the process whereby multiple polypeptide isoforms are generated from a single gene, and involves the splicing together of nonconsecutive exons during the processing of some, but not all, transcripts of the gene. Thus a particular exon may be connected to any one of several alternative exons to form messenger RNAs. The alternatively- spliced mRNAs produce polypeptides ("splice variants") in which some parts are common while other parts are different.

[73] The term "signal sequence" refers to a sequence of amino acids at the N-terminal portion of a protein that facilitates the secretion of the mature form of the protein outside the cell. The mature form of the extracellular protein lacks the signal sequence that is cleaved off during the secretion process. [74] By the term "host cell" is meant a cell that contains a vector and supports the replication, and/or transcription or transcription and translation (expression) of the expression construct. Host cells for use in the present disclosure can be prokaryotic cells, such as *E. coli*, or eukaryotic cells such as yeast, plant, insect, amphibian, or mammalian cells. In general, host cells are filamentous fungi. [75] The term "filamentous fungi" means any and all filamentous fungi recognized by those of skill in the art. A preferred fungus is selected from the group consisting of *Aspergillus*, *Trichoderma*, *Fusarium*, *Chrysosporium*, *Penicillium*, *Humicola*, *Neurospora*, or alternative sexual forms thereof such as *Emericella*,

Hypocrea. It has now been demonstrated that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina* (See, Kuhls et al., PNAS, 93:7755-7760, 1996).

[76] The term "cellooligosaccharide" refers to oligosaccharide groups containing from 2-8 glucose units and having β -1,4 linkages, e.g., cellobiose.

[77] The terms "cellulase" "cellulolytic enzymes" or "cellulase enzymes" refer to a category of enzymes capable of hydrolyzing cellulose polymers to shorter cello-oligosaccharide oligomers, cellobiose and/or glucose. Numerous examples of cellulases, such as exoglucanases, exocellobiohydrolases, endoglucanases, and glucosidases have been obtained from cellulolytic organisms, particularly including fungi, plants and bacteria. The enzymes made by these microbes are mixtures of proteins with three types of actions useful in the conversion of cellulose to glucose: endoglucanases (EG), cellobiohydrolases (CBH), and β -glucosidase. These three different types of cellulase enzymes act synergistically to convert cellulose and its derivatives to glucose.

[78] Many microbes make enzymes that hydrolyze cellulose, including the wood rotting fungus *Trichoderma*, the compost bacteria *Thermomonospora*, *Bacillus*, and *Cellulomonas*; *Streptomyces*; and the fungi *Humicola*, *Aspergillus* and *Fusarium*. [79] CBH2 from *Hypocrea jecorina* is a member of the Glycosyl Hydrolase Family 6 (hence Cel6) and, specifically, was the first member of that family identified in *Hypocrea jecorina* (hence Cel6A). The Glycosyl Hydrolase Family 6 contains both Endoglucanases and Cellobiohydrolases/exoglucanases, and that CBH2 is the latter. Thus, the phrases CBH2, CBH2-type protein and Cel6 cellobiohydrolases may be used interchangeably herein. [80] The term "cellulose binding domain" as used herein refers to portion of the amino acid sequence of a cellulase or a region of the enzyme that is involved in the cellulose binding activity of a cellulase or derivative thereof. Cellulose

binding domains generally function by non-covalently binding the cellulase to cellulose, a cellulose derivative or other polysaccharide equivalent thereof. Cellulose binding domains permit or facilitate hydrolysis of cellulose fibers by the structurally distinct catalytic core region, and typically function independent of the catalytic core. Thus, a cellulose binding domain will not possess the significant hydrolytic activity attributable to a catalytic core. In other words, a cellulose binding domain is a structural element of the cellulase enzyme protein tertiary structure that is distinct from the structural element which possesses catalytic activity. Cellulose binding domain and cellulose binding module may be used interchangeably herein. [81] As used herein, the term "surfactant" refers to any compound generally recognized in the art as having surface active qualities. Thus, for example, surfactants comprise anionic, cationic and nonionic surfactants such as those commonly found in detergents. Anionic surfactants include linear or branched alkylbenzenesulfonates; alkyl or alkenyl ether sulfates having linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates; olefinsulfonates; and alkanesulfonates. Ampholytic surfactants include quaternary ammonium salt sulfonates, and betaine-type ampholytic surfactants. Such ampholytic surfactants have both the positive and negative charged groups in the same molecule. Nonionic surfactants may comprise polyoxyalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, fatty acid glycerine monoesters, and the like. [82] As used herein, the term "cellulose containing fabric" refers to any sewn or unsewn fabrics, yarns or fibers made of cotton or non-cotton containing cellulose or cotton or non-cotton containing cellulose blends including natural cellulose and manmade cellulose (such as jute, flax, ramie, rayon, and lyocell). [83] As used herein, the term "cotton-containing fabric" refers to sewn or unsewn fabrics, yarns or fibers

made of pure cotton or cotton blends including cotton woven fabrics, cotton knits, cotton denims, cotton yarns, raw cotton and the like.

[84] As used herein, the term "stonewashing composition" refers to a formulation for use in stonewashing cellulose containing fabrics. Stonewashing compositions are used to modify cellulose containing fabrics prior to sale, i.e., during the manufacturing process. In contrast, detergent compositions are intended for the cleaning of soiled garments and are not used during the manufacturing process.

[85] As used herein, the term "detergent composition" refers to a mixture which is intended for use in a wash medium for the laundering of soiled cellulose containing fabrics. In the context of the present disclosure, such compositions may include, in addition to cellulases and surfactants, additional hydrolytic enzymes, builders, bleaching agents, bleach activators, bluing agents and fluorescent dyes, caking inhibitors, masking agents, cellulase activators, antioxidants, and solubilizers. [86] As used herein, the term "decrease or elimination in expression of the cbh2 gene" means that either that the cbh2 gene has been deleted from the genome and therefore cannot be expressed by the recombinant host microorganism; or that the cbh2 gene or transcript has been modified such that a functional CBH2 enzyme is not produced by the host microorganism or at levels that are significantly less than the unmodified cbh2 gene or transcript. [87] The term "variant cbh2 gene" means that the nucleic acid sequence of the cbh2 gene from *H. jecorina* has been altered by removing, adding, and/or manipulating the coding sequence.

[88] As used herein, the term "purifying" generally refers to subjecting transgenic nucleic acid or protein containing cells to biochemical purification and/or column chromatography. [89] As used herein, the terms "active" and "biologically active" refer to a biological activity associated with a particular protein and are used interchangeably

herein. For example, the enzymatic activity associated with a protease is proteolysis and, thus, an active protease has proteolytic activity. It follows that the biological activity of a given protein refers to any biological activity typically attributed to that protein by those of skill in the art. [90] As used herein, the term "enriched" means that the cellulase such as CB H2 is found in a concentration that is greater relative to the CB H2 concentration found in a wild-type, or naturally occurring, fungal cellulase composition. The terms enriched, elevated and enhanced may be used interchangeably herein.

[91] A wild type fungal cellulase composition is one produced by a naturally occurring fungal source and which comprises one or more BGL, CBH and EG components wherein each of these components is found at the ratio produced by the fungal source. Thus, an enriched CBH composition would have CBH at an altered ratio wherein the ratio of CBH to other cellulase components (i.e., EGs, beta-glucosidases and other endoglucanases) is elevated. This ratio may be increased by either increasing CBH or decreasing (or eliminating) at least one other component by any means known in the art.

[92] The term "isolated" or "purified" as used herein refers to a nucleic acid or amino acid that is removed from at least one component with which it is naturally associated.

[93] Thus, to illustrate, a naturally occurring cellulase system may be purified into substantially pure components by recognized separation techniques well published in the literature, including ion exchange chromatography at a suitable pH, affinity chromatography, size exclusion and the like. For example, in ion exchange chromatography (usually anion exchange chromatography), it is possible to separate the cellulase components by eluting with a pH gradient, or a salt gradient, or both a pH and a salt gradient. The purified CBH may then be added to the enzymatic solution resulting in an enriched CBH solution. It is also possible to elevate the amount of CBH produced

by a microbe using molecular genetics methods to overexpress the gene encoding CBH, possibly in conjunction with deletion of one or more genes encoding other cellulases.

[94] Fungal cellulases may contain more than one CBH component. The different components generally have different isoelectric points which allow for their separation via ion exchange chromatography and the like. Either a single CBH component or a combination of CBH components may be employed in an enzymatic solution. [95]

When employed in enzymatic solutions, the homolog or variant CBH2 component is generally added in an amount sufficient to allow the highest rate of release of soluble sugars from the biomass. The amount of homolog or variant CB H2 component added depends, upon the type of biomass to be saccharified, which can be readily determined by the skilled artisan when employed, the weight percent of the homolog or variant CB H2 component present in the cellulase composition is from preferably between 1 and 100 with illustrative examples being about 1, preferably about 5, preferably about 10, preferably about 15, or preferably about 20 weight percent to preferably about 25, preferably about 30, preferably about 35, preferably about 40, preferably about 45 or preferably about 50 weight percent. Furthermore, preferred ranges may be about 0.5 to about 15 weight percent, about 0.5 to about 20 weight percent, from about 1 to about 10 weight percent, from about 1 to about 15 weight percent, from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45

weight percent, from about 10 to about 50 weight percent, from about 15 to about 60 weight percent, from about 15 to about 65 weight percent, from about 15 to about 70 weight percent, from about 15 to about 75 weight percent, from about 15 to about 80 weight percent, from about 15 to about 85 weight percent, from about 15 to about 95 weight percent. However, when employed, the weight percent of the homolog or variant CB H2 component relative to any EG type components present in the cellulase composition is from preferably about 1, preferably about 5, preferably about 10, preferably about 15, or preferably about 20 weight percent to preferably about 25, preferably about 30, preferably about 35, preferably about 40, preferably about 45 or preferably about 50 weight percent. Furthermore, preferred ranges may be about 0.5 to about 15 weight percent, about 0.5 to about 20 weight percent, from about 1 to about 10 weight percent, from about 1 to about 15 weight percent, from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about 50 weight percent, from about 15 to about 20 weight percent, from about 15 to about 25 weight percent, from about 15 to about 30 weight percent, from about 15 to about 35 weight percent, from about 15 to about 40 weight percent, from about 15 to about 45 weight percent, from about 15 to about 50 weight percent.

II. Cellulases

[96] Cellulases are known in the art as enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. As set forth above, cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases (EC 3.2.1.21) ("BG").

[97] Certain fungi produce complete cellulase systems which include exocellobiohydrolases or CBH-type cellulases, endoglucanases or EG-type cellulases and beta-glucosidases or BG-type cellulases. However, sometimes these systems lack CBH-type cellulases and bacterial cellulases also typically include little or no CBH-type cellulases. In addition, it has been shown that the EG components and CBH components synergistically interact to more efficiently degrade cellulose. The different components, i.e., the various endoglucanases and exocellobiohydrolases in a multi-component or complete cellulase system, generally have different properties, such as isoelectric point, molecular weight, degree of glycosylation, substrate specificity and enzymatic action patterns. [98] It is believed that endoglucanase-type cellulases hydrolyze internal beta-1,4-glucosidic bonds in regions of low crystallinity of the cellulose and exocellobiohydrolase-type cellulases hydrolyze cellobiose from the reducing or non-reducing end of cellulose. It follows that the action of endoglucanase components can greatly facilitate the action of exocellobiohydrolases by creating new chain ends which are recognized by exocellobiohydrolase components. Further, beta-glucosidase-type cellulases have been shown to catalyze the hydrolysis of alkyl and/or aryl beta-D-glucosides such as methyl beta-D-glucoside and p-nitrophenyl glucoside as well as glycosides containing only carbohydrate residues, such as cellobiose. This yields glucose as the sole product for the microorganism and reduces or eliminates cellobiose which inhibits cellobiohydrolases and endoglucanases. [99] Cellulases also find a

number of uses in detergent compositions including to enhance cleaning ability, as a softening agent and to improve the feel of cotton fabrics (Hemmpel, ITB Dyeing/Printing/Finishing 3:5-14, 1991; Tyndall, Textile Chemist and Colorist 24:23-26, 1992; and Kumar et al., Textile Chemist and Colorist, 29:37-42, 1997). While the mechanism is not part of the disclosure, softening and color restoration properties of cellulase have been attributed to the alkaline endoglucanase components in cellulase compositions, as exemplified by U.S. Pat. Nos. 5,648,263, 5,691,178, and 5,776,757, which disclose that detergent compositions containing a cellulase composition enriched in a specified alkaline endoglucanase component impart color restoration and improved softening to treated garments as compared to cellulase compositions not enriched in such a component. In addition, the use of such alkaline endoglucanase components in detergent compositions has been shown to complement the pH requirements of the detergent composition (e.g., by exhibiting maximal activity at an alkaline pH of 7.5 to 10, as described in U.S. Pat. Nos. 5,648,263, 5,691,178, and 5,776,757).

[100] Cellulase compositions have also been shown to degrade cotton-containing fabrics, resulting in reduced strength loss in the fabric (U.S. Pat. No. 4,822,516), contributing to reluctance to use cellulase compositions in commercial detergent applications. Cellulase compositions comprising endoglucanase components have been suggested to exhibit reduced strength loss for cotton-containing fabrics as compared to compositions comprising a complete cellulase system.

[101] Cellulases have also been shown to be useful in degradation of cellulase biomass to ethanol (wherein the cellulase degrades cellulose to glucose and yeast or other microbes further ferment the glucose into ethanol), in the treatment of mechanical pulp (Pere et al., In Proc. Tappi Pulping Conf., Nashville, Tenn., 27-31, pp. 693-696, 1996), for use as a feed additive (WO 91/04673) and in grain wet milling.

[102] Most CBHs and EGs have a multidomain structure consisting of a core domain separated from a cellulose binding domain (CBD) by a linker peptide (Suurnakki et al., 2000). The core domain contains the active site whereas the CBD interacts with cellulose by binding the enzyme to it (van Tilbeurgh et al., FEBS Lett. 204:223-227, 1986; Tomme et al., Eur. J. Biochem. 170:575-581, 1988). The CBDs are particularly important in the hydrolysis of crystalline cellulose. It has been shown that the ability of cellobiohydrolases to degrade crystalline cellulose clearly decreases when the CBD is absent (Linder and Teeri, J. Biotechnol. 57:15-28, 1997). However, the exact role and action mechanism of CBDs is still a matter of speculation. It has been suggested that the CBD enhances the enzymatic activity merely by increasing the effective enzyme concentration at the surface of cellulose (Stahlberg et al., Bio/Technol. 9:286-290, 1991), and/or by loosening single cellulose chains from the cellulose surface (Torino et al., EMBO J. vol. 15, no. 21, pp. 5739-5751, 1996). Most studies concerning the effects of cellulase domains on different substrates have been carried out with core proteins of cellobiohydrolases, as their core proteins can easily be produced by limited proteolysis with papain (Tomme et al., 1988). Numerous cellulases have been described in the scientific literature, examples of which include: from *Trichoderma reesei*: Shoemaker, S. et al., Bio/Technology, 1:691-696, 1983, which discloses CBH1; Teeri, T. et al., Gene, 51:43- 52, 1987, which discloses CBH2. Cellulases from species other than *Trichoderma* have also been described e.g., Ooi et al., Nucleic Acids Research, vol. 18, no. 19, 1990, which discloses the cDNA sequence coding for endoglucanase F1-CMC produced by *Aspergillus aculeatus*; Kawaguchi T et al., Gene 173(2):287-8, 1996, which discloses the cloning and sequencing of the cDNA encoding beta-glucosidase 1 from *Aspergillus aculeatus*; Sakamoto et al., Curr. Genet. 27:435-439, 1995, which discloses the cDNA sequence encoding the endoglucanase CMCase-1 from *Aspergillus kawachii*

IFO 4308; Saarilahti et al., Gene 90:9-14, 1990, which discloses an endoglucanase from *Erwinia carotovora*; Spilliaert R, et al., Eur J Biochem. 224(3):923-30, 1994, which discloses the cloning and sequencing of *bglA*, coding for a thermostable beta-glucanase from *Rhodothermus marinus*; and Halldorsdottir S et al., Appl Microbiol Biotechnol. 49(3):277-84, 1998, which discloses the cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12. However, there remains a need for identification and characterization of novel cellulases, with improved properties, such as improved performance under conditions of thermal stress or in the presence of surfactants, increased specific activity, altered substrate cleavage pattern, and/or high level expression in vitro. [103] The development of new and improved cellulase compositions that comprise varying amounts CBH-type, EG-type and BG-type cellulases is of interest for use: (1) in detergent compositions that exhibit enhanced cleaning ability, function as a softening agent and/or improve the feel of cotton fabrics (e.g., "stone washing" or "biopolishing"); (2) in compositions for degrading wood pulp or other biomass into sugars (e.g., for bio-fuel production); and/or (3) in feed compositions.

[104] Also provided herein are whole cellulase preparation comprising cellulase variants. As used herein, the phrase "whole cellulase preparation" refers to both naturally occurring and non-naturally occurring cellulase containing compositions. A "naturally occurring" composition is one produced by a naturally occurring source and which comprises one or more cellobiohydrolase-type, one or more endoglucanase-type, and one or more beta- glucosidase components wherein each of these components is found at the ratio produced by the source. A naturally occurring composition is one that is produced by an organism unmodified with respect to the cellulolytic enzymes such that the ratio of the component enzymes is unaltered from that produced by the native

organism. A "non-naturally occurring" composition encompasses those compositions produced by: (1) combining component cellulolytic enzymes either in a naturally occurring ratio or non-naturally occurring, i.e., altered, ratio; or (2) modifying an organism to overexpress or underexpress one or more cellulolytic enzyme; or (3) modifying an organism such that at least one cellulolytic enzyme is deleted. Accordingly, in some embodiments, the whole cellulase preparation can have one or more of the various EGs and/or CBHs, and/or beta-glucosidase deleted. For example, EGI may be deleted alone or in combination with other EGs and/or CBHs.

[105] In general, the whole cellulase preparation includes enzymes including, but are not limited to: (i) endoglucanases (EG) or 1,4- β -d-glucan-4-glucanohydrolases (EC 3.2.1.4), (ii) exoglucanases, including 1,4- β -d-glucan glucanohydrolases (also known as cellodextrinases) (EC 3.2.1.74) and 1,4- β -d-glucan cellobiohydrolases (exo-cellobiohydrolases, CBH) (EC 3.2.1.91), and (iii) β -glucosidase (BG) or β -glucoside glucohydrolases (EC 3.2.1.21). [106] In the present disclosure, the whole cellulase preparation can be from any microorganism that is useful for the hydrolysis of a cellulosic material. In some embodiments, the whole cellulase preparation is a filamentous fungi whole cellulase. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota. In some embodiments, the whole cellulase preparation is a *Acremonium*, *Aspergillus*, *Emericella*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Scytalidium*, *Thielavia*, *Tolypocladium*, or *Trichoderma* species, whole cellulase. In some embodiments, the whole cellulase preparation is an *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, or *Aspergillus oryzae* whole cellulase. In another aspect, whole cellulase preparation is a *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*,

Fusarium gramineum, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, or *Fusarium venenatum* whole cellulase. In another aspect, the whole cellulase preparation is a *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Scytalidium thermophilum*, or *Thielavia terrestris* whole cellulase. In another aspect, the whole cellulase preparation is a *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* e.g., RL-P37 (Sheir-Neiss et al., Appl. Microbiol. Biotechnology, 20 (1984) pp. 46-53; Montenecourt B.S., Can., 1-20, 1987), QM9414 (ATCC No. 26921), NRRL 15709, ATCC 13631, 56764, 56466, 56767, or *Trichoderma viride* e.g., ATCC 32098 and 32086, whole cellulase. In some embodiments, the whole cellulase preparation is a *Trichoderma reesei* RutC30 whole cellulase, which is available from the American Type Culture Collection as *Trichoderma reesei* ATCC 56765. [107] Examples of commercial cellulase preparations suitable for use in the present disclosure include, for example, CELLUCLAST™ (available from Novozymes A/S) and LAMINEX™, IndiAge™ and Primafast™ LAMINEX BG enzyme (available Genencor Division, Danisco US. Inc.)

[108] In the present disclosure, the whole cellulase preparation can be from any microorganism cultivation method known in the art resulting in the expression of enzymes capable of hydrolyzing a cellulosic material. Fermentation can include shake flask cultivation, small- or large-scale fermentation, such as continuous, batch, fed-batch, or solid state fermentations in laboratory or industrial fermenters performed in a suitable medium and under conditions allowing the cellulase to be expressed or isolated.

[109] Generally, the microorganism is cultivated in a cell culture medium suitable for

production of enzymes capable of hydrolyzing a cellulosic material. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable culture media, temperature ranges and other conditions suitable for growth and cellulase production are known in the art. As a non-limiting example, the normal temperature range for the production of cellulases by *Trichoderma reesei* is 24°C to 28°C.

[110] Generally, the whole cellulase preparation is used as is produced by fermentation with no or minimal recovery and/or purification. For example, once cellulases are secreted by a cell into the cell culture medium, the cell culture medium containing the cellulases can be used. In some embodiments the whole cellulase preparation comprises the unfractionated contents of fermentation material, including cell culture medium, extracellular enzymes and cells. Alternatively, the whole cellulase preparation can be processed by any convenient method, e.g., by precipitation, centrifugation, affinity, filtration or any other method known in the art. In some embodiments, the whole cellulase preparation can be concentrated, for example, and then used without further purification. In some embodiments the whole cellulase preparation comprises chemical agents that decrease cell viability or kills the cells. In some embodiments, the cells are lysed or permeabilized using methods known in the art.

III. Molecular Biology

[111] In one embodiment this disclosure provides for the expression of variant *cbh2* genes under control of a promoter functional in a filamentous fungus. Therefore, this disclosure relies on routine techniques in the field of recombinant genetics (See, e.g., Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd ed., 1989; Kriegler, *Gene Transfer and Expression: A Laboratory Manual*, 1990; and Ausubel et al., eds.,

CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Greene Publishing and Wiley-Interscience, New York, 1994).

Methods of Mutating *cbh2* Nucleic Acid Sequences

[112] Any method known in the art that can introduce mutations is contemplated by the present disclosure.

[113] The present disclosure relates to the expression, purification and/or isolation and use of variant CBH2. These enzymes are preferably prepared by recombinant methods utilizing the *cbh2* gene from *H. jecorina*. The fermentation broth may be used with or without purification. [114] After the isolation and cloning of the *cbh2* gene from *H. jecorina*, other methods known in the art, such as site directed mutagenesis, are used to make the substitutions, additions or deletions that correspond to substituted amino acids in the expressed CB H2 variant. Again, site directed mutagenesis and other methods of incorporating amino acid changes in expressed proteins at the DNA level are known in the art (Sambrook et al., supra; and Ausubel et al., supra).

[115] DNA encoding an amino acid sequence variant of the *H. jecorina* CB H2 is prepared by a variety of methods known in the art. These methods include, but are not limited to, preparation by site-directed (or oligonucleotide-mediated) mutagenesis, PCR mutagenesis, and cassette mutagenesis of an earlier prepared DNA encoding the *H. jecorina* CB H2. [116] Site-directed mutagenesis is a preferred method for preparing substitution variants. This technique is well known in the art (see, e.g., Carter et al. Nucleic Acids Res. 13:4431- 4443 (1985) and Kunkel et al., Proc. Natl. Acad. Sci. USA 82:488 (1987)). Briefly, in carrying out site-directed mutagenesis of DNA, the starting DNA is altered by first hybridizing an oligonucleotide encoding the desired mutation to a single strand of such starting DNA. After hybridization, a DNA polymerase is used to synthesize an entire second strand, using the hybridized oligonucleotide as a primer, and

using the single strand of the starting DNA as a template. Thus, the oligonucleotide encoding the desired mutation is incorporated in the resulting double-stranded DNA.

[117] PCR mutagenesis is also suitable for making amino acid sequence variants of the starting polypeptide, i.e., H. jecorina CB H2. See Higuchi, in PCR Protocols, pp. 177-183 (Academic Press, 1990); and Valletta et al., Nuc. Acids Res. 17:723-733 (1989). See, also, for example Cadwell et al., PCR Methods and Applications, Vol 2, 28-33 (1992). Briefly, when small amounts of template DNA are used as starting material in a PCR, primers that differ slightly in sequence from the corresponding region in a template DNA can be used to generate relatively large quantities of a specific DNA fragment that differs from the template sequence only at the positions where the primers differ from the template. [118] Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al., Gene 34:315-323 (1985). The starting material is the plasmid (or other vector) comprising the starting polypeptide DNA to be mutated. The codon(s) in the starting DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the starting polypeptide DNA. The plasmid DNA is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures, wherein the two strands of the oligonucleotide are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 5' and 3' ends that are compatible with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated

DNA sequence. [119] Alternatively, or additionally, the desired amino acid sequence encoding a variant CB H2 can be determined, and a nucleic acid sequence encoding such amino acid sequence variant can be generated synthetically.

[120] The variant CBH2(s) so prepared may be subjected to further modifications, oftentimes depending on the intended use of the cellulase. Such modifications may involve further alteration of the amino acid sequence, fusion to heterologous polypeptide(s) and/or covalent modifications.

IV. cbh2 Nucleic Acids And CBH2 Polypeptides

A. Variant cbh2-Type Nucleic acids

[121] The nucleic acid sequence for the wild type cbh2 is shown in SEQ ID NO:1. The disclosure encompasses a nucleic acid molecule encoding the variant cellulases described herein. The nucleic acid may be a DNA molecule. [122] After DNA sequences that encode the CBH2 variants have been cloned into DNA constructs, the DNA is used to transform microorganisms. The microorganism to be transformed for the purpose of expressing a variant CBH2 according to the present disclosure may advantageously comprise a strain derived from *Trichoderma* sp. Thus, a preferred mode for preparing variant CBH2 cellulases according to the present disclosure comprises transforming a *Trichoderma* sp. host cell with a DNA construct comprising at least a fragment of DNA encoding a portion or all of the variant CBH2. The DNA construct will generally be functionally attached to a promoter. The transformed host cell is then grown under conditions so as to express the desired protein. Subsequently, the desired protein product may be purified to substantial homogeneity. [123] However, it may in fact be that the best expression vehicle for a given DNA encoding a variant CB H2 may differ from *H. jecorina*. Thus, it may be that it will be most advantageous to express a protein in a transformation host that bears phylogenetic similarity to the source

organism for the variant CB H2. In an alternative embodiment, *Aspergillus niger* can be used as an expression vehicle. For a description of transformation techniques with *A. niger*, see WO 98/31821, the disclosure of which is incorporated by reference in its entirety. [124] Accordingly, the present description of an *Aspergillus* spp. expression system is provided for illustrative purposes only and as one option for expressing the variant CB H2 of the disclosure. One of skill in the art, however, may be inclined to express the DNA encoding variant CBH2 in a different host cell if appropriate and it should be understood that the source of the variant CBH2 should be considered in determining the optimal expression host. Additionally, the skilled worker in the field will be capable of selecting the best expression system for a particular gene through routine techniques utilizing the tools available in the art.

B. Variant CBH2 Polypeptides

[125] The variant CBH2's of this disclosure have amino acid sequences that are derived from the amino acid sequence of a precursor CBH2. The amino acid sequence of the CB H2 variant differs from the precursor CBH2 amino acid sequence by the substitution, deletion or insertion of one or more amino acids of the precursor amino acid sequence. In a preferred embodiment, the precursor CBH2 is *Hypocrea jecorina* CBH2. The mature amino acid sequence of *H. jecorina* CBH2 is shown in Example 2 (SEQ ID NO:3). Thus, this disclosure is directed to CBH2 variants which contain amino acid residues at positions which are equivalent to the particular identified residue in *H. jecorina* CB H2. A residue (amino acid) of an CB H2 homolog is equivalent to a residue of *Hypocrea jecorina* CB H2 if it is either homologous (i.e., corresponding in position in either primary or tertiary structure) or is functionally analogous to a specific residue or portion of that residue in *Hypocrea jecorina* CB H2 (i.e., having the same or similar functional capacity to combine, react, or interact chemically or structurally). As used

herein, numbering is intended to correspond to that of the mature CBH2 amino acid sequence (SEQ ID NO:3).

[126] Alignment of amino acid sequences to determine homology is preferably determined by using a "sequence comparison algorithm." Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson &

Lipman, *Proc. Nat'l Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection, Visual inspection may utilize graphics packages such as, for example, MOE by Chemical Computing Group, Montreal Canada.

[127] An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul, et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes

to zero or below, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M⁵, N⁻⁴, and a comparison of both strands. [128] The BLAST algorithm then performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, an amino acid sequence is considered similar to a protease if the smallest sum probability in a comparison of the test amino acid sequence to a protease amino acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. [129] For purposes of the present disclosure, the degree of identity may be suitably determined by means of computer programs known in the art, such as GAP provided in the GCG program package (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wis., USA 53711)

(Needleman, S. B. and Wunsch, C. D., (1970), Journal of Molecular Biology, 48, 443-45), using GAP with the following settings for polynucleotide sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3. [130] A structural alignment between a *T. reesei* CB H2 and other cellulases may be used to identify equivalent/corresponding positions in other cellulases having a moderate to high degree of homology, e.g. , about 50%, 55%, 60 %, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

or even 99%, with *T. reesei* CB H2 (SEQ ID NO: 3). One method of obtaining the structural alignment is to use the Pile Up programme from the GCG package using default values of gap penalties, i.e., a gap creation penalty of 3.0 and gap extension penalty of 0.1. Other structural alignment methods include the hydrophobic cluster analysis (Gaboriaud et al. , FEBS Letters, 224:149-155, 1987) and reverse threading (Huber and Torda, Protein Science, 7:142-149, 1998). [131] An exemplary alignment of the mature form of various reference cellulases is provided as Figure 1. The reference cellulases include: *Hypocrea jecorina* (also known as *T. reesei*) CBH2 (SEQ ID NO:3), *Hypocrea koningii* CBH2 (SEQ ID NO:4), *Humicola insolens* CBH2 (SEQ ID NO:5), *Acremonium cellulolyticus* CBH2 (SEQ ID NO:6), *Agaricus bisporus* CBH2 (SEQ ID NO:7), *Fusarium osysporum* EG (SEQ ID NO: 8), *Phanerochaete chrysosporium* CBH2 (SEQ ID NO:9), *Talaromyces emersonii* CBH2 (SEQ ID NO: 10), *Thermobifida fusca* 6B/E3 CBH2 (SEQ ID NO: 11), *Thermobifida fusca* 6A/E2 EG (SEQ ID NO: 12), and *Cellulomonas fimi* CenA EG (SEQ ID NO: 13). Sequences were aligned using the ClustalW and MUSCLE multiple sequence alignment algorithms. A matrix showing the percent identity of cellulases of the sequence alignment of Figure 1 is provided in Table 1.

*Numbers in the top row and left column correspond to the SEQ ID NOS of the aligned sequences of Figure 1.

[132] Sequence searches are typically carried out using the BLASTN program when evaluating a given nucleic acid sequence relative to nucleic acid sequences in the GenBank DNA Sequences and other public databases. The BLASTX program is preferred for searching nucleic acid sequences that have been translated in all reading frames against amino acid sequences in the GenBank Protein Sequences and other public databases. Both BLASTN and BLASTX are run using default parameters of an

open gap penalty of 11.0, and an extended gap penalty of 1.0, and utilize the BLOSUM-62 matrix. (See, e.g., Altschul, et al., 1997.)

V. Expression Of Recombinant CBH2 Variants [133] The methods of the disclosure rely on the use cells to express variant CB H2, with no particular method of CBH2 expression required. The variant CB H2 is preferably secreted from the cells. The disclosure provides host cells which have been transduced, transformed or transfected with an expression vector comprising a variant CBH2-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the parental host cell prior to transduction, transformation or transfection and will be apparent to those skilled in the art.

[134] In one approach, a filamentous fungal cell or yeast cell is transfected with an expression vector having a promoter or biologically active promoter fragment or one or more (e.g., a series) of enhancers which functions in the host cell line, operably linked to a DNA segment encoding variant CB H2, such that variant CBH2 is expressed in the cell line.

A. Nucleic Acid Constructs/Expression Vectors.

[135] Natural or synthetic polynucleotide fragments encoding variant CBH2 ("CBH2-encoding nucleic acid sequences") may be incorporated into heterologous nucleic acid constructs or vectors, capable of introduction into, and replication in, a filamentous fungal or yeast cell. The vectors and methods disclosed herein are suitable for use in host cells for the expression of variant CB H2. Any vector may be used as long as it is replicable and viable in the cells into which it is introduced. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Cloning and expression vectors are also described in Sambrook et al., 1989, Ausubel F M et al., 1989, and Strathern et al., *The Molecular Biology of the Yeast*

Saccharomyces, 1981, each of which is expressly incorporated by reference herein. Appropriate expression vectors for fungi are described in van den Hondel, C. A. M. J. J. et al. (1991) In: Bennett, J. W. and Lasure, L. L. (eds.) More Gene Manipulations in Fungi. Academic Press, pp. 396-428. The appropriate DNA sequence may be inserted into a plasmid or vector (collectively referred to herein as "vectors") by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by standard procedures. Such procedures and related sub-cloning procedures are deemed to be within the scope of knowledge of those skilled in the art.

[136] Recombinant filamentous fungi comprising the coding sequence for variant CBH2 may be produced by introducing a heterologous nucleic acid construct comprising the variant CBH2 coding sequence into the cells of a selected strain of the filamentous fungi. [137] Once the desired form of a variant cbh2 nucleic acid sequence is obtained, it may be modified in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, trans versions, deletions, and insertions may be performed on the naturally occurring sequence.

[138] A selected variant cbh2 coding sequence may be inserted into a suitable vector according to well-known recombinant techniques and used to transform filamentous fungi capable of CBH2 expression. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used to clone and express variant CBH2. Therefore it is appreciated that such substitutions in the coding region fall within the sequence variants covered by the present disclosure. Any and all of these sequence variants can be utilized in the same way as described herein for a parent CBH2-encoding nucleic acid sequence. [139] The present disclosure also includes recombinant

nucleic acid constructs comprising one or more of the variant CBH2-encoding nucleic acid sequences as described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the disclosure has been inserted, in a forward or reverse orientation.

[140] Heterologous nucleic acid constructs may include the coding sequence for variant cbh2. (i) in isolation; (ii) in combination with additional coding sequences; such as fusion protein or signal peptide coding sequences, where the cbh2 coding sequence is the dominant coding sequence; (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; and/or (iv) in a vector or host environment in which the cbh2 coding sequence is a heterologous gene. [141] In one aspect of the present disclosure, a heterologous nucleic acid construct is employed to transfer a variant CBH2-encoding nucleic acid sequence into a cell in vitro, with established filamentous fungal and yeast lines preferred. For long-term, production of variant CB H2, stable expression is preferred. It follows that any method effective to generate stable transformants may be used in practicing the disclosure. [142] Appropriate vectors are typically equipped with a selectable marker-encoding nucleic acid sequence, insertion sites, and suitable control elements, such as promoter and termination sequences. The vector may comprise regulatory sequences, including, for example, non-coding sequences, such as introns and control elements, i.e., promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in host cells (and/or in a vector or host cell environment in which a modified soluble protein antigen coding sequence is not normally expressed), operably linked to the coding sequence. Large numbers of suitable

vectors and promoters are known to those of skill in the art, many of which are commercially available and/or are described in Sambrook, et al., (supra).

[143] Exemplary promoters include both constitutive promoters and inducible promoters, examples of which include a CMV promoter, an SV40 early promoter, an RSV promoter, an EF-I. alpha, promoter, a promoter containing the tet responsive element (TRE) in the tet-on or tet-off system as described (ClonTech and BASF), the beta actin promoter and the metallothionine promoter that can upregulated by addition of certain metal salts. A promoter sequence is a DNA sequence which is recognized by the particular filamentous fungus for expression purposes. It is operably linked to DNA sequence encoding a variant CB H2 polypeptide. Such linkage comprises positioning of the promoter with respect to the initiation codon of the DNA sequence encoding the variant CBH2 polypeptide in the disclosed expression vectors. The promoter sequence contains transcription and translation control sequence which mediate the expression of the variant CBH2 polypeptide. Examples include the promoters from the *Aspergillus niger*, *A. awamori* or *A. oryzae* glucoamylase, alpha- amylase, or alpha-glucosidase encoding genes; the *A. nidulans* *gpdA* or *trpC* Genes; the *Neurospora crassa* *cbhl* or *trpl* genes; the *A. niger* or *Rhizomucor miehei* aspartic proteinase encoding genes; the *H. jecorina* (*T. reesei*) *cbhl*, *cbh2*, *egl1*, *egl2*, or other cellulase encoding genes.

[144] The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art. Typical selectable marker genes include *argB* from *A. nidulans* or *T. reesei*, *amdS* from *A. nidulans*, *pyr4* from *Neurospora crassa* or *T. reesei*, *pyrG* from *Aspergillus niger* or *A. nidulans*. Additional exemplary selectable markers include, but are not limited to *trpc*, *trpl*, *oliC31*, *niaD* or *Ieu2*, which are included in heterologous nucleic acid constructs used to transform a mutant strain such as *trp-*, *pyr-*, *leu-* and the like. [145] Such selectable

markers confer to transformants the ability to utilize a metabolite that is usually not metabolized by the filamentous fungi. For example, the *amdS* gene from *H. jecorina* which encodes the enzyme acetamidase that allows transformant cells to grow on acetamide as a nitrogen source. The selectable marker (e.g. *pyrG*) may restore the ability of an auxotrophic mutant strain to grow on a selective minimal medium or the selectable marker (e.g. *olic31*) may confer to transformants the ability to grow in the presence of an inhibitory drug or antibiotic.

[146] The selectable marker coding sequence is cloned into any suitable plasmid using methods generally employed in the art. Exemplary plasmids include pUC18, pBR322, pRAX and pUC100. The pRAX plasmid contains AMAL sequences from *A. nidulans*, which make it possible to replicate in *A. niger*.

[147] The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Sambrook et al., 1989; Freshney, *Animal Cell Culture*, 1987; Ausubel, et al., 1993; and Coligan et al., *Current Protocols in Immunology*, 1991.

B. Host Cells and Culture Conditions For CBH2 Production (i) Filamentous Fungi

[148] Thus, the present disclosure provides filamentous fungi comprising cells which have been modified, selected and cultured in a manner effective to result in variant CBH2 production or expression relative to the corresponding non-transformed parental fungi. [149] Examples of species of parental filamentous fungi that may be treated and/or modified for variant CBH2 expression include, but are not limited to *Trichoderma*, e.g., *Trichoderma reesei*, *Trichoderma longibrachiatum*, *Trichoderma viride*, *Trichoderma koningii*; *Penicillium* sp., *Humicola* sp., including *Humicola insolens*; *Aspergillus* sp., *Chrysosporium* sp., *Fusarium* sp., *Hypocrea* sp., and

Emmericella sp. [150] CB H2 expressing cells are cultured under conditions typically employed to culture the parental fungal line. Generally, cells are cultured in a standard medium containing physiological salts and nutrients, such as described in Pourquie, J. et al., *Biochemistry and Genetics of Cellulose Degradation*, eds. Aubert, J. P. et al., Academic Press, pp. 71-86, 1988 and Ilmen, M. et al., *Appl. Environ. Microbiol.* 63:1298-1306, 1997. Culture conditions are also standard, e.g., cultures are incubated at 28.degree. C. in shaker cultures or fermenters until desired levels of CB H2 expression are achieved.

[151] Preferred culture conditions for a given filamentous fungus may be found in the scientific literature and/or from the source of the fungi such as the American Type Culture Collection (ATCC; www.atcc.org/). After fungal growth has been established, the cells are exposed to conditions effective to cause or permit the expression of variant CBH2. [152] In cases where a CBH2 coding sequence is under the control of an inducible promoter, the inducing agent, e.g., a sugar, metal salt or antibiotics, is added to the medium at a concentration effective to induce CBH2 expression.

[153] In one embodiment, the strain comprises *Aspergillus niger*, which is a useful strain for obtaining overexpressed protein. For example *A. niger* var *awamori* dgr246 is known to secrete elevated amounts of secreted cellulases (Goedegebuur et al., *Curr. Genet* (2002) 41: 89-98). Other strains of *Aspergillus niger* var *awamori* such as GCD AP3, GCD AP4 and GAP3-4 are known Ward et al (Ward, M, Wilson, L. J. and Kodama, K. H., 1993, *Appl. Microbiol. Biotechnol.* 39:738-743).

[154] In another embodiment, the strain comprises *Trichoderma reesei*, which is a useful strain for obtaining overexpressed protein. For example, RL-P37, described by Sheir-Neiss, et al., *Appl. Microbiol. Biotechnol.* 20:46-53 (1984) is known to secrete elevated amounts of cellulase enzymes. Functional equivalents of RL-P37 include

Trichoderma reesei strain RUT- C30 (ATCC No. 56765) and strain QM9414 (ATCC No. 26921). It is contemplated that these strains would also be useful in overexpressing variant CB H2. [155] Where it is desired to obtain the variant CB H2 in the absence of potentially detrimental native cellulolytic activity, it is useful to obtain a *Trichoderma* host cell strain which has had one or more cellulase genes deleted prior to introduction of a DNA construct or plasmid containing the DNA fragment encoding the variant CB H2. Such strains may be prepared by the method disclosed in U.S. Pat. No. 5,246,853 and WO 92/06209, which disclosures are hereby incorporated by reference. By expressing a variant CBH2 cellulase in a host microorganism that is missing one or more cellulase genes, the identification and subsequent purification procedures are simplified. Any gene from *Trichoderma* sp. which has been cloned can be deleted, for example, the *cbhl*, *cbh2*, *egl1*, and *egl2* genes as well as those encoding EG III and/or EGV protein (see e.g., U.S. Pat. No. 5,475,101 and WO 94/28117, respectively). [156] Gene deletion may be accomplished by inserting a form of the desired gene to be deleted or disrupted into a plasmid by methods known in the art. The deletion plasmid is then cut at an appropriate restriction enzyme site(s), internal to the desired gene coding region, and the gene coding sequence or part thereof replaced with a selectable marker. Flanking DNA sequences from the locus of the gene to be deleted or disrupted, preferably between about 0.5 to 2.0 kb, remain on either side of the selectable marker gene. An appropriate deletion plasmid will generally have unique restriction enzyme sites present therein to enable the fragment containing the deleted gene, including flanking DNA sequences, and the selectable marker gene to be removed as a single linear piece. [157] A selectable marker must be chosen so as to enable detection of the transformed microorganism. Any selectable marker gene that is expressed in the selected microorganism will be suitable. For example, with *Aspergillus* sp., the

selectable marker is chosen so that the presence of the selectable marker in the transformants will not significantly affect the properties thereof. Such a selectable marker may be a gene that encodes an assayable product. For example, a functional copy of an *Aspergillus* sp. gene may be used which if lacking in the host strain results in the host strain displaying an auxotrophic phenotype. Similarly, selectable markers exist for *Trichoderma* sp.

[158] In one embodiment, a *pyrG*- derivative strain of *Aspergillus* sp. is transformed with a functional *pyrG* gene, which thus provides a selectable marker for transformation. A *pyrG*- derivative strain may be obtained by selection of *Aspergillus* sp. strains that are resistant to fluoroorotic acid (FOA). The *pyrG* gene encodes orotidine-5'-monophosphate decarboxylase, an enzyme required for the biosynthesis of uridine. Strains with an intact *pyrG* gene grow in a medium lacking uridine but are sensitive to fluoroorotic acid. It is possible to select *pyrG*- derivative strains that lack a functional orotidine monophosphate decarboxylase enzyme and require uridine for growth by selecting for FOA resistance. Using the FOA selection technique it is also possible to obtain uridine-requiring strains which lack a functional orotate pyrophosphoribosyl transferase. It is possible to transform these cells with a functional copy of the gene encoding this enzyme (Berges & Barreau, *Curr. Genet.* 19:359-365 (1991), and van Hartingsveldt et al., (1986) Development of a homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. *Mol. Gen. Genet.* 206:71-75). Selection of derivative strains is easily performed using the FOA resistance technique referred to above, and thus, the *pyrG* gene is preferably employed as a selectable marker. [159] In a second embodiment, a *pyr4*- derivative strain of *Hyprocrea* sp. (*Hyprocrea* sp. (*Trichoderma* sp.)) is transformed with a functional *pyr4* gene, which thus provides a selectable marker for transformation. A *pyr4.sup.-*

derivative strain may be obtained by selection of *Hyprocrea* sp. (*Trichoderma* sp.) strains that are resistant to fluoroorotic acid (FOA). The *pyr4* gene encodes orotidine-5'-monophosphate decarboxylase, an enzyme required for the biosynthesis of uridine. Strains with an intact *pyr4* gene grow in a medium lacking uridine but are sensitive to fluoroorotic acid. It is possible to select *pyr4*.sup.- derivative strains that lack a functional orotidine monophosphate decarboxylase enzyme and require uridine for growth by selecting for FOA resistance. Using the FOA selection technique it is also possible to obtain uridine-requiring strains which lack a functional orotate pyrophosphoribosyl transferase. It is possible to transform these cells with a functional copy of the gene encoding this enzyme (Berges & Barreau, 1991). Selection of derivative strains is easily performed using the FOA resistance technique referred to above, and thus, the *pyr4* gene is preferably employed as a selectable marker. [160] To transform *pyrG*.- *Aspergillus* sp. or *pyr4*- *Hyprocrea* sp. (*Trichoderma* sp.) so as to be lacking in the ability to express one or more cellulase genes, a single DNA fragment comprising a disrupted or deleted cellulase gene is then isolated from the deletion plasmid and used to transform an appropriate *pyr*- *Aspergillus* or *pyr*- *Trichoderma* host. Transformants are then identified and selected based on their ability to express the *pyrG* or *pyr4*, respectively, gene product and thus compliment the uridine auxotrophy of the host strain. Southern blot analysis is then carried out on the resultant transformants to identify and confirm a double crossover integration event that replaces part or all of the coding region of the genomic copy of the gene to be deleted with the appropriate *pyr* selectable markers. [161] Although the specific plasmid vectors described above relate to preparation of *pyr*- transformants, the present disclosure is not limited to these vectors. Various genes can be deleted and replaced in the *Aspergillus* sp. or *Hyprocrea* sp. (*Trichoderma* sp.) strain using the above techniques. In addition, any available

selectable markers can be used, as discussed above. In fact, any host, e.g., *Aspergillus* sp. or *Hyprocrea* sp., gene that has been cloned, and thus identified, can be deleted from the genome using the above-described strategy. [162] As stated above, the host strains used may be derivatives of *Hyprocrea* sp. (*Trichoderma* sp.) that lack or have a nonfunctional gene or genes corresponding to the selectable marker chosen. For example, if the selectable marker of *pyrG* is chosen for *Aspergillus* sp., then a specific *pyrG*- derivative strain is used as a recipient in the transformation procedure. Also, for example, if the selectable marker of *pyr4* is chosen for a *Hyprocrea* sp., then a specific *pyr4*- derivative strain is used as a recipient in the transformation procedure. Similarly, selectable markers comprising *Hyprocrea* sp.

(*Trichoderma* sp.) genes equivalent to the *Aspergillus nidulans* genes *amdS*, *argB*, *trpC*, *niaD* may be used. The corresponding recipient strain must therefore be a derivative strain such as *argB*-, *trpC*-, *niaD*-, respectively. [163] DNA encoding the CBH2 variant is then prepared for insertion into an appropriate microorganism. According to the present disclosure, DNA encoding a CBH2 variant comprises the DNA necessary to encode for a protein that has functional cellulolytic activity. The DNA fragment encoding the CBH2 variant may be functionally attached to a fungal promoter sequence, for example, the promoter of the *glaA* gene in *Aspergillus* or the promoter of the *cbhl* or *eglI* genes in *Trichoderma*.

[164] It is also contemplated that more than one copy of DNA encoding a CBH2 variant may be recombined into the strain to facilitate overexpression. The DNA encoding the CBH2 variant may be prepared by the construction of an expression vector carrying the DNA encoding the variant. The expression vector carrying the inserted DNA fragment encoding the CBH2 variant may be any vector which is capable of replicating autonomously in a given host organism or of integrating into the DNA of the host,

typically a plasmid. In preferred embodiments two types of expression vectors for obtaining expression of genes are contemplated. The first contains DNA sequences in which the promoter, gene-coding region, and terminator sequence all originate from the gene to be expressed. Gene truncation may be obtained where desired by deleting undesired DNA sequences (e.g., coding for unwanted domains) to leave the domain to be expressed under control of its own transcriptional and translational regulatory sequences. A selectable marker may also be contained on the vector allowing the selection for integration into the host of multiple copies of the novel gene sequences.

[165] The second type of expression vector is preassembled and contains sequences required for high-level transcription and a selectable marker. It is contemplated that the coding region for a gene or part thereof can be inserted into this general-purpose expression vector such that it is under the transcriptional control of the expression cassettes promoter and terminator sequences.

[166] For example, in *Aspergillus*, pRAX is such a general-purpose expression vector. Genes or part thereof can be inserted downstream of the strong *glac* promoter. [167] For example, in *Hypocrea*, pTEX is such a general-purpose expression vector. Genes or part thereof can be inserted downstream of the strong *cbh1* promoter. [168] In the vector, the DNA sequence encoding the CBH2 variant of the present disclosure should be operably linked to transcriptional and translational sequences, i.e., a suitable promoter sequence and signal sequence in reading frame to the structural gene. The promoter may be any DNA sequence that shows transcriptional activity in the host cell and may be derived from genes encoding proteins either homologous or heterologous to the host cell. An optional signal peptide provides for extracellular production of the CBH2 variant. The DNA encoding the signal sequence is preferably that which is naturally associated with the gene to be expressed, however the signal sequence from any suitable source, for

example an exo- cellobiohydrolase or endoglucanase from *Trichoderma*, is contemplated in the present disclosure.

[169] The procedures used to ligate the DNA sequences coding for the variant CBH2 of the present disclosure with the promoter, and insertion into suitable vectors are well known in the art. [170] The DNA vector or construct described above may be introduced in the host cell in accordance with known techniques such as transformation, transfection, microinjection, microporation, biolistic bombardment and the like.

[171] In the preferred transformation technique, it must be taken into account that the permeability of the cell wall to DNA in *Hyphocrea* sp. (*Trichoderma* sp.) is very low. Accordingly, uptake of the desired DNA sequence, gene or gene fragment is at best minimal. There are a number of methods to increase the permeability of the *Hyphocrea* sp. (*Trichoderma* sp.) cell wall in the derivative strain (i.e., lacking a functional gene corresponding to the used selectable marker) prior to the transformation process. [172] The preferred method in the present disclosure to prepare *Aspergillus* sp. or *Hyphocrea* sp. (*Trichoderma* sp.) for transformation involves the preparation of protoplasts from fungal mycelium. See Campbell et al. Improved transformation efficiency of *A. niger* using homologous *niaD* gene for nitrate reductase. *Curr. Genet.* 16:53-56; 1989. The mycelium can be obtained from germinated vegetative spores. The mycelium is treated with an enzyme that digests the cell wall resulting in protoplasts. The protoplasts are then protected by the presence of an osmotic stabilizer in the suspending medium. These stabilizers include sorbitol, mannitol, potassium chloride, magnesium sulfate and the like. Usually the concentration of these stabilizers varies between 0.8 M and 1.2 M. It is preferable to use about a 1.2 M solution of sorbitol in the suspension medium. [173] Uptake of the DNA into the host strain, (*Aspergillus* sp. or *Hyphocrea* sp. (*Trichoderma* sp.)), is dependent upon the calcium ion concentration. Generally between about 10 mM

CaCl₂ and 50 mM CaCl₂ is used in an uptake solution. Besides the need for the calcium ion in the uptake solution, other items generally included are a buffering system such as TE buffer (10 mM Tris, pH 7.4; 1 mM EDTA) or 10 mM MOPS, pH 6.0 buffer (morpholinepropanesulfonic acid) and polyethylene glycol (PEG). It is believed that the polyethylene glycol acts to fuse the cell membranes thus permitting the contents of the medium to be delivered into the cytoplasm of the host cell, by way of example either *Aspergillus* sp. or *Hyphocrea* sp. strain, and the plasmid DNA is transferred to the nucleus. This fusion frequently leaves multiple copies of the plasmid DNA integrated into the host chromosome.

[174] Usually a suspension containing the *Aspergillus* sp. protoplasts or cells that have been subjected to a permeability treatment at a density of 10^5 to 10^6 /mL, preferably 2×10^5 /mL are used in transformation. Similarly, a suspension containing the *Hyphocrea* sp. (*Trichoderma* sp.) protoplasts or cells that have been subjected to a permeability treatment at a density of 10^8 to 10^9 /mL, preferably 2×10^8 /mL are used in transformation. A volume of 100 μ L of these protoplasts or cells in an appropriate solution (e.g., 1.2 M sorbitol; 50 mM CaCl₂) are mixed with the desired DNA. Generally a high concentration of PEG is added to the uptake solution. From 0.1 to 1 volume of 25% PEG 4000 can be added to the protoplast suspension. However, it is preferable to add about 0.25 volumes to the protoplast suspension. Additives such as dimethyl sulfoxide, heparin, spermidine, potassium chloride and the like may also be added to the uptake solution and aid in transformation.

[175] Generally, the mixture is then incubated at approximately 0°C. for a period of between 10 to 30 minutes. Additional PEG is then added to the mixture to further enhance the uptake of the desired gene or DNA sequence. The 25% PEG 4000 is generally added in volumes of 5 to 15 times the volume of the transformation mixture;

however, greater and lesser volumes may be suitable. The 25% PEG 4000 is preferably about 10 times the volume of the transformation mixture. After the PEG is added, the transformation mixture is then incubated either at room temperature or on ice before the addition of a sorbitol and

CaCl₂ solution. The protoplast suspension is then further added to molten aliquots of a growth medium. This growth medium permits the growth of transformants only. Any growth medium can be used in the present disclosure that is suitable to grow the desired transformants. However, if Pyr^{sup}+ transformants are being selected it is preferable to use a growth medium that contains no uridine. The subsequent colonies are transferred and purified on a growth medium depleted of uridine.

[176] At this stage, stable transformants may be distinguished from unstable transformants by their faster growth rate and, in *Trichoderma*, for example, the formation of circular colonies with a smooth, rather than ragged outline on solid culture medium lacking uridine. Additionally, in some cases a further test of stability may be made by growing the transformants on solid non-selective medium (i.e. containing uridine), harvesting spores from this culture medium and determining the percentage of these spores which will subsequently germinate and grow on selective medium lacking uridine.

[177] In a particular embodiment of the above method, the CB H2 variant(s) are recovered in active form from the host cell after growth in liquid media as a result of the appropriate post translational processing of the CB H2 variant.

(ii) Yeast

[178] The present disclosure also contemplates the use of yeast as a host cell for CB H2 production. Several other genes encoding hydrolytic enzymes have been expressed in various strains of the yeast *S. cerevisiae*. These include sequences encoding for two

endoglucanases (Penttila et al., Yeast vol. 3, pp 175-185, 1987), two cellobiohydrolases (Penttila et al., Gene, 63: 103-112, 1988) and one beta-glucosidase from *Trichoderma reesei* (Cummings and Fowler, Curr. Genet. 29:227-233, 1996), a xylanase from *Aureobasidium pullulans* (Li and Ljungdahl, Appl. Environ. Microbiol. 62, no.1, pp. 209-213, 1996), an alpha-amylase from wheat (Rothstein et al., Gene 55:353-356, 1987), etc. In addition, a cellulase gene cassette encoding the *Butyrivibrio fibrisolvens* endo-[beta]-1,4-glucanase (END1), *Phanerochaete chrysosporium* cellobiohydrolase (CBH1), the *Ruminococcus flavefaciens* cellodextrinase (CEL1) and the *Endomyces fibrilizer* cellobiase (BgII) was successfully expressed in a laboratory strain of *S. cerevisiae* (Van Rensburg et al., Yeast, vol. 14, pp. 67-76, 1998).

C. Introduction of a CBH2-Encoding Nucleic Acid Sequence into Host Cells. [179] The disclosure further provides cells and cell compositions which have been genetically modified to comprise an exogenously provided variant CBH2-encoding nucleic acid sequence. A parental cell or cell line may be genetically modified (i.e., transduced, transformed or transfected) with a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc, as further described above.

[180] The methods of transformation of the present disclosure may result in the stable integration of all or part of the transformation vector into the genome of the filamentous fungus. However, transformation resulting in the maintenance of a self-replicating extra-chromosomal transformation vector is also contemplated.

[181] Many standard transfection methods can be used to produce *Trichoderma reesei* cell lines that express large quantities of the heterologous protein. Some of the published methods for the introduction of DNA constructs into cellulase-producing strains of

Trichoderma include Lorito, Hayes, DiPietro and Harman, 1993, *Curr. Genet.* 24: 349-356; Goldman, VanMontagu and Herrera-Estrella, 1990, *Curr. Genet.* 17:169-174; Penttila, Nevalainen, Ratto, Salminen and Knowles, 1987, *Gene* 6: 155-164, for *Aspergillus* Yelton, Hamer and Timberlake, 1984, *Proc. Natl. Acad. Sci. USA* 81: 1470-1474, for *Fusarium* Bajar, Podila and Kolattukudy, 1991, *Proc. Natl. Acad. Sci. USA* 88: 8202-8212, for *Streptomyces* Hopwood et al., 1985, The John Innes Foundation, Norwich, UK and for *Bacillus* Brigidi, DeRossi, Bertarini, Riccardi and Matteuzzi, 1990, *FEMS Microbiol. Lett.* 55: 135-138). [182] Other methods for introducing a heterologous nucleic acid construct (expression vector) into filamentous fungi (e.g., *H. jecorina*) include, but are not limited to the use of a particle or gene gun, permeabilization of filamentous fungi cells walls prior to the transformation process (e.g., by use of high concentrations of alkali, e.g., 0.05 M to 0.4 M CaCl_2 or lithium acetate), protoplast fusion or *Agrobacterium* mediated transformation. An exemplary method for transformation of filamentous fungi by treatment of protoplasts or spheroplasts with polyethylene glycol and CaCl_2 is described in Campbell, E. I. et al., *Curr. Genet.* 16:53-56, 1989 and Penttila, M. et al., *Gene*, 63:11-22, 1988.

[183] Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, biolistics, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., *supra*). Also of use is the *Agrobacterium*-mediated transfection method described in U.S. Pat. No. 6,255,115. It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the heterologous gene. [184] In

addition, heterologous nucleic acid constructs comprising a variant CBH2- encoding nucleic acid sequence can be transcribed in vitro, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection. [185] The disclosure further includes novel and useful transformants of filamentous fungi such as *H. jecorina* and *A. niger* for use in producing fungal cellulase compositions. The disclosure includes transformants of filamentous fungi especially fungi comprising the variant CBH2 coding sequence, or deletion of the endogenous *cbh* coding sequence. [186] Following introduction of a heterologous nucleic acid construct comprising the coding sequence for a variant *cbh2*, the genetically modified cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying expression of a variant CBH2-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the host cell selected for expression, and will be apparent to those skilled in the art.

[187] The progeny of cells into which such heterologous nucleic acid constructs have been introduced are generally considered to comprise the variant CBH2-encoding nucleic acid sequence found in the heterologous nucleic acid construct. [188] The disclosure further includes novel and useful transformants of filamentous fungi such as *H. jecorina* for use in producing fungal cellulase compositions. *Aspergillus niger* may also be used in producing the variant CB H2. The disclosure includes transformants of filamentous fungi especially fungi comprising the variant *cbh 2* coding sequence, or deletion of the endogenous *cbh2* coding sequence. [189] Stable transformants of filamentous fungi can generally be distinguished from unstable transformants by their faster growth rate and, in *Trichoderma*, for example, the formation of circular colonies with a smooth rather than ragged outline on solid culture medium. Additionally, in some cases, a further test of stability can be made by growing the transformants on solid non-

selective medium, harvesting the spores from this culture medium and determining the percentage of these spores which will subsequently germinate and grow on selective medium.

VI. Isolation And Purification Of Recombinant CBH2 Protein

[190] In general, a variant CBH2 protein produced in cell culture is secreted into the medium and may be purified or isolated, e.g., by removing unwanted components from the cell culture medium. However, in some cases, a variant CBH2 protein may be produced in a cellular form necessitating recovery from a cell lysate. In such cases the variant CBH2 protein is purified from the cells in which it was produced using techniques routinely employed by those of skill in the art. Examples include, but are not limited to, affinity chromatography (Tilbeurgh et al., FEBS Lett. 16:215, 1984), ion-exchange chromatographic methods (Goyal et al., Bioresource Technol. 36:37-50, 1991; Fliess et al., Eur. J. Appl. Microbiol. Biotechnol. 17:314-318, 1983; Bhikhabhai et al., J. Appl. Biochem. 6:336-345, 1984; Ellouz et al., J. Chromatography 396:307-317, 1987), including ion-exchange using materials with high resolution power (Medve et al., J. Chromatography A 808: 153-165, 1998), hydrophobic interaction chromatography (Tomaz and Queiroz, J. Chromatography A 865: 123-128, 1999), and two-phase partitioning (Brumbauer, et al., Bioseparation 7:287-295, 1999).

[191] Typically, the variant CBH2 protein is fractionated to segregate proteins having selected properties, such as binding affinity to particular binding agents, e.g., antibodies or receptors; or which have a selected molecular weight range, or range of isoelectric points. [192] Once expression of a given variant CB H2 protein is achieved, the CB H2

protein thereby produced is purified from the cells or cell culture. Exemplary procedures suitable for such purification include the following: antibody-affinity column chromatography, ion exchange chromatography; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; and gel filtration using, e.g., Sephadex G-75. Various methods of protein purification may be employed and such methods are known in the art and described e.g. in Deutscher, *Methods in Enzymology*, vol. 182, no. 57, pp. 779, 1990; Scopes, *Methods Enzymol.* 90: 479-91, 1982. The purification step(s) selected will depend, e.g., on the nature of the production process used and the particular protein produced.

VII. Utility of cbh2 and CBH2

[193] It can be appreciated that the variant cbh nucleic acids, the variant CBH2 protein and compositions comprising variant CB H2 protein activity find utility in a wide variety applications, some of which are described below.

[194] New and improved cellulase compositions that comprise varying amounts BG-type, EG-type and variant CBH-type cellulases find utility in detergent compositions that exhibit enhanced cleaning ability, function as a softening agent and/or improve the feel of cotton fabrics (e.g., "stone washing" or "biopolishing"), in compositions for degrading wood pulp into sugars (e.g., for bio-ethanol production), and/or in feed compositions. The isolation and characterization of cellulase of each type provides the ability to control the aspects of such compositions. [195] In one approach, the cellulase of the disclosure finds utility in detergent compositions or in the treatment of fabrics to improve the feel and appearance. [196] Since the rate of hydrolysis of cellulosic products may be increased by using a transformant having at least one additional copy of the cbh gene inserted into the genome, products that contain cellulose or

heteroglycans can be degraded at a faster rate and to a greater extent. Products made from cellulose such as paper, cotton, cellulosic diapers and the like can be degraded more efficiently in a landfill. Thus, the fermentation product obtainable from the transformants or the transformants alone may be used in compositions to help degrade by liquefaction a variety of cellulose products that add to the overcrowded landfills.

[197] Separate saccharification and fermentation is a process whereby cellulose present in biomass, e.g., corn stover, is converted to glucose and subsequently yeast strains convert glucose into ethanol. Simultaneous saccharification and fermentation is a process whereby cellulose present in biomass, e.g., corn stover, is converted to glucose and, at the same time and in the same reactor, yeast strains convert glucose into ethanol. Thus, in another approach, the variant CBH type cellulase of the disclosure finds utility in the degradation of biomass to ethanol. Ethanol production from readily available sources of cellulose provides a stable, renewable fuel source.

[198] Cellulose-based feedstocks are comprised of agricultural wastes, grasses and woods and other low-value biomass such as municipal waste (e.g., recycled paper, yard clippings, etc.). Ethanol may be produced from the fermentation of any of these cellulosic feedstocks. However, the cellulose must first be converted to sugars before there can be conversion to ethanol.

[199] A large variety of feedstocks may be used with the inventive variant CBH and the one selected for use may depend on the region where the conversion is being done. For example, in the Midwestern United States agricultural wastes such as wheat straw, corn stover and bagasse may predominate while in California rice straw may predominate. However, it should be understood that any available cellulosic biomass may be used in any region. [200] The methods of the present disclosure can be used in the production of monosaccharides, disaccharides, and polysaccharides as chemical or fermentation

feedstocks for microorganism for the production of organic products, chemicals and fuels, plastics, and other products or intermediates. In particular, the value of processing residues (dried distillers grain, spent grains from brewing, sugarcane bagasse, etc.) can be increased by partial or complete solubilization of cellulose or hemicellulose. In addition to ethanol, some chemicals that can be produced from cellulose and hemicellulose include, acetone, acetate, glycine, lysine, organic acids (e.g., lactic acid), 1,3-propanediol, butanediol, glycerol, ethylene glycol, furfural, polyhydroxyalkanoates, cis, cis-muconic acid, animal feed and xylose. [201] A cellulase composition containing an enhanced amount of cellobiohydrolase finds utility in ethanol production. Ethanol from this process can be further used as an octane enhancer or directly as a fuel in lieu of gasoline which is advantageous because ethanol as a fuel source is more environmentally friendly than petroleum derived products. It is known that the use of ethanol will improve air quality and possibly reduce local ozone levels and smog. Moreover, utilization of ethanol in lieu of gasoline can be of strategic importance in buffering the impact of sudden shifts in non-renewable energy and petrochemical supplies. [202] Ethanol can be produced via saccharification and fermentation processes from cellulosic biomass such as trees, herbaceous plants, municipal solid waste and agricultural and forestry residues. However, the ratio of individual cellulase enzymes within a naturally occurring cellulase mixture produced by a microbe may not be the most efficient for rapid conversion of cellulose in biomass to glucose. It is known that endoglucanases act to produce new cellulose chain ends which themselves are substrates for the action of cellobiohydrolases and thereby improve the efficiency of hydrolysis of the entire cellulase system. Therefore, the use of increased or optimized cellobiohydrolase activity may greatly enhance the production of ethanol.

[203] Thus, the inventive cellobiohydrolase finds use in the hydrolysis of cellulose to its sugar components. In one embodiment, a variant cellobiohydrolase is added to the biomass prior to the addition of a fermentative organism. In a second embodiment, a variant cellobiohydrolase is added to the biomass at the same time as a fermentative organism. Optionally, there may be other cellulase components present in either embodiment.

[204] In another embodiment the cellulosic feedstock may be pretreated. Pretreatment may be by elevated temperature and the addition of either of dilute acid, concentrated acid or dilute alkali solution. The pretreatment solution is added for a time sufficient to at least partially hydrolyze the hemicellulose components and then neutralized. [205] The major product of CBH2 action on cellulose is cellobiose which is available for conversion to glucose by BG activity (for instance in a fungal cellulase product). Either by the pretreatment of the cellulosic biomass or by the enzymatic action on the biomass, other sugars, in addition to glucose and cellobiose, can be made available from the biomass. The hemi-cellulose content of the biomass can be converted (by hemi-cellulases) to sugars such as xylose, galactose, mannose and arabinose. Thus, in a biomass conversion process, enzymatic saccharification can produce sugars that are made available for biological or chemical conversions to other intermediates or end-products. Therefore, the sugars generated from biomass find use in a variety of processes in addition to the generation of ethanol. Examples of such conversions are fermentation of glucose to ethanol (as reviewed by M. E. Himmel et al. pp 2-45, in "Fuels and Chemicals from Biomass", ACS Symposium Series 666, ed B. C. Saha and J. Woodward, 1997) and other biological conversions of glucose to 2,5-diketo-D-gluconate (U.S. Pat. No. 6,599,722), lactic acid (R. Datta and S-P. Tsai pp 224-236, *ibid*), succinate (R. R. Gokarn, M. A. Eiteman and J. Sridhar pp 237-263, *ibid*), 1,3

-propanediol (A- P. Zheng, H. Biebl and W-D. Deckwer pp 264-279, *ibid*), 2,3-butanediol (C. S. Gong, N. Cao and G. T. Tsao pp 280-293, *ibid*), and the chemical and biological conversions of xylose to xylitol (B. C. Saha and R. J. Bothast pp 307-319, *ibid*). See also, for example, WO 98/21339. [206] The detergent compositions of this disclosure may employ besides the cellulase composition (irrespective of the cellobiohydrolase content, i.e., cellobiohydrolase-free, substantially cellobiohydrolase-free, or cellobiohydrolase enhanced), a surfactant, including anionic, non-ionic and ampholytic surfactants, a hydrolase, building agents, bleaching agents, bluing agents and fluorescent dyes, caking inhibitors, solubilizers, cationic surfactants and the like. All of these components are known in the detergent art. The cellulase composition as described above can be added to the detergent composition either in a liquid diluent, in granules, in emulsions, in gels, in pastes, and the like. Such forms are well known to the skilled artisan. When a solid detergent composition is employed, the cellulase composition is preferably formulated as granules. Preferably, the granules can be formulated so as to contain a cellulase protecting agent. For a more thorough discussion, see U.S. Pat. No. 6,162,782 entitled "Detergent compositions containing cellulase compositions deficient in CB H2 type components," which is incorporated herein by reference.

[207] Preferably the cellulase compositions are employed from about 0.00005 weight percent to about 5 weight percent relative to the total detergent composition. More preferably, the cellulase compositions are employed from about 0.0002 weight percent to about 2 weight percent relative to the total detergent composition.

[208] In addition the variant CB H2 nucleic acid sequence finds utility in the identification and characterization of related nucleic acid sequences. A number of techniques useful for determining (predicting or confirming) the function of related

genes or gene products include, but are not limited to, (A) DNA/RNA analysis, such as (1) overexpression, ectopic expression, and expression in other species; (2) gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing (VIGS, see Baulcombe, 100 Years of Virology, Calisher and Horzinek eds., Springer- Verlag, New York, N.Y. 15:189-201, 1999); (3) analysis of the methylation status of the gene, especially flanking regulatory regions; and (4) in situ hybridization; (B) gene product analysis such as (1) recombinant protein expression; (2) antisera production, (3) immunolocalization; (4) biochemical assays for catalytic or other activity; (5) phosphorylation status; and (6) interaction with other proteins via yeast two- hybrid analysis; (C) pathway analysis, such as placing a gene or gene product within a particular biochemical or signaling pathway based on its overexpression phenotype or by sequence homology with related genes; and (D) other analyses which may also be performed to determine or confirm the participation of the isolated gene and its product in a particular metabolic or signaling pathway, and help determine gene function.

EXPERIMENTAL

[209] The present disclosure is described in further detail in the following examples, which are not in any way intended to limit the scope of the disclosure as claimed. The attached figures are meant to be considered as integral parts of the specification and description of the disclosure. The following examples are offered to illustrate, but not to limit the claimed disclosure

[210] In the experimental disclosure which follows, the following abbreviations apply: M (molar); mM (millimolar); μ M (micromolar); nM (nanomolar); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); gm (grams); mg (milligrams); μ g (micrograms); pg (picograms); L (liters); ml and mL (milliliters); μ l and μ L (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers);

U (units); V (volts); MW (molecular weight); sec (seconds); min(s) (minute/minutes); h(s) and hr(s) (hour/hours); °C (degrees Centigrade); QS (quantity sufficient); ND (not done); NA (not applicable); rpm (revolutions per minute); H₂O (water); dH₂O (deionized water); HCl (hydrochloric acid); aa (amino acid); bp (base pair); kb (kilobase pair); kD (kilodaltons); cDNA (copy or complementary DNA); DNA (deoxyribonucleic acid); ssDNA (single stranded DNA); dsDNA (double stranded DNA); dNTP (deoxyribonucleotide triphosphate); RNA (ribonucleic acid); MgCl₂ (magnesium chloride); NaCl (sodium chloride); w/v (weight to volume); v/v (volume to volume); g (gravity); OD (optical density); CNPG (chloro-nitro- phenyl-beta-D-glucoside); CNP (2-chloro-4-nitrophenol); APB (acid-pretreated bagasse); PASC (phosphoric acid swollen cellulose) PCS (acid-pretreated corn stover); Pi or PI (performance index); HPLC (high pressure liquid chromatography); PAGE (polyacrylamide gel electrophoresis); PCR (polymerase chain reaction); RT-PCR (reverse transcription PCR); and SEL (site evaluation library).

EXAMPLE 1

Assays

[211] The following assays were used in the examples described below. Any deviations from the protocols provided below are indicated in the examples. In these experiments, a spectrophotometer was used to measure the absorbance of the products formed after the completion of the reactions. A. Hexokinase Assay For Measurement Of Residual Glucose

[212] Residual glucose from *H. jecorina* culture supernatants expressing CBH2 variants was measured using a hexokinase assay. A volume of 5 µl of supernatant was added to 195 µl glucose hexokinase assay (Instrumentation Laboratory, Breda, Netherlands) in a 96-well microtiterplate (Costar Flat Bottom PS). The plates were incubated at room

temperature for 15 min. Following incubation, the absorbance was measured at 340 nm OD. Supernatants of cultures expressing residual glucose were excluded from pooling for further studies.

B. HPLC Assay for Protein Content Determination [213] The concentration of CBH2 variant proteins from pooled culture supernatants was determined using an Agilent 1100 (Hewlett Packard) HPLC equipped with a Proswift RP 2H column (Dionex). Ten microliters of sample, mixed with 50 μ l of 10% acetonitrile in filtered demineralized water was injected following equilibration of the HPLC column with 10% acetonitrile containing 0.01% trifluoroacetic acid. Compounds were eluted using a gradient of 10% to 30% acetonitrile from 0.3 min to 1 min, followed by a gradient of 30% to 65% from 1 min to 4 mins. Protein concentrations of CB H2 variants were determined from a calibration curve generated using purified wild-type CBH2 (6.25, 12.5, 25, 50 μ g/ml). To calculate performance index (P_i or PI), the ratio of the (average) total protein produced by a variant and (average) total protein produced by the wild-type at the same dose were averaged.

C. Specific Activity Determination by Phosphoric Acid Swollen Cellulose (PASC) Hydrolysis Assay

[214] Cellulose hydrolysis: Phosphoric acid swollen cellulose (PASC) was prepared from Avicel according to a published method (Walseth, Tappi 35:228, 1971; and Wood, Biochem J, 121:353-362, 1971). This material was diluted with buffer and water to achieve a 1% w/v mixture such that the final concentration of sodium acetate was 50mM, pH 5.0. One hundred microliters of a 1% suspension of PASC in 50 mM sodium acetate buffer (pH5.0) was dispensed in a 96-well microtiterplate (Costar Flat Bottom PS). Ten microliters of a 5 mg/ml culture supernatant from a CBH2 deleted strain was added to the PASC, and 5, 10, 15, or 20 μ l of pooled culture supernatants

from *H. jecorina* cells expressing either wild-type CBH2 or CB H2 variants were added to it. Deletion of the CBH2 gene from *Hypocrea jecorina* (also referred to as *Trichoderma reesei*) is described in U. S. Pat. Nos. 5,861,271 and 5,650,322. Compensating volumes of acetate buffer were added to make up for differences in total volume. The microtiterplate was sealed and incubated in a thermostatted incubator at 50 °C under continuous shaking at 900 rpm. After two hours, the hydrolysis reaction was stopped by the addition of 100 µl glycine buffer, pH 10 to each well. To calculate performance index (P_i or PI), the ratio of the (average) total sugar produced by a variant and (average) total sugar produced by the wild-type at the same dose were averaged. The hydrolysis reaction products were analyzed with the PAHBAH assay. PAHBAH assay: Aliquots of 150 µl of PAHBAH reducing sugar reagent (5% w/v p-hydroxybenzoic acid hydrazide (PAHBAH, Sigma # H9882, dissolved in 0.5 N HCl), (Lever, Anal Biochem, 47:273-279, 1972) were added to all wells of an empty microtiter plate. Ten microliters of the hydrolysis reaction supernatants were added to the PABA reaction plate. All plates were sealed and incubated at 69°C under continuous shaking of 900 rpm. After one hour the plates were placed on ice for five minutes and centrifuged at 720 x g at room temperature for five minutes. Samples of 80 µL of the developed PAHBAH reaction mixtures were transferred to a fresh (read) plate and absorbance was measured at 410 nm in a spectrophotometer. A cellobiose standard was included as control. A dose response curve was generated for wild-type CB H2 protein.

D. Specific Activity Determination By Hydrolysis Of Dilute Acid Pretreated Corn Stover (PCS)

[215] Pretreated Corn Stover (PCS): Corn stover was pretreated with 2% w/w H₂SO₄ as described (Schell et al., J Appl Biochem Biotechnol, 105:69-86, 2003) and followed

by multiple washes with deionized water to obtain a paste having a pH of 4.5. Sodium acetate buffer (pH 5.0) was then added to a final concentration of 50mM sodium acetate and, if necessary, this mixture was then titrated to pH 5.0 using 1N NaOH. The cellulase concentration in the reaction mixture was approximately 7%. Sixty-five microliters of this cellulose suspension was added per well to a 96-well microtiterplate (Nunc Flat Bottom PS). Ten microliters of a 5 mg/ml culture supernatant from a CBH2 deleted strain was added to the PCS, and 5, 10, 15, or 20 μ l of pooled culture supernatants from *H. jecorina* cells expressing either wild-type CBH2 or CBH2 variants were added to it. Compensating volumes of acetate buffer were added to make up for differences in total volume. After sealing of the plate, the plates were placed in a thermostatted incubator at 50 °C under continuous shaking of 1300 rpm for 5 minutes. The plates were then incubated at 50 °C while shaking at 220 rpm under 80% humidity to prevent drying. After 7 days the plates were put on ice for 5 min and the hydrolysis reaction was stopped by the addition of 100 μ l glycine buffer, pH 10 to each well. The hydrolysis reaction products were analyzed with the PAHBAH assay. To calculate performance index (PI or PI), the ratio of the (average) total sugar produced by a variant and (average) total sugar produced by the wild-type at the same dose were averaged.

PAHBAH assay: Aliquots of 150 μ l of PAHBAH reducing sugar reagent (5% w/v p-hydroxybenzoic acid hydrazide (PAHBAH, Sigma # H9882, dissolved in 0.5 N HCl), (Lever, Anal Biochem, 47:273-279, 1972) were added to all wells of an empty microtiter plate. Ten microliters of the hydrolysis reaction supernatants were added to the PABA reaction plate. All plates were sealed and incubated at 69°C under continuous shaking of 900 rpm. After one hour the plates were placed on ice for five minutes and centrifuged at 720 x g at room temperature for five minutes. Samples of 80 μ L of the developed PAHBAH reaction mixtures were transferred to a fresh (read) plate

and absorbance was measured at 410 nm in a spectrophotometer. A cellobiose standard was included as control. A dose response curve was generated for wild-type CB H2 protein.

E. Specific Activity Determination By Hydrolysis Of Ammonia Pretreated Corncob (CC)

[216] Corn cob was ground to pass through a 0.9mm screen then pretreated as described in Example 4 of patent application WO2006110901 (herein incorporated by reference for this method). Pretreated corn cob was used as a 7% cellulose suspension in 50 mM sodium acetate pH 5.0. Sixty-five microliters of the suspension was added per well to a 96- well microtiterplate (Nunc Flat Bottom PS). To each well, 20 μ l of 3.4 mg/ml supernatant from a strain (Δ egl1, Δ egl2, Δ cbh1, Δ cbh2) supplemented with 0.23 mg/ml purified *T. reesei* beta- glucosidase 1 (Cel3 A) was added. Twenty or forty microliters of pooled culture supernatants from *H. jecorina* cells expressing either wild-type CB H2 or CB H2 variants were added. Compensating volumes of acetate buffer were added to make up for differences in total volume. The plate was incubated at 50 °C while shaking at 220 rpm under 80% humidity to prevent drying. After 3 days the plate was put on ice for 5 min and 100 μ l of 100 mM glycine pH 10.0 was added. After mixing, the plate was centrifuged at 3000 rpm for 5 min. A volume of 10 μ l supernatant was diluted in 190 μ l water. Twenty μ l of the diluted solution was transferred to a new 96-well microtiterplate (Costar Flat Bottom PS) containing 100 μ l ABTS glucose assay mixture (2.74 mg/ml 2,2' azino-bis(3-ethylbenzo-thiazoline-6-sulfonic acid, 0.333 U/ml horseradish peroxidase type VI, 1 U/ml glucose oxidase) and increase in A₄₂₀ was recorded in a microtiterplate spectrophotometer (Spectramax Plus 384, Molecular Devices). A range of glucose concentrations was included as a standard on each plate (6.25, 12.5, 25, 50, 100, 200, 400 mM). Assays were done in duplicate. A dose response

curve was generated for the wild-type CBH2 by fitting the data with the Langmuir equation ($y = (x \cdot a) / (x + b)$) and the activities of the CB H2 variants were divided by a calculated activity of wild-type CBH2 of the same plate to yield a performance index.

F. Stability Of CBH2 Variants In Presence Of Ethanol

[217] The stability of wild-type CB H2 and CB H2 variants was tested in the presence of 4.5% ethanol (EtOH) at 49°C. Pooled culture supernatants (80 µL) of *H. jecorina* cells expressing CBH2 variants were added to a 96-well plate (Greiner V-bottom PS) containing 10 µl of 40.5% EtOH per well. The plates were sealed and incubated in a thermostated incubator at 49°C for 16 hours with shaking at 900 rpm. Following incubation, the plates were placed on ice for 5 minutes. Residual CBH2 activity was determined using the phosphoric acid swollen cellulose (PASC) hydrolysis assay as described above. [218] To calculate residual activity, the value of the product formed by the addition of 5, 10, 15 and 20 µl of EtOH-incubated CB H2 to the residual activity PASC assay was divided by the value of the product formed by the addition of 5, 10, 15 and 20 µl of EtOH-free CB H2 to the PASC assay. The individual values of these four ratios were then averaged to give the average residual activity. To determine PI value for the variant, the value of average residual activity for the variants was then divided by the average of the residual activity values of the wild-type CB H2 controls.

G. Thermostability Of CBH2 Variants

[219] The thermostability of wild-type CB H2 and CBH2 variants was tested at 53°C. Pooled culture supernatant (80 µL) of *H. jecorina* cells expressing CB H2 variants were added to a 96-well plate (Greiner V-bottom PS). The plates were sealed and incubated in a thermostatted incubator at 53°C for 16 hours with shaking at 900 rpm. Following incubation, the plates were placed on ice for 5 minutes. Residual CBH2 activity was determined using the phosphoric acid swollen cellulose (PASC) hydrolysis assay as

described above. [220] To calculate residual activity, the value of the product formed by the addition of 5, 10, 15 and 20 µl of heat-treated CBH2 to the residual activity PASC assay was divided by the value of the product formed by the addition of 5, 10, 15 and 20 µl of unheated CB H2 to the PASC assay. The individual values of these four ratios were then averaged to give the average residual activity. To determine PI value for the variant, the value of average residual activity for the variants was then divided by the average of the residual activity values of the wild-type CB H2 controls.

EXAMPLE 2

Generation of Hypocreajecorina CBH2 Site Evaluation and Combinatorial Libraries

[221] The pTTTpyr-cbh2 plasmid containing the Hypocrea jecorina CB H2 protein encoding sequence (SEQ ID NO:1) was sent to BASEClear (Leiden, The Netherlands), Gene Art AG (Regensburg, Germany), and Sloning BioTechnology GmbH (Puchheim, Germany) for the generation of Site Evaluation Libraries (SELs). A map of pTTTpyr-cbh2 is provided in Figure 3. A request was made to the vendors for the generation of positional libraries at each of the sites in Hypocrea jecorina CBH2 mature protein (SEQ ID NO:3) shown in Table 2-1. The amino acid sequence of CB H2 full length protein is shown in SEQ ID NO:2. [222] SEQ ID NO : 1 sets forth the reference Hypocrea

jecorina	CBH2	coding	DNA	sequence:
				atgattgtcggcattctcaccacgctggctacgctggccacactcgagctagtgtgcct
				ctagaggagcggcaagcttgctcaagcgt
				ctggggccaatgtggtggccagaattggtcgggtccgacttgctgtgcttcggaagcac
				atcgtctactccaacgactattactccc
				agtgtcttcccggcgctgcaagctcaagctcgtccacgcgcgccgcgtcgacgactctc
				gagtatcccccaacaatcccggtcga
				gctccgcgacgcctccacctggttctactactaccagagtacctccagtcggatcgggaa

ccgctacgtattcaggcaaccctttgttg
gggtcactccttgggccaatgcatattacgcctctgaagtagcagcctcgctattccta
gcttgactggagccatggccactgctgcag
cagctgtcgcaaaggttccctctttatgtggctagatactcttgacaagacccctctca
tggagcaaaccttggccgacatccgcaccg
ccaacaagaatggcggttaactatgccggacagtttgggtgatgacttgccggatcgcg
attgcgctgcccttgcctcgaatggcgaa
tactctattgccgatgggtggcgtcgccaaatataagaactatatcgacaccattcgtcaa
attgtcgtggaatattccgatatccggaccct
cctggttattgagcctgactcttggccaacctggtgaccaacctcgggtactccaaagt
tgccaatgctcagtcagcctaccttgagtgc
atcaactacgccgtcacacagctgaaccttccaaatgttgcgatgtatttggacgctggc
catgcaggatggcttggctggccggcaaa
ccaagacccggccgctcagctatttgcaaagtgttacaagaatgcatcgtctccgagagc
tcttcgaggattggcaaccaatgtcgcca
actacaacgggtggaacattaccagcccccatcgtacacgcaaggcaacgctgtctaca
acgagaagctgtacatccacgctattg
gaccttcttggcaatcacggctggccaacgccttcttcatcactgatcaaggctgat
cgggaaagcagcctaccggacagcaaca
gtggggagactgggtgcaatgtgatcggcaccggatttggattcgccatccgcaaacac
tggggactcgttgctggattcgtttgtctg
ggtaagccaggcggcgagtgtagcggcaccagcgacagcagtgcgccacgatttgactc
ccactgtgcgctcccagatgccttgc
aaccggcgccctcaagctgggtgcttgggtccaagcctacttgtgcagcttctcaciaaacg caaacccatcgttcctgtaa.

[223] SEQ ID NO:2 sets forth the *Hypocrea jecorina* CBH2 full length protein sequence:

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYS
NDY YSQCL
PGAASSSSSTRAASTTSRVSP TTSRSSSATPPPGSTTTRVPPVGS GTATYSGNPFVG
VTP WA
NAYYASEVSSLAIPSLTGAMATAAAVAKVPSFMWLDTLDKTPLMEQTLADIRT
ANKNGG NY
AGQFVVYDL PDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYS DIRTLL
VIEPDS LA
NLVTNLGTPKCANAQSAYLECYAVTQLNLPNVAMYLDAGHAGWLGW PANQ
DPAAQLFA NV
YKNASSPRALRGLATNVANYNGWNITSPPSYTQGN AVYNEKLYIHAIG PLLANH
GWSNAF FI
TDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGEC
DGTSDSSA PR FDSHCALPDALQPAPQAGAWFQAYFVOLLTNANPSFL.

[224] SEQ ID NO:3 sets forth the Hypocreajecorina CBH2 mature protein sequence:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAAST
TSRVSP T

RSSSATPPPGSTTTTRVPPVGSGTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTG
AMAT A

AAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDC
AALASNGE Y

IADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLNLVTNLGTPKCANAQSAY
LECI NY

AVTQLNLPNVAMYLDAGHAGWLGWLPANQDPAAQLFANVYKNASSPRALRGLA
TNVANYNG WN
ITSPPSYTQGNNAVYNEKLYIHAIGPLLANHGWSNAFFITDQGRSGKQPTGQQQW
GDWCNV IG
TGFGIRPSANTGDSLLDSFVWVKPGGECGTSDSSAPRFDSHCALPDALQPAPQ
AGAWFQAYFVQLLTNANPSFL.

[225] For each of the 162 sites listed in Table 2-1, an average of 18 substitution variants were generated. The libraries were received as purified plasmids encoding CBH2 variant proteins.

Table 2-1: Selected Positions of the Mature CBH2 Protein

58

59 [226] Four synthetic CBH2 combinatorial libraries were also produced by Sloning Biotechnology GmbH (Puchheim, Germany) and BaseClear (Leiden, The Netherlands). Tables 2-2 to 2-5 list the substitutions that could be present in members of the synthetic CBH2 combinatorial libraries (numbered according to the CBH2 mature amino acid sequence).

Production of CBH2 Variant Proteins [227] Purified pTTTpyr-cbh2 plasmids (p c b h i, Amp^R, Acetamide^R) containing open reading frames encoding CBH2 variant sequences were obtained from the vendors specified above. Protoplasts of a quad-deleted *H. jecorina* strain (Aegl1, Δegl2, Acbh1 Acbh2) were transformed with the pTTTpyrG constructs and grown on selective agar containing acetamide at 28°C for 7 days as described (WO 2009/048488). Genes encoding cellobiohydrolase I (CBHI, Cel7a), cellobiohydrolase II (CBHII, Cel6a), endoglucanase I (EGI, Cel7b), and endoglucanase II (EGII, Cel5a) have been inactivated in the quad-deleted strain. Spores were harvested, replated on acetamide agar, and incubated at 28°C for 7 days. In

addition, spores were harvested in 15% glycerol and stored at -20°C for further use. For CB H2 variant protein production, a volume of 10 μl spore suspension was added to 200 μl glycine minimal medium supplemented with 2% glucose/sophorose mixture in a PVDF filter plate: 6.0 g/L glycine, 4.7 g/L $(\text{NH}_4)_2\text{SO}_4$; 5.0 g/L KH_2PO_4 ; 1.0 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 33.0 g/L PIPPS; pH 5.5; with post sterile addition of -2% glucose/sophorose mixture as the carbon source, 10 ml/L of 100g/L of CaCl_2 , 2.5 ml/L of T. reesei trace elements (400X): 175g/L Citric acid anhydrous; 200g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$; 16g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$; 3.2 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$; 1.4 g/L $\text{MnSO}_4 \cdot \text{H}_2\text{O}$; 0.8 g/L H_3BO_3 . Each CB H2 variant was grown in quadruplicate. After sealing the plate with an oxygen permeable membrane, the plates were incubated at 28°C for 6 days, while shaking at 220 rpm. Supernatant was harvested by transferring the culture medium to a microtiter plate under low pressure and tested for residual glucose using the hexokinase assay as described in Example 1.

EXAMPLE 3

Expression, Activity and Stability of CBH2 Variants

[228] H. jecorina CBH2 SEL and combinatorial variant proteins were tested various properties of interest. In particular, the cellulase variants were tested for protein expression using the HPLC assay (HPLC), specific activity using the PASC hydrolysis assay (Act. PASC) and the PCS hydrolysis assay (Act. PCS), stability in the presence of ethanol (EtOH ratio) and thermostability (heat ratio) as described in Example 1. Combinatorial variants were also tested for specific activity by hydrolysis of ammonia pretreated corncob (Sp. Act. CC) as described in Example 1. Performance data for CB H2 SEL variants are shown in Table 3-1, and performance data for the CB H2 combinatorial variants are shown in Table 3-2. Rows of Table 3-2 lacking performance data correspond to CBH2 combinatorial variants that were not expressed in initial tests.

[229] Performance index (Pi or PI) is the ratio of performance of the variant to the parent or reference cellulase. Various terms set forth below are used to describe the mutation: up mutations have a $P_i > 1$; neutral mutations have a $P_i > 0.5$, non-deleterious mutations have a $P_i > 0.05$; deleterious mutations have a $P_i = 0.05$; combinable mutations are those mutations for which the variant has PI values = 0.5 for at least one property, and >0.05 for all properties. Combinable mutations are mutations that can be combined to deliver proteins with appropriate PIs for one or more desired properties. Positions at which mutations occur are classed as follows: Non-restrictive positions have $> 20\%$ neutral mutations for at least one property; and Restrictive positions have $< 20\%$ neutral mutations for activity and stability. [230] This data may be used to engineer any CBH. Even if the CBH to be engineered has an amino acid different from that of CB H2 at a particular position, this data may be used to find a substitution that will alter the desired properties by identifying the best choice substitution, including substitution to the CB H2 wild type amino acid [231] Table 3-1 shows performance index values (Pi or PI) for 2,828 variants of *Hypocrea jecorina* CBH2 at 162 positions. Performance indices less than or equal to 0.05 were fixed to 0.05 and indicated in bold italics in the table.

[232] Non Combinable variants are those for which all P_i values are < 0.05 . For *Hypocrea jecorina* CBH2, of the 2,828 variants, 6 are non-combinable. They are A121Q, L243H, A322Q, L328C, L328Y, V436Y. Any CBH2 which has one of the above substitutions relative to *Hypocrea jecorina* CBH2 can be improved by mutating that amino acid to one of the combinable substitutions at that position, or to the amino acid present in *Hypocrea jecorina* CBH2 at position 121, 243, 322, 328, or 463.

EXAMPLE 4 Classification of CBH2 Positions

[233] As described throughout, functionality of cellulase variants was quantified as a performance index (PI), which is the ratio of performance of a variant to a parent or reference cellulase. Various terms set forth below are used to describe the mutation: up mutations have a $PI > 1$; neutral mutations have a $PI > 0.5$, non-deleterious mutations have a $PI > 0.05$; deleterious mutations have a $PI = 0.05$; combinable mutations are those mutations for which the variant has Performance index values = 0.5 for at least one property, and >0.05 for all properties. Combinable mutations are mutations that can be combined to deliver proteins with appropriate performance indices for one or more desired properties. Positions at which mutations occur are classed as follows: Non-restrictive positions have $> 20\%$ neutral mutations for at least one property; and Restrictive positions have $< 20\%$ neutral mutations for activity and stability.

[234] As shown in Table 4-1, for the 162 positions tested in CB H2, all 162 are Non-restrictive. Non-restrictive positions are the positions that are most suitable for use in constructing combinatorial libraries, since they have a large number of combinable mutations.

EXAMPLE 5 Effect of Charge Change on the Activity of CBH2 Variants

[236] In this example, the effect of charge change on the activity of CBH2 in PCS and PASC assays was assessed. Briefly, the number of PCS and PASC winners in the CBH2 SELs was determined as a property of net charge change. In Tables 5-1 and 5-2, the ratio of observed to expected (o/e) winners was determined in PCS and PASC assays respectively. Values in bold italics are significantly different from the average of 10 random distributions plus or minus the number of standard deviations (sd) listed in the respective columns.

[237] As shown in Table 5-1 and Figure 2A, decreasing charge (e.g., -1, -2) results in a significantly higher frequency of CB H2 winners in the PCS assay, while increasing

charge (e.g., +1) results in a significantly lower frequency of CBH2 winners in the PCS assay. In contrast, as shown in Table 5-2 and Figure 2B, a significantly higher frequency of CBH2 winners is observed for variants without a charge change (e.g., 0), while a significantly lower frequency of CBH2 winners is observed for charge changes (e.g., -2, -1, +2). [238] In conclusion, CBH2 activity on PCS correlates with decrease in charge. CBH2 activity on PASC however, shows that this trend does not correlate with cellulose hydrolyzing activity, but is specific for PCS material.

EXAMPLE 6 Creation of *Trichoderma reesei* Screening Strains [239] Improved screening strains were created to increase the consistency of CB H2 variant expression in the presence of factors unrelated to the amino acid sequences of the enzyme variants. In particular, *T. reesei* screening strains were developed in combination with a targeting vector to force integration of *cbh2* variant genes (e.g., coding region in operable combination with a regulatory sequence). The new strains prepared during development of the present disclosure, combine several mutations that are advantageous for screening variant libraries. A schematic of the genetic engineering steps is shown in Figure 4. [240] Deletion of *ku80* from the *T. reesei* quad deleted derivative strain. A single orthologue of MUS52, the *N. crassa* orthologue of the human KU80, was identified by TBLASTN search in the genome sequence of *H. jecorina* QM6a (*Trichoderma reesei*) and was consequently named *T.reesei ku80*. protein id 58213; <http://genome.jgi-psf.org/Trire2/Trire2.home.html> The nucleotide sequence of the *T.reesei ku80* gene is provided as SEQ ID NO:23 :

```
ATGGCGGACAAGGAAGCAACCGTCTTCATCATCGACCTCGGCGCGTCCATGG
CA
GCTGTCAATGGGGGTCGAGAAGAATCCGACCTTGATTGGAGCATGAGCTACG
TC
```

TGGGACAAGATCAGCAACGTCGTGGCCTCGAATCGCAAGACGCTGTGCGTT
GGC
GTCGTGGGGTTCAGAACCGACGAGACAAACCACACGCTGAGCGAGGATGGG
TAC
GAGAACATCTCCATATTGCAGCCCCTGGGGCCGATGAGCATGTCCAGCCTCA
AG
GCTCTTCAGCCCAAGGTGAAGCCGAGCAGGACGGTGGAAGGCGATGCCATC
TCG
GCGATTGTCATTGCCGTCGACATGATTGACAAGTACACGAAGAAGAACAAAT
GG
AAGCGGCAGATTGTTCTCATTACCGACGGCCAAGGCGAGATTGATCCAGATG
AT
ATTGGCGACATTGCTAGAAAGATGCGCGACTCGAATATTGAATTGACAGTCTT
GT
GAGTTGGCGAGACCGTTTGGCGGACGGTAATGGTGCTGACGGTGATGCAAG
GGG
CGTCGACTTTGATGCTCCCGATTACGGCTTCAAAGAGGAGGACAAACCTTCA
GTC
AAGGTACTCCATATGTTCACTTCTTTTCTTTTCTTCTTTATTTTCTTTTCTTT
GAA
GCTTTCATTAACCTCTTCGTTAGAAGCAAAACGAAGAGACCCTAAAAAAGCT
CGT
GGATGGCTGTGGCGACGACTCAAGGTTGCCTCCATGGTCGAGGCCATTGAC
GA
CTTGAATGAGCCACGAGCAAAGTCGGTCAAGCCTTACAAAACGTACGAAGG

TCT

CTTGACCTTGGGAGATCCGAAAAACGCTCCCGCAGTGGTGGAAATCCGCGTC

GA

GAGATACTTCAAGACCCATCTAGCCAGGCCACCTGCCGCCAGCACCGTGGTG

GTC

AAGGAGGAGCAAGCTGGGCCGTCTCAGGCAGACGAGGACGAACAGATGGA

CGG

AGCGGAACTTACAGCTGTGAGGCAGGCCAGGACATACAAGGTCAATGATCC

AGA

TGCCCCTGGCGGTAAGCGTGACGTTGAGTTTGAGTCTCTGGCCAAAGGGTAC

GAG

TACGGCAGGACGGCAGTCCACATCAGCGAGTCTGATCAAAACGTCACCAAG

CTC

GCGACAGAAAAGAGCTTCAAGATCATCGGCTTCGTCCAGAAAGAAAAGGTA

TTG

GCTTGGCTCTCAGCATTTGACCCGTTGCTCTTGGCTAACCCTTGTTTAGTATG

AAA

TGCTCCTTAATCTTGGCGAAACCTGCGTTACCGTTGCATCCAAGTACGATGAA

AA

GTCTGAGCTGGCTTTTAGCTCTCTGGTGTGGGCGCTCTCGGAGCTCGACGCC

TAC

GCCGTGGCCCGCCTAGTAACTAAGGACCAAAAGGACCCCATGCTGGTGTTAC

TG

ATGCCGTATATGGAGCCTGATTATGTTTGTCTCTATGATGTGCCTCTGCCTTTC

GC

AGAGGACATCAGGACGTACCAGTTTCCTCCCTTGGACAGAGTCGTTACCGTC
AGT
GGCCAAACGCTCACCAACCATCGCCTATTGCCATCCGACGAGCTCAACCAAG
CG
ATGAGCGACTACGTAGATGCCATGGACATTTCAAGTTATGGTATCGATGAAGA
TG
GGTGAGTATAGAAGATGATTGTTCAAATCTTCACTTCTAAGCATTGCTTCTGA
TC
TAGGCAACCGGCTGAATATGCCACCATCGATGAGTTATAACAACCCTGCGATAC
AT
CGCATAGGCCATGCGATCAAACAACGAGCGATCCACCCAGAGAAACCCGTG
CCC
GAGATCCCCCAGTCTTGCTTAGATTTCGCAGCACCCCCGACAGAACTCGTCG
AGA
CTGTGCAGCCTCATATCGATGCACTGATTCACGCTGCAGACGTGAAGAAAGG
TAC
TGATTCCATTACATATGCTTCTCTGCACACTGATGTTTGATTTGTGCTAACGCC
CC
CCTTAGTGCCGCCCAAGGCCAAGGGCAAGCGCCAAAGAGAAACAGTTAAAC
CCA
TCTCGGGACTGGATGTGGATGCCCTTCTGGGAGAAGAGCAGAAAGGTTCCAT
TA
GTCCGGAGAATGCCATTCCGGACTTCAAACGAGCCCTCAACTCGTCCGAAGA
AG
TCGAGCAGATTGCCGACGCCACAAAACAAATGGGGGCCATTGTGCGGTCTCT

CA

TTACGGACAGCTTCGGGGATAGCAAATATGCCCAGGCAATGGAAGGCATTGG

TG

CGATGCGTGAGGAGCTGATCAACCTGGAAGAGCCTGGCCTGTACAACGACTT

TG

TGCGCGACTTGAAGAAAAGTTTGCTATCTGGAGCCTTGGGTGGTGACAGGCG

AG

ATTTCTGGTTCAAGATGAGGTGGGCGAAGCTGGGCCTGATTGACAAGAAAC

AGT

CGGAGGTGTCTTCGGTCACTCTTGAGGAGGCGGACGAGGTGAGTGGTGACAG

CAT

GCTGTCGGATTATACGGACGTTGTTTGCTAACTTGTGGGATAGTTTTACAAGT

CG

AGGTGAGGTATCTACGTTGACCAAGAATGGGACCATGTATATGAGCGGTGTA

AC AACAGAATCCTGTGCTTTGAGCATTGTATGA

[241] The *T. reesei* ku80 gene was deleted from the quad deleted derivative strain using standard methods of the art (WO 2005/001036). Briefly, a ku80 deletion cassette was utilized that employed a selectable marker flanked between 1.3 kb of 5' ku80 sequence and 2.3 kb of 3' ku80 sequence, as schematically shown in figure 5. The variant *T. reesei* ah, which confers resistance to the herbicide chlorimuron ethyl, was used as selectable marker as described in WO 2008/039370. The nucleotide sequence of the ku80 knockout cassette is 7685 base pairs in length: bases 1-1271 correspond to the 5' ku80 homologous region; bases 1280-7685 correspond to the α /s-chlorimuron ethyl resistant variant (A190D); and bases 5381-7685 correspond to the 3' ku80 homologous region. The nucleotide sequence of the ku80 knockout cassette is provided as SEQ ID NO: 14:

ggccgcctcaacacccacactcgaggcacacgagttcatcggcggcttccccacaagct
ctcggccaacctgctaccggctctctc
gcgagacttcccaaagcctacaaacgaggtcgacgtcaaggaggccctcgagcgccagcc
cggcagatggagcctccagggccca
gatcaaggccaacaacatgagagcccagagcgccgcactccggctcgacgacaaggaggg
caaggcgagagcctttgaggagg
ccaagcgcgagctactggcgtatcaccacagcgccctgcggaagccttccggcgcaagat
aatgagcttgatcgcaatgacgagttc acgtacgctttgccatatgtgtgtgctttttgttggctctacatgtacggcgcatgg
ttgggaggatataccacggagagtgtccgagt
ggcttctgggatttagagcgtcattagcaggatagagatggttggccaggggaatggaat
tgacttttactacaaggaacttgttcactc
tggtgtigattccattgcgtgactggtagtagggaggaatgcttttactttgtgccact
agaccgcagagaagggttggtgcaagcgg
ggctccgtgtataccgaccaagagtgatgggcatacagcaacgtttctgaacgacttcatt
ttgtccgagtctactggatgcgagatgccca
gcgtgaagccgtacgccaccaggggcgacgaactcgacaagggtgacgaggaggagatgc
cgtgcagcatgccaacttctgttg agggcacgcatctcatccgactgtgcatccttgcataccactcctttccgtctcgttg
gctggtgggagggttcaacaaatccatcgtc
agccatccgggggtctcaaatcaatggcgtgcatgcggagtcgggcttgaggctaaccttg
tccatggcggtccttcatggtcttgacag
tggcgggaagcagcacggcgaggttgacgaggccgctgacgaacatggttgcgatgggca
ccaaggagctccacttgttgggagc
gtcgacgaggccgccgatccgcccttgatgccaagagggcgttccgggggaacgtgag
ggcgagcagcgcggggatggccgt
ctgcatgccaaagtagatggggaacagcttgctctggatggcggagaaggaggggccggct

gacggtgcggaacatgacgatgccg
ttgacgaaggactgcagtagcgtagtgtgatggttaagcagctggccggcgcgcctgagac
aatggccggcaatggtaaaaaggacc
aagatgtactaggtagtgtgcaatgtggcttattacctactacctggtaggcaccta
ctaggtacttgggtagacggacaatgaaattt
gaagtcgggggtgcaggaaa gcaggcgctggacacattgtgcttcaggcgggtaccgctc
gtcatcgtcagccaatgtcgaggccc
ggcagcccgaggagcgcagacaacctggccggaggagcccgcaggtacctgccaaagcgc
ggctgggtacctctcaacctctcag
gcctgttggatgccctatgacatgccctgggggatgcagctgttccccggccccgcaact
ttcgggtgaccgcgaggctgctgattgg
ctggttgccacgggctggcggtccctgaagtgttgccatctgaactctgtcggcgctg
gcgtcggctgcgcccgaatgggaggcga
gacaactcagggtactagaatcactgacagaagaagagaatcgaagtaggtagacagcc
aattcgttgcatggcaggcaaccgca
caggagaaaaattgactacccacaaatcaggcacagtaagtagggcacagtacgtatgta
cagacaaggcgcgaagcgatactgcgc
gacccgggtacctcgccggttgacacgtgcgacaggctactttactagtattcgagcgg
cgggtcgcgcattattacatgtactgtgc
cgccatttgatgactgggctgctgcagtattagtagatctgccggcatcgccttccat
gggcgcgacccgggactggaccctctga
ctctacctacatgtacctaggccggcgccgggcttggtgactttgtccgatcaggtcgtt
cgctggctacctattatttcttcttctctc
catcctgcttctggccttgcaattcttctgccactectccctcttccccgcgatac
ccttgaattcgtcagagaggaaaagacgaga

aaaaaaagggcagcagagacgtcggctcggctcacgtgctgcatctctgcgcactctcat
ttttttatgtccgacccctccctcaacctt ctccttcgttgacaggctaagccttgcttcgacgctctctcttgaattttctacttct
accttcttttcttgcgtgttaccaccatagctcgat
tcacgatgctccgaagtcgccaagtcacagccagggccgtccgggctctgggccaggcgc
gcgcctttacctcgacgaccaagcct
gtcatgatccagagcagccagaggaaacaggccaacgccagcgctgctccgtaagtcgcc
cattgccattgcatcttctgtttgatata
tacttcctgctgcttgcgtggcgtcgtctcggttatgcgtgtcaaggaccagggtgtgt
tcgcatcgtggttttccagcgccgattaccgg
gggacgaattttggctgctcaactcgcgcgcgcgcatctgattcttcgtttcaatct
tgagcgacaactggctaacataatggccattg
gcaattgcttcacacagacaagtccgcccgtaccgagccctgcttcaacgctgaagac
aaagaccgcagccatgtgcagcctctgg
tcaaccgctcgaagcccgcacatggatgaatcgtatgtccacgtcccctcgtcccgcccta
caaatgaacacgattacaccagaattttt
gcaacaatcgacacttctataacagaccaattgagctttgttctgaccaatcatgttgct
ctagattcattggcaaaaccggaggcgaaat
cttcacgagatgatgctgcgacagggtgtaagcacattgtaggttccgatgccggcc
gcccacacgggctccatccttgctccatc
tctccagctaggcaaatctcgtaaccttgagtcaccatccagtcggataccctggcggc
gctatcctgcccgtcttcgacgccatctac
aactcaaaacacttcgacttcacctgccccgtcatgagcaggagctggccatatggcc
gagggtatgcccgtgcctcgggcaaa
cccgggtgctgctcgtggtgacttccggccccgggtgctaccaatgtcatcacgcccatgcag
gatgccctgtcggacggaacgcccttgg

tcgtcttctgcggccagggtccccaccacggccatcggcagcgatgacttccaagaggccg
acgtcgtgggcatctcgcgggcctgc
accaagtggaacgtcatggtcaagagcgttgctgagctgccgcggagaatcaacgaggcc
tttgagattgccaccagcggccgcc
tggccccgtcctcgtcgacctgccaaggatgtcacggctggtatcctgaggagagccat
ccctacggagactgctctgccgtctctg
cccagtgccgcctcccgccgcccatggagctgagctccaagcagctcaacgcctccatc
aagegtgccgccgacctcatcaacat
cgccaagaagcccgtcatctacgccggtcaggggtgcatccagtcgagggcggcgttga
gctcctgaagcagctggcggacaag
gccctcatccccgtcaccaccacctccatggcctgggtgccttgatgagctggacgag
aagtcgctgcacatgctgggcatgcacg
gctcggcgtatgccaacatggccatgcagcaggccgacctcatcatcgccctcggcagcc
gattcgacgaccgtgttactctgaatgt
ctccaaatttgcgcctgcagccaggcaagctgctgccgagggccgcggcggcatcattca
ctttgagatcatgccaagaacatcaa
caagggtcatccaggcgaccgaggccgtcgagggcgacgtcgccaccaacctgaagcacct
cattccccagattgccgaaaagtcca
tggcggaccgaggagagtggttcggcctcatcaatgagtggaagaagaagtggcccctgt
caaaactaccagcgcgcggagcgggc
tggcctcatcaagccgcagacgggtcatggaggagattagcaacctgacggccaaccgaaa
ggacaagacgtacattgccacgggt
gtcggccagcaccagatgtgggttggccagcacttccgctggaggcaccctcgatccatg
attacctctggtggtctgggcaccatgg
gctacgggtctccccgcggccattggcgccaagggtggcccagcccacgctctcgtaatg

acgttgatggc gatgcctcgtttaacat
gacgctgacggagctgtcgactgctgcacagtcaacattggcgtcaagggtggttgct
caacaacgaggagcagggcatggtgac
gcagtggcagaacctcttttacgaggaccgatatgccacacgcaccagaagaaccccgga
cttcatgaagctggccgacgccatgg
gcgttcagcaccagcgcgtgacggagccggagaagctggtcgatgccctgacgtggctga
tcaacaccgatggcccggccctgttg
gagggtgtcacggacaagaagggtgcctgtcctgcccatggtgcccggcgatcgccctg
cacgagttcctcgtctttgaacctgggtg
agtctacttcagacatattgcttgcgcattgcagatactaactctcacagaaaaggat
aagcagcggcgtgagctgatgaaggagag
aacaaaggggtgtgcactcctaaagc gatgatgtctgcgaggggttcttcgttgaacccta
gttcaggcaccatcttacccctcttattttcc
cgtgggctttcattttgtgtcatccgagcatgacgtttaggggttgagtttcttcttt
ttatcttgcatttactggtacccataggcgcgag
actaggcttccatgttttgtttgcgactttcaaaaagtacttttagtggttggggcac
gacgaggggggggcaacctcttctgtcgaaaaa
gggtggctggatggatgagatgagatgagatgaggggtgaagatagatacctgcagtgttt
tgacgcgacgggatggcgatcgagca
ccccgcagaaactcgtcgagactgtgcagcctcataatcgatgcactgattcacgctgca
gacgtgaagaaaggctactgattccattac
atatgcttctcgcacactgatgttgatttgtgtaacgcccccttagtgccgccaa
ggccaagggcaagcgccaaagagaacag
ttaaacccatctcgggactggatgtggatgcccttctgggagaagagcagaaagggtcca
ttagtccggagaatgccattccggacttc

aaacgagccctcaactcgtccgaagaagtcgagcagattgccgacgccacaaaacaaatg
ggggccattgtgcggctctcattacg
gacagcttcggggatagcaaatatgccaggaatggaaggcattggtgcgatgcgtgag
gagctgatcaacctggaagagcctgg
cctgtacaacgacttltgtcgcgacttgaagaaaagtttctatctggagccttgggtgg
tgacaggcgagatttctggttcaagatgag
gtgggcgaagctgggcctgattgacaagaaacagtcggaggtgtcttcggtcactcttga
ggaggcggacgaggtgagtggtgcag
catgctgtcggattatacggacgttgttgctaacttggggatagtttacaagtcgag
gtgaggtatctacgttgaccaagaatgggac
catgtatatgagcgggtgaacaacagaatcctgtgctttgagcattgtatgatatgatta
ttgatgaaccggacaaaaggggtagggga
ttgatgccatcacgaccgattgaccagacctggattctcgacagcatggctgctgattt
tgttgaccttgcgacgtaacatccctgaaga
acaacctactattaacctatcatttagcagaagctctgtaaccttcttgattctgtatt
cagcttctgagtctgtcaaatgtaatcatttcgagg
ttgtgtaattccggccaagcaggcggccgtctgccagcgcctgcctaggctgcaccgcaa
tctgcccaatcagctgcccttcagtttcg tttgaccttgcagctgcccttcacctttatctgcacacaattcttttctctgctctg
cgcattcttctctctctctctctccttctcaagctcaa
cttcacctcatccgctccactacaagccctcccgtcgtcgtctcgcacacctcatctcgac
tgcggccagcaaaacaagcaaagccgtga
tcgatcctcagcatggctaccttcaacctcaccgtccgcctggagatgctcaaagaaatt
ggaatcacctccaatacggcgagcatgt
agcgaaagaagcagccagcaacgaagcagcgatggcattcgaagaagaagaagagttccc
cgccgttgtgccgccaaggcaga

acagcacgcctctgaacacgacgctggccacgatgcttgggacgcggctgccacatctc
 gacttcggcgcaagaacagcagaag
 cccagggagatggacgactcgtctatcgtgatgccgctggactactccaagttgtcgtt
 ggagagcctgcggacgaatccatcagctt
 ttgctcgtggaaggctcgtcaggcttatcctgaccagtttatcggcaaggcaaacaggcc
 tcgtgtatgtagcgattgctttctctgcatta
 tgggaatctcaagagagtatggtagaagataactgacaacttgcaaggccaagccgtactt
 tgacaagattttggaagacagagtctgg
 gatttgtgaggatcttgattgatgtgcatatggcgacatgcctgctaataatcattgtagc
 ttctatctctacaacccccgagaagccttcaga
 gaagcctcgcgtgctgggtgccactgttcagctcgaaggctttctcaaaagcatcaacag
 agcgctcgggtacttctctcaccattccag
 gagggggcaaaccaggaccgttttctgaggttcggccaggagacaccccaaggcctc
 gatatctacagaggtcgagagaccag
 aaatccctaaagattgaaacgttccccgatittcaacaggcggactacgacagcttagg
 aacgcgcatggcgccatccaggaggact
 ggttgaagaactggcagatgctggtacctcgccgagtttcgacaagaagaaaaatgcag
 acaaaagagcagccaagagaaggct
 cgagcgagagcgaatgcttcacaatacgcaggaatttctcatttggcaggttaaggggcaa aggggctgacgtgg.

[242] Creation of the Archy2 strain from the *T. reesei* Δ ku80 quad deleted derivative strain. The *pyr2* gene was deleted from the ku80 knockout strain. The *pyr2* deletion cassette contains the *T. reesei* cbhl promoter, a hygromycin resistance gene and a partial *amdS* selectable marker flanked by 5' and 3' *pyr2* sequences, schematically shown in figure 6. Use of this vector permits screening for resistance to hygromycin and fluoroorotic acid of *pyr2* knockout transformants. The partial *amdS* gene contains the 3'

portion of the gene, but lacks a promoter and the amino-terminal portion of the coding region, and is consequently nonfunctional. The nucleotide sequence of the pyr2 knockout cassette is 9259 base pairs in length: bases 1-1994 correspond to the pyr2 3' homologous region; bases 2002-3497 correspond to the *T. reesei* cbhl promoter; bases 3563-5449 correspond to the hygromycin resistance selectable marker; bases 5450-7440 correspond to the *A. nidulans* amdS 3' partial marker, and bases 7441-9259 correspond to the pyr2 5' homologous region. The nucleotide sequence of the pyr2 knockout cassette is provided as SEQ ID NO: 15:

```
atcacgccctcgcataaaagaccctcaagagtcctatgtgccctatctgcctgactctct
aaccttatttaacattggccctatcacacc
tagttcttctcagcctgcttgtcaacactgtcacggttcaactcaacgtaacagca
ggtagcaggacgaggatagggagagaaacg
aagagaagaaggagagaggaagaaaaaaaagaaaagaaaagggaaggaa
agaaggaggaaaagagaagaa
agtcagatgaagaagcaagaagacgccatggtagccaccgctgtcagggctccttagca
acgaacaactctagcttggggacttgtc
gatgtgtcgtttccttctaccatcagcaccaacgatgagttcgatatagacgaggacc
tcattggaagtagagaccattgggttcgaca
ggatctctcagtttcacttctatgaggtctgtcgtcggatgacttttgaggagcttcc
ccttctgcttcaaccccaactctcttctgaa
accgcagcacgttggcacggccgtgttgctggagcagtttgcttcgagcactctcagcg
tggttcagcagccactggtgagtggc
ctcctttgacgtccacaccttgctcctgtcgcgcgtatctggagggaacgactgctcc
aaggaggattgctaacgaggtttagggccg
aatacgcacagattctccggaaccttagctacggcctcttcaacatctgtgacatga
```

cggagcgcaagtactggtggtggcgacca
agatgcgcggctggaacatcgacggctgccccgaagacgtcaggagactcatgtttgttc
acatcatcgccacctgggatgcagcc
ccgtcgtgacggatgaagacatggactacccaagaactggcggaattctccacggtg
gagacagatatccgagtgaacctgtgg
gccaccggcctcatggcgccacctctgcctccactcgggtggccgtctgcctcgtctcc
agggcttgggtctcgggtactgcgactct
gaagtcgtatgtgcagcgcatgaacagcctcggcgccgcggaccgtgttgctctcgtttg
ccgcaagcccagacgagatttttgaa
agatgcggcttcaggaacagcggccggagtagtatcaagactctggtcggcgaatactac
aacatggtgtgtgctccacatcgactt ggccagactctatacgattttcaaacctcgctatacgatattgacttggttcttagg
tcttcgatttggccgggcccgaagactttatcga
ctggaatagcattgccgacgctgccaagaagatgtgaaccatttgactgatacgatgtgt
gctacgcatgtcgaccttctttgtttgttctt
tggcggctctttgtataccttgggacacggcagacgcatgtctatgtgaagaaaacgttc
acggcgctgtttgcatcaggaatatgatca
ttaaacatggagcgtaattggtattaatgatcaactagaaaaatggatgggaaggcgaga
ggcgcatcaacaaagcagccccggggc
atagtctggaagcagcaggaattggaagggaaggaagctgcacaatgaagggatatcg
tgagcggagtggctcacgagagtatc
aacagactggcgaaagcaagcaattgccaacgccggctattaggccataagatggcctgt
tgtgagtcccagttgcacgtatcccat
atgactgctctgtcgtgacttgaaaaaaataggaggataaaggagaaagaaagtgag
acaacccgtgagggacttgggtagta
ggagaacacatgggcaaccgggcaatacacgcgatgtgagacgagttcaacggcgaatgg

aaaatcttgaaaaacaaaataaaata
actgccctccatacgggtatcaaattcaagcagttgtacggaggctagctagagttgtga
agtcggtaatcccgctgtatagtaatacga
gtcgcatctaaatactccgaagctgctgcgaacccggagaatcgagatgtgctggaaagc
ttctagcgagcggctaaattagcatgaa
aggctatgagaaattctggagacggcttgttgaatcatggcgttcattcttcgacaagc
aaagcgttccgtcgcagtagcaggcactc
attcccgaaaaaactcggagattcctaagtagcgatggaaccggaataataataggca
atacattgagttgcctcgacggttgaat
gcaggggtactgagcttggacataactgttccgtacccacctcttctcaacctttggcg
tttcctgattcagcgtagccgtacaagtcg
taatcactattaaccagactgaccggacgtgtttgcccttcatttggaataaatgt
cattgcgatgtgtaattgcctgcttgaccgac
tggggctgttcgaagcccgaatgtaggattgttatccgaactctgctcgtagaggcatgt
tgtgaatctgtgtcgggcaggacacgcctc
gaagggtcacggcaagggaaccaccgatagcagtgcttagtagcaacctgtaaagccgc
aatgcagcatcacttgaaaatacaaa
ccaatggctaaaagtacataagttaatgcctaaagaagtcataataccagcggctaataat
tgtacaatcaagtggtctaaacgtaccgtaa
ttigccaacggcttgtggggttgagaagcaacggcaaagccccacttccccacgtttgt
ttcttcactcagtcgaatctcagctggat
ccccaattgggtcgcttgttgcgggtgaagtgaagaagacagaggaagaatgtc
tgactcggagcgtttgcatacaaccaag
ggcagtgatggaagacagtgaaatgttgacattcaaggagtatttagccagggatgcttg
agtgtatcgtgtaaggagggttgtctgccg

atacgacgaatactgtatagtcacttctgatgaagtggccatattgaaatgtaaagtcg
gcactgaacaggcaaaagattgagttgaaa
ctgcctaagatctcgggccctcgggccttcggcctttgggtgtacatgtttgtgctcgg
gcaaatacaaagtgtgtaggacgaaca
cactgtcgcctttaccaagcagctgagggtatgtgataaggcaaatgttcaggggccactg
catggtttcgaatagaaaagaagcttag
ccaagaacaatagccgataaagatagcctcattaaacggaatgagctagtaggcaaagtc
agcgaatgtgtatatataaagggtcag
gtccgtgcctccctcatgtctccccatctactcatcaactcagatcctccaggagactt
gtacaccatctttgaggcacagaaaccaa
tagtcaaccgcggactgcgcacatgtatcggaaagtggccgtcatctcggccttcttg
ccacacctcgtgctagactaggcgccca
ggaagcccgaaggtgaagtggattctcggcgtggctggagcaaccggtggattccagcg
tctccgacttggactgagcaattcagc
gtcacggattcacgatagacagctcagaccgtccacggctggcggcattattggttaac
ccggaaactcagtctccttggccccgtcc
cgaagggacccgacttaccaggctgggaaagccagggatagaatacactgtacgggcttc
gtacgggaggttcggcgtagggttgtt
cccaagttttacacaccccccaagacagctagcgcacgaaagacgcggagggtttggtga
aaaaagggcgaaaattaagcgggag acgtatttaggtgctagggccggttctctccccattttcttcggttccctttctctct
ggaagactttctctctctctctctctctctctctctctccatc
ctcagtccatcttcttcccatcatccatctcctcactccatctcaactccatcacat
cacaatcgatatgaaaaagcctgaactcaccg
cgacgtctgtcgagaagtcttgatcgaaaagttcgacagcgtctccgacctgatgcagc
tctcggaggggcgaagaatctcgtgctttc

agcttcgatgtaggagggcgtggatatgtcctgcgggtaaataagctgcgccgatggttc
tacaagatcgttatgtttatcggcacttg
catcgccgcgctcccgaftccggaagtgttgacattggggaattcagcgagagcctga
cctattgcatctcccgccgtgcacaggg
tgtcacgttgcaagacctgcctgaaaccgaactgcccgtgttctgcagccggcgcgga
ggccatggatgcgatcgctgcggccga
tcttagccagacgagcgggttcggccattcggaccgaaggaatcggtaatacactac
atggcgtgattcatatgcgcgattgctg
atccccatgtgtatcactggcaaacgtgatggacgacaccgtcagtcgctccgtcgcgc
aggctctcgatgagctgatgcttgggccc
gaggactgccccgaagtccggcacctcgtgcacgcggattcggctccaacaatgtcctg
acggacaatggccgcataacagcgggt
cattgactggagcgaggcgatgttcggggattccaatacaggctcgccaacatcttctt
ctggaggccgtgggtggcttgtatggagc
agcagacgcgctacttcgagcggaggcatccggagcttgaggatcgccgcggctccggg
cgtatatgctccgattggtcttgacc
aactctatcagagcttgggtgacggcaattc gatgatgcagcttgggcgcagggtcgat
gcgacgcaatcgtccgatccggagccgg
gactgtcgggcgtacacaaatcggcgagaagcgcggccgtctggaccgatggctgtgt
agaagtactcgccgatagtggaaacc
gacgccccagcactcgtccgagggcaaaggaatagagtagatgccgaccgggatccactt
aacgttactgaaatcatcaaacagctt
gacgaatctggatataagatcgttggtgtcgatgtcagctccggagttgagacaaatggt
gttcaggatctcgataagatacgttcattg
tccaagcagcaaagagtgccttctagtgtttaaatagctccatgtcaacaagaataaaac

gcgttcgggttacctctccagatacagc
tcctctgcaatgcattaatgcattggacctcgcaaccctagtagcccttcaggctccgg
cgaagcagaagaatagcttagcagagtct
atttcattttcgggagactagcattctgtaaagggcagcaatgccagcagttagta
gggtcccctctacctctcaggagatgtaa
caacgccaccttatgggactatcaagctgacgctggcttctgtgcagacaaactgcgccc
acgagttctccctgacgccgctctcgcg
caggcaagggaactcgatgaatactacgcaaagcacaagagacccgttggtccactccat
ggcctccccatctctctcaaagaccag
cttcgagtcaaggtagaccgttgcccctaagtcgtagatgtcccttttgcagctaac
atatgccaccagggctacgaaacatcaatgg
gctacatctcatggctaaacaagtagcagcgaaggggactcggttctgacaacctgctcc
gcaaagccgggtgccgtctctacgtcaa
gacctctgtcccgagaccctgatggctctgcgagacagtcaacaacatcatcggcgcac
cgtcaaccacgcaacaagaactggtc
gtgcggcggcagttctgggtggtgaggggtgcgacgttgggattcgtgggtggcgtcatcgg
ttaggaacggatacgggtggctcgatt
cgagtccggccgcgttcaacttcctgtacggctaaaggccgagtcagggcggctgccg
tatgcaaagatggcgaacagcatgga
gggtcaggagacgggtgcacagcgttgtcgggccgattacgcactctgttgagggtgagtc
cttcgcctcttcctcttttcctgctctata
ccaggcctccactgtcctcctttctgtttttatactataacgagaccggcagtcact
gatgaagtatgtagacctccgcctctcacca
aatccgtcctcggtcaggagccatggaaatacgactccaaggctatccccatgccctggc
gccagtccgagtcggacattattgcctc

caagatcaagaacggcgggctcaatatcggctactacaacttcgacggcaatgtccttc
acacctcctatcctgcgcggcgtggaa
accaccgtcgcgcactcgccaaagccggtcacaccgtgaccccgaggacgccatacaag
cacgatttcggccacgatctcatctcc
catatctacgggctgacggcagcgccgacgtaatgcgcgatatcagtgcacccggcgag
ccggcgattccaaatatcaaagacct
ctgaacccgaacatcaaagctgttaacatgaacgagctctgggacacgcatctccagaag
tggaattaccagatggagtaccttgaga
aatggcgggaggctgaagaaaaggccgggaaggaaactggacgccatcatcgcccgatta
cgcctaccgctgcggtaggcatga
ccagttccgggtactatgggtatgcctctgtgatcaacctgctggattcacgagcgtggt
tgttccggttaccttgcggataagaacatcg
ataagaagaatgagagttcaaggcggtagtgagcttgatgccctcgtgcaggaagagt
atgatccggaggcgtaccatggggcac
cggttgcagtgaggttatcgacggagactcagtgaagagaggacgttggcgattgcag
aggaagtggggaagttgctgggaaat
gtggtgactccatagctaataagtgtagatagcaattgcacaagaaatcaataccagc
aactgtaaataagcgctgaagtaccatg
ccatgctacgaaagagcagaaaaaacctgccgtagaaccgaagagatatgacacgcttc
catctctcaaaggagaatcccttcag
ggttgcgttccagtagtgattttaccgctgatgaaatgactggactccctcctctgct
cttatacgaaaaattgcctgactctgcaaaggt
tgttgtcttgaagatgatgtccccccatcgctcttatctcataccccgccatcttt
ctagattctcatcttcaacaagaggggcaatcc
atgatctgcgatccagatgtgcttctggcctcactctgccttcagggtgatgttact

taattggtgacgaattcagctgattgctgcagt
atgctttgtgttggttcttccaggcttgtgccagccatgagcgcttgagagcatgttg
tcacttataaactcgagtaacggccacatattg
ttcactacttgaatcacatacctaatttgatagaattgacatgtttaagagctgaggt
agctttaatgcctctgaagtattgtgacacagct
tctcacagagtgagaatgaaaagttggactccccctaatgaagtaaaagttcgtctctg
aacggtgaagagcatagatccggcatcaa
ctacctggctagactacgacgtcaattctgcggcctttgacctttatatatgtccatta
atgcaatagattcttttttttttttttttttttttttttttt
ttttttttgcccatttcgcagatcaaagtgagcgttatagcatcataactaagctca
gttgctgaggggaagccgtctactaccttagccc
atccatccagctccataccttgatacttiagacgtgaagcaattcacactgtacgtctcg
cagctctccttcccgtcttgcttccccactgg
ggtccatggtgcgtgtatcgccccccacaattctatgcatggtacctccagcttacc
aatgccccgctaacaagtcgcctctttgcctt
gatagcttatcgataaaacttttttccgccagaaaggctccgccacagacaagaaaaa
aaattcaccgcctagcctttggccccggc
atttggctaaacctcgagcctctctcccgtcttggggatcaggaagaaaagaaaaaat
ccatcgccaagggtgttttggcatcacca
cccgaacacagcacttcctcgatcaaaagttgcccgcctgaagaccacgtggaaggaca
tccttcgggtgcctacgcaccaggagt
ttctggacattgtgctgagcaggaccagcgcaaactgccactcagatccgtgccggct
tcaagattagcagaattcgaggtacgtc
gcattgcccatcgaggatgtctcattatcggggctccttggaacgatcatgattgcat
ggcgtatgctaacacatagacagccttctac

actcgaaaggtcaagttcacccaggagacgtttccgaaaagttcgctccatcctcgac
 agcttccctcgctccaggacatccaccc
 ctccacaaggaccttctcaacaccctctacgatgccgaccacttcaagattgcccttgg
 ccagatgtccactgccaagcacctggtcg
 agaccatctcgcgcgactacgtccgtctcttgaaatacgcccagtcgctctaccagtgc
 agcagctcaagcggggcgctctcggtcg
 catggccacgctggtcaagcgcctcaaggacccctgctgtacctggaccagggtccgcca
 gcatctcggccgtcttccctccatcgac
 cccaacaccaggaccctgctcatctgcgggtaccccaatgttggcaagtccagcttcctg
 cgaagtatcacccgcgccgatgtggacg
 tccagccctatgctttcaccaccaagagtctgtttgtcggccactttgactacaagtacc
 tgcgattccaggccattgatacccccggtatt
 ctggaccaccctcttgaggagatgaacactatcgaaatgcagaggtatgtggcgcggcta .

[243] Creation of the Archy3 strain from the Archy2 *T. reesei* strain. The Archy 2 strain was transformed with a vector to integrate at the same *pyr2* locus and replace the hygromycin resistance gene with the coding region of the *pyr2* gene. The hygromycin deletion cassette is shown in figure 7. This re-introduction of the *pyr2* gene back into the *pyr2* locus placed it between the *T. reesei* *cbhl* promoter and the partial *amdS* selectable marker. This strain could be selected for uridine prototrophy and sensitivity to hygromycin. The nucleotide sequence of the *hygR* knockout cassette is 9088 base pairs in length: bases 1-1994 correspond to the *pyr2* 3' homologous region; bases 1995-3497 correspond to the *T. reesei* *cbhl* promoter; bases 3564-5137 correspond to the *pyr2* selectable marker; bases 5280-7270 correspond to the *A. nidulans* *amdS* 3' partial marker; bases 7271-9088 correspond to the *pyr2* 5' homologous region. The nucleotide sequence of the *hygR* knockout cassette is provided as SEQ ID NO: 16:

atcacgccctcgcataaaagacctcaagagtccatgtgccctatctgcctgatcttct
aaccctatttaacattggccctatcacaacc
tagttcttctcagcctgctttgtcaacactgtcacggttcaactcaacgtaatcagca
ggtagcaggacgaggatagggagagaaacg
aagagaagaaggagagagggaagaaaaaaaaaagaaaagaaagaaaaaggaaaaggaa
agaaggaggaaaaagagaagaa
agtcagatgaagaagcaagaagacgccatggtagccaccgctgtcagggctccttagca
acgaacaactctagcttggggacttgtc
gatgtgtcgtttccttctacccatcagcaccaacgatgagttcgatatagacgaggacc
tcatggaagtagagaccattgggttcgaca
ggatctctcagtttcacttctatgaggctgtcgtcggatgacttttgaggagcttcc
ccttctgcttcaaccccaaactctcttctgaa
accgcagcacgttggcacggccgtgtgtcggagcagtttgccttcgagcactctcagcg
tggtttcagcagcccactggtgagtggc
ctcctttgacgtccacacctgtcctgtcgcgatcgtatctggtgggaacgactgtcc
aaggaggattgctaacgaggtttagggccg
aatatcgcacagattctccgtaaccttagctacggccttcaacatctgtgacatga
cggagcgcgaagtactggtggttggcgacca
agatgcgcggctggaacatcgacggctgccccgaagacgtcaggagactcatgtttgttc
acatcatcgccacctgggatgcagcc
ccgtcgtgacggatgaagacatggactacccaagaactgggcggcaattctccacggta
gagacagatatccgagtgaacctgtgg
gccaccggcctcatggcgcaccatctgcctccactcgggtggccgtctgccctcgtctcc
agggcttgggtctcgggtactgcgactct
gaagtcgtatgtgcagcgcataaacagcctcggcgccgggaccgtgtgtctctcgtttg

ccgcaagcccgagacgagatTTTTgaa
agatgcggcttcaggaacagcggccggagtagtatcaagactctggcggcgaatactac
aacatgggtgtgtcctccacatcgactt ggccagactctatacgaTTTcaaacctcgtatacgtcatattgacttgttctttagg
tcttcgatttccccggggcccaaagactttatcga
ctggaatagcattgccgacgctgccaagaagatgtgaaccattgactgatacgaTgtgt
gctacgcatgtcgaccttctttgtttgttctt
tggcggctctttgtataccttgggacacggcagacgcatgtctatgtgaagaaaacgttc
acggcgctgtttgcatcaggaatatgatca
ttaaacaatggagcgtaatggtattaatgatcaactagaaaaatggatggaggggcgaga
gggcgatcaacaaagcagccccggggc
atagtctggaagcagcaggaattggaagggaagggaagctgcacaatgaagggatatcg
tgagcggagtggctcacgagagtatc
aacagactggcgaaagcaagcaattgccaacgccggctattaggccataagatggcctgt
tgtgagtcccagttgcacgtatcccat
atgactgctctgtcgtgacttgaaaaaaataggaggataaaggagaaagaaagtgag
acaacccgtgagggacttggggtagta
ggagaacacatgggcaaccgggcaatacacgcgatgtgagacgagtcaacggcgaatgg
aaaatcttgaaaaacaaataaaata
actgccctccatacgggtatcaaattcaagcagttgtacggaggctagatagagttgtga
agtcggtaatcccgtgtatagtaatacga
gtcgcataaatactccgaagctgctgcgaacccggagaatcgagatgtcctggaaagc
ttctagcgagcggctaaattagcatgaa
aggctatgagaaattctggagacggcttgtgaatcatggcgttcattcttcgacaagc
aaagcgttccgtcgagtagcaggcactc
attcccgaaaaaactcggagattcctaagtagcgatggaaccggaataatataataggca

atacattgagttgcctcgacggttgaat
gcaggggtactgagcttgacataactgtccgtacccacctcttctcaacctttggcg
ttccctgattcagcgtacccgtacaagtcg
taatcactattaaccagactgaccggacgtgtttgcccttcatttgagaaataatgt
cattgcgatgtgtaattgcctgcttgaccgac
tggggctgttcgaagcccgaaatgtaggattgtatccgaactctgctcgtagaggcatgt
tgtgaatctgttcgggcaggacacgcctc
gaagggtcacggcaagggaaaccaccgatagcagtgctagtagcaacctgtaaagccgc
aatgcagcatcactggaaaatacaaa
ccaatggctaaaagtacataagttaatgcctaaagaagtcataaccagcggctaataat
tgtacaatcaagtggtctaaacgtaccgtaa
tttgccaacggcttggtgggtgcagaagcaacggcaaagccccacttccccacgtttgt
ttcttactcagtcgaatctcagctggtgat
ccccaattgggtcgcttgttgtccggtgaagtgaagaagacagaggtagaatgtc
tgactcggagcgtttgcatacaaccaag
ggcagtgatggaagacagtgaatgtgacattcaaggagtatttagccagggatgcttg
agtgtatcgtgtaaggagggttgctgccg
atacgacgaatactgtatagtcacttctgatgaagtggccatattgaaatgtaaagtcg
gcactgaacaggcaaaagattgagttgaaa
ctgcctaagatctcgggccctcgggccttggccttgggtgtacatgtttgtctccgg
gcaaatgcaaagtgtggtaggatcgaaca
cactgctgcctttaccaagcagctgagggtatgtgataaggcaaatgttcaggggccactg
catggttcgaatagaaagagaagcttag
ccaagaacaatagccgataaagatagcctcaftaaacggaatgagctagtaggcaaagtc
agcgaatgtgtatatataaagggttcgag

gtccgtgcctccctcatgctctccccatctactcatcaactcagatcctccaggagactt
gtacaccatctttgaggcacagaaaccaa
tagtcaaccgaggactgcgcacatgtatcgggaagtggccgtcatctcggccttcttg
ccacacctcgtgctagactaggcgcgtca
atatgtggccgttactcgagttataagtgaacaatgctctcaaagcgtcatggctgg
cacaagcctggaaagaaccaacacaaag
catactgcagcaaatcagctgaattcgtcaccaattaaagtgaacatcaacctgaaggcag
agtatgaggccagaagcacatctggatc
gcagatcatggattgccctcttgtgaagatgagaatctagaaagatggcggggtatga
gataagagcgatgggggggacatcatc
ttcaagacaaacaacctttgcagagtcaggcaattttcgtataagagcaggaggagg
agtccagtcattcatcagcggtaaaatc
actctagacaatctcaagatgagttctgccttgggtgacttatagccatcatcatacct
agacagaagcttgtgggatactaagaccaac
gtacaagctcgcactgtacgcttgacttccatgtgaaaactcgatacggcgcgcctcta
aattttatagctcaaccactccaatccaacc
tctgcatccctctcactcgtcctgatctactgttcaaatacagagaataaggacactatcc
aatccaacagaatggctaccacctccagc
tgctgcctacaagcaggacttctcaaataccgcatcgacggcggcgtcctcaagtttg
gcagcttcgagctcaagtccaagcgat
atccccctacttcttaacgcgggcgaattccacacggcgcgcctcgccggcgccatcgc
ctccgcctttgcaaagaccatcatcgag
gcccaggagaaggccggcctagagtcgacatcgtcttcggcccgccctacaagggcac
ccgctgtgctccgcatcaccatcaa
gtcggcgagctggcgccccagaacctggaccgcgtctcctactcgttgaccgcaagga

ggccaaggaccacggcgagggcgg
caacatcgtcggcgcttcgctcaagggaagagggtcctgattgtcgacgacgtcatcac
cgccggcaccgccaagagggaagcc
attgagaagatcaccaaggaggcgcatcgtcgccggcatcgtcgtggccctggaccgc
atggagaagctccccgctgcggatg
gcgacgactccaagcctggaccgagtgccattggcgagctgaggaaggagtacggcatcc
ccatctttgccatcctcactctggatg
acattatc gatggcatgaagggtttgctaccctgaggatatcaagaacacggaggatt
accgtgccaagtacaaggcgactgactg
attgaggcggtcaatgtcagaaggagagaaaagactgaaaaggtgaaagaagaggcaaa
ttgtgttattattattattctatctcgaat
cttctagatcttgcgtaaataaacaagcgtactagctagcctccgtacaactgcttga
atttgatacccgatggaggggcagttattttatt
ttgttttcaagattttccattcgccgttgaactcgtctcacatcgctgtattgcccgg
ttgcccattgtgtacgcgtttcgggtttacctcttcc
agatacagctcatctgcaatgcattaatgcattggacctcgcaaccctagtagcccttc
aggctccggcggaagcagaagaatagctta
gcagagtcatttttcatcttcgggagactagcattctgtaaaggcgagcaatgcccag
cagttagtaggggtcccctctacctctcagg
gagatgtaacaacgccaccttatgggactatcaagctgacgctggcttctgtgcagacaa
actgcgcccacgagttcttccctgacgcc
gctctcgcgcaggcaagggaactcgatgaatactacgcaaagcacaagagaccgttgg
ccactccatggcctccccatctctca
aagaccagcttcgagtcaaggtaaccgttggccctaagtcgttagatgtcccttttgt
cagctaacatatgccaccagggtacgaaa

catcaatgggctacatctcatggctaaacaagtacgacgaaggggactcggttctgacaa
ccatgctccgcaaagccggtgccgtctt
ctacgtcaagacctctgtcccgagacctgatggctcgcgagacagtcaacaacatcat
cggggcgaccgtcaaccacgcaacaa
gaactggctgctcggcggcagttctgggtgagggtgcgacgttgggattcgtgggtg
cgatcatcgggtgtaggaacggatatcgggt
ggctcgattcagtgccggccgcgtcaacttctgtacggctcgaaggccgagtcatggg
cggctgccgtatgcaaagatggcgaac
agcatggagggtcaggagacgggtgcacagcgttgcgggccgattacgcactctgttgag
ggtagtcccttcgctcttctctttcc tgctctataccaggcctccactgtcctcttcttgccttttatactatatacagaccg
gcagtcactgatgaagtatgtagacctccgct
cttcaccaaatacgcctcggcagggagccatggaaatacgaactccaaggatcatcccat
gccctggcgccagtcaggatcggacatt
attgcctccaagatcaagaacggcgggctcaatatcggctactacaacttcgacggcaat
gtccttcacacccctcctatcctgcgcgg
cgtggaaaccacgctgcgcgactcgccaaagccggtcacaccgtgaccccgaggacgcc
atacaagcacgatttcggccacgatc
tcctctcccatatctacgcggctgacggcagcgccgacgtaatgcgcgatatcagtgcat
ccggcgagccggcgattccaaatatcaa
agacctactgaacccgaacatcaaagctgttaacatgaacgagctctgggacacgcatct
ccagaagtgggaattaccagatggagtac
cttgagaaatggcgggaggctgaagaaaaggccgggaagggaactggacgccatcatcgcg
ccgattacgcctaccgctgcggtac
ggcatgaccagttccgggtactatgggtatgcctctgtgatcaacctgctggatttcacga
gcgtgggtgttcgggttacctttgcggataa

gaacatcgataagaagaatgagagttcaaggcggtagtgagcttgatgccctcgtgca
ggaagagtatgatccggaggcgtacat
ggggcaccgggtgcagtgacggttatcggacggagactcagtgaagagaggacgttggcg
attgcagaggaagtggggaagtgt
gggaaatgtggtgactccatagctaataagtgtcagatagcaattgcacaagaaatcaa
taccagcaactgtaaataagcgctgaagt
gaccatgcatgctacgaaagagcagaaaaaacctgccgtagaaccgaagagatatgac
acgcttccatctctcaaaggaagaatc
ccttcagggttgcgtttccagtagtgattttaccgctgatgaaatgactggactccctcc
tcctgctcttatacgaaaaattgcctgactctg
caaaggttgttgccttgaagatgatgtccccccatcgctcttatctcataccccgc
catcttctagattctcatcttcaacaagagggg
gcaatccatgatctgcgatccagatgtgcttctggcctcactctgccttcagggtgat
gttcacttaattggtgacgaattcagctgatttg
ctgcagtatgcttgtgttggcttctccaggcttgtgccagccatgagcgcttgagag
catgttgcacttataaaactcagtaacggcc
acatattgtcactactgaaatcacatacctaatttgatagaattgacatgtttaaga
gctgaggtagctttaatgcctctgaagtattgtga
cacagcttctcacagagtgagaatgaaaagttggactccccctaatgaagtaaaagttc
gtctctgaacgggtgaagagcatagatccg
gcatcaactacctggctagactacgacgtcaattctgcggcctttgacctttatatatg
tcattaatgcaatagattcttttttttttttttt
tttttttttttttttttgcctaattcgcatcaaagtggacgttatagcatcat
aactaagctcagttgctgaggggaagccgtctactacc
ttagcccatccatccagctccataccttgatactttagacgtgaagcaattcacactgta

cgtctcgagctctccttcccgtcttgcctcc
ccactgggggtccatgggtgcgtgtatcgccctccacaattctatgccatggtacctcca
gcttatcaatgccccgctaacaagtcgcctc
tttgccttgatagcttatcgataaaactttttccgccagaaaggctccgccacagac
aagaaaaaaattcaccgcctagcctttggc
ccgggcatttggttaaaccicgagcctctctcccgtcttgggggtatcaggaagaaaagaa
aaaaatccatcgccaagggtgtttggc
atcaccacccgaaaacagcacttcctcgatcaaaagttgccgccatgaagaccacgtgg
aaggacatccctccgggtgcctacgcac
caggagtttctggacatigtgctgagcaggaccagcgaaactgccactcagatccgt
gccggcttcaagattagcagaattcgag
gtacgtcgcatggccatcgaggatgtctcattatcggggtccttggagaacgatcatg
attgcatggc gatgctaacacatagacag
ccttctacactcgaaaggtaagttcaccaggagacgtttccgaaaagttcgctcca
tcctcgacagcttccctcgcctccaggaca
tccaccccttcacaaggaccttctcaacacctctacgatgccgaccactcaagattg
cccttggccagatgtccactgccaagcac
ctggtcgagaccatctcgcgcgactacgtccgtctcttgaaatacgcccagtcgctctac
cagtgcgaagcagctcaagcggggccgctc
tcggtcgcatggccacgctggtaagcgcctcaaggacccctgctgtacctggaccagg
tccgccagcatctcggccgtcttccctc
catcgaccccaacaccaggaccctgctcatctcggttaccccaatgttggcaagtccag
cttctgcgaagtatcaccgcgccgat
gtggacgtccagccctatgctttcaccaccaagagtctgttgcggccacttgactac

aagtacctgcgattccaggccattgataccc

ccggtattctggaccaccctcttgaggagatgaacactatcgaaatgcagaggtatgtgg cgcggt.

[244] Creation of the A5D strain from the Archy3 *T. reesei* strain. Native *T. reesei* *bgll* was deleted from the Archy 3 strain using a double recombination vector known in the art. Hygromycin resistance was used as the selectable marker for *bgll* deletion. In addition, the hygromycin resistance marker was flanked by loxP sites. The deletion cassette is shown in figure 8. Subsequent hygromycin resistant transformants were analyzed for *bgll* deletion. A strain confirmed for deletion of *bgll* was next transformed with a telomeric vector encoding *ere* recombinase and a functional *amdS* for selection of transformants to facilitate removal of the hygromycin by *ere* recombinase expression and loop out of the loxP sites. The telomeric vector encoding the *ere* recombinase is schematically shown in figure 10. Transformants were first obtained on acetamide media, then transferred to potato dextrose agar and replica plated onto hygromycin media to screen for hygromycin sensitivity. Strains sensitive to hygromycin, were again transferred to potato dextrose media and replica plated to acetamide media. A strain was selected which lost the ability to grow on acetamide, indicative of a loss of the telomeric vector. The nucleotide sequence of the *bgll* knockout cassette is provided as

SEQ ID NO: 17: aatggtaggaatgctgggatataggctctgtgctggcaagttgatggatcctcgaatgag
gccgccctgcaaggggaacatcagagat

ctaccattgcctccttggcccaatccactatcacctacctcatgatcattcctgcgaa

ggtctaccagtaaataatttcctcgtcccgtgtt.

catcatgtccagaacctcatctcgccaaattgactttgccacagtgtctggagctggga

agcagcgtgccaaggaattgtgtcgagtc

tgtgccaggcattgtgcccgaattgtgaacttcagccaggagaacttttcgatcgacc

tatgctgagcaccgtgggcatgcatggc

ttcaataatgcagttcgagagggagtggtcatgccctaaagctcattggccacctccac
aggctagctctacctgcatctgtatgga
ctttccttgctcctccttcagaaaacctcttggtcgctcgaggttaactgttgttgc
gtcattgttgacagtggatagccaaggcaaaa
ccgtctgctttcaacggaagcattcggcgggtgttgcacgtgtatcgatcgaccag
gagaaccagacgagtgtgttcgagagaa
tcacgacgatgtgaagaggcgacgactagtatctagaagattataatgaacaaatcag
cgttgtctgtcgggcgttgagggcgca
gttccccgcaaagcagcgtcgcaatatataggcagcgagagactgtcaacagccagccg
ccatgtgatcgatcgtagccgtcttc
cgatcttcctaaaccccttcttggggggcgggggcagcggcgttctaattttgctg
gctgtctggataacgtgaatggtagacatg
gtaatgttcggtctgcgaaacattgtacaattggagttacgatcgagatggaaggaaa
cgctccacaaactcggtagctgggtgcc
tcagggtctcagggcatagcgttctctgcaaataaggaaagagaatgcactagtga
gtgtgaatcacaatgaagaggaggtgt
tgccggaatgctttgagcagcgtcaaagtgaactgaagctatcacaattgcagggt
aagtacatgttggtgccagttgacagcac.
agtgcgcggagcggaggatgtcgcggaagaggcgcgacgctaaccgggccttcttca
gtgagcagaactcctgctgcaagagt
tccttctctctcgagatgacgtgagggccaatttcagcttcctcgaacaaggtgatt
gaacatctctctccctcacattcatcac
tacctctcaattcacttctgcttcggcgttcatcattcatgttactgctctgatgc
ctatcctgaagattgtatcctgcagtattcacgcc
atccaccttcggtcctcactcacagtcacaggtaaccgccttcacctcctcgcgatg

atgtcggcaatctggtggatcaatgtgcgg
ttgagggccgccgtagtgaggatgggcatggggaacgaggtcggccattcggccacagat
aaattcgtatagcatacattatacgaag
ttatcctgggcttgtgactggcgcgagctgccactaagtggggcagtaccattttatcg
gacccatccagctatgggacccactcgca
aattttacatcattttcttttggctcagtaacggccacctttgtaaagcgtaccagc
aaacaaattgcaattggcccgtagcaaggtagt
cagggccttatcgtgatggaggagaaggctatatcagcctcaaaaatatgttgccagctgg
cggaagcccgggaaggtaagtggattctt
cgccgtggctggagcaaccgggtgattccagcgtctccgacttggactgagcaattcagc
gtcacggattcacgatagacagctcag
accgctccacggctggcggcattattggttaacccggaaactcagtccttggccccgt
ccggaagggaacccgacttaccaggctgg
gaaagccagggatagaatacactgtacgggcttcgtacgggaggttcggcgtagggttgt
tcccaagttttacacaccccccaagaca
gctagcgcacgaaagacgcggagggtttggtgaaaaaggggcgaataaagcgggagac
gtatttaggtgctagggccggttccct ccccatttttcttcggttcccttctctcctggaagactttctctctctcttctctc
ttcttccatctcagtcctatcttcttcccatcatccat
ctctcacctccatctcaactccatcacatcacaatcgatatgaaaaagcctgaactcac
cgcgacgtctgtcgagaagttctgatcga
aaagttcgacagcgtctccgacctgatgcagctctcggagggcgaagaatctcgtgcttt
cagcttcgatgtaggagggcggtggatat
gtcctgcgggtaaatagctgcgccgatggtttctacaaagatcgttatgtttatcggcac
tttgcacggccgcgctcccgattccggaa
gtgcttgacattggggaattcagcgagagcctgacctattgcatctccgccgtgcacag

gggtgtcacgttgcaagacctgcctgaaac
cgaactgcccgtgttctgcagccggtcgcggaggccatggatgcgatcgtcggccga
tcttagccagacgagcgggttcggcc
cattcggaccgcaaggaatcggtaatacactacatggcgtgatttcatatgcgcgattg
ctgatcccatgtgtatcactggcaaactgt
gatggacgacaccgtcagtgccgtccgtcgcgcaggctctcgatgagctgatgcttgggc
cgaggactgccccgaagtcggcacct
cgtgtcacgcggatttcggctccaacaatgtcctgacggacaatggccgcataacagcgg
cattgactggagcgaggcgatgttcgg
ggattcccaatacagggtcgccaacatcttcttctggaggccgtggttgcttgatgga
gcagcagacgcgctacttcgagcggagg
catccggagcttgcaggatcgccgcggctccggcgatatatgctccgattggtcttgac
caactctatcagagcttggtgacggcaa
tttcgatgatgcagcttgggcgcagggtcgatgcgacgcaatcgtccgatccggagccgg
gactgtcgggcgtacacaaatcgccc
gcagaagcgcggccgtctggaccgatggctgtgtagaagtactcgccgatagtggaacc
gacgccccagcactcgtccgagggc
aaaggaatagagtagatgccgaccgggatccactaacgttactgaaatcatcaaacagc
ttgacgaatctggatataagatcgttgg
gtcgtatgtcagctccggagttgagacaaatggtgttcaggatctcgataagatacgttca
tttgtccaagcagcaaagagtgccttctagt
gatttaatagctccatgtcaacaagaataaaacgcgttcgggtttacctcttcagata
cagctcatctgcaatgcattaatgcattggac
ctcgcaccctagtagcccttcaggctccggcgaagcagaagaatagcttagcagagtc
tatttcatcttcgggagacgagatcaag

cagatcaacggctgtcaagagacctacgagactgaggaatccgctcttggtccacgcga
ctatatattgtctctaattgtacttgacat
gctcctcttcttactctgatagcttgactatgaaaatccgtcaccagcccctgggttc
gcaaagataattgcactgtttcttccttgaactct
caagcctacaggacacacattcatcgtagggtataaacctcgaaaatcattcctactaaga
tgggtatacaatagtaaccatgggtgccta
gtgaatgctccgtaacacccaatacggcggccgataacttcgtatagcatacattatacg
aagttatacttggcgcgcctagtgggaacac
gagcacataagcttttacctatgggtatcgcttgcacttacgcgccgttgatgggtggagg
atgggtggacgttcccagacccctacgagc
tgtggcatcgtcaaactgtgccacagaccttgtcttgccttcataacctcgaggagtg
ttccagactcatcatccatacacaagcagta
ttaatcaaagaaactcggtcgcaatggcaaaaatggttgcaaacagaaaactatggcct
cttcctattccatcataactactctacccgt
ttgtcataacaacatcattaaaaccttatgcgtcagggtgtagcatccttgatctgttgc
tcctccaacggccagttctcaatcgttacctctt
ctcccaccaactcaaactcaagcttcacagactcgtcgggtgtcaaggctagctcatact
tgccgggggtatacaatccggttccgtgag
aatcaacacgggcgagagcactgacagggatggggatgctgagcttggagagtgaccag
gcttgatgtcggcaagtcggtcgaat
ccgacgagccacttgtcgggtacggggctgggccagcgttgcttgtcgaacaaacagc
atggccgtatatggggactccgtcttgc
ccgagttcttgatgttggcctcgaagggtgaagacgggaatctgctcgctgtaagtgtatc
cgggggtgaggagcagagaggatcgaatga
gggtgtgaacttgaggctcttgggggtggctggcgagagtctccttgaagggtggttagaa

gagaccactgccaaactcgtagacgggt
 ttgccgggtgtaccagatgtaagtctgtccagggttgactttccatcgggtcggagggtc
 atgtcattctgggggaattggtgaacatactc
 agccgggtactgagtggtagaccagtcggccggcaggagcacgcctgccagagagaatgtc
 gaagaggggcaacgcctcccgactg
 gccgggatatccgcccagacgaggaggtgaccttctgttgctcttgagcaggatga
 gtctacctgaccaccgcccatttgagg
 acgacaagggtttgccgacctcgtgagctgcttgatgagatccagctgattaccgggc
 caagcaatgtccgtgcggtcagcgcct
 cctgttcaatggtgtgtcaattccaccgaggtagatgatggcatccgacttcttgccgg
 cagcaatggccttggcaaagccagtggg
 ctgttgccggcgaictctgtgccgagttcaaagttgacgtgatagccggccttcttagca
 gcttcagagggctgatgaggtatggggc
 agggccatagtagttgccttgcatgttggttggttgccatggatccgatcagagc
 aatgctgcgcaccttcttgacagaggga
 gagtgccatcgttcttgagcaggacgatgccctcaacagcagcctctgacgagatgtt.

[245] The nucleotide sequence of the telomeric vector, pTTT-cre, is provided as SEQ ID

NO:18: ttgtacaaagtggatgcgcgccgcgccagctccgtgcgaaagcctgacgcaccggta
 gattcttggtgagcccgtatcatgacg gcggcgggagctacatggccccgggtgattattttttgtatctacttctgaccctt
 tcaaatatacgggtcaactcatctttcactggaga
 tgcggcctgcttggtattgcgatgtgtcagcttgcaaatgtggcttgcgaaacaca
 aaacgattccttagtagccatgcattttaagat
 aacggaatagaagaaggaaattaaaaaaaaaaaaaaaaacaaacatcccgttcataac
 ccgtagaatcggcgctcttcgtgtatcc
 cagtaccagttattttgaatagctcggccgctggagagcatcctgaatgcaagtaacaa

ccgtagaggctgacacggcagggtgtgct
agggagcgtcgtgttctacaaggccagacgtcttcgcgggtgatatatatgtatgtttga
ctgcaggctgctcagcgacgacagtcaagt
tcgccctcgtcgttgtgcaataatcgacgtggggaagccacaccgtgactcccatctt
cagtaaagctctgttggtgttatcagcaat
acacgtaatttaaactcgttagcatggggctgatagcttaattaccgtttaccagtcca
tggttctgcagctticcttgccccgtaaaattc
ggcgaagccagccaatcaccagctaggcaccagctaaaccctataattagtctcttatca
acaccatccgctccccgggatcaatga
ggagaatgagggggatgcggggctaaagaagcctacataaccctcatgccaactcccagt
ttacactcgtcgagccaacatcctgac
tataagctaacacagaatgcctcaatcctgggaagaactggccgctgataagcgcgccccg
cctcgcaaaaaccatccctgatgaatg
gaaagtccagacgctgcctgcggaagacagcgttattgattcccaaagaaatcggggat
cctttcagaggccgaactgaagatcaca
gaggcctccgctgcagatcttgtccaagctggcggccggagagttgacctcgggtggaa
gttacgctagcattctgtaaaccgggcag
caatgcccagcagttagtaggggtccctctacctctcaggagatgtaacaacgccacc
ttatgggactatcaagctgacgctggctt
ctgtgcagacaaactgcgccacgagttctccctgacgccgctctcgcgcaggcaaggg
aactcgatgaatactacgcaaagcaca
agagacccgttggccactccatggcctccccatctctctcaaagaccagcttcgagtca
aggtacaccgttggccctaagtcgttagat
gtccctttttgtcagctaacatagccaccagggtacgaaacatcaatgggctacatct
catggctaaacaagtacgacgaaggggac

tcggttctgacaaccatgctccgcaaagccggtgccgtcttctacgtcaagacctctgtc
ccgcagaccctgatggctcgcgagacagt
caacaacatcatcgggcgcaccgtcaaccacgcaacaagaactggctcgtcggcggcag
ttctgggtggtgagggcgcatcgttg
ggattcgtggtggcgtcatcgggttaggaacggatatcggaggctcgattcagtgccgg
ccgcgttcaacttcctgtacgggtctaagg
ccgagtcatgggcggctgccgtatgcaaagatggcgaacagcatggagggtcaggagacg
gtgcacagcgttgctgggccgattacgcactctgttgaggggtgagtccttcgcctcttcttcttcttcttctgctctataaccaggcc
tccactgtcctccttcttctgcttttatactatatac
gagaccggcagtcactgatgaagtatgtagacctccgcctcttcaccaaaccgtctc
ggtcaggagccatggaaatacgaactcca
aggtcatcccatgccctggcgccagtcggagtcggacattattgcctccaagatcaaga
acggcgggctcaatatcggctactacaa
cttcgacggcaatgtccttcacacctcctatcctgcgcggcggtggaaccaccgtcgc
cgactcgccaaagccggtcacaccgt
gaccccggtggacgccatacaagcacgattcggccacgatctcatctccatactacgc
ggctgacggcagcggcgacgtaatgcg
cgatatcagtgcatccggcgagccggcgattccaaatatcaaagacctactgaacccgaa
catcaaagctgttaacatgaacgagctc
tgggacacgcatctccagaagtggaaattaccagatggagtagccttgagaaatggcgggag
gctgaagaaaaggccgggaaggaaac
tggacgccatcatcgcgccgattacgcctaccgctgcggtaggcatgaccagttccggt
actatgggtatgcctctgtgatcaacctg
ctggatttcacgagcgtggtgtccgggtaccttgcggataagaacatcgataagaag
aatgagagttcaaggcggtagtgagctt

gatgccctcgtgcaggaagatgatccggaggcgtaccatggggcaccggttcagtg
caggttatcggacggagactcagtga
agagaggacgttggcgattgcagaggaagtggggaagtigctgggaaatgtggtgactcc
atagctaataagtgtcagatagcaattt
gcacaagaaatcaataccagcaactgtaaataagcgtgaagtgacctgccatgctacg
aaagagcagaaaaaacctgccgtag
aaccgaagagatatgacacgcttccatctctcaaaggaagaatcccttcagggttcggtt
tccagtctagacacgtataacggcacaag
tgtctctaccaaattgggttatatctcaaattgtgatctaaggatggaaagcccagaatat
cgatcgcgcgcagatccatatatagggccc
gggttataattacctcaggaaatagctttaagtagcttattaaattataat
attttaataataactatatttctttaataaataggatt
ttaagctttatataataataataataataataatattatagctttttattaataa
ataaaaatagctaaaaatataaaaaaataagctttaa
atacttattttaattagaattttatatattttaatatataagatctttacttttta
taagcttcctacctaataataaattttacttttttactatttta
ctatatcttaataaaggctttaaaaataaaaaaatcttcttatattataagcta
taaggattatatatatatttttttaatttttaaagtaa
gtattaaagctagaattaaagttttaatttttaaggcttatttaaaaaaggcagtaa
tagcttataaaaagaaatttcttttctttatactaaa
agtactttttttaataagggttagggtagggtagggtagggtagggtagggtagg
gggttagggtagggtagggtagggtagggtagg
gttagggtagggtagggtaagggttaaacaaagccacgttgtgtctaaaatctctg
atgttacattgcacaagataaaaatatatcat
catgaacaataaaactgtctgcttacataaacagtaataacaaggggtgtatgagccata

ttcaacgggaaacgtcttgctcgaggccgc
gattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcg
ggcaatcaggtgcgacaatctatcgattgtat
gggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgat
gttacagatgagatggtcagactaaact
ggctgacggaatttatgccctctccgacatcaagcattttatccgtactcctgatgatg
catggttactcaccactgcgatccccgggaa
aacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgct
ggcagtgctcctgcgccggttgcatcgttc
ctgtttgtaattgtcctttaacagcgatcgctatttcgtctcgctcaggcgcaatcac
gaatgaataacggtttggttgatgcgagtgttt
tgatgacgagcgtaatggctggcctgttgacaagtctggaaagaaatgcataagctttt
gccattctcaccggattcagtcgtcactcat
ggtgatttctcacttgataaccttattttgacgaggggaaatiaataggttgattgat
gttggacgagtcggaatcgagaccgatacca
ggatcttgccatcctatggaactgcctcggtagtatttctccttcattacagaaacggct
tttcaaaaatatggtattgataatcctgatatg
aataaattgcagtttcatttgatgctcgatgagttttctaatcagaattggttaattgg
ttgtaacactggcagagcattacgctgacttgac
gggacggcggccttgttgataaaatgaacttttgctgagttgaaggatcagatcacgca
tcttcccgacaacgcagaccgttccgtgg
caaagcaaaagttcaaaatcaccaactgggtccacctacaacaaagctctcatcaaccgtg
gctccctcactttctggctggatgatggg
gcgattcaggcctggatgagtcagcaacaccttcttcacgaggcagacctcagcgggtt
aaacctaaccctaaccctaaccctaacc

taaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaacc
taacctaaccctaataagggtcgatctgaac
cgaggatgaggggttctatagactaatctacaggccgtacatggtgtgattgcagatgcga
cgggcaagggtgtacagtgtccagaagga
ggagagcggcataggattgtaatagaccagctttacataataatgcctgttgctactg
actgatgaccttcttccctaaccagtttccta
attaccactgcagtgaggataaccctaactcgtctggggttattattatactgattagc
aggtggcttatatagtgctgaagtactataag
agtttctgcgggaggaggtggaaggactataaactggacacagttagggatagagtgatg
acaagacctgaatgttatcctccggtgt
ggtatagcgaattggctgacctgcagatggaatggttaggcagggttttgagagg
gggacgagaacgcgttctgcgatttaacg
gctgctgccccaagctttacggttctctaattggcgggccgcctcaggtcgacgtcccat
ggccattcgaattcgtaatcatggcatag
ctgttctctgtgtgaaattgttatccgctcacaattccacacaacatacagccggaagc
ataaagtgtaaagcctggggtgcctaataga
gtgagctaactcacattaattgcgttgcgtcactgcccgtttccagtcgggaaacctg
tcgtgccagctgcattaatgaatcggccaa
cgcgcggggagaggcggtttgcgtattggcgctcttccgcttcctcgtcactgactcg
ctgcgctcggtcgttcggctcggcgag
cggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcag
gaaagaacatgtgagcaaaaggccag
caaaaggccaggaaccgtaaaaaggccgcgttgctggcgttttccataggctccgcccc
cctgacgagcatcacaaaaatcgacgc
tcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctgga

agctccctcgtgcgtctcctgttccga
ccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctc
atagctcacgctgtaggtatctcagttcgg
gtaggtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctg
cgcttatccggtaactatcgtttgagt
ccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagca
gagcgagggtatgtaggcggctgctaca
gagttcttgaaagggtggcctaactacggctacactagaagaacagtatttggtatctgc
gcctctgctgaagccagttaccttcggaaaa
agagttggtagctcttgatccggcaaacaaccaccgctggtagcgggtgtttttgtt
tgcaagcagcagattacgcgcagaaaaaa
aggatctcaagaagatcctttgatcttttctacgggggtctgacgctcagtggaacgaaaa
ctcacgttaagggattttggctatgagattat
caaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaataaatctaaa
gtatatatgagtaaacttggtctgacagttacca
atgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcacccatagttgc
ctgactccccgtcgtgtagataactacgatacg
ggagggttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccggc
tccagatttatcagcaataaaccagcc
agccggaagggccgagcgcagaagtggctcctgcaactttatccgcctccatccagtctat
taattgttgccgggaagctagagtaagta
gttcgccagttaatagtttgcgcaacgttgtgccattgctacaggcatcgtggtgtcac
gctcgtcgtttggtatggcttcattcagctccg
gttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcggttagct
ccttcggtcctccgatcgttgtcagaagta

agttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtca
tgccatccgtaagatgctttctgtgactggtg
agtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttggccgg
cgtcaatacgggataataccgcgccacat
agcagaactttaaaagtgtcatcattggaaaacgttctcggggcgaaaactctcaagg
atcttaccgctgttgagatccagttc gatgt
aaccctactgtgcaccaactgatcttcagcatctttactttcaccagcgtttctgggt
gagcaaaaacagggaaggcaaaatgccgcaa
aaaaggggaataagggcgacacggaaatgtigaatactcatactcttcttttcaatatt
attgaagcatttatcagggttattgtctcatga
gcggtacatatattgaatgtatttagaaaaataaacaatatgggggtccgcgcacatttc
ccgaaaaagtgccacctgacgtctaagaaa
ccattattatcatgacattaacctataaaaaataggcgatcacgaggcccttctgtctcg
cgcgtttcgggtgatgacggtgaaaacctctga
cacatgcagctcccgagacgggtcacagcttgtctgtaagcggatgccgggagcagacaa
gcccgtcagggcgcggtcagcgggtg
ttggcgggtgtcggggctggcttaactatcgggcatcagagcagattgtactgagagtgc
accataaaattgtaaacgttaatatttgtt
aaaattcgcgttaaattttgttaaatcagctcatttttaaccaataggccgaaatcgg
caaaatcccttataaatcaaaagaatagcccga
gatagggttgagtgttgttccagtttgaacaagagtccactattaaagaacgtggactc
caacgtcaaagggcgaaaaaccgtctatc
aggcgcatggcccactacgtgaaccatcacccaaatcaagtttttggggtcgaggtgcc
gtaaagcactaaatcggaaccctaaagg
gagccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaa

gaaagcgaaaggagcgggcgcta
gggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatg
cgccgctacagggcgcgtactatggt
tgctttgacgtatgcgggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatca
ggcgccattcgccattcaggctgcgcaa
ctgttgggaaggcgcatcgggtcggggcctcttcgctattacgccagctggcgaaaggggg
atgtgctgcaaggcgattaaagtgggt
aacgccagggtttccagtcacgacggtgtaaacgacggccagtgcccaagcttacta
gatgcatgctcgagcggccgccagtgt
gatggatatctgcagaattcgcccttgactagtctctctatcctgggtggcaggcgctcaa
gtaccagaggcagcagcgggcttagga
gcggcctgggttgttctccgcaccctctacatgctgggctatattatagcgacaagccg
aacggcaccggcaggtacaatggttcgct
gtacttgcttgcgcaagcgggtctttggggattgagcgcatttgggttgcaaaggattt
gatgtaaatgtagtcgacatcttagcacaga
ggggagagttgataaaatgtggtctgtttgaatgatagtcgggttcgtgacctatattcg
tgatagtggagataggctcgcgcctatcttat
cgggccggagcaaaaattccaccgcagcggggtaggtttcgttatacagccatcccact
tccagcttcaaattgtcagtttaatccagc
ccaattcaatcattggagaaccgccatcatgtcttcgaagtcaccctcccctacgcaat
tcgcgcaaccaaccatcccaaccctttaac
atctaaactcttctccatcgccgaggagaagaaaaccaacgtcaccgtctccgcagacgt
tactactccgccgagctcctcgatcttgc
tgaccgtacatcctgcaccaatgcccctccaggataacaatagctgatgcgtagtgagt
acaggcctaggcccctatatcgcagttct

gaaaaccacatcgacatcctcaccgatctcaccocgtgaccctttcctcgtccaatc
cctcgcgacaaagcacaacttcctcatctt
gaggaccgcaagttcatcgacatcggcaacaccgtgcaaaagcagtaccacgggtggcgct
ctccgcatctccgaatgggcacacat
catcaactgcgccatcctgccggcggaaggatcgtcaggccctcgacagacaaccaa
gtctcctgactttaagacgcgaatca
acgaggctcctgattctgccgagatgacgagtaagggatccttgcgacaggggagtc
acaggcacgctcggttgagtacgcgcg
gaagtataaggggttgtgatgggattcgtgagtacaagggcggttgagtgggtgctgcc
cgaacagaaaaggagagcgaggattt
tgtcgtctttacgactggggtgaatctgtcggataagggggataagctggggcagcagta
tcagacacctgggtcggcgggtgggcg
aggatcggactttatcattgcggtaggggcatctataaggcggacgatccagtcgaggc
ggttcagaggtaccgggaggaaggct
ggaaagcttacgagaaaagagttggactttgagtgtgagtggaaatgtgtaacggtattg
actaaaagggatccatatgtttattgcagc
cagcatagttaccagaaagagcctcactgacggctctagtagtattcgaacagatatt
attgtgaccagctctgaacgatatgctccct
aatctggtagacaagcactgatctacccttggaaacgcagcatctaggctctggctgtgc
tctaaccctaactagacgattgatcgcaga
ccatccaatactgaaaagtctctatcagaggaaatccccaacattgtagtagtcagggtc
ctttgtggctgggagagaattggttcgtcc
actgattccagttgagaaagtgggctagaaaaagtcttgaagattggagttgggctgtg
gttatctagtagtctcagagctctgtacatgt
ccggtcgcgacgtacgcgtatcgatggcgccagctgcaggcggccgcctgcagccacttg

cagtccttggaattctcacggtgaat
gtaggccttttagggtaggaattgtcactcaagcaccaccaacccattacgcctcc
cccatagagttccaatcagtgagtcattg
cactgttctcaaatagattggggagaagtgactccgccagagctgaaggcgcacaa
ccgcatgatatagggtcggcaacggca
aaaaagcacgtggctcaccgaaaagcaagatgttgcgatctaataccaggaacctgga
tacatccatcatcacgcacgaccactt
gatctgctggtaaactgtattgccctaaaccgaagtgcgtggtaaattacacgtggg
cccccttcgggtatactgcgtgtgtcttctta
gggtccattctttccctcctctagtggtgaattgttggagtcaggctgtaa
ctaccttgaaatctctggagaatggaggactaac
gactaccgtgcacctgcatcatgtatataatagtgatcctgagaaggggggttgagca
atgtgggactttgatggtcatcaacaaag
aacgaagacgcccttttgaaagtttgttcggctacgggaagaactggatacttgt
tgtgtctctgtgtattttgtggcaacaagag
gccagagacaatctattcaaacaccaagcttgccttttgagctacaagaacctgtgggg
tatatactagagttgtgaagtcggaatcc
cgctgtatagtaatacagatcgcattcaatactccgaagctgctgcgaacccggagaat
cgagatgtgctggaaagcttctagcgag
cggctaaattagcatgaaaggctatgagaattctggagacggcttgtgaatcatggcg
ttccattctcgacaagcaaagcggtccgt
cgcatgacggcactcattccgaaaaaactcggagattcctaagtagcgatggaaccg
gaataatataataggcaatacattgagt
gcctcgacgggtgcaatgcaggggtactgagcttgacataactgtccgtacccacct
cttctcaaccttggcggttccctgattcag

cgtacccgtacaagtcgtaatcactattaacccagactgaccggacgtgtttgcccttc
atttgagaaataatgtcattgcgatgtgtaa
tttgctgcttgaccgactggggctgttcgaagcccgaatgtaggattgtatccgaact
ctgctcgtagaggcatgttgtgaatctgtgtc
gggcaggacacgcctcgaagggtcacggcaagggaaccaccgatagcagtgtctagtag
caacctgtaaagccgcaatgcagcat
cactggaaaatacaaaccaatggctaaaagtacataagttaatgcctaaagaagtcatat
accagcggctaataattgtacaatcaagt
gctaaacgtaccgtaatttgccaacggcctgtgggggtgcagaagcaacggcaaagcccc
acttccccacgtttgtttcttactcagtc
caatctcagctgggtatcccccaattgggtcgcttgtttgtccggtgaagtgaaagaag
acagaggtgaagaatgtctgactcggagcg
tttgcatacaaccaagggcagtgatggaagacagtgaaatgttgacattcaaggagtat
ttagccagggatgcttgagtgtatcgtgta
aggagggttgtctgccgatacgacgaatactgtatagtcacttctgatgaagtggccat
attgaaatgtaaagtcggcactgaacaggc
aaaagattgagttgaaactgcctaagatctcgggccctcgggccttcggcctttgggtgt
acatgtttgtgctccgggcaaagtcaaaagt
gtggtaggatcgaacacactgctgcctttaccaagcagctgagggtatgtgataggcaaa
tgttcagggggccactgcatggttcgaat
agaaagagaagcttagccaagaacaatagccgataaagatagcctcattaaacggaatga
gctagtaggcaaagtcagcgaatgtgt
atatataaagggtcgaaggctcgtgcctccctcatgctctccccatctactcatcaactca
gatcctccaggagacttgtacaccatcttttga
ggcacagaaaccaatagtcaaccatcacaagtttgtacaaaaagcaggctcacatga

gcaacctgctcaccgtccaccagaacc
 tccctgccctccctgtcgagccacctctgacgaggtccgcaagaacctcatggacatgt
 tccgcgaccgccaggcctttagcgagca
 cacctggaagatgctcctcagcgtctgccgatcttgggccgcctggtgcaagctcaacaa
 ccgcaagtgggtccccgccgagccgga
 ggacgtccgcgactacctcctctacctgcaggcccaggcctggccgtcaagaccatcca
 gcagcacctcggccagctcaacatgc
 tccaccgacgctctggcctgcctcgcctagcgactctaacgccgtcagcctggtcatgc
 gccgcatccgcaaggagaacgtcgacg
 ctggcgagcgagccaagcaggccctcgccttcgagcgcaccgacttcgaccaggtccgca
 gcctcatggagaacagcgaccgctg
 ccaggatatccgcaacctcgcctttctcggcattgcctacaacacctgctccgattgc
 cgagatcgcgcgcatccggtcaaggac
 atctctcgaccgacggcgccgcatgctcattcacatcgccgcaccaagaccctcgtg
 tctaccgccggcgtcgagaaggccctc
 agcctcggcgtcaccaagctcgtcgagcgctggatttctgtctccggcgctcgctgacgac
 cccaacaactaccttcttctgccgctccg
 aaagaacggcgctcgccgcccttctgccacctctcagctcagcaccgagccctggaggg
 catctttgaggccaccacccgctcat
 ctacggcgccaaggacgactctggccagcgctacctcgcctggtctggccactctgcccg
 agtcggcgctgcccagacatggccc
 gagccggcgtcagcatccccgagattatgcaggccggcggtggaccaacgtcaacatcg
 tcatgaactacatccgcaccctcgact ctgagaccggcgccatggtccgactcctcgaggacggcgactaaaccagcttc.

[246] Creation of the MAD6 strain from A5D T. reesei strain. Native eg!3 was deleted from the A5D strain using the method previously described for bglI deletion. A

schematic of the deletion cassette is shown in figure 9. Hygromycin resistance was used as the selectable marker for egl3 deletion. The hygromycin resistance marker was flanked by loxP sites. A transformant was confirmed with a deletion of the egl3. The hygromycin marker was removed from this strain as described for creation of the A5D strain. The nucleotide sequence of the egl3 deletion cassette is provided as SEQ ID NO:

19: gggaggtaggcgagatacgggtgcatgggacccgaacccgtaaccggaacacgaccttat
cagccctccaactcacacccctctcgc
ctatcactatcctagatagttcatcggccaactcatgtaacctagctacctacctacctg
gtaagaatgcgggctatcatgtctcacggcg
cggtagatgtcgggtatctcgtgcttccccgcagggtgacgtcggaaatccatgcaagtac
tccctgaaatcgagacgacagagagaac
aaccaacgcgcttaaacgcttcattcatctaagaggcacattcgaagaactagcttaa
cacactagacctggcttttcgacccctcc
gcagaaagccgtttctcctcaatcctcccgggcttggctttgtcagtcctgacttgc
gcgctaacagagtcttgacgcagcgttgc
gcatcagtcctgcaggcgggtcacgggactaggacaacaggggatgtgacaggccggata
gtaattatgggttatccgggtaagca
gggaatttacgaggccgcttacgtgggggaacagccacttgcggggggaagaggagtag
taggcgactcggtcgatgagctcga
ggtgtctggttgacttgactgcagagcgtaggtaattgagatcgggcaacattatcgt
gttcggctcgggtatggccgagttgcgact
gcttggtcattcggcgaagctgatgtcgtggatcctgaagcatcgatcggaaacct
gatggtcagtcctatctgacgtgtgcggtga
caagcgagtcgggattttgtgacatgacgttcaacttcagtcaatgccttagggctcgat
aagattaagattgggttctggcagcggtcta

gaacaccgccacaaattctgtccattgaggagcgtgatgtctaggcgcatacactaacacg
gagctgtatgaccggcagctcaacgga
cttccttcgttcaacggcagctctatttgcggtacacgaatggatctttcttctggtct
tgaagtgccgcagtggcgtgcgaatgtatagat
gtctcgtacctagaaaaagctggctttctgacagggctccctccacctctcctaccaac
gacaaactgaacaagtatctggcggttccc
aacgccgaataggccagtcgccaatactccctccagccctgattgggcccctcgaagtat
cgccatgtctgtgtgtgagattattcga
ggacgtcactcccccaacctacaggaagagcaaaatgggagcagtggtctgcaatgagct
atataatagatcgctcgatctcatacaaa
ttgtatgctcagtcatacaacgagcgggtccaagatcccttctccaacgacctcga
cattgcaacccgggtgcagcctgaacttgt
cgtatagcctagaaagcgacgccatcttcatctttacgcgattagcctcatggctatt
gtgccgaagtgggagttgtatggtagcagtg
aggagattgtggctacgacacaggcgggttctcttgagcggcttacatctccgattagg
cctgcgtacgatccagatcatgggaaact
ttacaatggcttactcgtttatctcaacactgagcttccaattcactctatgcattgat
taacacgtttggtcatgtggttctcagctgtaaat
cttcagcttccaagaattgcaacctcgctgattgctaatagtgttgcacgttgcatc
ctggtgcggcagtgcaaaggagagtc
tagccggcagattaatttaagcttatcactcaggggtaaacagccgtaaaggacctt
tgatctaacatgccgatgtgtatgtatgatca
cgcaatgccaccatatcttggcagtcagattgtccgtggcgcgccaagtataactcg
tataatgtatgctatacgaagtatcggccg
gcgtattgggtgttacggagcattcactaggcaacctggttactattgtataccatct

tagtaggaatgatttcgaggtttatacctacg
atgaatgtgtgtcctgtaggcttgagagttcaaggaagaacagtgcattatctttgcg
aaccaggggctggtgacggaatttcata
gtcaagctatcagagtaaagaaggagcatgtcaaagtacaattagagacaaatatata
gtcgcgtggagccaagagcggattcctc
agtctcgtaggtctcttgacgaccgttgatctgcttgatctcgtctcccgaatgaaa
tagactctgctaagctattctctgcttcgccg
gagcctgaagggcgtagtaggggtgcgaggtccaatgcattaatgcattgcagatgagct
gtatctggaagaggtaaaccggaaacg
cgttttattctgttgacatggagctattaaatcactagaaggcactctttgctgcttgg
acaaatgaacgtatcttatcgagatcctgaacac
catttgtctcaactccggagctgacatcgacaccaacgatcttatatccagattcgtcaa
gctgtttgatgatttcagtaacgttaagtggat
ccgggtcggcatctactctattcctttgccctcggacgagtgcctggggcgtcggttcca
ctatcggcgagtactctacacagccatcg
gtccagacggccgcgttctgcgggcgatttgtgtacgcccagagtcgccggtccggat
cggacgattgcgtcgcacgcacctgc
gccaagctgcatcatcgaaattgccgtcaaccaagctctgatagagttggtcaagacca
atcgggagcatatacgcccggagccgc
ggc gatcctgcaagctccggatgcctccgctcgaagtagcgcgtctgctgctccatacaa
gccaaccacggcctccagaagaagatg
ttggcgacctcgtattgggaatccccgaacatcgctcgtccagtcaatgaccgctgtt
atcgggccattgtccgtcaggacattgttg
gagccgaaatccgcgtgcacgaggtgccggacttcggggcagtcctcggcccaaagcatc
agctcatcgagagcctgcgcgacgg

acgcactgacgggtgctccatcacagttgccagtatacacatggggatcagcaatcg
cgcatatgaaatcacgcatgtagtgatt
gaccgattccttgcgggccgaatgggccgaacccgctcgtctggctaagatcgccgcag
cgatcgcatccatggcctccgcgaccg
gctgcagaacagcgggcagttcggttcaggcaggtcttgaacgtgacaccctgtgcac
ggcgggagatgcaataggtcaggctct
cgctgaattcccaatgtcaagcacttccggaatcgggagcgcggccgatgcaaagtcc
gataaacataacgatctttagaaacc
atcggcgcagctatttaccgcaggacatatccacgccctcctacatcgaagctgaaagc
acgagattcttcgccctccgagagctgc
atcaggtcgggagacgctgtcgaactttcgtacagaaacttctcagacagctcgcggtg
agttcaggcttttcatatcgattgtgatgtg
atggagttgagatggaggtgaggagatggatgatgggaaaggagatggactgaggatgg
aagaagagaagaagagagagaga
gaaagtcttcaggagagaaagggaaccgaagaaaaatggggaggaaaccggccctagca
cctaaatacgtctcccgcttaatttc gccctttttcaccaaaccctccgcgtcttctgtcgttagctgtcttggggggtgtgta
aaacttgggaacaaccctacgccgaacctcc
cgtacgaagcccgtacagtgattctatccctggctttccagcctggttaagtcgggtcc
cttcgggacggggccaaggagactgagti
tccgggtaaccaataatgccgccagccgtggagcggctctgagctgtctatcgtgaatcc
gtgacgctgaattgctcagtcgaagtcgg
agacgctggaatccaccggttctccagccacggcgaagaatccacttaccttccgggt
tccgccagctggcaacataattttgaggc
tgatatagccttctcctccatcacgataagccctgactaccttgctacgggccaattgca
attgtttgctggttacgctttacaaaagggtgg

ccgttactgagcaaaaagaaaatgatgtaaaaatttgcgagtgggtcccatagctggatg
gggccgataaaatggtagtccccactta
gtggcagctcgcgaccagtcacaagcccaggataacttcgtataatgtatgctatacgaa
gttatctgtggcggttatgaataatagact
ggaaccggggccctttgattgacgactccatattttgtagatgtagcaactcggcaagagc
attatgtgcaatacatttgttaccatacaaa
ggcagctgccagacgacttgattgcgtacaattctcacggcaagctttccagggtttat
gcattatgcgcaaatgcttgatgcttaccgc
aggattaatctcggagaagcgctgcaagctatatgggtgtagtagatatgtagatgtac
caaccaatgaagaacatttatggtctagaa
cgtagtgatgaagggtttgagtaattgtatcaagtaagacgatattatgatataatac
caagcatatattcatgataaattacttggaaacca
cccttgcgtccggcctcacgagccttctcactgccgggctcgaaggagccactggaggcc
tgtccacccttgatgcgatttcctgca
ccctttccttgggcctgcacgtcgattagacatgattcaaatcgagatcttggaaatatct
tacctgtggcgaagccaccggtgtggctgg
actgtccgcccttctgcgcaatgcttgaacctcctccttggggctgtgtagaaaggttt
gttagcaacattactacaactctcaggactcg
gtggtcgtaccggttggcgaagtttccggggttatcgttgccagacattgtgtgattatt
tggtgtgcaaatgtgtgctatgtgtgtgtgtgct
gttggtgatgatgctgaagctgttgaaagcaggctgggtctgtgggagagacttgggata
tttatatccaaagttcggtcgtgttccttctg
gaagctcttctactccatacaatcatccaaagttgtcgtcattgagcgttgatcagta
gtagcctctgaggatcatcaccatgatccttccg
gccaacagtcggcactcatcaacagcaacaatcagccgccacaaacataggtacagtaag

gagtagatatcatgtagtcgtcgagta
 ctcgacatcatgacgtacaagcttggcagtgtaggtgcaagtatgatgatcgat
 ccgccgttggtcgatcgaacagagtgcggt
 cagattcacggttctctcaccttgaacattggatgcaattggattgatccacaatcctg
 gagaatggcttcaagctcactgctccagtcgc
 aagcttcagagcctattactaagggtagagctacctatgtcaagagttttcaaggtacct
 aagctacatgtgatagtcggcaagccat
 gaacgcagaccgtgaacggtgatgtaaaccgggatagacgccaagcgtgccgtgtcaa
 tgacgctagatacacctcgatttacgta
 gagtgaatgccagccaatggagtcatgcacataacccgcttagactctgctcggggcgat
 acccgatcgcagaggcagagccgctta
 aacgcgatcgcggtaacctgtaatcagagccagcgctcgatgaattgcatcatggaagcc
 attgatgtggaatgttgagcgataacaa
 cacgaattgaagacgacattgacttgcttcaagtgagtggagaattgccgggcagacaag
 ataggtaggctcttggtgcgctgtcacat caatccattcctttcctctgttcaatctctatgttgacattctgatagggatcatigga
 tgccaatgcaaagaacatgagagtgtggtctgc
 attcaagtatcctggctgtaagctgtggccatgggcgctgcggtaaggtaaatcgcgat
 gactaatcagtcctcggtgactctggggcg
 gtagaggcagtgctgtaaccaaagcttgagccgagggcaaaaacaacggcgcacaaac
 aatcaacgaaagcatcgtcaacagt gtctcttcccagtcattacttcgaaaaccttctcgatagaaccttcagacgatgaac
 aggccacgcaaccgtcagccgcgcccccc
 aggacagactcagcggcgggaggcagatcgtcacaccttggtcgacgagctc.

EXAMPLE 7 Analysis of Select Singly- Substituted CBH2 Variants in Activity and
 Thermostability Assays [247] Protoplasts of a six-fold deleted *T. reesei* strain (*Aegl1*,
 Δ *egl2*, Δ *egB*, *Acbh1*, *Acbh2*, *Abg1*) were transformed with selected CB H2 SEL variants

encoded on the pTTTpyrG vector as described (WO 2009/048488). Spores were harvested, replated on acetamide agar, and incubated at 28°C for 7 days. Next, CBH2 variants were produced by inoculating 25 ml of YEG culture (5 g/L yeast extract, 20 g/L glucose) with a sporulating *Trichoderma* strain on agar, followed by a two day incubation period at 28 °C, 200 rpm. Five ml of the YEG culture was used to inoculate 45 ml of glycine medium containing a 2% glucose sophorose mixture. The culture was dispensed over a 6- well microtiterplate and incubated stationary for five days at 28°C in an oxygen-enriched atmosphere. Cells were removed by filtration through a 0.2 µm filter. Supernatants were subsequently concentrated using Vivaspinn spin cells with a 10 kD molecular weight cutoff. The variants were tested for their ability to hydrolyze phosphoric acid swollen cellulose (PASC) at 50 °C, dilute acid pretreated corn stover (PCS) at 50 °C, dilute ammonia pretreated corncob (CC) at 50 °C and 57°C. In addition, the variants were tested for thermostability.

[248] PASC 50° C. Enzyme activity on phosphoric acid swollen cellulose was examined essentially as described in Example 1, Section C, with the following changes. To each well, 10 µl was added containing 4.9 µg protein in supernatant from a CBH2 deleted strain (Δ egl1, Δ egl2, Δ cbhl, Δ cbh2). The plate was incubated at 50 °C while shaking at 200 rpm. After 2 hours the plate was put on ice for 5 min and 100 µl of 100 mM glycine pH 10.0 was added. After mixing, the plate was centrifuged at 3000 rpm for 5 min. A volume of 40 µl supernatant was diluted in 160 µl water. Ten µl of the diluted solution was transferred to a new 96-well microtiterplate (Costar Flat Bottom PS) containing 100 µl ABTS glucose assay mixture (2.74 mg/ml 2,2' azino-bis(3-ethylbenzo-thiazoline-6-sulfonic acid, 1 U/ml horseradish peroxidase type VI, 1 U/ml glucose oxidase) and increase in A₄₂₀ was recorded in a microtiterplate spectrophotometer (Spectramax Plus 384, Molecular Devices). A range of glucose concentrations was included as a standard

on each plate (0; 0.008; 0.016; 0.031; 0.063; 0.125; 0.25; 0.5; 1 mg/ml). Assays were done in duplicate. A dose response curve was generated for the wild-type CB H2 by fitting the data with a Temkin isotherm equation ($y = a + b(\ln(1 + c \cdot x))$) and the activities of the CB H2 variants were divided by a calculated activity of wild-type CB H2 of the same plate to yield a performance index.

[249] PCS 50 °C. Enzyme activity on washed dilute acid pretreated cornstover was examined essentially as described in Example 1, Section D, with the following changes. To each well, 10 µl was added containing 49 µg protein in supernatant from a CB H2 deleted strain (Δegl1 , Δegl2 , Δcbh1 , Δcbh2). The plate was incubated at 50 °C while shaking at 200 rpm. After 2 days the plate was put on ice for 5 min and 100 µl of 100 mM glycine pH 10.0 was added. After mixing, the plate was centrifuged at 3000 rpm for 5 min. A volume of 10 µl supernatant was diluted in 190 µl water. Ten µl of the diluted solution was transferred to a new 96-well microtiter plate (Costar Flat Bottom PS) containing 100 µl ABTS glucose assay mixture and assayed and analyzed as described above for the PASC assay. [250] Corncob 50 °C. Enzyme activity on corncob at 50 °C was performed essentially as described in Example 1, Section E, with the following changes. To each well 10 µl solution was added containing 46.55 µg protein of supernatant from a CB H2 deleted strain (Δegl1 , Δegl2 , Δcbh1 , Δcbh2), supplemented with 4.90 µg *T. reesei* CBH1, 6.84 µg *T. reesei* Xyn2 Y5 (Xiong et al, *Extremophiles* 8:393-400, 2004), 2.28 µg *Fusarium verticillioides* (Fv) 51A, 5.32 µg Fv3A, 0.76 µg Fv43D, and 2.45 µg *T. reesei* BGL1. The *Fusarium verticillioides* enzymes have been described in US 61/245269 (herein incorporated by reference for the teaching of this method). Different volumes of supernatant of a transformed *T. reesei* strain expressing a CBH2 variant were added as described in Example 1. The plates were incubated for two days at 50 °C with shaking at 200 rpm. [251] Corncob 5TC. Corn cob was ground to

pass through a 0.9 mm screen then pretreated as described in Example 4 of WO 2006/110901 (herein incorporated by reference for the teaching of this method). Pretreated corn cob was used as a 7% cellulose suspension in 50 mM sodium acetate pH 5.0. Seventy microliters of the suspension was added per well to a 96-well microtiterplate (Nunc Flat Bottom PS). To each well 10 μ l solution containing 46.55 μ g protein of supernatant from a CBH2 deleted strain (Δ egl1, Δ egl2, Δ cbh1, Δ cbh2), 4.90 μ g CBH1 variant

(S8P/T41I/N89D/S92T/S113N/S196T/P227L/D249K/T255P/S278P/E

295K/T296P/T332Y/ V403D/S411F/T462I), 6.84 μ g *T. reesei* Xyn2 Y5 (Xiong et al, Extremophiles 8:393-400, 2004), 2.28 μ g Fv51A, 5.32 μ g Fv3A, 0.76 μ g Fv43D, 2.45 μ g *Talaromyces emersonii* beta- glucosidase were added. Up to twenty microliters of supernatant from *H. jecorina* cells expressing either wild-type CB H2 or a CBH2 variant was added. Compensating volumes of acetate buffer were added to make up for differences in total volume. The plate was incubated at 57°C while shaking at 200 rpm. After 2 days the plate was put on ice for 5 min and 100 μ l of 100 mM glycine pH 10.0 was added. After mixing, the plate was centrifuged at 3000 rpm for 5 min. A volume of 10 μ l supernatant was diluted in 190 μ l water. Ten μ l of the diluted solution was transferred to a new 96-well microtiterplate (Costar Flat Bottom PS) containing 100 μ l ABTS glucose assay mixture and assayed and analyzed as described above. [252]

CBH2 Thermostability Assay 2. To test the thermostability of the CB H2 variants, 50 μ l of supernatant was incubated in a PCR machine across a temperature range of 50-70 °C. The remaining activity was determined by incubating 20 μ l supernatant with 80 μ l 0.625 mg/ml cellotriose in 50 mM sodium acetate buffer pH5.0 with 0.0025% Tween80. After 1 hour, 40 μ l of 100 mM glycine pH10 was added. Twenty μ l was added to 80 μ l ABTS reagent (see above) and color development at OD420 was recorded for 5 min. A pseudo

melting temperature (T_m) was calculated by fitting the remaining activity at each temperature with the formula: $y = 1/(1+\exp(-(a*(1-b/c-\ln(c/b))+d*((1/c)-(1/b)))/e))$ in which a is the heat capacity ($\text{kcal/mol } ^\circ\text{K}^{-1}$), b is the melting temperature (K), c is the assay temperature (K), d is the enthalpy change ($\text{kcal/mol } ^\circ\text{K}^{-1}$), and e is the gas constant ($\text{kcal/mol } ^\circ\text{K}^{-1}$).

As determined by thermostability assay 2. The T_m of the reference CB H2 of SEQ ID NO:3 was 57.5°C under the test conditions.

[253] A category of mutations was devised, that is highly combinable mutations, to encompass mutations that have a $PI > 0.75$ for at least one property, and >0.05 for all properties. As illustrated in the succeeding examples, some variants having two highly combinable mutations, have activity and/or stability that is greater than the reference CB H2 enzyme. An accumulation of highly combinable mutations can be used to make a beneficial variant where each individual mutation is not sufficient to confer superior activity, thermostability and/or expression levels to a variant enzyme.

EXAMPLE 8 CBH2 Combinatorial Library Variants and Activities Thereof

[254] A synthetic CBH2 combinatorial library was produced by GeneOracle (Mountain View, CA). Table 8-1 lists the possible substitutions of members of the CB H2 combinatorial library (numbered according to the CB H2 mature amino acid sequence). This library was created by combining substitutions classified as up mutations and/or highly combinable mutations on the basis of their T_m and performance as listed in Table 7-1.

The library was received from the above-mentioned provider as purified PCR products in which primers GACCGGACGTGTTTTGCCCTTCAT (SEQ ID NO:20) and GTGTGACCGGCTTTGGCGAGTG (SEQ ID NO:21) were used to amplify the *cbh2* gene flanked upstream by about 1.1 kb of the *cbh1* promoter and downstream by about

1.85 kb of the amdS marker for forced integration in the pyr2 locus in the H. jecorina host strain. A schematic of the homologous recombination of the expression cassette into the screening strain is shown in figure 11. The nucleotide sequence of a PCR fragment (partial cbhl promoter, cbh2 gene, and partial amdS gene) amplified from pTTTpyrG-CBH2 using the primers of SEQ ID NO:20 and SEQ ID NO:21, is provided below as

SEQ ID NO:22: gaccggacgtgtttgcccttcatttggagaaataatgtcattgcgatgtgtaattgcc
tgcttgaccgactggggctgttcgaagcccg
aatgtaggattgttatccgaactctgctcgtagaggcatgttgtaatctgtgtcgggca
ggacacgcctcgaagggtcacggcaaggg
aaaccaccgatagcagtgtctagtagcaacctgtaaagccgaatgcagcatcactggaa
aatacaaaccaatggctaaaagtacata
agttaatgcctaaagaagtcataaccagcggctaataattgtacaatcaagtggctaaa
cgtaccgtaattgccaacggcttgtggggt
tgcagaagcaacggcaaagccccacttccccacgtttgttcttcactcagtccaatctc
agctggatcccccaattgggtcgcttgtt
gttcgggtgaagtgaagaagacagaggtgaagaatgtctgactcggagcgtttgcatac
aaccaagggcagtgatggaagacagt
aaatgttgacattcaaggagtatttagccagggatgcttgagtgtatcgtgtaaggaggt
ttgtctgccgatacgacgaatactgtatagt
cacttctgatgaagtggcatattgaaatgtaaagtcggcactgaacaggcaaaagatt
gagttgaaactgcctaagatctcgggccct
cgggccttcggccttgggtgtacatgtttgtgctccgggcaaagcaaaagtgtgtagg
atcgaacacactgctgccttaccagca
gctgagggtatgtgataggcaaatgttcaggggccactgcatggttcgaatagaaagag
aagcttagccaagaacaatagccgataa

agatagcctcattaacggaatgagctagtaggcaaagtcagcgaatgtgtatatataaa
ggttcgaggtccgtgcctccctcatgctct
ccccatctactcatcaactcagatcctccaggagactgtacaccatctttgaggcaca
gaaacccaatagtcacccatcacaaagttgt
acaaaaaagcaggctccgcggccgccccctcaccacccatgattgtcggcattctcacc
acgctggctacgctggccacactcgca
gctagtgtgcctctagaggagcggcaagcttgctcaagcgtctggttaattatgtgaaccc
tctcaagagacccaaatactgagatatgtc
aaggggccaatgtggtggccagaattggcgggtccgacttgctgtgcttccggaagcac
atgcgtctactccaacgactattactccc
agtgtctcccgcgctgcaagctcaagctcgtccacgcgcggcgctcgacgacttctc
gagtatccccacacatcccggtcga
gtccgcgacgcctccacctggttctactactaccagagtacctccagtcggatcgggaa
ccgctacgtattcaggcaaccctttgttg
gggtcactcctgggccaatgcatattacgcctctgaagttagcagcctcgctattccta
gcttgactggagccatggccactgctgcag
cagctgtcgcaaaggttccctctttatgtggctgtaggctctccggaaccaaggcaat
ctgttactgaaggctcatcttactgcaga
gatactcttgacaagacccctctcatggagcaaaccttgccgacatccgcaccgccaac
aagaatggcggtaactatccggacag
tttgtggtgtatgacttgccggatcgcgattgcgctgcccttgccctgaatggcgaatac
tctattgccgatggtggcgctgccaaatata
agaactatatcgacaccattcgtcaaattgtcgtggaatattccgatatccggacctcc
tggttattggTAtgagtttaaacacctgcct
cccccccccttcccttcttcccgccggcatcttgcgttgtgtaactattgttccc

tcttccagagcctgactctcttgccaacctgggtg
accaacctcgggtactccaaagtgtgccaatgctcagtcagcctaccttgagtgcataac
tacgccgtcacacagctgaaccttccaaa
tgttgcgatgtatttgacgctggccatgcaggatggcttggctggccggcacaaccaaga
ccggccgctcagctatttgcaaatgtt
acaagaatgcatcgtctccgagagctcttcgaggattggcaaccaatgtcgccaactaca
acgggtggaacattaccagcccccatc
gtacacgcaaggcaacgctgtctacaacgagaagctgtacatccacgctattggacctct
tcttgccaatcacggctggtccaacgcct
tcttcactgatcaaggctgatcgggaaagcagcctaccggacagcaacagtggggag
actggtgcaatgtgatcggcaccggat
ttgtattcgcccatccgcaaacactggggactcgttgcctggattcgtttgtctgggtca
agccaggcggcgagtgtgacggcaccag
cgacagcagtcgccacgattgactcccactgtgcgctcccagatgccttgcaaccggc
gcctcaagctggtgcttgggtccaagcc
tactttgtcagcttctcacaacgcaaaccatcgttctgtaaAagggtgggcgcgcc
gaccagcttctgtacaaagtgggtgatc
gcgccgcgcgccagctccgtgcgaaagcctgacgcaccggtagattcttgggtgagccgt
atcatgacggcggcgggagctacatg gccccgggtgatttattttttgtatctactctgacccttttcaatataggtcaac
tcatcttctactggagatgcggcctgcttggattg
cgatgtgtcagcttggcaaattgtggcttcgaaaacacaaaacgattccttagtagcc
atgcattttaagataacggaatagaagaaag
aggaaattaaaaaaaaaaaaaaaaacaaacatcccgttcataaccgtagaatcggcgtc
ttcgtgtatcccagtaccagtttatttgaa
tagctcgcccgtggagagcatcctgaatgcaagtaacaaccgtagaggctgacacggca

gggtgtgctagggagcgtcgtgttcta
caaggccagacgtcttcgcggtgatatatgtatgttgactgcaggctgctcagcga
cgacagtcaagttcgccctcgtgcttgtg
caataatcgcagtggggaagccacaccgtgactcccatcttcagtaaagctctgttgg
gtttatcagcaatacacgtaatttaaactcgt
tagcatggggctgatagcttaattaccgtttaccagtgccatgggtctgcagctttcctt
ggcccgtaaaatcggcgaagccagccaatc
accagctaggcaccagctaaaccctataattagctctttatcaacaccatccgctcccc
gggatcaatgaggagaatgagggggatg
cggggctaaagaagcctacataaccctcatgccaaactcccagttacactcgtcgagcca
acatcctgactataagctaacacagaatg
cctcaatcctgggaagaactggccgctgataagcgcgcccgccctcgcaaaaaccatccct
gatgaatggaaagtccagacgctgcct
gcggaagacagcgttattgatttcccaaagaaatcggggatccttcagaggccgaactg
aagatcacagaggcctccgctgcagatc
ttgtgtccaagctggcggccggagagttgacctcgggtggaagtacgctagcattctgta
aacgggcagcaatcgcccagcagttagt
agggtccctctacctctcagggaagtgaacaacgccaccttatgggactatcaagctg
acgctggcttctgtgcagacaaactgcgc
ccacgagttcttccctgacgccgctctcgcgcaggcaagggaactcgatgaatactacgc
aaagcacaagagacccgttgggtccact
ccatggcctccccatctctctcaaagaccagcttcgagtcaaggtaaccgttggcccta
agtcgttagatgtccctttttgtcagctaaca
tatgccaccagggttacgaaacatcaatgggctacatctcatggctaacaagtacgacg
aaggggactcgggtctgacaaccatgct

ccgcaaagccgggtgccgtctctacgtcaagacctctgtcccgagaccctgatggtctg
 cgagacagtcaacaacatcatcgggcg
 caccgtcaaccacgcaacaagaactggtcgtgcggcggcagttctggtggtgaggggtgc
 gatcgtgggattcgtggtggcgtcat
 cgggtgtaggaacggatatcgggtggctcgattcgagtgccggccggttcaacttcctgta
 cggctaaaggccgagtcatgggcggctg
 ccgtatgcaaagatggcgaacagcatggagggtcaggagacggtgcacagcgttgctggg
 ccgattacgcactctgttgagggtga gtccttcgcctcttctcttttctgctctataccaggcctccactgtcctcctttctt
 gctttttatactatatacgagaccggcagtcactga
 tgaagtatgtagacctccgcctcttcaccaaaccgtcctcgggtcaggagccatggaaa
 tacgactccaagggtcatcccatgccctg
 gcgccagtccgagtcggacattattgcctccaagatcaagaacggcgggctcaatatcgg
 ctactacaacttcgacggcaatgtccttc
 cacaccctcctatcctgcgcggcgtggaaaccaccgtcgccgcactcgccaaagccggtc acacc.

[255] Protoplasts of the AD5 *H. jecorina* strain (Aegl1, Δ egl2, Acbh1, Acbh2, Abgl1) described in Example 6 were transformed with the linear DNA library as described (US 2006/0094080) and grown on selective agar containing acetamide at 28°C for 7 days (0.6 g/L acetamide, 1.68 g/L CsCl, 20 g/L glucose, 6 g/L KH₂PO₄, 0.6 g/L MgSO₄·7H₂O, 0.6 g/L CaC₁₂·2H₂O, 0.5 g/L uridine, trace element salts, 10 g/L low melting point agarose). After 24 hours the agar was overlaid with selective agar supplemented with 1.2 g/L fluoroorotic acid (FOA). A total of 380 colonies were transferred to potato dextrose agar plates containing 1.2 g/L FOA and incubated at 28°C for 4-5 days. Spores were transferred to fresh potato dextrose agar plates, which were incubated at 28°C for 3 days. Alternatively, protoplasts of the MAD6 strain described in Example 6 can be employed instead of AD5 for expression of variant library members.

Likewise, protoplasts of derivatives of the MAD6 strain in which additional cellulases have been inactivated can be used for this purpose. Such derivatives would exhibit even less background cellulase activity. [256] For CBH2 variant protein production, spores were transferred using a 96-pin replicator to 200 μ l glycine minimal medium supplemented with 2% glucose/sophorose mixture in a PVDF filter plate: 6.0 g/L glycine, 4.7 g/L $(\text{NH}_4)_2\text{SO}_4$; 5.0 g/L KH_2PO_4 ; 1.0 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 33.0 g/L PIPPS; pH 5.5; with sterile addition of a 2% glucose/sophorose mixture as the carbon source, 10 ml/L of 100g/L of CaCl_2 , 2.5 ml/L of *T. reesei* trace elements (400X): 175g/L Citric acid anhydrous; 200g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$; 16g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$; 3.2 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$; 1.4 g/L $\text{MnSO}_4 \cdot \text{H}_2\text{O}$; 0.8 g/L H_3BO_3 . Each CBH2 variant was grown in quadruplicate. After sealing the plate with an oxygen permeable membrane, the plates were incubated at 28°C for 6 days, while shaking at 200 rpm. Supernatant was harvested by transferring the culture medium to a microtiter plate under low pressure. [257] A total of ten variants that showed improved activity on corn cob at 57°C were isolated. Genomic DNA of these strains was isolated and their *cbh2* gene sequences determined. The CBH2 variants were tested for properties of interest. The substitutions and activities of combinatorial library members on corncob and corn stover is shown in Table 8- 2. The specific activities for washed dilute acid pretreated cornstover (PCS 50 °C), for corncob at 50 °C (CC 50 °C), and for corncob at 57°C (CC 57 °C) were determined as described in Example 7.

EXAMPLE 9 CBH2 Singly and Doubly Substituted Variants and Activities Thereof

[258] Synthetic genes encoding CBH2 variants were constructed by GeneOracle (Mountain View, CA) in pTTTpyrG. Table 9-1 lists the single and double variants that were constructed (numbered according to the CBH2 mature amino acid sequence).

[259] The *cbh2* variant genes were received from the above-mentioned provider as purified PCR products in which primers GACCGGACGTGTTTTGCCCTTCAT (SEQ ID NO:20) and GTGTGACCGGCTTTGGCGAGTG (SEQ ID NO:21) were used to amplify the *cbh2* gene flanked upstream by about 1000 bp of the *cbh1* promoter and downstream by about 1000 bp of the *amdS* marker for forced integration in the *pyr2* locus in the *H. jecorina* host strain. The nucleotide sequence of a PCR fragment (partial *cbh1* promoter, *cbh2* gene, and partial *amdS* gene) amplified from pTTTpyrG-CBH2 using the primers of SEQ ID NO:20 and SEQ ID NO:21, is provided above in Example 8 as SEQ ID NO:2q22. [260] Protoplasts of the MAD6 *H. jecorina* strain (Aegl1, Aegl2, Aegl3, Acbh1, Acbh2, Abgl1) described in Example 6, were transformed with the linear DNA fragments as described (US 2006/0094080) and grown on selective agar containing acetamide at 28°C for 7 days as described. After 24 hours the agar was overlain with selective agar supplemented with 1.2 g/L fluoro-orotic acid (FOA). Colonies were transferred to potato dextrose agar plates containing 1.2 g/L FOA and incubated at 28°C for 4-5 days. Spores were transferred to potato dextrose agar plates, which were incubated at 28°C for 3 days.

[261] The variants were grown in microtiterplates as described in Example 8. For growth in shake flasks, 25 ml of YEG culture (5 g/L yeast extract, 20 g/L glucose) was inoculated with a sporulating *Trichoderma* strain on agar and incubated at 28°C at 200 rpm for 2 days. Next, 5 ml of the YEG culture was used to inoculate 45 ml of glycine medium containing a 2% glucose sophorose mixture in a shake flask. Following inoculation, the cultures were incubated in a resonant acoustic incubator (Applikon) at 28°C for 3 days. Cells were removed by filtration through a 0.2 µm filter. Supernatants were concentrated using Vivaspin spin cells with a 10 kD molecular weight cutoff.

[262] Samples grown in microtiter plate were analyzed for CBH2 production levels by HPLC as described in Example 1. Both microtiter plate and shake flask grown samples were tested for properties of interest. Table 9-2 lists the performance of the CB H2 variants for multiple properties of interest.

T_m = pseudo melting point temperature (°C) was determined by thermostability assay

2. IR = Performance index for stability as determined by thermostability assay 3. [263]

CBH2 Thermostability Assay 3. To analyze the thermostability of the CBH2 variants, 40 μ l of supernatant was incubated in a PCR machine 58°C in triplicate and samples were removed after 30, 60 and 120 min of incubation. The remaining activity was determined by incubating 10 μ l supernatant with 50 μ l of 1 mg/ml cellotriose in 50 mM sodium acetate buffer pH 5.0 with 0.0025% Tween20. After one hour, 40 μ l of 100 mM glycine pH 10 was added. Fifty μ l of 3X concentrated ABTS reagent (see above) was added and color development at OD420 was recorded for 3 min. The residual activity in time was fitted with a formula for exponential decay: $y = A_0 * \exp(-k * t)$, where A_0 is activity at $t=0$, t is time, and k is the decay constant. Performance indici (PI) were calculated by the following formula: $PI = -\log(k_{variant}) / -\log(k_{wt})$.

CLAIMS

We claim:

1. A cellulase variant, wherein the variant is a mature form having cellulase activity and comprising a substitution at one or more positions selected from the group consisting of: 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 and 447, wherein the positions are numbered by correspondence with the amino acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

2. The cellulase variant of Claim 1, wherein the variant comprises a further substitution at one or more further positions selected from:

(i) a first group consisting of 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 and 405; or

(ii) a second group consisting of 94, 98, 107, 120, 134, 144, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 291, 312, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, and 429,

wherein the further positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3.

3. A cellulase variant, wherein said variant is a mature form having cellulase activity and comprising a substitution at one or more positions selected from the group consisting of: 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 and 405, wherein the positions are numbered by correspondence with the amino

acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not a K/E, R/Q, N/D, Q/E, D/N or E/Q substitution.

4. The cellulase variant of Claim 3, wherein the variant comprises a further substitution at one or more further positions selected from:

(i) a first group consisting of 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 and 447; or

(ii) a second group consisting of 94, 98, 107, 120, 134, 144, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 291, 312, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, and 429,

wherein the further positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3.

5. A cellulase variant, wherein the variant is a mature form having cellulase activity and comprising a substitution at one or more positions selected from the group consisting of: 94, 98, 107, 120, 134, 144, 154, 178, 179, 206, 210, 214, 231, 232, 266, 272, 275, 291, 312, 316, 323, 343, 360, 380, 381, 386, 399, 410, 413, 416, 426, and 429, wherein the positions are numbered by correspondence with the amino acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not one of the group consisting of V94E, P98L, E107Q, M120L, M134G, M134L, M134V, L144G, L144R,

L144S, T154A, A178V, L179C, V206L, S210L, S210R, T214M, T214Y, G231P, G231I, G231A, G231N, G231S, G231T, T232V, H266S, H266A, W272A, W272D, W272Y, N275L, S316P, V323L, V323N, V323Y, G360R, S380T, A381T, E399N, E399D, R413Y, R413P, and A416G.

6. The cellulase variant of Claim 5, wherein the variant comprises a further substitution at one or more further positions selected from: (i) a first group consisting of 5, 18, 19, 28, 30, 32, 35, 38, 79, 80, 89, 100, 102, 103, 104, 105, 111, 117, 119, 121, 125, 126, 133, 137, 138, 139, 140, 141, 143, 150, 158, 162, 177, 180, 181, 182, 185, 186, 188, 190, 191, 192, 193, 196, 201, 207, 225, 226, 228, 229, 230, 233, 234, 236, 240, 243, 245, 251, 252, 258, 267, 268, 274, 292, 293, 303, 304, 306, 307, 313, 319, 322, 328, 331, 338, 340, 346, 361, 362, 363, 364, 365, 371, 384, 394, 396, 400, 406, 407, 414, 417, 422, 427, 431, 433, 436, 440, 441, 443, 444, 445 and 447; or

(ii) a second group consisting of 63, 77, 129, 146, 147, 151, 153, 157, 161, 189, 194, 197, 203, 204, 208, 211, 237, 239, 244, 247, 254, 277, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, 382 and 405,

wherein the further positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3.

7. The cellulase variant of any one of Claims 1-6, wherein the substitution at one or more positions is selected from the group consisting of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25 positions.

8. A cellulase variant, wherein the variant is a mature form having cellulase activity and comprising a substitution at from one to six positions selected from the group consisting of 111, 144, 154, 162, 410 and 413, wherein the cellulase variant comprises: leucine or serine at position 111; leucine, glutamine or tryptophan at position 144; threonine, cysteine or valine at position 154; tyrosine or asparagine at position 162; arginine or

serine at position 410; and serine, tryptophan or tyrosine at position 413; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not a S413Y substitution.

9. A cellulase variant, wherein the variant is a mature form having cellulase activity and comprising a substitution at from one to six positions selected from the group consisting of 98, 194, 313, 316, 384 and 443, wherein the cellulase variant comprises: proline or leucine at position 98; lysine, cysteine or glutamic acid at position 194; serine or threonine at position 313; serine or proline at position 316; glycine, cysteine or glutamine at position 384; and asparagine or isoleucine at position 443; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not one of the group selected from a P98L substitution, a K194E substitution, and a S316P substitution.

10. A cellulase variant, wherein the variant is a mature form having cellulase activity and comprising a substitution at from one to six positions selected from the group consisting of 153, 161, 203, 233, 422 and 444, wherein the cellulase variant comprises: arginine or glutamine at position 153; asparagine, alanine or tryptophan at position 161; arginine or histidine at position 203 ; proline or aspartic acid at position 233; glutamine or valine at position 422; and proline or glutamine at position 443; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase

variant consists of a single substitution, the single substitution is not a R153Q substitution.

11. A cellulase variant, wherein the variant is a mature form having cellulase activity and comprising a substitution at from one to seven positions selected from the group consisting of 98, 111, 144, 313, 316, 413 and 422, wherein the cellulase variant comprises: proline or leucine at position 98; leucine or serine at position 111; leucine or tryptophan at position 144; serine or threonine at position 313; serine or proline at position 316; serine or tryptophan at position 413; and glutamine or valine at position 422; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not one of the group selected from a P98L substitution, and a S316P substitution.

12. A cellulase variant comprising a glutamine at position 98, and a substitution selected from the group consisting of a T138C, a S316P, a S343Q, a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3.

13. A cellulase variant comprising a cysteine at position 138, and a substitution selected from the group consisting of a S316P, a S343Q, a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3.

14. A cellulase variant comprising a proline at position 316, and a substitution selected from the group consisting of a S343Q, a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CB H2) set forth as SEQ ID NO:3.

15. A cellulase variant comprising a glutamine at position 343, and a substitution selected from the group consisting of a Q362I, a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

16. A cellulase variant comprising an isoleucine at position 362, and a substitution selected from the group consisting of a S386C, a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

17. A cellulase variant comprising a cysteine at position 386, and a substitution selected from the group consisting of a C400S and a S406P, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

18. The cellulase variant of Claim 1, comprising a serine at position 400 and a proline at position 406.

19. A cellulase variant, wherein the variant is a mature form having cellulase activity and comprising a substitution at from one to seven positions selected from the group consisting of 98, 111, 182, 291, 316, 362 and 400, wherein the cellulase variant comprises: proline, leucine or glutamine at position 98; leucine or serine at position 111; asparagine or tryptophan at position 182; serine or a glutamic acid at position 291; serine or proline at position 316; glutamine, isoleucine or leucine at position 362; and cysteine or serine at position 400; wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein when the cellulase variant consists of a single substitution, the single substitution is not one of the group selected from a P98L substitution, and a S316P substitution.

20. The cellulase variant of Claim 19, wherein the variant further comprises a L439P substitution or a T74S substitution.
21. The cellulase variant of any one of Claims 1-20, wherein the variant is derived from a parent cellulase selected from the group consisting of *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, and *Cellulomonas fimi* CenA EG.
22. The cellulase variant of any one of Claims 1-21, wherein the variant is derived from a parent cellulase whose amino acid sequence is at least 75% 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a member of the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13.
23. The cellulase variant of any one of Claims 1-22, wherein the variant is isolated.
24. A composition comprising the cellulase variant of any one of Claims 1 to 23.
25. The composition of Claim 24, wherein the composition is enriched in the cellulase variant.
26. An isolated nucleic acid encoding the cellulase variant of any one of Claims 1 to 22.
27. An expression vector comprising the isolated nucleic acid of Claim 26 operably linked to a regulatory sequence.
28. A host cell comprising the expression vector of Claim 27.
29. A method for producing a cellulase variant, comprising culturing the host cell of Claim 28 in a culture medium, under suitable conditions to produce the cellulase variant.
30. A composition comprising the host cell of Claim 29 and culture medium.

31. A composition comprising the cellulase variant in supernatant of the culture medium of Claim 29.

32. The composition of Claim 24, further comprising at least one additional enzyme selected from the group consisting of a subtilisin, a neutral metalloprotease, a lipase, a cutinase, an amylase, a carbohydrase, a pectinase, a mannanase, an arabinase, a galactanase, a xylanase, an oxidase, and a peroxidase.

33. A method of cleaning or fabric care comprising contacting a surface or an article comprising a fabric with the composition of Claim 24 or 32.

34. A method for depilling and surface finishing a fabric comprising contacting a surface or an article comprising a fabric with the composition of Claims 24 or 32.

35. A method for converting biomass to sugars comprising contacting the biomass with the composition of Claims 24 or 32.

36. A method for producing a fuel comprising: contacting a biomass composition with an enzymatic composition comprising the cellulase variant of any one of Claims 1 to 22 to yield a sugar solution; and culturing the sugar solution with a fermentative microorganism under conditions sufficient to produce a fuel.

299

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
NIT : 800.176.089-2
- / -



RECIBO DE CAJA

No. 11 - 116518

Bogotá D.C., Noviembre 30 de 2011 - 13:12:09

RECIBIDO DE : CASTILLO PINEDA & ASOCIADOS

NI 800.253.423

*** Soporte del Pago ***

TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAGO	VR. PAGO
CONSIGNACION	BANCO DE BOGOTA	062754387	426539573	30/11/2011	32.302.000.00

*** Conceptos Pagados ***

CANT.	RENTISTICO	CONCEPTO	Vr. UNDITARIO	Vr. CONCEPTO
1	50005-01-01 SOLICITUDES	1 TRAMITES DE SOL. DE PATENTE DE INVENCIÓN	571.000.00	571.000.00
				\$571.000.00
=====				

SON: **QUINIENTOS SETENTA Y UN MIL PESOS MONEDA CORRIENTE***

Responsable: _____

Recibo de Caja Aplicado al Expediente No. _____

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-165857- -00000-0000

Fecha: 2011-12-01 16:35:55 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Evs: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 294

Act. 11 PRESENTACION Folios: 300

Sede Centro: Carrera 13 No. 27 - 00 Pisos 3,4,5 y 10 Bogotá, D.C.- Colombia

Web: www.sic.gov.co e-mail: Info@sic.gov.co Conmutador: (571) 5870000 Fax: (571) 5870284 Línea: 018000-910165 Call Center: (571) 6513240

Noviembre 30 de 2011 - 13:12:18

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
NIT : 800.176.089-2
- / -

Industria y Comercio
SUPERINTENDENCIA

RECIBO DE CAJA

No. 11 - 116519

Bogotá D.C., Noviembre 30 de 2011 - 13:12:09

RECIBIDO DE : CASTILLO PINEDA & ASOCIADOS NI 800.253.423

*** Soporte del Pago ***

TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAGO	VR. PAGO
CONSIGNACION	BANCO DE BOGOTA	062754387	426539573	30/11/2011	32.302.000.00

*** Conceptos Pagados ***

CANT.	RENTISTICO	CONCEPTO	Vr. LINDITARIO	Vr. CONCEPTO
26	50005-01-01 SOLICITUDES	1607 REIVINDICACION UNITARIA ADICIONAL A LAS 10 INICIALES	27.000.00	702.000.00
				\$702.000.00

SON: **SETECIENTOS DOS MIL PESOS MONEDA CORRIENTE**

Responsable:

Recibo de Caja Aplicado al Expediente No.

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-165857- -00000-0000

Fecha: 2011-12-01 16:35:55 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 294

Act. 411 PRESENTACION Folios: 300

Sede Centro: Carrera 13 No. 27 - 00 Pisos 3,4,5 y 10 Bogotá, D.C.- Colombia

Web: www.sic.gov.co e-mai: info@sic.gov.co Conmutador: (571) 5870000 Fax: (571) 5870284 Línea: 018000-910165 Call Center: (571) 6513240

Noviembre 30 de 2011 - 13:12:16



No. 11-165857-00000-0000

Fecha: 2011-12-01 16:35:55 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTE/III Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 294

LISTA DE CHEQUEO
ADMISIÓN A TRÁMITE - NUEVAS CREACIONES

Act. 411 PRESENTACION Folios: 300 //

PATENTE DE INVENCION ☐MODELO DE UTILIDAD ☐

Art 33 Decisión 486/00

☐	Indicación que se solicita una patente.
☐	Datos de identificación del solicitante o de la persona que presenta la solicitud
☐	Descripción de la invención
☐	Dibujos de ser estos pertinentes
☐	Comprobante de pago de las tasas establecidas (De ser el caso formato de descuento)
Completa ☐	Incompleta ☐

PATENTE DE INVENCION PCT ☒ MODELO DE UTILIDAD PCT ☐ Art.33 Decisión 486/00, Circular Única

☒	Indicación que se solicita una PCT
☒	Copia de la solicitud en español, tal como fue presentada inicialmente (capítulo descriptivo, reivindicatorio, resumen)
☐	Dibujos de ser estos pertinentes
☒	Comprobante de pago de las tasas establecidas (de ser el caso formato de descuento)
Completa ☒	Incompleta ☐

DISEÑO INDUSTRIAL ☐

(Art. 119 Decisión 486/00)

☐	Indicación que se solicita Diseño industrial
☐	Datos de identificación del solicitante o de la persona que presenta la solicitud
☐	Representación gráfica y fotográfica del Diseño industrial o muestra del material que incorpora el diseño
☐	Comprobante de pago de las tasas establecidas
Completa ☐	Incompleta ☐

ESQUEMA DE TRAZADO ☐

(Art. 92 Decisión 486/00)

☐	Indicación que se solicita un esquema de trazado
☐	Datos de identificación del solicitante o de la persona que presenta la solicitud
☐	Representación gráfica de un esquema de trazado
☐	Comprobante de pago de las tasas establecidas
Completa ☐	Incompleta ☐