

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 621 181**

51 Int. Cl.:

C12N 9/42 (2006.01)

C12N 15/52 (2006.01)

C12P 7/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.11.2009 PCT/US2009/065571**

87 Fecha y número de publicación internacional: **27.05.2010 WO10060056**

96 Fecha de presentación y número de la solicitud europea: **23.11.2009 E 09828362 (5)**

97 Fecha y número de publicación de la concesión europea: **18.01.2017 EP 2361311**

54 Título: **Levaduras que expresan celulasas para la sacarificación y fermentación simultáneas utilizando celulosa**

30 Prioridad:

21.11.2008 US 116981 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

03.07.2017

73 Titular/es:

**LALLEMAND HUNGARY LIQUIDITY
MANAGEMENT LLC (50.0%)
Nepfurdo utca 22, Bldg. B, 13th Floor
1138 Budapest, HU y
STELLENBOSCH UNIVERSITY (50.0%)**

72 Inventor/es:

**MCBRIDE, JOHN;
BREVNOVA, ELENA;
GHANDI, CHHAYAL;
MELLON, MARK;
FROELICH, ALAN;
DELEAULT, KRISTEN;
RAJGARHIA, VINEET;
FLATT, JIM;
VAN ZYL, EMILE;
DEN HAAN, RIAAN;
LAGRANGE, DANIE;
ROSE, SHAUNITA;
PENTTILA, MERJA;
ILMEN, MARJA;
SIIKA-AHO, MATTI;
UUSITALO, JAANA;
HAU, HEIDI HANSON;
RICE, CHARLES;
VILLARI, JEFF;
STONEHOUSE, EMILY A.;
GILBERT, ALAN;
KEATING, JEFFREY D.;
XU, HAOWEN;
WILLES, DEIDRE;**

ES 2 621 181 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 621 181**

72 Inventor/es:

**SHIKHARE, INDRANEEL;
THORNGREN, NAOMI;
WARNER, ANNE K. y
MURPHY, DAN**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 621 181 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Levaduras que expresan celulasas para la sacarificación y fermentación simultáneas utilizando celulosa

5 Antecedentes de la invención

10 La biomasa lignocelulósica es muy reconocida como una fuente prometedora de materia prima para la producción de combustibles y productos químicos renovables. El obstáculo principal que impide la producción más generalizada de energía a partir de materias primas de biomasa es la ausencia general de tecnología asequible para superar la recalcitrancia de estos materiales a la conversión en combustibles útiles. La biomasa lignocelulósica contiene fracciones de hidratos de carbono (por ejemplo, celulosa y hemicelulosa) que pueden convertirse en etanol. Para convertir estas fracciones, la celulosa y la hemicelulosa deben convertirse o hidrolizarse finalmente en monosacáridos; esto es, la hidrólisis que ha demostrado ser históricamente problemática.

15 Los procesos biológicamente mediados son prometedores para la conversión de energía, en particular, para la conversión de biomasa lignocelulósica en combustibles. Los esquemas de procesamiento de biomasa que implican la hidrólisis enzimática o microbiana normalmente implican cuatro transformaciones biológicamente mediadas: (1) la producción de enzimas sacarolíticas (celulasas y hemicelulasas); (2) la hidrólisis de componentes de hidratos de carbono presentes en biomasa pretratada en azúcares; (3) la fermentación de azúcares de hexosa (por ejemplo, glucosa, manosa y galactosa) y (4) la fermentación de azúcares de pentosa (por ejemplo, xilosa y arabinosa). Estas cuatro transformaciones se producen en una sola etapa en una configuración de proceso denominado bioprocesamiento consolidado (BPC), que se diferencia de otras configuraciones menos altamente integradas y que no implica una etapa de proceso especializada para la producción de celulosa y/o hemicelulosa.

25 El BPC ofrece la posibilidad de reducir costes y de aumentar la eficacia en comparación con los procesos que representan una producción de celulosa especializada. Los beneficios resultan en parte de evitar inversiones de capital, sustratos y otras materias primas e instalaciones asociadas con la producción de celulosa. Además, diversos factores apoyan la realización de tasas de hidrólisis más elevadas y por lo tanto un volumen de reactor reducido y una inversión de capital utilizando el BPC, incluyendo la sinergia entre microbios y enzimas y el uso de organismos termófilos y/o de sistemas de celulosa formando complejos. Además, probablemente los microorganismos celulófilos adherentes a celulosa compiten satisfactoriamente para productos de hidrólisis de celulosa con microbios no adherentes, por ejemplo, contaminantes, lo que podría aumentar la estabilidad de procesos industriales basándose en la utilización de celulosa microbiana. El progreso en el desarrollo de microorganismos que permiten el BPC se está realizando a través de dos estrategias: diseñando microorganismos celulófilos de origen natural para mejorar las propiedades relacionadas con el producto, tales como producción y título, y diseñando organismos no celulófilos que presenten altos rendimientos del producto y títulos para expresar un sistema de celulosa y hemicelulosa heteróloga que permita la utilización de celulosa y hemicelulosa.

40 Se requieren tres tipos principales de actividades enzimáticas para la degradación de celulosa nativa: el primer tipo son las endoglucanasas (1,4-β-D-glucano 4-glucanohidrolasas; EC 3.2.1.4). Las endoglucanasas cortan al azar en la cadena polisacáridica de celulosa de la celulosa amorfa, generando oligosacáridos de diversas longitudes y por consiguiente nuevos extremos de cadena. El segundo tipo son las exoglucanasas, incluyendo celodextrinasas (1,4-β-D-glucano glucanohidrolasas; EC 3.2.1.74) y celobiohidrolasas (1,4-β-D-glucano celobiohidrolasas; EC 3.2.1.91). Las exoglucanasas actúan de una manera procesiva sobre los extremos reductores o no reductores de cadenas polisacáridicas de celulosa, liberando bien glucosa (glucanohidrolasas) o celobiosa (celobiohidrolasa) como productos principales. Las exoglucanasas también pueden actuar sobre celulosa microcristalina, supuestamente desprendiendo cadenas de celosa de la estructura microcristalina. El tercer tipo son las β-glucosidasas (β-glucósido glucanohidrolasas; EC 3.2.1.21). Las β-Glucosidasas hidrolizan celodextrinas y celobiosa solubles en unidades de glucosa.

50 La levadura del panadero (*Saccharomyces cerevisiae*) sigue siendo el microorganismo preferido para la producción de etanol (Hahn-Hagerdal, B., et al., Adv. Biochem. Eng. Biotechnol. 73, 53-84 (2001)). Como atributos favorables de este microbio se incluyen (i) alta productividad a rendimientos casi teóricos (0,51 g de etanol producido/g de glucosa utilizada), (ii) alta tolerancia osmótica y al etanol, (iii) fuerza natural en procesos industriales y (iv) es generalmente reconocido como seguro (GRAS, siglas del inglés *being Generally Regarded As Safe*) debido a su gran asociación con la elaboración de vino y pan, y fabricación de cerveza. Adicionalmente, *S. cerevisiae* exhibe tolerancia a inhibidores normalmente encontrados en hidrolizados resultantes de pretratamiento de biomasa.

60 Un inconveniente principal de *S. cerevisiae* es su incapacidad de utilizar polisacáridos complejos, tales como celulosa, o sus productos de degradación, tales como celobiosa y celodextrinas. En un intento de abordar este problema, diversas celulasas heterólogas de fuentes bacterianas y fúngicas se han transferido a *S. cerevisiae*, permitiendo la degradación de derivados celulósicos (Van Rensburg, P., et al., Yeast 14, 67-76 (1998)), o cultivado en celobiosa (Van Rooyen, R., et al., J. Biotech. 120, 284-295 (2005)); McBride, J.E., et al., Enzyme Microb. Technol. 37, 93-101 (2005)). Sin embargo, niveles de expresión y de actividad específica normales de las celulasas expresadas de manera heteróloga en levaduras, son incluso insuficientes para permitir un crecimiento y una producción eficaz de etanol por levaduras en sustratos celulósicos sin enzimas añadidas externamente. Sigue

habiendo una necesidad significativa de mejorar la cantidad de actividad celulasa para conseguir el objetivo de obtener un sistema de bioprocesamiento consolidado (BPC) capaz de convertir, de un modo eficaz y asequible, sustratos celulósicos en etanol.

5 Otro inconveniente principal del uso de *S. cerevisiae* es que las celulasas que se producen externamente actúan óptimamente a una temperatura más alta que la temperatura a la cual actúa óptimamente *S. cerevisiae*. Por tanto, cada procesamiento debe realizarse en un proceso de dos etapas a dos temperaturas diferentes o puede seleccionarse una temperatura en la que ambos procesos actúen a algún grado, pero al menos uno de los procesos no se produzca a una eficiencia óptima.

10 Para abordar estas limitaciones, la presente invención proporciona la expresión heteróloga de combinaciones de celulasas heterólogas, de tipo silvestre y optimizadas con codones, en levaduras que permitan una producción eficiente de etanol a partir de fuentes de celulosa. La invención también proporciona la expresión de dichas celulasas heterólogas en levaduras termotolerantes y métodos de uso de dichas levaduras transformadas para la producción de etanol.

Breve descripción de la invención

20 La presente invención se refiere a células hospedadoras celulíticas. Las células hospedadoras de la invención expresan celulasas heterólogas y pueden producir etanol a partir de celulosa.

25 La invención proporciona una célula hospedadora de levadura termotolerante transformada que comprende (a) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una endoglucanasa; (b) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una β -glucosidasa; (c) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una celobiohidrolasa I; y (d) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una celobiohidrolasa II, en el que la célula hospedadora de levadura puede producir etanol cuando crece utilizando celulosa como fuente de carbono.

30 La invención también proporciona un cocultivo que comprende al menos dos células hospedadoras de levadura en el que (a) al menos una de las células hospedadoras comprende un primer oligonucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una endoglucanasa; (b) al menos una de las células hospedadoras comprende un segundo polinucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una β -glucosidasa; (c) al menos una de las células hospedadoras comprende un tercer polinucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una celobiohidrolasa I; (d) al menos una de las células hospedadoras comprende un cuarto polinucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una celobiohidrolasa II; en el que el primer polinucleótido, el segundo polinucleótido, el tercer polinucleótido y el cuarto polinucleótido no están en la misma célula hospedadora; y en el que el cocultivo puede producir etanol a partir de celulosa.

40 En algunas realizaciones particulares de la invención, la fuente de carbono de celulosa es celulosa insoluble, celulosa cristalina, celulosa procedente de lignocelulosa, de celulosa de madera dura, de celulosa hinchada con ácido fosfórico con o celulosa microcristalina, lodo de papel o forraje de maíz.

45 En algunas realizaciones, las células hospedadoras de la invención comprenden un polinucleótido heterólogo que comprende un ácido nucleico que codifica una primera celobiohidrolasa, un polinucleótido que comprende un ácido nucleico que codifica una endoglucanasa, un polinucleótido que comprende un ácido nucleico que codifica una β -glucosidasa y/o un polinucleótido que comprende un ácido nucleico que codifica una segunda celobiohidrolasa, como se reivindica en la reivindicación 1.

50 En algunas realizaciones, la celulasa, la endoglucanasa, la β -glucosidasa o la celobiohidrolasa es una celulasa, una endoglucanasa, una β -glucosidasa o una celobiohidrolasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* o de *Arabidopsis thaliana*.

60 En algunas realizaciones particulares, la celobiohidrolasa es una CBH1 de *H. grisea*, una CBH1 de *T. aurantiacus*, una CBH1 de *T. emersonii*, una CBH1 de *T. reesei*, una CBH2 de *T. emersonii*, una CBH2 de *C. lucknowense* o una CBH2 de *T. reesei*. En algunas realizaciones, el polinucleótido heterólogo comprende un ácido nucleico que codifica una celobiohidrolasa, que codifica una proteína de fusión que comprende una celobiohidrolasa y un módulo de unión a celulosa (MUC). En algunas realizaciones particulares, el MUC es el MUC de la CBH2 de *T. reesei*, el MUC de la CBH1 de *T. reesei*, o el MUC de la CBH2b de *C. lucknowense* o el MUC de la CBH1 de *Humicola grisea*. En algunas realizaciones particulares, el MUC está fusionado a la celobiohidrolasa mediante una secuencia enlazadora. En algunas realizaciones particulares, la célula hospedadora expresa una primera y una segunda celobiohidrolasa, en la

que la primera celobiohidrolasa es una fusión de la CBH1 de *T. emersonii* y el CBD (*Cellulose Binding Domain*, dominio de unión a celulosa), y la segunda celobiohidrolasa es una CBH2b de *C. lucknowense*.

5 En otras realizaciones particulares, la β -glucosidasa es una β -glucosidasa de *S. fibuligera*. En otra realización particular, la endoglucanasa es una endoglucanasa de *C. formosanus*. En otra realización en particular la endoglucanasa es una endoglucanasa de *T. reesei*, por ejemplo, EG2 de *T. reesei*.

10 En algunas realizaciones de la invención, al menos una o al menos dos de las celulasas están ligadas. En otras realizaciones de la invención, al menos una de las celulasas está secretada. En otra realización, al menos una de las celulasas está ligada y al menos una de las celulasas está secretada. En otra realización, todas las celulasas están secretadas.

En algunas realizaciones de la invención, el ácido nucleico que codifica una celulasa tiene codones optimizados.

15 En algunas realizaciones, la célula hospedadora puede ser una célula hospedadora termotolerante. En algunas realizaciones, la célula hospedadora es una célula hospedadora de *Issatchenkia orientalis*, *Pichia mississippiensis*, *Pichia mexicana*, *Pichia farinosa*, *Clavispora opuntiae*, *Clavispora lusitaniae*, *Candida mexicana*, *Hansenula polymorpha* o *Kluveryomyces*. Por ejemplo, en algunas realizaciones, la célula hospedadora es una célula hospedadora de *K. lactis* o de *K. marxianus*. En algunas realizaciones, la célula hospedadora termotolerante es una
20 célula hospedadora de *S. cerevisiae*, en la que la *S. cerevisiae* se selecciona por ser termotolerante.

En algunas realizaciones, la célula hospedadora puede ser una célula de levadura oleaginoso. En algunas realizaciones particulares, la célula de levadura oleaginoso es una célula de *Blakeslea*, *Candida*, *Cryptococcus*,
25 *Cunninghamella*, *Lipomyces*, *Mortierella*, *Mucor*, *Phycomces*, *Pythium*, *Rhodospiridium*, *Rhodotorula*, *Trichosporon* o de *Yarrowia*.

En algunas realizaciones, la célula hospedadora es una célula de *Saccharomyces cerevisiae*.

30 En algunas realizaciones particulares, la célula hospedadora puede producir etanol a partir de celulosa a temperaturas por encima de aproximadamente 30 °C, 35 °C, 37 °C, 40 °C, 42 °C, 45 °C o 50 °C.

En otra realización particular, la célula hospedadora puede producir etanol a una tasa de al menos aproximadamente 10 mg por hora por litro, al menos aproximadamente 30 mg por hora por litro, al menos aproximadamente 40 mg por
35 hora por litro, al menos aproximadamente 50 mg por hora por litro, al menos aproximadamente 60 mg por hora por litro, al menos aproximadamente 70 mg por hora por litro, al menos aproximadamente 80 mg por hora por litro, al menos aproximadamente 90 mg por hora por litro, al menos aproximadamente 100 mg por hora por litro, al menos aproximadamente 200 mg por hora por litro, al menos aproximadamente 300 mg por hora por litro, al menos aproximadamente 400 mg por hora por litro, al menos aproximadamente 500 mg por hora por litro, al menos
40 aproximadamente 600 mg por hora por litro, al menos aproximadamente 700 mg por hora por litro, al menos aproximadamente 800 mg por hora por litro, al menos aproximadamente 900 mg por hora por litro, o al menos aproximadamente 1 g por hora por litro.

45 La presente invención también proporciona métodos para el uso de las células hospedadoras y cocultivos de la invención. La presente invención también se refiere a un método para hidrolizar un sustrato celulósico, que comprende poner en contacto dicho sustrato celulósico con una célula hospedadora de la invención o con el cocultivo de la invención de acuerdo con la reivindicación 22. Los métodos comprenden adicionalmente poner en contacto el sustrato celulósico con enzimas celulasas producidas externamente.

50 En algunas métodos particulares de la invención, el sustrato celulósico es una biomasa lignocelulósica seleccionada del grupo que consiste en hierba, pasto varilla, espartina, raigrás, alpiste, miscato, residuos de procesamiento del azúcar, bagazo de caña de azúcar, restos agrícolas, paja de arroz, cascara de arroz, paja de cebada, mazorca de maíz, paja de cereales, paja de trigo, paja de colza, paja de avena, cáscaras de avena, fibra de maíz, forraje, forraje de soja, forraje de maíz, restos forestales, fibra de pulpa de madera reciclada, lodo de papel, serrín, madera dura, madera blanda, Agave y combinaciones de los mismos.

55 En algunas realizaciones particulares de la invención, la célula hospedadora o el cocultivo de la invención producen etanol. El etanol puede producirse a una tasa de al menos aproximadamente 10 mg por hora por litro, al menos aproximadamente 30 mg por hora por litro, al menos aproximadamente 40 mg por hora por litro, al menos aproximadamente 50 mg por hora por litro, al menos aproximadamente 60 mg por hora por litro, al menos
60 aproximadamente 70 mg por hora por litro, al menos aproximadamente 80 mg por hora por litro, al menos aproximadamente 90 mg por hora por litro, al menos aproximadamente 100 mg por hora por litro, al menos aproximadamente 200 mg por hora por litro, al menos aproximadamente 300 mg por hora por litro, al menos aproximadamente 400 mg por hora por litro, al menos aproximadamente 500 mg por hora por litro, al menos aproximadamente 600 mg por hora por litro, al menos aproximadamente 700 mg por hora por litro, al menos
65 aproximadamente 800 mg por hora por litro, al menos aproximadamente 900 mg por hora por litro, o al menos aproximadamente 1 g por hora por litro.

En otros métodos particulares de la invención, la célula hospedadora o los cocultivos de la invención pueden ponerse en contacto con una sustancia celulósica a una temperatura de al menos aproximadamente 37 °C, de al menos aproximadamente 42 °C, de aproximadamente 42 °C a aproximadamente 45 °C, o de aproximadamente 42 °C a aproximadamente 50 °C.

5 Breve descripción de los dibujos

La Figura 1 muestra una imagen de un ensayo en placa con CMC (carboximetilcelulosa) para detectar actividad de endoglucanasa I en cepas de *K. lactis* (colonias numeradas 1-8) y de *K. marxianus* (colonias numeradas 9-16) transformadas con celulasas heterólogas. Las cepas 8 y 16 son controles negativos no transformados. La placa de la izquierda muestra el crecimiento de las colonias y la placa de la derecha muestra la actividad CMCasa, indicada por la presencia de una zona de aclaramiento. Las zonas de aclaramiento aparecen como manchas blancas en la imagen.

La Figura 2 representa los resultados de un ensayo con MU-lac (4-Metilumbeliferil-β-D-lactósido) para detectar actividad de CBH1 en cepas de *K. marxianus* transformadas con celulasas heterólogas.

La Figura 3 representa el porcentaje de Avicel convertido por diversas cepas de *K. marxianus* que expresan celulasas heterólogas.

La Figura 4 representa la producción/consumo de etanol a partir de Avicel por diversas cepas de *K. marxianus* que expresan celulasas heterólogas.

La Figura 5 representa el crecimiento de *S. cerevisiae* que expresa celulasas heterólogas en celulosa microcristalina bacteriana (CMCB).

La Figura 6 representa la producción de etanol a partir de Avicel por una cepa de *S. cerevisiae* que expresa celulasas heterólogas.

La Figura 7 representa la producción de etanol a partir de madera dura pretratada (5 % en base a un porcentaje de peso seco) por una cepa de *S. cerevisiae* que expresa celulasas heterólogas.

La Figura 8 representa la producción de etanol a partir de madera dura pretratada (5 % en base a un porcentaje de peso seco) por una cepa de *S. cerevisiae* que expresa celulasas heterólogas en presencia de diversas concentraciones de celulasas añadidas de manera exógena.

La Figura 9 representa la producción de etanol a partir de Avicel por MO288 (círculos) y de una cepa de control (triángulos) en medios tanto YP como YNB.

La Figura 10 representa el rendimiento de etanol a partir de Avicel (15 % en base a un porcentaje de peso seco) mediante un proceso a pequeña escala de sacarificación y fermentación simultáneas (SFS) utilizando *S. cerevisiae* complementada con celulasas externas. El rendimiento de una cepa de levadura que expresa celulasas heterólogas (MO288) se compara con un rendimiento de una cepa de control (MO249) a diversas concentraciones de celulasa externa durante 150 horas. (Una carga de celulasa del 100 % indica 25 mg/g de sólidos totales; la concentración de sólidos inicial fue del 15 %).

La Figura 11 representa el rendimiento teórico de etanol a partir de un proceso de sacarificación y fermentación simultáneas (SFS) utilizando *S. cerevisiae* complementada con celulasas externas. El rendimiento de una cepa de levadura que expresa celulasas heterólogas (MO288) se comparó con el de una cepa de control (MO249).

La Figura 12 ilustra los ahorros previstos de la enzima celulasa basándose en el rendimiento de etanol a las 168 horas del proceso de sacarificación y fermentación simultáneas (SFS).

La Figura 13 muestra la actividad de una celulasa artificial en el ensayo de conversión de Avicel como se describe en el Ejemplo 9. La cepa MO429 se transformó en la secuencia consenso de la CBH1 "consCBH1", y la cepa MO419 se transformó con el vector pMU451 vacío como control negativo. En la Tabla 8 del Ejemplo 9 se encuentran descripciones de otras cepas.

La Figura 14 demuestra la actividad de levaduras que expresan diversas combinaciones de enzimas CBH1 y CBH2 en Avicel como se describe en el Ejemplo 10.

La Figura 15 demuestra actividad de levaduras que expresan diversas enzimas celulasas en Avicel como se describe en el Ejemplo 10.

La Figura 16 representa la producción de etanol a partir de Avicel mediante un cocultivo de cinco cepas de *S. cerevisiae* que expresan celulasas heterólogas.

La Figura 17 representa la producción de etanol a partir de Avicel mediante un cocultivo de cuatro cepas de *S. cerevisiae* que expresan celulasas heterólogas, así como la producción de etanol a partir de la cepa MO288, que está expresando cuatro celulasas.

La Figura 18 representa la producción de etanol a partir de Avicel mediante un cocultivo de cuatro cepas de *S. cerevisiae* que expresan celulasas heterólogas en combinación con celulasa añadida externamente.

La Figura 19 representa el ahorro calculado de enzima utilizando un cocultivo de cuatro cepas de *S. cerevisiae* que expresan celulasas heterólogas o MO288 en comparación con *S. cerevisiae* no transformada.

La Figura 20 representa la utilización de xilosa y la producción de etanol de una reserva de refrigerador M0509, de aislado en YPX y de aislado en YPD.

La Figura 21 representa el crecimiento de M1105 (marcado como "colonia C2") y de MO1046 en presencia del mismo medio y de 8 g/l de acetato a 40 °C.

La Figura 22 representa la producción de etanol por M1105 (triángulos) y M1088 (cuadrados) en TS MS419 al 18 %. El experimento con M1105 tuvo una dosis enzimática mucho más baja del 10 % y la mitad de la densidad celular inoculada, pero produjo un título de etanol más alto. El experimento con MO1105 se realizó a 40 °C y el experimento con M1088 se realizó a 35 °C.

La Figura 23 representa la producción de etanol de M1105 en el que la fermentación se inoculó solo con 0,5 g/l DCW (*Dry Cellular Weight*, peso seco de células) y produjo alguna acumulación de azúcar y etanol 29 g/l.

La Figura 24 representa la producción de etanol de M1254 en condiciones con IFM convencional (círculo) y con IFM amonio bajo (cuadrado).

5 La Figura 25 representa la tasa de crecimiento específico de colonias sencillas en comparación con M1254 y M1339 en medio de xilosa complejo complementado con una mezcla de inhibidor sintético (que incluía acetato 8 g/l) a 40 °C. Las colonias sencillas se exploraron en las mismas condiciones en las que se produjo la evolución. La colonia C1 se renombró como M1360.

10 La Figura 26 representa el comportamiento de fermentación de M1360 a 40 °C en un medio de fermentación industrialmente relevante complementado con glucosa. La fermentación se inoculó con 60 mg/l peso seco de células de M1360.

La Figura 27 representa la producción de etanol de ejecuciones de SFS en PHW (*pretreated hardwood*, madera dura pretratada) (sólidos 18 %, MS149 sin lavar) a 35 °C y a 40 °C mediante diversas cepas. Todas las reacciones se cargaron con “zoomerasa” 4 mg/g (Novozyme 22c).

15 La Figura 28 representa cultivos aplicados en placas SC^{URA} que contienen 0,2 % de CMC o liquenina o β -glucano de cebada. Las dos filas superiores de cada placa eran cultivos basados en Y294, y las dos filas inferiores contenían cepas basadas en MO749. Los números indican el plásmido contenido en cada cepa. pMU471 contenía la EG de C.f. y sirvió como un control positivo. Las placas se incubaron durante 24 horas a 30 °C (ilustrado a la izquierda), después de lo cual las colonias se lavaron y las placas se tiñeron con rojo Congo al 0,1 % y se destiñeron con NaCl al 1 % (ilustrado a la derecha).

20 La Figura 29 representa un análisis SDS-PAGE de los sobrenadantes de cepas productoras de celulasa Cel5. Una cepa que contenía un plásmido con gen no exógeno se usó como cepa de referencia (REF). También se incluyó la cepa que contenía el plásmido pMU471 que expresaba la EG de C.f., la EG más satisfactoria previamente encontrada.

25 La Figura 30 representa la actividad de cepas que expresan las EG en (A) PASC (del inglés *Phosphoric acid swollen cellulose*, celulosa hinchada con ácido fosfórico,) (2 horas) y (B) Avicel 24 horas. Como cepa de referencia (REF) se usó una cepa que contenía un plásmido con gen no exógeno y la cepa que expresaba la EG de C.f. (pMU471) se incluyó como control positivo.

30 La Figura 31 representa la distribución de la capacidad de conversión de Avicel de sobrenadantes de levadura a partir de la transformación con EG2Tr y CBH1Te con CBDTr adicional. La conversión de M1088 se presenta como una línea vertical oscura, y las líneas de puntos que flanquean esta línea representan la desviación típica de la medición.

35 La Figura 32 representa la conversión de Avicel en el ensayo de Avicel HTP (momento a 48 horas) por sobrenadantes de cepas de levadura que expresan celulasa. M0509 es el control negativo que no expresa celulasas. La cepa 1088 es la cepa precursora que solo expresa CBH1, CBH2, y BGL, mientras que 1179, 1180, y 1181 son transformantes de 1088 que también expresan la EG2 de Tr.

40 La Figura 33 representa la producción de etanol en lodo de papel por BPC/SFS con la cepa celulolítica M1403 y con la cepa de fondo no celulolítica M1254 con diversas cantidades de complementación enzimática comercial. Condiciones experimentales: semicontinuo sólidos 30 %, inoculación celular 10 g/l, pH 5,5 y temperatura 40 °C, Zoom = preparación de celulasa Novozymes 22C, BGL = AB Enzymes preparación BGL EL2008044L, Xil = AB Enzymes preparación de xilanasa EL2006020L.

45 La Figura 34 representa la fermentación de dos tipos de lodos de papel mediante BPC por cepa de levadura (M1179) y una cepa de control M0509, que no expresan celulasas. Condiciones experimentales: sólidos 18 %, células cargadas a 10 o a 1 g/l, pH 5,5, Temperatura: 35 °C, 1 mg/g BGL y un 1 mg/g de Xil cargados. BGL = AB Enzymes preparación BGL EL2008044L, Xil = Enzymes AB preparación de xilanasa EL2007020L.

La Figura 35 representa el comportamiento de la cepa de levadura celulolítica M0963 y de la cepa de control no celulolítica (M0509) en sólidos de madera dura pretratada (PHW, *pretreated hardwood*) no lavada al 22 % (MS 149) a diversas concentraciones de celulasa externa. Condiciones experimentales: semicontinuo sólidos 22 %, pH 5,4, temperatura 35 °C, todas las proteínas enzimáticas (PE) eran “zoomerasa” (Novozymes 22C).

50 La Figura 36 representa el comportamiento de la cepa de levadura celulolítica M1284 en sólidos de madera dura pretratada lavada al 30 % a diversas cargas celulares iniciales. Condiciones experimentales: semicontinuo sólidos 30 %, pH 5,0, temperatura 35 °C, EP 4 mg = BGL 0,25 mg + Xilanasa 0,25 mg + Pectinasa 0,25 mg + Zoomerasa 3,25 mg, EP 20 mg = BGL 1 mg + Xilanasa 1 mg + Pectinasa 1 mg + Zoomerasa 16,7 mg. Zoomerasa = Preparación de celulasa Novozymes 22C, BGL = preparación de BGL AB Enzymes EL2008044L, Xil = preparación de xilanasa AB Enzymes EL2007020L, Pectinasa = FE Pectinasa Genencor Multifect.

55 La Figura 37 representa la producción de etanol en forraje de maíz lavado mediante BPC/SFS con la cepa celulolítica M1284 y la cepa de fondo no celulolítica M0509 con diversas cantidades de complementación enzimática comercial. Condiciones experimentales: semicontinuo sólidos 18 %, inoculación celular 10 g/l, pH 5,0 y temperatura 35 °C, BGL 1 mg/g y xilanasa 1 mg/g cargada en cada caso. BGL = preparación de BGL AB Enzymes EL2008044L, Xil = preparación de xilanasa AB Enzymes EL2007020L.

60 La Figura 38 representa la actividad en Avicel (A, B) o en MULac (C, D) de sobrenadantes de cultivo de levadura que expresan diferentes genes de CBH1, y una concentración de CBH1 estimada (mg/l, E, F) basándose en MULac. La cepa hospedadora fue Y294 o M0749. Los genes de CBH1 fueron: *Te*, *Talaromyces emersonii*; *Ct*, *Chaetomium thermophilum*; *At*, *Acremonium thermophilum*; *Tr*, *Trichoderma reesei*; *Hg*, *Humicola grisea*; *Ta*, *Thermoascus aurantiacus*. Las levaduras se cultivaron en YPD por triplicado durante 3 días. Los datos son la media \pm desviación típica.

65

La Figura 39 muestra los genes modificados en la cepa de levadura M0509.

La Figura 40 muestra las cepas de levadura utilizadas para construir M0509 y las modificaciones genéticas relevantes.

La Figura 41 muestra la genealogía de la cepa de levadura M1105.

5 La Figura 42 muestra la genealogía de la cepa de levadura M1254.

Descripción detallada de la invención

10 Los métodos y materiales desvelados son generalmente útiles en el campo de la modificación de levaduras mediante ingeniería genética.

Definiciones

15 Un "vector", por ejemplo, un "plásmido" o "YAC" (cromosoma artificial de levadura, *Yeast Artificial Chromosome*) se refiere a un elemento extracromosómico que a menudo lleva uno o más agentes que no forman parte del metabolismo central de la célula, y que normalmente está en forma de una molécula de ADN bicatenaria circular. Dichos elementos pueden ser secuencias que se replican de manera autónoma, secuencias que se integran en el genoma, secuencias de fagos o de nucleótidos, lineares, circulares o superenrolladas de un ADN o ARN mono o bicatenario, procedentes de cualquier fuente, en la que diversas secuencias de nucleótidos se han unido o recombinao en una única construcción que puede introducir un fragmento promotor y una secuencia de ADN para un producto génico seleccionado junto con una secuencia 3' no traducida apropiada en una célula. Preferentemente, los plásmidos o vectores de la presente invención son estables y auto-replicantes.

20 Un "vector de expresión" es un vector que puede dirigir la expresión de genes con los que está asociado operativamente.

25 El término "heterólogo" como se indica en el presente documento, se refiere a un elemento de un vector, plásmido o célula hospedadora que procede de una fuente distinta de la fuente endógena. Por tanto, por ejemplo, una secuencia heteróloga sería una secuencia que procede de un gen o plásmido diferente del mismo hospedador, de una cepa de célula hospedadora diferente, o de un organismo de un grupo taxonómico diferente (por ejemplo, de diferente reino, filo, clase, orden, familia, género o especie, o de cualquier subgrupo dentro de estas clasificaciones). El término "heterólogo" también se usa en el presente documento de manera sinónima con el término "exógeno".

30 El término "dominio", como se usa en el presente documento, se refiere a una parte de una molécula o estructura que comparte características físicas o químicas comunes, por ejemplo, propiedades o dominios hidrófobos, polares, globulares, helicoidales, por ejemplo, un dominio de unión a ADN o un dominio de ADN a ATP. Los dominios pueden identificarse por su homología con motivos estructurales o funcionales conservados. Como ejemplos de dominios de celobiohidrolasa (CBH) se incluyen el dominio catalítico (DC) y el dominio de unión a celulosa (DBC).

35 Un "ácido nucleico", "polinucleótido", o "molécula de ácido nucleico" es un compuesto polimérico que consta de subunidades ligadas por enlace covalente denominadas nucleótidos. Los ácidos nucleicos incluyen ácido polirribonucleico (ARN) y ácido poldesoxirribonucleico (ADN), pudiendo ser ambos mono o bicatenarios. El ADN incluye ADNc, ADN genómico, ADN sintético y ADN semisintético.

40 Una "molécula de ácido nucleico aislada" o un "fragmento de ácido nucleico aislado" se refiere a la forma polimérica de éster fosfato de los ribonucleósidos (adenosina, guanina, uridina o citidina; "moléculas de ARN") o de los desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina o desoxicitidina; "moléculas de ADN"), o cualquiera de sus análogos fosfoéster, tales como fosforotioatos y tioésteres, en forma monocatenaria, o en una hélice bicatenaria. Son posibles hélices bicatenarias de ADN-ADN, ADN-ARN y ARN-ARN. La expresión molécula de ácido nucleico, y en particular molécula de ADN o de ARN, se refiere únicamente a la estructura primaria y secundaria de la molécula, y no se limita a cualquiera de las formas terciarias en particular. Por tanto, esta expresión incluye ADN bicatenario encontrado, entre otros, en moléculas de ADN lineal o circular (por ejemplo, fragmentos de restricción), plásmidos y cromosomas. En el análisis de la estructura de moléculas de ADN bicatenario particular, pueden describirse secuencias en el presente documento de acuerdo con la convención normal de proporcionar únicamente la secuencia en la dirección de 5' a 3' a lo largo de la cadena de ADN no transcrita (es decir, la cadena que tiene una secuencia homóloga a la del ARNm).

45 Un "gen" se refiere a un conjunto de nucleótidos que codifica un polipéptido e incluye ácidos nucleicos de ADNc y de ADN genómico. Un "gen" también se refiere a un fragmento de ácido nucleico que expresa una proteína específica, incluyendo secuencias interviniendo (intrones) entre segmentos codificantes individuales (exones), así como secuencias reguladoras que preceden (secuencias no codificantes en 5') y siguen (secuencias no codificantes en 3') la secuencia codificante. Un "gen nativo" se refiere a un gen como se encuentra en la naturaleza con sus propias secuencias reguladoras.

60 Una molécula de ácido nucleico puede "hibridarse" con otra molécula de ácido nucleico, tal como un ADNc, ADN genómico o ARN, cuando una forma monocatenaria de la molécula de ácido nucleico puede emparejarse con la otra

molécula de ácido nucleico en condiciones de temperatura y fuerza iónica de la solución apropiadas. Las condiciones de hibridación y de lavado son muy conocidas y se ilustran, por ejemplo, en Sambrook, J., Fritsch, E. F. y Maniatis, T. MOLECULAR CLONING: A LABORATORY MANUAL, segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularmente en el Capítulo 11 y en la Tabla 11.1 de este documento (el no sucesivo en el presente documento "Maniatis"). Las condiciones de temperatura y fuerza iónica determinan la "rigurosidad" de la hibridación. Las condiciones de rigurosidad pueden ajustarse para explorar fragmentos moderadamente similares, tales como secuencias homólogas de organismos remotamente relacionados, a fragmentos muy similares, tales como genes que duplican enzimas funcionales a partir de organismos estrechamente relacionados. Los lavados posteriores a la hibridación determinan condiciones de rigurosidad. Un conjunto de condiciones utiliza una serie de lavados que comienza con SSC 6X, SDS al 0,5 % a temperatura ambiente durante 15 minutos, después se repite con SSC 2X, SDS al 0,5 % a 45 °C durante 30 minutos y después se repite dos veces con SSC 0,2X, SDS al 0,5 % a 50 °C durante 30 minutos. Para condiciones más rigurosas, los lavados se realizan a temperaturas más altas en las que los lavados son idénticos a los anteriores excepto que la temperatura de los dos lavados finales de 30 minutos en SSC 0,2X, SDS al 0,5 % se aumenta a 60 °C. Otro conjunto de condiciones muy rigurosas utiliza dos lavados finales en SSC 0,1X, SDS al 0,1 % a 65 °C. Un conjunto adicional de condiciones muy rigurosas se define por una hibridación en SSC 0,1X, SDS al 0,1 %, 65 °C y lavados con SSC 2X, SDS al 0,1 % seguido de SSC 0,1X; SDS al 0,1 %.

La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque dependiendo de la rigurosidad de la hibridación, son posibles emparejamientos erróneos entre bases. La rigurosidad apropiada para la hibridación de ácidos nucleicos depende de la longitud de los ácidos nucleicos y del grado de complementación, que son variables muy conocidas en la técnica. Cuanto mayor sea el grado de similitud u homología entre dos secuencias de nucleótidos, mayor será el valor de la T_f (temperatura de fusión) para híbridos de ácidos nucleicos que tienen estas secuencias. La estabilidad relativa (correspondiente a una T_f más alta) de hibridaciones de ácido nucleico disminuye en el siguiente orden: ARN:ARN, ADN:ARN, ADN:ADN. Para híbridos de más de 100 nucleótidos de longitud, las ecuaciones para calcular la T_f se han derivado (véase, por ejemplo, Maniatis en 9.50-9.51). Para hibridaciones con ácidos nucleicos más cortos, es decir, oligonucleótidos, la posición de los emparejamientos erróneos llega a ser más importante y la longitud del oligonucleótido determina su especificidad (véase, por ejemplo, Maniatis, en 11.7-11.8). En una realización, la longitud para un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Preferentemente, una longitud mínima para un ácido nucleico hibridable es de al menos aproximadamente 15 nucleótidos; más preferentemente de al menos aproximadamente 20 nucleótidos; y lo más preferentemente la longitud es de al menos 30 nucleótidos. Además, el experto en la materia reconocerá que la temperatura y la concentración salina de la solución de lavado pueden ajustarse según sea necesario de acuerdo con factores tales como la longitud de la sonda.

La expresión "porcentaje de identidad", como se conoce en la técnica, es una relación entre dos o más secuencias polipeptídicas o dos o más secuencias polinucleotídicas, determinada comparando las secuencias. En la técnica, "identidad" también significa el grado de relación de secuencia entre secuencias polipeptídicas o polinucleotídicas, según sea el caso, determinada por la coincidencia entre las cadenas de dichas secuencias.

Como se sabe en la técnica, la "similitud" entre dos polipéptidos se determina comparando la secuencia de aminoácidos del polipéptido, y los sustitutos de aminoácidos conservados en ella, con la secuencia de un segundo polipéptido.

La "identidad" y la "similitud" pueden calcularse fácilmente por métodos conocidos, incluyendo, pero sin limitación, los descritos en: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, NY (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, NY (1993); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, NJ (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, NY (1991). Se diseñan métodos preferidos para determinar la identidad para proporcionar la mejor coincidencia entre las secuencias ensayadas. Los métodos para determinar la identidad y similitud están codificados en programas informáticos disponibles al público. Los alineamientos de secuencias y los cálculos del porcentaje de identidad pueden realizarse utilizando el programa Megalign del paquete informático de LASERGENE bioinformatics (DNASTAR Inc., Madison, Wis.). Los alineamientos múltiples de las secuencias desveladas en el presente documento se realizaron utilizando el método de alineamiento Clustal (Higgins y Sharp (1989) CABIOS. 5:151-153) con los parámetros por defecto (PENALIZACIÓN POR HUECO=10, PENALIZACIÓN POR LONGITUD DE HUECO=10). Los parámetros por defecto para alineamientos por pares utilizando el método Clustel fueron KTUPLA 1, PENALIZACIÓN POR HUECO=3, VENTANA=5 y DIAGONALS SAVED=5.

Las secuencias adecuadas de ácidos nucleicos, o sus fragmentos (polinucleótidos aislados de la presente invención), codifican polipéptidos que son al menos aproximadamente de 70 % a 75 % idénticas a las secuencias de aminoácidos indicadas en el presente documento, al menos aproximadamente 80 %, 85 % o 90 % idénticas a las secuencias de aminoácidos indicadas en el presente documento, o al menos aproximadamente 95 %, 96 %, 97 %, 98 %, 99 %, o 100 % idénticas a las secuencias de aminoácidos indicadas en el presente documento. Los fragmentos de ácidos nucleicos adecuados son al menos aproximadamente 70 %, 75 % u 80 % idénticos a las secuencias de ácidos nucleicos indicadas en el presente documento, al menos aproximadamente 80 %, 85 % o

90 % idénticos a las secuencias de ácidos nucleicos indicadas en el presente documento, o al menos aproximadamente 95 %, 96 %, 97 %, 98 %, 99 %, o 100 % idénticos a las secuencias de ácidos nucleicos indicadas en el presente documento. Los fragmentos de ácidos nucleicos adecuados no solamente tienen las identidades/similitudes anteriores, sino que codifican típicamente un polipéptido que tiene al menos 50 aminoácidos, al menos 100 aminoácidos, al menos 150 aminoácidos, al menos 200 aminoácidos o al menos 250 aminoácidos.

Una "región codificante" de ADN o ARNA es una molécula de ADN o ARN que se transcribe y/o traduce en un polipéptido en una célula *in vitro* o *in vivo* cuando se coloca bajo el control de secuencias reguladoras apropiadas. Las "regiones reguladoras adecuadas" se refieren a regiones de ácido nucleico localizadas cadena arriba (secuencias no codificantes en 5'), dentro, o cadena abajo (secuencias no codificantes en 3') de una región codificante, y que influyen en la transcripción, procesamiento o estabilidad del ARN, o en la traducción de la región codificante asociada. Las regiones reguladoras pueden incluir promotores, secuencias líder de traducción, sitios de procesamiento de ARN, sitios de unión efectores y estructuras en bucle-tallo. Los límites de la región codificante se determinan mediante un codón de inicio en el extremo 5' (amino) y un codón de terminación de la traducción en el extremo 3' (carboxilo). Una región codificante puede incluir, pero sin limitación con, regiones procariotas, ADNc de ARNm, moléculas de ADN genómico, moléculas de ADN sintético, o moléculas de ARN. Si la región codificante está destinada para la expresión en una célula eucariota, normalmente una señal de poliadenilación y la secuencia de terminación de la transcripción estarán localizadas en 3' con respecto a la región codificante.

Una "isoforma" es una proteína que tiene la misma función que otra proteína, pero que está codificada por un gen diferente y puede tener pequeñas diferencias en su secuencia.

Un "parólogo" es una proteína codificada con un gen relacionado con duplicación dentro de un genoma.

Un "ortólogo" es un gen de una especie diferente que ha evolucionado de un gen ancestral común por especiación. Normalmente, en el curso de la evolución, los ortólogos conservan la misma función que la del gen ancestral.

"Fase de lectura abierta", abreviada como ORF (del inglés *Open Reading Frame*), significa una longitud de ácido nucleico, ya sea ADN, ADNc o bien ARN, que comprende una señal de inicio o un codón de inicio de la traducción, tal como un codón ATG o AUG, y un codón de terminación y puede traducirse posiblemente en una secuencia polipeptídica.

"Promotor" se refiere a un fragmento de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En general, una región codificante se localiza en 3' con respecto a un promotor. Los promotores pueden proceder en su totalidad de un gen nativo, o pueden estar compuestos por elementos diferentes procedentes de diferentes promotores encontrados en la naturaleza, o incluso comprender segmentos de ADN sintéticos. Los expertos en la técnica saben que distintos promotores pueden dirigir la expresión de un gen en distintos tipos de tejidos o células, o a diferentes fases del desarrollo, o en respuesta a diferentes condiciones ambientales o fisiológicas. Los promotores que hacen que un gen se exprese en la mayoría de tipos de células en la mayoría de los casos, reciben comúnmente el nombre de "promotores constitutivos". También se reconoce que, aunque en la mayoría de los casos los límites exactos de las secuencias reguladoras no se han definido por completo, los fragmentos de ADN de diferentes longitudes pueden tener la misma actividad promotora. Generalmente un promotor está limitado en su extremo 3' por el sitio de inicio de la transcripción y se extiende cadena arriba (dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Dentro del promotor se encontrará un sitio de inicio de la transcripción (definido convenientemente, por ejemplo, mapeando con nucleasa S1), así como dominios de unión a proteína (secuencias consenso) responsables de la unión de la ARN polimerasa.

Una región codificante está "bajo el control" de elementos de control transcripcionales y traduccionales en una célula cuando la ARN polimerasa transcribe la región codificante en ARNm, que después se corta y empalma en trans-ARN (si la región codificante contiene intrones) y se traduce en la proteína codificada por la región codificante.

"Las regiones de control transcripcionales y traduccionales" son secuencias reguladoras de ADN, tales como promotores, potenciadores, terminadores y similares, que proporcionan la expresión de una región codificante en una célula hospedadora. En las células eucariotas, las señales de poliadenilación son regiones de control.

La expresión "asociado operativamente" se refiere a la asociación de secuencias de ácido nucleico en un solo fragmento de ácido nucleico de tal manera que la función de una está afectada por la otra. Por ejemplo, un promotor está asociado operativamente con una región codificante cuando puede afectar a la expresión de esa región codificante (es decir, que la región codificante está bajo el control transcripcionales del promotor). Las regiones codificantes pueden asociarse operativamente a regiones reguladoras en orientación sentido o antisentido.

El término "expresión", como se usa en el presente documento, se refiere a la transcripción y acumulación estable de ARN en sentido (ARNm) o antisentido procedente del fragmento de ácido nucleico de la invención. La expresión también puede referirse a la traducción de ARNm en un polipéptido.

Células hospedadoras que expresan celulasas heterólogas

Para abordar las limitaciones de sistemas previos, la presente invención proporciona células hospedadoras que expresan celulasas heterólogas que pueden utilizarse de manera efectiva y eficiente para producir etanol a partir de celulosa. Las células hospedadoras pueden ser una levadura. De acuerdo con la presente invención, la célula hospedadora de levadura puede ser, por ejemplo, de los géneros *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces* y *Yarrowia*. Las especies de levadura como células hospedadoras pueden incluir, por ejemplo, *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus*, o *K. fragilis*. En algunas realizaciones, la levadura se selecciona del grupo que consiste en *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Candida albicans*, *Pichia pastoris*, *Pichia stipitis*, *Yarrowia lipolytica*, *Hansenula polymorpha*, *Phaffia rhodozyma*, *Candida utilis*, *Arxula adeninivorans*, *Debaryomyces hansenii*, *Debaryomyces polymorphus*, *Schizosaccharomyces pombe* y *Schwanniomyces occidentalis*. En una realización particular, la levadura es *Saccharomyces cerevisiae*. En otra realización, la levadura es una *Saccharomyces cerevisiae* termotolerante. Se considera que la selección de un hospedador apropiado se incluye en el ámbito de los expertos en la materia a partir de las enseñanzas del presente documento.

En algunas realizaciones de la presente invención, la célula hospedadora es una célula oleaginosa. De acuerdo con la presente invención, la célula hospedadora oleaginosa puede ser una célula levadura oleaginosa. Por ejemplo, la célula hospedadora de levadura oleaginosa puede ser de los géneros *Blakeslea*, *Candida*, *Cryptococcus*, *Cunninghamella*, *Lipomyces*, *Mortierella*, *Mucor*, *Phycomyces*, *Pythium*, *Rhodospiridium*, *Rhodotorula*, *Trichosporon* o *Yarrowia*. De acuerdo con la presente invención, la célula hospedadora oleaginosa puede ser una célula hospedadora de microalga oleaginosa. Por ejemplo, la célula hospedadora de microalga oleaginosa puede ser de los géneros *Thraustochytrium* o *Schizochytrium*. Por tanto, el biodiésel puede producirse a partir del triglicérido producido por los organismos oleaginosos utilizando procesos de transesterificación de lípidos convencionales. En algunas realizaciones particulares, las células hospedadoras oleaginosas pueden inducirse para secretar lípidos sintetizados. Las realizaciones en las que se utilizan células hospedadoras oleaginosas son ventajosas porque pueden producir biodiésel a partir de materias primas lignocelulósicas que, en relación con sustratos de semillas oleaginosas, son más asequibles, pueden crecer más densamente, mostrar emisiones más bajas de dióxido de carbono en el ciclo de vida y pueden cultivarse en terrenos marginales.

En algunas realizaciones de la presente invención, la célula hospedadora es una célula hospedadora termotolerante. Las células hospedadoras termotolerantes pueden ser particularmente útiles en procesos de sacarificación y fermentación simultáneas permitiendo que las celulasas se produzcan externamente y que las células hospedadoras productoras de etanol se comporten óptimamente en intervalos de temperatura similares.

Las células hospedadoras termotolerantes de la invención pueden incluir, por ejemplo, células hospedadoras de *Issatchenkia orientalis*, *Pichia mississippiensis*, *Pichia mexicana*, *Pichia farinosa*, *Clavispora opuntiae*, *Clavispora lusitanae*, *Candida mexicana*, *Hansenula polymorpha* y de *Kluyveromyces*. En algunas realizaciones, la célula termotolerante es una cepa de *S. cerevisiae*, u otra cepa de levadura, que se ha adaptado para crecer a altas temperaturas, por ejemplo, por selección para el crecimiento a altas temperaturas en un citostato.

En algunas realizaciones particulares de la presente invención, la célula hospedadora es una célula hospedadora de *Kluyveromyces*. Por ejemplo, la célula hospedadora de *Kluyveromyces* puede ser una célula hospedadora de *K. lactis*, *K. marxianus*, *K. blattae*, *K. phaffii*, *K. yarrowii*, *K. aestuarii*, *K. dobzhanskii*, *K. wickerhamii* *K. thermotolerans*, o *K. waltii*. En una realización, la célula hospedadora es una célula hospedadora de *K. lactis* o *K. marxianus*. En otra realización, la célula hospedadora es una célula hospedadora de *K. marxianus*.

En algunas realizaciones de la presente invención la célula hospedadora termotolerante puede crecer a temperaturas por encima de aproximadamente 30 °C, de aproximadamente 31 °C, de aproximadamente 32 °C, de aproximadamente 33 °C, de aproximadamente 34 °C, de aproximadamente 35 °C, de aproximadamente 36 °C, de aproximadamente 37 °C, de aproximadamente 38 °C, de aproximadamente 39 °C, de aproximadamente 40 °C, de aproximadamente 41 °C o de aproximadamente 42 °C. En algunas realizaciones de la presente invención la célula hospedadora termotolerante puede producir etanol a partir de celulosa a temperaturas por encima de aproximadamente 30 °C, de aproximadamente 31 °C, de aproximadamente 32 °C, de aproximadamente 33 °C, de aproximadamente 34 °C, de aproximadamente 35 °C, de aproximadamente 36 °C, de aproximadamente 37 °C, de aproximadamente 38 °C, de aproximadamente 39 °C, de aproximadamente 40 °C, de aproximadamente 41 °C, de aproximadamente 42 °C, de aproximadamente 43 °C, de aproximadamente 44 °C, de aproximadamente 45 °C, o de aproximadamente 50 °C.

En algunas realizaciones de la presente invención, la célula hospedadora termotolerante puede crecer a temperaturas de aproximadamente 30 °C a 60 °C, de aproximadamente 30 °C a 55 °C, de aproximadamente 30 °C a 50 °C, de aproximadamente 40 °C a 60 °C, de aproximadamente 40 °C a 55 °C o, de aproximadamente 40 °C a 50 °C. En algunas realizaciones de la presente invención, la célula hospedadora termotolerante puede producir etanol a partir de celulosa a temperaturas de aproximadamente 30 °C a 60 °C, de aproximadamente 30 °C a 55 °C, de aproximadamente 30 °C a 50 °C, de aproximadamente 40 °C a 60 °C, de aproximadamente 40 °C a 55 °C o de aproximadamente 40 °C a 50 °C.

En algunos métodos descritos en el presente documento, la célula hospedadora tiene la capacidad de metabolizar xilosa. En las siguientes publicaciones puede encontrarse información detallada en lo que respecta al desarrollo de tecnología utilizando xilosa: Kuyper M et al. FEMS Yeast Res. 4: 655-64 (2004), Kuyper M et al. FEMS Yeast Res. 5:399-409 (2005), y Kuyper M et al. FEMS Yeast Res. 5:925-34 (2005).

Por ejemplo, la utilización de xilosa puede realizarse en *S. cerevisiae* expresando de manera heteróloga el gen de la xilosa isomerasa, *XilA*, por ejemplo, a partir del hongo anaerobio *Piromyces sp. E2*, sobreexpresando cinco enzimas de *S. cerevisiae* implicadas en la conversión de xilulosa a productos intermedios glucolíticos (xiluloquinasa, ribulosa 5-fosfato isomerasa, ribulosa 5-fosfato epimerasa, transcetolasa y transaldolasa) y delecionando el gen GRE3 que codifica la aldosa reductasa para minimizar la producción de xilitol.

De acuerdo con los métodos descritos en el presente documento, las células hospedadoras pueden contener marcadores antibióticos o pueden contener marcadores no antibióticos.

Las células hospedadoras se modifican por ingeniería genética (se transducen o se transforman o se transfectan) con los polinucleótidos que codifican las celulasas de esta invención que se describen con más detalle adelante. Los polinucleótidos que codifican las celulasas pueden introducirse en la célula hospedadora en un vector de la invención, que puede ser, por ejemplo, un vector de clonación o un vector de expresión que comprende una secuencia que codifique una celulosa heteróloga. Las células hospedadoras pueden comprender polinucleótidos de la invención como copias integradas o copias de plásmidos.

En determinados aspectos, la presente invención se refiere a células hospedadoras que contienen las construcciones de polinucleótidos descritas más adelante. Las células hospedadoras de la presente invención pueden expresar uno o más polipéptidos de celulasas heterólogas. En algunas realizaciones, la célula hospedadora comprende una combinación de polinucleótidos que codifican celulasas heterólogas o fragmentos, variantes o derivados de los mismos. La célula hospedadora puede comprender, por ejemplo, copias múltiples de la misma secuencia de ácido nucleico, por ejemplo, para aumentar los niveles de expresión, o la célula hospedadora puede comprender una combinación de polinucleótidos únicos. En otras realizaciones, la célula hospedadora comprende un solo polinucleótido que codifica una celulosa heteróloga o un fragmento, una variante o derivado de los mismos. En particular, dichas células hospedadoras que expresan una sola celulosa heteróloga pueden utilizarse en cocultivo con otras células hospedadoras de la invención que comprenden un polinucleótido que codifica al menos otra celulosa heteróloga distinta, o un fragmento, una variante o derivados de los mismos.

La introducción de un polinucleótido que codifica una celulosa heteróloga, en una célula hospedadora, puede realizarse por métodos conocidos en la técnica. La introducción de polinucleótidos que codifican celulasas heterólogas, por ejemplo, en células hospedadoras de levadura pueden verse afectadas por transformación con acetato de litio, transformación con esferoplastos, o transformación con electroporación, como se describe en Current Protocols in Molecular Biology, 13.7.1-13.7.10. La introducción de la construcción en otras células hospedadoras puede efectuarse mediante transfección con fosfato de calcio, transfección mediada con DEAE-Dextrano o electroporación (Davis, L., et al., Basic Methods in Molecular Biology, (1986)).

Las células hospedadoras transformadas o los cultivos celulares, como se ha descrito anteriormente, pueden examinarse con respecto al contenido de las proteínas endoglucanasa, celobiohidrolasa y/o β glucosidasa. Para el uso de las celulasas heterólogas secretadas, el contenido de las proteínas puede determinarse analizando los sobrenadantes de las células hospedadoras (por ejemplo, levaduras). En determinadas realizaciones, puede recuperarse material de alto peso molecular del sobrenadante de las células de levadura, bien mediante precipitación con acetona o tamponando las muestras con cartuchos desalinizantes desechables. Las proteínas, incluyendo las celulasas heterólogas ligadas, también pueden recuperarse y purificarse de cultivos de células de levaduras recombinantes por métodos que incluyen preparación y lisis de esferoplastos, alteración celular utilizando perlas de vidrio y alteración celular utilizando, por ejemplo, nitrógeno líquido. Como métodos de purificación de proteínas adicionales se incluyen precipitación con etanol o con sulfato de amonio, extracción con ácidos, cromatografía de intercambio aniónico o catiónico, cromatografía con fosfato de celulosa, cromatografía de interacción hidrófoba, cromatografía de afinidad, cromatografía de hidroxiapatita, filtración en gel o cromatografía de lectina. Las etapas de plegamiento de proteínas pueden utilizarse si fuera necesario, para completar la configuración de la proteína madura. Finalmente, para las etapas de purificación final, puede emplearse cromatografía líquida de alto rendimiento (HPLC).

Los métodos de análisis de proteínas incluyen métodos tales como el método de Lowry tradicional o el método de ensayo de proteínas de acuerdo con el protocolo del fabricante BioRad. Utilizando dichos métodos, puede estimarse el contenido proteico de las enzimas sacarolíticas. Adicionalmente, para medir con precisión la concentración proteica, una celulosa heteróloga puede expresarse con una etiqueta, por ejemplo, una etiqueta de His o una etiqueta de HA y purificarse por métodos convencionales utilizando, por ejemplo, anticuerpos contra la etiqueta, una técnica de purificación con resina de níquel convencional, o una estrategia similar.

Las células hospedadoras transformadas o los cultivos celulares, como se ha descrito anteriormente, pueden analizarse además por hidrólisis de celulosa (por ejemplo, mediante un ensayo de detección de azúcar), para un tipo

de actividad celulasa particular (por ejemplo, midiendo la actividad de endoglucanasa, celobiohidrolasa o β -glucosidasa individual) o para la actividad celulasa total. La actividad de endoglucanasa puede determinarse, por ejemplo, midiendo un aumento de extremos reductores en un sustrato de CMC específico de endoglucanasa. La actividad de celobiohidrolasa puede medirse, por ejemplo, utilizando sustratos celulósicos insolubles tales como la celulosa hinchada con ácido fosfórico (PASC), un sustrato amorfo, o celulosa microcristalina (Avicel) y determinando el grado de hidrólisis del sustrato. La actividad β glucosidasa puede medirse por diversos ensayos, o utilizando, por ejemplo, celobiosa.

Una actividad celulasa total, que incluye a actividad de endoglucanasa, celobiohidrolasa y β -glucosidasa, puede hidrolizar sinérgicamente celulosa cristalina. Por tanto, la actividad celulasa total puede medirse utilizando sustratos insolubles que incluyen sustratos celulósicos puros tales como papel de filtro Whatman No. 1, borra de algodón, celulosa microcristalina, celulosa bacteriana, celulosa de algas y sustratos que contienen celulosa tales como celulosa teñida, alfa-celulosa o lignocelulosa pretratada. La actividad específica de las celulasas también puede detectarse por métodos conocidos por un experto habitual en la técnica, tales como el ensayo con Avicel (descrito anteriormente) que podría normalizarse por la concentración de proteína (celulasa) medida para la muestra.

Por tanto, un aspecto de la invención se refiere a la producción eficiente de celulasas para ayudar en la digestión de celulosa y en la generación de etanol. Una celulasa puede ser cualquier enzima implicada en la digestión, en el metabolismo y/o en la hidrólisis de celulasa, incluyendo una endoglucanasa, una exoglucanasa o una β -glucosidasa.

En realizaciones adicionales, las células hospedadoras transformadas o cultivos celulares se ensayan con respecto a la producción de etanol. La producción de etanol puede medirse por técnicas conocidas por un experto habitual en la materia, por ejemplo, mediante un método convencional de índice refractario por HPLC.

Celulasas heterólogas

De acuerdo con la presente invención, la expresión de celulasas heterólogas en una célula hospedadora puede utilizarse, de manera ventajosa, para producir etanol a partir de fuentes celulósicas. Las celulasas de diversas fuentes pueden expresarse de manera heteróloga para aumentar satisfactoriamente la eficiencia de la producción de etanol. Por ejemplo, las celulasas pueden proceder de fuentes tales como hongos, bacterias, plantas, protozoos o termitas. En algunas realizaciones, la celulasa es una celulasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowens*, e *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, o de *Arabidopsis thaliana*.

En algunas realizaciones de la invención, celulasas múltiples de un solo organismo se coexpresan en la misma célula hospedadora. En algunas realizaciones de la invención, celulasas múltiples de diferentes organismos se coexpresan en la misma célula hospedadora. En particular, celulasas de dos, tres, cuatro, cinco, seis, siete, ocho, nueve o más organismos pueden coexpresarse en la misma célula hospedadora. De manera similar, la invención puede incluir cocultivos de cepas de levadura, en el que las cepas de levadura expresan diferentes celulasas. Los cocultivos pueden incluir cepas de levadura que expresan celulasas heterólogas del mismo organismo o de organismos diferentes. Los cocultivos pueden incluir cepas de levadura que expresen celulasas de dos, tres, cuatro, cinco, seis, siete, ocho, nueve o más organismos.

Las celulasas de la presente invención incluyen tanto endoglucanasas como exoglucanasas. Las celulasas pueden ser, por ejemplo, endoglucanasas, β -glucosidasas o celobiohidrolasas.

En determinadas realizaciones de la invención, la endoglucanasa(s) puede ser endoglucanasa I o una isoforma, un parólogo u ortólogo de una endoglucanasa II. En algunas realizaciones, la endoglucanasa expresada por las células hospedadoras de la presente invención puede ser una endo-1,4- β -glucanasa recombinante. En realizaciones particulares, la endoglucanasa es una endoglucanasa de *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *R. speratus*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces*, *Irpex lacteus*, *C. lucknowense*, *C. globosum*, *Aspergillus terreus*, *Aspergillus fumigatus*, *Neurospora crassa* o de *Acremonium thermophilum*. En una realización particular, la endoglucanasa comprende una secuencia de aminoácidos seleccionada de las SEQ ID NO: 30-39 o 52-56, como se muestra en la Tabla 1 a continuación. En determinadas otras realizaciones, la endoglucanasa comprende la secuencia de aminoácidos que es al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99 o 100% idéntica a una secuencia de aminoácidos seleccionada de las SEQ ID NO: 30-39 o 52-56.

En la práctica, si cualquier polinucleótido es al menos 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o 100% idéntico a un polinucleótido de la presente invención, puede determinarse convencionalmente utilizando programas informáticos conocidos. Los métodos para determinar el porcentaje de identidad, como se indica con más detalle

más adelante en relación a la identidad de polinucleótidos, son también relevantes para evaluar la identidad de secuencia del polipéptido.

5 En una realización particular, la endoglucanasa es una endoglucanasa ("eg1") de *Trichoderma reesei*. En determinadas realizaciones, la endoglucanasa comprende una secuencia de aminoácidos al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99 o 100% idéntica a SEQ ID NO:39.

10 En otra realización particular, la endoglucanasa es una endoglucanasa de *C. formosanus*. En determinadas realizaciones, la endoglucanasa comprende una secuencia de aminoácidos al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99 o 100% idéntica a SEQ ID NO:31.

15 En otra realización particular, la endoglucanasa es una endoglucanasa de *H. jecorina*. En determinadas realizaciones, la endoglucanasa comprende una secuencia de aminoácidos al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99 o 100% idéntica a SEQ ID NO:54.

20 En determinadas realizaciones, la β -glucosidasa es una β -glucosidasa I o una isoforma, un parálogo o un ortólogo de β -glucosidasa II. En determinadas realizaciones de la presente invención, la β -glucosidasa procede de *Saccharomycopsis fibuligera*. En realizaciones particulares, la β -glucosidasa comprende una secuencia de aminoácidos al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100% idéntica a SEQ ID NO:40.

25 En determinadas realizaciones de la invención, la celobiohidrolasa(s) puede ser una celobiohidrolasa I y/o una isoforma, un parálogo o un ortólogo de celobiohidrolasa II. En una realización particular, la celobiohidrolasa comprende una secuencia de aminoácidos seleccionada de las SEQ ID NO: 21-29 o 46, como se muestra en la siguiente Tabla 1. En realizaciones particulares de la presente invención, la celobiohidrolasa es una celobiohidrolasa I o II de *Trichoderma reesei*. En otra realización, una celobiohidrolasa comprende una secuencia de al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100% idéntica a SEQ ID NO:27 o SEQ ID NO:28.

35 En otras realizaciones particulares de la presente invención la celobiohidrolasa es una celobiohidrolasa I o II de *T. emersonii*. En otra realización, la celobiohidrolasa comprende una secuencia de al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100% idéntica a SEQ ID NO:23 o SEQ ID NO:24.

40 En otra realización, la celobiohidrolasa de la invención es una celobiohidrolasa de *C. lucknowense*. En una realización particular, la celobiohidrolasa es celobiohidrolasa Cbh2b de *C. lucknowense*. En una realización, la celobiohidrolasa comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100% idéntica a SEQ ID NO:25.

45 En algunas realizaciones particulares de la invención, la celulasa comprende una secuencia seleccionada de las secuencias indicadas en la Tabla 1 a continuación. Las celulasas de la invención también incluyen celulasas que comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99 o 50 100% idéntica a las secuencias de la Tabla 1.

Algunas realizaciones de la invención incluyen un polipéptido que comprende al menos 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, o 500 o más aminoácidos consecutivos de cualquiera de las SEQ ID NO: 21-40, 46 o 52-56, o dominios, fragmentos, variantes o derivados de las mismas.

55

Tabla 1: Celulasas usadas en los Ejemplos 1-11 como se describe a continuación

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas		
<i>Humicola grisea</i> <i>cbh1</i>	<p>GAAATTCATGAGAACCGCTAAGTTCGCTACCTTGGCTGCCTTGGTTGCTCT GCTGCTGCTCAACAAGCCTGTTCTTGACTACTGAACGTCACCCATCTTTG TCTTGGAAACAAGTGTACTGCTGGTGGTCAATGTCAAAGTGTCCAAAGCCTCC ATCACTTTGGACTCTAATTGGAGATGGACCCACCAAGTCTCTGCTAGTACT AACTGTTACACCGGTATAAAGTGGACACTTCTATTGTACTGACGCTAA GTCTTGTGCTCAAAATTTGTTGTTGATGGTGTGATTACACCTCCACTTA TGGTATTACCACCAACGGTGACTCTTTGTCTTGAAGTTCGTTACTAAAGG TCAACATTCACCAACGTCGGTCTAGAACCTACTTAATGGACGGTGAAG ACAAAGTACCAACCTTCGAATTTGGGTAATGAATTAACCTTCGATGTCG ATGTGCTAACATCGGTTGTGTTTGAACGGTCTTATACTTCGTTTCTAT GGACGCCGACGGTGGTTTGTCTCGTTACCCAGGTAATAAGGCTGGTGCCA AGTATGGTACCGGTTACTGTGATGCTCAATGCCCAAGAGACATTAAGTTC ATCAACGGTGAAGCTAACATTTGAAGTTGGACTGGTCTACCAACGACCC AAACGCTGGCGCCGTAGATACGGTACCTGTTGTTCCGAAATGGACATTT GGGAAGCCCAACAACATGGCTACTGCTTTTACTCCACACCCATGTACCATC ATTGGTCAATCCAGATGTGAAGGTGACTCCTCTGTGGCGGTACCTACTCCAA CGAAAGATACGCTGGTGTGTGATCCAGACGGTGTGACTTCAACTCCTA CAGACAAGGTACAAGACTTCTATGGTAAGGTATGACTGTCGATACCA CCAAGAAAGATCACCGTCCGTCACCCAAATCTTGAAGGACGCTAACGGTGAT TTAGGTGAAATTAAGAATTTCTACGTCCAAAGATGGTAAGATCATCCCAA CTCTGAATCTACCAATTCAGGTGTTGAAGGTAATTCATCACTCAAGACTG GTGTGACAGACAAAAGGTTGCCCTTCGGTGATATTTGACGACTTCAACAGAA AGGGTGGTATGAAAGCAAATGGTAAGGCTTTGGCCGGTCCAATGGTCTTG GTTATGCTATTTGGGACGATCACGCTTCCAACATGTTGTGTTGGACTCC ACCTTCCCAGTTGATGCTGGTAAGCCAGGTGCCGAAAGAGGTTGCTTG TCCAACACTTCCGGTGTCCCAGCTGAAAGTTGAAGCCGAAGCTCCAAAT CTAACGTTGTCTTCTAACATCAGATTCGGTCCAAATCGGTTCCACAGTCCG</p>	<p>CAA35159 MRTAKFATLAALVASAAAQQA CSLTERHPSLSWVKCTAGGQC QTVQASITLDSNWRWTHQVSGS TNCYTGNKWDTSICTDAKSCAQ NCCVDGADYTSYGITTINGDSLS LKFVTKGQHSNTVGSRTYLMDG EDKYQTFELLGNEFTFDVDVSN GCGLNGALYFVSMADGGLSR YPGNKAGAKYGTGYCDAQPR DIKFINGEANIEGWTTSTNDPNA GAGRYGTCCSEMIDIWEANNMA TAFTPHPTIIGQSRCEGDSCGGT YSNERYAGVCDPDPDGFNSYRQ GNKTFYKGGMTVDTTKKITVVT QFLKDANGDLGEIKRFYVQDQK IIPNSETIPGVEGNSITQDWCDR QKVAFGDIDDFNRKGGMKQMG KALAGPMVLVMSIWDHASN LWLDSTFPVDAAGKPGAERGAC PTTSGVPAEVEAEAPNSNVVFSN IRFGPIGSTVAGLPGAGNGGNG GNPPPTTTTSSAPATTTTASAGP KAGRWQQCGGIGFTGTQCEEP YICTKLNWDWYSQCL (SEQ ID NO:21)</p>

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas	<p>CTGGTTTGCCAGGTGCTGGTAATGGTGGTAATAACGGTGGTAACCCACCA CCACCAACCACTACCACTTCTCTGCCCCAGCTACTACCACCCGCTTCT GCTGGTCCA AAGGCTGGTAGATGGCAACAATGTGGTGGTATTGGTTTCAC CGGTCCAACCCAATGTGAAGAACAATACATCTGTACCAAGTTGAACGACT GGTACTCTCAATGTTTATAACTCGAG (SEQ ID NO:1)</p>	

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p> <p><i>Thermoplasma aurantiacum</i> <i>cbh 1</i></p>	<p>GAATTCATGTACCAAGAGCTCTATTGTTCTCCTTCTTCTTGGCCGCCGCT AGAGCTCATGAAAGCCGGTACTGTACCCGCCGAAACACCCCATCCTTGAC TTGGCAACAATGTTCTCTGTTCTTGTACTACTCAAAACGGGAAGGT TGTTATTGACGCTAACTGGAGATGGGTTACACTACCTCCGGTTACACCAA CTGTTACACTGTAAACACTTGGGATACTTCCATCTGTCCAGACGACGTTAC CTGTGCTCAAAACTGTGCTTTGGACGGTGTGACTACTCCGGTACTTACGG TGTCACCTACTGGCAACGGTTGAGATTGAACTTCGTACCCAACTTTC TGGTAAGAACAATCGGTTCTAGATTGTACTTGTGCAAGACGATACTACTTA CCAAATCTTCAAGTTGTTGGTCAAGAGTTCACTTTCGACGTTGATGTTTC CAACTTGCCCTTGTGTTGAACGGTGTCTTGTACTTCTGTTGCTATGGACGC CGACGGTAACCTATCCAAGTACCCAGGTAACAAGCCGGTGCCAAAGTACG GTACCCGGTTACTGTGATTCTCAATGTCCAAGAGACCTAAAATTCATTAAACG GTCAAGCTAACGTCGAAGTTGGCAACCACTGTCTAACGATCCAAACGCC GGTGTGGTAATCACGGTTCTCTCTGTGTGCTGAAATGGACGTTTGGGAAGC TAACTCTATCTCACCGCCGTCACCTCCACATCCATGTGATAACCCAGGTCA AACCATGTGTCAAGGTGATGATTGTGGTGGTACCTACTCTTCCACTAGATA CGCTGGTACCTGTGACACCCGACGGTGTGATTTCACCCCATACCAACCCAG GTAAACCACTTTCTACGGTCCAGGTAAGATTGTGATACTTCTTCTAAGT TCACTGTGTCACTCAATTCATTACCGACGATGGTACCCCATCTGGTACCC TAACTGAAATTAAGAGATTCTACGTCCAAACGGTAAAGTCATTCCACAA TCCGAAAGCACCAATTCGGGTGTTACCGGTAACCTCCATCACCACTGAATAC TGTACCGCTCAAAAGGCCGCCCTTGTGACAAACACCGGTTTCTTCCACCCATGGT GGTTTGCAAAAGATTCTCAAGCCTTGGCTCAAGGTATGGTTTGGTCAATG TCCCTTGTGGGATGACCACCGTGTAAACATGTTGTGGTTGGATTCTACTTAC CCAACTGACGCTGATCCAGACACCCAGGTTGTGCTAGAGGTACTTGTCC AAACCACTTCTGGTGTCCAGCTGACGTCGAATCTCAAAACCCCTAACTCTTA CGTTATCTACTTAACATCAAGGTGGGTCCAAATTAACCTCCACCTTCACTGC</p>	<p>AAL83303</p> <p>MYQRALLFSFFLAAARAHEAGT VTAENHPSLTWQCSSGGSCFT QNGKVVIDANWRVWHTTSGYT NCYTGTNTWDTICPDDVTC AQN CALDGADYSPTYGVTTSGNALR LNFVTQSSGKNIGSRLYLLQDDT TYQIFKLLGQEFDFDVSNLPC GLNGALYFVAMDADGNLSKYP GNKAGAKYGTGYCDSQCPRDL KFINQANVEGWQPSANDPNAG VGNHGSSCAEMDVWEANSISTA VTPHPCDTPGQIMCQGDCCGGT YSSTRYAGTCDTDGCDFNYPQP GNHSFYGPGKIVDTSSKFTVVVQ FITDDGTPSGTTLTEIKRFYVQNG KVIPQSESTISGVTGNSITTEYCT AQKAAFDNTGFFTHGGLQKISQ ALAQGMVLVMSLWDDHAANM LWLDSTYPTDADPDTPGVARGT CPTTSVGPADVESQNPNSYVIYS NIKVGPINSTIFTAN (SEQ ID NO:22)</p>
	<p>TAACCTAACTCGAG (SEQ ID NO:2)</p>	

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p> <p><i>Talaromyces emersoni</i> <i>cbh1</i></p>	<p>GAATTCATGCTAAGAAGAGCTTACTATATGAGCTCTTCTGCTATCTTTGGCC GTTAAGGCTCAACAAGCCGGTACCGCTACTGCTGAAACACCCCTCCATT GACCTGGCAAGAATGTACCGCTCCAGGTTCTTGACCCACCCAAACCGGTG CTGTCTCTTGGACGCTAACTGGAGATGGGTCCACGACGTCACCGTTAC ACTAACTGTTACACCGGTAACACCTGGACCCAACTTACTGTCCAGACGA CGAAACTTGGCTCAAAACTGTGCCTTGGACGGTGTGACTACGAAGGTA CTTACGGTGTACCTCCTCTGGTCTTCTTGAAGTTGAACCTTCTGCTCACTGG TTCTAAACGTCGGTCCAGATTGATTTGTTGCAAGATGACTCCACTTACCA AATCTTCAAGTTGTTGAACAGAGAAATTTCTTTCGACGTCGATGTGTCCAA CTTGGCTTGTGTTTGAACGGTGTCTATACTTCTGTTGCTATGGACGCTGA TGGTGGTGTTCCTCAAGTACCCAAACAACAAAGGCTGTGCCAAAATACGGTA CTGGTACTGTGACTCTCAATGTCCACGCTGACTTGAAGTTTATGATGGTG AAGCTAATGTCGAAGTTGGCAACCATCTTCTAACAAACGCTAACACTGGC ATCGGTGACCAACGGTCTTGTGTGCGGAAATGGACGTTTGGGAAGCCAA CTCCATTTCCAAACCGGTCACCTCCACACCCATGTGACACTCCAGGTCAAAC TATGTTCCGGCGATGACTGTGGTGGTACTTACTCTAACGATAGATACGC TGGTACCTGTGATCCAGACGGTGGGACTTCAATCCATACAGAAATGGGTA ACACTTCTTTTACGGTCCAGGCAAGATCATCGACACTACTAAGCCATTCA CTGTTGTCACCCAAATCTTGAACCGGACGATGGTACTGATACCGGTACTTGT CCGAATCAAGAGATTCTACATCCAAACTTAACTGATACCCCAACCA AATTCCGACATCTCTGGTGTCACTGGTAACTCCATTACCCAGGAAATTTGT ACCGCCAAAGCAAGCTTTCGGTGTGACACCGGACGACTTCTCTCAACACAGG TGTTTTGGCTAAGATGGTGTGCTGCTATGCAACAAGGATGTTTGGTGTGAT GTCTTTTGGGACGACTACGCTGCTCAAAATGTTGTTGGTGGACTCCGATTA CCCAACCGATGCCGACCCAAACCCCTGGTATCGCTAGAGGTACCTGTG CAACTGACTCTGGTGTCCATCTGACGTCGAATCCCAATCTCCAAACTCCT ACGTCACTTACTCCAAACATTAATTCGGTCCAAATCAACTCCACTTTCACGT CTTCTTAACCTCGAG (SEQ ID NO:3)</p>	<p>AAL89553</p> <p>MLRRALLSSSAILAVKAQQAG TATAENHPPLTWQECTAPGSCIT QNGAVVLDANWRWVHDVNGY TNCYTGNWDPYCPDDETCAQ NCALDGADYEGTYGVTSSGSSL KLNFTGSNVGSRLLYLQDDST YQIFKLLNREFSFDVDVSNLPCG LNGALYFVAMDADGGVSKYPN NKAGAKYGTGYCDSQCPRDLKF IDGEANVEGWQPSSNNANTGIG DHGSCCAEMDVWEANSISNAV PHPCDTPGQTMCSGDDCCGGTYS NDRYAGTCDPDGCFNRYRMG NTSFYGPVKIITTKP FTVVVTFQFLTDDGTDITGTLSEIKR FYIQNSNVIPQPNSDISGVTGNSI TTEFCTAQKQAFGDTDDDFSQHG GLAKMGAAMQQGMVLVMSLW DDYAAQMLWLDSDYPTDADPT TPGIARGTCTDSDGVPDVEESQSP NSYVTVYSNIKFGPINSTFTAS (SEQ ID NO:23)</p>

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas <i>Talaromyces emersoni</i> <i>nii cbh2</i>	GAATTCATGCCGTAACCTTGTGGCCCTTGGCTCCAGCCGCTTGTGGTTGGT GCTGCCGAGCTCAACAATCCCTTGTGGGGTCAATGCCGGTGGTTCCTCCTG GACTGGTGCAACTTCCTGTGCCCGCTGGTGCCACCTGTTCACCCATTAAACC	AAL78165 MRNLLALAP AALLVGAAEAQQS LWGQCGGSSWTGATSCAAGAT

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p> <p>ATACTACGCTCAATGTGTTCCAGCCACTGCCACTCCAACTACCTTGACTAC CACCACTAAGCCAACTCCACCGGTGGTGTCTGCTCCAACTCCACCACCCAC CAACTACTACCGGTACTACCACCTCTCCAGTCGTCACCAGACCTGCCTCCG CCTCCGGTAATCCATTCCGAAGGTTATCAATGTACGCTAACCCCTTACTACG CTTCTGAAGTCAATTCCTTGGCTATCCCATCTTTGAGCTCCGAGTTGGTCCC AAAGGCCTCCGAAGTTGCTAAGGTCCCCTTCAATTTGTCTGGTTAGATCAAGC TGCCAAAGGTTCCATCTATGGGTGATTACTTGAAGGATATTCATCTCAAAA CGCTGCTGGTGTGATCCACCAATCGCCGGTATTTTCGTTGTTACGATTT GCCAGATAGAGACTGTGCCCGCGTCTTCTAAACGGTGAATTTTCTATCGC CAACAACGGTGTGCTTTATACAACAATATATCGAATTCATAGAGAAC AATTAACCACTTACTCCGACGTCCTACCATCTTGGTTATCGAACCCAGACT CTTTGGCTAACGTTGTCACTAACTTGAACGTTCCAAAATGTGCTAACGCTC AAGATGCTTACTTGGAAATGTATCAACTACGCTATTACCCAAATGGACTTGC CAAACGTTGCTATGTACTTGGACGCTGGTCAACGCCGGTTGGTTGGGTTGGC AAGCCAACTTGGCCCCAGCTGCTCAATATTCGCTTCTGTTACAAGAAGC CCTCTCCCAAGCCTCTGTTAGAGGTTTGGCTACCAACGTTGGCTAACCTACA ACGCCCTGGTCCATTTCTAGATGTCCATCCTACACTCAAGGTGACGCTAACT GTGATGAAGAAGATTACGTTAACGCTTTGGGTCCATTTGTTCCAAAGAACAA GGTTTCCCAAGCTTACTTCATCATCGACACTTCCCCTAACGGTGCAGACCA ACTAAGCAATCTCAATGGGGTGACTGGTGAACGTTATTTGGTACCCGGTTTC GGTGTAGACCAACCACCGACACTGGTAAACCCATTGGAAGACGCTTTCGT TTGGGTCAAGCCAGGTGGTGAATCCGACGTTACCTCCCAACTACTAGCC CACGTTACGATTACCACTGTGGTTTGTCTGACGCTTTTGCAACCCAGCTCCAG AAGCTGGTACCTGGTTCCAAAGCCTACTTCCGAACAATTTGTTGACTAACGCC AACCCATTGTTCTAACTCGAG (SEQ ID NO:4)</p>	<p>CSTINPYAQCVPATATPTLLT TTKPTSTGGAAPTTPPTTTGTTT SPVTRPASASGNPFEGYQLYAN PYYASEVISLAIPSLSSELVPKAS EVAKVPSFVWLDQAARKVPSMG DYLKDIQSONAAGADPPIAGIFV VYDLPDRDCAAAAANGFESIAN NGVALYKQYIDSIREQLTTYSDV HTLVIEPDSLAVVTLNLNPKC ANAQDAYLECYAITQLDLPNV AMYLDAGHAGWLGWQANLAP AAQLFASVYKKNASSPASVRGLA TNVANYNAWSISRCPSYTQGDA NCDEEDYVNALGPLFQEQQFPA YFIIDTSRNGVRPTKQSQWGDW CNVIGTGFVVRPTTDTGNPLEDA FVWVKPGGESDGTSTNTTSPRYD YHCGLSDALQPAPEAGTWFQAY FEQLLTNANPLF (SEQ ID NO:24)</p>	

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas		
Chryso- sporium luckno- wense CBH2b	<p>ATGGCCAAGAAAGTTGTTTCATTACCGCTGCCCTTAGCTGCCCGCAGTGGCTTGGCT GCACCAAGTATCGAAGAGAGACAAAATTGCGGAGCCGTCTGGACACAGT GCGAGGCAACGGCTGGCAAGGCCAACATGTTGTGCTTCTGGCTCAACG TGCGTGGCACAGAACGAGTGGTATTCCCAAGTGCCCTCCAAACTCCCAGGT GACTTCTTCAACAACCCCGCTCAACGTCTACTTCAACAGAGATCCACAA GTACCTCTTCTAGCACAAACCAGAAAGTGGCTCATCCTCATCTAGCAGTACG ACCCCTCCACCCCGTATCAAGTCCCTGTCCACGAGTATCCCCTGGCGGAGCAAC</p>	<p>MAKKLFTAALAAAVLAAPVIEE RQNCGAVWTQCGNGWQGPTC CASGSTCVAQNEWYSQCLPNSQ VTSSTPSSTSTSQRSSTSSSTTR SGSSSSSTTPPPVSSPVTSPGGA TSTASYSGNPFSGVRLFANDYYR</p>

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p>	<p>CTCAACAGCCAGTTATCCGGCAATCCCTTCTCTGGAGTGAGATTATTTGC AAACGACTATTATAGATCAGAGGTTCAACAACCTTGCAATTCCCTTCTATGAC GGAAACCCTAGCCGCAAAAGGCTTCCGCCGTAGCAGAAAGTCCCTAGTTTCC AATGGCTTGACAGAAACGTTACAATAGATACACTTATGGTACAGACTTTA TCTCAGGTTAGAGCTTTGAATAAGGCCGGTGCCAAACCCACCTTATGCTGCC CAATTAGTAGTCTATGACTTGCCAGATAGAGACTGTGCTGCCGACGCTTCT AATGGTGAATTTTCCATCGCAAAATGGCGGAGCTGCAAACTATAGATCATA CATTGATGCAATAAGAAACACATCATTTGAGTATTTGATATTAGAAATAA TCCTTGTGATTGAACAGACTCCATGGCTAATATGGTTACCAACATGAATG TAGCCAAGTGTCTAACGCAGCTTCCACATACCATGAGCTAACCCGTATAT GCATTAACAACAACCTGAATCTACCTAACGTTGCTATGTACTTAGATGCCGGT CATGCCGATGGTTGGGCTGGCCTGCAAAATATCCAACCCGCAAGCTGAATT GTTCCGCTGGAATCTACAACGACGCCGAAAGCCCGCTGCCGTTAGAGGCT TAGCCACAATGTTGCAAAATTACAACGCTTGGTCAATTGCTAGTGCCCTT CTTATACCTCACCAATCCTAACCTACGATGAGAAACATTACATAGAAAGCA TTTTCCCATTTGTTAACTCCGCTGGATTCCTGCCAGATTCATCGTGGAT ACCGGTAGAAACGGCAAAACAACCACTGGACAACAACAATGGGGAGATT GGTGTAAACGTCAAGGGAACCGGCTTCGGCGTCAGGCCTACGGCAACACACC GGACACGAGCTAGTCGACGCTTTTGTATGGGTTAAGCCAGGTGGCGAAAG TGACGGAACAAGTGACACGAGTGTGCAAGATACGATTACCACTGTGGTC TGTCCGACGCTTACAGCCCCCCCCGAGGCTGGACAATGGTTCCAGGCT TATTTTGAACAATTTGTTAACGAAACGCAAAATCCACCATTCTAA (SEQ ID NO:5)</p>	<p>SEVHNLAIPSMTGTLAAKASAV AEVPSFQWLDNRNVTIDTLMVQT LSQVRALNKAGANPPYAAQLVV YDLPDRDCAAAASNGEFSIANG GAANYRSYIDAIRKHIIEYSDIRII LVIEPDSMANMVTNMMNVAKCS NAASTYHELVYALKQLNLPNV AMYLDAGHAGWLGWPANIQPA AELFAGIYNDAGKPAAVRGLAT NVANYNAWSIASAPSYTSPNPN YDEKHYIEAFSPLLNSAGFPARFI VDTGRNGKQPTGQQQWGDWC NVKGTGFGVRPTANTGHELVDVA FVWVVKPGGESDGTSDTSAARYD YHCGLSDALQPAPEAGQWFQAY FEQLLTNANPPF (SEQ ID NO:25)</p>

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas		
<p><i>Talaromyces emersonii</i> cbh1 con CBD</p>	<p>ATGCTAAGAAGAGCTTTACTATTGAGCTCTTCTGCTATCTTTGGCCGTTAAG GCTCAACAAGCCGGTACCGCTACTGCTGAAAACCAACCCCTCCATTGACCTG GCAAGAATGTACCGCTCCAGGTTCTGTACCAACCCAAAACGGTGTCTGTCG TCTTGGACGCTAACTGGAGATGGGTCCACGACGTCACCGGTTACACTAAC TGTTACACCCGGTAAACACCTGGGACCCAACTTACTGTCCAGACGACGAAAC TTGCGCTCAAAACTGTGCCTTGGACGGTGTGACTACGAAAGGTACTTACG GTGTTACCTCCTCTGGTTCTTCCCTTGAAGTTGAACTTCGTCACCTGGTCTAA CGTCGGTCCAGATTGTAATTTGTTGCAAGATGACTCCACTTACCAAACTTT CAAGTTGTTGAACACAGAGAAATTTTCTTTCGACCGTCCGATGTGTCCAACTTGCC</p>	<p>MLRRALLSSSAILAVKAQQAG TATAENHPPLTWQECTAPGSCCT QNGAVVLDANWRWVHDTVNGY TNCYTGNTWDPPTYCPDDETCAQ NCALDGADYEGTYGVTSSGSSL KLNFTGNSNVGSRLLYLQDDST YQIFKLLNREFSFDVDVSNLPCG LNGALYFVAMDADGGVSKYPN NKAGAKYGTGYCDSQCPRDLKF</p>

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p>	<p>TTGTGGTTGAACGGTGTCTATACTTCGTTGCTATGGACGCTGATGGTGG TGTTTCCAAAGTACCCAAACAACAAGCTGGTGCCAAATACGGTACTGGTT ACTGTGACTCTCAATGTCCACGTGACTTGAAGTTTATTGATGGTGAAGCTA ATGTCGAAAGGTTGGCAACCATCTTCTAACAAACGCTAACACTGGCATTCCGGT GACCACGGTTCTTGTGTGCCGAAATGGACGTTTGGGAAAGCCAACTCCAT TTCCAAACGGCGTCACTCCACACCCATGTGACACTCCAGGTCAAACTATGTG TTCCGGCGATGACTGTGGTACTTACTCTAACGATAGATACGCTGGTAC CTGTGATCCAGACGGTTGCGACTTCAATCCATACAGAATGGTAACACTT CCTTTACGGTCCAGGCAAGATCATCGACACTACTAAGCCATTCACTGTG TCACCCAAATCTTGACCGACGATGGTACTGATACCGGTACTTTGTCCGAAA TCAAGAGATTCTACATCCAAAATCTAACGTCATCCCAACCAAAATTC GACATCTTGGTGTCACTGGTAACTCCATTACCACCGAATTTGTACCGCC CAAAAGCAAGCTTTCGGTGACACCGGACGACTTCTCTCAACACGGTGGTTT GGCTAAGATGGGTGCTGCTATGCAACAAGGTATGGTTTGGTCAATGCTTTT GTGGACGACTACGCTGCTCAAATGTTGTGGTTGGACTCCGATTACCCAA CCGATGCCGACCCAAACCACCCCTGGTATCGCTAGAGGTACCTGTCCAACT GACTCTGGTGTCCATCTGACGTCGAATCCCAATCTCCAAACTCCTACGTC ACTTACTCCAACATTAAATTCGGTCCAAATCAACTCCACTTTCACTGCTTCT AACCCCTCCAGGTGGTAACAGAGGTACTACCCTACTCTCGTAGGCCAGCTAC TACAACTGGTTCTTCCCCAGGCCAAACCCAAATCCCCTACGGTCAATGTGG TGGTATCGGTTACTCTGTGTCACCGTCTGTGCTTCTGGTACTACCTGTCA AGTTTAAACCCATACTACTCTCAATGTTTGTAG (SEQ ID NO:6)</p>	<p>IDGEANVEGWQPSSNNANTGIG DHGCCAEMDVVWEANSISNAVT PHPCDTPGQTMCSGDDCCGGTYS NDRYAGTCDPDGCDNFYRIMG NTSFYGPCKIIDTTKPFVVTQFL TDDGTDGTLSEIKRFYIQNSNVI POPNSDISGVTGNSITTEFTAQK QAFGDTDDFSQHGGLAKMGAA MQQGMVLVMSLWDDYAAQML WLDSDYPTDADPTTPGIARGTCP TDSGVPSDVESQSPNSYVYTSNI KFGPINSFTASNPPGGNRGTTT RRPATTTGSSPGPTQSHYGQCCG IGYSGPTVCASGTTCCQVLNPPYS QCL (SEQ ID NO:26)</p>

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p> <p><i>Trichoderma reesei</i> CBH1</p>	<p>ATGGTCTCCCTCACCTCCCTGCTGGCCGGCGTTGCCCCTATCTCTGGTGIC CTAGCAGCCCTGCGCAGAAAGTTGAACCTGTGCGCAGTTGAGAAACGTGA GGCCGAAGCAGAAAGCTCAATCCGCTTGTACCCCTACAATCCGAAACTCACCC CACCATTGACCTGGCAAAAAGTGTCTAGCGGTGGAACCTGTACTCAACAA ACTGGTCTGTGTTATCGACGCTAACTGGAGATGGACACACGCCACTAA CTC'TTCTACCAACTGTTACGACGGTAACACTTGGTCTTCCACTTTATGTCC AGATAACGAAACTTGTGCTAAGAAATTGCTGTTTGGACGGTGCCGCCCTACG CTTCTACCTACGGTGTACCACCTCCGGTAACTCCTTGTCTATTGGTTTCGT CACTCAATCCGCTCAAAAGAACGTTGGTGTAGATTGTACTTGTATGGCTTC TGACACTACTTATCAAGAAATTACTTTTGTGGGTAAACGAATTTTCTTTTCGA</p>	<p>ACCESSION NO.: CAA49596</p> <p>MVSFTSLLAGVAAISGVLAAPA AEVEPVAVEKREAEAEAQSACT LQSETHPPLTWQKCSSGGTCTQ QTGSVVIDANWRWTHATNSSTN CYDGNWSSSTLCPDNETCAKNC CLDGAAAYASTYGVTTSGNSLSIG FVTQSAQKNVGARLYLMASDIT YQEFTLLGNEFSFDVDVSQLPCG</p>

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p> <p>TGTTGACGTTTCCCAATTGCCATGTGGCTTGAACGGTGCCTTTGTACTTTGTCTCTATGGATGCTGACGGTGGTGTCTAAGTACCCAACTAACACTGCCCGGTGCTAAGTACGGTACTGGTACTGTGATTCCTCAATGTCCACGTGACTTGAA GTTCAATTAACGGTCAAGCCAAACGTCGAAGTTGGAAACCATCCTCCAACAACGCTAACACCCGATATCGGTGATCGGTGTCACGGTTCCTGTGTTCCGAAATGGACATCTGGGAAGCTAACAGTATTTCTGAAGCTTTGACACCCACACCCATGCACACTGTCCGGTCAAGAAATTTGTGAAGGTGATGGATGGTGGAAACCTACTCTGATAACAGATAACGGTGTACTTGTGACCCAGACGGTGTGACTGGAAACCCATACAGATTGGGTAACACTTCTTCTATGGTCCAGGTCTTCTTTTACCTTGGATACCACCAAGATTGACTGTTGTTACCCAAATTCGAAACTTCTGGTGCTATCAACAGATACTACGTTCAAACGGTGTACCTTCCAAACAACCAACGCTGAATTGGGTTCTTACTCTGTAAATGAATTGAACGACGACTACTGTACCCTGTAAGAAGCTGAATTTGGTGGTTCCTCTTCTCCGACAAGGGTGGTTTACCCAAATTCAGAAAGCTACCTCCGGTGGTATGGTTTTGGTTATGTCCCTTGTGGATGATTACTACGCAACATGTTATGGTTAGACAGTACTTACCCCAACTAACGAAACCTCCTACTCCAGGTGCTGCAGAGGTTCCCTGTCTACCTCTTCTGTGTTCCAGCTCAAGTTGAATCTCAATCTCCAACGCTAAGGTCAC TTTCTCCAACATCAAGTTCGGTCCAATCGGTTCCACTGGTAATCCATCTGGTGGAAACCCCTCCAGGTGTAACAGAGGTACTACCCTACTCGTAGGCCAGCTACTACAACCTGGTCTTCCCAAGGCCAAACCCAAATCCCACTACGGTCAATGTGGTGTATCGGTTACTCTGGTCCAAACCGTCTGTGTTCTTGGTACTACTCTGTCATCTCAATGTTGTAA (SEQ ID NO:7)</p>	<p>LNGALYFVSMADGGVSKYPTN TAGAKYGTGYCDSQCPRDLKFI NGQANVEGWEPSSNNANTGIGG HGSCSEMDIWEANSISEALTPH PCTTVGQEICEGDGCGGTYS DN RYGGTCDPDGCDWNPYRLGNTS FYGPGSSFTLDTTKKLTVVTFE TSGAINRYVYQNGVTFQQPNAE LGSYSGNELNDDYCTAEEAEFG GSSFSKGGLTQFKKATSGGMV LVMSLWDDYYANMLWLDSTYP TNETSSTPGAVRGSCSTSSGVPA QVESQSPNAKVTFSNIKFPGIGST GNPSGGNPPGGNRGTTTTRRPAT TTGSSPGPTQSHYGQCCGGIGYSG PTVCASGTTCCQVLNPPYYSQCL (SEQ ID NO:27)</p> <p>[SEÑAL DE SECRECIÓN: 1 – 33 DOMINIO CATALÍTICO: 41 – 465 DOMINIO DE UNIÓN A CELULOSA: 503 – 535]</p>	

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas		
Tri-choderma reesei CBH2	<p>ATGGTCTCCTTCACCTCCCTGCTGGCCGGGTTGCCGGCTATCTCTGGTGTCTAGCAGCCCTGCCGAGAAAGTTGAACCTGTCCGAGTTGAGAAACGTGAGGCCGAAAGCAGAGCTGTCCCATTAGAAGAAAGACAAGCCTGCTCCCTCTGTTTTGGGGTCAAATGTGGTGGTCAAAAACCTGGTCTGGTCCAACTTGTGTGCTTCCGGTCTACCTGTGTTACTCCAACGACTACTATTCCCAAATGTTTGCCAGGTGCTGCTTCCCTCTTCAACTAGAGCTGCTTCTACAACTTCTAGGGTCTCCCCAACCACTCCAGATCCTTCTGTCTACTCCACCACCCAGGTTCTACTACCACTAGAGTTCCACCGGTTCCGGTACTGCTACTTACTCTGGTAAACCTTTCGTGGTGTACTCCATGGGCTAACGCTTACTACGCTTCTGAAAGTTCTTCTTTGGCTATCCCATCTTTGACTGGTGTCTATGGCTACCCGCTGCTGCT</p>	<p>ACCESSION NO.: AAA72922AAA34210</p> <p>MIVGILTLATLATALAASVPLEE RQACSSVWGQCGGQNWSPGTC CASGSTCVYSNDYYSQCLPGAA SSSSSTRAAASSTTSRVSPSTSRSSS ATPPPGSTTRVPPVGSATATYS GNPFVGVTPWANAYYASEVSSL AIPSLTGAMATAAAA VAKVPSF</p>

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p>	<p>GCTGTCGCCAAAAGTTCCATCCTTCCATGTTGGTGGACACCTTGGACAAAACCTCCATTAATGGAACAACCTTGGCAGACATAAGGACTGCTAACAAAGAACGGTAACACTACGCTGGTCAATTTGTTGTACGACTTGGCAGACAGAGACTGTGCTGCTTTGGCTTCCAAACGGTGAATACTCCATCGCTGACGGTGGTGCGCCAAGTACAAGAACTACATTTGATACCAATTAGACAATACTGTTGTCCGAATACTCTGACATCAGAACCTTGTAGTCAATCGAACCCAGATTCTTTAGCCAAATTTAGTCAACCAACTTGGTACTCCAAAGTGTGCTAACGCTCAATCTGCCACTTAGAATGATCAATATGCAAGTTACCCAAATTTGAACCTTGCCAAACGTTGCTATGTAATGGACGCTGGTCAACGCTGGTGGTTGGGTTGGCCAGCTAACCAAGACCCAGCCGCTCAATATTCGCCAACGTTTACAAGAATGCCCTCTTCTCCTAGAGCTTGGCTTGGCTACTAACGTCGCTACTACAACGGTTGGAACATCACTTCTCCACCATCTTACACCCAAAGTAAACGCTGTTTACACGAAAGTTGTACATTCACGCTATCGGTCATTTATGGCTAACCATGGTTGGTCTAACCGCTTCTCATCACCGACCAAGGTAGATCCGGTAAACAACCAACTGGTCAACAACAATGGGGTGAATTTGTTAAACGTCATCGGTACTGGTTTCGGTATCAGACCAATCCGCTAACACTGGTGAATTCCTTGTGGATTCCCTTCGICTGGGTTAAGCCAGGTGGTGAATGTGATGGCACCTCTGATTCCTCTGCTCCAAAGATTCGATTCCTCACTGGCCTTGGCAGACGCTTTGCAACCCAGCCCAAGCTGGTGCAATGGTTCCAAAGCTTACTTTGTCCAAATTTGTTGACCAACGCTAACCCCATCTTCTTGTA (SEQ ID NO:8)</p>	<p>MWLDLTKPLMEQTLADIRTA NKNGGNYAGQFVVYDLPRDRC AALASNGEYSIADGGVAKYKNY IDTIRQIVVEYSDIRTLVIEPDSL ANLVTNLGTPKCANAQSAYLEC INYAVTQLNLPNVAMYLDAGHA GWLGWPANQDPAAQLFANVYK NASSPRALRGLATNVANYNGW NITSPPSYTQGNNAVYNEKLYIHAI GRLLANHGWSNAFFITDQGRSG KQPTGQQQWGDWCNVIGTGFGI RPSANTGDSLLDSFVWVKPGGE CDGTSDSSAPRFDSHCALPDALQ PAAQAGAWFQAYFVQLLTNAN PSFL (SEQ ID NO:28)</p>

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas		
<i>Chaetomium thermophilum</i> CBH1	TTAATTAAACAATGATGTACAAGAAATTTGCAGCCCTAGCTGCTTTAGTIG CAGGAGCTTCCGCTCAACAGGCATGTTTCATTGACTGCCGAAATCATCCA TCCTTAACGTGGAAGAGATGCACCGTCAGGAGGTTTCATGCTCCACTGTAAA CGGAGCTGTCACAATAGATGCAAAATTTGGAGATGGACCCACACTGTGTCGG GTAGTACA AACTGCTACACCGGTAATCAATGGGATACGTCCTTTGTGTACA GATGGAAAGTCATGCGCTCAGACCTGTTGCGTGGATGGAGCAGACTACTC TTCTACTTACGGAATCACGACATCAGGTGACAGTCTTAATTTGAAATTCGT AACCAAGCACCAAGTACGGAAACAATGTAGGCTCCAGAGTACTTAATGG AGAACGATACCAAAATATCAAAATGTTCCGAGTTATTAGGCAATGAGTTTACC TTTGACGTAGACGTTAGCAAATTTGGGTTGCCGATTAAACGGCCCTTTAC TTCGTGCTATGGATGCTGACGGAGGTATGTCAAAGTATTCTGGTAACAA AGCCGGAGCAAAGTACGGTACAGGTTATTGTGACGCTCAGTGCCCTAGAG	AM711862 MMYKFAALAAALVAGASAQQA CSLTAENHPSLTWKRCTSGGSCS TVNGAVTIDANWRWTH TVSGSTNCYTGNOQWDTSLCTDG KSCAQTCVVDGADYSSTYGHITS GDSLNLKFVTKHQYG TNVGSRVYLMENDTKYQMFELL GNEFTFDVDVSNLGCGLNGALY FVSMADADGGMSKYSGN KAGAKYGTGYCDAQCPRLKFI NGEANVGNWTPSTNDANAGFG

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p>	<p>ATTTGAAGTTTATCAACGGAGAAGCCAAACGTTGGTAACCTGGACGCCAAGT ACTAACGACGCAACGCTGGATTCCGGCAGATACGGTAGTTGTTGCTCAGA AATGGACGTTGGAGGCCAAATAACATGGCAACCGCTTTACTCCTCACCC CATGTACAACACTGTTGGACAATCTAGATGTGAGCCGACACCGTGGGTGGC ACCTACAGTAGCGTATAGGTATGCAGGAGTATGTGATCCTGACCGTTGCCG TTTTCAATGCTTATAGACAAGGAGACAAACGTTTATGGTAAAGGTATGA CCGTCGATACTAACAAAGAAAGATGACTGTGGTTACCCAGTCCACAAGAAC TCAGCTGGAGTATTGCTGAAATTAAGATTTCTACGTCACGGATGGAAA GATTATTGCTAATGCCGAGAGTAAAGTACCCAGTAACCTGGAAATAGTA TCACACAGGAATACTGTGACCGCTCAGAAAGTAGCTTTTAGCAACACCGAT GACTTCAATAGAAAGGGTGGATGGCTCAATGAGTAAGGCTTTAGCCGG TCCAAATGGTGTGGTGTGCTGTTGGGATGATCATACTATGCAAAACATGCT TTGGCTTGACAGCACCTATCCTATCGACCAAGCCGGAGCCCAAGGTGCTG AAAGGGTGCATGTCCAACCCACGAGTGGTGTGCCCGCCGAGATTGAAGCT CAAGTGCCTAATAGTAACGTTATCTTTTCCAAATAAAGATTCCGACCAATC GGATCCACTGTTCCAGGTTTGGATGGATCTAATCCTGGCAACCCCAACAAC CACGGTAGTCCCTCCAGCTTCAACTTCCACAAGTAGACCAACAAGTTCAA CGTCCAGTCCAGTGTACTCCTACCCGACAAACCCAGGAGGCTGTACCACT CAGAAATGGGGTCAATGGGGTGGAAATGGCTATACAGGTTGTACGAAATG CGTTGCAGGAACCACTTGTACACAGTTAAACCCCTTGGTACTCACAATGCCT ATAAGGCGCGCC (SEQ ID NO:9)</p>	<p>RYGSCSEMDVWEANNM ATAFTPHPCITTVGQSRCEADTCG GTYSSDRYAGVCDPDDGCDFNAY RQGDKTFYGKGMTVD TNKMTVVTQFHKNSAGVLSI KRFYVQDQKILANAESKIPGNPG NSITQEYCDQAQKVAF SNTDDFNKGGMAQMSKALAG PMVLVMSVWDDHYANMLWLD STYPIDQAGAPGAERGACP TTSQVPAEIEAQVPNSNVIFSNI FGPIGSTVPLDGSNPNPNTTV VPPASTSTSRPTS STSSPVSTPTGQPGGCTTQKWGQ CGGIGYTGCTNCVAGTTCTQLN PWYSQCL (SEQ ID NO:29)</p>

Organismo donante/gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
Celobiohidrolasas		
Acremonium thermophilum CBH1	<p>ATGTAACCAAAATTTGCTGCATTGGCCGCTTTAGTTGCAACAGTAAGAGGT CAAGCCGCTTGTCTAACCAGCAAACTCACCCATCTCTACAAATGGCA GAAATGCACAGCCCCTGGATCTTGTACAACTGCTCCGGCCAAAGTCACCA TTGACGCTAATTGGAGATGGCTTCAACCAAACTAACTCTTCAAACGAAATTGTT ATACCGGTAACGAATGGGATACTTCCATAATGTTTCATCCGATACAGACTGC GCAACGAAATGTTTGTAGATGGAGCAGACTATACGGGAACCTTATGGTGT TACAGCCTCAGGTAATTCCCTAAACCTTAAGTTCGTAACCTCAAGGACCAT ATAGTAAGAATAATCGGCTCTAGAATGTACTTGTGATGGAAGTGAGAGCAAA TATCAGGGTTTTACGTTATTGGGACAAGAGTTTACATTTGATGTTGATGTG AGTAACTTAGGTTGGGCTAAACGGCCCTTGTACTTCGTTTCTATGGAT CTTGATGGAGGTGATCAAAAATACACGACCAACAAGGCTGGAGCCAAATA</p>	<p>MYTKFAALAAALVATVRGQAAAC SLTAETHPSLQWQKCTAPGSCTT VSGQVTIDANWRWLHQINSSTN CYTGNEWDTSCSSDTCATKC CLDGADYTGTYGVTASGNSLNL KFVTQGPYSKNIGSRMYLMESES KYQGFTLLGQEFDFDVSNLG CGLNGALYFVSMDDLDDGGVSKYT TNKAGAKYGTGYCDSQCPRDLK FINGQANIDGWQPSSNDANAGL GNHGSCCSEMDIWEANKVSAAY</p>

Organismo donante/ gen	Secuencia de ADN con codones optimizados utilizada	Número de registro y secuencia de aminoácidos
<p>Celobiohidrolasas</p>	<p>TGGTACGGGATATTGTGACAGCCAAATGCCCTAGAGACTTAAAGTTCATTA ACGGTCAGGCAAAATATTGACGGCTGGCAACCAAGCAGTAACGACGCTAAT GCCGGACTAGGTAACCATGGCTCATGTTGTTCCGAAATGGATATCTGGGA AGCCAATAAGGTGCCCTGCTACACCCCCATCCATGCACGACAATCG GTCAGACAATGTGTACCGGTGATGACTGTGGAGGCACATACTCAAGTGAT AGGTACGCCGGTATATGTGATCCTGACGGTTGCCGATTTCAACTCTTATAGA ATGGGAGATACATCCTTTACGGCCCCGGTAAACAGTTGATACGGGTAG TAA GTTCACTGTTGTTACTCAGTTCTTAA CAGGTT CAGACGGCAATCTTAG TGAAATCAAAGAATTCTACGTT CAGAATGGA AAGTCAATTCCTAATTCCG AGAGTAAGATTGCTGGTGTCTGTTAACAGTATCACGACCGACTTCTGT ACCGCCAAAAGACTGCCTTTGGAGATACGAATGTTTTCGAGGAAAAGGGG CGGTCTTGCTCAAATGGGCAAGGCTTTGGCCGAACCAATGGTATTAGTCC TATCCGTTTGGGATGATCATGCAGTGAATATGCTTTGGCTTGATAGCACCT ACCCTACTGACAGCACCAAGCCAGGAGCTGCCAGAGGTGACTGTCTATC ACAAGTGGCGTTCCAGCAGATGTAGAGGCCAAGCTCCAAACTCCAATGT GATCTATTCTAACATCAGATTTGGCCCCATT AATAGTACCTATACAGGAAC GCCCTCTGGTGGTAAACCCCTCCAGCGGAGGCACCACTACCCACGACCA CAACGACTTCAAAGCCTTCTGGCCCTACGACA ACTACCAATCCTTCCCGGA CCACAGCAA ACTCACTGGGT CAGTGTGGAGGCCAAGGATGGACGGGTC CTACCGTGTGTC AATCACCTTACACATGC AATAACAGT AATGACTGGTACT CTCAGTGT TATAA (SEQ ID NO:45)</p>	<p>TPHPCTTIGQTMCTGDDDCGGTYS SDRYAGICDPDGCDFNSYRMDG TSFYGPGKTVDTGSKFTVVTQFL TGSDGNLSEIKRFYVQNGKVIPN SESKIAGVSGNSITDFCTAQKT AFGDTNVFEERGGLAQMGKAL AEPMLVLSVWDDHAVNMLWL DSTYPTDSTKPGAARGDCPITSG VPADVESQAPNSNVIYSNIRFGPI NSTYTGTSPSGGNPPGGGTTTTT TTTTSKPSGPTTTTNPSPGPPQTHW GQCGGQGWGTGPTVCQSPYTCK YSNDWYSQCL (SEQ ID NO:46)</p>

Endoglucanasas		
Coptofer- mes lacteus EG	<p>ATGAGATTTCCTCCATATTCACCGCTGTTTTGTTCGCAGCCTCAAGTGCTT TAGCAGAAATGTACTAAGGGTGGATGTAACAAGAATGGATACATAGTT CATGATAAGCACGTCGTGACATCCAGAAATAGAGACACTTTGGACCCCTCC AGACTTAGATTATGAAAAGGACGCTGGGAGTAACCCGTGTCGGGTGGAACCC TTAGTCAAAGATTAGTCTCAACTTGGAAACGGTAAGAAAAGTCGTGGGAAGT AGATTGTATATTGTGGACGAAAGCCGACGAGAAATATCAATTAATCACATTT GTCGGTAAGGAGTTCACCTATACCCGTTGATATGTCCAGATCCAATGTGGA ATCAATGCCGCATTATACACAGTGGAAATGCCCTGCCGCTGGAAAAGACCCC TGGAGGTGTTAAGTATGGATATGGATATTGTGATGCCAACTGCCGTGGATG GAGATTGTTGTATGGAGTTCGATATCCAAGAAAGCTTCTAACAAAGGCAATC</p>	<p>GTTTACACACCCATTCCCTGTCAAAGTCAAACCTTCAGGTTGCGATACCTCA GGATGCGGTTACAAACCTTACAGAGACAGTGGTGACAAAGGCATTCCTGGGG AACAACTATAAACGTA AACCCAGCCTGTGACAAATTGTAACACACAGTTTATCG GTTCTGTAAGTTCCTTAACTGAAGTCAAAGATTGTGCGTGCAGAGTGGA AAGACCTTCCCTCCAGCCAAATCATTAACCCGACAGTTATTGTAATGCCAAC GACTATAGAAAGTTTGAGAACTATGGGTGCATCCATGGCTAGAGGACACCGT TGTGTGTTTTCTTTGTGGGATCTAATGGTATGATGGATGGATGGAGG TAACGCCGGTCCCTGTACCTCATATAATTTGAATCTTTGGAAATCCAGTCA GCCAAACTTAAAGGTCACATGGTCAAACGTTGAAATACGGGAGAGATCGATT CTCCTTATTAA (SEQ ID NO:10)</p>
	<p>MRFPSIFTAVLFAASSALAECTK GGCTNKNGYIVHDKHVGDIQNR DTLDPDDLDEKDVGVTVSSGGT LSQRLVSTWNGKKVVGSRLYIV DEADEKYQLFTFVKGKEYTYTVD MSQIQCGINAALYTVEMPAAGK TPGGVKYGYGYCDANCVDDGDC CMEFDIQEASNKAIVYTTTHSCQS QTSGCDTSGCGYNPYRDSGDKA FWGTTINVNQPVTVTQFIGSGSS</p>	<p>LTEVKRLCVQGGKTFPPAKSLT DSYCNANDYRSLRTMGASMAR GHVVVFSLWDSNGMSWMDGG NAGPCTSYNIESLESSQP NLKVT WSNVKYGEIDSPY (SEQ ID NO:30)</p>

<p><i>Endoglucanasas</i></p>	<p>ATGAGATTCCCTCCATTTTCACTGCTGTTTTGTTTGCAGCCTCAAGTGCTT TAGCAGCCTATGACTACAAGACAGTATTGAAGAACCTCCTTGTGTTCTACG AAGCTCAAAGAAAGTGGAATAATGCCTGCAGACCAGAAAGGTGACCTGGAG AAAAGATTCCGCATTAAACGACAAGGACAGAAAGGAGAGGACTTAACT GGAGGTTATTACGACCGCGAGACTTTGTGAAGTTCGGTTTTCCAAATGGCA TACACAGTTACCGTGTGGCCTGGGGTTTAGTCGATTTAGAACTGCTTAC AGTACTGGGGTGCCCTGGATGATGGTAGAAAGGCCCTTGAATGGGGTAC AGATTATTTCTTGAAGCACATACCGCTGCCAATGAGTTTTACGGACAGGT GGGTCAGGGAGATGTGGATCATGCTTACTGGGACGCTCTGAGGACATGA CTATGTCTAGACCAGCTTACAAGATCGATACATCAAAACCTGGTAGTGACT TAGCTGCAGAAACAGCAGCCGCTTAGCAGCAACCGCAATAGCTTACAAG TCAGCCGATTTCTACCTACAGTAACAACCTTAATTACTCATGCAAGCAGTTG TTCGATTTTGCAACAATAATAGAGGAAAGTACTCTGATAGTATTACCGAT GCCAAGAATTTCTATGCATCCGGTGATTAAGGACGAATTAGTAGTGGGCT GCAGCCTGGTTGTATAGAGCTACAATGATAACACTTACTTAAACCAAAGC CGAATCATTTGATAATAATGAAATTTGGTTTAGGATCTTGGAAACGGTGCATTCAA TTGGGATAACAAGATAATCCGGAGTTCAGGTCTTATTAGCCAAAATTGACATC CAAACAAGCATAACAAGATAAAGTTCAGGGTATGTTGATTACTTAGTCTC CTCTCAAAGAATAACTCCAAGGGATTGGTCTATATTGACCCAATGGGGAA CCTTAAGACACGCAGCTAATAGTGCCTTGATCGCTTTACAGGCCGCTGATT TGGGTATAAACGCTGCTAGTTATAGACAATACGCAAAAGAAAGCAAAATTGAT TATGCCCTTAGGTGACGGAGGTCGTTCTTACGTGGTCGGATTCCGGAACTAAC</p>	<p>BAB40697 MRPFSIFTAVLFAASSALAAAYDY KTVLKNLSLLFYEAQRSGLPAD QKVTWRKDSALNDKGQKGEDL TGGYYDAGDFVKFGFPMAYTV TVLAWGLVDYESAYSTAGALD DGRKALKWGTDYFLKKAHTAAN EFYQQVGGQDVDHAYWGRPED MTMSRPAYKIDTSKPGSDLAEE TAAALAATAIAYKSADSTYSNN LITHAKQLDFANNYRGKYSDSI TDAKNFYASGDYKDELVWAAA WLYRATNDNTYLTKAESLYNEF GLGSWNGAFNWDNKISGVQVL LAKLTSKQAYKDKVQGYVDYL VSSQKTKPKGLVYIDQWGTLRH AANSALIALQAADLGINAASYR QYAKKQIDYALGDGGRSYVVG FGTNPVVRPHRSSHSSCPDAPAC DWNTYNSAGPNAHVLTGALVG GPDSNDSYTDSDYISNEVATD YNAGFQSAVAGLLKAGV (SEQ ID NO:31)</p>
<p>Coplo- fermes for- mosa- nus EG</p>	<p>ATGAGATTCCCTCCATTTTCACTGCTGTTTTGTTTGCAGCCTCAAGTGCTT TAGCAGCCTATGACTACAAGACAGTATTGAAGAACCTCCTTGTGTTCTACG AAGCTCAAAGAAAGTGGAATAATGCCTGCAGACCAGAAAGGTGACCTGGAG AAAAGATTCCGCATTAAACGACAAGGACAGAAAGGAGAGGACTTAACT GGAGGTTATTACGACCGCGAGACTTTGTGAAGTTCGGTTTTCCAAATGGCA TACACAGTTACCGTGTGGCCTGGGGTTTAGTCGATTTAGAACTGCTTAC AGTACTGGGGTGCCCTGGATGATGGTAGAAAGGCCCTTGAATGGGGTAC AGATTATTTCTTGAAGCACATACCGCTGCCAATGAGTTTTACGGACAGGT GGGTCAGGGAGATGTGGATCATGCTTACTGGGACGCTCTGAGGACATGA CTATGTCTAGACCAGCTTACAAGATCGATACATCAAAACCTGGTAGTGACT TAGCTGCAGAAACAGCAGCCGCTTAGCAGCAACCGCAATAGCTTACAAG TCAGCCGATTTCTACCTACAGTAACAACCTTAATTACTCATGCAAGCAGTTG TTCGATTTTGCAACAATAATAGAGGAAAGTACTCTGATAGTATTACCGAT GCCAAGAATTTCTATGCATCCGGTGATTAAGGACGAATTAGTAGTGGGCT GCAGCCTGGTTGTATAGAGCTACAATGATAACACTTACTTAAACCAAAGC CGAATCATTTGATAATAATGAAATTTGGTTTAGGATCTTGGAAACGGTGCATTCAA TTGGGATAACAAGATAATCCGGAGTTCAGGTCTTATTAGCCAAAATTGACATC CAAACAAGCATAACAAGATAAAGTTCAGGGTATGTTGATTACTTAGTCTC CTCTCAAAGAATAACTCCAAGGGATTGGTCTATATTGACCCAATGGGGAA CCTTAAGACACGCAGCTAATAGTGCCTTGATCGCTTTACAGGCCGCTGATT TGGGTATAAACGCTGCTAGTTATAGACAATACGCAAAAGAAAGCAAAATTGAT TATGCCCTTAGGTGACGGAGGTCGTTCTTACGTGGTCGGATTCCGGAACTAAC</p>	<p>CCTCCAGTAAGACCTCATCATAGATCCAGTTCCTGTCTGACGCACCCAGCC GCTTGCGACTGGAATACTTACAACCTGCCCGGACCAAATGCCACACGCTTGG ACCGGAGCCTTAGTAGGTGGACCAAGATTTCCAACGATAGTTACACAGATTC ACGTTCTGATTATATCAGTAACGAAAGTCCGCTACTGATTACAATGCCGGTTT CCAACTGACAGTTGCTGTTTTGTTGAAAGCCGGAGTATAA (SEQ ID NO:11)</p>

<p>Endoglucanasas</p>	<p>Nasuti- fermes ta- kasa- goensis EG</p> <p>ATGAGATTTCCATCTATTTTCACTGCCGTCCTTATTTGCAGCCTCCAGTGCAT TAGCAGCCTATGATTATAACAAGTTTGTGAGAGATTCCCTTATTGTTCTACG AAGCTCAGAGAAGCGGTAGATTACCAGCAGACCAGAAAGTCACTTGGAG AAAAGATTCAGCCTTGAATGATCAGGGAGATCAAGGTCAAGACTTAAACCG GAGGTTATTTGACGCCGGTGATTTTGTGAAATTTGGTTTCCCAATGGCAT ATACTGCTACCGTCTTGGCTGGGTTTAAATCGATTTGAGGCAGGATACA GTTCCGCTGCTGCTTGGATGACGGTAGAAAAGCAGTAAAGTGGCAACT GATTACTTTATAAAGGCCCACACTTCACAGAATGAGTTTACGGACAAGTC GGTCAGGGTGACGCTGATCACGCTTCTGGGGACGTCCTGAAGATATGAC CATGGCTAGACCAGCCTACAAGATTGACACCAGCAGACCAGGTAGTGACT TAGCGGGTGAACCCGACGGCATTGGCAGCTGCCAGTATCGTGTTTAGA AATGTTGATGGTACATACTTAACTTAACTTACTTACTCATGCCAGACAATTA TTTGACTTTGCAAAATAACTACAGAGGAAATACTCAGATTCCATAACCGA CGCTAGAAACTTTTACGCCAGTGCAGATTACCGTGACGAAATTGGTTTGGGC TGCCGCATGGTTGTACAGAGCTACAATGACAACACTTACTTGAATACCG CAGAACTCCTGTGATGATGAATTTGGATTGCAGAACTGGGGTGGAGGGTTA AACTGGGATTCAAAAGTGTCTGGTGTCCAGGCTTGTTAGCAAAATTGACC AACAAACAGGCTTACAAGATACTGTGCAGTCTTACGTGAATACCTGATT AATAACCAGCAAAAGACCCCAAAAGGATTGTTATACATTGATATGTGGGG TACATTGAGACACGCCGCAAAATGCTGCATTCATCATGTTGGAAAGTGCCT AGTTGGGTTTATCCGCATCATCTTACAGACAGTTTGTCTCAAACTCAGATCG ACTACGCTTTGGGTGACGGTGGAAAGTTTCGCTGTGGTTTTGGTTCAA ACCTCCTACAAGACCACATCATCGTTCTTCCAGTTGCCCGCTGCCCCAG CAACTTGTGACTGGAATACATTCAACTCACTGACCCAAATACACACGTGT TATCTGGAGCTTTGGTAGGAGGCCAGATCAAAACGATAATTATGTGGAT GATAGATCCGACTACGTCCATAACGAAAGTGGCAACCGACTACAACGCCGG ATTTACAGAGTGTCTTTGGCAGCCTTAGTTGCTTTGGGTTATTAA (SEQ ID</p>	<p>NO:12)</p>
	<p>MRFPSIFTAVLFAAASSALAAAYDY KQVLRDSLFLFYEAQRSGLPAD QKVTWRKDSALNDQGDQGGDL TGGYFDAGDFVKFGFPMAYTAT VLAWGLIDFEAGYSSAGALDDG RKA VKWA TDYFIKAHTSQNEFY GQVGGDADHAFWGRPEDMT MARPAYKIDTSRPGSDLAGETA AALAAAIVFRNVDTYSNNLL THARQLFDFANNYRGKYSDSIT DARNFYASADYRDELVWAAA W LYRATNDNTYLN TAESLYDEFG LQNWGGGLNWD SKVSGVQVLL AKLTNKQAYKDTVQSYVNYLIN NQKTPKGLLYIDMWGTLRHA ANAAFIMLEAAELGLSASSYRQF AQTOIDYALGDGGRSFVCGFGS NPPTRPHHRSSSCPPAPATCDWN TFNSPDPNYHVLSGALVGGPDQ NDNYVDDRSDYVHNEVATDYN AGFQSAALAAALVALGY (SEQ ID NO:32)</p>	

<p>Endoglucanasas</p>	<p>ATGAGATTCCCTAGTATTTTCACTGCCGCTTATTGACGCCAGTTCIGCTI TAGCCGCATATGATTATACCACAGTTTGAAGTTCCTTATTGTTCTACG AAGCTCAAAGATCCGGTAAGTTGCCAGCCAGCAGAAAGTCACTTGGAGA AAAGATTACAGCAATTAGACGATAAAGGAATAATGGAGAGGACTTAAACAG GAGGTTATTATGACGCTGGTATTTGTGAAGTTTGGTTTTCCTTTAGCATA CACCGTACTGTTTAGCCTGGGTTTGGTGGACTATGAAGCGGTTACTC ATCCGCTGGAGCCACAGATGACGGTAGAAAGCAGTGAATGGGCAACC GACTATTTGTTGAAGGCACATACTGCCGCTACCGAGTTATACGGACAGGTC GGGACGGTGACCGGATCACGCATATTGGGGACGTCCTGAAGATATGAC TATGGCTAGACCCAGCATACAAGATCGACGCTAGCAGACCAGGATCTGACT TAGCGGTGAACCGCTGCCGCTTAGCCGCTGCATCCATAGTTTTCAAAAG GTGTAGATTCTTCAATATTCTGACAACTTGTAGCTCACGCTAAACAGTTAT TTGATTTCCGTGACAATTATAGAGGAAATACAGTGTATCCATAACACAA GCTTCAAACTTTTACGCTCCGGAGATTACAAGACGAGTTAGTCTGGCT GCCACTTGGTTGTACAGAGCAACCAACGATAATACATAATTGACCAAAGC AGAATCCTTGTACAAACGAGTTCGGATTAGGAAACTGGAACGGAGCCTTAA ATTGGACAACAAGGTGCCGGTGTCAAGTGTGTTAGCCAAATTGACCT CCAAGCAGGCTTATAAAGACACCCGTTCAAGGATACGTCGATTATTGATTA ACAATCAGCAAAGACCCCAAGGGTTTGTATACATAGACCAAATGGGG ACCTTGAGACACGCAAGTAAATGCTGCCTTAAATACTTACAGGCTGCTGAT TTGGGTATTTCTGCCGACAGTTATAGACAATTCGCAAGAAAGCAATAAGA TTACGCTTTAGGTGACGGAGGTAGATCATATGATGTTGGTTTGGAGACAA TCCTCCAACACATCCTCATCACCGTTCTTCCCTCATGCCCTGACGCCCCAGC AGTATGCGAATTGGAATACTTTCAATTCACCTGATCCAAACTTTCATGTCTT AACCGGAGCTTTAGTGGGAGGTCTGATCAGAACGATAACTACGTTGATG ATCGTTCTGACTACGTGTCACACGAGGTTGCAACCCGACTATAATGCAGGAT TCCAAAGTGTGCTGGCCGCTTTAGTTACTTTAGGAGTTTAA (SEQ ID NO:13)</p>
<p>Copto- termes acinaci- formis EG</p>	<p>MRFPSIFTAVLFAASSALAAAYDY TTVLKSSLLFYEAQRSGLPADQ KVTWRKDSALDDDKGNNGEDLT GGYDDAGDFVKFGFPPLAYTATV LAWGLVDYEAGYSSAGATDDG RKAVKWATDYLLKAHTAATEL YGQVGDGDADHAYWGRPEM TMAPAYKIDASRRPGSDLAGET AAALAAASIVFKGVDDSSYSDNL LAHAKQLFDFADNYRGKYSDSI TQASNFYASGDYKDELVWAAT WLYRATNDNTYLTKAESLYNEF GLGNWNGAFNWDNKVSGVQV LLAKLTSKQAYKDTVQGYVDY LNNQKTPKGLLYIDQWGTLR HAANAALIIQAAADLGSADSYR QFAKKQIDYALGDGGRSYVVG GDNPTPHHRSSSCPDAVVC DWNFTFNSPDPNFHVLTGALVGG PDQNDNYVDDRSYVVSNEVAT DYNAGFQSAVAALVTLGV (SEQ ID NO:33)</p>

<p><i>Endoglucanasas</i></p>	<p>MRFPSIFTAVLFAASSALAAAYDY NDVLTKSLLFYEAQRSGLPSD QKVTWRKDSALNDKQNGEDL TGGYYDAGDYVKFGFPMAYTA TVLAWGLVDHPAGYSSAGVLD DGRKAVKWVTDYLKKAHVSKN ELYGQVGDGDADHAYWGRPED MTMARPAYKIDTSRPGSDLAGE TAAALAAAASIVFKSTDSNYANT LLTHAKQLDFDFANNYRKGYS ITQASNFYSSDYKDELVWAAV WLYRATNDQTYLTTAEKLYSDL GLQSWNGGFTWDTKISGVEVLL AKITGKQAYKDKVKGYCDYISG SQKTPKGLVYIDKWGSLRMA ANAAAYICAVAADVGGISSTAYRQ FAKTQINYILGDAGRSFVVGYG NNPPTHPHRRSSSCPDAFATCD WNNYNSANPNPHVLYGALVGG PDSNDNYQDLRSDYVANEVAT DYNAAFQSLALIVDLGL (SEQ ID NO:34)</p>
<p>Mastofe- mes darwini sis EG</p>	<p>ATGAGATCCCAAAGTATATTTACTGCTGTTTTGTTCCGACGCCAGTTCTGCTT TAGCAGCCTATGATTACAATGACGTATTAACCAAAGTTTGTGTTCTACG AAGCTCAAAGATCCGGTAAGTTACCTTCTGATCAGAAAAGTACACCTGGAGA AAAGATTCAGCATTAACGATAAGGGACAAAATGGTGAGGACTTAACCTGG TGGATATTATGACGCCGGTGATTACGTGAAGTTTGGTTTTCCAAATGGCATA TACTGCTACCGTTTTGGCTTGGGTTTAGTGGACCATCTGCCGGATACAG TTCTGCGGGTCTTGGATGATGTAAGAAAGCTGTGAAGTGGGTTACCG ATTACTTAATCAAAGCCACGTAATCAAAGAACGAAATTAACGGACAGGTC GGTGACGGTGACGCAGATCACGCTTATTGGGACGTCAGAGGATATGAC AATGGCAAGACCAGCATACAAAATAGACACTTCAAGACCAGTTCCGACT TAGCGGGTGAAACCGCAGCGGCAATTGGCTGCTGCACTAATTGTGTTAAAGT CAACAGATTCTAATTACGCCAACACCTTATTGACCCACGCAAAAACAATTAT TCGACTTTGCCAAATAACTATAGAGGTAAGTATAGTGATTCATAACACAG GCATCTAATTTCTACAGTAGTCCGACTATAAGATGAATTGGTTGGGCA GCTGTATGTTGTACAGAGCCACTAACGATCAGACCTATTGACAACCTGCA GAGAAAGTTACTCAGACTTGGGATTACAGTCTGGAACGGAGTTTCAC ATGGGACACCAAATTAGTGGAGTAGAAGTGTATTGGCTAAGATTACTG GTAAACAGGCATATAAGGACAAAAGTAAAGGATATTGTGATTATCTCA GGATCTCAGCAGAAAACACCTAAAGGATTAGTTTACATAGATAAGTGGGG TTCCTTAAGAAATGGCCGCAACCCGCATATATTGGCTGTAGCCGCAGA CGTCGGAATCAGTTCAACAGCTTACAGACAGTTCGCCAAAACACAGATTA ATTACATAATTGGGTGATGCCGGACGTTCTTTTGTGGTTGGTTACGGAAACA ACCCACCTACACACCCACATCACAGATCCAGTTTCATGTCTTGACGCCCCAG CAACATGCGATTGGAATAACTACAACAGTGCTAACCCATAATCCACATGTTT TATACGGTGCAATTAGTTGGTGACCAGATTCCAAACGATAATTATCAAGACT TAAGATCAGATTAATGTGCCAAACGAAAGTGGCAACAGACTACAATGCAGCC TTCCAGTCATTGTTAGCATTAAATCGTGACTTAGGTTTGTA (SEQ ID NO:14)</p>

<p>Endoglucanasas</p>	<p>ATGAGATTTCATCTATTTTCACTGCCGTCCTTATTTGTCAGCCCTCAAAGTGCTT TAGCAGCCTATGATTACAAACAAGTATTGAGAGATTCCCTTATTGTTCTACG AAGCTCAGAGAAAGCGGTAGATTACCAGCAGACAGCAGAGAGGTCACCTGGAG AAAAGATCCGCCCTTGAATGATCAGGGAGCAAGGTCAAAGACTTAACCCG GAGGTTATTTGACGCCGGTGATTTTGTGAAGTTGGATTCCCAATGGCTT ATACAGCAACCGTTTGGCCCTGGGTTTAAATCGACTTTGAAGCCGGTTACT CTTCTGCTGGTCCCTGGACGATGGTAGAAAAGCAGTAAAGTGGGCTACT GATTACTTTATAAAAGCCCATACTTCTCAAACGAGTTTACGGACAAGTC</p>	<p>MRFPSTAVLFAASSALAAAYDY KQVLRDSSLFFYEAQRSGRLPAD QKVTWRKDSALNDQGEQQDL TGGYFDAGDFVKFGFPMAYTAT VLAWGLIDFEAGYSSAGALDDG RKA VKWATDYFIKAHTSQNEFY GQVGGQGDVDHAYWGRPDMT MARPAYKIDTSRPGSDLAGETA</p> <p>AALAAASIVFKNVDGTYSNLL THARQLFDFAANNYRGKYSDSIT DARNFYASADYRDELVWAAAW LYRATNDNSYLNTAESLYNEFG LQNWGGGLNWDKSVSGVQVLL AKLTKQEKYKDTIQSYVNYLIN NQKTPKGLLYIDMWGTLRHA ANA AFIMLEAADLGLSASSYRQ FAQTQIDYALGDGGRSFVCGFG SNPPTRPHHRSSSCPPAPATCDW NTFNSPDPNYNVLSGALVGGPD QNDNYVDDRSDYVHNEVATDY NAGFQSALAAALVALGY (SEQ ID NO:35)</p>
<p><i>Nasuti-fermes walkeri</i> EG</p>	<p>GGTCAGGGTGACGTAGATCACGCATATTGGGGACGTCCTGAAGATATGAC AATGGCTAGACCAGCCTACAAGATTGATACCGAGACCAAGGTAGTGACT TAGCAGGAGAAACTGCTGCAGCTTTGGCTGCCCATCCATCGTTTCAAGA ATGTAGATGGTACATAATCCAAACAACCTTACTTACTCATGCTAGACAGTTGT TTGATTTCCGCAACAATTACAGAGGAAATACTCTGATAGTATTACCGATG CAAGAAACTTTACGCTAGTGCCGACTATAGAGATGAGTTAGTCTGGGCA GCTGCCGTGTTGACAGCAACCAACGACAATTTACTTGAACACTGCT GAATCATTATACAACGAGTTTGGATTGCAAAAATTTGGGTGGAGGGTTAAA CTGGGATTTCTAAAGTGAGTGTGTTCAAGTTTTTGTAGCCAAAGTTGACCAA CAAAACAAGATATAAGGACACTATTCAATCATACTACGTGAATTAATCA ATAACCAACAGAAAACCTCAAAGGGATTGTTATACATTGACATGTGGGG ACCTTGAGACACGCAGCTAACGCAGCTTTATAATGTTAGAGCTGCCGA CTTAGGTTTATCCGCTTCATCTTATAGACAGTTCGCCCAAACACAATAAGA CTACGCATTTGGGGACGGTGGACGTTCTTTTGTCTGTGGTTTCGGTTCTAA TCCTCCAACCTAGACCTCATAGATCCAGTTCATGCCCGCCTGCTCCAGC TACCTGTGATTGGAATACATTCAAATTCCTGACCCAAACTACAATGTTT ATCCGGTCCCTTGGTTGGTGTCTGACCCAGAAATGATAACTACGTGGACG ATAGAAGTGATTATGTCCATAATGAGGTAGCAACTGACTACAATGCCGGT TTCCAATCAGCCTTAGCCGCTTAGTCGCCTTAGGTTACTAA (SEQ ID NO:15)</p>	<p>AAALAAASIVFKNVDGTYSNLL THARQLFDFAANNYRGKYSDSIT DARNFYASADYRDELVWAAAW LYRATNDNSYLNTAESLYNEFG LQNWGGGLNWDKSVSGVQVLL AKLTKQEKYKDTIQSYVNYLIN NQKTPKGLLYIDMWGTLRHA ANA AFIMLEAADLGLSASSYRQ FAQTQIDYALGDGGRSFVCGFG SNPPTRPHHRSSSCPPAPATCDW NTFNSPDPNYNVLSGALVGGPD QNDNYVDDRSDYVHNEVATDY NAGFQSALAAALVALGY (SEQ ID NO:35)</p>

<p>Endoglucanasas</p>	<p>ATGAGATCCCAAAGTATATTTACTGCCGCTTAAITTTGACGCCICCAAGTGCATGCTTATGACTACAAAACAGTATTGTCCAAATTCCTTGTGTTCTACTGAAAGCTCAAAGATCCGGTAAAGTTACCTTCTGACCAGAAAGTGACCTGGAGAAAGGATTCAGCATTAAACGACAAAAGGACAAAAGGGTGAGGACTTAACCAGGTGATATTACGACCGGAGACTTTGTGAAATTTGGTTTTCCAAATGGCTTACACAGTTACCGTATTGGCATGGGTGTTATTGATTACGAAATCCGCCTACTCTGCCGAGGAGCTTAGATTACAGGTAGAAAAGGCCTTGAATAATATGGGACGACTATTTCTTAAAGGCACATACAGCAGCTAACGAGTTTACGGACAGGTGGGTCAAGGTGACGTTGACCCAGCATACTGGGACGTCCTGAAAGATATGACCATGAGCAGACCAGCATACAAAATAGACACTTCTAAGCCTGGTTCCTGACTTAGCTGCAGAGACTGCAGCTGCATTAGCAGCCACAGCTATTGCATACAAATCTGCCGATGCACAATATTCCAAACAATTTGATAACACATGCAAAACAA</p>	<p>AB019095</p> <p>MRPFSIFTAVLFAASSALAAAYDY KTVLSNLLFYEAQRSGLKLPDQ KVTWRKDSALNDKQKQGEDLT GGYYDAGDFVKFGFPMAYTVT VLAWGVIDYESAYSAAAGALDSG RKALKYGTDYFLKAHTAANEFY GQVGGQDGDVDHAYWGRPEDMT MSRPAYKIDTSKPGSDLAETA AALAATAIAYKSADATYSNNLIT HAKQLFDFANNYRKGKYSDSITD</p>
<p><i>Reticuli- termes- peratus</i> EG</p>	<p>TTATTCGACTTTGGCCAAACAATTACAGAGGAAAATAATTCGGATAGTATTACC GATGCCAAGAACTTTTATGCTTCTGGTGATTAACAAGACGAAATGGGTATGG GCCGCTGCATGGTTGTACAGAGCAACCAATGACAAACACATAATTTGACTAA GGCAGAAATCCTTATACAATGAATTTGGTTTGGGAAACTTCAATGGTGCCTT CAATTGGGATAACAAGTCTCCGGAGTCCAGGTGTTATTGGCCAAAGTTAA CCTCAAAAACAAGTGTATAAGGATAAGGTACAGTCTTACGTGGACTATTTG ATCTCCTCACAAAAGACACCCAAAAGGTTTATGTGTACATCGATCAATG GGTACTTTAAGACACCGCAGCTAATCTGCTTTGATCGCTTTGCAGGCAGC TGACTTAGGAATTAAACGTGCTACTTACAGAGCCTACGCAAGAAGCAAA TCGACTATGCTTTGGGTGATGGTGGAAAGATCCTATGTTATTGGATTTGGGA CCAAACCTCCAGTAAGACCACATCACAGAAGTTCATCTTGCCAGATGCA CCAGCTGTCTGCGATTGGAAACACCTATAACTCCGCTGGTCCAAACGCCAC GTGTTAAACCGGTGCAATGGTTGGAGGACCTGATAGTAATGATAGTTATACC GATGCTCGTTCTGACTACATATCCAAACGAAAGTGGCAACTGATTACAATGCG GGTTTCCAATCCGCTGCTCGCTGGATTATTGAAGGCGGGTGTCTAA (SEQ ID NO:16)</p>	<p>AKNFYASGDYKDELVWAAA WL YRATNDNTYLTKAESLYNEFGL GNFNGAFNWDNKVSGVQVLLA KLTSKQVYKDKVQSYVDYLISS QKKTPKGLVYIDQWGTLRHAA NSALIALQAADLGINAATYRAY AKKQIDYALGDGGRSYVIGFGT NPPVRRPHRRSSSCPDAVAVCDW NTYNSAGPNHVLGTGALVGGPD SNDSYTDARSDYISNEVATDYN AGFQSAVAGLLKAGV (SEQ ID NO:36)</p>

<p>Endoglucanasas</p>	<p>Neosar- forya fischeri EG</p>	<p>ATGAGATTTCCATCTATTTTCACTGCAGTTTGTTCGCAGCCAGTTCCGCTT TGGCCCAACAGATCGGGTCCATCGCCGAAATCATCTCCTGAGTTGACAAAC TATAGATGCTCCTCAAGCTGGATCGGTAGCACAGAGTACTTCCGTCGTG TTAGATATTACGCTCATTTGGATTTCATCAAAACGGTGCCCAACAAGTTGC ACTACCTCAAGTGGATTGGACCCTTCAATTGTGCCCTGATAAAGTCACTGT TCTCAGAACTGCGTAGTCGAAAGGAATAACCGACTACTCATCTTTTGGTGTG CAAAACTCCGGAGATGCAATGACATTAAGACAGTATCAAGTTCAAAAATGG ACAGATCAAAAACATTCGCTCTAGAGTGTACTTGTAGCTGAGGATGGAA TCAATTACTCCAATTTGCAAGTTGTTGAACCAAGAGTTTACTTTCGATGTGG ACGCTTCCAAATTTGCCCTTGTGGTATGAATGGAGCTTTATATTTGTCAGAAA TGGATGCTTCTGGTGACGTTCTGCCCTTGAACCCAGCGGGTGCCACATATG GACAGGTTACTGTGATGCCAGTGTCTCAACCCAGGTCCATGGATAAAT GGAGAAGCAATACTGTGGAGCCGGTGCATGTTGCCAAGAGATGGACTT ATGGGAAGCCAACTCCCGTTCTACCCATTTTTCAGTCTCACCCATGTACAAC TGCCGGTTTGTATGCCCTGTACTGGAGCTGAGTGTACTCAATCTGTGACGG TTATGGTTGCACCTTACAACCCCTTATGAATTAGGAGCCAAAGATTACTATGG</p>	<p>XM_001258277 MRFPSTAVLFAASSALAAQJG SIAENHPELTYRCSSQAGCVAQ STSVVLDINAHWIHQNGAQTSC TTSSGLDPSLCPDKVTCSQNCVV EGITDYSSFGVQNSGDAMTLRQ YQVQNGQIKTLRPRVYLLAEDG INYSKQLLNQEFDFDASKLP CGMNGALYLSEMDASGGRSAL NPAGATYGTGYCDAQCFNPGP WINGEANTAGAGACCQEMDLW EANSRSTIFSPHPCTTAGLYACT GAECYSICDGYGCTYNPYELGA KDYGYGLTIDTAKPITVVTFQF MTADNTATGTLAEIRRLYVQDQ KVIQNTAVAMTEAFCSSTRFEE</p>
	<p>TTACGGTTTGACTATTGACACCCGCAAGCCAAATAACAGTGGTTACTCAGTT TATGACCCGCTGATAATACAGCAACCCGTACATTAGCAGAGATCAGAAGAT TATATGTTCAAGATGGTAAAGTAAATCGGAAATACAGCCGTGGCCATGACC GAGGCATTTTGTAGTCTAGTAGAACAATTTGAAGAGTTAGGTGGTTTGCAA AGAAATGGGAGAAAGCTTTAGGTAGAGGAATGGTGCCAGTTTCTCAATAATG GGACGATCCTGTTTGTGGATGCATTGGTTAGATTCTGACGGTGCAGGACC TTGTGGTAATACTGAAGGTGATCCTGCCCTTTCATTCAGGCTAACTACCCAAA TACCCCGGTAACATTTCTCCAAGGTGAGATGGGGAGATAICGGTTCTACCTA TAGTTCCTTAA (SEQ ID NO:17)</p>	<p>LGGLQRMGEALGRGMVPVFSI WDDPGLWMHWLSDSDGAGPCG NTEGDPAFIQANYPNTAVTFISKV RWGDIGSTYSS (SEQ ID NO:37)</p>	

<p>Endoglucanasas</p>	<p>ATGAGATTTCCATCTATTTTCACTGCTGTTTGTTCGCAGCCCTCAAGTGCTT TAGCACAAATGGATGCAGATCGGTGGTAAAGCAGAAATATCCTGCCTTTAAG CCAGGTGCTAAGTACGGAAAGAGGTTATTGTGACGGACAGTGCCCTCACGA CATGAAGGTGCTAGTGGAAAGAGCAAAACGTTGACGGATGGAAGCCACAA GACAAACGACGAAATAAGTGGAAATGGAAATAATGGGTACATGTTGCTGGGA GATGGATAATATGGGAAAGGAAACTTAGTGTCCCAAGCCTACACCCTCACG CTGGTTCCAAAGTCCGGACAATATGAGTGTACTGGAACACAATGCCGGTGAC ACCGACAGTGGTGAAGAATTCAGGGAAACATGCCGATAAAGATGGTTGTGA TTTCGCAAGTTACAGATGGGGAGCTACAGACTATTACGGTCTGGAAAGA CCGTGGACACCAAAACAGCCAATGACAGTCTGACCCAGTTCATTGGTGAC CCTTTGACTGAGATAAAGAGAGTTTATGTACAAGGAGGAAAAGTCATAAA CAATTCCAAAACATCTAACTTAGTTTCAAGTGTACGATTCCTTTGACTGAGGC CTTCTGCGATGACACCAACAGGTTACAGGTGATACAAATGACTTTAAGG CTAAAGGAGGTATGCTGGATTCTCCAAAGAACTTAGACACCCCAAGTTT TGGTGATGCTTTATGGGATGACCATACAGCTAATATGTTATGGTTAGATT CTACTTATCCTACCGATAGTACAAAGCCAGGTGCCCGCAAGAGGTACTTGT GCCGTCACCTCCGGGACCCCTAAAGATGTGGAAATCCAAGCAAAGCCAACTC TCAGGTAGTTTACAGTGACATTAAGTTTGGTCCCTATTAAATTCACACATACAA AGCAAATTAA (SEQ ID NO:18)</p>
<p><i>Reticuli- termes flavipes</i> EG</p>	<p>DQ014512 MRPFSIFTAVLFAASSALAQWM QJGGKQKYPAFKPGAKYGRGYC DGQCPHDMKVSSGRANVDGWK PQNDENSGNGLGTCCWEMD IWEGNLVSQAYTVHAGSKSGQY ECTGTCCGDTDSGERFKGTCDK DGCDFASYRWWGATDYYGPGKT VDTKQPMVVVTFIGDPLTEIKR VYVQGGKVINNSKTSNLGSVYD SLTEAFCDDTKQVTDGTDNDFKA KGGMSGFKNLDTQVLMVMSL WDDHTANMLWLDSTYPTDSTK PGAARGTCAVTSGDPKDVESKQ ANSQVVVYSDIKFGPINSTYKAN (SEQ ID NO:38)</p>

<p>Endoglucanasas</p>	<p>AB003694 MVSFTSLLAGVAAISGVLAAAPA AEVEPVAVEKREAEAEAAQQPGT STPEVHPKLTYYKCTKSGGCVA QDTSVLDWNRYRWMHDDANYN SCTVNGGVNNTLCPDEATCGKN CFIEGVDYAASGVTTSGSSLTMN QYMPSSGGYSSVSPRLYLDDSD GEYVMLKLNQELSFVDLSAL PCGENGLYLSQMDENGGANQ YNTAGANYGSGYCDAAQCPVQT WRNGTLNTHQGFCCNEMDILE GNSRANALTPHSCATACDSAG CGFNPYGSGYKSYGPGDVTDT SKTFTHIQFNTDNGSPSGLVSI TRKYQQNGVDIPSAQPGGDTISS CPSASAYGGLATMGKALSSGM VLVFSIWNDSQYMNWLDSDGN AGPCSSTEGNPSNILANNPNTHV VFSNIRWGDIGSTTNSTAPPPPA SSTTFSTTRRSSTTSSSPSCTQTH WQCGGIGYSGCKTCTSGTTCQ YSNDYYSQC (SEQ ID NO:39)</p>
<p>Tri- choder- ima ree- sei EG/</p> <p>ATGGTCTCCTTCACTCCCTGCTGGCCGGGTTGCCGCTATCTCTGGTGTCC TAGAGCCCTGCGCAGAAAGTTGAACCTGTGCGAGTTGAGAAACGTGAG GCCGAAGCAGAAAGCTCAACAACCCAGGAACATCAACACCAGAAGTCCATC CAAAGTTAACAACTATAAATGTAATAAGAGTGGAGGTTGTAGCCGAG GACACAAGTGTGTTAGACTGGAATTAATCGTTGGATGATGATGCCAAT TATAATCCCTGACTGTTAACGGCGGTGTTAACACTACGTTATGCCCCGAT GAAGCGACTTGTGGTAAGAATGTTTTATTGAAGGGGTTGACTACGCCGCT AGTGGTGTACGACGAGTGGTCACTCCTTGACGATGAATCAATACATGCC TCTTCTAGTGGTGGTATTCTCTGTGCTCACAAGCTGTAATTTATTGGATT CCGATGGGGAATATGTTATGTTAAATTAATGAATGGCAAGAACTGAGTTTT GATGTGATCTACTGCATACCTTGTGGAGAAAATGGTAGTCTTTAATTA TCACAAATGGACGAAAACGGCGGAGCCAAATCAGTACAATACAGCTGGTGC TAATATGGTTCAGGCTATTGTGATGCTCAATGTCCAGTGCAGACTTGGAG GAATGGCACCTTAAACACATCACATCAAGGATTTTGCTGTAAACGAAATGG ACATATTAGAAGTAATTCAAAGACTAATGCCTAATCCCGACTCTTGTA CTGGACCGCATGTGATTTCTGCCGTTGTGGTTTCAACCTTATGGTCTG GTATAAGAGTTACTACGGTCCGGGAGACACCGTGGATACGTCAAAGACC TTCACTATAATCACTCAGTTTAAACACAGATAACGGATCTCCGAGTGGTAAT TTGGTGAGTATTACTAGGAAATATCAGCAGAACCGGTGTTGATATCCGTCC GCGCAGCCAGCGGTGACACTATATCTAGCTGTCTCCGCCCAGTGCCTAT GGCGGACTTGCTACAATGGTAAGGCATTTGCTCCTCAGGTATGGTCCTAGTA TTTTCTATTTGGAAATGATAATTCACAATACATGAATGGCTGGATTCTGGT AATGCAGGCCCTTGCTCCTCTACAGAAAGGTAACCCAAAGCAATATACTAGC TAATAACCCAAAATACTCATGTTGTCTTTAGTAATAATTAAGATGGGGGATAT AGGTAGCACTACGACAGTACCGCACCTCCTCCACCTGCTAGCTCCAC GACATTTCCCACTACTAGAAGTCCAGCACTACCACTCATCACCATCTTG TACTCAAACCCATTGGGGACAGTGTGGTGTATAGGTTACAGCGGTTGCA AAACCTGCACATCTGGTACTACATGCCCCAATACAGTAATGACTATTACTCAC AATGTTAA (SEQ ID NO:19)</p>	

<p>Endoglucanases</p>	<p>As- pergillus ka- wachii EgA</p>	<p>TTAATTAATAAGAGAAATTTCTAACTTGATTGTTGCTGCTTCTGCTGCTACTA TGGTTTCTGCTTTGCCAATCTAGACAAATGAAAAGAGGGATTCTGGTTTTA AATGGGTTGGTACTTCTGAATCTGGTCTGAAATTTGGTTCTGCTTTACCAG GTACTTTGGTACTGATTATACTTGGCCAGAAACTTCTAAAATTCAAAGTTT TGAGAAAACAAGGTATGAACAATTTTAGAATACCAATCTTGATGGAAGA TTAACTCCAGATGGTTGACTGGTTCTTTTGTCTTACTTACTTGTCTGATT TGAAGTCAACTGTTGAATTTGTTACTAATCTGGTGTCTTATGCTGTTTTAGA TCCACATAATACGGTAGATTCGATGGTTCTAATATTGAATCTACTTCTGAT TTAAGACTTGGTGGAAAATGTTGCTACTGAAATTTGCTGATAACGATAAG GTTAATTTTCGATACAACAACGAAATATCATGATATGGAACAATCTTTGGTT TTGAAATTTGAACCAAGCTGCTAATTAATGGTAATAGAGCTGCTGGTGTACT ACTCAATACAATTTTCGTTGAAGGTAAATGCTTATACTGGTGTGGGATTTGG ACTACTTACAATGATGATTTGCTGTGTTTAACTGATTTCTGAAAGATAAGAIA ATATACGAAATGCATCAATACTTGGATTTCTGATTTCTTCTGGTACATCTGAA ACTTGTGTTTCTTACTAATGGTAAAGAAAGAAATGAAAAGGCTACTGAA TGGTTGAAAACATAACAACGAAGGTAATTTGTTGTTGTTGTTGTTGTTGTTG TGTTAATTTCTGTTTGAAGAGGCTGTTGAAGGAATGTTGGCTTATATGTC TGAAAATTTCTGATGTTTGGGTTGGTGTCTTCTGGTGGTCTGCTGGTCCATG GTGGGTACTTACATGTAATCTTTGGAAAATACTTCCCACTGTTGATGCTTCAITC TACTTATTTGCCAATTTGGAAAATACTTCCCACTGTTGATGCTTCAITCA TCTTCACTGCTTCCAGTTCAGTTGCCAGCCGCTACTTCTGCTGTTTCTACTA CTACTACAGCTGCAATTTGAACAATACTACTCCAGCTACTCAAGTTGAAA TTGCTTCTTCTTCACTTTCATCATCAGCTGTTGCTGCTTCAACAATACTTT GTCTAAGGTTAAGTCTAAATCTAAATCTCCATGTAATTTGTCATCTGCTAC TTCATCTGCTGTTTTCATCAGCTGCTGCAGTTACTACACCTGCAGTTGCAGCT ACAACTCCAGCTGCTGCTCCAACTTCTTCTTCTGTTGCTTTTGGTACTACTT CTGTTTACGTTCCAACTACTACTGCTGCTGCACCACTCTCAAGTTTCACTCTIC AGCTGCAGCTTCACTTCCAGGTTGTTGGTGTGTTCTGATCCACAAGGTC ATCTGCTACTAATTTCTGCTGGTGAAGTTAATCAATATTACCAATGTTGGTGG TATTAATTTGGACTGGTCCAACTGTTTGTGCTTCTCCATATACTTGTAAAGTT CAAAAACGATTACTACTATCAATGTTGCTGCTGAATTAAGGCGCGCC (SEQ ID NO:47)</p>	<p>MRISNLV AASAA TMVSALPSRQ MKKRD SGFKWVGTSESGAEFGS ALPGTLGTDYTWPETSKIQVLR NKG MNFRIPFLMERLTPDGLTG SFASTYLSDLKSTVEFVNSGAY AVLDPHNYGRFDSIESTSDFK TWWKNVATEFADNDKVFIDTN NEYHDMEQSLVNLNLQAANGI RAAGATTQYIFVEGNAYTGAW DWTTYNDDLSGLTDSSEDKIYE MHQYLDSDSSGTSETCVSSTIGK ERIEKATEWLKTNKQGIHGEFA GGVNSVCBEAEV EGMLAYMSEN SDVVWVGASWWSAGPWVWGTYM YSLEPTDGTAYSTYLPILEKYFPS GDASSSSASASVAAATS AVSTT TTAAFEQTTTPATQVEIASSSSS SAVAASQTTLSKVSKSKSPCKL SSATSSAVSSAAA VTTPAVAATT PAAAPTSSVAFATTSVYVPTTT AAAPSQVSSSAAA SSSGVVGV DPQGPSATNSAGEVNQY YQC INWGTPTVCASPYTCKVQNDYY YQCVAE (SEQ ID NO:52)</p>
------------------------------	---	---	--

<p>Endoglucanasas</p>	<p>MHWADVACSRPPWPRDSVKAL KCNWNAVIRGAMGVDEGGYL SDANTAYNLMVAVIEAAISNGIY VIVDWHAHNAHPDEAVKFFTRI AQYGSYLHILYEDFNEPLDVS WTDVLVPYHKKVIAAIRADKK NVILGTPKWSQDQDVAASQNPIK DYQNLMYTLHFYASSHFTSDLG AKLKTAVNNGLPVVFVTEYGTCE ASGNLNTDSMSSWWTLSDS LKISYANWAISDKSEACSAALSPG TTAVNVGVSSRWTSNGNMVAS YYKKKSTGISCSSSSSSSSSSSS GSSGTSSSSSSSSSSSSSSSSSS SGSSSSSSSSSSSSSSSSSSSSSS NGGGRVNFEEKNTGSPVPLCGVV FSVSLPSGTTLGGSWNMESAGS GQYSLPSWVRIEAGKSSKDAAGL TFNGKDKPTAKIVTTTKKC (SEQ ID NO:53)</p>
<p>Hetero- dera schachatii Engl</p>	<p>TTAATTAATAATGCAATGGGCTGATGTTGCTTGTCTAGACCACCATGGCCA AGAGATTCTGTTAAAGCTTTGAAGTGAATTTGGAACGCTAATGTTATTAGA GGTGCTATGGGTGTTGATGAAGTGGTTAATTTGCTCTGATGCTAATACTGCT TACAAATTTGATGGTTCGTTAATGAAGCTGCTAATTTCTAATGGTATCTACG TTAATGTTGATTTGGCATGCTCATAATGCTCATCCAGATGAAGCTGTTAAAT TCTTTACTAGAAATGCTCAAGCTTATGGTTCTTACTTGCCATATTTTGTACGA AGATTTCAATGAACCAATGGATGTTTCTTGGACTGATGTTTGGTTCCATA CCATAAAAAGTTAATGCTGCCATTAGAGCTATTGATAAGAAGAACGTTA TTATCTTGGGTACTCCAAAATGGTCACAAGATGTTGATGTTGCTTCTCAA ATCCAAATTAAGGATTACCAAAAACCTTGATGTACACTTTGCAATTTTACGCTT CATCTCATTTTACATCTGATTTGGGTGCTAAATTAAGAAACTGCTGTTAACA ATGGTTTGCCAGTTTGTACTGATAATGGTACTTGTGAAGCTTCTGGTA ATGGTAAATTTGAATACTGATTTCTATGTCATCTTGGTGGACTTTGTTGGATT TTTGAAAATTTCTTACGCTAAATGGGCTAATTTCTGATAAATCTGAAGCTTGT TCTGCTTGTCTCCAGGTACTACTGCTGTTAATGTTGGTGTCTTCTTAGAT GGACTTCTCTGGTAATATGGTTGCTTCTTACTACAAAATAAGTCCACTG GTATTTCTTGTCTGGTAGTTCTCAGGTTCTTCAAGTGGTTCATCTAGTGG TTCCTCCGGTACATCTTCTGTTCTAGTGGTTCATCTAGTGGTACTTCTTCC GGTAGTTCTAGTGGTAGTTCTGGTTCAGTTCCTGTTCCCTCCTCGTTCTG GTTCTGCATCTAATTTCTGTTGTTCCATCTAATACTTGGAAATGGTGGTAG AGTTAATTTGAAATTAAGAACAACACTGGTTCCTGTTCCATTTGTGGTGTGTT TTTTCTGTTTCTTTGCCAATCTGGTACTACTTGGGTGTTCTTGGAAATATGG AATCTGCTGGTTCGGTCAATATTTCTTTACCATCTTGGGTTAGAAATGAAG CTGGTAAATCTTCTAAAGATGCTGGTTTGACTTTAATGGTAAAGATAAGC CAACTGCTAAAATTTGTTACCACCAAGAAAGTGTCTTATAAGGCGCGCC (SEQ ID NO:48)</p>

<p>Endoglucanasas</p>	<p>TTAATTAATAAAGTCTGTTGCTCCATTGTTGTTGGCTGCTTCTATTT TGTATGGTGTGCTGTTGCTCAACAACACTGTTTGGGGTCAATGTTGGTGGTA TTGGTTGGTCTGGTCCAACTAATGTTGCTCCAGGTTCTGCTTGTCTACTTT GAATCCATAATATGCTCAATGTAATCCAGGTGCTACTACTATACTACTTCT ACTAGACCACCATCTGGTCCACAACACTACTAGAGCTACTTCTACATCT TCTTCTACTCCACCAACTTCACTCTGGTGTAGATTTGCTGTGTAAACATG CTGGTTTGTATTTGGTTGTACTACTGATGTAATGTTGTTACTTCTAAAGT TTACCCACCATTGAAAAATTTCACTGGTTCTAACAATATCCAGATGGTAT TGGTCAAAATGCAACATTTTGTAAACGAAGATGGTATGACTATTTTAGATT GCCAGTTGGTTGGCAATAATTTGGTTAACAAACAATTTGGGTGGTAAATTTGGA TTCTACTTCTATTTCTAAGTACGATCAATTTGGTTCAAGGTTGTTGCTTTG GGTGTACTGTATTTGTTGATATTCATAATATGCTAGATGGAATGGTGGT ATTATTGGTCAAGGTGGTCCACAATGCTCAATTTACTTCTTTGTGGTCA CAATTGGCTTCAAAAATATGCTTCTCAATCTAGAGTTTGGTTTGGTATTATG AATGAACCACATGATGTTAACATTAATACTTTGGGCTGCTACTGTTCAAGAA GTTGTTACTGCTATTAGAAATGCTGGTGTACTTCTCAATTCATTTCTTTGC CAGGTAAATGATTGGCAATCTGCTGGTGTCTTTATTTCTGATGGTCTGCTGC TGCCTTGTCTCAAGTTACTAATCCAGATGGTCTACTACTAAATTTGATCTTC GATGTTCAATAAGTACTTGGATTTCTGATAATTTCTGTAATCATGCTGAATGT ACTACAACAATAATGATGGTGTCTTTTCTCCATTTGGTACTTGGTTGAGA CAAAACAATAGACAAGCTATTTTGTACTGAACTGGAAGTGGTGGTAAATGTTCA ATCTTGTATCCAAGATATGTGCCAACAAATTCATACTTGAACCAAAATTC TGATGTTTATTTGGGTTACGTTGGTTGGGTTGGGTTCTTTTGTACTACT TACGTTTAACTGAACTCCAATCTTCTGGTAAATCTTGGACTGATACTT CTTTGGTTTCTTCAATGTTTGGCTAGAAAGTTATAAGGGCGGCC (SEQ ID NO:49)</p>
<p><i>Hypoc- rea jeco- rina (ana- morph: Tri- choder- ma ree- sel) Eg2</i></p>	<p>MNKSVAPLLLAASILYGGAVAQ QTVWQCGGIGWSGPTNCAPGS ACSTLNPYYAQCPGATTTTSTR PPSGPTTTTRATSTSSSTPPTSSG VRFAGVNIAGFDGCTTDGTCV TSKVYPLKNFTGSNNYPDGIGQ MQHFVNEDGMTIFRLPVGWQY LVNNNLGGNLDSTSISKYDQLV QGCLSLGAYCIVDIHNYARWNG GIHQGGPTNAQFTSLWSQLASK YASQSRVWFIMNEPHD VNINI WAATVQEVVTAIRNAGATSQFI SLPGNDWQSA GAFISDGSAAAL SQVTNPDGSTTNLIFD VHXYLDS DNSGTHAECTTNNDGAFSPLAT WLRQNNRQAILTETGGGNVQSC IQDMCQQIQYLNQNSDVYLYGV GWGAGSFDSTYVLTETPTSSGN SWTDTSLVSSCLARK (SEQ ID NO:54)</p>

<p>Endoglucanasas</p>	<p>Orpino- myces sp. PC-2 CellB</p>	<p>TTAATTAATAAGAGTCTTGAACCTCTTGTCTTTTGTGGGTTTGGTTAATIG CTGGTTGTGAAGCTATGAGAAACATTTCTTAAGAATGGTTAAAGAAT TGACTATTGGTTGGTCTTTGGGTAATACTTTGGATGCTTCTTGTGTTGAAAC TTTGAACTACTCTAAAGATCAAACTGCTTCTGAAACTTGTGGGGTAAATGT TAAAACTACTCAAGAATTGTACTACAAATTGTCTGATTTGGGTTTCAATAC TTTCAGAAATACCAACTACTTGGTCTGGTCAATTTGGTGATGCTCCAGATTA CAAAATTTCTGATGTTGGATGAAAGAGTTCCAGAAAGTTGTTGATTAATGC TTTGAATACTGGTGTACGCTAATTTGAACATTCATCATGAAACTTGGAA TTACGCTTTTCAAAAGAAATTTGGAATCTGCTAAAAGAAATTTTGGTTGCTAT TTGGAAACAATTTGCTGCTGAATTTGGTGATTACGATGAACATTTGATTTT TGAAAGGTATGAATGAACCAAGAAAGTTGGTGATCCAGCTGAATGGACTG GTGGTGATCAAGAAGTTGGAAATTTTGTAAATGAAATGAACGCTTTTGTTCG TTAAAACATAAGAGCTACTGGTGTAAACAATGCTAATAGACATTTGATGA TTCCAACCTATGCTGCTTCTGTTAATGATGGTCTAATTAACAATTTAAGTA CCCAATGGTGATGATAAGTTATTGTTCTTTGCACTTCTACTCCTCCATAC AATTTTGGCTTTGAACAATGGTCCAGGTGCTAATTTCTAATTTCTACGATGGT AACGAAATTTGATGGGTTATGAACACTAATTAACCTCTTCAATTTCTTAAG GGTATCCAGTTAATTTGGTGAATTTGGTCTATGAACAGAGATAATGAA GATGATAGAGAAAGATGGCAAGAACTACTACATTAATAAGGCTACTGCTTT GGGTATCCCATGTGTTAATTTGGGATAATGGTTAATTTGAAGGTGAAGGTGA AAGATTTGGTATTATTGATAGAAAGTCTTTTGAACGTTAATTTCCCAAAGTT GATTAATGGTTTGTAGAAAGGTTTGGGTGATGAAAAACCAAAACTACTA TTAGAAGAACTACTACTACAGTTCAAAGTTCAACCAACTATTAAACAACG AATGTTTCTCTACTAGATTGGGTTAATTTGTTGTTAATGGTTTCGATGTTT GTACACTGATAATGATGGTCAATGGGGTGTGAAAATGGTAATTTGGTGTG GTATTAATCTCTTGTGGTAACAATCAAGACAAATGTTGGTCTGAAAGAT TAGGTTATCCATGTTGTCATACACTACTAATGCTGAATATACAGACAACG ACGGTAGATGGGGTGTAGAAAACGGTAACGTTAAGTGGTGGGGAATATACTTGTAA GGCGCGCC (SEQ ID NO:50)</p>	<p>MKFLNSLSLLGLVIAGCEAMRNI SSKELVKELTIGWLSLGNLTLDASC VETLNYSKDQTASET CWGNVKT TQELYKLSLDLGFNTRFPIPTWS GHFGDAPDYKISDVWMKRVHE VVDYALNTGGYAILNIHHETWN YAFQKNLES AKKILVAIWKQIA AEFGDYDEHLIFEGMNEPRKVG DPAEWTGGDQEGWNFVNEMN ALFVKTRATGGNNANRHLMIP TYAASVNDGSINNFKYPNGDDK VIVSLHSYSPYNFALNNGPGAIS NFYDGNEDWVMNTINSSFISKG IPVIIGFVAMNRDNEDDRERW QEYYIKKATALGIPCVIWDNGYF EGEGERFGIHDRKSLNVFPPKLIN GLMKGLGDEKPKTTRRTTTT VQVQPTINNECFSTRLGYSCCNG FDVLYTDNDGQWGVENGNWC GIKSSCGNNQRQCWSERLGYPC CQYTTNAEYTDNDGRWGVENG NWCGIY (SEQ ID NO:55)</p>
------------------------------	---	---	---

<p>Endoglucanasas</p>	<p><i>Irpex lacteus</i> En1</p> <p>TTAATTAATAAGAGTCCTTTGTTGTTGCTGTCTGTCTACTTTGGCTTTAT CTACTCCAGCTTTTCTGTCTGTCTGGGTCAATGTGGTGGTATTTGGTTT TACTGGTTCTACTACTGTGATGCTGGTACTTCTTGTGTTCATTTGAACGAT TACTACTTTCAATGTCAACCAGGTGCTACTTCTACTGTTCACCAACT ACTACTGCTTCTACTACTTCTTCTGCTGCAGCTCCATCTTCTTCAGGTAATG CTGTTTGTCTGCTACTAGAAACAAGTTTAAAGTTCTTCCGGTGTAAATGAAT CTGGTGTGAATTTGGTAACAATGTTATTCAGGTACTTTGGGTACTGATT ATACTTGGCCATCTCCATCTTCTATTGATTTTTTCGTGGTAAGGGTTTTAA TACTTTCAGAGTTCCATTTTTGATGGAAGATTGCTCCACCTGCTACTGGT TTGACTGGTCCATTTGATTTCTACTTATTTGCAAGGTTTGAACAATATTGTT CTTACATTAAGGTAAGGTTATGCTTTGGTTGATCCACATAACTTTA TGATTTACAACGGTCTACTATTTCTGATACTAATGCTTTTCAAACTTGGTG GCAAAATTTGGCTGCTCAATTTAAGACTGATTTCTCAIGTTGTTTTTCGATGTT ATGAATGAACCACATGATATTCAGCTCAAACTGTTTTTAACTTGAACCAA GCTGCTAATAAGAAATAGAGCTTCTGGTCTACTTCTCAATCTATTTGG TTGAAGGTACTTCTTATACTGGTCTTGGACTTGGACTACTACTTCTGGTA ATTCTCAAGTTTTTGGTCTATTATGATCCAAACAACAATGTTGCTATTG AAATGCATCAATACTTGGATTTCTGATGTTCTGGTACTTCTCCAACTTGTG TTTCTCCAACTATTGGTGTGAAAGATTGCAAGCTGCTACTCAATGGTTGC AACAAACAATTTGAAAGGTTTCTTGGGTGAAATGGTGTGGTTCTAATG CTGATTTGATTTCTGCTGTTCAAGGTGCTTTGTGTGAAATGCAACAATCTG ATGTTTGGTGGTGTGTTGTGGTGGCTGCTGGTCCATGGTGGGTGATT ATTTTCAATCTATTGAACCACCATCTGGTGTGCTGTTTCTTCTATTTTGCC ACAAGCTTTGGAAACCATTTTTTGTTATAAGGGCGGCC (SEQ ID NO:51)</p>	<p>β-Glucosidasas</p>
	<p>MKSLLSAAAATLALSTPAFSVSV WQCQGGIGFTGSTTCDAGTSCV HLNDYFQCQPGAATSTVQPTT TASSTSSAAAPSSGNAVCSGTR NKFKFFGVNESGAIEFGNNVIPGT LGTDYTWPSPIDFFVVKGFNT FRVPFLMERLSPPATGLTGFDS TYLQGLKTVSYITKGGYALV DPHNFMIYNGATISDTINAFQTW WQNLAQAQFKTDSHVVFDMNE PHDIPAQTVFNLNQAAINRIRAS GATSQSILVEGTSYTGAWTWT TSGNSQVFGAIHDPNANNVAIEM HQYLDSDSGTSPTCVSPTIGAE RLQAATQWLQNNLKGFLGEIG AGSNADCISAVQCALCEMQQSD VWLGALWWAAGPWVWGDYFQS IEPPSGVAVSSILPQALEPFL (SEQ ID NO:56)</p>	

<p>Endoglucanasas BGLI Sf</p>	<p>FJ028723</p> <p>MVSFTSLLAGVAAISGVLAAPA AEVESVAVEKRSDSRVPIQNYT QSPSRDESSQWVSPHYYPQPQ GGRLQDVWQEAYARAKAIVGQ MTIVEKVNLTGTGTGWQLDPCVG NTGSVPRFGIPNLCIQDGPLGVR FADFVTGYPSGLATGATFNKDL FLQRQALGHEFNSKGVHIALG PAVGPLGVKARGGRNFEAFGSD PYLQGTAAAATIKGLQENNVMA CVKHFIGNEQEKYRQDDINPAT NQTTKEAISANIPDRAMHELYL WPFADSVRAGVGSVMCSYNRV NNTYACENSYMMNHLLKEELG FQGFVSDWGAQLSGVYSAISG LDMSMPGEVYGGWNTGTFSFWG QNLTKAIYNETVPIERLDDMATR ILAALYATNSFPTEDHLPNFSSW TTKEYGNKYADNTTEIVKVNY HVDPSNDFTEDTALKVAEESIVL LKNENNTLPIPEKAKRLLLSGIA AGPDPGYQCEDQSCNTGALFQ GWGSGSVGSPKYQVTPFEEISYL ARKNKMQFDYIRESYDLAQVTK VASDAHLSIVVVAASGEGYITV DGNQDRRRLTLWNNGDKLIET VAENCANTVVVVTSTGQINFEG FADHPNVTAIWVWAGPLGDRSGT AIANILFGKANPSGHLPTIAKTD DDYIPIETYSPSSGEPEDNHLVEN DLLVDYRYFEEKNIEPRYAFGY</p>	<p>ATGGTCTCCTTCACTCCCTCCTCGCCGGCGGTGCGCCGCAATCTCGGGCGTCT TTGGCCGCTCCCGCCGAGGTGCAATCCCGTGGCTGTGGAGAAGCGCTC GGACTCGCGAGTCCCAATTCAAAACATAACCCAGTCTCCATCCAGAGAG ATGAGAGCTCCCAATGGGTGAGCCCGCATTATTATCCAACTCCACAAAGGT GGTAGGCTCCAAGACGTCTGGCAAGAAGCATATGCTAGAGCAAAAAGCCCAT CGTTGGCCAGATGACTATTGTTGAAAAGGTCAATTTGACCACTGGTACCCG TTGGCAATTAGATCCATGTGTGGTAATACCGGTTCTGTTCCAAGATTCCG CATCCAAAACCTTTGCCACAAGATGGCCATTTGGTGTTCGATTCCGCTGA CTTTGTTACTGGCTATCCATCCGCTTTGCTACTCGGTCATGAGTTCAACATAA GGATTTGTTTCTTCAAAGAGGTCAAAGTCTCGGTCATGAGTTCAACAGCAA AGGTGTACATAATTGCGTTGGCCCTGCTGTTGGCCCACTTGGTGTCAAAGC CAGAGGTGGCAGAAAATTCGAAGCCTTTGGTTCGACCCCATATCTCCAAG GTAAGTGTGCTGTGCAACCATCAAAGGTCTCCAAAGAGATAAATGTTATG GCTTGTGTCAAAGCACTTTATTGGTAACGAACAAGAAAGTACAGACAGCC AGATGACATAAACCCTGCCACCACAACAACTACTAAAGAAGCTATTAGTG CCAACATTCAGACAGAGCCATGATGAGTTGTACTTGTGGCCATTTGGCCG ATTCGGTTCGAGCAGGTGTGGTTCIGTTATGTCCTTATAACAGAGTCA ACAACACTTACGCTTGGAAAACCTTACATGATGAACCACTTGTCTTAAAG AAGAGTTGGGTTTTCAAGGCTTTGTTGTTCCGACTGGGTGCACAATTA GTGGGTTTATAGCGCTATCTCGGCTTAGATATGCTATGCTGCTGGTGAAG TGTATGGGGATGGAACACCGGCACGCTTTCTGGGTCAAACCTTGACG AAAGCTATTTACAATGAGACTGTCCGATTGAAAGATTAGATGATATGGC AACCAAGGATCTGGCTGCTTTGTATGCTACCAATAGTTTCCCAACAGAAGA TCACCTTCCAAAATTTTCTTCATGGACAACGAAAAGAAATATGGCAATAAATA TTATGCTGACAACACTACCGAGATTGTCAAAGTCAAACACTACCATGTGGACCC ATCAAAATGACTTTACGGGAGACACAGCTTTGAAGGTTGCTGAGGAATCTA TTGTGCTTTTAAAAAATGAAAACAACACTTTGCCCAATTTCTCCCGAAAAGG CTAAAAGATTACTATTGTCGGGTATTGCTGCAGGCCCTGATCCGATAGGTT ATCAGTGTGAAGATCAATCTTGCACAAAATGCGCTTTGTTTCAAGGTTGGG GTCTGGCAGTGTGGTCTCCAAAATATCAAGTCACTCCATTTGAGGAAA TTTCTTATCTTGCAAGAAAACAAGATGCAATTTGATTATATTCGGGAGT CTTACGACTTAGCTCAAGTTACTAAAGTAGCTTCCGATGCTCATTTGTCTA</p>
--	---	--

<p><i>Endoglucanasas</i></p>	<p>GLSYNEYEVSNAKVSAAKKVVDE ELPEPATYLSEFSYQNAKDSKNP SDAFAPTDLNRVNEYLYPYLDS NVTLKDGNYEYDPGYSTEQRRT PIQPGGLGGNDALWEVAYKVE VDVQNLGNSTDKFVPLYLKHP EDGKFETPIQLRGFEEKVELSPGE KKTVEFELLRRDLVWVDTTRQS WIVESGTYEALIGVAVNDIKTSV LFTI (SEQ ID NO:40)</p>
	<p>TAGTTGTTGCTCTGCTGCAAGCGGTGAGGGTTATATAACCGTTGACGGTA ACCAAGGTGACAGAAAGAAATCTCACTTGTGGAAACAACGGTGATAAATTG ATTGAACAGTTGCTGAAAACCTGTGCCAATACTGTTGTTGTTACTTCTI ACTGGTCAAATAATTTGAAGGCTTTGCTGATCACCCAAATGTTACCGCA ATTGCTGGCGGCCCATTAGGTACAGATCCGGGACTGCTATCGCCAAAT ATTCTTTTGGTAAAGCGAACCCATCAGTCACTTCCACTTACTATTGCTA AGACTGACGATGATTACATTCCAATTGAAACCTACAGTCCATCGAGTGGT GAACCTGAAGACAACCACTTGGTTGAAAATGACTTGTGTTGACTATAG ATATTTTGAAGAGAAGAAATATTGAGCCAAAGATACGCATTTGGTTATGGCTT GTCTTACAAATGAGTATGAAAGTTAGCAATGCAAAAGGTCTCGGCAGCCAAA AAGTTGATGAGGAGTTGCCTGAACCAGCTACCTACTTATCGGAGTTTAGCT ATCAAAAATGCAAAAGACAGCAAAAATCCAAAGTGTGCTTTTGTCCCAACA GATTTAAACAGAGTTAATGAGTACCTTTATCCATAATTTAGATAGCAATGTT ACCTTAAAGACGGAAACTATGAGTATCCCGATGGCTACAGCACTGAGCA AAGAACAACACCTATCCAACTGGGGCGGCTTGGGAGGCAACCGATGCTT TGTGGGAGGTCGCTTATAAAGTTGAAGTGGACGTTCAAAAACCTTGGGTAAC TCCACTGATAAGTTTGTTCACACAGTTGTATTTGAAAACACCCCTGAGGATGGC AAGTTTGAAAACCCCTATTCAATTGAGAGGGTTTGAAAAGGTTGAGTTGTCC CCGGTGAGAAGAAGACAGTTGAGTTGAGCTTTTGAGAAAGAGATCTTAG TGTGTGGGATACCAACAGACAGTCTTGGATCGTTGAAATCTGGTACTTATGA GGCCTTAATTGGTGTGCTGTTAATGATATCAAGACATCTGTCCCTGTTACT ATT (SEQ ID NO:20)</p>

En determinados aspectos de la invención, los polipéptidos y polinucleótidos de la presente invención se proporcionan en una forma aislada, por ejemplo, purificada hasta la homogeneidad.

5 La presente invención también incluye polipéptidos que comprenden, o como alternativa consisten en, una secuencia de aminoácidos que es al menos aproximadamente 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% similar a la del polinucleótido de cualquiera de las SEQ ID NO: 21-40, 46, o 52-56 y a partes de dicho polipéptido, conteniendo dicha parte del polipéptido generalmente al menos 30 aminoácidos y más preferentemente al menos 50 aminoácidos.

10 Como se sabe en la técnica, la "similitud" entre dos polipéptidos se determina comparando la secuencia de aminoácidos del polipéptido, y los sustitutos de aminoácidos conservados en ella, con la secuencia de un segundo polipéptido.

15 La presente invención también se refiere a un dominio, fragmento, variante, derivado, o análogo del polipéptido de cualquiera de las SEQ ID NO: 21-40, 46, o 52-56.

20 Los fragmentos o partes de los polipéptidos de la presente invención pueden emplearse para la producción del polipéptido de longitud completa correspondiente mediante síntesis peptídica. Por lo tanto, los fragmentos pueden emplearse como productos intermedios para la producción de polipéptidos de longitud completa.

25 Los fragmentos de polipéptidos de celobiohidrolasa, endoglucanasa o beta-glucosidasa, incluyen dominios, fragmentos proteolíticos, fragmentos de delección y, en particular, fragmentos de polipéptidos de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o de *Arabidopsis thaliana*, que conservan cualquier actividad biológica específica de las proteínas de celobiohidrolasa, endoglucanasa o beta-glucosidasa. Adicionalmente los fragmentos de polipéptidos incluyen cualquier parte del polipéptido que conserve una actividad catalítica de las proteínas de celobiohidrolasa, endoglucanasa o beta-glucosidasa.

35 La variante, el derivado o el análogo del polipéptido de cualquiera de las SEQ ID NO: 21-40, 46, o 52-56, puede ser (i) uno en el que uno o más de los restos de aminoácidos se sustituye con un resto de aminoácido conservado o no conservado (preferentemente un resto de aminoácido conservado) y dicho resto de aminoácido conservado puede estar, o no, codificado por el código genético, o (ii) uno en el que uno o más de los restos de aminoácido incluye un grupo sustituyente, o (iii) uno en el que el polipéptido maduro se fusiona con otro compuesto, tal como un compuesto para aumentar la semivida del polipéptido (por ejemplo, polietilenglicol), o (iv) uno en el que los aminoácidos adicionales se fusionan con el polipéptido maduro para la purificación del polipéptido o (v) uno en el que un fragmento del polipéptido sea soluble, es decir, no esté unido a la membrana, sino que se una a ligandos en el receptor unido a la membrana. Se considera que, a partir de las enseñanzas del presente documento, dichas variantes, derivados y análogos, están al alcance de los expertos en la técnica.

45 Los polipéptidos de la presente invención incluyen adicionalmente variantes de los polipéptidos. Una "variante" del polipéptido puede ser una variante conservativa, o una variante alélica. Como se indica en el presente documento, una variante conservativa se refiere a alteraciones en la secuencia de aminoácidos que no afectan de manera adversa a las funciones biológicas de la proteína. Se dice que una sustitución, una inserción o una delección afecta de un modo adverso a la proteína cuando la secuencia alterada impide o altera una función biológica asociada con la proteína. Por ejemplo, las propiedades hidrófilas-hidrófobas o estructurales, de carga general de la proteína, pueden alterarse sin afectar de un modo adverso a una actividad biológica. Por consiguiente, la secuencia de aminoácidos puede alterarse, por ejemplo, para hacer que el péptido sea más hidrófilo o hidrófobo, sin afectar de un modo adverso a las actividades biológicas de la proteína.

55 Por una "variante alélica" se entiende formas alternativas de un gen que ocupa un locus determinado en un cromosoma de un organismo. Genes II, Lewin, B., ed., John Wiley & Sons, Nueva York (1985). Las variantes de origen no-natural pueden producirse utilizando técnicas de mutagénesis conocidas en la materia. Las variantes alélicas, aunque poseen una secuencia de aminoácidos ligeramente diferente a las indicadas anteriormente en el presente documento, seguirán teniendo las mismas funciones biológicas, o similares, asociadas con la proteína celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o de *Arabidopsis thaliana*.

Las variantes alélicas, las variantes de sustitución conservativa, y miembros de las familias de las proteínas endoglucanasa, celobiohidrolasa o β -glucosidasa, pueden tener una secuencia de aminoácidos que tenga una identidad de secuencia de al menos 75%, al menos 80%, al menos 90%, al menos 95% con una secuencia de aminoácidos de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *R. flavipes*, o *Neosartorya fischeri*, expuesta en una cualquiera de las SEQ ID NO: 21-40, 46, o 52-56. En el presente documento, la identidad u homología con respecto a dichas secuencias se define como el porcentaje de restos de aminoácidos en la secuencia candidata que son idénticos a los péptidos conocidos, después de alinear las secuencias e introducir huecos, si fuera necesario, para obtener un porcentaje de homología máximo, y no considerando ninguna de las sustituciones conservativas como parte de la identidad de secuencia. En la secuencia peptídica, las extensiones, deleciones o inserciones, en N-terminal, C-terminal o internas, no se interpretará como que afectan a la homología.

Por tanto, las proteínas y los péptidos de la presente invención incluyen moléculas que comprenden la secuencia de aminoácidos de las SEQ ID NO: 21-40, 46 y 52-56, o fragmentos de las mismas, que tienen una secuencia consecutiva de al menos aproximadamente 3, 4, 5, 6, 10, 15, 20, 25, 30, 35 o más restos de aminoácidos de las secuencias polipeptídicas de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *R. flavipes*, o de *Neosartorya fischeri*; variantes de las secuencias de aminoácidos de dichas secuencias en las que al menos un resto de aminoácido se ha insertado en el N o C terminal, o en el interior de la secuencia desvelada; variantes de secuencias de aminoácidos de las secuencias desveladas, o sus fragmentos como se ha definido anteriormente, que se han sustituido por otro resto. Las variantes contempladas incluyen adicionalmente aquellas que contienen mutaciones predeterminadas mediante, por ejemplo, recombinación homóloga, mutagénesis por PCR o dirigida a sitio y las proteínas correspondientes de otras especies animales, incluyendo, pero sin limitación, especies de bacterias, hongos, insecto, conejo, rata, porcino, bovino, ovino, equino y primates no humanos, los alelos u otras variantes de origen natural de la familia de proteínas; y derivados en los que la proteína se ha modificado covalentemente por sustitución, medios químicos, enzimáticos u otros medios apropiados con un radical que no sea un aminoácido de origen natural (por ejemplo, un radical detectable tal como una enzima o un radioisótopo).

Utilizando métodos conocidos de modificación por ingeniería genética de proteínas y tecnología de ADN recombinante, pueden generarse variantes para mejorar o alterar las características de los polipéptidos de celulasa. Por ejemplo, uno o más aminoácidos puede deleccionarse del extremo N o del extremo C de la proteína secretada sin que haya una pérdida sustancial de función biológica.

Por tanto, la invención incluye adicionalmente, variantes polipeptídicas de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* o de *Arabidopsis thaliana*, que muestran actividad biológica sustancial. Dichas variantes incluyen deleciones, inserciones, inversiones, repeticiones y sustituciones seleccionadas de acuerdo con normas generales conocidas en la técnica de manera que tengan poco efecto sobre la actividad.

El experto en la técnica es plenamente consciente de las sustituciones de aminoácidos que son menos probables que afecten, o que no es probable que afecten, significativamente a la función de la proteína (por ejemplo, reemplazando un aminoácido alifático con un segundo aminoácido alifático), como se describe adicionalmente a continuación.

Por ejemplo, en Bowie et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990), se proporciona orientación sobre cómo hacer sustituciones de aminoácidos fenotípicamente silenciosas, donde los autores indican que hay dos estrategias principales para estudiar la tolerancia al cambio de una secuencia de aminoácidos.

La primera estrategia aprovecha la tolerancia de las sustituciones de aminoácidos por selección natural durante el proceso de la evolución. Comparando secuencias de aminoácidos en diferentes especies, pueden identificarse aminoácidos conservados. Estos aminoácidos conservados son probablemente importantes para el funcionamiento de las proteínas. En cambio, las posiciones de aminoácidos en las que se han tolerado sustituciones por selección natural indican que estas posiciones no son críticas para el funcionamiento de las proteínas. De este modo, las posiciones que toleran la sustitución de aminoácidos podrían modificarse conservando al mismo tiempo la actividad biológica de la proteína.

La segunda estrategia utiliza modificación por ingeniería genética para introducir cambios de aminoácidos en posiciones específicas de un gen clonado para identificar regiones críticas para el funcionamiento de las proteínas. Por ejemplo, puede utilizarse mutagénesis dirigida a sitio o mutagénesis con alanina (introducción de mutaciones de una sola alanina en cada resto en la molécula). (Cunningham y Wells, Science 244:1081-1085 (1989).) La actividad biológica de las moléculas mutantes resultantes puede después someterse a ensayo.

Como indican los autores, estas dos estrategias han revelado que, sorprendentemente, las proteínas a menudo toleran sustituciones de aminoácidos. Además, los autores indican cuales son los cambios de aminoácidos posiblemente permisivos en determinadas posiciones de aminoácidos en la proteína. Por ejemplo, los restos de aminoácido más ocultos (dentro de la estructura terciaria de la proteína) requieren cadenas laterales no polares, mientras que generalmente se conservan algunas características de las cadenas laterales en la superficie. Además, las sustituciones de aminoácidos conservativas toleradas implican el reemplazo de los aminoácidos alifáticos o hidrófobos Ala, Val, Leu e Ile; el reemplazo de los restos de hidroxilo Ser y Thr; el reemplazo de los restos ácidos Asp y Glu; el reemplazo de los restos amida Asn y Gln, el reemplazo de los restos básicos Lys, Arg, e His; el reemplazo de los restos aromáticos Phe, Tyr, y Trp y el reemplazo de los aminoácidos de pequeño tamaño Ala, Ser, Thr, Met y Gly.

Los términos “derivado” y “análogo” se refieren a un polipéptido que difiere del polipéptido de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o de *Arabidopsis thaliana*, pero que conserva sus propiedades esenciales. Generalmente, los derivados y análogos son en conjunto muy similares, y, en muchas regiones, idénticos a los polipéptidos de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o de *Arabidopsis thaliana*. Cuando los términos “derivado” y “análogo” se refieren a los polipéptidos de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* o de *Arabidopsis thaliana* incluyen cualquiera de los polipéptidos que conservan al menos algo de la actividad del polipéptido nativo correspondiente, por ejemplo, la actividad exoglucanasa, o la actividad del dominio catalítico.

Los derivados de polipéptidos de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o de *Arabidopsis thaliana*, son polipéptidos que se han alterado de manera que muestran características adicionales no encontradas en el polipéptido nativo. Los derivados pueden modificarse covalentemente por sustitución, medios enzimáticos, químicos u otros medios apropiados con un radical distinto de un aminoácido de origen natural (por ejemplo, un radical detectable, tal como una enzima o un radioisótopo). Los ejemplos de derivados incluyen proteínas de fusión.

Un análogo es otra forma de un polipéptido de celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o de *Arabidopsis thaliana* de la presente invención. Un “análogo” también conserva sustancialmente la misma función o actividad biológica que la del polipéptido de interés, por ejemplo, funciona como una celobiohidrolasa. Un análogo incluye una proproteína que puede activarse por escisión de la parte proproteína para producir un polipéptido maduro activo.

El polipéptido de la presente invención puede ser un polipéptido recombinante, un polipéptido natural o un polipéptido sintético. En algunas realizaciones particulares, el polipéptido es un polipéptido recombinante.

En la presente invención también se proporcionan variantes alélicas, ortólogos y/u homólogos de especies. Utilizando información de las secuencias desveladas en el presente documento o los clones depositados en la ATCC, pueden utilizarse procedimientos conocidos en la técnica para obtener genes de longitud completa, variantes alélicas, variantes de corte y empalme, partes codificantes de longitud completa, ortólogos y/u homólogos de especies de genes correspondientes a cualquiera de las SEQ ID NO: 1-40. Por ejemplo, construyendo sondas o cebadores adecuados a partir de las secuencias proporcionadas en el presente documento y explorando una fuente de ácido nucleico adecuada para las variantes alélicas y/o el homólogo deseado, pueden aislarse e identificarse variantes alélicas y/u homólogos de especies.

Secuencias consenso de celulasas

En algunas realizaciones de la presente invención, las células hospedadoras expresan al menos una celulasa heteróloga que no procede de ningún organismo particular, sino que tiene una secuencia de aminoácidos adicional que es una secuencia celulasa consenso. La secuencia celulasa consenso puede ser una secuencia consenso endoglucanasa, una secuencia consenso β -glucosidasa o una secuencia consenso celobiohidrolasa.

En una realización particular, la celulasa heteróloga es una secuencia consenso de CBH1. Por lo tanto, en una realización, la invención se refiere a una secuencia polipeptídica que comprende una secuencia que es al menos 80 %, 85 %, 90 %, 95 %, 98 % o 99 % idéntica a la secuencia consenso de CBH1 de SEQ ID NO: 43. En algunas realizaciones, la invención se refiere a un polipéptido que comprende la secuencia de SEQ ID NO: 43.

La invención también se refiere a células hospedadoras que comprende una secuencia polipeptídica que comprende una secuencia que es al menos 80 %, 85 %, 90 %, 95 %, 98 % 99 % o 100 % idéntica a la secuencia consenso de CBH1 de SEQ ID NO: 43. La invención se refiere adicionalmente a células hospedadoras que comprenden un polinucleótido que codifica una secuencia polipeptídica que comprende una secuencia que es al menos 80 %, 85 %, 90 %, 95 %, 98 % 99 % o 100 % idéntica a la secuencia consenso de CBH1 de SEQ ID NO:43. En algunas realizaciones, la célula hospedadora comprende al menos un polinucleótido que codifica una secuencia polipeptídica que comprende una secuencia que es al menos 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % idéntica a la secuencia consenso de CBH1 de SEQ ID NO:43 y al menos un segundo polinucleótido que codifica una celulasa heteróloga. El segundo polinucleótido puede codificar una endoglucanasa, una β -glucosidasa, una celobiohidrolasa, una secuencia consenso de endoglucanasa, una secuencia consenso de β -glucosidasa o una secuencia consenso de celobiohidrolasa. En algunas realizaciones, la célula hospedadora que comprende el polinucleótido que codifica una secuencia polipeptídica que comprende una secuencia que es al menos 80 %, 85 %, 90 %, 95 %, 98 %, 99 %, o 100 % idéntica a la secuencia consenso de CBH1 de SEQ ID NO: 43, puede producir etanol cuando se cultiva utilizando celulosa como fuente de carbono.

Combinaciones de celulasas

En algunas realizaciones de la presente invención las células hospedadoras expresan una combinación de celulasas heterólogas. Por ejemplo, la célula hospedadora puede contener al menos dos celulasas heterólogas, al menos tres celulasas heterólogas, al menos cuatro celulasas heterólogas, al menos cinco celulasas heterólogas, al menos seis celulasas heterólogas, al menos siete celulasas heterólogas, al menos ocho celulasas heterólogas, al menos nueve celulasas heterólogas, al menos diez celulasas heterólogas, al menos once celulasas heterólogas, al menos doce celulasas heterólogas, al menos trece celulasas heterólogas, al menos catorce celulasas heterólogas o al menos quince celulasas heterólogas. Las celulasas heterólogas en la célula hospedadora pueden ser de la misma especie o de especies diferentes.

En algunas realizaciones de la presente invención, las células hospedadoras expresan una combinación de celulasas heterólogas que incluye al menos una endoglucanasa, al menos una β -glucosidasa y al menos una celobiohidrolasa. En otra realización de la invención, las células hospedadoras expresan una combinación de celulasas heterólogas que incluye al menos una endoglucanasa, al menos una β -glucosidasa y al menos dos celobiohidrolasas. Las al menos dos celobiohidrolasas pueden ser ambas celobiohidrolasa I, ambas celobiohidrolasa II, o una puede ser celobiohidrolasa I y la otra celobiohidrolasa II.

En una realización particular de la invención, las células hospedadoras expresan una combinación de celulasas que incluye una endoglucanasa I de *C. formosanus* y una β -glucosidasa I de *S. fibuligera*. En otra realización de la invención, las células hospedadoras expresan una combinación de celulasas que incluye una celobiohidrolasa I de *T. emersonii* y una celobiohidrolasa II de *T. reesei*.

En otra realización más, las células hospedadoras expresan una combinación de celulasas que incluye una endoglucanasa I de *C. formosanus* una β -glucosidasa I de *S. fibuligera*, una celobiohidrolasa I de *T. emersonii* y una celobiohidrolasa II de *C. lucknowense*. En otra realización más, las células hospedadoras expresan una combinación de celulasas que incluye una endoglucanasa I de *C. formosanus*, una β -glucosidasa I de *S. fibuligera*, una

celobiohidrolasa I de *T. emersonii* y una celobiohidrolasa II de *T. reesei*. En otra realización adicional las células hospedadoras expresan una combinación de celulasas que incluye una endoglucanasa 2 de *H. jecorina* una β -glucosidasa I de *S. fibuligera*, una celobiohidrolasa I de *T. emersonii* y una celobiohidrolasa II de *T. reesei*. En otra realización adicional, las células hospedadoras expresan una combinación de celulasas que incluye una endoglucanasa 2 de *H. jecorina*, una β -glucosidasa I de *S. fibuligera*, una celobiohidrolasa I de *T. emersonii* y una celobiohidrolasa II de *C. lucknowense*.

Celulasas ligadas y secretadas

De acuerdo con la presente invención, las celulasas pueden estar ligadas o secretadas. Como se indica en el presente documento, una proteína está "ligada" a la superficie celular de un organismo si al menos un extremo de la proteína está unido, por unión covalente y/o electrostática, por ejemplo, a la membrana celular o a la pared celular. Se apreciará que una proteína ligada puede incluir una o más regiones enzimáticas que pueden unirse a uno o más tipos de regiones distintas a nivel de ácido nucleico y de proteína (por ejemplo, un promotor, un terminador, un dominio de anclaje, un enlazador, una región de señalización, etc.). Sin embargo, aunque la una o más regiones enzimáticas pueden no estar directamente unidas a la membrana celular o a la pared celular (por ejemplo, tal como cuando se produce la unión mediante un dominio de anclaje), la proteína se considera una "enzima ligada" de acuerdo con la presente memoria descriptiva.

El ligamiento puede realizarse, por ejemplo, incorporando un dominio de anclaje en una proteína recombinante que se expresa de una manera heteróloga en una célula o por prenilación, enlace acil graso, anclajes de glicosil fosfatidil inositol u otros anclajes moleculares adecuados que puedan anclar la proteína ligada a la membrana celular o a la pared celular de la célula hospedadora. Una proteína ligada puede estar ligada en su extremo amino terminal u opcionalmente en su extremo carboxilo terminal.

Según se indica en el presente documento "secretada" significa liberada en el entorno extracelular, por ejemplo, en el medio. Aunque las proteínas ligadas pueden tener señales de secreción como parte de su secuencia de aminoácidos inmadura, estas se mantienen unidas a la superficie celular y no se incluyen dentro del ámbito de las proteínas secretadas tal y como se indica en el presente documento.

Como se indica en el presente documento, una "secuencia enlazadora flexible" se refiere a una secuencia de aminoácidos que enlaza dos secuencias de aminoácidos, por ejemplo, una secuencia de aminoácidos de anclaje a la pared celular con una secuencia de aminoácidos que contiene la actividad enzimática deseada. La secuencia enlazadora flexible permite libertad necesaria a la secuencia de aminoácidos que contiene la actividad enzimática deseada que tiene impedimento estérico reducido con respecto a la proximidad a la célula y también puede facilitar un plegamiento apropiado de la secuencia de aminoácidos que contiene la actividad enzimática deseada.

En algunas realizaciones de la presente invención, las enzimas celulasas ligadas están ligadas mediante una secuencia enlazadora flexible ligada a un dominio de anclaje. En algunas realizaciones, el dominio de anclaje es de CWP2 (para el anclaje carboxilo terminal) o de FLO1 (para el anclaje amino terminal) de *S. cerevisiae*.

En algunas realizaciones, pueden añadirse señales de secreción heterólogas a los vectores de expresión para facilitar la expresión extracelular de las proteínas celulasas. En algunas realizaciones, la señal de secreción heteróloga es la señal de secreción de Xyn2 de *T. reesei*.

Proteínas de fusión que comprenden celulasas

La presente invención también incluye proteína de fusión. Por ejemplo, las proteínas de fusión pueden ser una fusión de una celulasa heteróloga y de un segundo péptido. La celulasa heteróloga y el segundo péptido pueden fusionarse directa o indirectamente, por ejemplo, a través de una secuencia enlazadora. La proteína de fusión puede comprender, por ejemplo, un segundo péptido que está en posición N terminal con respecto a la celulasa heteróloga y/o un segundo péptido que está en posición C terminal con respecto a la celulasa heteróloga. Por tanto, en determinadas realizaciones, el polipéptido de la presente invención comprende un primer polipéptido y un segundo polipéptido, en el que el primer polipéptido comprende una celulasa heteróloga.

De acuerdo con la presente invención, la proteína de fusión puede comprender un primer y un segundo polipéptido en el que el primer polipéptido comprende una celulasa heteróloga y el segundo polipéptido comprende una secuencia señal. De acuerdo con otra realización, la proteína de fusión puede comprender un primer y un segundo polipéptido, en el que el primer polipéptido comprende una celulasa heteróloga y el segundo polipéptido comprende un polipéptido que se usa para facilitar la purificación o la identificación de un péptido indicador. El polipéptido que se utiliza para facilitar la purificación o identificación del péptido indicador puede ser, por ejemplo, una etiqueta de HIS, una etiqueta de GST, una etiqueta de HA, una etiqueta de FLAG, una etiqueta de MYC o una proteína fluorescente.

De acuerdo con otra realización adicional, la proteína de fusión puede comprender un primer y un segundo polipéptido, en el que el primer polipéptido comprende una celulasa heteróloga y el segundo polipéptido comprende un péptido de anclaje. En algunas realizaciones, el dominio de anclaje es de CWP2 (para el anclaje carboxilo

terminal) o de FLO1 (para el anclaje amino terminal) de *S. cerevisiae*.

De acuerdo con otra realización adicional, la proteína de fusión puede comprender un primer y un segundo polipéptido, en el que el primer polipéptido comprende una celulasa heteróloga y el segundo polipéptido comprende un módulo de unión a celulosa (MUC). En algunas realizaciones, el MUC es, por ejemplo, de Cbh1 o Cbh2 de *T. reesei*, de Cbh1 de *H. grisea* o de Cbh2b de *C. lucknowense*. En algunas realizaciones particulares, el MUC está fusionado a una celobiohidrolasa. En una realización particular, la proteína de fusión comprende un primer y un segundo polipéptido, en el que el primer polipéptido comprende una celobiohidrolasa heteróloga y el segundo polipéptido comprende un MUC. En otra realización particular, la celobiohidrolasa es celobiohidrolasa I de *T. emersonii* y el MUC es un MUC de celobiohidrolasa de *T. reesei*. En otra realización particular más, la celobiohidrolasa es celobiohidrolasa I de *T. emersonii* y el MUC es un MUC de celobiohidrolasa de *H. grisea*. En algunas realizaciones, el MUC de *H. grisea* comprende los aminoácidos 492-525 e SEQ ID NO: 21.

En determinadas realizaciones, el polipéptido de la presente invención incluye una proteína de fusión que comprende un primer polipéptido y un segundo polipéptido, en el que el primer polipéptido es una celobiohidrolasa y el segundo polipéptido es un dominio o un fragmento de una celobiohidrolasa. En determinadas realizaciones, el polipéptido de la presente invención incluye una proteína de fusión que comprende un primer polipéptido, en el que el primer polipéptido es una Cbh1 de *T. emersonii*, una Cbh1 de *H. grisea*, una Cbh1 de *T. aurantiacus*, una Cbh2 de *T. emersonii*, una Cbh1 de *T. reesei*, una Cbh2 de *T. reesei*, una Cbh2b de *C. lucknowense*, o un dominio, un fragmento, una variante o un derivado de los mismos, y un segundo polipéptido, en el que el segundo polipéptido es una Cbh1 de *T. emersonii*, una Cbh1 de *H. grisea* o una Cbh1 de *T. aurantiacus*, una Cbh2 de *T. emersonii*, una Cbh1 de *T. reesei* o una Cbh2 de *T. reesei*, una Cbh2b de *C. lucknowense*, o un dominio, un fragmento, una variante o un derivado de los mismos. En realizaciones particulares el primer polipéptido es la Cbh1 de *T. emersonii* y el segundo polipéptido es un MUC de Cbh1 o Cbh2 de *T. reesei* o de Cbh2b de *C. lucknowense*. En realizaciones adicionales, el primer polipéptido está en posición N terminal o C terminal con respecto al segundo polipéptido. En determinadas otras realizaciones, el primer polipéptido y/o el segundo polipéptido están codificados por polinucleótidos con codones optimizados, por ejemplo, polinucleótidos con codones optimizados para *S. cerevisiae* o *Kluyveromyces*. En realizaciones particulares, el primer polinucleótido es una cbh1 de *T. emersonii* con codones optimizados y el segundo polipéptido codifica un MUC con codones optimizados de Cbh1 o Cbh2 de *T. reesei*. En otras realizaciones particulares, el primer polinucleótido es una cbh1 de *T. emersonii* con codones optimizados y el segundo polinucleótido codifica un MUC con codones optimizados de *C. lucknowense* o Cbh2b.

En determinadas otras realizaciones, el primer polipéptido y el segundo polipéptido están fusionados mediante una secuencia enlazadora. La secuencia enlazadora puede, en algunas realizaciones, estar codificada por un polinucleótido con codones optimizados. (Los polinucleótidos con codones optimizados se describen más adelante con más detalle). Una secuencia de aminoácidos correspondiente a un enlazador 1 con codones optimizados de acuerdo con la invención es una fusión de enlazador flexible - etiqueta de strep - sitio TEV- FLAG - enlazador flexible fusión y corresponde a la secuencia GGGGSGGGGS AWHPQFGG ENLYFQG DYKDDDK GGGGSGGGGS (SEQ ID NO:57)

La secuencia de ADN es como se indica a continuación:

**GGAGGAGGTGGTTCAGGAGGTGGTGGGTCTGCTTGCCAT
CCACAATTTGGAGGAGGCGGTGGTGAAAATCTGTATTC
CAGGGAGGCGGAGGTGATTACAAGGATGACGACAAAGG
AGGTGGTGGATCAGGAGGTGGTGGCTCC (SEQ ID NO:41)**

Una secuencia de aminoácidos correspondiente a un enlazador 2 con codones optimizados es un enlazador flexible - etiqueta de strep - enlazador - sitio TEV - enlazador flexible y corresponde a la secuencia GGGGSGGGGS WSHQFEK GG ENLYFQG GGGGSGGGGS (SEQ ID NO:58). La secuencia de ADN es como se indica a continuación:

**ggtggcgggtggaatctggaggaggcgggttcttggctcaccacaatttgaaaagggtggagaaaactgtactttcaa
ggcgggtggagggttctggcggagggtgctccggctca (SEQ ID NO:42)**

Cocultivos

La presente invención también se refiere a cocultivos que comprenden al menos dos células hospedadoras de levadura, en el que las al menos dos células hospedadoras de levadura comprenden un polinucleótido aislado que codifica una celulasa heteróloga. Como se indica en el presente documento, un "cocultivo" se refiere al crecimiento conjunto de dos cepas o especies diferentes de células hospedadoras en el mismo recipiente. En algunas realizaciones de la invención, al menos una célula hospedadora del cocultivo comprende un polinucleótido heterólogo que comprende un ácido nucleico que codifica una endoglucanasa, comprendiendo al menos una célula hospedadora del cocultivo un polinucleótido heterólogo que comprende un ácido nucleico que codifica una β -glucosidasa y al menos una célula hospedadora comprende un polinucleótido heterólogo que comprende un ácido nucleico que codifica una celobiohidrolasa. En una realización adicional, el cocultivo comprende además una célula hospedadora que comprende un polinucleótido heterólogo que comprende un ácido nucleico que codifica una segunda celobiohidrolasa.

El cocultivo puede comprender dos o más cepas de células hospedadoras de levadura y las celulasas heterólogas pueden expresarse en cualquier combinación en las dos o más cepas de células hospedadoras. Por ejemplo, de acuerdo con la presente invención, el cocultivo puede comprender dos cepas: una cepa de las células hospedadoras que expresa una endoglucanasa y una segunda cepa de células hospedadoras que expresa una β -glucosidasa, una celobiohidrolasa y una segunda celobiohidrolasa. De acuerdo con la presente invención, el cocultivo también puede comprender cuatro cepas: una cepa de células hospedadoras que expresa una endoglucanasa, una cepa de células hospedadoras que expresa una β -glucosidasa, una cepa de células hospedadoras que expresa una primera celobiohidrolasa, y una cepa de células hospedadoras que codifica una segunda celobiohidrolasa. De manera similar, el cocultivo puede comprender una cepa de células hospedadoras que expresa dos celulasas, por ejemplo, una endoglucanasa y una β -glucosidasa y una segunda cepa de células hospedadoras que expresa dos celulasas, por ejemplo, una endoglucanasa y una beta-glucosidasa y una segunda cepa de células hospedadoras que expresa una o más celulasas, por ejemplo, una o más celobiohidrolasas. El cocultivo también puede incluir, además de las al menos dos células hospedadoras que comprenden celulasas heterólogas, otras células hospedadoras que no comprenden celulasas heterólogas.

En el cocultivo las diversas cepas de células hospedadoras pueden estar presentes en igual número, o una cepa o especie de célula hospedadora puede superar significativamente en número a otra segunda cepa o especie de célula hospedadora. Por ejemplo, en un cocultivo que comprende dos cepas o especies de células hospedadoras la proporción de una célula hospedadora con respecto a otra puede ser de aproximadamente 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:100, 1:500 o 1:1000. De manera similar, en un cocultivo que comprende tres o más cepas o especies de células hospedadoras, las cepas o especies de células hospedadoras pueden estar presentes en igual o distinto número.

Los cocultivos de la presente invención pueden incluir celulasas ligadas, celulasas secretadas o celulasas tanto ligadas como secretadas. Por ejemplo, en algunas realizaciones de la invención, el cocultivo comprende al menos una célula hospedadora de levadura que comprende un polinucleótido que codifica una celulasa heteróloga secretada. En otra realización, el cocultivo comprende al menos una célula hospedadora de levadura que comprende un polinucleótido que codifica una celulasa heteróloga ligada. En otra realización, todas las celulasas heterólogas en el cocultivo se secretan y en otra realización, todas las celulasas heterólogas en el cocultivo están ligadas. Además, en el cocultivo puede haber otras celulasas, tales como células externamente añadidas.

Polinucleótidos que codifican celulasas heterólogas

La presente invención también incluye polinucleótidos aislados que codifican celulasas de la presente invención. Por tanto, los polinucleótidos de la invención pueden codificar endoglucanasas o exoglucanasas. Los polinucleótidos pueden codificar endoglucanasas, β -glucosidasas o celobiohidrolasas.

En algunas realizaciones particulares de la invención, el polinucleótido codifica una endoglucanasa que es una endo-1,4- β -glucanasa. En realizaciones particulares, el polinucleótido codifica una endoglucanasa I de *Trichoderma reesei*. En determinadas otras realizaciones, la endoglucanasa está codificada por un polinucleótido que comprende la secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, por 100 % idéntica a la SEQ ID NO:19. En realizaciones particulares, el polinucleótido codifica una endoglucanasa I de *C. formosanus*. En determinadas otras realizaciones, la endoglucanasa está codificada por un polinucleótido que comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100 % idéntica a la SEQ ID NO: 11. En realizaciones particulares, el polinucleótido codifica una endoglucanasa I de *Trichoderma reesei*. En determinadas otras realizaciones, la endoglucanasa está codificada por un polinucleótido que comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100 % idéntica a la SEQ ID NO:19. En realizaciones particulares, el polinucleótido codifica una endoglucanasa 2 de *H. jecorina*. En determinadas otras realizaciones, la endoglucanasa está codificada por un polinucleótido que comprende una

secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100 % idéntica a la SEQ ID NO:54.

5 En determinadas realizaciones, el polinucleótido codifica una isoforma, un parálogo o un ortólogo de una β -glucosidasa I o una β -glucosidasa II. En determinadas realizaciones de la presente invención, el polinucleótido codifica una β -glucosidasa derivada de *Saccharomycopsis fibuligera*. En realizaciones particulares, la β -glucosidasa está codificada por un polinucleótido que comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100 % idéntica a la SEQ ID NO:20.

15 En determinadas realizaciones de la invención, el polinucleótido codifica una isoforma, un parálogo o un ortólogo de celobiohidrolasa I y/o celobiohidrolasa II. En realizaciones particulares de la presente invención, el polinucleótido codifica una celobiohidrolasa I o II de *Trichoderma reesei*. En realizaciones particulares de la presente invención, el polinucleótido codifica una celobiohidrolasa I o II de *Trichoderma emersonii*. En otra realización, a la celobiohidrolasa está codificada por un polinucleótido que comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100 % idéntica a la SEQ ID NO:7 o SEQ ID NO:8. En realizaciones particulares de la presente invención, el polinucleótido codifica una celobiohidrolasa de *C. lucknowense*. En otra realización, la celobiohidrolasa está codificada por un polinucleótido que comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100 % idéntica a la SEQ ID NO:5.

25 En realizaciones adicionales el polinucleótido es un polipéptido que comprende una secuencia al menos aproximadamente 70, aproximadamente 80, aproximadamente 90, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98, aproximadamente 99, o 100 % idéntica a una secuencia de nucleótidos indicada en la Tabla 1. En determinados aspectos el polinucleótido puede codificar una endoglucanasa, una celobiohidrolasa o β -glucosidasa procedente, por ejemplo, de una fuente fúngica, bacteriana, protozoaria o de termita.

30 En determinados aspectos, la presente invención se refiere a un polinucleótido que comprende un ácido nucleico que codifica un dominio funcional o estructural de Cbh1 o Cbh2 de *T. emersonii*, *H. grisea*, *T. aurantiacus*, *C. lucknowense* o *T. reesei*. Por ejemplo, los dominios de Cbh1 de *T. reesei* incluyen, sin limitación: (1) una secuencia señal, del aminoácido 1 al 33 de SEQ ID NO: 27; (2) un dominio catalítico (DC) desde aproximadamente el aminoácido 41 a aproximadamente el aminoácido 465 de SEQ ID NO: 27; y (3) un módulo de unión a celulosa (MUC) desde aproximadamente el aminoácido 503 a aproximadamente el aminoácido 535 de SEQ ID NO: 27. Los dominios de Cbh2 de *T. reesei* incluyen, sin limitación: (1) una secuencia señal, desde el aminoácido 1 al 33 de SEQ ID NO: 27; (2) un dominio catalítico (DC) desde aproximadamente el aminoácido 141 a aproximadamente el aminoácido 454 de SEQ ID NO: 27; y (3) un módulo de unión a celulosa (MUC) desde aproximadamente el aminoácido 52 a aproximadamente el aminoácido 83 de SEQ ID NO: 27.

45 La presente invención también incluye un polinucleótido aislado que comprende un ácido nucleico que es al menos aproximadamente 70 %, 75 % u 80 % idéntico, al menos de aproximadamente 90 % a aproximadamente 95 % idéntico, o al menos aproximadamente 96 %, 97 %, 98 %, 99 % o 100 % idéntico, a un ácido nucleico que codifica un dominio de Cbh1 o Cbh2 de *T. emersonii*, *H. grisea*, *T. aurantiacus*, *C. lucknowense* o *T. reesei*, como se ha descrito anteriormente.

50 La presente invención también incluye variantes de los genes de celulasa, como se ha descrito anteriormente. Las variantes pueden contener alteraciones en las regiones codificantes, en las regiones no codificantes o en ambas regiones. Son ejemplos variantes polinucleotídicas que contiene alteraciones que producen sustituciones, adiciones o deleciones silenciosas, pero que no alteran las propiedades o las actividades del polipéptido codificado. En determinadas realizaciones, las variantes nucleotídicas se producen por sustituciones silenciosas debido a la generación del código genético. En realizaciones adicionales, pueden producirse variantes polinucleotídicas de celulasas de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, y *Arabidopsis thaliana* por diversas razones, por ejemplo, para optimizar con codones la expresión de un hospedador particular. Más adelante se analizan adicionalmente polinucleótidos con codones optimizados de la presente invención.

65 La presente invención también incluye un polinucleótido aislado que codifica una proteína de fusión. En determinadas realizaciones el ácido nucleico que codifica una proteína de fusión comprende un primer polinucleótido que codifica una Cbh1 de *T. emersonii*, una Cbh1 de *H. grisea*, una CBH1 de *T. aurantiacusi* o una Cbh1 de *T. emersonii* y un segundo polinucleótido que codifica el dominio del MUC de la cbh1 de *T. reesei* o de la cbh2 de *T.*

reesei o de la *cbh2b* de *C. lucknowense*. En realizaciones particulares del ácido nucleico que codifica una proteína de fusión, el primer polinucleótido codifica una *cbh1* de *T. emersonii* y el segundo polinucleótido codifica un MUC de la *Cbh1* o de la *Cbh2* de *T. reesei*.

5 En realizaciones adicionales, el primer y segundo polinucleótido están en la primera orientación, o el segundo polinucleótido está en orientación inversa con respecto al primer polinucleótido. En realizaciones adicionales, el primer polinucleótido codifica un polipéptido que está en posición N terminal o C terminal con respecto al polipéptido codificado por el segundo polipéptido. En determinadas otras realizaciones, el primer polinucleótido y/o el segundo polinucleótido están codificados por polinucleótidos con codones optimizados, por ejemplo, polinucleótidos con codones optimizados para *S. cerevisiae*, *Kluyveromyces* o tanto para *S. cerevisiae* como para *Kluyveromyces*. En realizaciones particulares del ácido nucleico que codifica una proteína de fusión, el primer polinucleótido es una *cbh1* de *T. emersonii* con codones optimizados y el segundo polinucleótido codifica un MUC con codones optimizados de *Cbh1* o *Cbh2* de *T. reesei*.

15 En la presente invención también se proporcionan variantes alélicas, ortólogos y/u homólogos de especies. Para obtener genes de longitud completa, variantes alélicas, variantes de corte y empalme, partes codificantes de longitud completa, ortólogos y/u homólogos de especies de genes correspondientes a cualquiera de las SEQ ID NO: 1-20, pueden utilizarse procedimientos conocidos en la técnica utilizando información de las secuencias desveladas en el presente documento o de los clones depositados en la ATCC. Por ejemplo, a partir de las secuencias proporcionadas en el presente documento, pueden aislarse variantes alélicas y/u homólogos de especies, y pueden identificarse para construir sondas o cebadores adecuados y explorar una fuente de ácido nucleico que sea adecuada para las variantes alélicas y/o el homólogo deseado.

25 Por un ácido nucleico que tiene una secuencia de nucleótidos al menos, por ejemplo, 95 % "idéntica" a una secuencia de nucleótidos de referencia de la presente invención, se entiende que la secuencia de nucleótidos del ácido nucleico es idéntica a la secuencia de referencia excepto que la secuencia de nucleótidos puede incluir hasta cinco mutaciones puntuales por cada 100 nucleótidos de la secuencia de nucleótidos de referencia que codifica el polipéptido particular. En otras palabras, para obtener un ácido nucleico que tenga una secuencia de nucleótidos al menos 95 % idéntica a una secuencia de nucleótidos de referencia, hasta el 5 % de los nucleótidos en la secuencia de referencia puede delecionarse o sustituirse por otro nucleótido, o por diversos nucleótidos, hasta el 5 % del total de nucleótidos en la secuencia de referencia pueden insertarse en la secuencia de referencia. La secuencia problema puede ser una secuencia entera mostrada de cualquiera de las SEQ ID NO: 1-20, o cualquier fragmento o dominio especificado como se describe en el presente documento.

35 Como un asunto práctico, utilizando programas informáticos conocidos puede determinarse de manera convencional si cualquier molécula de ácido nucleico o polipéptido particular es al menos 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntica a la secuencia de nucleótidos o polipéptido de la presente invención. Un método para determinar la mejor coincidencia total entre una secuencia problema (una secuencia de la presente invención) y una secuencia objeto, también denominado alineamiento global de secuencias, puede determinarse utilizando el programa informático FASTDB basado en el algoritmo de Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245.) En un alineamiento de secuencias, tanto la secuencias problema como la secuencia objeto son secuencias de ADN. Una secuencia de ARN puede compararse convirtiendo el U en T. El resultado de dicho alineamiento de secuencia global es un porcentaje de identidad. Los parámetros preferidos utilizados en un alineamiento FASTDB de secuencias de ADN para calcular el porcentaje de identidad son: Matriz=Unitaria, k-tupla=4, Penalización por emparejamiento erróneo=1, Penalización por unión=30, Longitud de grupo de aleatorización=0, Puntuación de límite=1, Penalización por hueco=5, Penalización por tamaño de hueco 0,05, Tamaño de ventana=500 o la longitud de la secuencia de nucleótidos objeto, por muy corta que sea.

50 Si la secuencia objeto es más corta que la secuencia problema debido a delecciones 5' o 3', no debido a delecciones internas, puede realizarse una corrección manual en los resultados. Esto es porque el programa FASTDB no explica los truncamientos 5' y 3' de la secuencia objeto cuando se calcula el porcentaje de identidad. Para secuencias objeto truncadas en los extremos 5' o 3', con respecto a la secuencia problema, el porcentaje de identidad se corrige calculando el número de bases de la secuencia problema que están en 5' y 3' de la secuencia objeto, que no están emparejadas/alienadas, como un porcentaje del total de bases de la secuencia problema. Si un nucleótido está emparejado/alineado se determina por resultados de la secuencia FASTDB del alineamiento de la secuencia FASTDB. Este porcentaje se resta después del porcentaje de identidad, calculado por el programa FASTDB anterior utilizando los parámetros especificados, para llegar a una puntuación de porcentaje de identidad final. Esta puntuación corregida es la que se utiliza para los fines de la presente invención. Únicamente las bases que están fuera de las bases en 5' y 3' de las secuencias objeto, como se muestra en alineamiento FASTDB, que no están emparejadas/alineadas con la secuencia problema, se calculan para los fines de ajustar manualmente la puntuación del porcentaje de identidad.

65 Por ejemplo, para determinar el porcentaje de identidad, una secuencia objeto de 90 bases se alinea con una secuencia problema de 100 bases. Las delecciones se producen en el extremo 5' de la secuencia objeto y, por lo tanto, el alineamiento FASTDB no muestra un emparejamiento/alineamiento de las 10 primeras bases en el extremo 5'. Las 10 bases no apareadas representan el 10 % de la secuencia (número de bases en los extremos 5' y 3' no

emparejados/número total de bases en la secuencia problema) de tal manera que el 10 % se resta de la puntuación del porcentaje de identidad calculado por el programa FASTDB. Si las 90 bases restantes fuesen perfectamente coincidentes el porcentaje de identidad final debería ser 90 %. En otro ejemplo, una secuencia objeto de 90 bases se compara con una secuencia problema de 100 bases. Esta vez las deleciones son deleciones internas de tal manera que no hay bases en el 5' o 3' de la secuencia objeto que no están emparejadas/alineadas con la secuencia problema. En este caso, el porcentaje de identidad calculado por FASTDB no se corrige manualmente. De nuevo, únicamente las bases en 5' y 3' de la secuencia objeto que no están emparejadas/alineadas con la secuencia problema se corrigen manualmente. Para los fines de la presente invención no se realiza ninguna otra corrección manual.

Algunas realizaciones de la invención incluyen una molécula de ácido nucleico que comprende al menos 10, 20, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, u 800 nucleótidos consecutivos o más de cualquiera de las SEQ ID NO: 1-20, o dominios, fragmentos, variantes, o derivados de las mismas.

El polinucleótido de la presente invención puede estar en forma de ARN o de ADN, cuyo ADN incluye ADNc, ADN genómico y ADN sintético. El ADN puede ser bicatenario o monocatenario, y si es monocatenario, puede ser la cadena codificante o no codificante (antisentido). La secuencia codificante que codifica el polipéptido maduro puede ser idéntica a la secuencia codificante que codifica las SEQ ID NO: 21-40, 46 o 52-56, o puede ser una secuencia codificante diferente, cuya secuencia codificante, como resultado de la redundancia o degeneración del código genético, codifica el mismo polipéptido maduro que el ADN de una cualquiera de las SEQ ID NO: 21-40, 46 o 52-56.

En determinadas realizaciones, la presente invención proporciona un polinucleótido aislado que comprende un fragmento de ácido nucleico que codifica al menos 10, al menos 20, al menos 30, al menos 40, al menos 50, al menos 60, al menos 70, al menos 80, al menos 90, al menos 95, o al menos 100 o más aminoácidos contiguos de las SEQ ID NO: 21-40, 46 o 52-56.

El polinucleótido que codifica al polipéptido maduro de las SEQ ID NO: 21-40, 46 o 52-56 puede incluir: únicamente la secuencia codificante del polipéptido maduro; la secuencia codificante de cualquier dominio del polipéptido maduro; y la secuencia codificante del polipéptido maduro (o secuencia codificante del dominio) junto con la secuencia no codificante, tal como intrones o la secuencia no codificante 5' y/o 3' de la secuencia codificante del polipéptido maduro.

Por tanto, la expresión "polinucleótido que codifica un polipéptido" incluye un polipéptido que incluye solo secuencias que codifican al polipéptido, así como un polinucleótido que incluye secuencias codificantes y/o no codificantes adicionales.

En aspectos adicionales de la invención, las moléculas de ácido nucleico tienen secuencias al menos aproximadamente 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idénticas a la secuencias de ácido nucleico desveladas en el presente documento, codifican un polipéptido que tiene actividad funcional celobiohidrolasa ("Cbh"), endoglucanasa ("Eg") o beta-gluconasa ("Bgl"). Por "un polipéptido que tiene actividad funcional Cbh, Eg o Bgl" se entiende polipéptidos que presentan actividad similar, pero no necesariamente idéntica, a una actividad funcional de los polipéptidos Cbh, Eg o Bgl de la presente invención, medida, por ejemplo, en un ensayo biológico particular. Por ejemplo, una actividad funcional Cbh, Eg o Bgl puede medirse habitualmente determinando la capacidad de una Cbh, Eg o Bgl de hidrolizar celulosa, o midiendo el nivel de actividad de Cbh, Eg o Bgl.

Por supuesto, debido a la degeneración del código genético, un experto habitual en la técnica reconocerá inmediatamente que una gran parte de las moléculas de ácido nucleico que tienen una secuencia al menos 90 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntica a la secuencia de ácido nucleico de cualquiera de las SEQ ID NO: 1-20, o fragmentos de las mismas, codificarán polipéptidos que tienen actividad funcional Cbh, Eg o Bgl. De hecho, dado que todas las variantes degeneradas de cualquiera de estas secuencias de nucleótidos, codificarán el mismo polipéptido, en muchos casos, esto estará claro para el experto en la técnica incluso sin realizar el ensayo de comparación descrito anteriormente. Adicionalmente se reconocerá en la técnica que, para que dichas moléculas de ácido nucleico, que no sean variantes degeneradas, un número razonable también codificará un polipeptido que tenga actividad funcional Cbh, Eg o Bgl.

Los polinucleótidos de la presente invención también comprenden ácidos nucleicos que codifican una celulasa, o dominio, fragmento, variante, o derivado de los mismos, de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* o de *Arabidopsis thaliana*, fusionados a un polipéptido que codifica una secuencia marcadora que permite la detección del polinucleótido de la presente invención. En una realización de la invención, la expresión del marcador es independiente de la expresión de la celulasa. La secuencia marcadora puede ser un marcador de selección de levadura seleccionado del grupo que consiste en URA3, HIS3, LEU2, TRP1, LYS2 o ADE2. Casey,

G.P. et al., "A convenient dominant selection marker for gene transfer in industrial strains of *Saccharomyces yeast*: SMR1 encoded resistance to the herbicide sulfometuron methyl," *J. Inst. Brew.* 94:93-97 (1988).

Polinucleótidos con codones optimizados

5 De acuerdo con una realización de la invención, los codones de los polinucleótidos que codifican celulasas heterólogas pueden optimizarse. Como se usa en el presente documento, la expresión "región codificante con codones optimizados" significa una región codificante de ácido nucleico que se ha adaptado para la expresión en las células de un organismo determinado, reemplazando al menos uno, o más de uno, o un número significativo de
10 codones, por uno o más codones que se utilizan más frecuentemente en los genes de ese organismo.

En general, los genes altamente expresados en un organismo están sesgados hacia codones que reconocen la mayoría de las especies de ARNt abundantes en ese organismo. Una medición de este sesgo es el "índice de adaptación de codones" o "IAC", que mide el grado al cual los codones utilizados para codificar cada aminoácido en un gen particular son aquellos que se producen más frecuentemente en un conjunto de referencia de genes altamente expresados de un organismo.

El IAC de secuencias con codones optimizados de la presente invención corresponde a entre aproximadamente 0,8 y 1,0, entre aproximadamente 0,8 y 0,9 o aproximadamente 1,0. Una secuencia con codones optimizados puede modificarse adicionalmente para la expresión en un organismo particular, dependiendo de las limitaciones biológicas del organismo. Por ejemplo, pueden retirarse grandes series de "As" o "Ts" (por ejemplo, series mayores de 4, 4, 5, 6, 7, 8, 9, o 10 bases consecutivas) de las secuencias si se sabe que estas afectan negativamente la transcripción. Además, pueden retirarse sitios de enzimas de restricción específicos con fines de clonación molecular. Como ejemplos de dichos sitios de enzimas de restricción se incluyen PaeI, AclI, BamHI, BglII, EcoRI y XhoI.
25 Adicionalmente, la secuencia de ADN puede verificarse con respecto a repeticiones directas, repeticiones invertidas o repeticiones especulares con longitudes de diez bases o más largas, que pueden modificarse manualmente reemplazando codones con "segundos mejores" codones, es decir, codones que se producen en la segunda frecuencia más alta dentro del organismo particular para el que se está optimizando la secuencia.

Las desviaciones en la secuencia de nucleótidos que comprende los codones que codifican los aminoácidos de cualquier cadena polipeptídica, permite variaciones en la secuencia codificante para el gen. Dado que cada codón consta de tres nucleótidos, y que los nucleótidos que constituyen el ADN están limitados a cuatro bases específicas, hay 64 posibles combinaciones de nucleótidos, 61 de las cuales codifican aminoácidos (los tres codones restantes codifican señales de finalización de la traducción). El "código genético", que muestra cuales son los codones que codifican los aminoácidos, se reproduce en el presente documento como Tabla 2. Como resultado, muchos aminoácidos se designan mediante más de un codón. Por ejemplo, los aminoácidos alanina y prolina están codificados por cuatro tripletes, la serina y arginina por seis, mientras que el triptófano y la metionina están codificados por un solo triplete. Esta degeneración permite que la composición de bases del ADN varíe sobre un amplio intervalo sin alterar la secuencia de aminoácidos de las proteínas codificadas por el ADN.

Tabla 2: El código genético convencional

	T	C	A	G
T	TTT Phe (F) TTC" TTA Leu (L) TTG"	TCT Ser (S) TCC" TCA" TCG"	TAT Tyr (Y) TAC" TAA Ter TAG Ter	TGT Cys (C) TGC TGA Ter TGG Trp (W)
C	CTT Leu (L) CTC" CTA" CTG"	CCT Pro (P) CCC" CCA" CCG"	CAT His (H) CAC" CAA Gln (Q) CAG"	CGT Arg (R) CGC" CGA" CGG"
A	ATT Ile (I) ATC" ATA" ATG Met (M)	ACT Thr (T) ACC" ACA" ACG"	AAT Asn (N) AAC" AAA Lys (K) AAG"	AGT Ser (S) AGC" AGA Arg (R) AGG"
G	GTT Val (V) GTC" GTA" GTG"	GCT Ala (A) GCC" GCA" GCG"	GAT Asp (D) GAC" GAA Glu (E) GAG"	GGT Gly (G) GGC" GGA" GGG"

Muchos organismos presentan un sesgo para el uso de codones particulares para codificar la inserción de un aminoácido particular en una cadena peptídica en crecimiento. La degeneración del código genético permite establecer diferencias de preferencia codónica o sesgo codónico, en el uso de codones entre organismos, y está bien documentada en muchos organismos. El sesgo codónico a menudo se correlaciona con la eficacia de la traducción del ARN mensajero (ARNm) que, a su vez, se piensa que depende, entre otras cosas, de las propiedades de los codones que se están traduciendo y de la disponibilidad de moléculas de ARN de transferencia (ARNt) particulares. La predominancia de los ARNt seleccionados en una célula es generalmente un reflejo de los codones más frecuentemente utilizados en la síntesis peptídica. Por consiguiente, los genes pueden adaptarse para una expresión génica óptima en un organismo determinado basándose en la optimización de los codones.

Dado el gran número de secuencias génicas disponibles para una amplia variedad de especies de animales, plantas y microbios, es posible calcular las frecuencias relativas del uso de codones. Las tablas del uso de codones se encuentran fácilmente disponibles, por ejemplo, en la <http://phenotype.biosci.umbc.edu/codon/sgd/index.php> (visitada el 7 de mayo de 2008) o en la <http://www.kazusa.or.jp/codon/> (visitada el 20 de marzo de 2008) y estas tablas pueden adaptarse de diversas maneras. Véase nakamura, Y., et al "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Las tablas del uso de codones para levaduras, calculadas del GenBank Release 128.0 [15 de febrero de 2002], se reproducen en la Tabla 3 más adelante. Esta Tabla utiliza nomenclatura de ARNm, y por tanto en lugar de timina (T) que se encuentra en el ADN, las tablas utilizan uracilo (U) que se encuentra en el ARN. La Tabla se ha adaptado de manera que estas frecuencias se calculan para cada aminoácido, en lugar de para los 64 codones.

Tabla 3: Tabla de uso de codones para genes de *Saccharomyces cerevisiae*

Aminoácido	Codón	Número	Frecuencia por ciento
Phe	UUU	170666	26,1
Phe	UUC	120510	18,4
Total			
Leu	UUA	170884	26,2
Leu	UUG	177573	27,2
Leu	CUU	80076	12,3
Leu	CUC	35545	5,4
Leu	CUA	87619	13,4
Leu	CUG	68494	10,5
Total			
Ile	AUU	196893	30,1
Ile	AUC	112176	17,2
He	AUA	116254	17,8
Total			
Met	AUG	136805	20,9
Total			
Val	GUU	144243	22,1
Val	GUC	76947	11,8
Val	GUA	76927	11,8
Val	GUG	70337	10,8
Total			

ES 2 621 181 T3

Aminoácido	Codón	Número	Frecuencia por ciento
Ser	UCU	153557	23,5
Ser	UCC	92923	14,2
Ser	UCA	122028	18,7
Ser	UCG	55951	8,6
Ser	AGU	92466	14,2
Ser	AGC	63726	9,8
Total			
Pro	CCU	88263	13,5
Pro	CCC	44309	6,8
Pro	CCA	119641	18,3
Pro	CCG	34597	5,3
Total			
Thr	ACU	132522	20,3
Thr	ACC	83207	12,7
Thr	ACA	116084	17,8
Thr	ACG	52045	8,0
Total			
Ala	GCU	138358	21,2
Ala	GCC	82357	12,6
Ala	GCA	105910	16,2
Ala	GCG	40358	6,2
Total			
Tyr	UAU	122728	18,8
Tyr	UAC	96596	14,8
Total			
His	CAU	89007	13,6
His	CAC	50785	7,8
Total			
Gln	CAA	178251	27,3
Gln	CAG	79121	12,1
Total			

ES 2 621 181 T3

Aminoácido	Codón	Número	Frecuencia por ciento
Asn	AAU	233124	35,7
Asn	AAC	162199	24,8
Total			
Lys	AAA	273618	41,9
Lys	AAG	201361	30,8
Total			
Asp	GAU	245641	37,6
Asp	GAC	132048	20,2
Total			
Glu	GAA	297944	45,6
Glu	GAG	125717	19,2
Total			
Cys	UGU	52903	8,1
Cys	UGC	31095	4,8
Total			
Trp	UGG	67789	10,4
Total			
Arg	CGU	41791	6,4
Arg	CGC	16993	2,6
Arg	CGA	19562	3,0
Arg	CGG	11351	1,7
Arg	AGA	139081	21,3
Arg	AGG	60289	9,2
Total			
Gly	GGU	156109	23,9
Gly	GGC	63903	9,8
Gly	GGA	71216	10,9
Gly	GGG	39359	6,0
Total			

Aminoácido	Codón	Número	Frecuencia por ciento
Terminación	UAA	6913	1,1
Terminación	UAG	3312	0,5
Terminación	UGA	4447	0,7

Utilizando esta tabla, o tablas similares, un experto habitual en la técnica puede aplicar las frecuencias a cualquier secuencia polipeptídica determinada, y producir un fragmento de ácido nucleico de una región codificante con codones optimizados, que codifica el polipéptido, pero que utiliza codones óptimos para una especie determinada. Pueden diseñarse regiones codificantes con codones optimizados mediante diversos métodos diferentes.

En un método, se utiliza una tabla de uso de codones para encontrar el codón sencillo más frecuente utilizado para cualquier aminoácido determinado, y ese codón se utiliza cada vez que aparece ese aminoácido particular en la secuencia polipeptídica. Por ejemplo, referente a la Tabla 3 anterior, para la leucina, el codón más frecuente es UUG, que se utiliza el 27,2 % de las veces. Por tanto, todos los restos de leucina en una secuencia de aminoácidos determinada se asignarían al codón UUG.

En otro método, las frecuencias reales de los codones se distribuyen al azar a lo largo de la secuencia codificante. Por tanto, utilizando este método para la optimización, si una secuencia polipeptídica hipotética tuviese 100 restos de leucina, en referencia a la Tabla 3 para la frecuencia de uso en *S. cerevisiae*, aproximadamente 5, o el 5 % de los codones de leucina serían CUC, aproximadamente 11, o el 11 % de los codones de leucina serían CUG, aproximadamente 12, o el 12 % de los codones de leucina serían CUU, aproximadamente 13, o el 13 % de los codones de leucina serían CUA, aproximadamente 26, o el 26 % de los codones de leucina serían UUA, y aproximadamente 27, o el 27 % de los codones de leucina serían UUG.

Estas frecuencias se distribuirían al azar en todos los codones de leucina en la región codificante que codifica el polipéptido hipotético. Como entenderán los expertos habituales en la técnica, la distribución de codones en la secuencia puede variar significativamente utilizando este método; sin embargo, la secuencia siempre codifica el mismo polipéptido.

Cuando se utilizan los métodos anteriores, el término "aproximadamente" se usa precisamente para representar porcentajes fraccionados de frecuencias de codones para un aminoácido determinado. Como se usa en el presente documento, "aproximadamente" se define como un aminoácido más o un aminoácido menos que el valor determinado. El valor numérico íntegro de aminoácidos se redondea al alza si la frecuencia fraccionada de uso es de 0,50 o mayor, y se redondea a la baja si la frecuencia fraccionada de uso es de 0,49 o menor. Utilizando de nuevo el ejemplo de la frecuencia de uso de leucina en genes humanos para un polipéptido hipotético que tiene 62 restos de leucina, la frecuencia fraccionada de uso de codones podría calcularse multiplicando 62 por las frecuencias para los diversos codones. Por tanto, 7,28 % de 62 es igual a 4,51 codones UAA, o "aproximadamente 5", es decir, 4, 5 o 6 codones UAA, 12,66 % de 62 es igual a 7,85 codones UUG o "aproximadamente 8", es decir, 7, 8 o 9 codones UUG, 12,8 % de 62 es igual a 7,98 codones CUU, o "aproximadamente 8", es decir, 7, 8 o 9 codones CUU, 19,56 % de 62 es igual a 12, 13 codones CUC o "aproximadamente 12", es decir, 11, 12 o 13 codones CUC, 7,00 % de 62 es igual a 4,34 codones CUA o "aproximadamente 4", es decir, 3, 4 o 5 codones CUA, y 42,62 % de 62 es igual a 25,19 codones CUG, o "aproximadamente 25", es decir, 24, 25 o 26 codones CUG.

La asignación al azar de codones a una frecuencia optimizada para codificar una secuencia polipeptídica determinada, puede realizarse manualmente calculando las frecuencias de codones para cada aminoácido, y después asignando los codones a la secuencia polipeptídica al azar. Adicionalmente, los expertos en la técnica disponen de diversos algoritmos y programas informáticos. Por ejemplo, la función "EditSeq" en el paquete informático Lasergene, disponible en DNASTar., Madison, WI, la función de retrotraducción en el conjunto de programas VectorNTI, disponible en InforMax, Inc., Bethesda, MD, y la función "retrotraductora" en el paquete informático CGC Wisconsin, disponible en Accelrys, Inc., San Diego, CA. Además, diversas fuentes se encuentran disponibles al público para optimizar los codones de secuencias de regiones codificantes, es decir, la función "retrotraductora" en la <http://www.entelechon.com/bioinformatics/backtranslation.php?lang=eng> (visitada el 15 de abril de 2008) y la función "retrotransec" disponible en la <http://ibioinfo.pbi.nrc.ca:8090/EMBOSS/index.html> (visitada el 9 de julio de 2002). La construcción de un algoritmo rudimentario para asignar codones, basándose en una frecuencia determinada, también puede realizarla fácilmente un experto habitual en la técnica con funciones matemáticas básicas.

Se dispone de diversas opciones para sintetizar regiones codificantes con codones optimizados diseñadas por cualquiera de los métodos descritos anteriormente, utilizando manipulaciones biológicas moleculares convencionales y habituales bien conocidas por los expertos habituales en la técnica. En una estrategia, una serie de pares de oligonucleótidos complementarios, cada uno de 80-90 nucleótidos de longitud y que abarca la longitud de la secuencia deseada, se sintetiza por métodos convencionales. Estos pares de oligonucleótidos se sintetizan de tal manera que, después del emparejamiento, forman fragmentos bicatenarios de 80-90 pares de bases, que contienen

extremos cohesivos, por ejemplo, cada oligonucleótido en el par se sintetiza para extender 3, 4, 5, 6, 7, 8, 9, 10, o más bases más allá de la región que es complementaria al otro oligonucleótido en el par. Los extremos monocatenarios de cada par de oligonucleótidos se diseñan para emparejarse con el extremo monocatenario de otro par de oligonucleótidos. Se permite que los pares de oligonucleótidos se emparejen, y aproximadamente de cinco a seis de estos fragmentos bicatenarios se deja después que se emparejen entre sí mediante los extremos monocatenarios cohesivos y después se ligan entre sí y se clonan en un vector de clonación bacteriano convencional, por ejemplo, un vector TOPO® disponible en Invitrogen Corporation, Carlsbad, CA. Después, la construcción se secuencía mediante métodos convencionales. Diversas de estas construcciones constan de 5 a 6 fragmentos de 80 a 90 fragmentos de pares de bases ligados entre sí, es decir, se preparan fragmentos de aproximadamente 500 pares de bases, de tal manera que toda la secuencia deseada se representa en una serie de construcciones plasmídicas. Después, los insertos de estos plásmidos se cortan con enzimas de restricción apropiadas y se ligan entre sí para formar la construcción final. La construcción final se clona después en un vector de clonación bacteriano convencional, y se secuencía. Para el experto habitual en la técnica pueden ser inmediatamente obvios métodos adicionales. Además, en el comercio se dispone fácilmente de síntesis de genes.

En determinadas realizaciones, los codones de una secuencia polipeptídica completa, o de un fragmento, o variante, o sus derivados, se optimizan mediante cualquiera de los métodos descritos en el presente documento. Diversos fragmentos, variantes o derivados deseados, se diseñan y los codones de cada uno de ellos se optimizan individualmente. Además, las regiones codificantes con codones parcialmente optimizados de la presente invención pueden diseñarse y construirse. Por ejemplo, la invención incluye un fragmento de ácido nucleico de una región codificante con codones optimizados que codifica un polipéptido, en la que al menos aproximadamente el 1 %, 2 %, 3 %, 4 %, 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, o 100 % de las posiciones de codones, tiene codones optimizados para una especie determinada. Es decir, contiene un codón que se usa preferentemente en los genes de una especie deseada, por ejemplo, una especie de levadura tal como *Saccharomyces cerevisiae* o *Kluyveromyces*, en lugar de un codón que normalmente se utiliza en la secuencia nativa de ácido nucleico.

En realizaciones adicionales, los codones de una secuencia polipeptídica de longitud completa se optimizan para una especie determinada, dando como resultado una región codificante con codones optimizados que codifica todo el polipéptido, y después los fragmentos de ácido nucleico de la región codificante con codones optimizados, que codifica fragmentos, variantes y derivados del polipéptido, se preparan a partir de la región codificante original con codones optimizados. Como entenderán los expertos habituales en la técnica, si los codones se han asignado al azar en la región codificante de longitud completa basándose en su frecuencia de uso en una especie determinada, los fragmentos de ácido nucleico que codifican fragmentos, variantes y derivados no estarán necesariamente completamente optimizados con codones para la especie determinada. Sin embargo, dichas secuencias están mucho más próximas al uso de codones de la especie deseada que al uso de codones nativo. La ventaja de esta estrategia es que sintetizando fragmentos de ácido nucleico con codones optimizados que codifican cada fragmento, variante y derivado de un polipéptido determinado, aunque rutinaria, llevaría mucho tiempo y daría como resultado un gasto significativo.

Las regiones codificantes con codones optimizados puede ser, por ejemplo, versiones codificantes de una celobiohidrolasa, endoglucanasa o beta-glucosidasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense* *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o *Arabidopsis thaliana*, o de dominios, fragmentos, variantes, o derivados de las mismas.

La optimización de codones se realiza para una especie particular mediante métodos descritos en el presente documento, por ejemplo, en determinadas realizaciones, los polipéptidos que codifican regiones codificantes con codones optimizados de celulasas, o dominios, fragmentos, variantes o derivados de los mismos, de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense* *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o *Arabidopsis thaliana*, están optimizados de acuerdo con el uso de codones de levaduras, por ejemplo, *Saccharomyces cerevisiae*, *Kluyveromyces lactis* y/o *Kluyveromyces marxianus*. También se proporcionan polinucleótidos, vectores y otras construcciones de expresión que comprenden regiones codificantes con codones optimizados que codifican polipéptidos de celulasas o dominios, fragmentos, variantes o derivados de los mismos de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense* *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*,

Chaetomium globosum, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, o *Arabidopsis thaliana*, y diversos métodos de uso de dichos polinucleótidos, vectores y otras construcciones de expresión.

5 En determinadas realizaciones descritas en el presente documento, una región codificante con codones optimizados que codifica cualquiera de las SEQ ID NO: 21-40, 46, o 52-56 o un dominio, fragmento, variante o derivado de las mismas, está optimizada de acuerdo con el uso de codones en levaduras (*Saccharomyces cerevisiae*, *Kluyveromyces lactis* o *Kluyveromyces marxianus*). En algunas realizaciones, los codones de las secuencias están optimizados específicamente para la expresión en *Saccharomyces cerevisiae*. En algunas realizaciones, los codones de las secuencias están optimizados para la expresión en *Kluyveromyces*. En algunas realizaciones, los codones de una secuencia están simultáneamente optimizados para una expresión óptima tanto en *Saccharomyces cerevisiae* como en *Kluyveromyces*. Como alternativa, una región codificante con codones optimizados que codifica cualquiera de las SEQ ID NO: 21-40, 46, o 52-56 puede optimizarse de acuerdo con el uso de codones en cualquier especie vegetal, animal o microbiana.

15 Vectores y métodos del uso de vectores en células hospedadoras

La presente invención también se refiere a vectores que incluyen polinucleótidos de la presente invención, a células hospedadoras que están modificadas genéticamente con vectores de la invención y a la producción de polipéptidos de la invención mediante técnicas recombinantes.

Las células hospedadoras están modificadas genéticamente (transducidas o transformadas o transfectadas) con los vectores de la presente invención que pueden ser, por ejemplo, un vector de clonación o un vector de expresión. El vector puede estar, por ejemplo, en forma de un plásmido, una partícula vírica, un fago, etc. Las células hospedadoras modificadas por ingeniería genética pueden cultivarse en medios con nutrientes convencionales modificados según convenga para la activación de promotores, selección de transformantes o amplificación de los genes de la presente invención. Las condiciones del cultivo, tales como la temperatura, el pH y similares, son las previamente utilizadas con la célula hospedadora seleccionada para la expresión, y serán obvias para el experto habitual en la técnica.

Los polinucleótidos de la presente invención pueden emplearse para la producción de polipéptidos mediante técnicas recombinantes. Por tanto, por ejemplo, el polinucleótido puede incluirse en una cualquiera de una diversidad de vectores de expresión para expresar un polipéptido. Dichos vectores incluyen secuencias de ADN cromosómico, no cromosómico y sintético, por ejemplo, derivados de SV40; plásmidos bacterianos; y plásmidos de levaduras. Sin embargo, puede utilizarse cualquier otro vector siempre que este sea capaz de replicarse y sea viable en la célula hospedadora.

La secuencia de ADN apropiada puede insertarse en el vector mediante diversos procedimientos. En general, la secuencia de ADN se inserta en uno o más sitios de endoglucanasas de restricción apropiados por procedimientos conocidos en la técnica. Se considera que dichos procedimientos y otros están dentro del ámbito del experto en la técnica.

La secuencia de ADN en el vector de expresión está asociada operativamente a una o más secuencias de control de expresión apropiadas (promotor) para dirigir la síntesis de ARNm. Son ejemplos representativos de dichos promotores los siguientes:

Gen	Organismo	Nombre sistemático	Razón de uso/beneficio
PGK1	<i>S. cerevisiae</i>	YCR012W	Fuerte promotor constitutivo
ENO1	<i>S. cerevisiae</i>	YGR254W	Fuerte promotor constitutivo
TDH3	<i>S. cerevisiae</i>	YGR192C	Fuerte promotor constitutivo
TDH2	<i>S. cerevisiae</i>	YJR009C	Fuerte promotor constitutivo
TDH1	<i>S. cerevisiae</i>	YJL052W	Fuerte promotor constitutivo
ENO2	<i>S. cerevisiae</i>	YHR174W	Fuerte promotor constitutivo
GPM1	<i>S. cerevisiae</i>	YKL152C	Fuerte promotor constitutivo
TPI1	<i>S. cerevisiae</i>	YDR050C	Fuerte promotor constitutivo

Adicionalmente, las secuencias promotoras de genes de respuesta al estrés y a la inanición, son útiles en la presente invención. En algunas realizaciones, pueden utilizarse las regiones promotoras de los genes de *S. cerevisiae* GAC1, GET3, GLC7, GSH1, GSH2, HSF1, HSP12, LCB5, LRE1, LSP1, NBP2, PIL1, PIM1, SGT2, SLG1,

WHI2, WSC2, WSC3, WSC4, YAP1, YDC1, HSP104, HSP26, ENA1, MSN2, MSN4, SIP2, SIP4, SIP5, DPL1, IRS4, KOG1, PEP4, HAP4, PRB1, TAX4, ZPR1, ATG1, ATG2, ATG10, ATG11, ATG12, ATG3, ATG14, ATG15, ATG16, ATG17, ATG18, y ATG19. Puede utilizarse cualquier promotor adecuado para dirigir la expresión génica de las células hospedadoras de la invención. Adicionalmente, puede utilizarse el promotor lac o trp de *E. coli*, así como otros promotores que se sabe que controlan la expresión de genes en células procariotas o eucariotas interiores.

Además, para proporcionar un rasgo fenotípico para la selección de células hospedadoras transformadas, los vectores de expresión pueden contener uno o más genes marcadores de selección, tales como *URA3, HIS3, LEU2, TRP1, LYS2* o *ADE2*, dihidrofolato, resistencia a neomicina (G418) o resistencia a zeocina, para cultivo de células eucariotas, o resistencia a tetraciclina o ampicilina en *E. coli*.

El vector de expresión también puede contener un sitio de unión a ribosoma para el inicio de la traducción y/o un terminador de la transcripción. El vector también puede incluir secuencias apropiadas para amplificar la expresión, o puede incluir regiones reguladoras adicionales.

El vector que contiene la secuencia de ADN apropiada como se indica en el presente documento, así como un promotor o una secuencia de control apropiada, pueden emplearse para transformar un hospedador apropiado para permitir que el hospedador exprese la proteína.

Por tanto, en determinados aspectos, la presente invención se refiere a células hospedadoras que contienen las construcciones descritas anteriormente. La célula hospedadora puede ser una célula hospedadora como se describe en cualquier parte en la solicitud. La célula hospedadora puede ser, por ejemplo, una célula eucariota inferior, tal como una célula de levadura, por ejemplo, *Saccharomyces cerevisiae* o *Kluyveromyces*, o la célula hospedadora puede ser una célula procariota, tal como una célula bacteriana.

Como ejemplos representativos de hospedadores apropiados, pueden mencionarse: células bacterianas, tales como *E. coli, Streptomyces, Salmonella typhimurium*; bacterias termófilas o mesófilas; células fúngicas, tales como levaduras; y células vegetales, etc. Se considera que a partir de las enseñanzas del presente documento, la selección de un hospedador apropiado está dentro del ámbito de los expertos en la materia.

Los hospedadores fúngicos apropiados incluyen levaduras. En determinados aspectos de la invención la levadura se selecciona del grupo que consiste en *Saccharomyces cerevisiae, Kluyveromyces lactis, Schizosaccharomyces pombe, Candida albicans, Pichia pastoris, Pichia stipitis, Yarrowia lipolytica, Hansenula polymorpha, Phaffia rhodozyma, Candida utilis, Arxula adenivorans, Debaryomyces hansenii, Debaryomyces polymorphus, Schwanniomyces occidentalis, Issatchenkia orientalis, Kluyveromyces marxianus, Blakeslea, Candida, Cryptococcus, Cunninghamella, Lipomyces, Mortierella, Mucor, Phycocetes, Pythium, Rhodosporidium, Rhodotorula, Trichosporon y Yarrowia.*

Métodos de uso de células hospedadoras para producir etanol

La presente invención también se refiere al uso de células hospedadoras y a cocultivos para producir etanol a partir de sustratos celulósicos. Dichos métodos pueden realizarse, por ejemplo, poniendo en contacto un sustrato celulósico con una célula hospedadora o con un cocultivo de la presente invención.

De acuerdo con la presente invención pueden utilizarse numerosos sustratos celulósicos. Para los ensayos de actividad de celulosa los sustratos pueden dividirse en dos categorías, sustratos solubles e insolubles, basándose en su solubilidad en agua. Los sustratos solubles incluyen celodextrinas o derivados, carboximetil celulosa (CMC) o hidroximetil celulosa (HEC). Los sustratos insolubles incluyen celulosa cristalina, celulosa microcristalina (Avicel), celulosa amorfa, tal como celulosa hinchada con ácido fosfórico (PASC), celulosa teñida o fluorescente y biomasa lignocelulósica pretratada. Generalmente estos sustratos son material celulósico altamente organizado y son por tanto solo moderadamente solubles.

Se apreciará que un material lignocelulósico apropiado puede ser cualquier materia prima que contenga celulosa soluble y/o insoluble, pudiendo estar la celulosa insoluble en forma cristalina o no cristalina. En diversas realizaciones, la biomasa lignocelulósica comprende, por ejemplo, madera, maíz, forraje de maíz, serrín, corteza, hojas, restos agrícolas y forestales, hierba, tal como pasto varilla, productos de la digestión de rumiantes, basura urbana, aguas residuales de papeleras, periódicos, cartón o combinaciones de los mismos.

En algunas realizaciones, la invención se refiere a un método para hidrolizar un sustrato celulósico, por ejemplo, un sustrato celulósico como se ha descrito anteriormente, poniendo en contacto el sustrato celulósico con una célula hospedadora de la invención. En algunas realizaciones, la invención se refiere a un método para hidrolizar un sustrato celulósico, por ejemplo un sustrato celulósico como se ha descrito anteriormente, poniendo en contacto el sustrato celulósico con un cocultivo que comprende células de levadura que expresan celulasas heterólogas.

En algunas realizaciones, la invención se refiere a un método para fermentar celulosa. Dichos métodos pueden realizarse, por ejemplo, cultivando una célula hospedadora o cocultivo en un medio que contenga celulosa insoluble

para permitir la sacarificación y fermentación de la celulosa.

La producción de etanol puede realizarse, de acuerdo con la presente invención, a temperaturas de al menos aproximadamente 30 °C, aproximadamente 31 °C, aproximadamente 32 °C, aproximadamente 33 °C, aproximadamente 34 °C, aproximadamente 35 °C, aproximadamente 36 °C, aproximadamente 37 °C, aproximadamente 38 °C, aproximadamente 39 °C, aproximadamente 40 °C, aproximadamente 41 °C, aproximadamente 42 °C, aproximadamente 43 °C, aproximadamente 44 °C, aproximadamente 45 °C, aproximadamente 46 °C, aproximadamente 47 °C, aproximadamente 48 °C, aproximadamente 49 °C o aproximadamente 50 °C. En algunas realizaciones de la presente invención, la célula hospedadora termotolerante puede producir etanol a partir de celulosa a temperaturas por encima de aproximadamente 30 °C, aproximadamente 31 °C, aproximadamente 32 °C, aproximadamente 33 °C, aproximadamente 34 °C, aproximadamente 35 °C, aproximadamente 36 °C, aproximadamente 37 °C, aproximadamente 38 °C, aproximadamente 39 °C, aproximadamente 40 °C, aproximadamente 41 °C, aproximadamente 42 °C, o aproximadamente 43 °C, o aproximadamente 44 °C o aproximadamente 45 °C, o aproximadamente 50 °C. En algunas realizaciones de la presente invención, la célula hospedadora termotolerante puede producir etanol a partir de celulosa a temperaturas de aproximadamente 30 °C a 60 °C, de aproximadamente 30 °C a 55 °C, de aproximadamente 30 °C a 50 °C, de aproximadamente 40 °C a 60 °C, de aproximadamente 40 °C a 55 °C o de aproximadamente 40 °C a 50 °C.

En algunas realizaciones, los métodos de producción de etanol pueden comprender poner en contacto un sustrato celulósico con una célula hospedadora o cocultivo de la invención y adicionalmente poner en contacto el sustrato celulósico con enzimas celulasas producidas externamente. En el comercio están disponibles ejemplos de enzimas celulasas producidas externamente y los expertos en la técnica las conocen.

Por lo tanto, la invención también se refiere a métodos de reducción de la cantidad de enzimas celulasas producidas externamente necesarias para producir una cantidad determinada de etanol a partir de celulosa que comprende poner en contacto la celulosa con celulasas producidas externamente y con una célula hospedadora o cocultivo de la invención. En algunas realizaciones, utilizando al menos aproximadamente un 5 %, 10 %, 15 %, 20 %, 25 %, 30 % o 50 % menos de celulasas producidas externamente, puede obtenerse la misma cantidad de producción de etanol. En algunas realizaciones, no se añade celulosa externa, o menos de aproximadamente un 5 % de la celulosa es celulosa añadida externamente, o menos de aproximadamente un 10 % de la celulosa es celulosa añadida externamente, o menos de aproximadamente un 15 % de la celulosa es celulosa añadida externamente.

En algunas realizaciones, los métodos comprenden la producción de etanol a una tasa particular. Por ejemplo, en algunas realizaciones, el etanol se produce a una tasa de al menos aproximadamente 0,1 mg por hora por litro, al menos aproximadamente 0,25 mg por hora por litro, al menos aproximadamente 0,5 mg por hora por litro, al menos aproximadamente 0,75 mg por hora por litro, al menos aproximadamente 1,0 mg por hora por litro, al menos aproximadamente 2,0 mg por hora por litro, al menos aproximadamente 5,0 mg por hora por litro, al menos aproximadamente 10 mg por hora por litro, al menos aproximadamente 15 mg por hora por litro, al menos aproximadamente 20 mg por hora por litro, por lo menos aproximadamente 30 mg por hora por litro, por lo menos aproximadamente 30 mg por hora por litro, por lo menos aproximadamente 50 mg por hora por litro, al menos aproximadamente 100 mg por hora por litro, al menos aproximadamente 300 mg por hora por litro, al menos aproximadamente 400 mg por hora por litro, o al menos aproximadamente 500 mg por hora por litro.

En algunas realizaciones, las células hospedadoras de la presente invención pueden producir etanol a una tasa de al menos aproximadamente 0,1 mg por hora por litro, al menos aproximadamente 0,25 mg por hora por litro, al menos aproximadamente 0,5 mg por hora por litro, al menos aproximadamente 0,75 mg por hora por litro, al menos aproximadamente 1,0 mg por hora por litro, al menos aproximadamente 2,0 mg por hora por litro, al menos aproximadamente 5,0 mg por hora por litro, al menos aproximadamente 10 mg por hora por litro, al menos aproximadamente 15 mg por hora por litro, al menos aproximadamente 20,0 mg por hora por litro, al menos aproximadamente 25 mg por hora por litro, al menos aproximadamente 30 mg por hora por litro, al menos aproximadamente 50 mg por hora por litro, al menos aproximadamente 100 mg por hora por litro, al menos aproximadamente 200 mg por hora por litro, al menos aproximadamente 300 mg por hora por litro, al menos aproximadamente 400 mg por hora por litro, o al menos aproximadamente 500 mg por hora por litro más que una cepa de control (que carece de celulasas heterólogas), y cultivada en las mismas condiciones. En algunas realizaciones, el etanol puede producirse en ausencia de cualquier celulosa añadida externamente.

La producción de etanol puede medirse utilizando cualquier método conocido en la técnica. Por ejemplo, la cantidad de etanol en muestras de fermentación puede evaluarse utilizando análisis HPLC. En el comercio se dispone de muchos kits de ensayo de etanol que utilizan, por ejemplo, ensayos basados en la enzima alcohol oxidasa. A partir de las enseñanzas del presente documento, los métodos para determinar la producción de etanol están al alcance de los expertos en la técnica.

Las siguientes realizaciones de la invención se describirán ahora con más detalle a través de estos ejemplos no limitantes.

Ejemplos

La presente invención presenta diversas etapas directas importantes para crear una levadura capaz de realizar un bioprocesamiento consolidado. La invención describe levaduras celulolíticas mejoradas creadas expresando combinaciones de celulasas heterólogas. La presente invención demuestra por primera vez, la capacidad de *Kluyveromyces* transformada para producir etanol a partir de celulosa, la capacidad de cepas de levadura para expresar solamente celulasas heterólogas secretadas para producir etanol a partir de celulosa, y la capacidad de cocultivos de cepas de levadura múltiples que expresan diferentes celulasas para producir etanol a partir de celulosa. Además dichas cepas de levaduras y cocultivos de cepas de levaduras pueden aumentar la eficacia de los procesos de sacarificación y fermentación simultáneas (SFS).

Protocolos generales

Cultivos de cepas y medios generales

La cepa de *Escherichia coli* DH5 α (Invitrogen), o NEB 5 alfa (New England Biolabs) se usó para la transformación y propagación de plásmido. Las células se cultivaron en medio LB (extracto de levadura 5 g/l, NaCl 5 g/l, triptona 10 g/l) complementado con ampicilina (100 mg/l), kanamicina (50 mg/l) o zeocina (200 mg/l). Cuando se deseaba la selección con zeocina, el medio LB se ajustaba a un pH de 7,0. Además, cuando se deseaban medios sólidos, se añadían 15 g/l de agar.

Para la selección, las cepas de levadura se cultivaron habitualmente en medios YPD (extracto de levadura 10 g/l, peptona 20 g/l, glucosa 20 g/l), YPC (extracto de levadura 10 g/l, peptona 20 g/l, celobiosa 20 g/l) o YNB + glucosa (base de nitrógeno para levadura 6,7 g/l sin aminoácidos, y complementado con aminoácidos apropiados para la cepa, glucosa 20 g/l) con G418 (250 mg/l a menos que se especifique) o zeocina (20 mg/l a menos que se especifique). Para medios sólidos, se añadieron 15 g/l de agar.

Métodos moleculares

Después de los protocolos convencionales se realizaron manipulaciones de ADN (Sambrook *et al.*, 1989). La PCR se realizó utilizando polimerasa Phusion (New England Biolabs) para la clonación, y Taq polimerasa (New England Biolabs) para la exploración de transformantes, y en algunos casos polimerasa Advantage (Clontech) para la PCR de genes para corregir auxotrofías. Se siguieron las pautas proporcionadas por los fabricantes. Las enzimas de restricción se adquirieron en New England Biolabs y se prepararon digeridos de acuerdo con las directrices proporcionadas. Los ligamientos se realizaron utilizando el kit de ligamiento Quick (New England Biolabs) según lo especificado por el fabricante. La purificación en gel se realizó utilizando kits de investigación de Qiagen o Zymo, las purificaciones de los productos de la PCR y de los digeridos se realizaron utilizando kits de investigación Zymo y para la purificación de ADN plasmídico se utilizaron kits de Qiagen midi y miniprep. La secuenciación se realizó en la Molecular Biology Core Facility en Dartmouth College. Se utilizó ligamiento mediado por levaduras (YML, por las siglas en inglés *Yeast Mediated Ligation*) para crear algunas construcciones (Ma *et al.*, Gene 58: 201-216 (1987)). Esto se realizó creando fragmentos de ADN a clonar con 20-40 pb de homología con las otras piezas a combinar y/o con el vector estructural. Un vector estructural (pRS426), capaz de replicarse en levaduras, y con el gen Ura3 para selección, se transformó después en levadura mediante métodos convencionales con las secuencias diana para clonación. La levadura transformada recombina estos fragmentos para formar una construcción completa y el plásmido resultante permite la selección en medios sin uracilo.

Vectores

En la Tabla 4 se resumen vectores de construcciones plasmídicas en los experimentos detallados más adelante y en la Tabla 5 se muestran los cebadores utilizados en la construcción de los vectores.

Tabla 4. Plásmidos utilizados

Plásmido	Genotipo
pBKD1-BGLI	<i>bla KanMX PGK1_{P-S.f.} bglI - PGK1_T</i>
pBKD2-sEGI	<i>bla KanMX ENO1_{P-sT.r.} eg1- ENO1_T</i>
pBKD1-BGLI-sEGI	<i>bla KanMX ENO1_{P-sTr.} eg1- ENO1_T & PGK1_{P-S.f.} bglI- PGK1_T</i>
YE _p ENO-BBH	<i>bla URA3 ENO1_{PT}</i>
pJC1 La Grange <i>et al.</i> (1996)	<i>bla URA3 PGK_{PT}</i>
pRDH101	<i>bla URA3 ENO1_{P-sT.r.} cbh1- ENO1_T</i>
pRDH103	<i>bla URA3 ENO1_{P-sH.g.} cbh1- ENO1_T</i>

ES 2 621 181 T3

Plásmido	Genotipo
pRDH104	<i>bla URA3 ENO1_{P-sTa}.cbh1-ENO1_T</i>
pRDH105	<i>bla URA3 ENO1_{P-sT.e}.cbh1-ENO1_T</i>
pRDH106	<i>bla URA3 ENO1_{P-sT.e}.cbh2- ENO1_T</i>
pRDH107	<i>bla URA3 PGK1_{P-sT.r}.cbh2- PGK1_T</i>
pRDH108	<i>bla URA3 PGK1_{P-sT.r}.cbh2- PGK1_T & ENO1_{P-sT.e}.cbh1- ENO1_T</i>
pRDH118	<i>bla URA3 PGK1_{P-sT.r}.cbh2- PGK1_T & ENO1_{P-sH.g}.cbh1-ENO1_T</i>
pRDH120	<i>bla URA3 PGK1_{P-sT.r}.cbh2- PGK1_T & ENO1_{P-sTa}.cbh1- ENO1_{mT}</i>
pDF1 La Grange <i>et al.</i> (1996)	<i>bla furl::LEU2</i>
pCEL5 Den Haan <i>et al.</i> 2007	vector de 2 µm (micrómetros) para la expresión de BGLI S.f. y de EGI T.r. (secuencia nativa)
pMU185	pUG66 (loxp-zeo-loxp)
pKLAC1 New England Biolabs	Vector de expresión de <i>K. lactis</i> para la integración en el locus lac4, selección con acetamida
pRS426	Vector de 2 µm para ligamiento mediado por levadura (YML)
pMU289	pRS426 con parte de pKLAC1 para la inserción de EG1 T.r. (a partir de pBKD_11621, como se detalla en el ejemplo 1) en el locus lac4 creado por YML
pMU291	pRS426 con parte de pKLAC1 para la inserción de CBH2Tr (a partir de pBZD_20641, como se detalla en el ejemplo 1) en el locus lac4 creado por YML
pMU398	<i>ENO1_{P-sTe}.cbh1-ENO1_T</i> a partir de pRDH105 en pMU289 (clonación por YML)
pMU451	pRDH105 con enlazador PacI/Ascl (formado utilizando cebadores) insertado en EcoRI/XhoI
pMU458	construcción sintética para EG N.f. insertada en su interior
	pMU451 (PacI/Ascl digerido de ambas partes)
pMU463	TrEG1 a partir de pBKD1-BGLI-sEGI en pMU451 (PacI/Ascl digerido de ambas partes)
pMU465	construcción sintética para EG C.I.(a) insertada en pMU451 (PacI/Ascl digerido de ambas partes)
pMU469	construcción sintética para EG R.f. insertada en pMU451 (PacI/Ascl digerido de ambas partes)
pict471	construcción sintética para EG C.f. insertada en pMU451 (PacI/Ascl digerido de ambas partes)
pMU472	construcción sintética para EG N.t. insertada en pMU451 (PacI/Ascl digerido de ambas partes)
pMU473	construcción sintética para EG C.a. insertada en eMu451 (PacI/Ascl digerido de ambas partes)
pMU475	construcción sintética para CBH2 T.r. derivada de pBKD 20641 con contacto retirado (del ejemplo 1) insertada en pMU451 (PacI/Ascl digerido de ambas partes)
pMU499	construcción sintética para EG M.d. insertada en pMU451 (PacI/Ascl digerido de ambas partes)
pMU500	construcción sintética para EG R.s insertada en pMU451 (PacI/Ascl digerido de ambas partes)
pMU503	construcción sintética para EG N.w. insertada en pMU451 (PacI/Ascl digerido de ambas partes)

ES 2 621 181 T3

Plásmido	Genotipo
pMU624/pMI529	vector de 2 µm para la expresión de CBH1 T.e. con CBD (fragmentos de PCR para enzima quimérica con pRDH105 digerido con <i>PmlI-XhoI</i>)
pMU326	construcción sintética para EG R.s. de Codon Devices
pMU784/pMI574	vector de 2 µm para la expresión de CBH2 C.I.(b) (construcción sintética para) CBH2 C.1.(b) insertada en pMU624 digerido con <i>PacI/Ascl</i>)
pMU562	pBKD_2 con <i>loxp-zeo-loxp</i> insertado (<i>NotI</i> digerido de ambas partes)
pMU576	ENO1p- T.r.cbh1-ENO1 _T (a partir de pMU291) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU577	ENO1p- T.e.cbh1-ENO1 _T en (a partir de pMU398) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
emu661	ENO1p- T.r. EG1-ENO1 _T (a partir de pMU463) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU662	ENO1p- C.1.(a) EG1-ENO1 _T (a partir de pMU465) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU663	ENO1p- C.f. EG1-ENO1 _T (a partir de pMU471) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU664	ENO1p- N.t. EG1-ENO1 _T (a partir de pMU472) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU665	ENO1p- C.a. EGI -ENO 1 (a partir de pMU473) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU666	ENO1p- T.r.CBH2-ENO1 _T (a partir de pMU475) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU667	ENO1p- M.d.-EG1-ENO1 _T (a partir de pMU499) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU668	ENO1p- N.w.-EG1-ENO1 _T (a partir de pMU503) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU755	ENO1p- T.e.CBHI con/CBD-ENO1 _T (a partir de pMU624) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU750	ENO1p- R.s.-EG2-ENO1 _T (a partir de pMU326) en pMU562 (<i>PacI/Ascl</i> digerido de ambas piezas)
emu809	ENO1p- C.1.(b) CBH2b-ENO1 _T (a partir de pMU784) en pMU562 (<i>PacI/Ascl</i> digerido de ambas partes)
pMU721	pMU562 con el gen <i>hph</i> (marcador de resistencia a higromicina) reemplazando el marcador de zeocina (<i>NotI</i> digerido de ambos fragmentos)
pMU760	ENO 1 p- T.e.CBHI 1 con CBD-ENO 1 a partir de pMU624 en pMU721 (<i>MheI/Ascl</i> digerido de ambos fragmentos)
pMU761	ENO1p- T.r.CBH2-ENO1 _T a partir de pMU291 en pMU721 (<i>PacI/Ascl</i> digerido de ambos fragmentos)
pMI553	vector de 2 µm para la expresión de CBH2 T.r. y de CBH1 T.e.+MUC
pMI568	vector de 2 µm para la expresión de EG1 T.r., por favor véase el texto de la descripción de cómo se construyó esta construcción.
pMI574	vector de 2 µm para la expresión de CBH2 C.I.(b)
pMI577	vector de 2 µm para la expresión de CBH2 T.r y de CBH1 H.g.

Plásmido	Genotipo
pMI578	vector de 2 µm para la expresión de CBH2 T.r. y de CBH1 T.e
pMI579	vector de 2 µm para la expresión de CBH2 T.r. y de CBH1 C.l.(b)
pMI580	vector de 2 µm para la expresión de CBH2 C.l.(b) y de CBH1 T.e. +MUC
pMI581	vector de 2 µm para la expresión de CBH2 C.l.(b) y de CBH1 T.e.
pMI582	vector de 2 µm para la expresión de CBH2 C.l.(b) y de CBH1 H.g.
pMI583	vector de 2 µm para la expresión de CBH2 C.l.(b) y de CBH1 C.t.

Abreviaturas: *ENO1_{P/T}* = promotor/terminador del gen de enolasa 1; *PGK1_{P/T}* = promotor y terminador del gen de la fosfoglicerato quinasa 1; *T.r.* = *Trichoderma reesei*; *H.g.* = *Humicola grisea*; *T.a.* = *Thermoascus aurantiacus*; *T.e.* = *Talaromyces emersonii*; *Sf.* = *Saccharomycopsis fibuligera*; *C.l. (a)* = *Coptotermes lacteus*; *C.f.* = *Coptotermes formosa-nus*; *N.t.* = *Nasutitermes takasagoensis*; *C.a.* = *Coptotermes acinaciformis*; *M.d.* = *Mastotermes darwinensis*; *N.w.* = *Nasutitermes walkeri*; *R.s.* = *Reticulitermes speratus*; *C.l. (b)* = *Chrysosporium lucknowense*; *N.f.* = *Neosartorya fischeri*; *R.f.* = *Reticulitermes flavipes*; *C.t.* = *Chaetomium thermophilum*

Tabla 5: cebadores utilizados

sCBHI/2-L	GACTGAATTCATAATGGTCTCCTCACCTCC
sCBHI-R	GACTCTCGAGTTACAAACATTGAGAGTAGTATGG
sCBH2-R	CAGTCTCGAGTTACAAGAAAGATGGGTTAGC
395 Te cbh1 Sint1 PacI-ATG	GCGTTGGTACCGTTTAAACGGGGCCCTTAATTAACAAT GCTAAGAAGAGCTTACTATTGAG
398 Te cbh1 núcleo sint Smal	CCTCCCCCGGGTTAGAAGCAGTGAAAGTGGAGTTGATTG G
399Trcbh1 sint MUC5 MlyIHincII	GCGACGAGTCAACCC TCCAGGTGGTAACAG AGGTACTACCAC
400 Trcbh1 sint MUC AsclXhoI	GCGACTCGAGGGCGCGCCTACAAACATTGAGAGTAGTA TGGGTTTA
379 ScPGK1prom -786 SacI+ApaI	GCGTTGAGCTCGGGCCCTAATTTTATTTAGATTCCTGA CTTCAAC
380 ScPGK1prom EcoRI-PacI	GCGTTGAATTCTTAATTAAGTAAAAAGTAGATAATTACT TCCTTG
381 CBH2 WT EcoRI-PacI- ATG	GCGTTGAATTCTTAATTAACAATGATTGTCGGCATTCT CACCACGC
386 CBH2 WT TAA-Ascl- EcoRI	gcgatgaattcggcggccTTACAGGAACGATGGGTTTGCGTTTG

El vector de expresión de levadura YEpENO-BBH se creó para facilitar la expresión heteróloga bajo el control del promotor y terminador del gen de enolasa 1 (*ENO1*) de *S. cerevisiae*. El vector también fue útil porque el casete de expresión de este vector pudo escindirarse simplemente utilizando una digestión con BamHI, BglII. YEpENO1 (Den Haan *et al.*, *Metabolic Engineering*, 9: 87- 942007) contiene la estructura YEp352 con las secuencias promotoras y terminadoras del gen de ENO1 clonadas en los sitios BamHI y HindIII. Este plásmido se digirió con BamHI y el saliente se cargó con polimerasa Klenow y con los dNTP para retirar el sitio BamHI. El plásmido volvió a ligarse para generar YEpENO-B. Utilizando el mismo método, los sitios BglII y después HindIII se destruyeron posteriormente para crear el molde YEpENO-BBH. El molde YEpENO-BB se utilizó como molde para una reacción de PCR con los cebadores ENOB-izquierda (5'-GATCGGATCCCAATTAATGT-GAGTTACCTCA-3') y ENOB-derecha (5'-GTACAAGCTTAGATCTCCTATGCGGTGTTAGATCTCCTATGCGGTGTGAAATA-3') en el que el casete ENO1 se amplificó junto con una región flanqueante de 150 pb cadena arriba y de 220 pb cadena abajo. Este producto se digirió con BamHI y HindIII y los salientes se cargaron por tratamiento con polimerasa Klenow y con los dNTP y se clonó entre los dos sitios PvuII en yENO1 reemplazando de un modo eficaz el casete ENO1 original y generando YEpENO-BBH.

Se diseñaron las versiones con codones optimizados de la *cbh1* de *Humicola grisea* (*cbh1Hg*), de la *cbh1* de *Thermoascus aurantiacus* (*cbh1Ta*) y de las *cbh1* y *cbh2* de *Talaromyces emersonii* (*cbh1Te* y *cbh2Te*) y se encargaron genes sintetizados en la GenScript Corporation (Piscataway, NJ, USA). Estos cuatro genes codificantes de *cbh* sintéticos recibidos de la GenScript Corporation, se clonaron en el plásmido pUC57. Los vectores resultantes se digirieron con EcoRI y XhoI para escindir los genes *cbh* que se clonaron posteriormente en un vector YE_pENO-BBH digerido con EcoRI y XhoI. Esto creó los plásmidos pRDH103 (con *cbh1Hg*), pRDH104 (con *cbh1Ta*), pRDH105 (con *cbh1Te*) y pRDH106 (con *cbh2Te*) con los genes codificantes de *cbh* bajo el control transcripcional del promotor y terminador de ENO1. Adicionalmente, se creó pRDH101 que expresaba la CBH1 de *T. reesei* de pBZD_10631_20641. Se utilizó la enzima ExTaq de Takara como se indica y para amplificar la *cbh1Tr* de pBZD_10631_20641 utilizando los cebadores sCBH1/2L y sCBH1R. Después, el fragmento se aisló y se digirió con EcoRI y XhoI. YE_pENO-BBH también se digirió con EcoRI y XhoI y las bandas relevantes se aislaron y ligaron. Se amplificó un fragmento de 1494 pb que codificaba el gen *cbh2* de *T. reesei* del plásmido pBZD_10631_20641, con los cebadores sCBH1/2-L y sCBH2R (5'-CAGTCTCGAGTTACAAGAAAGATGGGTTAGC-3'), se digirió con EcoRI y XhoI y se clonó en los sitios EcoRI y XhoI de pJC1 (Crouse *et al.*, Curr. Gen. 28: 467-473 (1995)) colocándolo bajo el control transcripcional del promotor y terminador del gen de la fosfoglicerato quinasa 1 (PGK1) de *S. cerevisiae*. Este plásmido se denominó pRDH107. Posteriormente, los casetes de expresión de pRDH103, pRDH104 y pRDH105 se escindieron por digestión con BamHI y BglII y se clonaron en el sitio BamHI de pRDH107 para dar pRDH118, pRDH120, pRDH108 y pRDH109, respectivamente. PRDH109 contiene los mismos casetes de expresión que pRDH108 pero en pRDH108 los casetes de expresión génicos están en orientación inversa entre sí. Estos plásmidos y sus genotipos básicos se resumen en la Tabla 4.

También se crearon dos vectores adicionales de 2 micrómetros para la expresión de la CBH2b de *Chrysosporium lucknowense* y de la CBH1 de *T. emersonii* con una fusión c terminal del MUC de la CBH1 de *T. reesei*. La fusión entre *cbh1* de *T. emersonii* y el MUC de *cbh1* de *T. reesei* se generó por ligamiento de tres fragmentos. En la Tabla 5 se indican los oligonucleótidos utilizados para estas construcciones. Se amplificó un producto de la PCR con los oligonucleótidos 395 *Te cbh1 Sint1* Pacl-ATG y 398 *Te cbh1* núcleo *sint* Smal utilizando como molde pRDH105, se digirió con PmlI y Smal y se aisló el fragmento de 800 pb. Se amplificó un segundo producto de la PCR con los oligonucleótidos 399 *Trcbh1 sint* MUC5 M1yIHincII y 400 *Trcbh1 sint* MUC AsclXhoI con pRDH101 como molde, se digirió con MlyI y XhoI y se aisló el fragmento de 180 pb. Los dos fragmentos de la PCR se ligaron con el fragmento PmlI-XhoI de 6,9 kb de pRDH105 produciendo pMU624.

La secuencia de ADN genómico de 3900 pb del gen *cbh2b* de *Chrysosporium lucknowense* (descrito en la Solicitud de Patente de Estados Unidos publicada N.º: 2007/0238155) se analizó para supuestos intrones utilizando el Servidor NetAspGene 1.0 (<http://www.cbs.dtu.dk/Services/NetAspGene/>). La retirada de los intrones citados de la secuencia genómica dio como resultado una fase de lectura abierta de 482 aminoácidos que se sintetizó en Codon Devices y sus codones se optimizaron para la expresión en *S. cerevisiae* y se clonaron en el vector pUC57. El plásmido pAJ401 (Saloheimo *et al.*, Mol. Microbiol., 13: 219-228, 1994), que contenía el promotor y terminador de *PGK1*, se modificó para la expresión de *cbh2* de *T. reesei* entre los sitios de restricción Pacl y Ascl. El promotor de *PGK1* se amplificó con los cebadores 379 *ScPGK1prom* -786 Sacl + Apal y 380 *ScPGK1prom* de EcoRI-Pacl y pAJ410 como molde y se digirieron con Pacl y EcoRI. La ORF de *cbh2* de *T. reesei* se amplificó a partir de pTTC01 (Teeri *et al.*, Gene 51: 43-52, 1987) con los oligonucleótidos 381 CBH2 WT EcoRI-Pacl-ATG y 386 CBH2 WT TAA-Ascl-EcoRI, se digirió con Pacl y EcoRI, y se ligó con el pAJ401 digerido con Sacl-EcoRI dando como resultado pMI508. El fragmento Pacl-Ascl en pMI508 se reemplazó por un gen *egl1* de *T. reesei* de 1,4 kb sintético dando como resultado pMI522. El fragmento de 1,9 kb de pMI522 se digirió con PmlI y XhoI y se ligó con el fragmento PmlI-XhoI de 6,4 kb de pRDH107 dando como resultado pMI568. pMI568 se digirió con Pacl y Ascl y el fragmento de 7 kb se ligó con el fragmento de 1,5 kb de pMI558 produciendo pMU784 para la expresión de *cbh2b* de *C. lucknowense*.

También se construyó un conjunto de vectores de 2 micrómetros para la expresión de endoglucanasas en *S. cerevisiae*, así como plásmidos relacionados para actuar como controles. Se creó pMU451 como un vector de control y para la clonación de las celulasas bajo el control del promotor y terminador de *ENO1*. Esto se realizó añadiendo un enlazador Pacl/Ascl en el sitio EcoRI/XhoI de pMU451. Los genes sintéticos encargados en Codon Devices y recibidos en pUC57 se clonaron en este vector como fragmentos Pacl/Ascl. Los vectores creados de este modo e indicados en la Tabla 4 son: pMU458, pMU463, pMU465, pMU469, pMU471, pMU472, pMU473, pMU475, pMU499, pMU500 y pMU503.

A partir de las construcciones pBKD_1 y pBKD_2, se crearon vectores para la integración de versiones de celulasas secretadas en los sitios de integración delta en *S. cerevisiae*, o para la integración en el genoma de *K. marxianus*. La BGL1 de *S. fibuligera* (BGLISf) se clonó por PCR a partir de ySFI (van Rooyen *et al.*, J. Biotechnol., 120: 284-95 (2005)). La endoglucanasa (EGITr) utilizada tenía la secuencia dada en la Tabla 1. Los genes que codifican celulasas se clonaron mediante PCR (utilizando los sitios Pacl y Ascl) en pBKD_1 y pBKD_2 - para crear pBKD1-BGL1 y pBKD2-sEG1. El casete ENO1P-sEG1-ENO1T de pBKD2-sEG1 se subclonó posteriormente como un fragmento SpeI, NotI con pBKDI-BGL1 para crear pBKD1-BGL1-sEG1.

El pMU562, utilizado para la integración de celulasas en *K. marxianus*, se generó cortando con pMU185 (pUG66) con Not1 y aislando un inserto que contenía lox P ZeoR de 1190 pb. Este inserto se ligó en un vector de integración delta de 4,5 Kb digerido con Not1 para producir pMU562. pMU576 se generó cortando el plásmido pMU291 que

5 contenía la CBH2 de *T. reesei* con Asc1/Pac1, aislando un gen de CBH2 de 1491 pb y ligándolo en el vector de integración delta pMU562 cortado con Asc1/Pac1. pMU577 se generó cortando CBH1 de *T. emersonii* de pMU398 con Asc1/Pac1, aislando un gen CBH1 de 1380 pb y ligándolo en el vector de integración delta pMU562 cortado con Asc1/Pac1. De manera similar, un conjunto de construcciones de celulasas recombinantes (pMU661 a pMU668 y pMU750, pMU755, pMU809-véase Tabla 4), incluyendo una variedad de endoglucanasas y celobiohidrolasas, se incorporó en pMU562 para la cotransformación. Secuencias sintéticas para estos genes de celulasa se obtuvieron originalmente en Codon Devices y posteriormente se clonaron en vectores de expresión de 2 µ para su uso en *S. cerevisiae*. Después se transfirieron desde estos vectores a los vectores de integración como se detalla (incluyendo las digestiones utilizadas) en la Tabla X. En su conjunto estas construcciones formaron una biblioteca que puede transformarse individualmente o conjuntamente y después explorarse mediante un ensayo de actividad. Las construcciones se digirieron con enzimas que cortaban dentro o muy cerca de las secuencias delta para la integración. También se construyeron construcciones similares para celulasas de integración utilizando el marcador higromicina (pMU721, pMU760 y pMU761).

15 Transformación en levaduras

Para la transformación rutinaria de plásmidos completos en *S. cerevisiae*, se utilizó transformación química convencional (Sambrook *et al.*, Molecular cloning: A laboratory manual, Nueva York: Cold Spring Harbor Laboratory Press, 1989). Para algunas transformaciones, se utilizó un protocolo modificado descrito por Hill *et al.* (Nucleic Acids Res. 19: 5791 (1991)).

25 Se desarrolló un protocolo para la electrotransformación de levaduras basado en Cho *et al.* (1999) y en Ausubel *et al.* (1994). Se crearon fragmentos lineales de ADN digiriendo pBD1-BGL1-sEG1 con Accl. Accl tenía un único sitio en la secuencia δ. Los fragmentos se purificaron por precipitación con NaAc 3 M y etanol enfriado en hielo, posteriormente se lavó con etanol al 70 % y se resuspendió en dH2O USB (agua estéril sin ADNasa ni ARNasa) después de secar en un horno al vacío a 70 °C.

30 Para la transformación, se prepararon células de *S. cerevisiae* cultivando hasta la saturación en cultivos YPD de 5 ml. Se muestrearon 4 ml del cultivo, se lavaron 2X con agua destilada fría y se resuspendieron en agua destilada fría 640 µl. Se añadieron 80 µl de Tris-HCl 100 mM, EDTA 10 mM, pH 7,5 (esterilizado con filtro tampón TE 10X) y 80 µl de acetato de litio 1 M, pH 7,5 (esterilizado con filtro 10X LiAc), y la suspensión celular se incubó a 30 °C durante 45 min con agitación suave. Se añadieron 20 µl de DTT 1 M y la incubación continuó durante 15 min. Después, las células se centrifugaron, se lavaron una vez con agua destilada fría y una vez con tampón de electroporación (sorbitol 1 M, HEPES 20 mM) y finalmente se resuspendió en tampón de electroporación 267 µl. Se utilizó el mismo protocolo para la transformación de las cepas de *K. lactis* y *K. marxianus*, excepto que se inocularon 50 ml de YPD con 0,5 ml de un cultivo durante una noche, se cultivó durante 4 horas a 37 °C, y después se centrifugó y se preparó como se ha indicado anteriormente. Adicionalmente, se realizaron etapas de incubación y de recuperación a 37 °C.

40 Para la electroporación, 10 µg de ADN linealizado (medido por estimación en un gel) se combinaron con 50 µl de la suspensión celular en un tubo de microcentrifuga estéril de 1, 5 ml. Después, la mezcla se transfirió a una cubeta de electroporación de 0,2 cm, y a la muestra se le aplicó un pulso de 1,4 kV (200 µ, 25 µF) utilizando el dispositivo Gene Pulser de Biorad. En la cubeta se puso 1 ml de YPD con sorbitol 1 M ajustado a pH 7,0 (YPDS) y dejó que las células se recuperasen durante ~3 h. En placas de agar YPDS con antibiótico apropiado, se diseminaron 100-200 µl de suspensión celular y se incubó a 30 °C durante 3-4 días hasta que aparecieron las colonias.

Cepas de levadura

50 Las cepas de levadura indicadas en la Tabla 6 se crearon utilizando los vectores y los protocolos de transformación descritos.

Tabla 6: Cepas de levadura.

Nombre	Cepa de fondo	Genes expresados y/o inactivados	Construcciones
M0013	<i>Saccharomyces cerevisiae</i> Y294 (ATCC 201160)	Genotipo: α, <i>leu2-3,112 ura3-52 his3 trp1-289</i>	Ninguna
M0243	M0013	BGLISf, EGITr	pBKD1-BGLI-sEGI
M0244	M0013	BGLISf, EGITr (secuencia nativa)	pCEL5
M0247	M0013	TeCBH1; delta FUR1	pRDH105
M0248	M0013	TrCBH2, TeCBH1; delta FUR1	pRDH108; pDF1

ES 2 621 181 T3

Nombre	Cepa de fondo	Genes expresados y/o inactivados	Construcciones
M0249	M0013	Ninguno (control); delta FUR1	pJC1; pDF1
M0265	M0013	CBHIHg; delta FUR1	pRDH103; pDF1
M0266	M0013	CBHITa; delta FUR1	pRDH104; pDF1
M0282	M0248	BGLISf, EGITr, CBH2Tr, CBH1Te; delta FUR1	pBKD1-BGLI-sEGl; pRDH108; pDF1
M0284	M0243	BGLISf, EGITr, CBH2Tr, CBH1Hg; delta FUR1	pBKD1-BGLI-sEGl; pRDH118;pDF1
M0286	M0243	BGLISf, EGITr, CBH2Tr, CBH1Ta; delta FUR1	pBKD1-BGLI-sEGl; pRDH120; pDF1
M0288	M0243	BGLISf, EGITr, CBH2Tr, CBH1Te; delta FUR1	pBKD1-BGLI-sEGl; pRDH108; pDF1
M0289	M0013	CBH2Tr, CBH1Hg; delta FUR1	pRDH118; pDF1
M0291	M0013	CBH2Tr, CBH1Ta; delta FUR1	pRDH120; pDF1
M0358	M0282	BGLISf, EGITr, TrCBH2, CBH1Te; delta FUR1; Trp1; His3	pBKD1-BGLI-sEGl; pRDH108; pDF1
M0359	M0288	BGLISf, EGITr, TrCBH2, CBH1Te; delta FUR1; Trp1; His3	pBKD1-BGLI-sEGl; pRDH108; pDF1
M0361	M0249	Ninguno (control); delta FUR1; Trp1; His3	pJC1; pDF1
M0157	<i>Kluyveromyces marxianus</i> (ATCC N.º 10606)	Ninguno	Ninguna
M0158	<i>Kluyveromyces lactis</i> (ATCC N.º 34440)	Ninguno	Ninguna
M0411	M0158 (colonia N.º 1)	BGLISf, EGITr	pBKD1-BGLI-sEGl;
M0412	M0158 (colonia N.º 2)	BGLISf, EGITr	pBKD1-BGLI-sEGl;
M0413	M0157 (colonia N.º 1)	BGLISf, EGITr	pBKD1-BGLI-sEGl;
M0414	M0157 (colonia N.º 2)	BGLISf, EGITr	pBKD1-BGLI-sEGl;
M0491	M0414	BGLISf, EGITr, TeCBH1, TrCBH2	pBKD1-BGLI-sEGl; pMU576andpMU577
M0599	M0414	BGLISf, EGITr, TeCBH1, TrCBH2	pBKD1-BGLI-sEGl; pMU760 and pMU761
M0600	M0414	BGLISf, EGITr, TeCBH1, TrCBH2	pBKD1-BGLI-sEGl; pMU760 and pMU761
de M0601 a M0604; de M0611 a M0617	M0414 (11 colonias muestran mayor actividad avicelasa)	BGLISf, EGITr, C1(a)EG, EGCf, EGNt, EGCa, EGMd, EGNw, EGRs, CBH1Te, CBH1Te+CBD, CBH2Tr, CBH2 C1(b)	pBKD1-BGLI-sEGl; pMU663, pMU755, pMU809, pMU576, pMU661, pMU662, pMU664, pMU665, pMU667, pMU668, pMU750, pMU577
de M0618 a M0625	M0157(8 colonias muestran mayor actividad avicelasa)	C1(a)EG, EGCf, EGNt, EGCa, EGMd, EGNw, EGRs, CBH1Te, CBH1Te+CBD, CBH2Tr, CBH2 C1(b)	pMU663, pMU755, pMU809, pMU576, pMU661, pMU662, pMU664, pMU665, pMU667, pMU668, pMU750, pMU577
yENO1	M0013	ENO1P/T	YEENO-BBH; pDF1
M0419	M0013	ENO1P/T	pMU451
M0420	M0013	CBH1Te	pMU272
M0423	M0013	EG1Tr	pMU463
M0424	M0013	BGL1Sf	pMU464
M0426	M0013	EGRf	pMU469
M0446	M0013	EG C1(a)	pMU465

Nombre	Cepa de fondo	Genes expresados y/o inactivados	Construcciones
M0449	M0013	EGCf	pMU471
M0450	M0013	EGNt	pMU472
M0460	M0013	EGMd	pMU499
M0461	M0013	EGRs	pMU500
M0464	M0013	EGNw	pMU503
M0476	M0013	EGNf	pMU458
Y294/pMI529 fur1Δ	M0013	CBH1Te+MUC	pMU624
Y294/pMI553 fur1Δ	M0013	CBH2Tr, CBH1Te+MUC	pMI553
Y294/pMI574 fur1Δ	M0013	CBH2C1(b)	pMI574
Y294/pMI577 fur1Δ	M0013	CBH2Tr, CBH1Hg	pMI577
Y294/pMI578 fur1Δ	M0013	CBH2Tr, CBH1Te	pMI578
Y294/pMI579 fur1Δ	M0013	CBH2Tr, CBH1 C1(b)	pMI579
Y294/pMI580 fur1Δ	M0013	CBH2 C1(b), CBH1Te+MUC	pMI580
Y294/pMI581 fur1Δ	M0013	CBH2 C1(b), CBH1Te	pMI581
Y294/pMI582 fur1Δ	M0013	CBH2 C1(b), CBH1Hg	pMI582
Y294/pMI583 fur1Δ	M0013	CBH2 C1(b), CBH1 C1(b)	pMI583

El plásmido pBKD1-BGL1-sEG1 (pMU276) se digirió con Accl y se transformó en Y294 de *S. cerevisiae* por electrotransformación para crear una cepa con copias integradas en delta en BGLISf y EGITr, denominada M0243. Después, plásmidos episomales se transformaron en Y294 y/o M0243 de *S. cerevisiae*.

5 Para crear cepas autoselectivas de *S. cerevisiae*, es decir, cepas que pueden cultivarse en medio sin requerir presión selectiva para mantener el plásmido episomal, las cepas se transformaron con pDF1 digeridas con Nsil y NcoI y se seleccionaron en placas con SC-ura-leu. Esto condujo a la alteración del gen FUR1 de *S. cerevisiae*. Se utilizó PCR para confirmar la alteración de FUR1 con los cebadores FUR1-izquierdo (5'-ATTTCTTCTTGAACCATGAAC-3') y FUR1-derecho (5'-CTTAATCAAGACTTCTGTAGCC-3'), donde un tramo de 2568 pb indicó una alteración.

10 M0282 se creó transformando M0248 con pBKD1-BGL1-sEGI digerido con Accl, como se ha descrito anteriormente, excepto que la mezcla de transformación se esparció sobre la placa que contenía celulosa microcristalina bacteriana, CMCB, con extracto de levadura 10 g/l y peptona 20 g/l.

15 La presencia de genes integrados se verificó por PCR para colonias para cepas de *Kluyveromyces*. Las cepas de levadura seleccionadas se volvieron prototróficas por transformación con productos de PCR para genes para complementar sus auxotrofías.

20 Sustratos celulósicos para ensayos enzimáticos

25 La celulosa microcristalina bacteriana (CMCB) fue un obsequio de la compañía CP Kelco. La CMCB recibida se agitó durante una noche a 4 °C en agua. Después de rehidratar el sustrato, se lavó 6 veces con agua y se resuspendió en agua. El peso seco del sustrato se midió secando las muestras a 105 °C hasta que se obtuvo un peso constante.

Se usó Avicel PH105 (FMC Biopolymers) proporcionado tal cual por el fabricante.

30 Se generaron maderas duras mixtas pretratadas por autohidrólisis del sustrato a 160 PSI durante 10 minutos. El material pretratado se lavó 5 veces para retirar inhibidores y azúcares solubles y se resuspendió en agua destilada. Las muestras se secaron durante una noche a 105 °C para determinar el peso seco. El análisis del contenido de azúcar por sacarificación cuantitativa mostró un contenido de glucano de 50 %.

Se preparó celulosa hinchada con ácido fosfórico (PASC) como desvelan Zhang y Lynd (2006), con algunas modificaciones. Se hidrató Avicel PH105 (10 g) con 100 µl de agua destilada en un matraz de 4 l. Se añadieron 800 ml de ácido fosfórico al 86,2 % lentamente al matraz con una primera adición de 300 ml seguido de mezclado y posteriores adiciones de partes alícuotas de 50 ml. La solución transparente se mantuvo a 4 °C durante 1 hora para permitir la solubilización completa de la celulosa, hasta que no quedaran grumos en la mezcla de reacción. A continuación, 2 l de agua destilada enfriada en hielo se añadieron en partes alícuotas de 500 ml mezclando entre las adiciones. Las partes alícuotas de 300 ml de la mezcla se centrifugaron a 5.000 rpm durante 20 minutos a 2 °C y el sobrenadante se retiró. La adición de 300 ml de agua destilada enfriada y posterior centrifugación se repitió 4 veces. Se añadieron 4,2 ml de carbonato de sodio 2 M y 300 ml de agua a la celulosa, seguido de 2 o 3 lavados con agua destilada, hasta que el pH final fue de ~6. Las muestras se secaron hasta obtener un peso constante en un horno a 105 °C para medir el peso seco.

Ensayos enzimáticos

La actividad β-glucosidasa se midió de una manera similar a la de McBride, JE, *et al.*, (Enzyme Microb, Techol. 37: 93-101 (2005)) excepto que el volumen del ensayo se redujo y la reacción se realizó en una placa de microtitulación. En resumen, se cultivaron cepas de levadura hasta saturación en medio YPD o YPC con o sin antibióticos apropiados, se midió la densidad óptica a 600 nm (DO (600)) y se tomó una muestra de los cultivos de 0,5 ml. Esta muestra se centrifugó, el sobrenadante se separó y se guardó y el sedimento celular se lavó 2 veces con tampón citrato 50 mM, pH 5,0. Se construyeron reacciones para sobrenadantes de muestra de 50 µl, tampón citrato 50 µl y 50 µl de sustrato p-nitrofenil-β-D-glucopiranosido (PNPG) 20 mM. Las reacciones con las células lavadas consistían en 25 µl de células, tampón citrato 75 µl y sustrato PNPG 50 µl. Si la actividad era demasiado alta para el intervalo de la curva patrón, se usaba una concentración celular más baja y el ensayo volvía a ejecutarse. La curva patrón consistía en una serie de dilución con factor 2 de patrones de nitrofenol (PNP), comenzando a 500 nM y finalizando a 7,8 nM y se incluyó un blanco de tampón. Después se prepararon diluciones apropiadas de sobrenadante o de células, la placa de microtitulación se incubó a 37 °C durante 10 minutos junto con el sustrato de reacción. La reacción se realizó añadiendo el sustrato, incubando durante 30 minutos y deteniendo con 150 µl de Na₂CO₃ 2 M. Después, la placa se centrifugó a 2500 rpm durante 5 minutos, y 150 µl de sobrenadante se transfirieron a otra placa. En cada pocillo, se leyó la absorbancia a 405 nm.

La actividad endoglucanasa se detectó cualitativamente observando zonas aclaradas en medios completos sintéticos (como se indicó anteriormente, pero incluyendo glucosa 20 g/l) con carboximetil celulosa (CMC) al 0,1 % teñida con rojo Congo (Beguin, Anal. Biochem. 333-6 (1983)). Las células se cultivaron durante 2-3 días en las placas y se retiraron por lavado de la placa con tampón Tris-HCl 1 M pH 7,5. Después, las placas se tiñeron después durante 10 minutos con una solución de rojo Congo al 0,1 %, y posteriormente el colorante extra se retiró por lavado con NaCl 1 M.

La actividad CBH1 se detectó utilizando el sustrato 4-metilumbeliferil-β-D-lactósido (MULac). Los ensayos se realizaron mezclando 50 µl de sobrenadante de levadura con 50 µl de una solución de sustrato MULac 4 mM, preparada en tampón citrato 50 mM, pH 5,5. La reacción se dejó continuar durante 30 minutos y después se detuvo con Na₂CO₃ 1 M. La fluorescencia en cada pocillo se leyó en un lector de placa de microtitulación (ex. a 355 nm y em. a 460 nm).

Cuantificación de la actividad enzimática

La actividad enzimática en PASC y Avicel se midió utilizando el protocolo descrito en Den Haan *et al.*, Enzyme and Microbial Technology 40: 1291-1299 (2007). En resumen, sobrenadantes de levadura se incubaron con celulosa a 4 °C para unir la celulosa. Después la celulosa se filtró del sobrenadante de levadura, se resuspendió en tampón citrato y azida sódica y se incubó a 37 °C. La acumulación de azúcar se midió en la reacción muestreando y realizando un ensayo con fenol y ácido sulfúrico (véase el Ejemplo 10 y la Tabla 9).

También se generaron niveles de actividad Avicel utilizando un método en placa de 96 pocillos. (Véase el Ejemplo 2). Las cepas a ensayar se cultivaron en YPD en placas de 96 pocillos de pocillo profundo a 35 °C con agitación a 900 rpm. Después del cultivo, las placas se centrifugaron a 4000 rpm durante 10 minutos. Se añadió sustrato 300 µl (avicel al 2 %, tampón acetato sódico 50 mM, azida sódica 0,02 %, β-glucosidasa 1 µl por ml) a una placa nueva de 96 pocillos de pocillo profundo, sin dejar que el avicel se sedimente. Se añadieron 300 µl de sobrenadante de levadura a este sustrato, y se extrajeron 100 µl para una muestra inicial. La placa de ensayo se incubó a 35 °C con agitación a 800 rpm y se extrajeron muestras a las 24 y 48 horas. Las muestras se colocaron en placas de PCR de 96 pocillos y se centrifugaron a 2000 rpm durante 2 minutos. Después, se añadieron 50 µl de sobrenadante a 100 µl de reactivo DNS previamente colocado en una placa de PCR de 96 pocillos distinta, se mezcló y se calentó a 99 °C durante 5 minutos en un aparato de PCR, seguido de enfriamiento a 4 °C. Se transfirieron 50 µl a una placa de microtitulación y se midió la absorbancia a 565 nm. La conversión de avicel se calculó de la siguiente manera:

$$Y = \frac{(\text{DO}(T=24 \text{ o } 48) - \text{DO}(T=0)) \times 100\%}{S \times A} = \frac{\Delta \text{DO} \times 100}{S \times A} = \frac{\Delta \text{DO} \times 100}{0,1 \times 10}$$

Y - % de Avicel convertido a las 24 o 48 h

S – calibrado de pendiente con DNS/glucosa que es de 0,1 para DNS a 565 nm

A - concentración de Avicel a T=0 que es de 10 g/l para Avicel al 1 %

5

Ejemplo 1: producción *Kluyveromyces* que expresa β -glucosidasa y endoglucanasa heterólogas

10

Para ensayar la capacidad de *Kluyveromyces* para expresar celulasas heterólogas funcionales, dos cepas de *Kluyveromyces*, *Kluyveromyces marxianus* (cepa ATCC n.º 10606; MO157) y *Kluyveromyces lactis* (cepa ATCC n.º 34440) se transformaron con vectores que codificaban celulasas heterólogas.

15

Los vectores, que contenían secuencias de integración delta de levadura, el marcador KanMX y las secuencias codificantes de BGLI S.f. y de EGI T.r. (pBKD-BFLI-sEG1), se transformaron en *Kluyveromyces* de acuerdo con el protocolo de transformación de levaduras como se ha descrito anteriormente y se seleccionaron en G418. Los transformantes se verificaron por PCR y después se ensayaron evaluando con CMC. En la Figura 1 se muestran los resultados. La presencia de la actividad de celulasa heteróloga se indica mediante una zona aclarada en la placa con CMC. Como se muestra en la Figura 1, ni la cepa de *K. lactis* no transformada (colonia 8) ni la cepa de *K. marxianus* no transformada (colonia 16) mostró actividad endoglucanasa. Sin embargo, 6 de 7 colonias de *K. lactis* transformadas mostraron actividad de CMCasa, y las 7 colonias de *K. marxianus* transformadas mostraron actividad CMCasa. MO413 y MO414 se identificaron como dos colonias de *K. marxianus* que mostraban actividad CMCasa.

20

Ejemplo 2: producción de *Kluyveromyces* que expresa CBH1 y CBH2

25

También se examinó la capacidad de *Kluyveromyces* para expresar celobiohidrolasas heterólogas funcionales. En estos experimentos, *K. marxianus* (MO157) se transformó con construcciones que contenían CBH2 de *T. reesei*, CBH1 de *T. emersonii* o ambas. De manera similar, MO414 (*K. marxianus* transformada con BGLI S.f. y con EGI T.r.) se transformó con construcciones que contenían CBH2 de *T. reesei*, CBH1 de *T. emersonii* o ambas.

30

Las transformaciones se realizaron como se ha descrito anteriormente. Después, la actividad CBH1 se detectó utilizando el sustrato 4-metilumbeliferil- β -D-lactósido (MU-Lac) como se ha descrito anteriormente. El ensayo se realizó en ocho colonias de cada transformante y se estableció el promedio de las tres colonias que mostraron la actividad más elevada. Los resultados se muestran en la Figura 2 y demuestran que las cepas transformadas con CBH1 de *T. emersonii* tenían una actividad MU-lac elevada.

35

También se evaluó la actividad de cepas de *Kluyveromyces* que expresaban celobiohidrolasas heterólogas en Avicel. En un experimento, MO413 se transformó con vectores que contenían las secuencias codificantes de CBH2 de *T. reesei* y de CBH1 de *T. emersonii* junto con un marcador de zeocina. Mediante esta transformación se creó una nueva cepa MO491 y mostró actividad MU-lactósido. En un segundo experimento, MO413 se transformó con vectores que contenían las secuencias codificantes de CBH2 de *T. reesei* y de CBH1 de *T. emersonii* junto con un marcador de higromicina, y de esta transformación se aislaron las cepas MO599 y MO600. La actividad en Avicel se evaluó a las 48 horas como se ha descrito anteriormente, y los resultados, mostrados en la Figura 3, demuestran que las cepas de *Kluyveromyces* que expresan celulasas heterólogas tienen actividad Avicelasa a 35 °C. También se demostró actividad Avicelasa a 45 °C (datos no mostrados).

40

45

Ejemplo 3: producción de *Kluyveromyces* que expresa una biblioteca de celulasas

50

También se crearon cepas de *Kluyveromyces* transformando levadura con una biblioteca de celulasas (la creación de bibliotecas se ha descrito anteriormente). Por ejemplo, MO413 se transformó con una biblioteca de celulasas que contenían un marcador de zeocina para producir nuevas cepas MO601-MO604 y MO611-MO617. Además, MO157 (*K. marxianus*) se transformó con la misma biblioteca y se identificaron nuevas cepas MO618-MO625. La actividad en Avicel se evaluó a las 48 horas como se ha descrito anteriormente, y los resultados, mostrados en la Figura 3, demuestran que las cepas de *Kluyveromyces*, transformadas con una biblioteca de celulasas heterólogas, también tenían actividad Avicelasa a 35 °C. Los transformantes de MO157 con la biblioteca mostraron la actividad más elevada. También se demostró actividad Avicelasa a 45 °C (datos no mostrados).

55

Ejemplo 4: producción de etanol por *Kluyveromyces* transformada

60

Para determinar si *Kluyveromyces*, que expresa celulasas heterólogas, podía producir etanol a partir de Avicel, se cultivaron precultivos durante 24 horas en YPD (YPD como indicado anteriormente, con glucosa 20 g/l; 25 ml en un matraz agitador de 250 ml) con agitación a 300 rpm a 35 °C. Después de 24 y 48 horas, se añadieron 40 g/l de glucosa adicional. Al cabo de 72 horas, el pH de los cultivos se ajustó a ~5,0 con tampón citrato (el pH inicial del tampón fue de 5,5 y la concentración final fue de 50 mM) y el cultivo se añadió a un matraz agitador de plástico hermético que contenía 5,5 gramos de Avicel (concentración final 10 % (p/v)). Se usó Avicel PH105 (FMC

Biopolymers) tal como proporcionó el fabricante. El cultivo se incubó a 35 °C con agitación a 150 rpm.

La cuantificación de etanol en muestras de fermentación se realizó por análisis HPLC, y las concentraciones de etanol iniciales en frascos (de precultivos) se restaron de todos los puntos de datos posteriores (las concentraciones de etanol iniciales variaban entre 0 y aproximadamente 6 g/l). La concentración de glucosa inicial en todas las cepas, excepto el MO603, fue de 0,000 g/l. Para esta cepa la concentración fue de 0,069 g/l, lo que produciría un máximo en 0,035 g/l de etanol a partir del azúcar inicial.

Los resultados, como se muestra en la Figura 4, demuestran que las cepas de *K. marxianus* modificadas por ingeniería genética también pudieron producir etanol directamente a partir de Avicel. La cepa MO157, la cepa de control no transformada, mostró una disminución constante en la concentración de etanol durante el experimento. Esto se debe al consumo de etanol por la cepa debido a la presencia de una pequeña cantidad de oxígeno en los matraces.

De las dos cepas transformadas con CBH2 de *T. reesei* y CBH1 de *T. emersonii* con el marcador de higromicina (MO599 y M600), una de ellas (MO599) mostró producción de etanol. Además, de las cinco cepas transformadas con CBH2 de *T. reesei* y con CBH1 de *T. emersonii* con el marcador de zeocina, cuatro (MO601, MO602, MO604 y MO491) mostraron producción de etanol. Esto demuestra que las cepas de *K. marxianus* termotolerantes modificadas por ingeniería genética podían producir etanol directamente a partir de la celulosa cristalina recalcitrante, Avicel.

Ejemplo: producción de *S. cerevisiae* que expresa celulasas heterólogas

También se produjeron cepas de *S. cerevisiae* que expresaban celulasas heterólogas y se sometieron a ensayo con respecto a su capacidad de cultivo en medios que contenían celulosa microcristalina bacteriana (CMCB). En estos experimentos, las condiciones microaerobias se mantuvieron cultivando las cepas en CMCB en tubos Hungate herméticos con una atmósfera de aire.

Las cepas que expresaban CBH1 de *T. emersonii* y CBH2 de *T. reesei* (MO248) se transformaron con una construcción que permitía la expresión de EGI de *T. reesei* y de BGLI de *S. fibuligera* (pKD-BGLI-sEGI). Esa transformación se sembró en una placa con agar sólido y CMCB y, después de 7 días, aparecieron cinco colonias en la placa (datos no mostrados). La levadura de la más grande de las cinco colonias se aisló como cepa MO282. (MO282 se ha descrito con más detalle anteriormente). Las tres cepas de control se sometieron a ensayo con respecto al crecimiento en las mismas placas. Una cepa expresó CBH1 de *T. emersonii* y CBH2 de *T. reesei* y dos cepas expresaron EGI de *T. reesei* y BGLI de *S. fibuligera*. No aparecieron colonias en las placas con cepas de levadura de control (datos no mostrados).

La capacidad de MO282 para crecer en CMCB también se sometió a ensayo utilizando medios líquidos. La Figura 5 muestra que MO282, que expresa las 4 celulasas secretadas, creció en CMCB a un grado mucho más alto que un plásmido solo de control (MO249), que una cepa que expresaba solo CBH1 de *T. emersonii* y CBH2 de *T. reesei* (MO249) y que una cepa que expresaba 4 celulasas ligadas (MO144).

Estos resultados indican que las levaduras que expresan CBH1 de *T. emersonii* secretada, CBH2 de *T. reesei*, EGI de *T. reesei* y BGLI de *S. fibuligera*, secretadas de manera heteróloga, pueden crecer en celulosa microcristalina bacteriana.

Ejemplo 6: *S. cerevisiae*, que expresa celulasas heterólogas, puede producir etanol a partir de Avicel y de madera dura pretratada.

Para determinar si *S. cerevisiae* transformada podía producir etanol directamente a partir de celulosa sin enzimas celulasas añadidas de manera exógena, se cultivaron cepas transformadas en Avicel como única fuente de carbono. Se usó Avicel PH105 (FMC Biopolymers) tal como como proporcionó el fabricante.

Los medios Avicel se prepararon utilizando componentes no glucosídicos de medio completo sintético para levadura incluyendo, base de nitrógeno de levadura sin aminoácidos, 6,7 g/l y complementado con una mezcla de aminoácidos completa (mezcla complementaria completa). En algunos casos, como complemento se utilizó extracto de levadura (10 g/l) y peptona (20 g/l) (YP) en los experimentos de cultivo. Las condiciones de cultivo fueron anaerobias y se mantuvieron enjuagando frascos de vidrio herméticos con N₂ después de la adición de la fuente de carbono y antes esterilizar en autoclave. Después de esterilizar en autoclave, se añadieron componentes de medios no carbónicos como diluciones de 10X por esterilización con filtro. La inoculación en cultivos Avicel se realizó al 20 % en volumen. La cuantificación de etanol en muestras de fermentación se realizó mediante análisis HPLC y las concentraciones de etanol iniciales en frascos (a partir de precultivos) se restaron de todos los puntos de datos posteriores.

Como se muestra en la Figura 6, cuando se utilizaban los componentes de medios YNB, la cepa MO288 (que expresaba BGLI de *S. fibuligera*, EGI de *T. reesei*, CBH2 de *T. reesei* y CBH1 de *T. emersonii*) pudo producir etanol

directamente a partir de avicel PH105 en comparación con la cepa de control (M0249).

La capacidad de MO288 para producir etanol a partir de celulosa también se demostró utilizando maderas duras pretratadas. Se generaron maderas duras mixtas pretratadas por autohidrólisis del sustrato a 160 PSI durante 10 minutos. El material previamente tratado se lavó 5 veces para retirar inhibidores y azúcares solubles y se resuspendió en agua destilada. Las muestras se secaron durante una noche a 105 °C para determinar el peso seco. El análisis del contenido de azúcar por sacarificación cuantitativa mostró un contenido de glucano al 50 %. Los medios y las condiciones de los cultivos eran como se ha descrito anteriormente para los experimentos con Avicel, excepto que los cultivos se inocularon al 10 % en volumen.

Los datos presentados en la Figura 7 demuestran que MO288 también podía fabricar etanol a partir de maderas duras pretratadas sin adición enzimática. La cepa fabricó ~0,5 g/l más que la cepa de control cuando se utilizaba medio YP y ~0,2 g/l cuando el medio que se utilizaba era YNB.

Estos datos demuestran que las levaduras que expresan CBH1 de *T. emersonii* secretada, CBH2 de *T. reesei*, EGI de *T. reesei* y BGLI de *S. fibuligera* heterológamente pueden producir etanol a partir de celulosa sin la adición de celulasas exógenas.

Ejemplo 7: las cepas de levadura transformadas y las celulasas añadidas externamente actúan sinérgicamente para producir etanol a partir de maderas duras mixtas pretratadas.

La producción de etanol a partir de biomasa se realiza normalmente utilizando un tipo de proceso de SFS en el que las enzimas celulasas se añaden exógenamente a una reacción que contiene biomasa celulósica pretratada, medios de crecimiento de levadura y levaduras. Para determinar si la levadura que expresa celulasas recombinantes podría mejorar este proceso, las levaduras recombinantes que expresan celulasas secretadas se cultivaron en presencia de una serie de concentraciones de celulosa añadida exógenamente. Las condiciones y los medios del cultivo fueron como se ha descrito en los ejemplos anteriores.

En estos experimentos, se comparó una cepa de levadura recombinante que expresaba cuatro celulasas secretadas (MO288) directamente con la cepa de control (M0249) en las mismas condiciones. Se añadieron celulasas externas a concentraciones de 25 mg de celulosa por gramo de celulosa (100 %), 22,5 mg de celulosa por gramo de celulosa (90 %), 18,75 mg de celulosa por gramo de celulosa (75 %) o 6,25 mg de celulosa por gramo de celulosa (25 %). También se realizaron experimentos sin añadir celulasas externas (0 %). Como fuente de celulosa se utilizaron maderas duras mixtas pretratadas (preparadas como se describe en los ejemplos anteriores) a una concentración de sólidos inicial de 5 %. En la Figura 8 se presentan los datos. A partir de estos datos, está claro que la cepa productora de celulasas elabora etanol adicional con respecto a la cepa de control para cada una de las concentraciones de carga de celulosa ensayadas.

Para examinar este efecto con más detalle, se evaluó la producción de etanol a diferentes concentraciones de celulosa externa en dos tipos de medios diferentes utilizando madera dura mixta tratada previamente. En la Figura 9 se muestran los resultados. En medios YP, MO288 elabora 6-9 % más etanol a las cargas de celulosa más elevadas, únicamente 1 % más a una carga de 25 % y 100 % más cuando no se carga celulosa. En medios YNB, MO288 elabora 20-40 % más etanol a bajas cargas de celulosa, y ~10 % más de etanol a cargas de celulosa más altas. Estos resultados pueden utilizarse para determinar la cantidad de celulosa que puede retirarse del proceso obteniendo la misma producción global de etanol. Para medios YP, la carga de celulosa puede reducirse ~15 % en comparación con el control, y para medios YNB, la carga de celulosa puede reducirse ~5 %. A cargas de celulosa que no sean cero, la productividad del etanol aumentó entre un 5 y un 20 % para cepas que expresaban celulasas en medios YP en comparación con el control. Esto aumentó entre 10 y 20 % para cepas cultivadas en medios YNB en comparación con el control.

Estos datos demuestran que los procesos de SFS previos pueden mejorarse en cuanto a la producción de etanol a partir de biomasa y productividad de etanol si las cepas que expresan celulasas secretadas se usan en combinación con celulasas añadidas exógenamente. De manera similar, las cargas de celulosa necesarias para conseguir un porcentaje particular de rendimiento teórico de etanol pueden reducirse cuando se añaden cepas que expresan celulasas recombinantes.

Ejemplo 8: las cepas de levadura transformadas también aumentan la eficacia de las celulasas añadidas externamente en la producción de etanol a partir de Avicel.

Para ensayar si esta misma tendencia se mantenía a altas concentraciones de sustrato, estos experimentos se repitieron utilizando Avicel PH105 al 15 % como sustrato en lugar de madera dura mixta pretratada al 5 %. En las Figuras 10 y 11 se muestran los resultados. La cepa que fabrica celulasas (MO288) produjo habitualmente más etanol a partir de Avicel que la cepa de levadura de control (M0249) en las mismas condiciones, incluso a concentraciones de etanol aumentadas (Figura 10). Por ejemplo, cuando se cargaban 25 mg de celulosa por gramo de celulosa en la reacción de SFS, la cepa de ensayo (M0288) produjo 54 g/l, mientras que la de control (M0249) produjo 50 g/l.

Para examinar el desplazamiento de celulasa se determinó el porcentaje del rendimiento teórico de etanol conseguido a diferentes cargas de celulasa. Los resultados presentados en la Figura 12 se repitieron por triplicado para M0288 y M0249, permitiendo calcular las desviaciones típicas para los rendimientos de etanol aumentados. En la Figura 12 se presentan los datos que pueden utilizarse para calcular el desplazamiento de celulasa. La Figura 12 presenta ahorros de enzima celulasa basándose en el rendimiento de etanol teórico al cabo de 168 horas en un experimento de SFS. El experimento de sacarificación y fermentación simultáneas, SFS, se realizó en 30 ml de YP purgado nitrógeno más Avicel al 15 % en frascos a presión. Se utilizó la mezcla de celulasa externa a una proporción de 5 Spyzme:1 Novozyme-188. El experimento continuó durante 168 horas y el muestreo se realizó cada día para la estimación del etanol mediante HPLC. En la figura las flechas representan la carga de celulasa necesaria para conseguir la misma producción de etanol a partir de celulosa como control. Esta carga es consistentemente menor que para el control (es decir, la producción de etanol es consistentemente más elevada). Para datos al cabo de 168 horas, el desplazamiento promedio de celulasa (cantidad menor que la necesaria para cargar) es de 13,3 % \pm 4,9 %.

Ejemplo 9: utilización de Cbh1 artificial para producir etanol

Para diseñar una proteína CBH1 con actividad celulasa eficaz, se alinearon 17 secuencias de la proteína CBH1 de la base de datos del NCBI (Tabla 7).

Tabla 7. Genes de CBH1 fúngicos utilizados para el alineamiento.

Organismo	Genbank n.º
<i>Neosartorya fischeri</i>	<u>XM 001258277</u>
<i>Gibberella zeae</i>	<u>AY196784</u>
<i>Penicillium janthinellum</i>	<u>X59054</u>
<i>Nectria haematococca</i>	<u>AY502070</u>
<i>Fusarium poae</i>	<u>AY706934</u>
<i>Chaetomium thermophilum</i>	<u>AY861347</u>
<i>Aspergillus terreus</i>	<u>XM 001214180</u>
<i>Penicillium chrysogenum</i>	<u>AY790330</u>
<i>Neurospora crassa</i>	<u>X77778</u>
<i>Trichoderma viride</i>	<u>AY368686</u>
<i>Humicola grisea</i>	<u>X17258</u>
<i>Thermoascus aurantiacus</i>	<u>AF421954</u>
<i>Talaromyces emersonii</i>	<u>AAL89553</u>
<i>Trichoderma reesei</i>	<u>P62694</u>
<i>Phanerochaete chrysosporium</i>	<u>Z29653</u>
<i>Aspergillus niger</i>	<u>XM 001391971</u>
<i>Aspergillus niger</i>	<u>XM 001389539</u>

Se diseñó la secuencia de proteína artificial como una secuencia consenso (la más común) para estas proteínas. La secuencia señal predicha se cambió por la secuencia preseñal del factor de apareamiento alfa de *S. cerevisiae*, y la secuencia de la proteína CBH1 consenso se muestra a continuación. Las letras mayúsculas indican la secuencia preseñal del factor de apareamiento alfa de *S. cerevisiae*.

MRFPSIFTAVLFAASSAL Aqqaglttaethpsltwqkctsggscttvngsvvidanwrwvhatsgst
 ncytgntwdttlcpddvtcaqncaldgadysstygvttsgnslrlnfvvtqgsqknvgsrlylmeddttymfklg
 qeftfdvdsnlpcglnalyfvamdadggmskypgnkagakygtgycdsqcprdlkfingqanvegweps
 sndanagignhgccaemdiweansistaftphpcdtigtmccegdscggyssdryggctdpdgcdfnpyr
 mgnktfygpgktvdttkvtvvtqfitgssgtlseikrfyvqngkvipnsestisgvsngsittdfetaqktafgdtd
 dfakkgglegmgkalaqgmvlvmslwdhaanmlwldstyptdatsstpgaargscdtssgvpadvansp
 nsyvtfstnikfpgigstfig (SEQ ID NO:43).

Se desarrolló una secuencia de *S. cerevisiae* y *K. lactis* con codones optimizados para expresar la secuencia
 consenso de CBH1 (SEQ ID NO: 44) y se muestra a continuación.

5

atgagattccttcaatcttactgctgtttgtcgcagcctcaagtgccttagcacaacaggccggaacattgacag
 cagaaactcaiccttcttaacctggcaaaagtgcacttctggaggctcatgcactacagtgaatggatctgctg
 atcgatgcaaactggagatgggttcacgcaactcaggttctaccaactgtataccggaacactgggacacca
 cattgtgccagatgacgtcacgtgcgctcagaactgtgcttggatggagctgattacagttaacctatggtgia
 actacatccggaactctttgagattaacttcgttactcaaggaagtcaaaagaacgttggttctagattgtactaa
 tggaggacgatacaacctatcaaatgttcaaatgttaggtcaggagttcaccttgacgtagatgacagtaactgc
 catgtgggttaaacggagcttatacttggcaatggatgctgacgggtggaatgtccaagtatccaggaacaa
 agccgggtgcaaagtacggtacaggatattgtattcacagtcacctagagattgaagttcattaacggtaagca
 aatgtggagggttgggaacctctagtaacgatgccaatgcgggtattggaatcatgggtcctgttgcgctgaga
 tggatatctgggagggccaactcaatactactgcctttaccctcacccatgcgatacaattggtcaaacatgtg
 aggggtgattcatgtggtggaacctactcctctgatagatacggaggtacatgcgatccagatggttgcgactta
 ccatacagaatgggaacaaaacctttacggtcctggaaagacagttgatactaccaagaaagtaacagtcgtg
 acccagttatcaccggtagttctggaacctatccgaaatcaaaagattctacgttcagaacggtaaaagtaattcca
 aacagtgaatctacaatttcaggagtgagtggaattctattactaccgactttgtacagctcagaaaacagcattg
 gtgacaccgatgacttctagaagggtggattagaaggtatgggtaaaagcttggcccagggaatggtgtagt
 tatgtctttatgggatgacacgccgcaaatatggtatggttgattcaacatatccaactgatgccacaagtagtaca
 cctggagctgccagaggttcttggatacatctccggtgtccagccgatgtagaagcaaattctcctaactcctat
 gttaccttctccaatataaagtttggccaatcgggtcaacattcactggttaa (SEQ ID NO:44)

10

La secuencia con codones optimizados se insertó en el vector de expresión de levadura episomal (pMU451) bajo el
 control del promotor y terminador de ENO1 en sitios PacI/AscI. Las construcciones de expresión resultantes
 (pMU505) se transformaron en la cepa hospedadora MO375 que procedía de Y294 (MO013) en la que las
 auxotrofias His3 y Trp1 se rescataron mediante transformación con los productos PCR His3 y Trp1 de *S. cerevisiae*.
 La cepa resultante que expresaba la secuencia consenso de CBH1 se denominó MO429.

15

Para determinar si MO429 tenía actividad celulasa, se realizó un ensayo de conversión con Avicel como se describe
 anteriormente y se midió a las 24 horas. Como se muestra en la Figura 14, *S. cerevisiae* que expresaba la secuencia
 Cbh1 consenso (MO429) mostró actividad celulasa en comparación con un control negativo transformado con un
 vector vacío (MO419). La actividad celulasa de MO429 también se comparó con la de las cepas de levadura que

expresaban otras celulasas heterólogas. En la Tabla 8 siguiente se resumen las cepas ensayadas.

Tabla 8: cepas celulolíticas utilizadas en el ensayo de conversión con Avicel

Cepa n.º	Descripción	Celulasa	Familia	Organismo	Actividad	Señal
M0419	MO375+pM U451	ninguna			ninguna	
M0420	MO375+pM U272	CBH1	Hongo	<i>Talaromyces emersonii</i>	exo	nativa
M0429	MO375+pM U505	consenso CBH1 fúngica	Hongo	N/A	exo	αMFpre de S.c
M0445	MO375+pM U459	CBH1	Hongo	<i>Neosartorya fischeri</i>	exo	αMFpre de S.c
M0456	MO375+pM U495	CBH1	Hongo	<i>Chaetomium thermophilum</i>	exo	αMFpre de S.c
M0457	MO375+pM U496	CBH1	Hongo	<i>Aspergillus terreus</i>	exo	αMFpre de S.c
M0458	MO375+pM U497	CBH1	Hongo	<i>Penicillium chrysogenum</i>	exo	αMFpre de S.c

5 Todas las cepas de la Tabla 8 procedían de la misma cepa precursora MO375 y se transformaron con un vector de expresión episomal. MO420, MO429, MO445, MO456, MO457 y MO458 se crearon utilizando vectores de levadura episomales que contenían los genes de celulasa heteróloga indicados en la tabla, cuyos codones estaban optimizados para la expresión en *S. cerevisiae* y *K. lactis*. Las celulasas en MO429, MO445, MO456, MO457 y MO458 se expresaron bajo el control del promotor y terminador de ENO1 de *S. cerevisiae*. La CBH1 de *T. emersonii* se expresó con su propia secuencia señal nativa. Como se muestra en la Figura 14, la actividad secretada en Avicel de la CBH1 consenso fue comparable con la actividad de otras CBH1 fúngicas expresadas en el mismo vector y en la misma cepa hospedadora.

15 Ejemplo 10: comparación de la actividad celulasa en *S. cerevisiae*

Las cepas de *S. cerevisiae* se transformaron con polinucleótidos que codificaban diversas celobiohidrolasas heterólogas diferentes y su actividad en PASC y Avicel se evaluó como se ha descrito anteriormente. Los resultados se muestran en la tabla siguiente.

20 Tabla 9: actividad celobiohidrolasa en *S. cerevisiae*.

Plásmido	Casete(s) de Expresión	Act. (PASC) (mU/gDCW)	Act. (Avicel) (mU/gDCW)
yENO1	<i>ENO1</i> p/t	2,6861,1	2,9960,7
M0265	<i>ENO1</i> p/t-sH.g.cbh1	32,8266,5	34,8562,0
M0266	<i>ENO1</i> p/t-sT.a.cbh1	38,5665,9	38,1564,1
M0247	<i>ENO1</i> p/t-sT.e.cbh1	75,60613,1	21,4266,1
M0248	<i>PGK1</i> p/t-sT.r.cbh2 & <i>ENO1</i> p/t-sT.e.cbh1	174,3566,5	40,564,9
M0289	<i>PGK1</i> p/t-sT.r.cbh2 & <i>ENO1</i> p/t-sH.g.cbh1	No medida	106,266,8
M0291	<i>PGK1</i> p/t-sT.r.cbh2 & <i>ENO1</i> p/t-sTa.cbh1	No medida	32,765,7

25 Además, la actividad en Avicel se ensayó utilizando un ensayo de 96 placas, y los resultados se muestran en la Figura 14. En la Figura, para cada cepa, la primera barra indica la liberación de azúcar a las 24 horas y la segunda barra indica la liberación de azúcar a las 48 horas. Las CBH1 expresadas individualmente, o en combinación con CBH2 de *T. reesei*, mostraron alguna actividad avicel que alcanzó una conversión de avicel del 10 % al cabo de 48 horas. Las combinaciones de CBH1 con CBH2 de *C. lucknowense* alcanzaron conversiones de avicel mucho más altas de aproximadamente 22 % de conversión a las 48 horas en combinación con la CBH1 de *T. emersonii* con CBD unido.

30 En la Figura 15 se muestran los datos de actividad avicel para las endoglucanasas ensayadas en *S. cerevisiae*. Los datos demuestran que entre las EG ensayadas, la EG de *C. formosanus* demostró la mayor actividad avicel cuando se expresaba en *S. cerevisiae*.

Ejemplo 11: cocultivos de cepas de levadura que expresan diferentes celulasas heterólogas producen etanol a partir

de Avicel

Un cocultivo de diversas cepas de levadura productoras de celulasa también mostró la capacidad de fabricar etanol a partir de Avicel PH105 en medios YNB (Figura 16). En este experimento se mezclaron 5 cepas que producían independientemente CBH1 de *T. emersonii* (M0247), CBH1 de *T. aurantiacus* (M0266), CBH1 de *H. grisea* (M0265), una combinación de CBH1 de *T. emersonii* y de CBH2 de *T. reesei* (M0248) y una combinación de EGI de *T. reesei* y de BGLI de *S. fibuligera* (M0244) en la misma proporción por volumen y después se inocularon al 20 % en volumen. Cada una de las celulasas expresadas heterológamente en cada una de estas cepas se secretó. Los medios y las condiciones de cultivo eran como se ha descrito anteriormente para los experimentos con Avicel. Los datos de la Figura 16 demuestran que las celulasas heterólogas no requieren expresarse en una cepa de levadura individual para producir etanol a partir de celulosa. En lugar de esto, las cepas de levadura que expresan diferentes celulasas heterólogas secretadas pueden cultivarse conjuntamente para producir etanol a partir de celulosa sin la adición de ninguna celulasa exógena.

También se evaluó un cocultivo utilizando una combinación de celulasas diferente. En este conjunto de experimentos de cocultivo, se cultivaron cuatro cepas de levadura conjuntamente: M0566 (M0424 con delección FUR): BGLISf secretada; M0592 (M0449 con delección FUR): EGICf secretada; M0563 (igual que fur1Δ Y294/pMI574): secretada CBH2b CI y M0567 (al igual que fur1Δ Y294/pMI529): secretada CBH1Te + CBD. Estas cepas se cultivaron en líquido YPD durante 3 días, hasta que el cultivo se saturó para precultivo. En este momento se utilizaron para inocular experimentos en los que se utilizó avicel (10 %) como sustrato, y las 4 cepas se mezclaron al mismo volumen antes de la inoculación.

La Figura 17 demuestra que las cepas cocultivadas pueden producir etanol directamente a partir de avicel sin que se añada ninguna enzima celulasa. El cocultivo produce aproximadamente 4 veces más etanol después de 168 horas en comparación con la cepa de control y aproximadamente 3 veces más que M0288.

Este cocultivo también se utilizó en experimentos de SFS en los que se utilizó el coctel de enzima celulasa Zoomerasa a 5 cargas diferentes (10 mg de proteína/g de avicel, 7,5 mg/g, 5 mg/g y 2,5 mg/g y 0 mg/g) y las cepas se inocularon al 10 % en volumen.

La Figura 18 presenta los datos sin procesar para la producción de etanol a diversas cargas de celulasa por el cocultivo, M0288 y M0249. La Figura 18A muestra que a todas las cargas de celulasa ensayadas, las cepas cocultivadas produjeron significativamente más etanol que un control no productor de celulasa. La Figura 18B muestra que a todas las cargas de celulasa ensayadas, el cocultivo produjo más etanol que la cepa previamente ensayada M0288. La Figura 19 muestra el porcentaje de rendimiento teórico de etanol que podría obtenerse con cada uno de estos cultivos después de 168 horas de SFS utilizando diversas cargas de celulasa. Los datos demuestran que las cepas cocultivadas conseguirían aproximadamente una reducción de 2 veces en celulasa con respecto a la cepa de control y aproximadamente una reducción del 35 % en comparación con M0288.

Estos datos demuestran que la combinación de celulasas en este cocultivo es altamente eficaz en la producción de etanol.

Ejemplo 12: construcción de una cepa fuerte que utiliza xilosa

La cepa M0509 (designación de depósito ATCC _____, depositada el 23 de noviembre del 2009) es una cepa de *Saccharomyces cerevisiae* que combina la capacidad de metabolizar xilosa con la fuerza requerida para fermentar azúcares en presencia de inhibidores de madera dura pretratada. M0509 se creó en un proceso de tres etapas. En primer lugar, cepas industriales de *S. cerevisiae* se evaluaron por comparación para identificar cepas que poseían un nivel de fuerza/vigor suficiente para la sacarificación y fermentación simultáneas (SFS) de sustratos de madera dura mixtos pretratados. La cepa M0086, una cepa diploide de cepas de *S. cerevisiae*, cumplió este primer requisito. En segundo lugar, M0086 se modificó genéticamente con la capacidad de utilizar xilosa, dando como resultado la cepa M0407. En tercer lugar, M0407 se adaptó durante varias semanas en un quimiostato que contenía medios con xilosa con inhibidores de pretratamiento, generando la cepa M0509.

La cepa M0407 se modificó genéticamente a partir de M0086 para utilizar xilosa. Esta modificación genética requirió siete modificaciones genéticas. La modificación principal fue la expresión funcional del gen de xilosa isomerasa heterólogo, XilA, aislado del hongo anaerobio *Piromyces* sp. E2. Los genes estructurales de *S. cerevisiae* que codifican las cinco enzimas implicadas en la conversión de xilulosa en productos intermedios glucolíticos también se sobreexpresaron: xiluloquinasa, ribulosa 5-fosfato isomerasa, ribulosa 5-fosfato epimerasa, transcetolasa y transaldolasa. Además, el gen *GRE3* que codifica la aldosa reductasa se delecionó para minimizar la producción de xilitol. Los siete genes modificados se indican en la Figura 39. Las modificaciones genéticas en los locus *GRE3*, *RKI1*, *RPE1*, *TAL1* y *TKL1* se diseñaron para dejar atrás el ADN del vector mínimo y sin marcadores antibióticos. Cada ADN del locus se secuenció para confirmar los resultados esperados. Cada una de las siete modificaciones genéticas se introdujo secuencialmente en la cepa M0086. La Figura 40 muestra la progresión de las modificaciones desde la parte superior a la inferior junto con las designaciones para la cepa en cada etapa en el proceso, comenzando con M0086 y finalizando con M0407

La delección de *GRE3* y la expresión aumentada de *RKI1*, *RPE1*, *TAL1* y *TKL1* implica modificaciones de los locus de *S. cerevisiae* endógenos. En el caso de *GRE3*, ambos alelos se deleccionaron. Para los otros cuatro locus, solamente se modificó un único alelo. Todas las modificaciones de los locus endógenos requirió el uso de marcadores antibióticos seleccionables incluyendo *kan^r* del transposón de *Escherichia coli Tn903* (que confiere resistencia a G418), *nat1* de *Streptomyces noursei* (que confiere resistencia a clonNAT/nourseotricina) y *dsdA* de *Escherichia coli* (que confiere resistencia a D-serina). Después de la selección para una modificación genómica deseada, el marcador antibiótico se escindió del genoma utilizando el sistema loxP/cre recombinasa. La cre recombinasa mantuvo en el plásmido pMU210 que contenía un marcador de resistencia a zeocina. La pérdida de pMU210 así como de todos los marcadores antibióticos se ensayó en los medios selectivos apropiados. El genotificado por PCR posterior y la secuenciación de ADN confirmó la retirada de los marcadores antibióticos de los locus genómicos modificados.

La sobreexpresión de *RKI1*, *RPE1*, *TAL1* y *TKL1* se consiguió colocando el promotor de la triosa fosfato isomerasa *TPI*, de *S. cerevisiae*, inmediatamente en 5' de cada una de las cuatro ORF. Para *TAL1* y *RKI1*, se deleccionaron pequeñas partes de sus promotores endógenos. Para impedir la alteración de las ORF adyacentes y de posibles elementos reguladores transcripcionales, la introducción del promotor de *TPI* en los locus *RPE1* y *TKL1* se realizó de tal manera que los locus *RPE1* y *TKL1* se duplicaron con las copias duplicadas de ambos locus que regulaba el promotor de *TPI*.

Para reforzar la utilización de xilosa de M0407 y aumentar su tolerancia inhibitoria al pretratamiento, la cepa se mantuvo en un quimiostato durante cuatro semanas con las siguientes condiciones secuenciales descritas en la Tabla 10

Tabla 10: condiciones para mejorar M0407.

Duración (días)	Tiempo de Residencia (h)	Medios
5	24	YPX, 20 g/l xilosa
5	18	YPX, 20 g/l xilosa
7	24	YPX + 25 % de un washate al 30 % MS129 (21,5 g/l xilosa)
14	24	YPX + 75 % de un washate al 30 % MS129 (~22 % equivalentes sólidos)

Una alícuota de cultivo del quimiostato adaptado se sembró en placas en YPXi50 % y nueve colonias "adaptadas" de M0407 se exploraron en medios YPDxi (glucosa 100 g/l, xilosa 50 g/l, MS149 pressate al 25 %). M0407 y M0228 (una cepa que utiliza xilosa creada en Mascoma que contiene XliA y XKS1 en plásmidos) se incluyeron como controles. A las 24 horas, todas las cepas había consumido por completo la glucosa. M0407 y M0228 habían utilizado 30 y 25 g/l de xilosa respectivamente. Las nueve colonias M0407 "adaptadas" habían utilizado más de 44 g/l de xilosa. La cantidad más alta de xilosa consumida fue de 48 g/l. Esta cepa se denominó M0509.

La secuenciación de ADNr 18S se utilizó para confirmar la cepa M0509 como *Saccharomyces cerevisiae* (Kurtzman CP y Robnett, CME, FEMS Yeast Research 3 (2003) 417-432). Un fragmento de 1774 pb que abarcaba el ADNr 18S se amplificó del ADN genómico de M0509 y se solicitó su secuenciación. La secuencia de 1753 pb del ADNr 18S de M0509 presentó una coincidencia del 100 % con la secuencia del NCBI de 18S de *S. cerevisiae* (registro de nucleótido n.º Z75578).

Dado que la cepa M0509 se obtuvo cultivando M0407 en un quimiostato durante cuatro semanas, la duración de la separación del cultivo de las dos cepas proporciona un medio para evaluar la estabilidad de las modificaciones modificadas por ingeniería genética. La comparación de la secuencia de ADN de M0407 y M0509 en los locus *GRE3*, *RKI1*, *RPE1*, *TAL1* y *TKL1* no mostró cambios. Esto sugiere que las modificaciones genéticas en estos locus son genéticamente estables, al menos en las condiciones de cultivo utilizadas.

Se utilizó análisis de PCR en tiempo real para estimar el número de copias de integraciones del vector *XliA/XKS1*. M0407 tenía aproximadamente 10 copias del vector, mientras que M0509 tenía aproximadamente 20 copias. Esto sugiere que el número de copias del vector *XliA/XKS1* puede aumentarse ampliando el cultivo en medios con xilosa.

Para evaluar adicionalmente la estabilidad de las integraciones de *XliA/XKS1*, M0509 se cultivó durante ~50 generaciones en medios líquidos con glucosa o xilosa como la única fuente de carbono. Después de 50 generaciones, se aisló una colonia individual de cada cultivo y se cuantificó el número de integraciones de *XliA/XKS1* y se comparó con la reserva de refrigerador M0509 original. La colonia aislada del cultivo con xilosa tenía ~20 copias de *XliA/XKS1*, las mismas que la reserva de refrigerador. La colonia cultivada con glucosa exhibió un número de copias ligeramente disminuido, ~16.

La ligera disminución en el número de copias de *XliA/XKS1* de la colonia con glucosa genera la cuestión del

comportamiento de la cepa. Para abordar parcialmente esta cuestión, el consumo de xilosa se comparó entre el aislado con xilosa, el aislado con glucosa y la reserva de refrigerador. Todos los aislados propagados del refrigerador y con xilosa utilizaron xilosa en 24 horas y produjeron las mismas cantidades de etanol, pero la cepa propagada con glucosa consumió únicamente como mucho la mitad de xilosa. Figura 20.

Ejemplo 13: selección de una cepa termotolerante, fuerte, que utiliza xilosa

M1105 es capaz de realizar la fermentación a temperaturas por encima de 40 °C en presencia de acetato 8 g/l. M1105 se construyó en un fondo de M0509 y es por lo tanto una cepa industrialmente fuerte capaz de convertir tanto glucosa como xilosa en etanol.

M1105 se aisló siguiendo cuatro rondas de selección/adaptación en un citostato como se indica en la Figura 41 y como se describe a continuación. La temperatura se elevó de 38 °C a 41 °C durante el curso del experimento. M1017 (designación de depósito ATCC _____, depositada el 23 de noviembre de 2009) se aisló de este primer proceso de criostato y se confirmó posteriormente mediante PCR del locus GRE3 que era un descendiente de M0509. M1017 se utilizó para inocular un segundo proceso citostático utilizando medios YMX (base de nitrógeno de levadura, xilosa 2 g/l) a 41 °C. M1046 se aisló de este segundo proceso citostático. A 42 °C en YPX50, M1046 creció lentamente incluso con un tiempo de generación 36 % más corto que el de M1017. M1080 se aisló de un citostato inoculado con M1046 y medios YMX a 40 °C. M1080 creció con una tasa de crecimiento específico de 0,22 h⁻¹ en YMX a 40 °C. M1105 se aisló de M1080 basándose en la selección del citostato utilizando medio YPD2X10+acetato (glucosa 2 g/l, xilosa 10 g/l, acetato 8 g/l, pH 5,4) a 39 °C.

M1105 creció 10-20 % más rápido que M0509 en medio rico a 35 °C. Además, M1105 tenía una tolerancia al acetato aumentada ya que la cepa podía crecer más rápidamente que sus cepas ancestrales en presencia de acetato. Figura 21. Aunque las cepas precursoras requirieron glucosa para la tolerancia a acetato a altas temperaturas, M1105 no requirió glucosa o componentes de medio complejo para crecer en presencia de acetato 7 g/l a pH 5,4.

Para ensayar el comportamiento de fermentación, M1105 se inoculó a aproximadamente 0,7 g/l DCW en MS419 al 18 % utilizando Zoomerasa/g de materia prima 3,8 mg a 40 °C. M1105 produjo etanol 3,55 % (p/v) a las 168 horas. La evolución se presenta en la Figura 22 junto con un proceso similar realizado con M1088 (descrito más adelante) para comparación. Un proceso similar utilizando solo 0,15 g/l DCW para inóculo dio como resultado etanol 2,9 % (p/v) y alguna acumulación de azúcar durante el experimento. Figura 23.

Ejemplo 14: Adaptación de una cepa termotolerante, fuerte, que utiliza xilosa

M1254 puede realizar fermentación a temperaturas por encima de 40 °C en presencia de 12 g/l de acetato, mostrando una fuerza aumentada con respecto a la cepa termotolerante M1105.

M1254 se aisló después de tres rondas de selección/adaptación en un citostato como se indica en la Tabla 11 y en la Figura 42 y como se describe a continuación. El primer proceso en citostato se inoculó con M1105. Se utilizó medio YMX (base de nitrógeno de levadura sin aminoácidos, 20 g/l de xilosa) más 8 g/l de acetato a pH 5,5 y a 40 °C. M1155 se aisló de este primer proceso en citostato y se utilizó para inocular un segundo citostato que contenía medio YPD (extracto de levadura, peptona, 20 g/l de glucosa) más 12 g/l de acetato, a pH 5,4 y a 41 °C. M1202 se aisló de este segundo proceso de citostato. M1254 se aisló de un tercer proceso de citostato inoculado tanto con M1155 como con M1202 en medio de base de nitrógeno de levadura sin aminoácidos + hidrolizado de MS419 equivalente a 5 % de sólidos, a pH 4,8, a 39 °C.

Tabla 11: Condiciones progresivas para generar M1254 a partir de M1105.

Cepa(s) precursor(a)s	Condición progresiva	Nueva Cepa
M1105	Xilosa mínima + acetato 8 g/l, pH 5,5, 40 °C	M1155
M1155	Complejo de Glucosa + acetato 12 p/l, pH 5,4, 41 °C	M1202
M1155+M1202	Hidrolizado de MS419 equivalente a 5 % de sólidos +base de nitrógeno de levadura sin aminoácidos, pH 4,8, 39 °C	M1254

M1254 crece 7,3 ± 0,9 % más rápido que M1202 y 17 ± 2,0 % más rápido que M1155 en hidrolizado de MS419 equivalente a 5 % de sólidos que es la condición en la que se seleccionó la cepa M1254 Sin embargo, el medio de fermentación convencional limita el comportamiento de la fermentación. Por consiguiente, el uso de esta cepa sería con concentraciones de amonio más bajas, tales como fosfato diamónico (DAP) 1,1 g/l o inferior a DAP 3 g/l. La Figura 24 demuestra la tasa de fermentación más alta utilizando la concentración de DAP más baja. Las fermentaciones se realizaron utilizando MS149 al 18 %, celulasa externa 4 mg/g TS, 40 °C, inoculación de M1254 0,5 g/l DCW y pH 5,4. El pH se controló utilizando hidróxido de potasio 5 M, y con cada suministro de sólidos se proporcionaba carbonato de magnesio 1 g/l. Todas las enzimas se cargaron de frente, mientras que los sólidos se

suministraron a cinco momentos (0, 3, 6, 24 y 48 horas) en suministros del mismo tamaño de TS al 3,6 %.

MO1360 se creó a partir de M1254 utilizando las condiciones progresivas descritas en la Tabla 12 siguiente.

5 Tabla 12: Condiciones para generar M1360 a partir de M1254.

Cepa(s) precursora(s)	Condición progresiva	Nueva Cepa
M1254	Complejo con bajo contenido de xilosa + acetato 8 g/l, pH 5,4, 40 °C	M1339
M1339	Complejo de xilosa + mezcla de inhibidor sintético (incluyendo acetato 8 g/l), pH 5,4, 40 °C	M1360

10 M1360, aunque aun sustancialmente inhibida por la mezcla de inhibidor sintético, crecía a 40 °C con un tiempo de generación de aproximadamente 5 horas. Figura 25. En un medio industrialmente relevante, M1360 puede generar más de 60 g/l de etanol a partir de glucosa junto con 5 g/l de peso seco de células en 48 horas a 40 °C comenzando solo con 60 mg/l de peso seco de células. Figura 26.

15 Se sabe que la actividad enzimática aumenta a medida que aumenta la temperatura, y por eso es deseable tener cepas de *S. cerevisiae* termotolerantes. La Figura 27 muestra tres SFS equivalentes con sólidos de madera dura pretratada, PHW, cargados al 18 %. Las reacciones realizadas a 40 °C muestran aproximadamente un 17 % más de etanol producido en comparación con la reacción de control realizada a 35 °C, cuando ambas reacciones se realizaron a la misma carga enzimática externa (4 mg/g). Este comportamiento aumentado representa un ahorro sustancial del proceso.

20 Ejemplo 15: expresión de celulasas en una cepa fuerte que utiliza xilosa

25 M1088 puede secretar tres enzimas celulolíticas distintas: β glucosidasa de *S. fibuligera* (BGLSf), celobiohidrolasa 2b de *C. lucknowense* (CBH2b C1) y celobiohidrolasa I de *T. emersonii* fusionada al dominio de unión de celulosa de celobiohidrolasa I de *T. reesei* (CBH1Te + CBDCBH1Tr). El genoma de M1088 también contiene genes que codifican polipéptidos que pueden proporcionar resistencia a los siguientes antibióticos: kanamicina, nourseotricina e higromicina B. El plásmido pMU624, que también está presente en M1088, contiene un gen que codifica un polipéptido que puede proporcionar resistencia a ampicilina. Las etapas que se utilizaron para generar M1088 y M0963 a partir de M0509 se resumen en la Tabla 13 siguiente.

30 Tabla 13: cepas utilizadas para generar las cepas M1088 y M0963

Cepa	Genotipo	Precursor	Descripción
M0509	<i>gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1+//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1- XKS::delta</i>		
M0539	<i>URA-3/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1+//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i>	M0509	Una sola copia del gen URA-3 genómico se delecionó y se reemplazó con un casete kanMX. El casete genico KanMX proporciona resistencia a kanamicina (un antibiótico aminoglucósido).
M0544	<i>ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1+//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1- XKS::delta</i>	M0539	La segunda copia del gen URA-3 genómico se delecionó y se reemplazó con un casete kanMX.
M0749	<i>ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1 +//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta (pMU782)fur1::nat</i>	M0544	Una sola copia del gen FUR-1 genómico se delecionó y reemplazó con un casete nat1 de <i>Streptomyces noursei</i> . El casete del gen nat1 proporciona resistencia al antibiótico nourseotricina/clonNAT (un antibiótico aminoglucósido).
M0867	<i>FUR-1/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1 +//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1- XKS::delta (pMU782)fur1::nat; [pMU624]</i>	M0749	El plásmido pMU624 se transformó en la cepa. pMU624 puede replicarse en <i>S. cerevisiae</i> (ori 2 micrómetros y URA-3) y <i>E. coli</i> (ori pBMR y gen de resistencia a ampicilina: antibiótico beta lactámico). pMU624 también lleva el gen CBH1 de

Cepa	Genotipo	Precursor	Descripción
			<i>T.emersonii</i> +CBDCBH1Tr regulado por el promotor y terminador ENO1.
M0759	<i>fur-1::hyg/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1RKI1+//loxP-PTPI-RKI1 RPE1+//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i> (pMU782) <i>fur1::nat</i> ; [pMU624]; (pMU1037) <i>fur1::hyg</i>	M0867	La segunda copia del gen FUR-1 genómico se deletó y se reemplazó con un casete <i>hygMX</i> . El casete del gen <i>hygMX</i> codifica una higromicina B fosfotransferasa que confiere resistencia a higromicina B (un antibiótico aminoglucósido).
M1088	<i>fur-1::hyg/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1 +//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i> [pMU624] (pMU1260) <i>delta::PGKprom-SfBGL-PGKterm, ENO1prom-TeCBH+TrCBD-ENO1term</i> (pMU1169) <i>delta::PGKprom-SfBGL-PGKterm, ENO1prom-CICBH2-ENO1term</i>	M0759	Dos casetes de integración distintos se transformaron en la cepa y se integraron múltiples copias en el genoma en el sitio <i>delta</i> . Un casete contenía los genes celulolíticos BGL de <i>S. fibuligeria</i> y CBH2b de <i>C. lucknowense</i> . El otro casete contenía los genes celulolíticos BGL de <i>S. fibuligeria</i> y CBH1 quimérica de <i>T. emersonii</i> .
M0963	<i>fur-1::hyg/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1+//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i>	M0759	El ADN lineal de los 4 plásmidos mostrados se transformó en M0759. 24 de las colonias resultantes se cultivaron después en pases durante una semana en medio YPD que contenían zeocina a bajo nivel (50 ug/ml) y se sometió a ensayo. La cepa resultante M0963 fue la mejor de las
	(pMU782) <i>fur1::nat</i> ; [pMU624]; (pMU1037) <i>fur1::hyg</i> ; (pMU755) <i>delta::ZeoMX, ENO1prom-TeCBH1w/TrCBD-ENO1term</i> ; (pMU809) <i>delta::ZeoMX, ENO1prom-C1CBH2b-ENO1term</i> ; (pMU663) <i>delta::ZeoMX, ENO1prom-CfEG-ENO1term</i> ; (pMU864) <i>delta::ZeoMX, ENO1prom-SfBGL-ENO1term</i>		encontradas en el ensayo con avicel.

Ejemplo 16: selección de una endoglucanasa para la expresión en una cepa fuerte que utiliza xilosa

5 Las endoglucanasas aumentan la actividad de las celobiohidrolasas, y por lo tanto, se investigó la capacidad de endoglucanasas de 5 familias para complementar la CBH1 y CBH2 previamente identificadas. Se seleccionaron cinco endoglucanasas de cinco familias y se clonaron y clonaron bajo el control del promotor/terminador ENO1 utilizando el plásmido de expresión pRDH122 como se muestra en la Tabla 14.

Tabla 14: endoglucanasas de 5 familias expresadas en *S. cerevisiae*.

Organismo y gen:	Domino del Módulo de Unión a Celulosa, MUC:	Plásmido de expresión:	Tamaño teórico de la enzima Da*
<i>Aspergillus kawachii</i> <i>ega</i>	MUC1 C terminal	pRDH145	55034,58
<i>Heterodera schachtii</i> <i>eng1</i>	MUC2 C terminal	pRDH146	43739,46
<i>Hypocrea jecorina</i> (anamorfo: <i>Trichoderma reesei</i>) <i>eg2</i>	MUC1 N terminal	pRDH147	44226,91
<i>Orpinomyces</i> sp.PC-2 <i>celB</i>	MUC10 2x C terminal	pRDH148	53103,40
<i>Irpex lacteus</i> <i>en1</i>	MUC1 N terminal	pRDH149	42357,15

Todos los plásmidos que expresaban las 5 nuevas celulasas de tipo EG2 se transformaron en Y294 (una cepa de

laboratorio) y en M0749 (una cepa fuerte que utiliza xilosa; descrita anteriormente) y los transformantes se confirmaron por PCR. La Figura 28 muestra diversas de las cepas M0749 que se aplicaron en placas con SC^{URA} que contenían 0,2 % de cualquiera de CMC o liquenina o β -glucano de cebada. Como puede observarse en la figura 28, la cepa de referencia M0749 produjo pequeñas zonas en las placas que contenían CMC. Las dos cepas basadas en pMU471 (EG de *Coptotermes formosanus*) y en pRDH147, produjeron zonas aclaradas muy buenas en todos los sustratos ensayados.

Junto con la cepa de referencia y una cepa que expresaba la EG de *Coptotermes formosanus* (pMU471), las cinco cepas que expresaban eg2 se ensayaron con respecto a la hidrólisis en avicel y en PASC mientras que las cepas que expresaban *cbh2* se ensayaron con respecto a la actividad en avicel. Las cepas se cultivaron en medio SC^{URA} de doble fuerza (YNB 3,4 g/l; grupo con pérdida de aminoácidos 3 g/l sin uracilo; sulfato de amonio 10 g/l, glucosa 20 g/l) que se tamponó a un pH de 6 (ácido succínico 20 g/l, NaOH 12 g/l, pH ajustado a 6 con NaOH). Durante 3 días crecieron cultivos en matraces Erlenmeyer de 125 ml a 30 °C. Se inocularon tres matraces de cada cepa. Después de la incubación, se extrajeron muestras para realizar análisis en gel y medir la actividad. Después de la centrifugación de las muestras, se extrajeron 12 μ l de cada una de ellas, se añadieron a 5 μ l de tampón de carga de proteína y se hirvieron durante 5 minutos. A continuación las muestras se cargaron en un SDS-PAGE al 10 % y se separaron, seguido por tinción de plata. Los resultados se muestran en la Figura 29. No todas las cepas produjeron bandas visibles en el intervalo de tamaño esperado. La EG de C.f. apareció como una banda de aproximadamente 55 kDa como anteriormente se observa pero la banda producida por M0749 parecía ser ligeramente más grande que la producida por Y294. No se observaron bandas para los productos de *eng1* de *H. schachtii*, *celB* de *Orpinomyces* o *en1* de *L. lacteus*. La EG2 de *H. jecorina* producida por Y294 y M0749 fue visible como bandas de ~57 kDa. El aumento de peso comparado con el tamaño previsto de 44 kDa puede representar hiperglicosilación. La EGA de *A. kawachii* producida por Y294 fue visible como una banda de ~42 kDa. Sin embargo, la EGA de *A. kawachii* producida por M0749 fue claramente visible como una banda de ~120 kDa. El peso extra puede significar hiperglicosilación.

Todas las cepas se ensayaron con respecto a su actividad utilizando el método de conversión de avicel de alto rendimiento como se ha descrito anteriormente. Las cepas que expresaban endoglucanasas también se ensayaron con respecto a su actividad en PASC. El DNS utilizado para el procedimiento del ensayo, contenía fenol que, de acuerdo con la bibliografía, produce mayor sensibilidad. En la Figura 30 pueden observarse datos de actividad.

La cepa M0749, que expresaba la *eg2* de *H.j.* (pRDH147), produjo los niveles más altos de actividad secretada medidos en PASC o avicel de las EG2 sometidas a ensayo. La actividad de esta enzima fue mayor en PASC y en avicel que la EG de C.f. (pMU471). La EGA de A.k. sintética (pRDH145) también produjo actividad apreciable en ambos sustratos. Parece que este producto ha sido producido a niveles más altos en M0749 que en Y294 y rindieron mayor actividad que la EG de C.f. en avicel y PASC cuando se producía en esta cepa.

Ejemplo 17: expresión de una endogluconasa en levadura fuerte que utiliza xilosa

Se crearon varias cepas para ensayar el impacto de la coexpresión de EG2Tr con las CBH en un fondo de cepa fuerte que utilizaba xilosa. M1088 se transformó con una construcción para integrar la EG2 Tr en el locus de ADN_r utilizando el gen Sh-ble como marcador (pMU1409). Se realizó una transformación similar, pero integrando la CBH1 de Te con el CBD de Tr para aumentar el número de copias de ese gen. Se cultivaron 43 transformantes de ambas transformaciones junto con cultivos M1088 por duplicado en YPD que contenía zeocina 20 μ g/ml y se realizó el ensayo con avicel. La Figura 31 muestra los resultados de estos ensayos. Los datos muestran que una gran proporción de cepas transformadas con la construcción EG2Tr tenían una capacidad de conversión de avicel significativamente aumentada, mientras que los transformantes con copias adicionales de CBH1Te con CBDTr solo tenían mejoras insignificantes en hidrólisis con avicel.

De las cepas ensayadas, se seleccionaron los 9 candidatos superiores y sembraron en estrías para obtener colonias individuales. Estas colonias individuales se cultivaron después en YPD con 2 transferencias que correspondían a un total de 18 generaciones. La transferencia final (los datos de los pases se observan en la Figura 32) se comparó con el primer cultivo en YPD (los datos originales se observan en la Figura 32). Los datos confirman que había un aumento de ~50 % en cuanto a la capacidad del sobrenadante de levadura para convertir avicel cuando se sobreexpresaba la EG2 de Tr.

Además, la cepa M1403, que contenía los genes heterólogos que codifican la BGL de *S. fibuligera* (BGLSf), la celobiohidrolasa 2b de *C. lucknowense* (CBH2BC1), la celobiohidrolasa I de *T. emersonii* fusionada con el dominio de unión a celulosa de la celobiohidrolasa I de *T. reesei* (CBH1Te + CBD_{CBH1}Tr) y la *eng1* de *Heterodera schachtii*, se produjo en el fondo M1254. La cepa M1284, que contenía los genes heterólogos que codifican las mismas cuatro celulasas, se produjo en el fondo M0509. Las cepas M1284 y M1403 se describen con más detalle en la Tabla 15.

Tabla 15: cepas de levaduras que expresan endogluconasa.

Cepa	Genotipo	Precursor	Descripción
M1403	(pMU1339) delta::MET3prom-SfBGL-PGKterm, ENO1prom-TeCBH1+TrCBD-ENO1term (pMU1260) delta::PGKprom-SfBGL-PGKterm, ENO1prom-TeCBH+TrCBD-ENO1term (pMU1169) delta::PGKprom-SfBGL-PGKterm, ENO1prom-CICBH2-ENO1term (pMU1409) rDNA::ZeoMX, ENO1prom-HjEG2-ENO1term	M1254	Casets de ADN lineales creados por digestiones de restricción de plásmidos se integraron en múltiples copias en el genoma en los sitios Ty1 delta y en sitios de ADNr.
M0991	gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1 +//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PTPI-XKS LEU2/leu2D::hph	M0509	Una sola copia del gen LEU-2 genómico se delecionó y se reemplazó con un casete hygMX.
M0992	gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1 +//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PTPI-XKS leu2D::hph/ leu2D::nat	M0991	La segunda copia del gen LEU-2 genómico se delecionó y reemplazó con un casete nat1 de <i>Streptomyces noursei</i> .
M1162	gre3::loxP/gre3::loxP TAL1+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1+//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI xylA PTPI-XKS leu2D::hph/ leu2D::nat (pMU1379) delta::leu2-19, ENO1prom-TeCBH+TrCBD-ENO1 term	M0992	Un casete de ADN lineal creado por digestiones de restricción del plásmido pMU1379 se integró en múltiples copias en el genoma en los sitios Ty1 delta.
M1284	gre3::loxP/gre3::loxP TALI+//loxP-PTPI-TAL1 RKI1+//loxP-PTPI-RKI1 RPE1+//loxP-PTPI-RPE1 TKL+//loxP-PTPI-TKL delta::PTPI-xylA PTPI-XKS leu2D::hph/ leu2D::nat (pMU1379) delta::leu2-19, ENO1prom-TeCBH+TrCBD-ENO1term (pMU1169) delta::PGKprom-SfBGL-PGKterm, ENO1prom-CICBH2-ENO1term (pMU1409)rDNA::ZeoMX, ENO1prom-HjEG2-ENO1term	M1162	Casets de ADN lineales creados por digestiones de restricción de los plásmidos pMLT1169 y pMU1409 se integraron en múltiples copias en el genoma en los sitios Ty1 delta y en sitios de ADNr.

Ejemplo 18: Conversión de sustratos lignocelulósicos mediante cepas de levadura CBP

- 5 La expresión de celulasas en levaduras, particularmente CBH1 (CBH1 de *T. emersonii* con el CBD de *T. reesei* unido), CBH2 (CBH2b de *C. lucknowense*), EG2 (EG2 de *T. reesei*) y BGL (BGL de *S. fibuligera*) reduce drásticamente la necesidad de añadir externamente enzimas durante la conversión enzimática de lignocelulosa a etanol. Para ensayar el efecto de sobreexpresión de estas enzimas, se construyeron diversas cepas y se ensayaron en diversos sustratos.
- 10 La Figura 33 presenta datos de una fermentación con BPC (bioprocesamiento consolidado) en lodo de papel mediante una cepa hospedadora de *S. cerevisiae* termotolerante modificada por ingeniería genética (cepa precursora M1254, derivado celololítico M1403). Los datos para M1254 en solitario demuestran que se requiere la adición de celulasa (es decir, zoomerasa) para la producción de etanol a partir de lodo de papel. Los datos para M1430 donde no se añadía celulasa externa (cuadrados color naranja), demuestran que esta cepa puede convertir una fracción sustancial (~80 %) del sustrato “convertible” debido a sus celulasas expresadas. Las fermentaciones con celulasa externa adicional añadida a la cepa M1403, demuestran el extremo potencial de conversión enzimática para el sustrato de lodo de papel. Inspecciones visuales demuestran que la cepa sin BPC no pudo licuar el sustrato, mientras que sí podía la cepa con BPC.
- 20 Además, la cepa M1179 con BPC, que expresaba CBH1, CBH2, EG2 y BGL, podía convertir lodo de papel a un mayor grado sin enzima celulasa añadida. Figura 34. En esta reacción, la cepa de control M0509, fabricó solamente una pequeña cantidad de etanol durante esta reacción. Los datos también muestran que M1179 puede convertir este material cuando se carga a una densidad celular más baja (1 g/l) a diferencia de cuando se carga una densidad celular más alta (10 g/l) utilizada en otras reacciones. Esto implica que la cepa puede cultivarse y producir celulasa durante los experimentos de fermentación.
- 25 La madera dura pretratada (PHW) también puede convertirse por cepas mediante BPC. La Figura 35 muestra el efecto del uso de una cepa que expresa celulasa (M0963) en comparación con una cepa de control que no expresa celulasa (M0509) durante la fermentación de PHW. La comparación demuestra que la cepa con BPC puede obtener la misma producción de etanol a partir de PHW cuando solamente se cargan 2 mg/g de enzima externa en comparación con cuando se cargan 4 mg/g de M0509 en el proceso. Esta doble reducción en cuanto a la necesidad de enzima externa representa un gran potencial de reducción de costes en el proceso.
- 30 Las cepas de BPC pueden producir también altas cantidades de etanol a partir de PHW. La Figura 36 muestra que
- 35

una fermentación de sólidos lavados al 30 % puede generar cantidades de etanol de hasta aproximadamente 70 g/l con carga mínima externa de enzima de 4 mg/g y un inóculo celular relativamente bajo (2 g/l). La capacidad del cultivo de densidad celular baja para alcanzar eventualmente y pasar el cultivo de densidad celular alta indica que la cepa crece y continúa elaborando enzima durante la fermentación.

5 Además de la PHW, el forraje de maíz se ha implicado como un buen sustrato para la conversión a etanol mediante una sacarificación enzimática. La Figura 37 demuestra que el forraje de maíz pretratado puede convertirse bien mediante cepas de levadura con BPC. En este experimento, la cepa con BPC pudo convertir aproximadamente el 82 % de lo que podía convertirse con una carga enzimática alta (15 FPU, o aproximadamente 20 mg/g). La cepa sin BPC fabricó aproximadamente 60 % de etanol en comparación con lo que pudo conseguir la cepa con CBP.

Ejemplo 19: comparación de celulasas CBH1

15 Para proporcionar datos adicionales sobre los niveles de expresión de diferentes enzimas CBH1, se cultivaron cepas seleccionadas en medio YPD y se ensayaron las actividades en MULac y Avicel. Se estudiaron los dos transformantes Y294 y M0749 y los resultados se muestran en la Figura 38.

Adicionalmente la invención se describe más adelante en las reivindicaciones.

20 LISTADO DE SECUENCIAS

<110> Mascoma Corporation

McBride, John

25 Brevnova, Elena

Ghandi, Chhayal

Mellon, Mark

Froehlich, Alan

Deleault, Kristen

30 Rajgharia, Vineet

Flatt, Jim

Van Zyl, Emile

Den Haan, Riaan

LaGrange, Danie

Rose, Shaunita

35 Penttila, Merja

Siika-Aho, Matti

Uusitalo, Jaana

Ilmen, Marja

Hau, Heidi

40 Rice, Charles

Villari, Jeff

Stonehouse, Emily

Gilbert, Alan

Keating, Jeffrey

45 Xu, Haowen

Willes, Deidre

Shikhare, Indraneel

Thorngren, Naomi

Warner, Anne

50 Murphy, Dan

<120> Levaduras que expresan celulasas para la sacarificación y fermentación simultáneas utilizando celulosa

<130> 2608.025PC01/PAJ/CLD

55

<140> A asignar

<141> Adjunta

<150> 61/116,981

60 <151> 21-11-2008

<160> 58

<170> PatentIn versión 3.3

65

<210> 1

ES 2 621 181 T3

<211> 1590
 <212> ADN
 <213> *Humicola grisea*

5 <400> 1

gaattcatga gaaccgctaa gttcgetacc ttggctgect tggttgcctc tgctgctgct 60
 caacaagcct gttccttgac tactgaacgt caccatcctt tgtcttgaa caagtgtact 120
 gctggtggtc aatgtcaaac tgtccaagcc tccatcactt tggactctaa ttggagatgg 180
 acccaccaag tctctggtag tactaactgt tacaccggtg ataagtggga cacttctatt 240
 tgtactgacg ctaagtcttg tgctcaaaat tgttgtggtg atggtgctga ttacacctcc 300
 acttatggta ttaccaccaa cgggtgactct ttgtccttga agttcgttac taaaggtcaa 360
 cattccacca acgtcgggtc tagaacctac ttaatggacg gtgaagaaa gtaccaaacc 420
 ttogaattgt tgggtaaatga atttaccttc gatgtcgatg tgtctaacad cggttgtggt 480
 ttgaacgggtg ctttataactt cgtttctatg gacgccgacg gtggtttgtc tcgttaccca 540
 ggtaataagg ctggtgccaa gtatggtacc gggtactgtg atgctcaatg cccaagagac 600
 attaagttca tcaacggtga agctaacatt gaaggttga ctggttctac caacgacca 660
 aacgctggcg ccggtagata cgggtacgtt tgttccgaaa tggacatttg ggaagccaac 720
 aacatggcta ctgcttttac tccacacca tgtaccatca ttggtcaatc cagatgtgaa 780
 ggtgactcct gtggcggtag ctactccaac gaaagatag ctggtggttg tgatccagac 840
 ggttgtgact tcaactccta cagacaaggt aacaagactt tctatggtaa gggtagact 900
 gtcgatacca ccaagaagat caccgtcgtc acccaattct tgaaggacgc taacggtgat 960
 ttaggtgaaa ttaaaagatt ctacgtccaa gatggtaaga tcatccaaa ctctgaatct 1020
 accattccag gtgttgaagg taattccatc actcaagact ggtgtgacag acaaaaggtt 1080
 gccttcgggtg atattgacga cttcaacaga aaggggtgta tgaagcaaat gggtaaggct 1140
 ttggccggtc caatggtctt gggtatgtct atttgggacg atcacgcttc caacatgttg 1200
 tggttggact ccacctccc agttgatgct gctggtaagc caggtgccga aagaggtgct 1260
 tgtccaacta cttccgggtg cccagctgaa gttgaagccg aagctccaaa ttctaactgt 1320
 gtcttctcta acatcagatt cgggtccaatc gggtccacag tcgctggttt gccaggtgct 1380
 ggtaatgggtg gtaataacgg tggtaaccca ccaccaccaa ccaactaccac ttcttctgcc 1440
 ccagctacta ccaccaccgc ttctgctggt ccaaaggctg gtagatggca acaatgtggt 1500
 ggtattggtt tcaccggtcc aaccfaatgt gaagaacat acatctgtac caagttgaac 1560
 gactggtact ctcaatggtt ataactcgag 1590

10 <210> 2
 <211> 1383
 <212> ADN
 <213> *Thermoascus aurantiacus*

ES 2 621 181 T3

<400> 2

```

gaattcatgt accaaagagc tctattgttc tcttcttct tggccgccc tagagctcat      60
gaagccggta ctgtcaccgc cgaaaaccac ccatccttga cttggcaaca atgttcctct      120
gggtggttctt gtactactca aaacgggaag gttgttattg acgctaactg gagatggggt      180
cacactacct cgggttacac caactgttac actggtaaca cttgggatac ttccatctgt      240
ccagacgacg ttacctgtgc tcaaaactgt gctttggaag gtgctgacta ctccggtact      300

tacgggtgca ctacctctgg caacgcggtg agattgaact tcgtcaccca atcttctggt      360
aagaacatcg gttctagatt gtacttggtg caagacgata ctacttacca aatcttcaag      420
ttgttgggtc aagagttcac ttctgacggt gatgtttcca acttgccttg tggtttgaac      480
gggtgctttgt acttcggtgc tatggaogcc gacggtaact tatccaagta cccaggtaac      540
aaggccggtg ccaagtacgg taccggttac tgtgattctc aatgtccaag agacctaaaa      600
ttcattaacg gtcaagctaa cgtcgaaggt tggcaaccat ctgctaacga tccaaacgcc      660
gggtgcggta atcacggttc ctctgtgct gaaatggacg tttgggaagc taactctatc      720
tccaccgccc tcaactccaca tccatgtgat accccaggtc aaacctatgt tcaaggtgat      780
gattgtgggtg gtacctactc ttccactaga tacgctggta cctgtgacac cgacggttgt      840
gatttcaacc cataccaacc aggtaaccac tctttctacg gtccaggtaa gattgtcgat      900
acttcttcta agttcactgt tgtcactcaa ttcattaccg acgatggtag cccatctggt      960
accctaactg aaattaagag attctacgtc caaaacggta aagtcattcc acaatccgaa     1020
agcaccattt ccggtgttac cggtaactcc atcaccactg aatactgtac cgctcaaaag     1080
gccgcctttg acaacaccgg tttcttcacc catgggtggtt tgcaaaagat ttctcaagcc     1140
ttggtcaag gtatggtttt ggtcatgtcc ttgtgggatg accacgctgc taacatggtg     1200
tggttggatt ctacttacc aactgacgct gatccagaca cccaggtgt tgctagaggt     1260
acttgtccaa ccacttctgg tgttccagct gacgtcgaat ctcaaaaccc taactcttac     1320
gttatctact ctaacatcaa ggtgggtcca attaactcca ccttactgc taactaactc     1380

gag                                                                           1383

```

5 <210> 3
 <211> 1380
 <212> ADN
 <213> *Talaromyces emersonii*

10 <400> 3

ES 2 621 181 T3

gaattcatgc taagaagagc tttactattg agctcttctg ctatcttggc cgttaaggct 60
 caacaagccg gtaccgctac tgctgaaaac caccctccat tgacctggca agaatgtacc 120
 gctccagggt cttgtaccac ccaaaacggg gctgtcgtct tggacgctaa ctggagatgg 180
 gtccacgacg tcaacgggta cactaactgt tacaccggta acacctggga cccaacttac 240
 tgtccagacg acgaaacttg cgctcaaaac tgtgccttgg acggtgctga ctacgaagg 300
 acttacggtg ttacctctc tggttcttcc ttgaagttga acttcgtcac tggttctaac 360
 gtcggtcca gattgtatth gtgcaagat gactccactt accaaatctt caagttgttg 420
 aacagagaat tttctttcga cgtcgtatgt tccaacttgc cttgtggttt gaacggtgct 480
 ctatacttgc ttgctatgga cgctgatggg ggtgtttcca agtacccaaa caacaaggct 540

 ggtgccaaat acggtactgg ttactgtgac tctcaatgtc cacgtgactt gaagtttatt 600
 gatggtgaag ctaatgtcga aggttgga ccatcttcta acaacgctaa cactggcacc 660
 ggtgaccacg gttcttctg tgccgaaatg gacgtttggg aagccaactc catttccaac 720
 gccgtcactc cacaccatg tgacactcca ggtcaaacta tgtgttccgg cgatgactgt 780
 ggtggtactt actctaacga tagatacgtt ggtacctgtg atccagacgg ttgcgacttc 840
 aatccataca gaatgggtaa cacttccttt tacgggtccag gcaagatcat cgacactact 900
 aagccattca ctgttgtcac ccaattcttg accgacgatg gtactgatac cgggtactttg 960
 tccgaaatca agagattcta catccaaaac tctaacttca tcccacaacc aaattccgac 1020
 atctctggtg tcaactgtaa ctccattacc accgaatttt gtaccgcca aaagcaagct 1080
 ttcggtgaca ccgacgactt ctctcaacac ggtggtttgg ctaagatggg tgctgctatg 1140
 caacaaggta tggttttggg catgtctttg tgggacgact acgctgctca aatggtgtgg 1200
 ttggactccg attacccaac cgatgccgac ccaaccacc ctggtatcgc tagaggtacc 1260
 tgtccaactg actctggtgt tccatctgac gtcgaatccc aatctccaaa ctctacgtc 1320
 acttactcca acattaaatt cgggtccaatc aactccactt tcaactgcttc ttaactcgag 1380

<210> 4
 <211> 1392
 <212> ADN
 <213> *Talaromyces emersonii*

 <400> 4

5

ES 2 621 181 T3

gaattcatgc gtaacttggt ggccttggt ccagccgctt tggtggtgg tgctgccgaa 60
 gctcaacaat ccttggtggg tcaatgoggt ggctcctcct ggactggtgc aacttctctg 120
 gccgctggtg ccacctgttc caccattaac ccatactacg ctcaatgtgt tccagccact 180
 gccactccaa ctaccttgac taccaccact aagccaacct ccaccggtgg tgctgctcca 240
 accactccac caccaactac taccggtact accacctctc cagtctcac cagacctgcc 300
 tccgctccg gtaatccatt cgaaggttat caattgtacg ctaaccctta ctacgcttct 360
 gaagtcattt ccttggtat cccatctttg agctccgagt tggccccaaa ggcctccgaa 420
 gttgctaagg tcccttcatt tgtctgggta gatcaagctg ccaaggttcc atctatgggt 480
 gattacttga aggatattca atctcaaac gctgctggtg ctgatccacc aatgccggt 540
 attttctgtt tttacgattt gccagataga gactgtgccg ccgctgcttc taacggtgaa 600
 ttttctatcg ccaacaacgg tgtcgttta tacaacaat atatcgattc cattagagaa 660
 caattaacca ctactcoga cgtccatacc atcttgggta tcgaaccaga ctctttggct 720
 aacggtgtca ctaacttgaa cgttccaaaa tgtgctaacg ctcaagatgc ttacttggaa 780
 tgtatcaact acgctattac ccaattggac ttgccaaacg ttgctatgta cttggacgct 840

 ggtcacgccg gttggttggg ttggcaagcc aacttggccc cagctgctca attattegct 900
 tctggttaca agaacgcctc tccccagcc tctgtagag gtttggctac caacgtggct 960
 aactacaacg cctggtccat ttctagatgt ccatactaca ctcaaggtga cgctaactgt 1020
 gatgaagaag attacgttaa cgctttgggt ccattgttcc aagaacaagg tttcccagct 1080
 tacttcatca tcgacacttc ccgtaacggt gtcagaccaa ctaagcaatc tcaatgggg 1140
 gactggtgta acgttattgg taccggtttc ggtgtagac caaccaccga cactggtaac 1200
 ccattggaag acgctttcgt ttgggtcaag ccaggtggtg aatccgacgg tacctccaac 1260
 actactagcc cacgttacga ttaccactgt ggtttgtctg acgctttgca accagctcca 1320
 gaagctggta cctggttcca agcctacttc gaacaattgt tgactaacgc caaccattg 1380
 ttctaactcg ag 1392

<210> 5
 <211> 1449
 <212> ADN
 <213> *Chrysosporium lucknowense*

5

<400> 5

ES 2 621 181 T3

atggccaaga agttgttcat tacogctgcc ttagctgcog cagtgtttgc tgcaccagtg 60
 atcgaagaga gacaaaattg cggagccgtc tggacacagt gggaggcaa cggctggcaa 120
 ggcccaacat gttgtgcttc tggetcaacg tgcgtggcac agaacgagtg gtattcccag 180
 tgccttcaa actcccaggt gacttcttca acaacccccca gctcaacgtc tacttcacag 240
 agatccacaa gtacctcttc tagcacaacc agaagtggct catcctcctc tagcagtacg 300
 acccctccac ccgtatcaag tcctgtcacg agtatccctg gggagcaac ctcaacagcc 360
 agttattccg gcaatccttt ctctggagtg agattatttg caaacgacta ttatagatca 420
 gaggttcaca acottgcaat tccttctatg acgggaacc ctagccgcaa ggcttcogcc 480
 gtagcagaag tccttagttt ccaatggctt gacagaaacg ttacaataga tacacttatg 540
 gtacagactt tatctcaggt tagagctttg aataaggccg gtgccaaacc accttatgct 600
 gcccaattag tagtctatga cttgccagat agagactgtg ctgccgcagc ttctaattgt 660
 gaattttcca tcgcaaatgg cggagctgca aactatagat catacattga tgcaataaga 720
 aaacacatca ttgagtattc tgatattaga ataatccttg tgattgaacc agactccatg 780
 gctaatatgg ttaccaacat gaatgtagcc aagtgttcta acgcagcttc cacataccat 840
 gagctaaccg tatatgcatt aaaacaactg aatctaccta acgttgctat gtacttagat 900
 gccggtcatg ccggatgggt gggctggcct gcaaatatcc aaccgcagc tgaattgttc 960
 gctggaatct acaacgagc cggaaagccc gctgccgta gaggcttagc cacaaatgtt 1020
 gcaaattaca acgcttggtc aattgctagt gcccttctt atacctcacc aaatcctaac 1080
 tacgatgaga aacattacat agaagcattt tccccattgt taaactcogc tggattccct 1140
 gccagattca tcgtggatac cggtagaaac ggcaacaac caactggaca acaacaatgg 1200
 ggagattggt gtaacgtcaa gggaaaccggc ttccggcgtca ggcctacggc aaacaccgga 1260
 cagcagctag tcgacgcttt tgtatgggtt aagccaggtg gcgaaagtga cggacaagt 1320
 gacacgagtg ctgcaagata cgattaccac tgtggtctgt ccgacgcttt acagcccgcc 1380
 cccgaggctg gacaatgggt ccaggcttat tttgaacaat tgtaacgaa cgcaaatcca 1440
 ccattctaa 1449

<210> 6
 <211> 1551
 <212> ADN
 <213> Secuencia artificial

5

<220>
 <223> cbh1 de T.emersonni con CBD

10

<400> 6

ES 2 621 181 T3

atgctaagaa gagctttact attgagctct tctgctatct tggccgtaa ggctcaacaa 60
 gccggtaccg ctactgctga aaaccacocct ccattgacot ggcaagaatg taccgctcca 120
 ggttcttgta ccaccocaaaa cgggtgctgto gtcttggaog ctaactggag atgggtccac 180
 gacgtcaacg gttacactaa ctgttacacc ggtaacacct gggacccaac ttactgtcca 240
 gacgacgaaa cttgcgctca aaactgtgcc ttggacggtg ctgactacga aggtacttac 300
 ggtgttacct cctctgggtc ttccttgaag ttgaacttcg tcaactgggtc taacgtcgggt 360
 tocagattgt atttgttgca agatgactcc acttaccaaa tottcaagtt gttgaacaga 420
 gaattttctt tcgacgtcga tgtgtocaaac ttgccttggtg gtttgaacgg tgctctatac 480
 ttcgttgcta tggacgctga tgggtggtgt tccaagtacc caaacaacaa ggctggtgcc 540
 aaatacggta ctggttactg tgactctcaa tgtccacgtg acttgaagtt tattgatggt 600
 gaagctaattg tcgaaggttg gcaaccatct tctaacaacg ctaaacactgg catcggtgac 660
 cacggttctt gctgtgccga aatggacgtt tgggaagcca actccatttc caacgccctc 720
 actccacacc catgtgacac tocaggtcaa actatgtgtt ccggcgatga ctgtggtggt 780
 acttactota acgatagata cgctggtacc tgtgatccag acggttgoga cttcaatcca 840
 tacagaatgg gtaacacttc cttttacgggt ccaggcaaga tcatcgacac tactaagcca 900
 ttcactggtg tcaccaatt cttgaccgac gatggtactg ataccggtac tttgtccgaa 960
 atcaagagat tctacatcca aaactctaac gtcateccac aaccaaattc cgacatctct 1020
 ggtgtcactg gtaactccat taccacogaa ttttgtaccg cccaaaagca agctttcgggt 1080
 gacacogacg acttctctca acacgggtggt ttggctaaga tgggtgctgc tatgcaacaa 1140
 ggtatggttt tggatcatgtc tttgtgggac gactacgctg ctcaaagtgt gtggttgac 1200
 tocgattacc caaccgatgc cgacccaacc acccctggtg togctagagg tacctgtcca 1260
 actgactctg gtgttccatc tgaogtoga tcccaatctc caaactccta cgtcacttac 1320
 tocaacatta aattcgggtc aatcaactcc actttcactg cttctaacco tccaggtggt 1380
 aacagaggta ctaccactac tcgtaggcca gctactacaa ctggttcttc cccaggccca 1440
 acccaatccc actacggtca atgtggtggt atcggttact ctggtccaac cgtctgtgct 1500
 tctggtacta cctgtcaagt tttaaaccca tactactctc aatgtttgta g 1551

<210> 7
 <211> 1608
 <212> ADN
 <213> *Trichoderma reesei*
 <400> 7

5

ES 2 621 181 T3

atggtctcct tcacctcct gctggccggc gttgccgcta tctctggtgt cctagcagcc 60
cctgccgcag aagttgaacc tgtcgcagtt gagaaacgtg aggccgaagc agaagctcaa 120
tccgcttgta ccctacaatc cgaaactcac ccaccattga cctggcaaaa gtgttctagc 180
ggtggaactt gtactcaaca aactggttct gttgttatcg acgctaactg gagatggaca 240
cacgccacta actctttctac caactgttac gacggtaaca cttggtcttc cactttatgt 300
ccagataacg aaacttgtgc taagaattgc tgtttggacg gtgccgcta cgcttctacc 360
tacggtgta ccacctccgg taactccttg tctattggtt tcgtcaactca atccgctcaa 420
aagaacgttg gtgctagatt gtacttgatg gcttctgaca ctacttatca agaatttact 480
ttgttgggta acgaattttc tttcgatggt gacgtttccc aattgccatg tggcttgaac 540
ggtgctttgt actttgtctc tatggatgct gacgggtggg tttctaagta cccaactaac 600
actgccggtg ctaagtacgg tactggttac tgtgattctc aatgtccacg tgacttgaag 660
ttcattaacg gtcaagccaa cgtcgaaggt tgggaaccat cctccaacaa cgctaacacc 720
ggtatcggtg gtcacggttc ctggtgttcc gaaatggaca tctgggaagc taacagtatt 780
tctgaagctt tgacaccaca cccatgcacc actgtcggtc aagaaatttg tgaagtgat 840
ggatgtggtg gaacctactc tgataacaga tacgggtgga cttgtgaccc agacggttgt 900
gactggaacc catacagatt gggtaacact tctttctatg gtccaggttc ttctttcacc 960
ttggatacca ccaagaagtt gactgttgtt acccaattcg aaacttctgg tgctatcaac 1020
agatactacg ttcaaaacgg tgtcaccttc caacaaccaa acgctgaatt gggttcttac 1080
tctggtaatg aattgaacga cgactactgt accgctgaag aagctgaatt tgggtggtcc 1140
tctttctccg acaaggggtg tttgacccaa ttcaagaagg ctacctccgg tggatgggtt 1200
ttggttatgt ccttgtggga tgattactac gcaaacatgt tatggttaga cagtacttac 1260
ccaactaacg aaacctctc tactccagggt gctgtcagag gttcctgttc tacctcttct 1320
ggtgttccag ctcaagttga atctcaatct ccaaacgcta aggtcacttt ctccaacatc 1380
aagttcggtc caatcggttc cactggtaat ccatctggtg gaaaccctcc aggtggtaac 1440
agaggtacta cactactcg taggccagct actacaactg gttcttcccc aggccaacc 1500
caatcccact acggtcaatg tgggtgtatc ggttactctg gtccaaccgt ctgtgcttct 1560
ggtactacct gtcaagtttt aaaccatac tactctcaat gtttgtaa 1608

<210> 8
<211> 1479
<212> ADN
<213> *Trichoderma reesei*

5

<400> 8

ES 2 621 181 T3

atggtctcct tcacctccct gctggccggc gttgccgcta tctctggtgt cctagcagcc 60
 cctgccgcag aagttgaacc tgtcgcagtt gagaaacgtg aggccgaagc agaagctgtc 120
 ccattagaag aaagacaagc ctgctcctct gtttggggtc aatgtggtgg tcaaaactgg 180
 tctggtocaa ctgtttgtgc ttccggttct acctgtgttt actccaacga ctactattcc 240
 caatgtttgc caggtgctgc ttctcttcc tottcaacta gagctgcttc tacaacttct 300
 agggctctccc caaccacttc cagatcctct tctgctactc caccaccagg ttctactacc 360
 actagagttc caccagtcgg ttccggctact gctacttact ctggtaaccc ttcctgctgg 420
 gttactccat gggctaacgc ttactacgct tctgaagttt cttctttggc tatcccatct 480
 ttgactgggtg ctatggctac cgctgctgct gctgtcgcca aagttccatc cttcatgtgg 540
 ttggacacct tggacaaaac tccattaatg gaacaaacct tggcagacat aaggactgct 600
 aacaagaacg gcggtaaacta cgctggtaaa tttgttgtgt acgacttgcc agacagagac 660
 tgtgctgctt tggcttccaa cgggtgaatac tccatcctgt acggtggtgt cgccaagtac 720
 aagaactaca ttgatccat tagacaaatc gttgtcgaat actctgacat cagaaccttg 780
 ttagtcatcg aaccagattc tttagccaat ttagtcacca acttgggtac tccaaagtgt 840
 gctaacgctc aatctgccta cttagaatgt atcaattatg cagttaccca attgaacttg 900
 ccaaacgttg ctatgtactt ggacgctggt cacgccggtt ggttggggtg gccagctaac 960
 caagaccag ccgctcaatt attcgccaac gtttacaaga atgcctcttc tcctagagcc 1020
 ttgcgtgggtt tggctactaa cgctgctaac tacaacggtt ggaacatcac ttctccacca 1080
 tcttacacco aaggtaacgc tgtttacaac gaaaagttgt acattcacgc tatcgggtcca 1140
 ttattggcta accatggttg gtctaacgcc ttcttcatca ccgaccaagg tagatccggt 1200
 aaacaaccaa ctggtcaaca acaatggggt gattggtgta acgtcatcgg tactggtttc 1260
 ggtatcagac catccgctaa cactggtgat tccttgttgg attccttctg ctgggttaag 1320
 ccagtggtg aatgtgatgg cacctctgat tcctctgctc caagattcga ttcccactgc 1380
 gccttgccag acgctttgca accagcocca caagctggtg catggttcca agcttacttt 1440
 gtccaattgt tgaccaacgc taacctatct ttcttgtaa 1479

<210> 9
 <211> 1618
 <212> ADN
 <213> *Chaetomium thermophilum*
 <400> 9

5

ES 2 621 181 T3

ttaattaaac aatgatgtac aagaaatTTG cagccctagc tgctttagtt gcaggagctt 60
 ccgctcaaca ggcatgttca ttgactgccg aaaatcatcc atccttaacg tggaagagat 120
 gcacgtcagg aggttcatgc tccactgtaa acggagctgt cacaatagat gcaaattgga 180
 gatggacca cactgtgtcc ggtagtacaa actgctacac cggtaatcaa tgggatacgt 240
 ctttgtgtac agatggaaag tcatgcgctc agacctgTTG cgtggatgga gcagactact 300
 cttctactta cggaatcacg acatcaggtg acagtcttaa tttgaaattc gtaaccaagc 360
 accagtacgg aacaaatgta ggctccagag tgtacttaat ggagaacgat accaaatac 420
 aatgttoga gttattaggc aatgagtTTA cctttgacgt agacgttagc aatttgggtt 480
 gcggattaaa cggcgccctt tacttcgtgt ctatggatgc tgacggaggt atgtcaaagt 540
 attctggtaa caaagccgga gcaaagtacg gtacaggTTA ttgtgacgct cagtgcccta 600
 gagatttGAA gtttatcaac ggagaagcca acgTTggtaa ctggacgcca agtactaacg 660
 acgcaaacgc tggattcggc agatacggta gttgTTgctc agaaatggac gtgtgggagg 720
 ccaataacat ggcaaccgct tttactcctc acccatgtac aactgttGGA caatctagat 780
 gtgaagccga cacgtgCGGT ggcacctaca gtagcgatag gtatgcagga gtatgtgatc 840
 ctgacggtg cGatTTcaat gcttatagac aaggagacaa aacgTTttat ggtaaaggta 900
 tgaccgtcga tactaacaag aagatgactg tggTTaccca gttccacaag aactcagctg 960
 gagtattgtc tgaaattaaa agattctacg tccaggatgg aaagattatt gctaattGCC 1020
 agagtaagat accaggtaac cctggaaata gtatcacaca ggaataactgt gacgctcaga 1080
 aggtagcttt tagcaacacc gatgacttca atagaaaggg tggaaTggct caaatgagta 1140
 aggctttagc cggTccaatg gtgTTgtgga tgtctgTTTg ggatgatcac tatgcaaaca 1200
 tgctttggct tgacagcacc taccctatcg accaagccgg agccccaggt gctgaaaggg 1260
 gtgcatgtcc aaccacgagt ggtgtgcccc cggagattga agctcaagtG cctaatagta 1320
 acgTTatctt ttccaatata agattcggac caatcggatc cactgttcca ggTTtgatg 1380
 gatctaatcc tggcaaccca acaaccacgg tagtccctcc agcttcaact tccacaagta 1440
 gaccaacaag ttcaacgtcc agtccagtgt ctactcctac cggacaacca ggaggctgta 1500
 ccactcagaa atggggTcaa tgcggtgGaa ttggctatac aggttgtacg aattgcgTTg 1560

 caggaaccac ttgtacacag ttaaaccctt ggtactcaca atgcctataa ggcgcgcc 1618

<210> 10
 <211> 969
 <212> ADN
 <213> *Coptotermes lacteus*

5

<400> 10

ES 2 621 181 T3

atgagatttc cttccatatt cacogctggt ttgttcgcag cctcaagtgc tttagcagaa 60
 tgtactaagg gtggatgtac taacaagaat ggatacatag ttcatagataa gcacgtcgggt 120
 gacatccaga atagagacac tttggaccct ccagacttag attatgaaaa ggacgtggga 180
 gtaaccgtgt ccggtggaac ccttagtcaa agattagtct caacttgga cggtaagaaa 240
 gtcgtgggaa gtagattgta tattgtggac gaagccgaag agaaatatca attattcaca 300
 tttgtcggta aggagttcac ctataccggt gatatgtccc agatccaatg tggaatcaat 360
 gccgcattat acacagtgga aatgcctgcc gctggaaaga cccctggagg tggttaagtat 420
 ggatatggat atttgtgatgc caactgcgtg gatggagatt gttgtatgga gttcgatata 480
 caagaagctt ctaacaaggc aatcgtttac accaccatt cctgtcaaag tcaaacttca 540
 ggttgcgata cctcaggatg cggttacaac ccttacagag acagtgggta caaggcattc 600
 tggggaacaa ctataaacgt aaaccagcct gtgacaattg taacacagtt tatcggttct 660
 ggtagttcct taactgaagt caaaagattg tgcgtgcaag gtggaaagac ottcctcca 720
 gccaaatcat taaccgacag ttattgtaat gccaacgact atagaagttt gagaactatg 780
 ggtgcatcca tggetagagg acacgttggt gtgttttctt tgtgggattc taatggtatg 840
 agttggatgg atggaggtaa cgcggctcct tgtacctcat ataattga atctttggaa 900
 tccagtcagc caaacttaa ggtcacatgg tcaaactgga aatacggaga gatcgattct 960
 ccttattaa 969

5 <210> 11
 <211> 1356
 <212> ADN
 <213> *Coptotermes formosanus*

10 <400> 11

atgagattcc cttccatttt cactgctggt ttgttcgcag cctcaagtgc tttagcagcc 60
 tatgactaca agacagtatt gaagaactcc ttgttgttct acgaagctca aagaagtgga 120
 aaattgcctg cagaccagaa ggtgacctgg agaaaagatt ccgcattaaa cgacaagga 180
 cagaagggag aggacttaac tggaggttat tacgacgccg gagactttgt gaagttcgggt 240
 tttccaatgg catacacagt taccgtgttg gcctgggggt tagtcgatta tgaatctgct 300
 tacagtaactg cgggtgcctt ggatgatggt agaaaggcct tgaaatgggg tacagattat 360

ES 2 621 181 T3

ttottgaaag cacataccgc tgccaatgag ttttacggac aggtgggtca gggagatgtg 420
 gatcatgctt actggggacg toctgaggac atgactatgt ctagaccagc ttacaagatc 480
 gatacatcaa aacctggtag tgacttagct gcagaaacag cagccgcttt agcagcaacc 540
 gcaatagctt acaagtcagc cgattctacc tacagtaaca acttaattac tcatgcaaag 600
 cagttgttcg attttgcaaa caattataga ggaaagtact ctgatagtat taccgatgcc 660
 aagaatttct atgcatccgg tgattataag gacgaattag tatgggctgc agcctggttg 720
 tatagagcta caaatgataa cacttactta accaaagccg aatcattgta taatgaattt 780
 ggtttaggat cttggaacgg tgcattcaat tgggataaca agatatccgg agttcaggtc 840
 ttattagcca aattgacatc caaacaagca tacaagata aagttoaggg ttatgttgat 900
 tacttagtct cctctcaaaa gaaaactcca aagggattgg tctatattga ccaatgggga 960
 acottaagac acgcagctaa tagtgccttg atcgctttac aggcogctga tttgggtata 1020
 aacgctgcta gttatagaca atacgcaaag aagcaaattg attatgcctt aggtgacgga 1080
 ggtcgttcct acgtggtcgg attoggaact aaccctccag taagacctca tcatagatcc 1140
 agttcctgtc ctgacgcaac agccgcttgc gactggaata cttacaactc tgccggacca 1200
 aatgccacg tcttgaccgg agccttagta ggtggaccag attccaacga tagttacaca 1260
 gattcacggt ctgattatat cagtaacgaa gtcgctactg attacaatgc cggtttccaa 1320
 tctgcagttg ctggtttggt gaaagccgga gtataa 1356

<210> 12
 <211> 1356
 <212> ADN
 <213> *Nasutitermes takasagoensis*
 <400> 12

5

ES 2 621 181 T3

atgagatttc catctatddd cactgccgtc ttattdgcag cctccagtgc attagcagcc 60
 tatgattata aacaagtdtt gagagattcc ttattdgttct acgaagctca gagaagcggd 120
 agattaccag cagaccagaa ggtcactdgg agaaaagatt cagccttgaa tgatcagggd 180
 gatcaaggtc aagacttaac cggaggttat ttdgacgccg gtgatttdgt gaaatttdgt 240
 ttdccaatgg catatactgc taccgtcttg gcctggggtd taatcgattt tgaggcagga 300
 tacagtdccg ctggdgcctt ggatgacggd agaaaagcag taaagtdggc aactgattac 360
 ttdataaagg cccacacttc acagaatgag ttdtacggac aagtdcgtca gggdgcagct 420
 gatcaagcct tctggggagc tctgaagat atgacctdgg ctgacctcgc ctacaagatt 480
 gacaccagca gaccaggtag tgacttagcg ggtgaaaccg cagcggcatt ggcagctgcc 540
 agtatcgtgt ttagaaatgt tgatggtaca tactctaaca acttacttac tcatgccaga 600
 caattattdg acttdgcaaa taactacaga ggaaaact cagattccat aaccgacgct 660
 agaaacttdt acgccagtgc agattaccgt gacgaattgg ttdgggctgc cgcattggtg 720
 tacagagcta caaatgacaa cacttacttg aataccgcag aatcctdgtg tgatgaattt 780
 ggattgcaga actggggdgg aggdtdaaac tgggattcaa aggtgtctgg tgtccaggtc 840
 ttdgttagcaa aattgaccaa caaacaggtc tacaagata ctgtgcagtc ttaactgaat 900
 taactgatta ataaccagca aaagacccca aaaggattgt tatacattga tatgtggggd 960
 acattgagac acgccgcaaa tgctgcattc atcatgtdg aagctgccga gtdgggtdta 1020
 tccgcatoat cttacagaca gtdtgcata actcagatcg actacgcttd ggdgacggd 1080
 ggaagaagtd tcgtctgtgg ttdtgdtdca aacctccta caagaccaca tcatcgttdt 1140
 tccagtdgcc cgcctgcccc agcaactdgt gactggaata cattcaactc acctgaacca 1200
 aattaccacg tgtdatctgg agcttdgtda ggaggaccag atcaaaacga taattatgtg 1260
 gatgatagat ccgactacgt ccataacgaa gtggcaaccg actacaacgc cggatttdcag 1320
 agtdcttdtg cagccttagt tgcttdgggt tatta 1356

<210> 13
 <211> 1356
 <212> ADN
 <213> *Coptotermes acinaciformis*

5

<400> 13

ES 2 621 181 T3

atgagattcc ctagtatfff cactgccgtc ttatttgcag ccagttctgc tttagccgca 60
 tatgattata ccacagtttt gaaaagttcc ttattgttct acgaagctca aagatccggt 120
 aagttgccag ccgaccagaa ggtcacttgg agaaaagatt cagcattaga cgataaagga 180
 aataatggag aggacttaac aggaggttat tatgacgctg gtgattttgt gaagtttggg 240
 tttccttttag catacaccgc tactgtttta gcctgggggt tgggtggacta tgaagcgggt 300
 tactcatocg ctggagccac agatgacggt agaaaggcag tgaaatgggc aaccgactat 360
 ttgttgaagg cacatactgc cgctaccgag ttatacggac aggtcgggga cggtgacgcc 420
 gatcacgcat attggggacg tcctgaagat atgactatgg ctagaccagc atacaagatc 480
 gacgctagca gaccaggatc tgacttagcg ggtgaaaccg ctgccgcttt agccgctgca 540
 tccatagttt tcaaaggtgt agattcttca tattctgaca acttgttagc tcacgctaaa 600
 cagttatftg atttgcgtga caattataga ggaaaataca gtgattccat aacacaagct 660
 tcaaactfff acgcctccgg agattacaaa gaogagttag tctgggctgc cacttggttg 720
 tacagagcaa ccaacgataa tacatatttg accaaagcag aatccttgta caacgagttc 780
 ggattaggaa actggaacgg agcctttaat tgggacaaca aggtgtccgg tgttcaggtg 840
 ttgttagcca aattgacctc caagcaggt tataaagaca ccgttcaagg atacgtcgat 900
 tatttgatta acaatcagca aaagaccoca aagggtttgt tatacataga ccaatggggg 960

 accttgagac acgcagctaa tgctgcctta ataatcttac aggtctgctga tttgggtatt 1020
 tctgccgaca gttatagaca attcgcaaag aagcaaatag attacgcttt aggtgacgga 1080
 ggtagatcat atgtagttgg ttttgagac aatcctccaa cacatcctca tcaccgttct 1140
 tcctcatgcc ctgacgccc agcagtatgc gattggaata ctttcaattc acctgatcca 1200
 aactttcatg tcttaaccgg agctttagtg ggaggctctg atcagaacga taactacgtt 1260
 gatgatogtt ctgactacgt gtccaacgag gttgcaaccg actataatgc aggattccaa 1320
 agtgcctggg ccgctttagt tactttagga gtttaa 1356

<210> 14
 <211> 1356
 <212> ADN
 <213> *Mastotermes darwinensis*

5

<400> 14

ES 2 621 181 T3

atgagattcc caagtatatt tactgctggt ttgttcgcag ccagttctgc ttagcagcc 60
 tatgattaca atgacgtatt aaccaaaagt ttgttggtct acgaagctca aagatccggt 120
 aagttacott ctgatcagaa agtcacctgg agaaaagatt cagcattaaa cgataagggg 180
 caaaatggtg aggacttaac tggatgatgat tatgacgccg gtgattacgt gaagtttggg 240
 tttccaatgg catatactgc taccgttttg gcttgggggt tagtggacca tctgcccggg 300
 tacagttctg cgggtgtcct ggatgatggt agaaaagctg tgaagtgggt taccgattac 360
 ttaatcaaag cccacgtatc aaagaacgaa ttatacggac aggtcgggtga cggtgacgca 420
 gatcacgctt attggggacg tccagaggat atgacaatgg caagaccagc atacaaaata 480
 gacacttcaa gaccaggttc cgacttagcg ggtgaaaccg cagcggcatt ggctgctgca 540
 tctattgtgt ttaagtcaac agattctaata tacgcccaaca ccttattgac ccacgcaaaa 600
 caattattcg actttgccaa taactataga ggtaagtata gtgattccat aacacaggca 660
 tctaatttct acagtagttc cgactataaa gatgaattgg tttgggcagc tgtatggttg 720
 tacagagcca ctaacgatca gacctattg acaactgcag agaagttata ctcagacttg 780
 ggattacagt cctggaacgg aggtttcaca tgggacacca aaattagtgg agtagaagtg 840
 ttattggcta agattactgg taaacaggca tataaggaca aagtaaaggg atattgtgat 900
 tatatctcag gatctcagca gaaaacacct aaaggattag tttacataga taagtggggg 960
 tccttaagaa tggccgcaaa cgccgcatat atttgcgctg tagccgcaga cgtcggaatc 1020
 agttcaacag cttacagaca gttcgccaaa acacagatta attacatatt ggggtgatgcc 1080
 ggacgttctt ttgtggttg ttacggaaac aaccaccta cacaccaca tcacagatcc 1140
 agttcatgtc ctgacgccc agcaacatgc gattggaata actacaacag tgctaaccct 1200
 aatccacatg ttttatacgg tgcattagtt ggtggaccag attccaacga taattatcaa 1260
 gacttaagat cagattatgt cgccaacgaa gtggcaacag actacaatgc agccttccag 1320
 tcattgtag cattaatcgt ggacttaggt ttgtaa 1356

<210> 15
 <211> 1356
 <212> ADN
 <213> *Nasutitermes walkeri*
 <400> 15

5

ES 2 621 181 T3

atgagatttc catctatfff cactgcccgc ttatttgacg cctcaagtgc tttagcagcc 60
 tatgattaca aacaagtatt gagagattcc ttattgttct acgaagctca gagaagcggc 120
 agattaccag cagaccagaa ggtcacctgg agaaaagatt ccgccttgaa tgatcagggc 180
 gagcaaggtc aagacttaac cggagggtat tttgacgccg gtgattttgt gaagtttgga 240
 ttcccaatgg cttatacagc aaccgttttg gcctgggggt taatcgactt tgaagccggc 300
 tactcttctg ctggtgcctt ggacgatggt agaaaagcag taaagtgggc tactgattac 360
 tttataaaag cccatacttc tcaaaacgag ttttacggac aagtcggctc gggtgacgta 420
 gatcacgcat attggggacg tcctgaagat atgacaatgg ctagaccagc ctacaagatt 480
 gataccagca gaccaggtag tgacttagca ggagaaactg ctgcagcttt ggctgccgca 540
 tccatcgttt tcaagaatgt agatggtaca tattccaaca acttacttac tcatgctaga 600
 cagttgtttg atttcgcaa caattacaga ggaaaatact ctgatagtat taccgatgca 660
 agaaactttt acgctagtgc cgactataga gatgagttag tctgggcagc tgccctggtg 720
 tacagagcaa ccaacgacaa ttcttacttg aacactgctg aatcattata caacgagttt 780
 ggattgcaaa attgggggtg aggggttaaac tgggattcta aagtgagtgg tgttcaagtt 840
 ttgtagcca agttgaccaa caacaagag tataaggaca ctattcaatc atacgtgaat 900
 tacttaatca ataaccaaca gaaaactcca aagggtattgt tatacattga catgtggggg 960
 accttgagac acgcagctaa cgcagccttt ataatgtag aagctgccga cttaggttta 1020
 tccgcttcat cttatagaca gttcgcccaa acacaaatag actacgcatt gggggacggc 1080
 ggacgttctt ttgtctgtgg tttcggttct aatcctcaa ctagacctca tcatagatcc 1140
 agttcatgcc cgctgctcc agctacctgt gattggaata cattcaattc tctgaccca 1200
 aactacaatg ttttatccgg tgccttggtt ggtggtcctg accagaatga taactacgtg 1260
 gacgatagaa gtgattatgt ccataatgag gtagcaactg actacaatgc cggtttccaa 1320
 tcagccttag ccgctttagt cgccttaggt tactaa 1356

<210> 16
 <211> 1356
 <212> ADN
 <213> *Reticulitermes speratus*

5

<400> 16

ES 2 621 181 T3

atgagattcc caagtatatt tactgccgtc ttatttgcag cctccagtgc attagccgct 60
 tatgactaca aaacagtatt gtccaattcc ttgttgttct acgaagctca aagatccggt 120
 aagttacctt ctgaccagaa agtgacctgg agaaaggatt cagcattaaa cgacaaagga 180
 caaaagggtg aggacttaac cgggtgatat tacgacgccg gagactttgt gaaatttgggt 240
 tttccaatgg cttacacagt taccgtattg gcatgggggtg ttattgatta cgaatccgcc 300
 tactctgccg caggagcttt agattcagggt agaaaggcct tgaaatatgg gaccgactat 360
 ttcttaaagg cacatacagc agctaacgag ttttacggac aggtgggtca aggtgacggt 420
 gaccaogcat actggggagc tcttgaagat atgaccatga gcagaccagc atacaaaata 480
 gacacttcta agcctgggtc cgacttagct gcagagactg cagctgcatt agcagccaca 540
 gctattgcat acaaactctgc cgatgcaaca tattccaaca atttgataac acatgcaaaa 600
 caattattcg actttgccaa caattacaga ggaaaatatt ccgatagtat taccgatgcc 660
 aagaactttt atgcttctgg tgattacaaa gacgaattgg tatgggccgc tgcattgggtg 720
 tacagagcaa ccaatgacaa cacatatttg actaaggcag aatccttata caatgaattt 780
 ggtttgggaa acttcaatgg tgccttcaat tgggataaca aagtctccgg agtccagggtg 840
 ttattggcca agttaacctc aaaacaagtg tataaggata aggtacagtc ttacgtggac 900
 tatttgatct cctcacaataa aaagacacca aaaggtttag tgtacatcga tcaatgggggt 960
 actttaagac acgcagctaa ttctgctttg atcgctttgc aggcagctga cttaggaatt 1020
 aacgctgcta cttacagagc ctacgcaaag aagcaaactg actatgcttt gggatgatgg 1080
 ggaagatcct atgttattgg atttgggacc aaccctccag taagaccaca tcacagaagt 1140
 tcatcttgcc cagatgcacc agctgtctgc gattggaaca cctataactc cgctgggtcca 1200
 aacgcccacg tgtaaccgg tgcatgggtt ggaggacctg atagtaatga tagttatacc 1260
 gatgctogtt ctgactacat atccaacgaa gtggcaactg attacaatgc gggtttccaa 1320
 tccgctgtcg ctggattatt gaaggcgggt gtctaa 1356

<210> 17
 <211> 1227
 <212> ADN
 <213> *Neosartorya fischeri*

<400> 17

5

ES 2 621 181 T3

```

atgagatttc catctatfff cactgcagtt ttgttcgcag ccagttccgc tttggcccaa      60
cagatcgggt ccatcgccga aatcatcct gagttgacaa cctatagatg ctctctcaa      120
gctggatgcg tagcacagag tacttccgtc gtgtagata ttaacgctca ttggattcat      180
caaacgggtg cccaaacaag ttgcactacc tcaagtggat tggacccttc attgtgcctt      240
gataaagtca cctgtttctca gaactgcgta gtcgaaggaa taaccgacta ctcatctfff      300

gggtgtgcaaa actccggaga tgcaatgaca ttaagacagt atcaagttca aaatggacag      360
atcaaaacat tgcgtcctag agtgtacttg ttagctgagg atggaatcaa ttactccaaa      420
ttgcagttgt tgaaccaaga gtttactttc gatgtggacg ctccaaatt gccttgtggt      480
atgaatggag ctttatatff gtcagaaatg gatgcttctg gtggacgctc tgccttgaac      540
ccagcgggtg ccacatatgg aacaggttac tgtgatgcc agtgcttcaa cccagggtcca      600
tgataaatg gagaagcaaa tactgctgga gccgggtgcat gttgccaaga gatggactta      660
tgggaagcca actcccgctc taccatfff agtcctcacc catgtacaac tgcggggttg      720
tatgcctgta ctggagctga gtgctactca atctgtgacg gttatggttg cacttacaac      780
ccttatgaat taggagccaa agattactat ggttacggtt tgactattga caccgcaaag      840
ccaataacag tggttactca gtttatgacc gctgataata cagcaaccgg tacattagca      900
gagatcagaa gattatatgt tcaagatggt aaagtaatcg gaaatacagc cgtggccatg      960
accgagggcat tttgtagttc tagtagaaca tttgaagagt taggtggttt gcaaagaatg     1020
ggagaagcct taggtagagg aatggtgcca gttttctcaa tatgggacga tcctgggttg     1080
tggatgcatt ggtagattc tgacggtgca ggaccttgtg gtaatactga aggtgatcct     1140
gccttcattc aggctaacta cccaatacc gccgtaacat tctccaaggt gagatgggga     1200
gatatcggtt ctacctatag ttcttaa                                           1227

```

<210> 18
 <211> 915
 <212> ADN
 <213> *Reticulitermes flavipes*
 <400> 18

5

ES 2 621 181 T3

atgagatttc	catctatfff	cactgctggt	ttgttcgcag	cctcaagtgc	tttagcacia	60
tggatgcaga	tcggtggtaa	gcagaaatat	cctgccttta	agccaggtgc	taagtacgga	120
agaggttatt	gtgacggaca	gtgccctcac	gacatgaagg	tgtctagtgg	aagagcaaac	180
gttgacggat	ggaagccaca	agacaacgac	gaaaatagtg	gaaatggaaa	attgggtaca	240
tgttgctggg	agatggatat	atgggaagga	aacttagtgt	ccaagccta	caccgttcac	300
gctggttcca	agtccggaca	atatgagtg	actggaacac	aatgcggtga	caccgacagt	360
ggtgaaagat	tcaagggaac	atgcgataaa	gatggttgtg	atttcgcaag	ttacagatgg	420
ggagctacag	actattaagg	tcctggaaag	accgtggaca	ccaaacagcc	aatgacagtc	480
gtgaccagtc	tcattgggta	ccctttgact	gagataaaga	gagtttatgt	acaaggagga	540
aaagtcataa	acaattccaa	aacatctaac	ttaggttcag	tgtacgattc	tttgactgag	600
gccttotgcg	atgacaccaa	acaggttaca	ggtgatacaa	atgactttaa	ggctaaagga	660
ggtatgtctg	gattotccaa	gaacttagac	acccacaag	ttttggtgat	gtctttatgg	720
gatgaccata	cagctaatat	gttatgggta	gattctactt	atcctaccga	tagtaciaaag	780
ccaggtgccg	caagaggtac	ttgtgccgtc	acctccgggg	accctaaaga	tgtggaatcc	840
aagcaagcca	actctcaggt	agtttacagt	gacattaagt	ttggtcctat	taattcaaca	900
taciaaagcaa	attaa					915

<210> 19
 <211> 1428
 <212> ADN
 <213> *Trichoderma reesei*

5

<400> 19

ES 2 621 181 T3

atggtctcct tcacctccct gctggccggc gttgccgcta tctctggtgt cctagcagcc 60
 cctgccgcag aagttgaacc tgctgcagtt gagaaacgtg aggccgaagc agaagctcaa 120
 caaccaggaa catcaacacc agaagtccat ccaaagttaa caacctataa atgtactaag 180
 agtggagggt gtgtagcgca ggacacaagt gtggtcttag actggaatta tcgttgatg 240
 catgatgcca attataattc ctgtaactgtt aacggcgggtg ttaacactac gttatgcccc 300
 gatgaagoga cttgtggtaa gaattgtttt attgaagggg ttgactacgc cgctagtgg 360
 gttacgacga gtgggtcatc cttgaogatg aatcaatata tgccttcttc tagtgggtggg 420
 tattcctctg tgtctccaag gctgtattta ttggattccg atggggaata tgttatgtta 480
 aaattaaatg ggcaagaact gagttttgat gtggatctat ctgcattacc ttgtggagaa 540
 aatggtagtc tttatttatc acaaatggac gaaaacggcg gagccaatca gtacaatata 600
 gctggtgcta attatggttc aggctattgt gatgctcaat gtccagtgca gacttgagg 660
 aatggcacct taaacacatc acatcaagga ttttgctgta acgaaatgga catattagaa 720
 ggtaattcaa gagctaatgc actaactccg cactcttgta ctgcgaccgc atgtgattct 780
 gccggttggt gtttcaacc ttatggttct ggttataaga gttactacgg tccgggagac 840
 accgtggata cgtcaaagac cttcactata atcactcagt ttaacacaga taacggatct 900
 ccgagtggta atttgggtgag tattactagg aaatcagc agaacgggtg tgatattccg 960
 tccgcgcagc caggcgggtga cactatatct agctgtcctt ccgccagtgc ctatggcgga 1020
 cttgctacaa tgggtaaggc attgtcctca ggtatgggtc tagtattttc tatttggaat 1080
 gataattcac aatacatgaa ttggctggat tctggtaatg caggcccttg ctccctctaca 1140
 gaaggaacc caagcaatat actagctaat aaccxaaata ctcatgttgt ctttagtaat 1200
 attagatggg gcgatatagg tagcactacg aacagtaccg cacctcctcc tccacctgct 1260
 agctccacga cattttccac tactagaagg tccagcacta ccagctcatc accatcttgt 1320
 actcaaacc attggggaca gtgtgggtgt ataggttaca gcggttgcaa aacttgaca 1380
 tctggtacta catgccaata cagtaatgac tattactcac aatgtaa 1428

<210> 20
 <211> 2688
 <212> ADN
 <213> *Saccharomycopsis fibuligera*
 <400> 20

5

ES 2 621 181 T3

atggtctoct tcacctccct cctcgccggc gtgcgcccca tctcgggcgt cttggccgct 60
 cccgccgcgc aggtcgaatc cgtggctgtg gagaagcgcg cggactcgcg agtcccaatt 120
 caaaactata cccagtctcc atcccagaga gatgagagct cccaatgggt gagcccgcac 180
 tattatccaa ctccacaagg tggtaggctc caagacgtct ggcaagaagc atatgctaga 240
 gcaaaagcca tcgttgGCCa gatgactatt gttgaaaagg tcaatttgac cactgggtacc 300
 ggttggcaat tagatccatg tgttggtaat accggttctg ttccaagatt cggcatocca 360
 aacctttgcc tacaagatgg gccattgggt gttcgattcg ctgactttgt tactggctat 420
 ccatccggtc ttgctactgg tgcaacgctc aataaggatt tgtttcttca aagaggtcaa 480
 gctctcggtc atgagttcaa cagcaaagggt gtacatattg cgttgggccc tgctggtggc 540
 ccacttggtg tcaaagccag aggtggcaga aatttcgaag cctttggttc cgaccatat 600
 ctccaaggta ctgctgctgc tgcaaccatc aaaggtctcc aagagaataa tgttatggct 660
 tgtgtcaagc actttattgg taacgaacaa gaaaagtaca gacagccaga tgacataaac 720
 cotgccacca accaaactac taaagaagct attagtGCCa acattccaga cagagccatg 780
 catgagttgt acttgtggcc atttgocgat tccggttcgag caggtggttg ttctggtatg 840
 tgctcttata acagagtcaa caacacttac gcttgCGaaa actcttacat gatgaaccac 900
 ttgcttaaag aagagttggg ttttcaaggc tttgttgttt cggactgggg tgcaacaatta 960
 agtgggggtt atagcgtat ctccggctta gatatgtcta tgccctggtga agtgtatggg 1020
 ggatggaaca ccggcacgtc tttctgggggt caaaacttga cgaagctat ttacaatgag 1080
 actgttccga ttgaaagatt agatgatatg gcaaccagga tcttgctgc tttgtatgct 1140
 accaatagtt tcccaacaga agatcacctt ccaaattttt cttcatggac aacgaaagaa 1200
 tatggcaata aatattatgc tgacaacact accgagattg tcaaagtcaa ctaccatgtg 1260
 gaccatcaa atgactttac ggaggacaca gctttgaagg ttgctgagga atctattgtg 1320
 cttttaaaaa atgaaaacaa cactttgcca atttctcccg aaaaggctaa aagattacta 1380
 ttgtcgggta ttgctgcagg cctgatccg ataggttatc agtgtgaaga tcaatcttgc 1440
 acaaatggcg ctttgtttca aggttgggggt tctggcagtg ttggttctcc aaaatatcaa 1500
 gtcactccat ttgaggaaat ttcttatctt gcaagaaaaa acaagatgca atttgattat 1560
 attcgggagt cttacgactt agctcaagtt actaaagtag cttccgatgc tcatttgtct 1620
 atagttggtg tctctgctgc aagcggtgag ggttatataa ccggtgacgg taaccaaggt 1680

ES 2 621 181 T3

gacagaagaa atctcacttt gtggaacaac ggtgataaat tgattgaaac agttgctgaa 1740
 aactgtgcc aactgttgt tgttgttact tctactggtc aaattaatth tgaaggcttt 1800
 gctgatcacc caaatgttac ogcaattgtc tgggccggcc cattaggtga cagatccggg 1860
 actgctatcg ccaatattct ttttggtaaa gccaaccat caggatcatc tccattcact 1920
 attgctaaga ctgacgatga ttacattcca attgaaacct acagtccatc gagtggtgaa 1980
 cctgaagaca accacttggg tgaaaatgac ttgcttgttg actatagata ttttgaagag 2040
 aagaatattg agccaagata ogcatttggg tatggcttgg cttacaatga gtatgaagtt 2100
 agcaatgcaa aggtctcggc agccaaaaaa gttgatgagg agttgcctga accagctacc 2160
 tacttatcgg agtttagcta tcaaaatgca aaagacagca aaaatccaag tgatgctttt 2220
 gctccaacag atttaaacag agttaatgag tacctttatc catatttaga tagcaatggt 2280
 accttaaaag acggaaacta tgagtatccc gatggctaca gcactgagca aagaacaaca 2340
 cctatccaac ctggggggcg cttggggagg aacgatgctt tgtgggaggg cgcttataaa 2400
 gttgaagtgg acgttcaaaa cttgggtaac tccactgata agtttgttcc acagttgtat 2460
 ttgaaacacc ctgaggatgg caagtttgaa acccctattc aattgagagg gtttgaagag 2520
 gttgagttgt ccccggtgga gaagaagaca gttgagtttg agcttttgag aagagatctt 2580
 agtgtgtggg ataccaccag acagtcttgg atcgttgaat ctggtactta tgaggcotta 2640
 attggtggtg ctgttaatga tatcaagaca tctgtcctgt ttactatt 2688

<210> 21
 <211> 525
 <212> PRT
 <213> *Humicola grisea*

5

<400> 21

Met Arg Thr Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Ser Ala
 1 5 10 15
 Ala Ala Gln Gln Ala Cys Ser Leu Thr Thr Glu Arg His Pro Ser Leu
 20 25 30
 Ser Trp Asn Lys Cys Thr Ala Gly Gly Gln Cys Gln Thr Val Gln Ala
 35 40 45
 Ser Ile Thr Leu Asp Ser Asn Trp Arg Trp Thr His Gln Val Ser Gly
 50 55 60
 Ser Thr Asn Cys Tyr Thr Gly Asn Lys Trp Asp Thr Ser Ile Cys Thr
 65 70 75 80

10

ES 2 621 181 T3

Asp Ala Lys Ser Cys Ala Gln Asn Cys Cys Val Asp Gly Ala Asp Tyr
 85 90 95
 Thr Ser Thr Tyr Gly Ile Thr Thr Asn Gly Asp Ser Leu Ser Leu Lys
 100 105 110
 Phe Val Thr Lys Gly Gln His Ser Thr Asn Val Gly Ser Arg Thr Tyr
 115 120 125
 Leu Met Asp Gly Glu Asp Lys Tyr Gln Thr Phe Glu Leu Leu Gly Asn
 130 135 140
 Glu Phe Thr Phe Asp Val Asp Val Ser Asn Ile Gly Cys Gly Leu Asn
 145 150 155 160
 Gly Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Leu Ser Arg
 165 170 175
 Tyr Pro Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp
 180 185 190
 Ala Gln Cys Pro Arg Asp Ile Lys Phe Ile Asn Gly Glu Ala Asn Ile
 195 200 205
 Glu Gly Trp Thr Gly Ser Thr Asn Asp Pro Asn Ala Gly Ala Gly Arg
 210 215 220
 Tyr Gly Thr Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Asn Met
 225 230 235 240
 Ala Thr Ala Phe Thr Pro His Pro Cys Thr Ile Ile Gly Gln Ser Arg
 245 250 255
 Cys Glu Gly Asp Ser Cys Gly Gly Thr Tyr Ser Asn Glu Arg Tyr Ala
 260 265 270
 Gly Val Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Gln Gly
 275 280 285
 Asn Lys Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Thr Lys Lys
 290 295 300
 Ile Thr Val Val Thr Gln Phe Leu Lys Asp Ala Asn Gly Asp Leu Gly
 305 310 315 320
 Glu Ile Lys Arg Phe Tyr Val Gln Asp Gly Lys Ile Ile Pro Asn Ser
 325 330 335

ES 2 621 181 T3

Glu Ser Thr Ile Pro Gly Val Glu Gly Asn Ser Ile Thr Gln Asp Trp
 340 345 350

Cys Asp Arg Gln Lys Val Ala Phe Gly Asp Ile Asp Asp Phe Asn Arg
 355 360 365

Lys Gly Gly Met Lys Gln Met Gly Lys Ala Leu Ala Gly Pro Met Val
 370 375 380

Leu Val Met Ser Ile Trp Asp Asp His Ala Ser Asn Met Leu Trp Leu
 385 390 395 400

Asp Ser Thr Phe Pro Val Asp Ala Ala Gly Lys Pro Gly Ala Glu Arg
 405 410 415

Gly Ala Cys Pro Thr Thr Ser Gly Val Pro Ala Glu Val Glu Ala Glu
 420 425 430

Ala Pro Asn Ser Asn Val Val Phe Ser Asn Ile Arg Phe Gly Pro Ile
 435 440 445

Gly Ser Thr Val Ala Gly Leu Pro Gly Ala Gly Asn Gly Gly Asn Asn
 450 455 460

Gly Gly Asn Pro Pro Pro Pro Thr Thr Thr Thr Ser Ser Ala Pro Ala
 465 470 475 480

Thr Thr Thr Thr Ala Ser Ala Gly Pro Lys Ala Gly Arg Trp Gln Gln
 485 490 495

Cys Gly Gly Ile Gly Phe Thr Gly Pro Thr Gln Cys Glu Glu Pro Tyr
 500 505 510

Ile Cys Thr Lys Leu Asn Asp Trp Tyr Ser Gln Cys Leu
 515 520 525

<210> 22
 <211> 456
 <212> PRT
 <213> *Thermoascus aurantiacus*
 <400> 22

Met Tyr Gln Arg Ala Leu Leu Phe Ser Phe Phe Leu Ala Ala Ala Arg
 1 5 10 15

Ala His Glu Ala Gly Thr Val Thr Ala Glu Asn His Pro Ser Leu Thr
 20 25 30

5

10

ES 2 621 181 T3

Trp Gln Gln Cys Ser Ser Gly Gly Ser Cys Thr Thr Gln Asn Gly Lys
 35 40 45

Val Val Ile Asp Ala Asn Trp Arg Trp Val His Thr Thr Ser Gly Tyr
 50 55 60

Thr Asn Cys Tyr Thr Gly Asn Thr Trp Asp Thr Ser Ile Cys Pro Asp
 65 70 75 80

Asp Val Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr Ser
 85 90 95

Gly Thr Tyr Gly Val Thr Thr Ser Gly Asn Ala Leu Arg Leu Asn Phe
 100 105 110

Val Thr Gln Ser Ser Gly Lys Asn Ile Gly Ser Arg Leu Tyr Leu Leu
 115 120 125

Gln Asp Asp Thr Thr Tyr Gln Ile Phe Lys Leu Leu Gly Gln Glu Phe
 130 135 140

Thr Phe Asp Val Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala
 145 150 155 160

Leu Tyr Phe Val Ala Met Asp Ala Asp Gly Asn Leu Ser Lys Tyr Pro
 165 170 175

Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln
 180 185 190

Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly
 195 200 205

Trp Gln Pro Ser Ala Asn Asp Pro Asn Ala Gly Val Gly Asn His Gly
 210 215 220

Ser Ser Cys Ala Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser Thr
 225 230 235 240

Ala Val Thr Pro His Pro Cys Asp Thr Pro Gly Gln Thr Met Cys Gln
 245 250 255

Gly Asp Asp Cys Gly Gly Thr Tyr Ser Ser Thr Arg Tyr Ala Gly Thr
 260 265 270

Cys Asp Thr Asp Gly Cys Asp Phe Asn Pro Tyr Gln Pro Gly Asn His
 275 280 285

ES 2 621 181 T3

Ser Phe Tyr Gly Pro Gly Lys Ile Val Asp Thr Ser Ser Lys Phe Thr
 290 295 300

Val Val Thr Gln Phe Ile Thr Asp Asp Gly Thr Pro Ser Gly Thr Leu
 305 310 315 320

Thr Glu Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Gln
 325 330 335

Ser Glu Ser Thr Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Thr Glu
 340 345 350

Tyr Cys Thr Ala Gln Lys Ala Ala Phe Asp Asn Thr Gly Phe Phe Thr
 355 360 365

His Gly Gly Leu Gln Lys Ile Ser Gln Ala Leu Ala Gln Gly Met Val
 370 375 380

Leu Val Met Ser Leu Trp Asp Asp His Ala Ala Asn Met Leu Trp Leu
 385 390 395 400

Asp Ser Thr Tyr Pro Thr Asp Ala Asp Pro Asp Thr Pro Gly Val Ala
 405 410 415

Arg Gly Thr Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu Ser
 420 425 430

Gln Asn Pro Asn Ser Tyr Val Ile Tyr Ser Asn Ile Lys Val Gly Pro
 435 440 445

Ile Asn Ser Thr Phe Thr Ala Asn
 450 455

<210> 23
 <211> 455
 <212> PRT
 <213> *Talaromyces emersonii*

<400> 23

Met Leu Arg Arg Ala Leu Leu Leu Ser Ser Ser Ala Ile Leu Ala Val
 1 5 10 15

Lys Ala Gln Gln Ala Gly Thr Ala Thr Ala Glu Asn His Pro Pro Leu
 20 25 30

Thr Trp Gln Glu Cys Thr Ala Pro Gly Ser Cys Thr Thr Gln Asn Gly
 35 40 45

5

10

ES 2 621 181 T3

Ala Val Val Leu Asp Ala Asn Trp Arg Trp Val His Asp Val Asn Gly
50 55 60

Tyr Thr Asn Cys Tyr Thr Gly Asn Thr Trp Asp Pro Thr Tyr Cys Pro
65 70 75 80

Asp Asp Glu Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr
85 90 95

Glu Gly Thr Tyr Gly Val Thr Ser Ser Gly Ser Ser Leu Lys Leu Asn
100 105 110

Phe Val Thr Gly Ser Asn Val Gly Ser Arg Leu Tyr Leu Leu Gln Asp
115 120 125

Asp Ser Thr Tyr Gln Ile Phe Lys Leu Leu Asn Arg Glu Phe Ser Phe
130 135 140

Asp Val Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala Leu Tyr
145 150 155 160

Phe Val Ala Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Asn Asn
165 170 175

Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys Pro
180 185 190

Arg Asp Leu Lys Phe Ile Asp Gly Glu Ala Asn Val Glu Gly Trp Gln
195 200 205

Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Asp His Gly Ser Cys
210 215 220

Cys Ala Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser Asn Ala Val
225 230 235 240

Thr Pro His Pro Cys Asp Thr Pro Gly Gln Thr Met Cys Ser Gly Asp
245 250 255

Asp Cys Gly Gly Thr Tyr Ser Asn Asp Arg Tyr Ala Gly Thr Cys Asp
260 265 270

Pro Asp Gly Cys Asp Phe Asn Pro Tyr Arg Met Gly Asn Thr Ser Phe
275 280 285

Tyr Gly Pro Gly Lys Ile Ile Asp Thr Thr Lys Pro Phe Thr Val Val

290	295	300
Thr Gln Phe Leu Thr Asp Asp Gly Thr Asp Thr Gly Thr Leu Ser Glu 305	310	315 320
Ile Lys Arg Phe Tyr Ile Gln Asn Ser Asn Val Ile Pro Gln Pro Asn 325	330	335
Ser Asp Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Thr Glu Phe Cys 340	345	350
Thr Ala Gln Lys Gln Ala Phe Gly Asp Thr Asp Asp Phe Ser Gln His 355	360	365
Gly Gly Leu Ala Lys Met Gly Ala Ala Met Gln Gln Gly Met Val Leu 370	375	380
Val Met Ser Leu Trp Asp Asp Tyr Ala Ala Gln Met Leu Trp Leu Asp 385	390	395 400
Ser Asp Tyr Pro Thr Asp Ala Asp Pro Thr Thr Pro Gly Ile Ala Arg 405	410	415
Gly Thr Cys Pro Thr Asp Ser Gly Val Pro Ser Asp Val Glu Ser Gln 420	425	430
Ser Pro Asn Ser Tyr Val Thr Tyr Ser Asn Ile Lys Phe Gly Pro Ile 435	440	445
Asn Ser Thr Phe Thr Ala Ser 450	455	

<210> 24
 <211> 459
 <212> PRT
 <213> *Talaromyces emersonii*
 <400> 24

5

ES 2 621 181 T3

Met Arg Asn Leu Leu Ala Leu Ala Pro Ala Ala Leu Leu Val Gly Ala
1 5 10 15

Ala Glu Ala Gln Gln Ser Leu Trp Gly Gln Cys Gly Gly Ser Ser Trp
 20 25 30

Thr Gly Ala Thr Ser Cys Ala Ala Gly Ala Thr Cys Ser Thr Ile Asn
 35 40 45

Pro Tyr Tyr Ala Gln Cys Val Pro Ala Thr Ala Thr Pro Thr Thr Leu

ES 2 621 181 T3

50						55						60					
Thr	Thr	Thr	Thr	Lys	Pro	Thr	Ser	Thr	Gly	Gly	Ala	Ala	Pro	Thr	Thr		
65					70					75					80		
Pro	Pro	Pro	Thr	Thr	Thr	Gly	Thr	Thr	Thr	Ser	Pro	Val	Val	Thr	Arg		
				85					90					95			
Pro	Ala	Ser	Ala	Ser	Gly	Asn	Pro	Phe	Glu	Gly	Tyr	Gln	Leu	Tyr	Ala		
			100					105					110				
Asn	Pro	Tyr	Tyr	Ala	Ser	Glu	Val	Ile	Ser	Leu	Ala	Ile	Pro	Ser	Leu		
		115					120					125					
Ser	Ser	Glu	Leu	Val	Pro	Lys	Ala	Ser	Glu	Val	Ala	Lys	Val	Pro	Ser		
	130					135					140						
Phe	Val	Trp	Leu	Asp	Gln	Ala	Ala	Lys	Val	Pro	Ser	Met	Gly	Asp	Tyr		
145					150					155					160		
Leu	Lys	Asp	Ile	Gln	Ser	Gln	Asn	Ala	Ala	Gly	Ala	Asp	Pro	Pro	Ile		
				165						170				175			
Ala	Gly	Ile	Phe	Val	Val	Tyr	Asp	Leu	Pro	Asp	Arg	Asp	Cys	Ala	Ala		
			180					185					190				
Ala	Ala	Ser	Asn	Gly	Glu	Phe	Ser	Ile	Ala	Asn	Asn	Gly	Val	Ala	Leu		
		195					200					205					
Tyr	Lys	Gln	Tyr	Ile	Asp	Ser	Ile	Arg	Glu	Gln	Leu	Thr	Thr	Tyr	Ser		
	210					215					220						
Asp	Val	His	Thr	Ile	Leu	Val	Ile	Glu	Pro	Asp	Ser	Leu	Ala	Asn	Val		
225					230					235					240		
Val	Thr	Asn	Leu	Asn	Val	Pro	Lys	Cys	Ala	Asn	Ala	Gln	Asp	Ala	Tyr		
				245					250					255			
Leu	Glu	Cys	Ile	Asn	Tyr	Ala	Ile	Thr	Gln	Leu	Asp	Leu	Pro	Asn	Val		
			260					265					270				
Ala	Met	Tyr	Leu	Asp	Ala	Gly	His	Ala	Gly	Trp	Leu	Gly	Trp	Gln	Ala		
		275					280					285					
Asn	Leu	Ala	Pro	Ala	Ala	Gln	Leu	Phe	Ala	Ser	Val	Tyr	Lys	Asn	Ala		
	290					295					300						

ES 2 621 181 T3

Ser Ser Pro Ala Ser Val Arg Gly Leu Ala Thr Asn Val Ala Asn Tyr
305 310 315 320

Asn Ala Trp Ser Ile Ser Arg Cys Pro Ser Tyr Thr Gln Gly Asp Ala
325 330 335

Asn Cys Asp Glu Glu Asp Tyr Val Asn Ala Leu Gly Pro Leu Phe Gln
340 345 350

Glu Gln Gly Phe Pro Ala Tyr Phe Ile Ile Asp Thr Ser Arg Asn Gly
355 360 365

Val Arg Pro Thr Lys Gln Ser Gln Trp Gly Asp Trp Cys Asn Val Ile
370 375 380

Gly Thr Gly Phe Gly Val Arg Pro Thr Thr Asp Thr Gly Asn Pro Leu
385 390 395 400

Glu Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Thr
405 410 415

Ser Asn Thr Thr Ser Pro Arg Tyr Asp Tyr His Cys Gly Leu Ser Asp
420 425 430

Ala Leu Gln Pro Ala Pro Glu Ala Gly Thr Trp Phe Gln Ala Tyr Phe
435 440 445

Glu Gln Leu Leu Thr Asn Ala Asn Pro Leu Phe
450 455

<210> 25

<211> 482

<212> PRT

<213> *Chrysosporium lucknowense*

<400> 25

Met Ala Lys Lys Leu Phe Ile Thr Ala Ala Leu Ala Ala Ala Val Leu
1 5 10 15

Ala Ala Pro Val Ile Glu Glu Arg Gln Asn Cys Gly Ala Val Trp Thr
20 25 30

Gln Cys Gly Gly Asn Gly Trp Gln Gly Pro Thr Cys Cys Ala Ser Gly
35 40 45

Ser Thr Cys Val Ala Gln Asn Glu Trp Tyr Ser Gln Cys Leu Pro Asn
50 55 60

5

10

ES 2 621 181 T3

Ser Gln Val Thr Ser Ser Thr Thr Pro Ser Ser Thr Ser Thr Ser Gln
65 70 75 80

Arg Ser Thr Ser Thr Ser Ser Ser Thr Thr Arg Ser Gly Ser Ser Ser
85 90 95

Ser Ser Ser Thr Thr Pro Pro Pro Val Ser Ser Pro Val Thr Ser Ile
100 105 110

Pro Gly Gly Ala Thr Ser Thr Ala Ser Tyr Ser Gly Asn Pro Phe Ser
115 120 125

Gly Val Arg Leu Phe Ala Asn Asp Tyr Tyr Arg Ser Glu Val His Asn
130 135 140

Leu Ala Ile Pro Ser Met Thr Gly Thr Leu Ala Ala Lys Ala Ser Ala
145 150 155 160

Val Ala Glu Val Pro Ser Phe Gln Trp Leu Asp Arg Asn Val Thr Ile
165 170 175

Asp Thr Leu Met Val Gln Thr Leu Ser Gln Val Arg Ala Leu Asn Lys
180 185 190

Ala Gly Ala Asn Pro Pro Tyr Ala Ala Gln Leu Val Val Tyr Asp Leu
195 200 205

Pro Asp Arg Asp Cys Ala Ala Ala Ala Ser Asn Gly Glu Phe Ser Ile
210 215 220

Ala Asn Gly Gly Ala Ala Asn Tyr Arg Ser Tyr Ile Asp Ala Ile Arg
225 230 235 240

Lys His Ile Ile Glu Tyr Ser Asp Ile Arg Ile Ile Leu Val Ile Glu
245 250 255

Pro Asp Ser Met Ala Asn Met Val Thr Asn Met Asn Val Ala Lys Cys
260 265 270

Ser Asn Ala Ala Ser Thr Tyr His Glu Leu Thr Val Tyr Ala Leu Lys
275 280 285

Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala Gly His Ala
290 295 300

Gly Trp Leu Gly Trp Pro Ala Asn Ile Gln Pro Ala Ala Glu Leu Phe
305 310 315 320

Ala Gly Ile Tyr Asn Asp Ala Gly Lys Pro Ala Ala Val Arg Gly Leu
 325 330 335

Ala Thr Asn Val Ala Asn Tyr Asn Ala Trp Ser Ile Ala Ser Ala Pro
 340 345 350

Ser Tyr Thr Ser Pro Asn Pro Asn Tyr Asp Glu Lys His Tyr Ile Glu
 355 360 365

Ala Phe Ser Pro Leu Leu Asn Ser Ala Gly Phe Pro Ala Arg Phe Ile
 370 375 380

Val Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Gln Gln Trp
 385 390 395 400

Gly Asp Trp Cys Asn Val Lys Gly Thr Gly Phe Gly Val Arg Pro Thr
 405 410 415

Ala Asn Thr Gly His Glu Leu Val Asp Ala Phe Val Trp Val Lys Pro
 420 425 430

Gly Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Ala Arg Tyr Asp
 435 440 445

Tyr His Cys Gly Leu Ser Asp Ala Leu Gln Pro Ala Pro Glu Ala Gly
 450 455 460

Gln Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Thr Asn Ala Asn Pro
 465 470 475 480

Pro Phe

<210> 26
 <211> 516
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> cbh1 de *T.emersonni* con CBD

Met Leu Arg Arg Ala Leu Leu Leu Ser Ser Ser Ala Ile Leu Ala Val
 1 5 10 15

Lys Ala Gln Gln Ala Gly Thr Ala Thr Ala Glu Asn His Pro Pro Leu
 20 25 30

Thr Trp Gln Glu Cys Thr Ala Pro Gly Ser Cys Thr Thr Gln Asn Gly
 35 40 45

5

10

ES 2 621 181 T3

Ala Val Val Leu Asp Ala Asn Trp Arg Trp Val His Asp Val Asn Gly
50 55 60

Tyr Thr Asn Cys Tyr Thr Gly Asn Thr Trp Asp Pro Thr Tyr Cys Pro
65 70 75 80

Asp Asp Glu Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr
85 90 95

Glu Gly Thr Tyr Gly Val Thr Ser Ser Gly Ser Ser Leu Lys Leu Asn
100 105 110

Phe Val Thr Gly Ser Asn Val Gly Ser Arg Leu Tyr Leu Leu Gln Asp
115 120 125

Asp Ser Thr Tyr Gln Ile Phe Lys Leu Leu Asn Arg Glu Phe Ser Phe
130 135 140

Asp Val Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala Leu Tyr
145 150 155 160

Phe Val Ala Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Asn Asn
165 170 175

Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys Pro
180 185 190

Arg Asp Leu Lys Phe Ile Asp Gly Glu Ala Asn Val Glu Gly Trp Gln
195 200 205

Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Asp His Gly Ser Cys
210 215 220

Cys Ala Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser Asn Ala Val
225 230 235 240

Thr Pro His Pro Cys Asp Thr Pro Gly Gln Thr Met Cys Ser Gly Asp
245 250 255

Asp Cys Gly Gly Thr Tyr Ser Asn Asp Arg Tyr Ala Gly Thr Cys Asp
260 265 270

Pro Asp Gly Cys Asp Phe Asn Pro Tyr Arg Met Gly Asn Thr Ser Phe
275 280 285

Tyr Gly Pro Gly Lys Ile Ile Asp Thr Thr Lys Pro Phe Thr Val Val

ES 2 621 181 T3

290					295						300					
Thr	Gln	Phe	Leu	Thr	Asp	Asp	Gly	Thr	Asp	Thr	Gly	Thr	Leu	Ser	Glu	
305					310					315					320	
Ile	Lys	Arg	Phe	Tyr	Ile	Gln	Asn	Ser	Asn	Val	Ile	Pro	Gln	Pro	Asn	
				325					330					335		
Ser	Asp	Ile	Ser	Gly	Val	Thr	Gly	Asn	Ser	Ile	Thr	Thr	Glu	Phe	Cys	
			340					345					350			
Thr	Ala	Gln	Lys	Gln	Ala	Phe	Gly	Asp	Thr	Asp	Asp	Phe	Ser	Gln	His	
		355					360					365				
Gly	Gly	Leu	Ala	Lys	Met	Gly	Ala	Ala	Met	Gln	Gln	Gly	Met	Val	Leu	
	370					375						380				
Val	Met	Ser	Leu	Trp	Asp	Asp	Tyr	Ala	Ala	Gln	Met	Leu	Trp	Leu	Asp	
385					390					395					400	
Ser	Asp	Tyr	Pro	Thr	Asp	Ala	Asp	Pro	Thr	Thr	Pro	Gly	Ile	Ala	Arg	
				405					410					415		
Gly	Thr	Cys	Pro	Thr	Asp	Ser	Gly	Val	Pro	Ser	Asp	Val	Glu	Ser	Gln	
			420					425					430			
Ser	Pro	Asn	Ser	Tyr	Val	Thr	Tyr	Ser	Asn	Ile	Lys	Phe	Gly	Pro	Ile	
		435					440					445				
Asn	Ser	Thr	Phe	Thr	Ala	Ser	Asn	Pro	Pro	Gly	Gly	Asn	Arg	Gly	Thr	
	450					455						460				
Thr	Thr	Thr	Arg	Arg	Pro	Ala	Thr	Thr	Thr	Gly	Ser	Ser	Pro	Gly	Pro	
465					470					475					480	
Thr	Gln	Ser	His	Tyr	Gly	Gln	Cys	Gly	Gly	Ile	Gly	Tyr	Ser	Gly	Pro	
				485					490					495		
Thr	Val	Cys	Ala	Ser	Gly	Thr	Thr	Cys	Gln	Val	Leu	Asn	Pro	Tyr	Tyr	
			500					505					510			
Ser	Gln	Cys	Leu													
		515														

<210> 27
 <211> 535
 <212> PRT

<213> *Trichoderma reesei*

<400> 27

ES 2 621 181 T3

Met Val Ser Phe Thr Ser Leu Leu Ala Gly Val Ala Ala Ile Ser Gly
1 5 10 15

Val Leu Ala Ala Pro Ala Ala Glu Val Glu Pro Val Ala Val Glu Lys
20 25 30

Arg Glu Ala Glu Ala Glu Ala Gln Ser Ala Cys Thr Leu Gln Ser Glu
35 40 45

Thr His Pro Pro Leu Thr Trp Gln Lys Cys Ser Ser Gly Gly Thr Cys
50 55 60

Thr Gln Gln Thr Gly Ser Val Val Ile Asp Ala Asn Trp Arg Trp Thr
65 70 75 80

His Ala Thr Asn Ser Ser Thr Asn Cys Tyr Asp Gly Asn Thr Trp Ser
85 90 95

Ser Thr Leu Cys Pro Asp Asn Glu Thr Cys Ala Lys Asn Cys Cys Leu
100 105 110

Asp Gly Ala Ala Tyr Ala Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn
115 120 125

Ser Leu Ser Ile Gly Phe Val Thr Gln Ser Ala Gln Lys Asn Val Gly
130 135 140

Ala Arg Leu Tyr Leu Met Ala Ser Asp Thr Thr Tyr Gln Glu Phe Thr
145 150 155 160

Leu Leu Gly Asn Glu Phe Ser Phe Asp Val Asp Val Ser Gln Leu Pro
165 170 175

Cys Gly Leu Asn Gly Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly
180 185 190

Gly Val Ser Lys Tyr Pro Thr Asn Thr Ala Gly Ala Lys Tyr Gly Thr
195 200 205

Gly Tyr Cys Asp Ser Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly
210 215 220

Gln Ala Asn Val Glu Gly Trp Glu Pro Ser Ser Asn Asn Ala Asn Thr
225 230 235 240

ES 2 621 181 T3

Gly Ile Gly Gly His Gly Ser Cys Cys Ser Glu Met Asp Ile Trp Glu
 245 250 255

Ala Asn Ser Ile Ser Glu Ala Leu Thr Pro His Pro Cys Thr Thr Val
 260 265 270

Gly Gln Glu Ile Cys Glu Gly Asp Gly Cys Gly Gly Thr Tyr Ser Asp
 275 280 285

Asn Arg Tyr Gly Gly Thr Cys Asp Pro Asp Gly Cys Asp Trp Asn Pro
 290 295 300

Tyr Arg Leu Gly Asn Thr Ser Phe Tyr Gly Pro Gly Ser Ser Phe Thr
 305 310 315 320

Leu Asp Thr Thr Lys Lys Leu Thr Val Val Thr Gln Phe Glu Thr Ser
 325 330 335

Gly Ala Ile Asn Arg Tyr Tyr Val Gln Asn Gly Val Thr Phe Gln Gln
 340 345 350

Pro Asn Ala Glu Leu Gly Ser Tyr Ser Gly Asn Glu Leu Asn Asp Asp
 355 360 365

Tyr Cys Thr Ala Glu Glu Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp
 370 375 380

Lys Gly Gly Leu Thr Gln Phe Lys Lys Ala Thr Ser Gly Gly Met Val
 385 390 395 400

Leu Val Met Ser Leu Trp Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu
 405 410 415

Asp Ser Thr Tyr Pro Thr Asn Glu Thr Ser Ser Thr Pro Gly Ala Val
 420 425 430

Arg Gly Ser Cys Ser Thr Ser Ser Gly Val Pro Ala Gln Val Glu Ser
 435 440 445

Gln Ser Pro Asn Ala Lys Val Thr Phe Ser Asn Ile Lys Phe Gly Pro
 450 455 460

Ile Gly Ser Thr Gly Asn Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn
 465 470 475 480

Arg Gly Thr Thr Thr Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser
 485 490 495

ES 2 621 181 T3

Pro Gly Pro Thr Gln Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr
 500 505 510

Ser Gly Pro Thr Val Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn
 515 520 525

Pro Tyr Tyr Ser Gln Cys Leu
 530 535

<210> 28
 <211> 471
 <212> PRT
 <213> *Trichoderma reesei*

<400> 28

Met Ile Val Gly Ile Leu Thr Thr Leu Ala Thr Leu Ala Thr Leu Ala
 1 5 10 15

Ala Ser Val Pro Leu Glu Glu Arg Gln Ala Cys Ser Ser Val Trp Gly
 20 25 30

Gln Cys Gly Gly Gln Asn Trp Ser Gly Pro Thr Cys Cys Ala Ser Gly
 35 40 45

Ser Thr Cys Val Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Pro Gly
 50 55 60

Ala Ala Ser Ser Ser Ser Ser Thr Arg Ala Ala Ser Thr Thr Ser Arg
 65 70 75 80

Val Ser Pro Thr Thr Ser Arg Ser Ser Ser Ala Thr Pro Pro Pro Gly
 85 90 95

Ser Thr Thr Thr Arg Val Pro Pro Val Gly Ser Gly Thr Ala Thr Tyr
 100 105 110

Ser Gly Asn Pro Phe Val Gly Val Thr Pro Trp Ala Asn Ala Tyr Tyr
 115 120 125

Ala Ser Glu Val Ser Ser Leu Ala Ile Pro Ser Leu Thr Gly Ala Met
 130 135 140

Ala Thr Ala Ala Ala Ala Val Ala Lys Val Pro Ser Phe Met Trp Leu
 145 150 155 160

Asp Thr Leu Asp Lys Thr Pro Leu Met Glu Gln Thr Leu Ala Asp Ile
 165 170 175

ES 2 621 181 T3

Arg Thr Ala Asn Lys Asn Gly Gly Asn Tyr Ala Gly Gln Phe Val Val
180 185 190

Tyr Asp Leu Pro Asp Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly Glu
195 200 205

Tyr Ser Ile Ala Asp Gly Gly Val Ala Lys Tyr Lys Asn Tyr Ile Asp
210 215 220

Thr Ile Arg Gln Ile Val Val Glu Tyr Ser Asp Ile Arg Thr Leu Leu
225 230 235 240

Val Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu Gly Thr
245 250 255

Pro Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Ile Asn Tyr
260 265 270

Ala Val Thr Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala
275 280 285

Gly His Ala Gly Trp Leu Gly Trp Pro Ala Asn Gln Asp Pro Ala Ala
290 295 300

Gln Leu Phe Ala Asn Val Tyr Lys Asn Ala Ser Ser Pro Arg Ala Leu
305 310 315 320

Arg Gly Leu Ala Thr Asn Val Ala Asn Tyr Asn Gly Trp Asn Ile Thr
325 330 335

Ser Pro Pro Ser Tyr Thr Gln Gly Asn Ala Val Tyr Asn Glu Lys Leu
340 345 350

Tyr Ile His Ala Ile Gly Arg Leu Leu Ala Asn His Gly Trp Ser Asn
355 360 365

Ala Phe Phe Ile Thr Asp Gln Gly Arg Ser Gly Lys Gln Pro Thr Gly
370 375 380

Gln Gln Gln Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly
385 390 395 400

Ile Arg Pro Ser Ala Asn Thr Gly Asp Ser Leu Leu Asp Ser Phe Val
405 410 415

Trp Val Lys Pro Gly Gly Glu Cys Asp Gly Thr Ser Asp Ser Ser Ala
420 425 430

ES 2 621 181 T3

Pro Arg Phe Asp Ser His Cys Ala Leu Pro Asp Ala Leu Gln Pro Ala
 435 440 445

Ala Gln Ala Gly Ala Trp Phe Gln Ala Tyr Phe Val Gln Leu Leu Thr
 450 455 460

Asn Ala Asn Pro Ser Phe Leu
 465 470

- 5
 <210> 29
 <211> 532
 <212> PRT
 <213> *Chaetomium thermophilum*
 <400> 29

Met Met Tyr Lys Lys Phe Ala Ala Leu Ala Ala Leu Val Ala Gly Ala
 1 5 10 15

Ser Ala Gln Gln Ala Cys Ser Leu Thr Ala Glu Asn His Pro Ser Leu
 20 25 30

Thr Trp Lys Arg Cys Thr Ser Gly Gly Ser Cys Ser Thr Val Asn Gly
 35 40 45

Ala Val Thr Ile Asp Ala Asn Trp Arg Trp Thr His Thr Val Ser Gly
 50 55 60

Ser Thr Asn Cys Tyr Thr Gly Asn Gln Trp Asp Thr Ser Leu Cys Thr
 65 70 75 80

Asp Gly Lys Ser Cys Ala Gln Thr Cys Cys Val Asp Gly Ala Asp Tyr
 85 90 95

Ser Ser Thr Tyr Gly Ile Thr Thr Ser Gly Asp Ser Leu Asn Leu Lys
 100 105 110

Phe Val Thr Lys His Gln Tyr Gly Thr Asn Val Gly Ser Arg Val Tyr
 115 120 125

Leu Met Glu Asn Asp Thr Lys Tyr Gln Met Phe Glu Leu Leu Gly Asn
 130 135 140

Glu Phe Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn
 145 150 155 160

Gly Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Met Ser Lys
 165 170 175

ES 2 621 181 T3

Tyr Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp
 180 185 190
 Ala Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Val
 195 200 205
 Gly Asn Trp Thr Pro Ser Thr Asn Asp Ala Asn Ala Gly Phe Gly Arg
 210 215 220
 Tyr Gly Ser Cys Cys Ser Glu Met Asp Val Trp Glu Ala Asn Asn Met
 225 230 235 240
 Ala Thr Ala Phe Thr Pro His Pro Cys Thr Thr Val Gly Gln Ser Arg
 245 250 255
 Cys Glu Ala Asp Thr Cys Gly Gly Thr Tyr Ser Ser Asp Arg Tyr Ala
 260 265 270
 Gly Val Cys Asp Pro Asp Gly Cys Asp Phe Asn Ala Tyr Arg Gln Gly
 275 280 285
 Asp Lys Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Asn Lys Lys
 290 295 300
 Met Thr Val Val Thr Gln Phe His Lys Asn Ser Ala Gly Val Leu Ser
 305 310 315 320
 Glu Ile Lys Arg Phe Tyr Val Gln Asp Gly Lys Ile Ile Ala Asn Ala
 325 330 335
 Glu Ser Lys Ile Pro Gly Asn Pro Gly Asn Ser Ile Thr Gln Glu Tyr
 340 345 350
 Cys Asp Ala Gln Lys Val Ala Phe Ser Asn Thr Asp Asp Phe Asn Arg
 355 360 365
 Lys Gly Gly Met Ala Gln Met Ser Lys Ala Leu Ala Gly Pro Met Val
 370 375 380
 Leu Val Met Ser Val Trp Asp Asp His Tyr Ala Asn Met Leu Trp Leu
 385 390 395 400
 Asp Ser Thr Tyr Pro Ile Asp Gln Ala Gly Ala Pro Gly Ala Glu Arg
 405 410 415
 Gly Ala Cys Pro Thr Thr Ser Gly Val Pro Ala Glu Ile Glu Ala Gln

ES 2 621 181 T3

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
 1 5 10 15

Ala Leu Ala Glu Cys Thr Lys Gly Gly Cys Thr Asn Lys Asn Gly Tyr
 20 25 30

Ile Val His Asp Lys His Val Gly Asp Ile Gln Asn Arg Asp Thr Leu
 35 40 45

Asp Pro Pro Asp Leu Asp Tyr Glu Lys Asp Val Gly Val Thr Val Ser
 50 55 60

Gly Gly Thr Leu Ser Gln Arg Leu Val Ser Thr Trp Asn Gly Lys Lys
 65 70 75 80

Val Val Gly Ser Arg Leu Tyr Ile Val Asp Glu Ala Asp Glu Lys Tyr
 85 90 95

Gln Leu Phe Thr Phe Val Gly Lys Glu Phe Thr Tyr Thr Val Asp Met

ES 2 621 181 T3

			100					105						110			
Ser	Gln	Ile	Gln	Cys	Gly	Ile	Asn	Ala	Ala	Leu	Tyr	Thr	Val	Glu	Met		
			115					120						125			
Pro	Ala	Ala	Gly	Lys	Thr	Pro	Gly	Gly	Val	Lys	Tyr	Gly	Tyr	Gly	Tyr		
	130						135				140						
Cys	Asp	Ala	Asn	Cys	Val	Asp	Gly	Asp	Cys	Cys	Met	Glu	Phe	Asp	Ile		
145					150						155				160		
Gln	Glu	Ala	Ser	Asn	Lys	Ala	Ile	Val	Tyr	Thr	Thr	His	Ser	Cys	Gln		
					165					170					175		
Ser	Gln	Thr	Ser	Gly	Cys	Asp	Thr	Ser	Gly	Cys	Gly	Tyr	Asn	Pro	Tyr		
			180						185					190			
Arg	Asp	Ser	Gly	Asp	Lys	Ala	Phe	Trp	Gly	Thr	Thr	Ile	Asn	Val	Asn		
			195				200							205			
Gln	Pro	Val	Thr	Ile	Val	Thr	Gln	Phe	Ile	Gly	Ser	Gly	Ser	Ser	Leu		
	210						215					220					
Thr	Glu	Val	Lys	Arg	Leu	Cys	Val	Gln	Gly	Gly	Lys	Thr	Phe	Pro	Pro		
225					230						235				240		
Ala	Lys	Ser	Leu	Thr	Asp	Ser	Tyr	Cys	Asn	Ala	Asn	Asp	Tyr	Arg	Ser		
					245					250					255		
Leu	Arg	Thr	Met	Gly	Ala	Ser	Met	Ala	Arg	Gly	His	Val	Val	Val	Phe		
			260						265					270			
Ser	Leu	Trp	Asp	Ser	Asn	Gly	Met	Ser	Trp	Met	Asp	Gly	Gly	Asn	Ala		
			275				280						285				
Gly	Pro	Cys	Thr	Ser	Tyr	Asn	Ile	Glu	Ser	Leu	Glu	Ser	Ser	Gln	Pro		
	290					295						300					
Asn	Leu	Lys	Val	Thr	Trp	Ser	Asn	Val	Lys	Tyr	Gly	Glu	Ile	Asp	Ser		
305					310					315					320		
Pro	Tyr																

<210> 31
 <211> 451
 <212> PRT
 <213> *Coptotermes formosanus*

ES 2 621 181 T3

<400> 31

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
 1 5 10 15

Ala Leu Ala Ala Tyr Asp Tyr Lys Thr Val Leu Lys Asn Ser Leu Leu
 20 25 30

Phe Tyr Glu Ala Gln Arg Ser Gly Lys Leu Pro Ala Asp Gln Lys Val
 35 40 45

Thr Trp Arg Lys Asp Ser Ala Leu Asn Asp Lys Gly Gln Lys Gly Glu
 50 55 60

Asp Leu Thr Gly Gly Tyr Tyr Asp Ala Gly Asp Phe Val Lys Phe Gly
 65 70 75 80

Phe Pro Met Ala Tyr Thr Val Thr Val Leu Ala Trp Gly Leu Val Asp
 85 90 95

Tyr Glu Ser Ala Tyr Ser Thr Ala Gly Ala Leu Asp Asp Gly Arg Lys
 100 105 110

Ala Leu Lys Trp Gly Thr Asp Tyr Phe Leu Lys Ala His Thr Ala Ala
 115 120 125

Asn Glu Phe Tyr Gly Gln Val Gly Gln Gly Asp Val Asp His Ala Tyr
 130 135 140

Trp Gly Arg Pro Glu Asp Met Thr Met Ser Arg Pro Ala Tyr Lys Ile
 145 150 155 160

Asp Thr Ser Lys Pro Gly Ser Asp Leu Ala Ala Glu Thr Ala Ala Ala
 165 170 175

Leu Ala Ala Thr Ala Ile Ala Tyr Lys Ser Ala Asp Ser Thr Tyr Ser
 180 185 190

Asn Asn Leu Ile Thr His Ala Lys Gln Leu Phe Asp Phe Ala Asn Asn
 195 200 205

Tyr Arg Gly Lys Tyr Ser Asp Ser Ile Thr Asp Ala Lys Asn Phe Tyr
 210 215 220

Ala Ser Gly Asp Tyr Lys Asp Glu Leu Val Trp Ala Ala Ala Trp Leu
 225 230 235 240

ES 2 621 181 T3

Tyr Arg Ala Thr Asn Asp Asn Thr Tyr Leu Thr Lys Ala Glu Ser Leu
 245 250 255

 Tyr Asn Glu Phe Gly Leu Gly Ser Trp Asn Gly Ala Phe Asn Trp Asp
 260 265 270

 Asn Lys Ile Ser Gly Val Gln Val Leu Leu Ala Lys Leu Thr Ser Lys
 275 280 285

 Gln Ala Tyr Lys Asp Lys Val Gln Gly Tyr Val Asp Tyr Leu Val Ser
 290 295 300

 Ser Gln Lys Lys Thr Pro Lys Gly Leu Val Tyr Ile Asp Gln Trp Gly
 305 310 315 320

 Thr Leu Arg His Ala Ala Asn Ser Ala Leu Ile Ala Leu Gln Ala Ala
 325 330 335

 Asp Leu Gly Ile Asn Ala Ala Ser Tyr Arg Gln Tyr Ala Lys Lys Gln
 340 345 350

 Ile Asp Tyr Ala Leu Gly Asp Gly Gly Arg Ser Tyr Val Val Gly Phe
 355 360 365

 Gly Thr Asn Pro Pro Val Arg Pro His His Arg Ser Ser Ser Cys Pro
 370 375 380

 Asp Ala Pro Ala Ala Cys Asp Trp Asn Thr Tyr Asn Ser Ala Gly Pro
 385 390 395 400

 Asn Ala His Val Leu Thr Gly Ala Leu Val Gly Gly Pro Asp Ser Asn
 405 410 415

 Asp Ser Tyr Thr Asp Ser Arg Ser Asp Tyr Ile Ser Asn Glu Val Ala
 420 425 430

 Thr Asp Tyr Asn Ala Gly Phe Gln Ser Ala Val Ala Gly Leu Leu Lys
 435 440 445

 Ala Gly Val
 450

<210> 32
 <211> 451
 <212> PRT
 <213> *Nasutitermes takasagoensis*
 <400> 32

ES 2 621 181 T3

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15

Ala Leu Ala Ala Tyr Asp Tyr Lys Gln Val Leu Arg Asp Ser Leu Leu
20 25 30

Phe Tyr Glu Ala Gln Arg Ser Gly Arg Leu Pro Ala Asp Gln Lys Val
35 40 45

Thr Trp Arg Lys Asp Ser Ala Leu Asn Asp Gln Gly Asp Gln Gly Gln
50 55 60

Asp Leu Thr Gly Gly Tyr Phe Asp Ala Gly Asp Phe Val Lys Phe Gly
65 70 75 80

Phe Pro Met Ala Tyr Thr Ala Thr Val Leu Ala Trp Gly Leu Ile Asp
85 90 95

Phe Glu Ala Gly Tyr Ser Ser Ala Gly Ala Leu Asp Asp Gly Arg Lys
100 105 110

Ala Val Lys Trp Ala Thr Asp Tyr Phe Ile Lys Ala His Thr Ser Gln
115 120 125

Asn Glu Phe Tyr Gly Gln Val Gly Gln Gly Asp Ala Asp His Ala Phe
130 135 140

Trp Gly Arg Pro Glu Asp Met Thr Met Ala Arg Pro Ala Tyr Lys Ile
145 150 155 160

Asp Thr Ser Arg Pro Gly Ser Asp Leu Ala Gly Glu Thr Ala Ala Ala
165 170 175

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

Asn Asn Leu Leu Thr His Ala Arg Gln Leu Phe Asp Phe Ala Asn Asn
195 200 205

Tyr Arg Gly Lys Tyr Ser Asp Ser Ile Thr Asp Ala Arg Asn Phe Tyr
210 215 220

Ala Ser Ala Asp Tyr Arg Asp Glu Leu Val Trp Ala Ala Ala Trp Leu
225 230 235 240

Tyr Arg Ala Thr Asn Asp Asn Thr Tyr Leu Asn Thr Ala Glu Ser Leu
245 250 255

Tyr Asp Glu Phe Gly Leu Gln Asn Trp Gly Gly Gly Leu Asn Trp Asp
 260 265 270
 Ser Lys Val Ser Gly Val Gln Val Leu Leu Ala Lys Leu Thr Asn Lys
 275 280 285
 Gln Ala Tyr Lys Asp Thr Val Gln Ser Tyr Val Asn Tyr Leu Ile Asn
 290 295 300
 Asn Gln Gln Lys Thr Pro Lys Gly Leu Leu Tyr Ile Asp Met Trp Gly
 305 310 315 320
 Thr Leu Arg His Ala Ala Asn Ala Ala Phe Ile Met Leu Glu Ala Ala
 325 330 335
 Glu Leu Gly Leu Ser Ala Ser Ser Tyr Arg Gln Phe Ala Gln Thr Gln
 340 345 350
 Ile Asp Tyr Ala Leu Gly Asp Gly Gly Arg Ser Phe Val Cys Gly Phe
 355 360 365
 Gly Ser Asn Pro Pro Thr Arg Pro His His Arg Ser Ser Ser Cys Pro
 370 375 380
 Pro Ala Pro Ala Thr Cys Asp Trp Asn Thr Phe Asn Ser Pro Asp Pro
 385 390 395 400
 Asn Tyr His Val Leu Ser Gly Ala Leu Val Gly Gly Pro Asp Gln Asn
 405 410 415
 Asp Asn Tyr Val Asp Asp Arg Ser Asp Tyr Val His Asn Glu Val Ala
 420 425 430
 Thr Asp Tyr Asn Ala Gly Phe Gln Ser Ala Leu Ala Ala Leu Val Ala
 435 440 445
 Leu Gly Tyr
 450

<210> 33
 <211> 451
 <212> PRT
 <213> *Coptotermes acinaciformis*
 <400> 33

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
 1 5 10 15

5

10

ES 2 621 181 T3

Ala Leu Ala Ala Tyr Asp Tyr Thr Thr Val Leu Lys Ser Ser Leu Leu
20 25 30

Phe Tyr Glu Ala Gln Arg Ser Gly Lys Leu Pro Ala Asp Gln Lys Val
35 40 45

Thr Trp Arg Lys Asp Ser Ala Leu Asp Asp Lys Gly Asn Asn Gly Glu
50 55 60

Asp Leu Thr Gly Gly Tyr Tyr Asp Ala Gly Asp Phe Val Lys Phe Gly
65 70 75 80

Phe Pro Leu Ala Tyr Thr Ala Thr Val Leu Ala Trp Gly Leu Val Asp
85 90 95

Tyr Glu Ala Gly Tyr Ser Ser Ala Gly Ala Thr Asp Asp Gly Arg Lys
100 105 110

Ala Val Lys Trp Ala Thr Asp Tyr Leu Leu Lys Ala His Thr Ala Ala
115 120 125

Thr Glu Leu Tyr Gly Gln Val Gly Asp Gly Asp Ala Asp His Ala Tyr
130 135 140

Trp Gly Arg Pro Glu Asp Met Thr Met Ala Arg Pro Ala Tyr Lys Ile
145 150 155 160

Asp Ala Ser Arg Pro Gly Ser Asp Leu Ala Gly Glu Thr Ala Ala Ala
165 170 175

Leu Ala Ala Ala Ser Ile Val Phe Lys Gly Val Asp Ser Ser Tyr Ser
180 185 190

Asp Asn Leu Leu Ala His Ala Lys Gln Leu Phe Asp Phe Ala Asp Asn
195 200 205

Tyr Arg Gly Lys Tyr Ser Asp Ser Ile Thr Gln Ala Ser Asn Phe Tyr
210 215 220

Ala Ser Gly Asp Tyr Lys Asp Glu Leu Val Trp Ala Ala Thr Trp Leu
225 230 235 240

Tyr Arg Ala Thr Asn Asp Asn Thr Tyr Leu Thr Lys Ala Glu Ser Leu
245 250 255

Tyr Asn Glu Phe Gly Leu Gly Asn Trp Asn Gly Ala Phe Asn Trp Asp
260 265 270

ES 2 621 181 T3

Asn Lys Val Ser Gly Val Gln Val Leu Leu Ala Lys Leu Thr Ser Lys
 275 280 285

Gln Ala Tyr Lys Asp Thr Val Gln Gly Tyr Val Asp Tyr Leu Ile Asn
 290 295 300

Asn Gln Gln Lys Thr Pro Lys Gly Leu Leu Tyr Ile Asp Gln Trp Gly
 305 310 315 320

Thr Leu Arg His Ala Ala Asn Ala Ala Leu Ile Ile Leu Gln Ala Ala
 325 330 335

Asp Leu Gly Ile Ser Ala Asp Ser Tyr Arg Gln Phe Ala Lys Lys Gln
 340 345 350

Ile Asp Tyr Ala Leu Gly Asp Gly Gly Arg Ser Tyr Val Val Gly Phe
 355 360 365

Gly Asp Asn Pro Pro Thr His Pro His His Arg Ser Ser Ser Cys Pro
 370 375 380

Asp Ala Pro Ala Val Cys Asp Trp Asn Thr Phe Asn Ser Pro Asp Pro
 385 390 395 400

Asn Phe His Val Leu Thr Gly Ala Leu Val Gly Gly Pro Asp Gln Asn
 405 410 415

Asp Asn Tyr Val Asp Asp Arg Ser Asp Tyr Val Ser Asn Glu Val Ala
 420 425 430

Thr Asp Tyr Asn Ala Gly Phe Gln Ser Ala Val Ala Ala Leu Val Thr
 435 440 445

Leu Gly Val
 450

- <210> 34
- <211> 451
- <212> PRT
- <213> *Mastotermes darwinensis*
- <400> 34

5

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
 1 5 10 15

Ala Leu Ala Ala Tyr Asp Tyr Asn Asp Val Leu Thr Lys Ser Leu Leu
 20 25 30

10

ES 2 621 181 T3

Phe Tyr Glu Ala Gln Arg Ser Gly Lys Leu Pro Ser Asp Gln Lys Val
35 40 45

Thr Trp Arg Lys Asp Ser Ala Leu Asn Asp Lys Gly Gln Asn Gly Glu
50 55 60

Asp Leu Thr Gly Gly Tyr Tyr Asp Ala Gly Asp Tyr Val Lys Phe Gly
65 70 75 80

Phe Pro Met Ala Tyr Thr Ala Thr Val Leu Ala Trp Gly Leu Val Asp
85 90 95

His Pro Ala Gly Tyr Ser Ser Ala Gly Val Leu Asp Asp Gly Arg Lys
100 105 110

Ala Val Lys Trp Val Thr Asp Tyr Leu Ile Lys Ala His Val Ser Lys
115 120 125

Asn Glu Leu Tyr Gly Gln Val Gly Asp Gly Asp Ala Asp His Ala Tyr
130 135 140

Trp Gly Arg Pro Glu Asp Met Thr Met Ala Arg Pro Ala Tyr Lys Ile
145 150 155 160

Asp Thr Ser Arg Pro Gly Ser Asp Leu Ala Gly Glu Thr Ala Ala Ala
165 170 175

Leu Ala Ala Ala Ser Ile Val Phe Lys Ser Thr Asp Ser Asn Tyr Ala
180 185 190

Asn Thr Leu Leu Thr His Ala Lys Gln Leu Phe Asp Phe Ala Asn Asn
195 200 205

Tyr Arg Gly Lys Tyr Ser Asp Ser Ile Thr Gln Ala Ser Asn Phe Tyr
210 215 220

Ser Ser Ser Asp Tyr Lys Asp Glu Leu Val Trp Ala Ala Val Trp Leu
225 230 235 240

Tyr Arg Ala Thr Asn Asp Gln Thr Tyr Leu Thr Thr Ala Glu Lys Leu
245 250 255

Tyr Ser Asp Leu Gly Leu Gln Ser Trp Asn Gly Gly Phe Thr Trp Asp
260 265 270

Thr Lys Ile Ser Gly Val Glu Val Leu Leu Ala Lys Ile Thr Gly Lys

ES 2 621 181 T3

275		280		285											
Gln	Ala	Tyr	Lys	Asp	Lys	Val	Lys	Gly	Tyr	Cys	Asp	Tyr	Ile	Ser	Gly
	290					295					300				
Ser	Gln	Gln	Lys	Thr	Pro	Lys	Gly	Leu	Val	Tyr	Ile	Asp	Lys	Trp	Gly
305					310					315					320
Ser	Leu	Arg	Met	Ala	Ala	Asn	Ala	Ala	Tyr	Ile	Cys	Ala	Val	Ala	Ala
				325					330					335	
Asp	Val	Gly	Ile	Ser	Ser	Thr	Ala	Tyr	Arg	Gln	Phe	Ala	Lys	Thr	Gln
			340					345					350		
Ile	Asn	Tyr	Ile	Leu	Gly	Asp	Ala	Gly	Arg	Ser	Phe	Val	Val	Gly	Tyr
		355					360					365			
Gly	Asn	Asn	Pro	Pro	Thr	His	Pro	His	His	Arg	Ser	Ser	Ser	Cys	Pro
	370					375					380				
Asp	Ala	Pro	Ala	Thr	Cys	Asp	Trp	Asn	Asn	Tyr	Asn	Ser	Ala	Asn	Pro
385					390					395					400
Asn	Pro	His	Val	Leu	Tyr	Gly	Ala	Leu	Val	Gly	Gly	Pro	Asp	Ser	Asn
				405					410					415	
Asp	Asn	Tyr	Gln	Asp	Leu	Arg	Ser	Asp	Tyr	Val	Ala	Asn	Glu	Val	Ala
			420					425					430		
Thr	Asp	Tyr	Asn	Ala	Ala	Phe	Gln	Ser	Leu	Leu	Ala	Leu	Ile	Val	Asp
		435					440						445		
Leu	Gly	Leu													
	450														

<210> 35
 <211> 451
 <212> PRT
 <213> *Nasutitermes walkeri*

<400> 35

5

ES 2 621 181 T3

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15

Ala Leu Ala Ala Tyr Asp Tyr Lys Gln Val Leu Arg Asp Ser Leu Leu
 20 25 30

Phe Tyr Glu Ala Gln Arg Ser Gly Arg Leu Pro Ala Asp Gln Lys Val

ES 2 621 181 T3

	35		40		45													
Thr	Trp	Arg	Lys	Asp	Ser	Ala	Leu	Asn	Asp	Gln	Gly	Glu	Gln	Gly	Gln			
	50					55					60							
Asp	Leu	Thr	Gly	Gly	Tyr	Phe	Asp	Ala	Gly	Asp	Phe	Val	Lys	Phe	Gly			
	65				70					75					80			
Phe	Pro	Met	Ala	Tyr	Thr	Ala	Thr	Val	Leu	Ala	Trp	Gly	Leu	Ile	Asp			
				85					90					95				
Phe	Glu	Ala	Gly	Tyr	Ser	Ser	Ala	Gly	Ala	Leu	Asp	Asp	Gly	Arg	Lys			
			100					105					110					
Ala	Val	Lys	Trp	Ala	Thr	Asp	Tyr	Phe	Ile	Lys	Ala	His	Thr	Ser	Gln			
		115					120					125						
Asn	Glu	Phe	Tyr	Gly	Gln	Val	Gly	Gln	Gly	Asp	Val	Asp	His	Ala	Tyr			
	130					135					140							
Trp	Gly	Arg	Pro	Glu	Asp	Met	Thr	Met	Ala	Arg	Pro	Ala	Tyr	Lys	Ile			
	145				150					155					160			
Asp	Thr	Ser	Arg	Pro	Gly	Ser	Asp	Leu	Ala	Gly	Glu	Thr	Ala	Ala	Ala			
				165					170					175				
Leu	Ala	Ala	Ala	Ser	Ile	Val	Phe	Lys	Asn	Val	Asp	Gly	Thr	Tyr	Ser			
			180					185					190					
Asn	Asn	Leu	Leu	Thr	His	Ala	Arg	Gln	Leu	Phe	Asp	Phe	Ala	Asn	Asn			
		195					200					205						
Tyr	Arg	Gly	Lys	Tyr	Ser	Asp	Ser	Ile	Thr	Asp	Ala	Arg	Asn	Phe	Tyr			
	210					215					220							
Ala	Ser	Ala	Asp	Tyr	Arg	Asp	Glu	Leu	Val	Trp	Ala	Ala	Ala	Trp	Leu			
	225				230					235					240			
Tyr	Arg	Ala	Thr	Asn	Asp	Asn	Ser	Tyr	Leu	Asn	Thr	Ala	Glu	Ser	Leu			
				245					250					255				
Tyr	Asn	Glu	Phe	Gly	Leu	Gln	Asn	Trp	Gly	Gly	Gly	Leu	Asn	Trp	Asp			
			260					265					270					
Ser	Lys	Val	Ser	Gly	Val	Gln	Val	Leu	Leu	Ala	Lys	Leu	Thr	Asn	Lys			
		275					280					285						

ES 2 621 181 T3

Gln Glu Tyr Lys Asp Thr Ile Gln Ser Tyr Val Asn Tyr Leu Ile Asn
 290 295 300

Asn Gln Gln Lys Thr Pro Lys Gly Leu Leu Tyr Ile Asp Met Trp Gly
 305 310 315 320

Thr Leu Arg His Ala Ala Asn Ala Ala Phe Ile Met Leu Glu Ala Ala
 325 330 335

Asp Leu Gly Leu Ser Ala Ser Ser Tyr Arg Gln Phe Ala Gln Thr Gln
 340 345 350

Ile Asp Tyr Ala Leu Gly Asp Gly Gly Arg Ser Phe Val Cys Gly Phe
 355 360 365

Gly Ser Asn Pro Pro Thr Arg Pro His His Arg Ser Ser Ser Cys Pro
 370 375 380

Pro Ala Pro Ala Thr Cys Asp Trp Asn Thr Phe Asn Ser Pro Asp Pro
 385 390 395 400

Asn Tyr Asn Val Leu Ser Gly Ala Leu Val Gly Gly Pro Asp Gln Asn
 405 410 415

Asp Asn Tyr Val Asp Asp Arg Ser Asp Tyr Val His Asn Glu Val Ala
 420 425 430

Thr Asp Tyr Asn Ala Gly Phe Gln Ser Ala Leu Ala Ala Leu Val Ala
 435 440 445

Leu Gly Tyr
 450

- <210> 36
- <211> 451
- <212> PRT
- <213> *Reticulitermes speratus*

5

<400> 36

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
 1 5 10 15

Ala Leu Ala Ala Tyr Asp Tyr Lys Thr Val Leu Ser Asn Ser Leu Leu
 20 25 30

Phe Tyr Glu Ala Gln Arg Ser Gly Lys Leu Pro Ser Asp Gln Lys Val
 35 40 45

10

ES 2 621 181 T3

Thr Trp Arg Lys Asp Ser Ala Leu Asn Asp Lys Gly Gln Lys Gly Glu
50 55 60

Asp Leu Thr Gly Gly Tyr Tyr Asp Ala Gly Asp Phe Val Lys Phe Gly
65 70 75 80

Phe Pro Met Ala Tyr Thr Val Thr Val Leu Ala Trp Gly Val Ile Asp
85 90 95

Tyr Glu Ser Ala Tyr Ser Ala Ala Gly Ala Leu Asp Ser Gly Arg Lys
100 105 110

Ala Leu Lys Tyr Gly Thr Asp Tyr Phe Leu Lys Ala His Thr Ala Ala
115 120 125

Asn Glu Phe Tyr Gly Gln Val Gly Gln Gly Asp Val Asp His Ala Tyr
130 135 140

Trp Gly Arg Pro Glu Asp Met Thr Met Ser Arg Pro Ala Tyr Lys Ile
145 150 155 160

Asp Thr Ser Lys Pro Gly Ser Asp Leu Ala Ala Glu Thr Ala Ala Ala
165 170 175

Leu Ala Ala Thr Ala Ile Ala Tyr Lys Ser Ala Asp Ala Thr Tyr Ser
180 185 190

Asn Asn Leu Ile Thr His Ala Lys Gln Leu Phe Asp Phe Ala Asn Asn
195 200 205

Tyr Arg Gly Lys Tyr Ser Asp Ser Ile Thr Asp Ala Lys Asn Phe Tyr
210 215 220

Ala Ser Gly Asp Tyr Lys Asp Glu Leu Val Trp Ala Ala Ala Trp Leu
225 230 235 240

Tyr Arg Ala Thr Asn Asp Asn Thr Tyr Leu Thr Lys Ala Glu Ser Leu
245 250 255

Tyr Asn Glu Phe Gly Leu Gly Asn Phe Asn Gly Ala Phe Asn Trp Asp
260 265 270

Asn Lys Val Ser Gly Val Gln Val Leu Leu Ala Lys Leu Thr Ser Lys
275 280 285

Gln Val Tyr Lys Asp Lys Val Gln Ser Tyr Val Asp Tyr Leu Ile Ser
290 295 300

ES 2 621 181 T3

Ser Gln Lys Lys Thr Pro Lys Gly Leu Val Tyr Ile Asp Gln Trp Gly
305 310 315 320

Thr Leu Arg His Ala Ala Asn Ser Ala Leu Ile Ala Leu Gln Ala Ala
325 330 335

Asp Leu Gly Ile Asn Ala Ala Thr Tyr Arg Ala Tyr Ala Lys Lys Gln
340 345 350

Ile Asp Tyr Ala Leu Gly Asp Gly Gly Arg Ser Tyr Val Ile Gly Phe
355 360 365

Gly Thr Asn Pro Pro Val Arg Pro His His Arg Ser Ser Ser Cys Pro
370 375 380

Asp Ala Pro Ala Val Cys Asp Trp Asn Thr Tyr Asn Ser Ala Gly Pro
385 390 395 400

Asn Ala His Val Leu Thr Gly Ala Leu Val Gly Gly Pro Asp Ser Asn
405 410 415

Asp Ser Tyr Thr Asp Ala Arg Ser Asp Tyr Ile Ser Asn Glu Val Ala
420 425 430

Thr Asp Tyr Asn Ala Gly Phe Gln Ser Ala Val Ala Gly Leu Leu Lys
435 440 445

Ala Gly Val
450

- <210> 37
- <211> 408
- <212> PRT
- <213> *Neosartorya fischeri*

5

<400> 37

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15

Ala Leu Ala Gln Gln Ile Gly Ser Ile Ala Glu Asn His Pro Glu Leu
20 25 30

Thr Thr Tyr Arg Cys Ser Ser Gln Ala Gly Cys Val Ala Gln Ser Thr
35 40 45

Ser Val Val Leu Asp Ile Asn Ala His Trp Ile His Gln Asn Gly Ala
50 55 60

10

ES 2 621 181 T3

Gln Thr Ser Cys Thr Thr Ser Ser Gly Leu Asp Pro Ser Leu Cys Pro
65 70 75 80

Asp Lys Val Thr Cys Ser Gln Asn Cys Val Val Glu Gly Ile Thr Asp
85 90 95

Tyr Ser Ser Phe Gly Val Gln Asn Ser Gly Asp Ala Met Thr Leu Arg
100 105 110

Gln Tyr Gln Val Gln Asn Gly Gln Ile Lys Thr Leu Arg Pro Arg Val
115 120 125

Tyr Leu Leu Ala Glu Asp Gly Ile Asn Tyr Ser Lys Leu Gln Leu Leu
130 135 140

Asn Gln Glu Phe Thr Phe Asp Val Asp Ala Ser Lys Leu Pro Cys Gly
145 150 155 160

Met Asn Gly Ala Leu Tyr Leu Ser Glu Met Asp Ala Ser Gly Gly Arg
165 170 175

Ser Ala Leu Asn Pro Ala Gly Ala Thr Tyr Gly Thr Gly Tyr Cys Asp
180 185 190

Ala Gln Cys Phe Asn Pro Gly Pro Trp Ile Asn Gly Glu Ala Asn Thr
195 200 205

Ala Gly Ala Gly Ala Cys Cys Gln Glu Met Asp Leu Trp Glu Ala Asn
210 215 220

Ser Arg Ser Thr Ile Phe Ser Pro His Pro Cys Thr Thr Ala Gly Leu
225 230 235 240

Tyr Ala Cys Thr Gly Ala Glu Cys Tyr Ser Ile Cys Asp Gly Tyr Gly
245 250 255

Cys Thr Tyr Asn Pro Tyr Glu Leu Gly Ala Lys Asp Tyr Tyr Gly Tyr
260 265 270

Gly Leu Thr Ile Asp Thr Ala Lys Pro Ile Thr Val Val Thr Gln Phe
275 280 285

Met Thr Ala Asp Asn Thr Ala Thr Gly Thr Leu Ala Glu Ile Arg Arg
290 295 300

Leu Tyr Val Gln Asp Gly Lys Val Ile Gly Asn Thr Ala Val Ala Met
305 310 315 320

Thr Glu Ala Phe Cys Ser Ser Ser Arg Thr Phe Glu Glu Leu Gly Gly
 325 330 335

Leu Gln Arg Met Gly Glu Ala Leu Gly Arg Gly Met Val Pro Val Phe
 340 345 350

Ser Ile Trp Asp Asp Pro Gly Leu Trp Met His Trp Leu Asp Ser Asp
 355 360 365

Gly Ala Gly Pro Cys Gly Asn Thr Glu Gly Asp Pro Ala Phe Ile Gln
 370 375 380

Ala Asn Tyr Pro Asn Thr Ala Val Thr Phe Ser Lys Val Arg Trp Gly
 385 390 395 400

Asp Ile Gly Ser Thr Tyr Ser Ser
 405

<210> 38
 <211> 304
 <212> PRT
 <213> *Reticulitermes flavipes*
 <400> 38

5

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
 1 5 10 15

Ala Leu Ala Gln Trp Met Gln Ile Gly Gly Lys Gln Lys Tyr Pro Ala
 20 25 30

Phe Lys Pro Gly Ala Lys Tyr Gly Arg Gly Tyr Cys Asp Gly Gln Cys
 35 40 45

Pro His Asp Met Lys Val Ser Ser Gly Arg Ala Asn Val Asp Gly Trp
 50 55 60

Lys Pro Gln Asp Asn Asp Glu Asn Ser Gly Asn Gly Lys Leu Gly Thr
 65 70 75 80

Cys Cys Trp Glu Met Asp Ile Trp Glu Gly Asn Leu Val Ser Gln Ala
 85 90 95

Tyr Thr Val His Ala Gly Ser Lys Ser Gly Gln Tyr Glu Cys Thr Gly
 100 105 110

Thr Gln Cys Gly Asp Thr Asp Ser Gly Glu Arg Phe Lys Gly Thr Cys
 115 120 125

10

ES 2 621 181 T3

Asp Lys Asp Gly Cys Asp Phe Ala Ser Tyr Arg Trp Gly Ala Thr Asp
 130 135 140

Tyr Tyr Gly Pro Gly Lys Thr Val Asp Thr Lys Gln Pro Met Thr Val
 145 150 155 160

Val Thr Gln Phe Ile Gly Asp Pro Leu Thr Glu Ile Lys Arg Val Tyr
 165 170 175

Val Gln Gly Gly Lys Val Ile Asn Asn Ser Lys Thr Ser Asn Leu Gly
 180 185 190

Ser Val Tyr Asp Ser Leu Thr Glu Ala Phe Cys Asp Asp Thr Lys Gln
 195 200 205

Val Thr Gly Asp Thr Asn Asp Phe Lys Ala Lys Gly Gly Met Ser Gly
 210 215 220

Phe Ser Lys Asn Leu Asp Thr Pro Gln Val Leu Val Met Ser Leu Trp
 225 230 235 240

Asp Asp His Thr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr
 245 250 255

Asp Ser Thr Lys Pro Gly Ala Ala Arg Gly Thr Cys Ala Val Thr Ser
 260 265 270

Gly Asp Pro Lys Asp Val Glu Ser Lys Gln Ala Asn Ser Gln Val Val
 275 280 285

Tyr Ser Asp Ile Lys Phe Gly Pro Ile Asn Ser Thr Tyr Lys Ala Asn
 290 295 300

<210> 39
 <211> 475
 <212> PRT
 <213> *Trichoderma reesei*
 <400> 39

5

Met Val Ser Phe Thr Ser Leu Leu Ala Gly Val Ala Ala Ile Ser Gly
 1 5 10 15

Val Leu Ala Ala Pro Ala Ala Glu Val Glu Pro Val Ala Val Glu Lys
 20 25 30

Arg Glu Ala Glu Ala Glu Ala Gln Gln Pro Gly Thr Ser Thr Pro Glu
 35 40 45

10

ES 2 621 181 T3

Val His Pro Lys Leu Thr Thr Tyr Lys Cys Thr Lys Ser Gly Gly Cys
50 55 60

Val Ala Gln Asp Thr Ser Val Val Leu Asp Trp Asn Tyr Arg Trp Met
65 70 75 80

His Asp Ala Asn Tyr Asn Ser Cys Thr Val Asn Gly Gly Val Asn Thr
85 90 95

Thr Leu Cys Pro Asp Glu Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu
100 105 110

Gly Val Asp Tyr Ala Ala Ser Gly Val Thr Thr Ser Gly Ser Ser Leu
115 120 125

Thr Met Asn Gln Tyr Met Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val
130 135 140

Ser Pro Arg Leu Tyr Leu Leu Asp Ser Asp Gly Glu Tyr Val Met Leu
145 150 155 160

Lys Leu Asn Gly Gln Glu Leu Ser Phe Asp Val Asp Leu Ser Ala Leu
165 170 175

Pro Cys Gly Glu Asn Gly Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn
180 185 190

Gly Gly Ala Asn Gln Tyr Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly
195 200 205

Tyr Cys Asp Ala Gln Cys Pro Val Gln Thr Trp Arg Asn Gly Thr Leu
210 215 220

Asn Thr Ser His Gln Gly Phe Cys Cys Asn Glu Met Asp Ile Leu Glu
225 230 235 240

Gly Asn Ser Arg Ala Asn Ala Leu Thr Pro His Ser Cys Thr Ala Thr
245 250 255

Ala Cys Asp Ser Ala Gly Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr
260 265 270

Lys Ser Tyr Tyr Gly Pro Gly Asp Thr Val Asp Thr Ser Lys Thr Phe
275 280 285

Thr Ile Ile Thr Gln Phe Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn

ES 2 621 181 T3

290	295	300																
Leu	Val	Ser	Ile	Thr	Arg	Lys	Tyr	Gln	Gln	Asn	Gly	Val	Asp	Ile	Pro			
305					310					315					320			
Ser	Ala	Gln	Pro	Gly	Gly	Asp	Thr	Ile	Ser	Ser	Cys	Pro	Ser	Ala	Ser			
				325					330					335				
Ala	Tyr	Gly	Gly	Leu	Ala	Thr	Met	Gly	Lys	Ala	Leu	Ser	Ser	Gly	Met			
			340					345					350					
Val	Leu	Val	Phe	Ser	Ile	Trp	Asn	Asp	Asn	Ser	Gln	Tyr	Met	Asn	Trp			
		355					360					365						
Leu	Asp	Ser	Gly	Asn	Ala	Gly	Pro	Cys	Ser	Ser	Thr	Glu	Gly	Asn	Pro			
	370					375						380						
Ser	Asn	Ile	Leu	Ala	Asn	Asn	Pro	Asn	Thr	His	Val	Val	Phe	Ser	Asn			
385					390					395					400			
Ile	Arg	Trp	Gly	Asp	Ile	Gly	Ser	Thr	Thr	Asn	Ser	Thr	Ala	Pro	Pro			
				405						410					415			
Pro	Pro	Pro	Ala	Ser	Ser	Thr	Thr	Phe	Ser	Thr	Thr	Arg	Arg	Ser	Ser			
			420					425					430					
Thr	Thr	Ser	Ser	Ser	Pro	Ser	Cys	Thr	Gln	Thr	His	Trp	Gly	Gln	Cys			
		435					440					445						
Gly	Gly	Ile	Gly	Tyr	Ser	Gly	Cys	Lys	Thr	Cys	Thr	Ser	Gly	Thr	Thr			
	450					455						460						
Cys	Gln	Tyr	Ser	Asn	Asp	Tyr	Tyr	Ser	Gln	Cys								
465					470					475								

<210> 40
 <211> 896
 <212> PRT
 <213> *Saccharomycopsis fibuligera*

5

<400> 40

ES 2 621 181 T3

Met Val Ser Phe Thr Ser Leu Leu Ala Gly Val Ala Ala Ile Ser Gly
1 5 10 15

Val Leu Ala Ala Pro Ala Ala Glu Val Glu Ser Val Ala Val Glu Lys
 20 25 30

Arg Ser Asp Ser Arg Val Pro Ile Gln Asn Tyr Thr Gln Ser Pro Ser

ES 2 621 181 T3

Thr Tyr Ala Cys Glu Asn Ser Tyr Met Met Asn His Leu Leu Lys Glu
 290 295 300

Glu Leu Gly Phe Gln Gly Phe Val Val Ser Asp Trp Gly Ala Gln Leu
 305 310 315

Ser Gly Val Tyr Ser Ala Ile Ser Gly Leu Asp Met Ser Met Pro Gly
 325 330 335

Glu Val Tyr Gly Gly Trp Asn Thr Gly Thr Ser Phe Trp Gly Gln Asn
 340 345 350

Leu Thr Lys Ala Ile Tyr Asn Glu Thr Val Pro Ile Glu Arg Leu Asp
 355 360 365

Asp Met Ala Thr Arg Ile Leu Ala Ala Leu Tyr Ala Thr Asn Ser Phe
 370 375 380

Pro Thr Glu Asp His Leu Pro Asn Phe Ser Ser Trp Thr Thr Lys Glu
 385 390 395 400

Tyr Gly Asn Lys Tyr Tyr Ala Asp Asn Thr Thr Glu Ile Val Lys Val
 405 410 415

Asn Tyr His Val Asp Pro Ser Asn Asp Phe Thr Glu Asp Thr Ala Leu
 420 425 430

Lys Val Ala Glu Glu Ser Ile Val Leu Leu Lys Asn Glu Asn Asn Thr
 435 440 445

Leu Pro Ile Ser Pro Glu Lys Ala Lys Arg Leu Leu Leu Ser Gly Ile
 450 455 460

Ala Ala Gly Pro Asp Pro Ile Gly Tyr Gln Cys Glu Asp Gln Ser Cys
 465 470 475 480

Thr Asn Gly Ala Leu Phe Gln Gly Trp Gly Ser Gly Ser Val Gly Ser
 485 490 495

Pro Lys Tyr Gln Val Thr Pro Phe Glu Glu Ile Ser Tyr Leu Ala Arg
 500 505 510

Lys Asn Lys Met Gln Phe Asp Tyr Ile Arg Glu Ser Tyr Asp Leu Ala
 515 520 525

Gln Val Thr Lys Val Ala Ser Asp Ala His Leu Ser Ile Val Val Val
 530 535 540

ES 2 621 181 T3

Ser Ala Ala Ser Gly Glu Gly Tyr Ile Thr Val Asp Gly Asn Gln Gly
545 550 555 560

Asp Arg Arg Asn Leu Thr Leu Trp Asn Asn Gly Asp Lys Leu Ile Glu
565 570 575

Thr Val Ala Glu Asn Cys Ala Asn Thr Val Val Val Val Thr Ser Thr
580 585 590

Gly Gln Ile Asn Phe Glu Gly Phe Ala Asp His Pro Asn Val Thr Ala
595 600 605

Ile Val Trp Ala Gly Pro Leu Gly Asp Arg Ser Gly Thr Ala Ile Ala
610 615 620

Asn Ile Leu Phe Gly Lys Ala Asn Pro Ser Gly His Leu Pro Phe Thr
625 630 635 640

Ile Ala Lys Thr Asp Asp Asp Tyr Ile Pro Ile Glu Thr Tyr Ser Pro
645 650 655

Ser Ser Gly Glu Pro Glu Asp Asn His Leu Val Glu Asn Asp Leu Leu
660 665 670

Val Asp Tyr Arg Tyr Phe Glu Glu Lys Asn Ile Glu Pro Arg Tyr Ala
675 680 685

Phe Gly Tyr Gly Leu Ser Tyr Asn Glu Tyr Glu Val Ser Asn Ala Lys
690 695 700

Val Ser Ala Ala Lys Lys Val Asp Glu Glu Leu Pro Glu Pro Ala Thr
705 710 715 720

Tyr Leu Ser Glu Phe Ser Tyr Gln Asn Ala Lys Asp Ser Lys Asn Pro
725 730 735

Ser Asp Ala Phe Ala Pro Thr Asp Leu Asn Arg Val Asn Glu Tyr Leu
740 745 750

Tyr Pro Tyr Leu Asp Ser Asn Val Thr Leu Lys Asp Gly Asn Tyr Glu
755 760 765

Tyr Pro Asp Gly Tyr Ser Thr Glu Gln Arg Thr Thr Pro Ile Gln Pro
770 775 780

Gly Gly Gly Leu Gly Gly Asn Asp Ala Leu Trp Glu Val Ala Tyr Lys
785 790 795 800

ES 2 621 181 T3

Val Glu Val Asp Val Gln Asn Leu Gly Asn Ser Thr Asp Lys Phe Val
 805 810 815

Pro Gln Leu Tyr Leu Lys His Pro Glu Asp Gly Lys Phe Glu Thr Pro
 820 825 830

Ile Gln Leu Arg Gly Phe Glu Lys Val Glu Leu Ser Pro Gly Glu Lys
 835 840 845

Lys Thr Val Glu Phe Glu Leu Leu Arg Arg Asp Leu Ser Val Trp Asp
 850 855 860

Thr Thr Arg Gln Ser Trp Ile Val Glu Ser Gly Thr Tyr Glu Ala Leu
 865 870 875 880

Ile Gly Val Ala Val Asn Asp Ile Lys Thr Ser Val Leu Phe Thr Ile
 885 890 895

5 <210> 41
 <211> 144
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> fusion con enlazador flexible
 <400> 41

ggaggagggtg gttcaggagg tgggtgggtct gcttggcctc cacaatttgg aggaggcggt 60
 ggtgaaaatc tgtatttcca gggaggcgga ggtgattaca aggatgacga caaaggaggt 120
 ggtggatcag gaggtggtgg ctcc 144

15 <210> 42
 <211> 117
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> fusion con enlazador flexible
 <400> 42

ggtggcggtg gatctggagg aggcggttct tggctctcacc cacaatttga aaagggtgga 60
 gaaaacttgt actttcaagg cgggtggtgga ggttctggcg gaggtggctc cggtca 117

25 <210> 43
 <211> 455
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> secuencia consenso de CBH1
 35 <400> 43

ES 2 621 181 T3

Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1 5 10 15

Ala Leu Ala Gln Gln Ala Gly Thr Leu Thr Ala Glu Thr His Pro Ser
20 25 30

Leu Thr Trp Gln Lys Cys Thr Ser Gly Gly Ser Cys Thr Thr Val Asn
35 40 45

Gly Ser Val Val Ile Asp Ala Asn Trp Arg Trp Val His Ala Thr Ser
50 55 60

Gly Ser Thr Asn Cys Tyr Thr Gly Asn Thr Trp Asp Thr Thr Leu Cys
65 70 75 80

Pro Asp Asp Val Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp
85 90 95

Tyr Ser Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Arg Leu
100 105 110

Asn Phe Val Thr Gln Gly Ser Gln Lys Asn Val Gly Ser Arg Leu Tyr
115 120 125

Leu Met Glu Asp Asp Thr Thr Tyr Gln Met Phe Lys Leu Leu Gly Gln
130 135 140

Glu Phe Thr Phe Asp Val Asp Val Ser Asn Leu Pro Cys Gly Leu Asn
145 150 155 160

Gly Ala Leu Tyr Phe Val Ala Met Asp Ala Asp Gly Gly Met Ser Lys
165 170 175

Tyr Pro Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp
180 185 190

Ser Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val
195 200 205

Glu Gly Trp Glu Pro Ser Ser Asn Asp Ala Asn Ala Gly Ile Gly Asn
210 215 220

His Gly Ser Cys Cys Ala Glu Met Asp Ile Trp Glu Ala Asn Ser Ile
225 230 235 240

ES 2 621 181 T3

Ser Thr Ala Phe Thr Pro His Pro Cys Asp Thr Ile Gly Gln Thr Met
 245 250 255

Cys Glu Gly Asp Ser Cys Gly Gly Thr Tyr Ser Ser Asp Arg Tyr Gly
 260 265 270

Gly Thr Cys Asp Pro Asp Gly Cys Asp Phe Asn Pro Tyr Arg Met Gly
 275 280 285

Asn Lys Thr Phe Tyr Gly Pro Gly Lys Thr Val Asp Thr Thr Lys Lys
 290 295 300

Val Thr Val Val Thr Gln Phe Ile Thr Gly Ser Ser Gly Thr Leu Ser
 305 310 315 320

Glu Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Asn Ser
 325 330 335

Glu Ser Thr Ile Ser Gly Val Ser Gly Asn Ser Ile Thr Thr Asp Phe
 340 345 350

Cys Thr Ala Gln Lys Thr Ala Phe Gly Asp Thr Asp Asp Phe Ala Lys
 355 360 365

Lys Gly Gly Leu Glu Gly Met Gly Lys Ala Leu Ala Gln Gly Met Val
 370 375 380

Leu Val Met Ser Leu Trp Asp Asp His Ala Ala Asn Met Leu Trp Leu
 385 390 395 400

Asp Ser Thr Tyr Pro Thr Asp Ala Thr Ser Ser Thr Pro Gly Ala Ala
 405 410 415

Arg Gly Ser Cys Asp Thr Ser Ser Gly Val Pro Ala Asp Val Glu Ala
 420 425 430

Asn Ser Pro Asn Ser Tyr Val Thr Phe Ser Asn Ile Lys Phe Gly Pro
 435 440 445

Ile Gly Ser Thr Phe Thr Gly
 450 455

<210> 44
 <211> 1368
 <212> ADN
 <213> Secuencia artificial
 <220>

ES 2 621 181 T3

<223> secuencia consenso de CBH1 con codones optimizados

<400> 44

```

atgagatttc cttcaatctt cactgctggt ttgttcgcag cctcaagtgc ttagcacaaa      60
caggccggaa cattgacagc agaaactcat ccttccttaa cctggcaaaa gtgcacttct      120
ggaggttcat gcactacagt gaatggatct gtcgtgatcg atgcaactg gagatgggtt      180
cacgcaactt caggttctac caactggtat accggaaaca cttgggacac cacattgtgc      240
ccagatgacg tcacgtgcgc tcagaactgt gctttggatg gagctgatta cagttcaacc      300
tatggtgtaa ctacatccgg aaactctttg agattaaact tcgttactca aggaagtcaa      360
aagaacgttg gttctagatt gtacttaatg gaggacgata caacctatca aatgttcaaa      420
ttgttaggtc aggagttcac ctttgacgta gatgtcagta acttgccatg tgggttaaac      480
ggagctttat actttgtggc aatggatgct gacgggtgaa tgtccaagta tccaggaaac      540
aaagccggtg caaagtacgg tacaggatat tgtgattcac agtgccttag agatttgaag      600
ttcattaacg gtcaagcaaa tgtggagggt tgggaaccat ctagtaacga tgccaatgcg      660
ggtattggta atcatgggtc ctggttgcgt gagatggata tctgggaggc caactcaata      720
tctactgcct ttaccctca cccatgcat acaattggtc aaactatgtg cgagggtgat      780
tcatgtggtg gaacctactc ctctgataga tacggaggta catgcatcc agatggttgc      840
gactttaatc catacagaat gggaaacaaa accttttacg gtcctggaaa gacagttgat      900
actaccaaga aagtaacagt cgtgaccocag tttatcaccg gtagttctgg aaccttatcc      960
gaaatcaaaa gattctacgt tcagaacggt aaagtaattc caaacagtga atctacaatt     1020
tcaggagtga gtggttaattc tattactacc gacttttgta cagctcagaa aacagcattt     1080
ggtgacaccg atgactttgc taagaagggt ggattagaag gtatgggtaa agctttggcc     1140
cagggaatgg tgtagttat gtctttatgg gatgatcacg ccgcaaatat gttatggttg     1200
gattcaacat atccaactga tgccacaagt agtacacctg gagctgccag aggttcttgt     1260
gatacatctt ccggtgttcc agccgatgta gaagcaaatt ctcttaactc ctatgttacc     1320
ttctccaata taaagtttgg tccaatcggg tcaacattca ctgggtaa                       1368

```

5

<210> 45

<211> 1572

<212> ADN

<213> *Acremonium thermophilum*

10

<400> 45

ES 2 621 181 T3

atgtatacca aatttgctgc attggccgct ttagttgcaa cagtaagagg tcaagccgct 60
 tgttctctaa ccgcagaaac tcacccatct ctacaatggc agaaatgcac agcccctgga 120
 tcttgtacaa ctgtctcogg ccaagtcacc attgacgcta attggagatg gcttcaccaa 180
 actaactctt caacgaattg ttataccggt aacgaatggg atacttccat atgttcatcc 240

 gatacagact gcgcaacgaa atgttgttta gatggagcag actatacggg aacttatggt 300
 gttacagcct caggaatc cctaaacctt aagttcgtaa ctcaaggacc atatagtaag 360
 aatatcggct ctagaatgta cttgatggaa agtgagagca aatatcaggg ttttacgta 420
 ttgggacaag agtttacatt tgatgttgat gtgagtaact taggttgctg cctaaacggc 480
 gccttgtact tcgtttctat ggatcttgat ggaggtgtat caaatacac gaccaacaag 540
 gctggagcca aatatggtac gggatattgt gacagccaat gccctagaga cttaaagttc 600
 attaacggtc aggcaaatat tgacggctgg caaccaagca gtaacgacgc taatgccgga 660
 ctaggtaacc atggctcatg ttgttcogaa atggatatct gggaagccaa taaggtgtcc 720
 gctgcctaca cccccatcc atgcacgaca atcggtcaga caatgtgtac cggatgatgac 780
 tgtggaggca catactcaag tgataggtac gccggtatat gtgatcctga cggttgcgat 840
 ttcaactctt atagaatggg agatacatcc ttttacggcc ccggtaaaac agttgatacg 900
 ggtagtaagt tcactgttgt tactcagttc ttaacaggtt cagacggcaa tcttagtgaa 960
 atcaaaagat tctacgttca gaatggaaa gtcattccta attccgagag taagattgct 1020
 ggtgtgtctg gtaacagtat cacgaccgac ttctgtaccg cccaaaagac tgcctttgga 1080
 gatacgaatg ttttcgagga aagggcggt cttgctcaaa tgggcaaggc tttggccgaa 1140
 ccaatggtat tagtcctatc cgtttgggat gatcatgcag tgaatatgct ttggcttgat 1200
 agcacctacc ctactgacag caccaagcca ggagctgcca gaggtgactg tcctatcaca 1260
 agtggcgttc cagcagatgt agagagccaa gctccaaact ccaatgtgat ctattctaac 1320
 atcagatttg gcccattaa tagtacctat acaggaacgc cctctggtgg taaccctcca 1380
 gggcgaggca ccacaactac cacgaccaca acgacttcaa agccttctgg ccctacgaca 1440
 actaccaatc cttccggacc acagcaaact cactggggtc agtgtggagg ccaaggatgg 1500
 acgggtccta ccgtgtgtca atcaccttac acatgcaaat acagtaatga ctggtactct 1560
 cagtgtttat aa 1572

5

<210> 46
 <211> 523
 <212> PRT
 <213> *Acremonium thermophilum*

<400> 46

ES 2 621 181 T3

Met Tyr Thr Lys Phe Ala Ala Leu Ala Ala Leu Val Ala Thr Val Arg
1 5 10 15

Gly Gln Ala Ala Cys Ser Leu Thr Ala Glu Thr His Pro Ser Leu Gln
20 25 30

ES 2 621 181 T3

Trp Gln Lys Cys Thr Ala Pro Gly Ser Cys Thr Thr Val Ser Gly Gln
 35 40 45
 Val Thr Ile Asp Ala Asn Trp Arg Trp Leu His Gln Thr Asn Ser Ser
 50 55 60
 Thr Asn Cys Tyr Thr Gly Asn Glu Trp Asp Thr Ser Ile Cys Ser Ser
 65 70 75 80
 Asp Thr Asp Cys Ala Thr Lys Cys Cys Leu Asp Gly Ala Asp Tyr Thr
 85 90 95
 Gly Thr Tyr Gly Val Thr Ala Ser Gly Asn Ser Leu Asn Leu Lys Phe
 100 105 110
 Val Thr Gln Gly Pro Tyr Ser Lys Asn Ile Gly Ser Arg Met Tyr Leu
 115 120 125
 Met Glu Ser Glu Ser Lys Tyr Gln Gly Phe Thr Leu Leu Gly Gln Glu
 130 135 140
 Phe Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn Gly
 145 150 155 160
 Ala Leu Tyr Phe Val Ser Met Asp Leu Asp Gly Gly Val Ser Lys Tyr
 165 170 175
 Thr Thr Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser
 180 185 190
 Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Ile Asp
 195 200 205
 Gly Trp Gln Pro Ser Ser Asn Asp Ala Asn Ala Gly Leu Gly Asn His
 210 215 220
 Gly Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Lys Val Ser
 225 230 235 240
 Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Thr Met Cys
 245 250 255
 Thr Gly Asp Asp Cys Gly Gly Thr Tyr Ser Ser Asp Arg Tyr Ala Gly
 260 265 270
 Ile Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asp
 275 280 285

ES 2 621 181 T3

Thr Ser Phe Tyr Gly Pro Gly Lys Thr Val Asp Thr Gly Ser Lys Phe
 290 295 300
 Thr Val Val Thr Gln Phe Leu Thr Gly Ser Asp Gly Asn Leu Ser Glu
 305 310 315 320
 Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Asn Ser Glu
 325 330 335
 Ser Lys Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Thr Asp Phe Cys
 340 345 350
 Thr Ala Gln Lys Thr Ala Phe Gly Asp Thr Asn Val Phe Glu Glu Arg
 355 360 365
 Gly Gly Leu Ala Gln Met Gly Lys Ala Leu Ala Glu Pro Met Val Leu
 370 375 380
 Val Leu Ser Val Trp Asp Asp His Ala Val Asn Met Leu Trp Leu Asp
 385 390 395 400
 Ser Thr Tyr Pro Thr Asp Ser Thr Lys Pro Gly Ala Ala Arg Gly Asp
 405 410 415
 Cys Pro Ile Thr Ser Gly Val Pro Ala Asp Val Glu Ser Gln Ala Pro
 420 425 430
 Asn Ser Asn Val Ile Tyr Ser Asn Ile Arg Phe Gly Pro Ile Asn Ser
 435 440 445
 Thr Tyr Thr Gly Thr Pro Ser Gly Gly Asn Pro Pro Gly Gly Gly Thr
 450 455 460
 Thr Thr Thr Thr Thr Thr Thr Thr Ser Lys Pro Ser Gly Pro Thr Thr
 465 470 475 480
 Thr Thr Asn Pro Ser Gly Pro Gln Gln Thr His Trp Gly Gln Cys Gly
 485 490 495
 Gly Gln Gly Trp Thr Gly Pro Thr Val Cys Gln Ser Pro Tyr Thr Cys
 500 505 510
 Lys Tyr Ser Asn Asp Trp Tyr Ser Gln Cys Leu
 515 520

<210> 47
<211> 1586

ES 2 621 181 T3

<212> ADN

<213> *Aspergillus kawachii*

<400> 47

5

```

ttaattaaaa tgagaatttc taacttgatt gttgctgctt ctgctgctac tatggtttct      60
gctttgccat ctagacaaat gaaaaagagg gattctgggt ttaaatgggt tgggtacttct      120
gaatctgggt ctgaatttgg ttctgcttta ccagggtactt tgggtactga ttatacttgg      180
ccagaaactt ctaaaattca agttttgaga aacaagggtg tgaacatttt tagaatacca      240
ttcttgatgg aaagattaac tccagatggt ttgactgggt cttttgcttc tacttacttg      300
tctgatttga agtcaactgt tgaatttgtt actaattctg gtgcttatgc tgttttagat      360
ccacataatt acggtagatt cgatggttct attattgaat ctacttctga ttttaagact      420
tgggtgaaaa atgttgctac tgaatttgtt gataacgata aggttatttt cgatacaaac      480
aacgaatata atgatatgga acaatctttg gttttgaatt tgaaccaagc tgctattaat      540
ggatttagag ctgctgggtc tactactcaa tacattttcg ttgaaggtaa tgcttatact      600
gggtgcttggg attggactac ttacaatgat gatttgtctg gtttaactga ttctgaagat      660
aagataatat acgaaatgca tcaatacttg gattctgatt cttctggtac atctgaaact      720
tgtgtttctt ctactattgg taaagaaaga attgaaaagg ctactgaatg gttgaaaact      780
aacaacaagc aaggatttat tgggtgaattt gcagggtggt ttaattctgt ttgtgaagag      840
gctgttgaag gaatgttggc ttatatgtct gaaaattctg atgtttgggt tgggtgcttct      900
tgggtggtctg ctggtccatg gtggggtact tacatgtatt ctttgaacc aactgatggt      960
actgcttatt ctacttattt gccaattttg gaaaaatact tcccatctgg tgatgcttca     1020
tcacttcat ctgcttcagc ttcagttgca gccgctactt ctgctgttcc tactactact     1080
acagctgcat ttgaacaaac tactactcca gctactcaag ttgaaattgc ttcttcttca     1140
tcttcatcat cagctgttgc tgcttcacaa actactttgt ctaagggtta gtctaaatct     1200
aaatctccat gtaaattgtc atctgctact tcactctgctg tttcatcagc tgctgcagtt     1260
actacacctg cagttgcagc tacaactcca gctgctgctc caacttcttc ttctgttgct     1320
tttgctacta cttctgttta cgttccaact actactgctg ctgcaccatc tcaagtttca     1380
tcttcagctg cagcttcate ttcaggtgtt gttggtggtt ctgatccaca aggtccatct     1440
gctactaatt ctgctggtga agttaatcaa tattaccaat gtggtggtat taattggact     1500
gggtccaactg tttgtgcttc tccatatact tgtaagggtc aaaacgatta ctactatcaa     1560
tgtgttgctg aattataagg cgcgcc

```

<210> 48

<211> 1280

<212> ADN

<213> *Heterodera schachtii*

<400> 48

10

ES 2 621 181 T3

ttaattaa	tgcaattgggc	tgatgttgct	tgttctagac	caccatggcc	aagagattct	60
gttaaagctt	tgaagtgtaa	ttggaacgct	aatgttatta	gaggtgctat	gggtgttgat	120
gaaggtgggtt	atttgtctga	tgctaatact	gcttacaatt	tgatggttgc	tgttattgaa	180
gctgctat	ctaatggat	ctacgttatt	gttgattggc	atgctcataa	tgctcatcca	240
gatgaagctg	ttaaattcct	tactagaatt	gctcaagctt	atggttctta	cttgcattat	300
ttgtacgaag	atttcaatga	accattggat	gtttcttggg	ctgatgtttt	ggttccatac	360
cataaaaaag	ttattgctgc	cattagagct	attgataaga	agaacgttat	tatcttgggt	420
actocaaaat	ggtcacaaga	tgttgatggt	gcttctcaaa	atccaattaa	ggattaccaa	480
aaacttgatgt	acactttgca	tttttacgct	tcactctcatt	ttacatctga	tttgggtgct	540
aaattgaaaa	ctgctgttaa	caatggtttg	ccagtttttg	ttactgaata	tggtacttgt	600
gaagcttctg	gtaatggtaa	tttgaatact	gattctatgt	catcttgggtg	gactttggtg	660
gattctttga	aaatttctta	cgctaattgg	gctatttctg	ataaatctga	agcttgttct	720
gctttgtctc	caggtactac	tgctgttaat	gttgggtggtt	cttctagatg	gacttcttct	780
ggtaaatatgg	ttgcttctta	ctacaaaaaa	aagtcactctg	gtatttcttg	ttctggtagt	840
tcttcagggtt	cttcaagtgg	ttcatctagt	ggttcttccg	gtacatcttc	tggttctagt	900
ggttcatcta	gtggtagttc	ttccggtagt	tctagtggta	gttctgggtc	aagttctgggt	960
tctcctctg	gttctgggtc	tgcatctatt	tctgttggtc	catctaatac	ttggaatgggt	1020
gggtgtagag	ttaattttga	aattaagaac	actggttctg	ttcattgtg	tggtgttggt	1080
tttctgttt	ctttgccatc	tggtactact	ttgggtggtt	cttgggaatat	ggaatctgct	1140
ggttctgggtc	aatattcttt	accatcttgg	gttagaattg	aagctggtaa	atcttctaaa	1200
gatgctgggtt	tgacttttaa	tggtaaagat	aagccaactg	ctaaaattgt	taccaccaag	1260
aagtgcttat	aaggcgcgcc					1280

<210> 49
 <211> 1277
 <212> ADN
 <213> *Hypocrea jecorina*
 <400> 49

5

ES 2 621 181 T3

ttaattaa	aaa	tgaacaagtc	tgttgctcca	ttggtggtg	ctgcttctat	tttgtatggt	60
ggtgctg	ttg	ctcaacaaac	tgtttgggg	caatgtggtg	gtattggtg	gtctggtcca	120
actaattg	tg	ctccaggctc	tgcttgctt	actttgaatc	catattatgc	tcaatgtatt	180
ccagggtg	ct	ctactattac	tacttctact	agaccacat	ctggtccaac	aactactact	240
agagctact	tt	ctacatcttc	ttctactcca	ccaacttc	cat	ctggtggttag	300
gtaaacatt	g	ctggttttga	ttttggtt	gt	actactgatg	gtacttgtgt	360
tacttctaaa	g	gtttaccac	cattgaaaa	tttcactggt	tctaacaatt	atccagatgg	420
tattggtcaa	g	atgcaacatt	ttgtaacga	agatggtatg	actattttta	gattgccagt	480
tggttgca	g	tatttggtta	acaacaattt	gggtggtaat	ttggattcta	cttctatttc	540
taagtacgat	g	caattggttc	aaggttggtt	gtctttgggt	gcttactgta	ttggtgat	600
tcataattat	g	gctagatgga	atggtggtat	tattggtcaa	ggtggtccaa	caaatgctca	660
atttacttct	g	ttgtggtcac	aattggcttc	aaaatatgct	tctcaatcta	gagtttggtt	720
tggtattatg	g	aatgaaccac	atgatgtaa	cattaatact	tgggctgcta	ctggtcaaga	780
agttgttact	g	gctattagaa	atgctggtgc	tacttctcaa	ttcatttctt	tgccaggtaa	840
tgattggcaa	g	tctgctggtg	cttttatttc	tgatggttct	gctgctgctt	tgtctcaagt	900
tactaatcca	g	gatggttcta	ctactaattt	gatcttcgat	gttcataagt	acttggttc	960
tgataattct	g	ggtactcatg	ctgaatgtac	tacaaacaat	attgatggtg	ctttttctcc	1020
attggctact	g	tggttgagac	aaaacaatag	acaagctatt	ttgactgaaa	ctggtggtgg	1080
taatgttcaa	g	tcttgatatcc	aagatatgtg	ccaacaaatt	caataactga	accaaaattc	1140
tgatgtttat	g	ttgggttacg	ttggttggg	tgctggttct	tttgattcta	cttacgtttt	1200
aactgaaact	g	ccaacttctt	ctggtaattc	ttggaactgat	acttctttgg	tttcttcatg	1260
tttggtctaga	g	aagttataag	gcgcgcc				1277

<210> 50
 <211> 1436
 <212> ADN
 <213> *Orpinomyces* sp.
 <400> 50

5

ES 2 621 181 T3

ttaattaa tgaagttctt gaactctttg tctttggtgg gtttggttat tgctggttgt 60
 gaagctatga gaaacatttc ttctaaagaa ttggttaaag aattgactat tggttggtct 120
 ttgggtaata ctttgatgc ttcttggtt gaaacttga actactctaa agatcaaact 180
 gcttctgaaa cttggtggg taatgttaa actactcaag aattgtacta caaattgtct 240
 gatttgggtt tcaatacttt cagaatacca actacttggc ctggtcattt tggatgatgct 300
 ccagattaca aaatttctga tgtttgatg aaaagagttc acgaagttgt tgattatgct 360
 ttgaatactg gtggttacgc tattttgaac attcatcatg aaacttggaa ttacgctttt 420
 caaaagaatt tggaatctgc taaaaagatt ttggttgcta tttggaaaca aattgctgct 480
 gaatttggtg attacgatga acatttgatt tttgaaggta tgaatgaacc aagaaaagtt 540
 ggtgatccag ctgaatggac tgggtggtgat caagaaggtt ggaattttgt taatgaaatg 600
 aacgctttgt tggtaaaac tattagagct actggtggtg acaatgctaa tagacatttg 660
 atgattccaa cttatgctgc ttctgttaat gatggttcta ttaacaattt taagtaccca 720
 aatggtgatg ataaagttat tgtttctttg cattcttact ctccatacaa ttttgctttg 780
 aacaatggtc caggtgctat ttctaatttc tacgatggtg acgaaattga ttgggttatg 840
 aacactatta actcttcatt ctttctaag ggtattccag ttattattgg tgaatttgtt 900
 gctatgaaca gagataatga agatgataga gaaagatggc aagaatacta cattaanaag 960
 gctactgctt tgggtattcc atgtgttatt tgggataatg gttattttga aggtgaaggt 1020
 gaaagatttg gtattattga tagaaagtct ttgaacgtta ttttccaaa gttgattaat 1080
 ggtttgatga aaggtttggg tgatgaaaa ccaaaaacta ctattagaag aactactact 1140
 actacagttc aagttcaacc aactattaac aacgaatggt tctctactag attgggttat 1200
 tcttgttga atggtttcga tgtttgtac actgataatg atggtcaatg ggggtgttga 1260
 aatggttaatt ggtgtggtat taaatcttct tgtggttaaca atcaaagaca atgttggctc 1320
 gaaagattag gttatccatg ttgtcaatac actactaatg ctgaatatac agacaacgac 1380
 ggtagatggg gtgtagaaaa cggttaactgg tgcggaatat acttgaagg cgcgcc 1436

<210> 51
 <211> 1220
 <212> ADN
 <213> *Irpex lacteus*

<400> 51

ES 2 621 181 T3

ttaattaa	aaa	tgaagtc	ttt	gttggtg	tct	gctgctg	cta	ctttggc	ttt	atctact	cca	60
gcttttctg	tttctg	tttg	gggtcaatg	ggtggtattg	gttttactg	gttctact	act	tttctact	act	tttctact	act	120
tgtgatgctg	gtacttctg	tgttcattg	aacgattact	actttcaatg	tcaaccaggt							180
gctgctactt	ctactgttca	accaactact	actgcttctt	ctacttcttc	tgctgcagct							240
coatcttctt	caggtaatgc	tgtttgttct	ggtactagaa	acaagtttaa	gttcttgggt							300
gttaatgaat	ctggtgctga	atttggtaac	aatggtattc	caggtacttt	gggtactgat							360
tatacttggc	catctccatc	ttctattgat	tttttctgtg	gtaagggttt	taatacttct							420
agagttccat	ttttgatgga	aagattgtct	ccacctgcta	ctggtttgac	tggtccattt							480
gattctactt	atttgcaagg	tttgaaaact	attgtttctt	acattactgg	taaaggtgggt							540
tatgctttgg	ttgatccaca	taactttatg	atttacaacg	gtgctactat	ttctgatact							600
aatgcttttc	aaacttggtg	gcaaaatttg	gctgctcaat	ttaagactga	ttctcatggt							660
gttttcgatg	ttatgaatga	accacatgat	attccagctc	aaactgtttt	taacttgaac							720
caagctgcta	ttaatagaat	tagagcttct	ggtgctactt	ctcaatctat	tttggttgaa							780
ggtacttctt	atactggtgc	ttggacttgg	actactactt	ctggtaattc	tcaagttttt							840
ggtgctattc	atgatccaaa	caacaatggt	gctattgaaa	tgcatcaata	cttggttct							900
gatggttctg	gtacttctcc	aacttgtggt	tctccaacta	ttggtgctga	aagattgcaa							960
gctgctactc	aatggttgca	acaaaacaat	ttgaaagggt	tcttgggtga	aattggtgct							1020
ggttcta	atg	ttctgctggt	caagggtgctt	tgtgtgaaat	gcaacaatct							1080
gatggttgg	tgggtgcttt	gtggtgggct	gctggtccat	ggtgggtga	ttattttcaa							1140
totattgaac	caccatctgg	tgttgctggt	tcttctattt	tgccacaagc	tttggaaacca							1200
tttttgttat	aaggcgcgcc											1220

<210> 52
 <211> 521
 <212> PRT
 <213> *Aspergillus kawachii*
 <400> 52

5

ES 2 621 181 T3

Met Arg Ile Ser Asn Leu Ile Val Ala Ala Ser Ala Ala Thr Met Val
 1 5 10 15

Ser Ala Leu Pro Ser Arg Gln Met Lys Lys Arg Asp Ser Gly Phe Lys
 20 25 30

Trp Val Gly Thr Ser Glu Ser Gly Ala Glu Phe Gly Ser Ala Leu Pro
 35 40 45

Gly Thr Leu Gly Thr Asp Tyr Thr Trp Pro Glu Thr Ser Lys Ile Gln
 50 55 60

Val Leu Arg Asn Lys Gly Met Asn Ile Phe Arg Ile Pro Phe Leu Met
 65 70 75 80

Glu Arg Leu Thr Pro Asp Gly Leu Thr Gly Ser Phe Ala Ser Thr Tyr
 85 90 95

Leu Ser Asp Leu Lys Ser Thr Val Glu Phe Val Thr Asn Ser Gly Ala
 100 105 110

Tyr Ala Val Leu Asp Pro His Asn Tyr Gly Arg Phe Asp Gly Ser Ile
 115 120 125

Ile Glu Ser Thr Ser Asp Phe Lys Thr Trp Trp Lys Asn Val Ala Thr
 130 135 140

Glu Phe Ala Asp Asn Asp Lys Val Ile Phe Asp Thr Asn Asn Glu Tyr
 145 150 155 160

His Asp Met Glu Gln Ser Leu Val Leu Asn Leu Asn Gln Ala Ala Ile
 165 170 175

ES 2 621 181 T3

Asn Gly Ile Arg Ala Ala Gly Ala Thr Thr Gln Tyr Ile Phe Val Glu
 180 185 190

Gly Asn Ala Tyr Thr Gly Ala Trp Asp Trp Thr Thr Tyr Asn Asp Asp
 195 200 205

Leu Ser Gly Leu Thr Asp Ser Glu Asp Lys Ile Ile Tyr Glu Met His
 210 215 220

Gln Tyr Leu Asp Ser Asp Ser Ser Gly Thr Ser Glu Thr Cys Val Ser
 225 230 235 240

Ser Thr Ile Gly Lys Glu Arg Ile Glu Lys Ala Thr Glu Trp Leu Lys
 245 250 255

Thr Asn Asn Lys Gln Gly Ile Ile Gly Glu Phe Ala Gly Gly Val Asn
 260 265 270

Ser Val Cys Glu Glu Ala Val Glu Gly Met Leu Ala Tyr Met Ser Glu
 275 280 285

Asn Ser Asp Val Trp Val Gly Ala Ser Trp Trp Ser Ala Gly Pro Trp
 290 295 300

Trp Gly Thr Tyr Met Tyr Ser Leu Glu Pro Thr Asp Gly Thr Ala Tyr
 305 310 315 320

Ser Thr Tyr Leu Pro Ile Leu Glu Lys Tyr Phe Pro Ser Gly Asp Ala
 325 330 335

Ser Ser Ser Ser Ser Ala Ser Ala Ser Val Ala Ala Ala Thr Ser Ala
 340 345 350

Val Ser Thr Thr Thr Thr Ala Ala Phe Glu Gln Thr Thr Thr Pro Ala
 355 360 365

Thr Gln Val Glu Ile Ala Ser Ser Ser Ser Ser Ser Ala Val Ala
 370 375 380

Ala Ser Gln Thr Thr Leu Ser Lys Val Lys Ser Lys Ser Lys Ser Pro
 385 390 395 400

Cys Lys Leu Ser Ser Ala Thr Ser Ser Ala Val Ser Ser Ala Ala Ala
 405 410 415

Val Thr Thr Pro Ala Val Ala Ala Thr Thr Pro Ala Ala Ala Pro Thr
 420 425 430

ES 2 621 181 T3

Ser Ser Ser Val Ala Phe Ala Thr Thr Ser Val Tyr Val Pro Thr Thr
 435 440 445

Thr Ala Ala Ala Pro Ser Gln Val Ser Ser Ser Ala Ala Ala Ser Ser
 450 455 460

Ser Gly Val Val Gly Val Ser Asp Pro Gln Gly Pro Ser Ala Thr Asn
 465 470 475 480

Ser Ala Gly Glu Val Asn Gln Tyr Tyr Gln Cys Gly Gly Ile Asn Trp
 485 490 495

Thr Gly Pro Thr Val Cys Ala Ser Pro Tyr Thr Cys Lys Val Gln Asn
 500 505 510

Asp Tyr Tyr Tyr Gln Cys Val Ala Glu
 515 520

<210> 53
 <211> 419
 <212> PRT
 <213> *Heterodera schachtii*
 <400> 53

5

Met His Trp Ala Asp Val Ala Cys Ser Arg Pro Pro Trp Pro Arg Asp
 1 5 10 15

Ser Val Lys Ala Leu Lys Cys Asn Trp Asn Ala Asn Val Ile Arg Gly
 20 25 30

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

Tyr Asn Leu Met Val Ala Val Ile Glu Ala Ala Ile Ser Asn Gly Ile
 50 55 60

Tyr Val Ile Val Asp Trp His Ala His Asn Ala His Pro Asp Glu Ala
 65 70 75 80

Val Lys Phe Phe Thr Arg Ile Ala Gln Ala Tyr Gly Ser Tyr Leu His
 85 90 95

Ile Leu Tyr Glu Asp Phe Asn Glu Pro Leu Asp Val Ser Trp Thr Asp
 100 105 110

Val Leu Val Pro Tyr His Lys Lys Val Ile Ala Ala Ile Arg Ala Ile
 115 120 125

10

ES 2 621 181 T3

Asp Lys Lys Asn Val Ile Ile Leu Gly Thr Pro Lys Trp Ser Gln Asp
 130 135 140

Val Asp Val Ala Ser Gln Asn Pro Ile Lys Asp Tyr Gln Asn Leu Met
 145 150 155 160

Tyr Thr Leu His Phe Tyr Ala Ser Ser His Phe Thr Ser Asp Leu Gly
 165 170 175

Ala Lys Leu Lys Thr Ala Val Asn Asn Gly Leu Pro Val Phe Val Thr
 180 185 190

Glu Tyr Gly Thr Cys Glu Ala Ser Gly Asn Gly Asn Leu Asn Thr Asp
 195 200 205

Ser Met Ser Ser Trp Trp Thr Leu Leu Asp Ser Leu Lys Ile Ser Tyr
 210 215 220

Ala Asn Trp Ala Ile Ser Asp Lys Ser Glu Ala Cys Ser Ala Leu Ser
 225 230 235 240

Pro Gly Thr Thr Ala Val Asn Val Gly Val Ser Ser Arg Trp Thr Ser
 245 250 255

Ser Gly Asn Met Val Ala Ser Tyr Tyr Lys Lys Lys Ser Thr Gly Ile
 260 265 270

Ser Cys Ser Gly Ser Ser Ser Gly Ser Ser Ser Gly Ser Ser Ser Gly
 275 280 285

Ser Ser Gly Thr Ser Ser Gly Ser Ser Gly Ser Ser Ser Gly Ser Ser
 290 295 300

Ser Gly Ser Ser Ser Gly Ser Ser Gly Ser Ser Ser Gly Ser Ser Ser
 305 310 315 320

Gly Ser Gly Ser Ala Ser Ile Ser Val Val Pro Ser Asn Thr Trp Asn
 325 330 335

Gly Gly Gly Arg Val Asn Phe Glu Ile Lys Asn Thr Gly Ser Val Pro
 340 345 350

Leu Cys Gly Val Val Phe Ser Val Ser Leu Pro Ser Gly Thr Thr Leu
 355 360 365

Gly Gly Ser Trp Asn Met Glu Ser Ala Gly Ser Gly Gln Tyr Ser Leu

ES 2 621 181 T3

370

375

380

Pro Ser Trp Val Arg Ile Glu Ala Gly Lys Ser Ser Lys Asp Ala Gly
385 390 395 400

Leu Thr Phe Asn Gly Lys Asp Lys Pro Thr Ala Lys Ile Val Thr Thr
405 410 415

Lys Lys Cys

<210> 54

<211> 418

<212> PRT

<213> *Hypocrea jecorina*

<400> 54

5

ES 2 621 181 T3

Met Asn Lys Ser Val Ala Pro Leu Leu Leu Ala Ala Ser Ile Leu Tyr
 1 5 10 15

Gly Gly Ala Val Ala Gln Gln Thr Val Trp Gly Gln Cys Gly Gly Ile
 20 25 30

Gly Trp Ser Gly Pro Thr Asn Cys Ala Pro Gly Ser Ala Cys Ser Thr
 35 40 45

Leu Asn Pro Tyr Tyr Ala Gln Cys Ile Pro Gly Ala Thr Thr Ile Thr
 50 55 60

Thr Ser Thr Arg Pro Pro Ser Gly Pro Thr Thr Thr Thr Arg Ala Thr
 65 70 75 80

Ser Thr Ser Ser Ser Thr Pro Pro Thr Ser Ser Gly Val Arg Phe Ala
 85 90 95

Gly Val Asn Ile Ala Gly Phe Asp Phe Gly Cys Thr Thr Asp Gly Thr
 100 105 110

Cys Val Thr Ser Lys Val Tyr Pro Pro Leu Lys Asn Phe Thr Gly Ser
 115 120 125

Asn Asn Tyr Pro Asp Gly Ile Gly Gln Met Gln His Phe Val Asn Glu
 130 135 140

Asp Gly Met Thr Ile Phe Arg Leu Pro Val Gly Trp Gln Tyr Leu Val
 145 150 155 160

Asn Asn Asn Leu Gly Gly Asn Leu Asp Ser Thr Ser Ile Ser Lys Tyr

Arg Lys

<210> 55
 <211> 471
 <212> PRT
 <213> *Orpinomyces* sp.

5

<400> 55

Met Lys Phe Leu Asn Ser Leu Ser Leu Leu Gly Leu Val Ile Ala Gly
 1 5 10 15

Cys Glu Ala Met Arg Asn Ile Ser Ser Lys Glu Leu Val Lys Glu Leu
 20 25 30

Thr Ile Gly Trp Ser Leu Gly Asn Thr Leu Asp Ala Ser Cys Val Glu
 35 40 45

Thr Leu Asn Tyr Ser Lys Asp Gln Thr Ala Ser Glu Thr Cys Trp Gly
 50 55 60

Asn Val Lys Thr Thr Gln Glu Leu Tyr Tyr Lys Leu Ser Asp Leu Gly
 65 70 75 80

Phe Asn Thr Phe Arg Ile Pro Thr Thr Trp Ser Gly His Phe Gly Asp
 85 90 95

Ala Pro Asp Tyr Lys Ile Ser Asp Val Trp Met Lys Arg Val His Glu
 100 105 110

Val Val Asp Tyr Ala Leu Asn Thr Gly Gly Tyr Ala Ile Leu Asn Ile
 115 120 125

His His Glu Thr Trp Asn Tyr Ala Phe Gln Lys Asn Leu Glu Ser Ala
 130 135 140

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

Asp Tyr Asp Glu His Leu Ile Phe Glu Gly Met Asn Glu Pro Arg Lys
 165 170 175

Val Gly Asp Pro Ala Glu Trp Thr Gly Gly Asp Gln Glu Gly Trp Asn
 180 185 190

Phe Val Asn Glu Met Asn Ala Leu Phe Val Lys Thr Ile Arg Ala Thr
 195 200 205

ES 2 621 181 T3

Gly Gly Asn Asn Ala Asn Arg His Leu Met Ile Pro Thr Tyr Ala Ala
 210 215 220

Ser Val Asn Asp Gly Ser Ile Asn Asn Phe Lys Tyr Pro Asn Gly Asp
 225 230 235 240

Asp Lys Val Ile Val Ser Leu His Ser Tyr Ser Pro Tyr Asn Phe Ala
 245 250 255

Leu Asn Asn Gly Pro Gly Ala Ile Ser Asn Phe Tyr Asp Gly Asn Glu
 260 265 270

Ile Asp Trp Val Met Asn Thr Ile Asn Ser Ser Phe Ile Ser Lys Gly
 275 280 285

Ile Pro Val Ile Ile Gly Glu Phe Val Ala Met Asn Arg Asp Asn Glu
 290 295 300

Asp Asp Arg Glu Arg Trp Gln Glu Tyr Tyr Ile Lys Lys Ala Thr Ala
 305 310 315 320

Leu Gly Ile Pro Cys Val Ile Trp Asp Asn Gly Tyr Phe Glu Gly Glu
 325 330 335

Gly Glu Arg Phe Gly Ile Ile Asp Arg Lys Ser Leu Asn Val Ile Phe
 340 345 350

Pro Lys Leu Ile Asn Gly Leu Met Lys Gly Leu Gly Asp Glu Lys Pro
 355 360 365

Lys Thr Thr Ile Arg Arg Thr Thr Thr Thr Val Gln Val Gln Pro
 370 375 380

Thr Ile Asn Asn Glu Cys Phe Ser Thr Arg Leu Gly Tyr Ser Cys Cys
 385 390 395 400

Asn Gly Phe Asp Val Leu Tyr Thr Asp Asn Asp Gly Gln Trp Gly Val
 405 410 415

Glu Asn Gly Asn Trp Cys Gly Ile Lys Ser Ser Cys Gly Asn Asn Gln
 420 425 430

Arg Gln Cys Trp Ser Glu Arg Leu Gly Tyr Pro Cys Cys Gln Tyr Thr
 435 440 445

Thr Asn Ala Glu Tyr Thr Asp Asn Asp Gly Arg Trp Gly Val Glu Asn
 450 455 460

Gly Asn Trp Cys Gly Ile Tyr
 465 470

<210> 56
 <211> 399
 <212> PRT
 5 <213> *Irpex lacteus*
 <400> 56

Met Lys Ser Leu Leu Leu Ser Ala Ala Ala Thr Leu Ala Leu Ser Thr
 1 5 10 15

Pro Ala Phe Ser Val Ser Val Trp Gly Gln Cys Gly Gly Ile Gly Phe
 20 25 30

Thr Gly Ser Thr Thr Cys Asp Ala Gly Thr Ser Cys Val His Leu Asn
 35 40 45

Asp Tyr Tyr Phe Gln Cys Gln Pro Gly Ala Ala Thr Ser Thr Val Gln
 50 55 60

Pro Thr Thr Thr Ala Ser Ser Thr Ser Ser Ala Ala Ala Pro Ser Ser
 65 70 75 80

Ser Gly Asn Ala Val Cys Ser Gly Thr Arg Asn Lys Phe Lys Phe Phe
 85 90 95

Gly Val Asn Glu Ser Gly Ala Glu Phe Gly Asn Asn Val Ile Pro Gly
 100 105 110

Thr Leu Gly Thr Asp Tyr Thr Trp Pro Ser Pro Ser Ser Ile Asp Phe
 115 120 125

Phe Val Gly Lys Gly Phe Asn Thr Phe Arg Val Pro Phe Leu Met Glu
 130 135 140

Arg Leu Ser Pro Pro Ala Thr Gly Leu Thr Gly Pro Phe Asp Ser Thr
 145 150 155 160

Tyr Leu Gln Gly Leu Lys Thr Ile Val Ser Tyr Ile Thr Gly Lys Gly
 165 170 175

Gly Tyr Ala Leu Val Asp Pro His Asn Phe Met Ile Tyr Asn Gly Ala
 180 185 190

Thr Ile Ser Asp Thr Asn Ala Phe Gln Thr Trp Trp Gln Asn Leu Ala
 195 200 205

ES 2 621 181 T3

Ala Gln Phe Lys Thr Asp Ser His Val Val Phe Asp Val Met Asn Glu
 210 215 220

Pro His Asp Ile Pro Ala Gln Thr Val Phe Asn Leu Asn Gln Ala Ala
 225 230 235 240

Ile Asn Arg Ile Arg Ala Ser Gly Ala Thr Ser Gln Ser Ile Leu Val
 245 250 255

Glu Gly Thr Ser Tyr Thr Gly Ala Trp Thr Trp Thr Thr Thr Ser Gly
 260 265 270

Asn Ser Gln Val Phe Gly Ala Ile His Asp Pro Asn Asn Asn Val Ala
 275 280 285

Ile Glu Met His Gln Tyr Leu Asp Ser Asp Gly Ser Gly Thr Ser Pro
 290 295 300

Thr Cys Val Ser Pro Thr Ile Gly Ala Glu Arg Leu Gln Ala Ala Thr
 305 310 315 320

Gln Trp Leu Gln Gln Asn Asn Leu Lys Gly Phe Leu Gly Glu Ile Gly
 325 330 335

Ala Gly Ser Asn Ala Asp Cys Ile Ser Ala Val Gln Gly Ala Leu Cys
 340 345 350

Glu Met Gln Gln Ser Asp Val Trp Leu Gly Ala Leu Trp Trp Ala Ala
 355 360 365

Gly Pro Trp Trp Gly Asp Tyr Phe Gln Ser Ile Glu Pro Pro Ser Gly
 370 375 380

Val Ala Val Ser Ser Ile Leu Pro Gln Ala Leu Glu Pro Phe Leu
 385 390 395

<210> 57
 <211> 42
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> secuencia enlazadora flexible

<400> 57

5

10

ES 2 621 181 T3

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Ala Trp His Pro Gln Phe
1 5 10 15

Gly Gly Glu Asn Leu Tyr Phe Gln Gly Asp Tyr Lys Asp Asp Asp Lys
20 25 30

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
35 40

<210> 58

<211> 37

<212> PRT

<213> Secuencia artificial

<220>

<223> secuencia enlazadora flexible

<400> 58

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Trp Ser His Pro Gln Phe
1 5 10 15

Glu Lys Gly Gly Glu Asn Leu Tyr Phe Gln Gly Gly Gly Gly Gly Ser
20 25 30

Gly Gly Gly Gly Ser
35

5

10

REIVINDICACIONES

1. Una célula hospedadora de levadura termotolerante transformada que comprende:
- 5 (a) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una endoglucanasa;
 (b) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una β -glucosidasa;
 (c) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una celobiohidrolasa I; y
 (d) al menos un polinucleótido heterólogo que comprende un ácido nucleico que codifica una celobiohidrolasa II;
- 10 en la que la célula hospedadora de levadura puede producir etanol cuando se cultiva utilizando celulosa como una fuente de carbono.
2. La célula hospedadora de la reivindicación 1, en la que la celulosa es celulosa cristalina, celulosa hinchada con ácido fosfórico, celulosa microcristalina, lodo de papel o forraje de maíz.
- 15 3. La célula hospedadora de la reivindicación 1 o reivindicación 2, en la que la β -glucosidasa es una β -glucosidasa de *Saccharomycopsis fibuligera*, la celobiohidrolasa I es un celobiohidrolasa I de *Talaromyces emersonii* y la celobiohidrolasa II es una celobiohidrolasa II de *Chrysosporium lucknowense*.
- 20 4. La célula hospedadora de una cualquiera de las reivindicaciones 1-3, en la que al menos una de, endoglucanasa, β -glucosidasa, celobiohidrolasa I y celobiohidrolasa II es secretada por la célula.
5. La célula hospedadora de la reivindicación 1, que adicionalmente comprende un polinucleótido heterólogo que comprende un ácido nucleico que codifica una segunda endoglucanasa.
- 25 6. La célula hospedadora de una cualquiera de las reivindicaciones 1-5, en la que la célula hospedadora es una célula de *Saccharomyces cerevisiae*.
7. La célula hospedadora de una cualquiera de las reivindicaciones 1-5, en la que la célula hospedadora es una célula de *Kluyveromyces*.
- 30 8. La célula hospedadora de cualquiera de las reivindicaciones 1-7, en la que la célula hospedadora termotolerante puede producir etanol a partir de celulosa a temperaturas por encima de aproximadamente 35 °C, por encima de aproximadamente 37 °C, por encima de aproximadamente 40 °C, por encima de aproximadamente 45 °C o por encima de aproximadamente 50 °C.
- 35 9. La célula hospedadora de una cualquiera de las reivindicaciones 1, 2, o 4-8, en la que la endoglucanasa, la β -glucosidasa o la celobiohidrolasa es una endoglucanasa, β -glucosidasa o celobiohidrolasa de *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpep lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* o de *Arabidopsis thaliana*.
- 40 10. La célula hospedadora de una cualquiera de las reivindicaciones 1-9, en la que dicha celobiohidrolasa I está fusionada a un módulo de unión a celulosa (MUC).
- 45 11. La célula hospedadora de la reivindicación 10, en la que el MUC es el MUC de la Cbh1 de *Trichoderma reesei* o de la Cbh1 de *Humicola grisea*.
- 50 12. La célula hospedadora de la reivindicación 1, en la que dicha endoglucanasa es una endoglucanasa de *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *R. speratus*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.* o de *Irpep lacteus*.
- 55 13. La célula hospedadora de una cualquiera de las reivindicaciones 1-12, en la que la célula hospedadora es una célula hospedadora que utiliza xilosa.
- 60 14. Un cocultivo que comprende al menos dos células hospedadoras de levadura en el que:
- 65 (a) al menos una de las células hospedadoras comprende un primer polinucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una endoglucanasa.
 (b) al menos una de las células hospedadoras comprende un segundo polinucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una β -glucosidasa;
 (c) al menos una de las células hospedadoras comprende un tercer polinucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una celobiohidrolasa I;

(d) al menos una de las células hospedadoras comprende un cuarto polinucleótido heterólogo que comprende un ácido nucleico que codifica una celulasa que es una celobiohidrolasa II;

5 en el que el primer polinucleótido, el segundo polinucleótido, el tercer polinucleótido y el cuarto polinucleótido no están en la misma célula hospedadora; y en el que el cocultivo produce etanol a partir de celulosa.

15. El cocultivo de la reivindicación 14, en el que al menos una de las células hospedadoras es una célula hospedadora termotolerante.

10 16. El cocultivo de la reivindicación 14 o de la reivindicación 15, en el que dicho polinucleótido que comprende un ácido nucleico que codifica una celobiohidrolasa I o II, codifica una proteína de fusión que comprende una celobiohidrolasa I o II y un módulo de unión a la celulosa (MUC).

15 17. El cocultivo de la reivindicación 16, en el que el MUC es el MUC de la Cbh1 de *Trichoderma reesei* o de la Cbh1 de *Humicola grisea*.

18. El cocultivo de una cualquiera de las reivindicaciones 14-17, en el que:

- 20 (a) dicha β -glucosidasa es una β -glucosidasa I de *S. fibuligera*;
(b) dicha celobiohidrolasa I es una celobiohidrolasa I de *T. emersonii*; y
(c) dicha celobiohidrolasa II es una celobiohidrolasa II de *C. lucknowense*.

25 19. El cocultivo de cualquiera de las reivindicaciones 14-18, en el que al menos una célula hospedadora es una célula hospedadora que utiliza xilosa.

20. La célula hospedadora de una cualquiera de las reivindicaciones 1-13, en la que todas las celulasas se secretan.

30 21. La célula hospedadora de la reivindicación 13, en la que la célula hospedadora que utiliza la xilosa expresa de manera heteróloga XilA en *Piromyces sp.* E2, sobreexpresa xiluloquinasa, ribulosa 5-fosfato isomerasa, ribulosa 5-fosfato epimerasa, transcetolasa, transaldolasa y no expresa el gen GRE3 que codifica la aldosa reductasa.

35 22. Un método para i) hidrolizar un sustrato celulósico, que comprende poner en contacto dicho sustrato celulósico con una célula hospedadora de acuerdo con una cualquiera de las reivindicaciones 1-13, o 20-21 o el cocultivo de acuerdo con una cualquiera de las reivindicaciones 14-19.

Figura 1

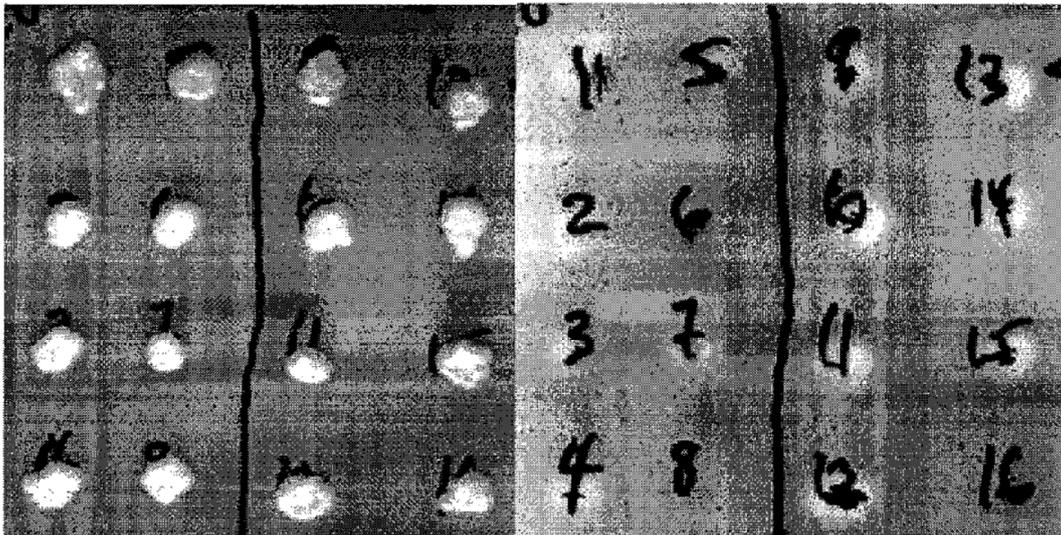
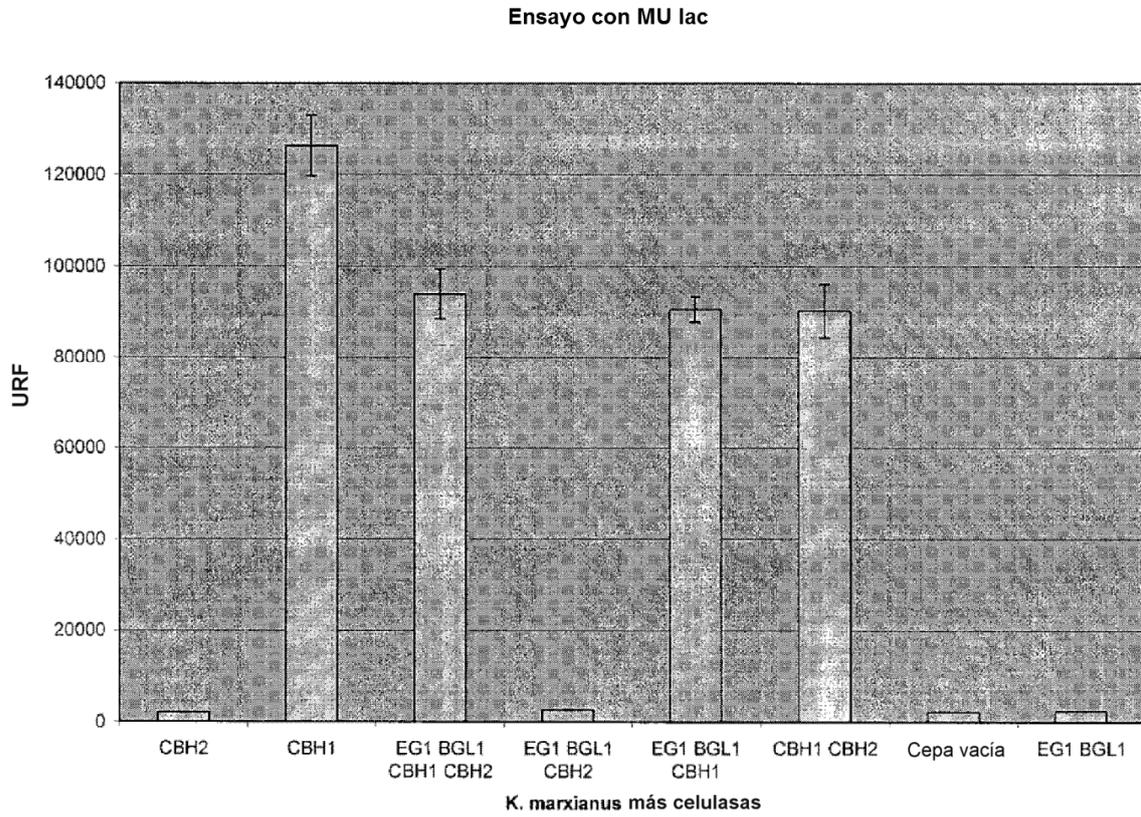


Figura 2



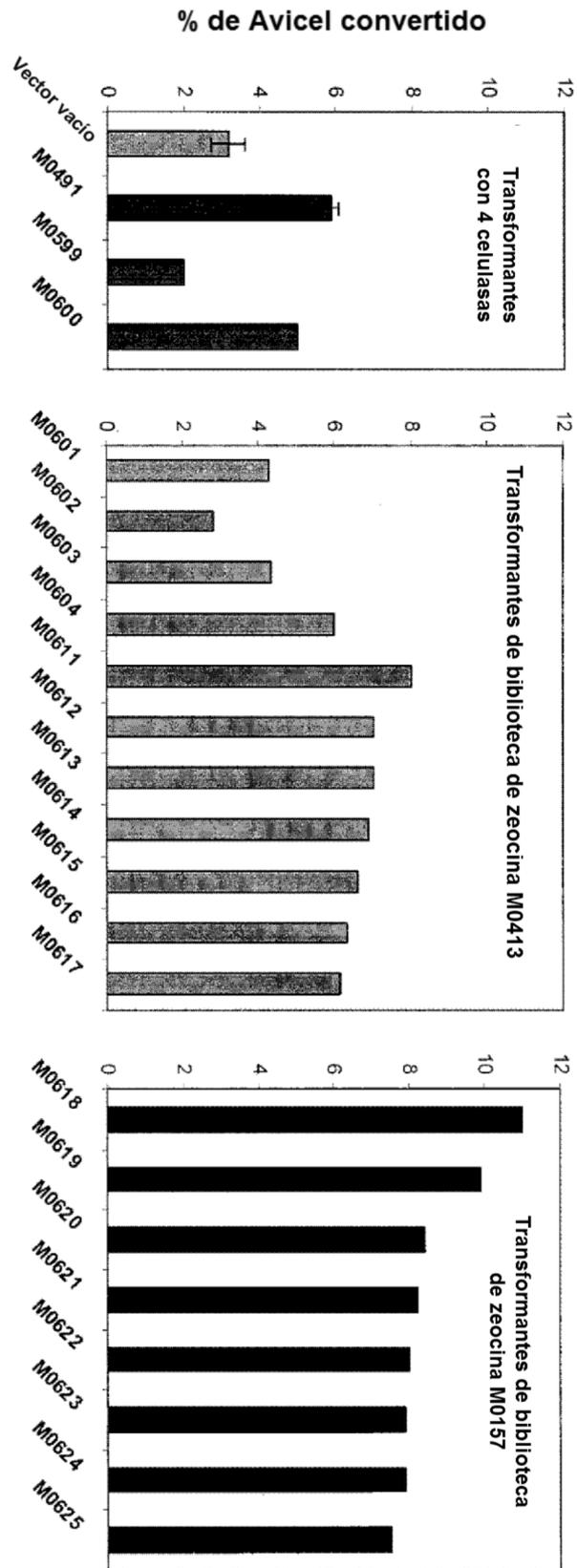


Figura 3

Figura 4

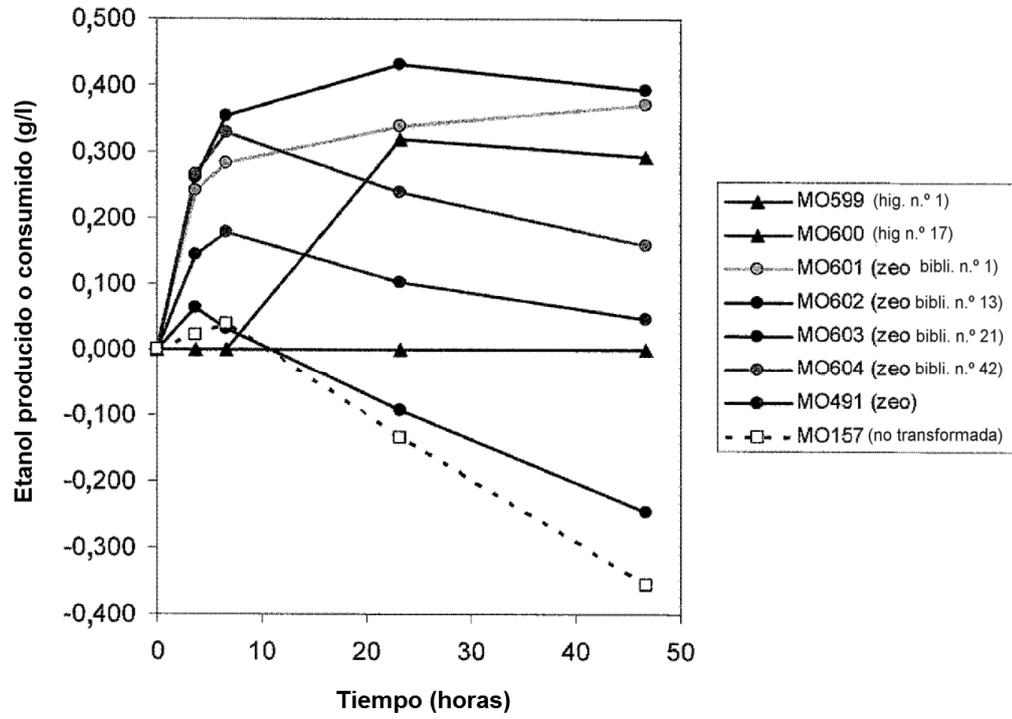


Figura 5

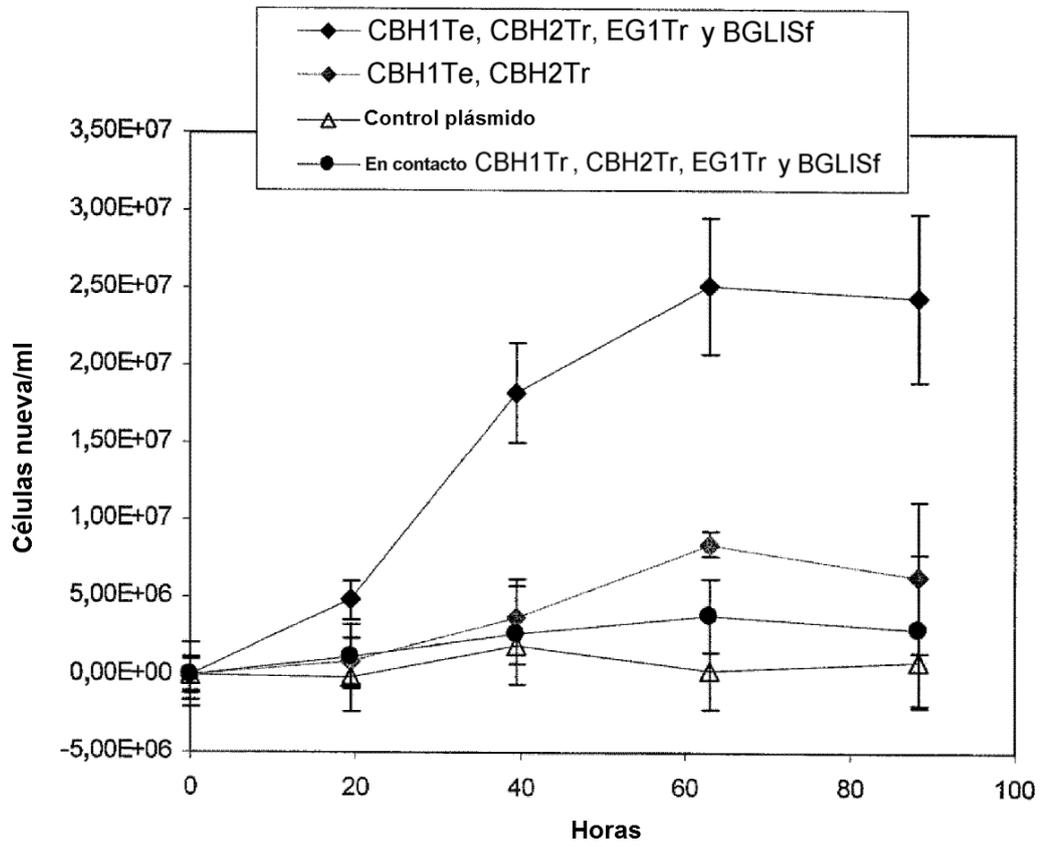


Figura 6

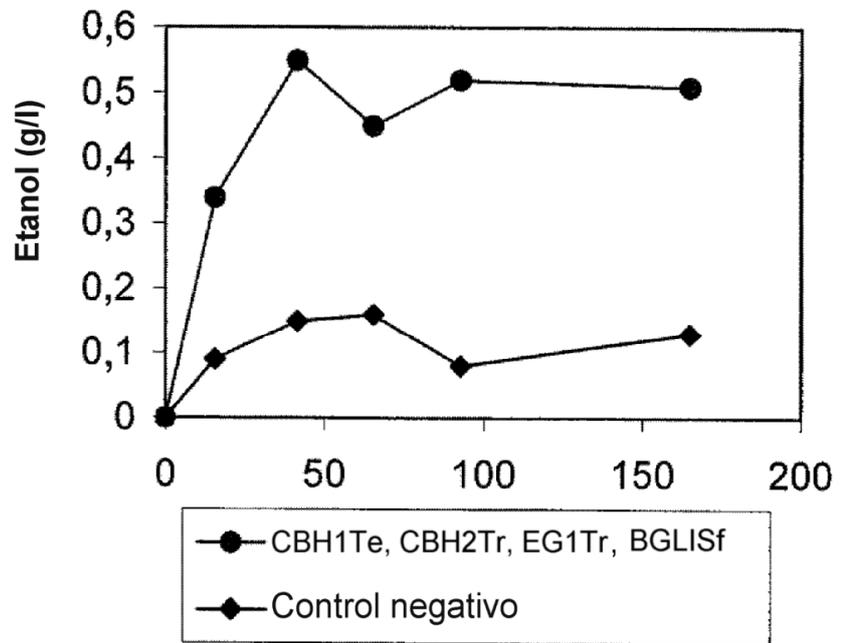


Figura 7

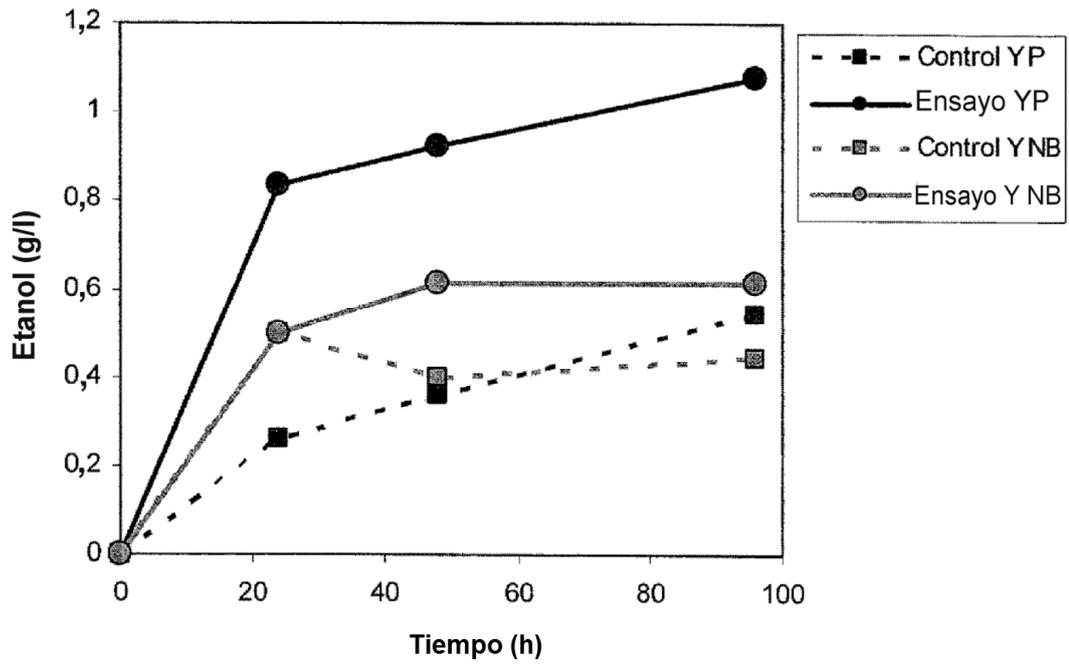


Figura 8

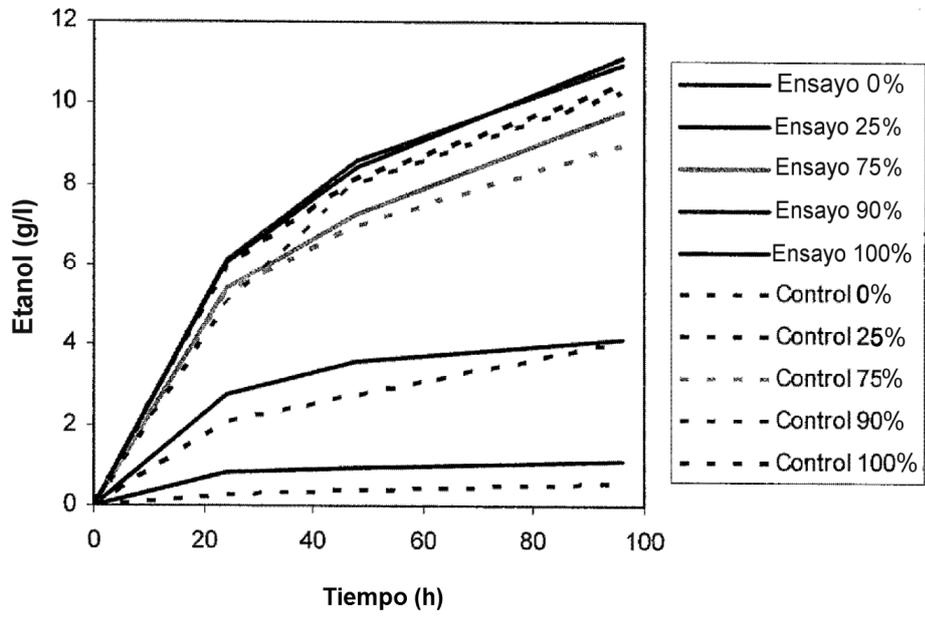


Figura 9

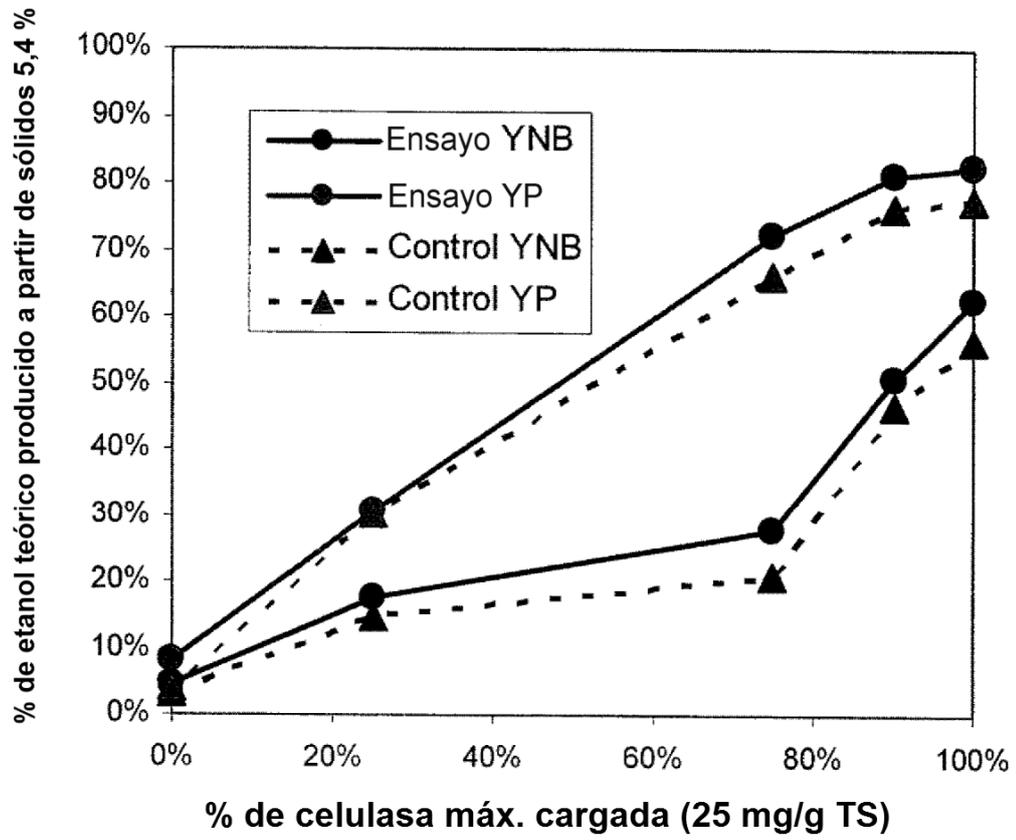


Figura 10

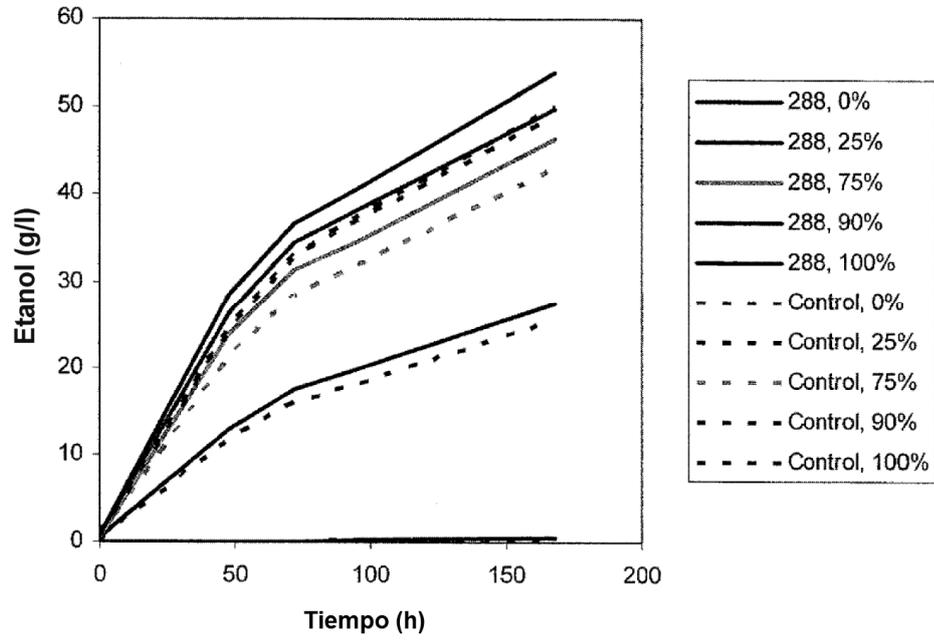


Figura 11

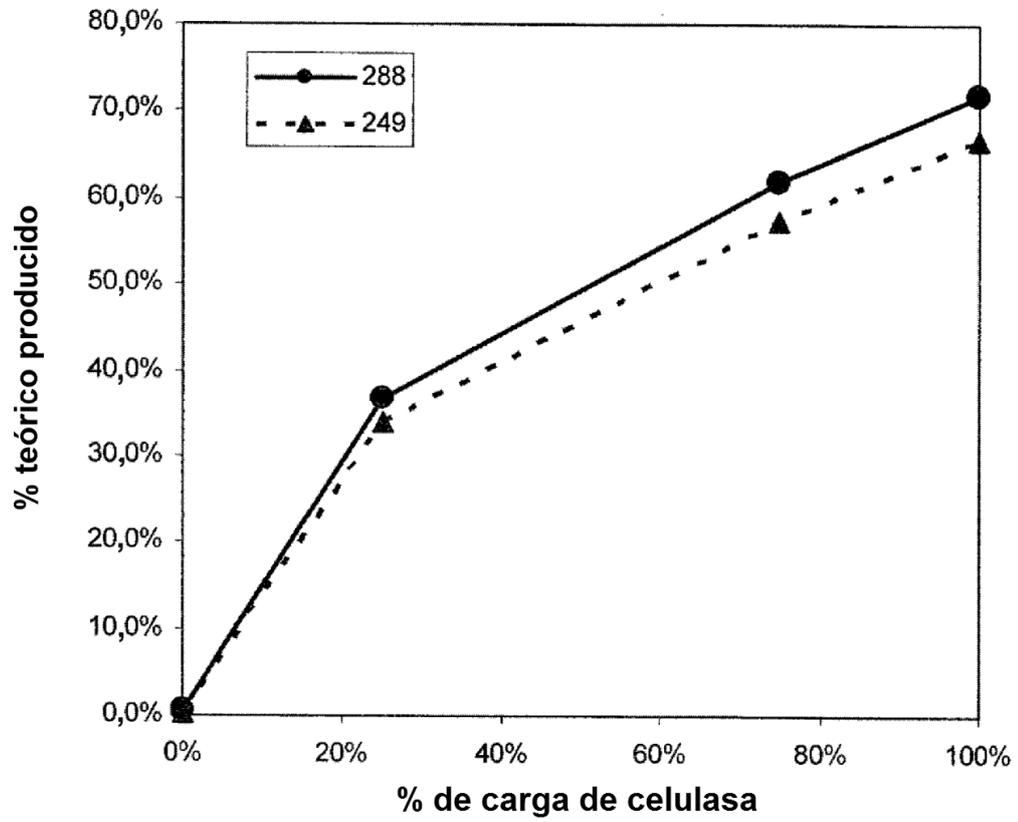


Figura 12

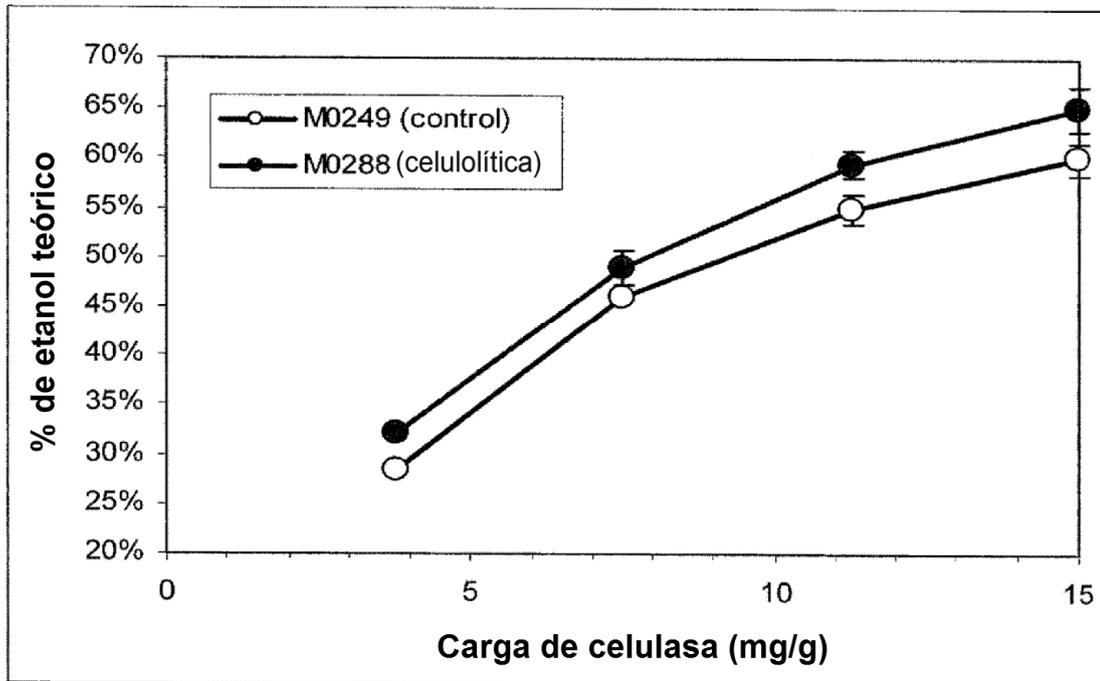
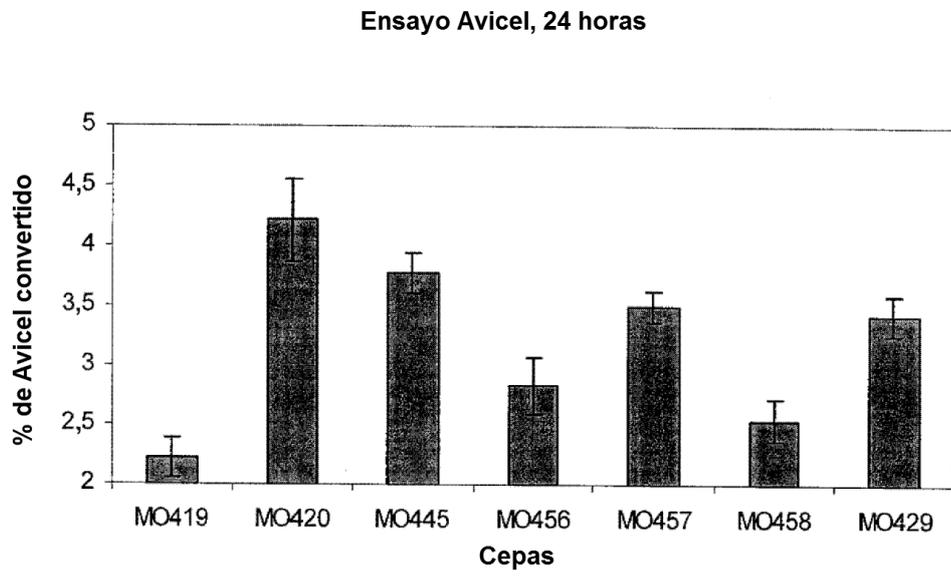


Figura 13



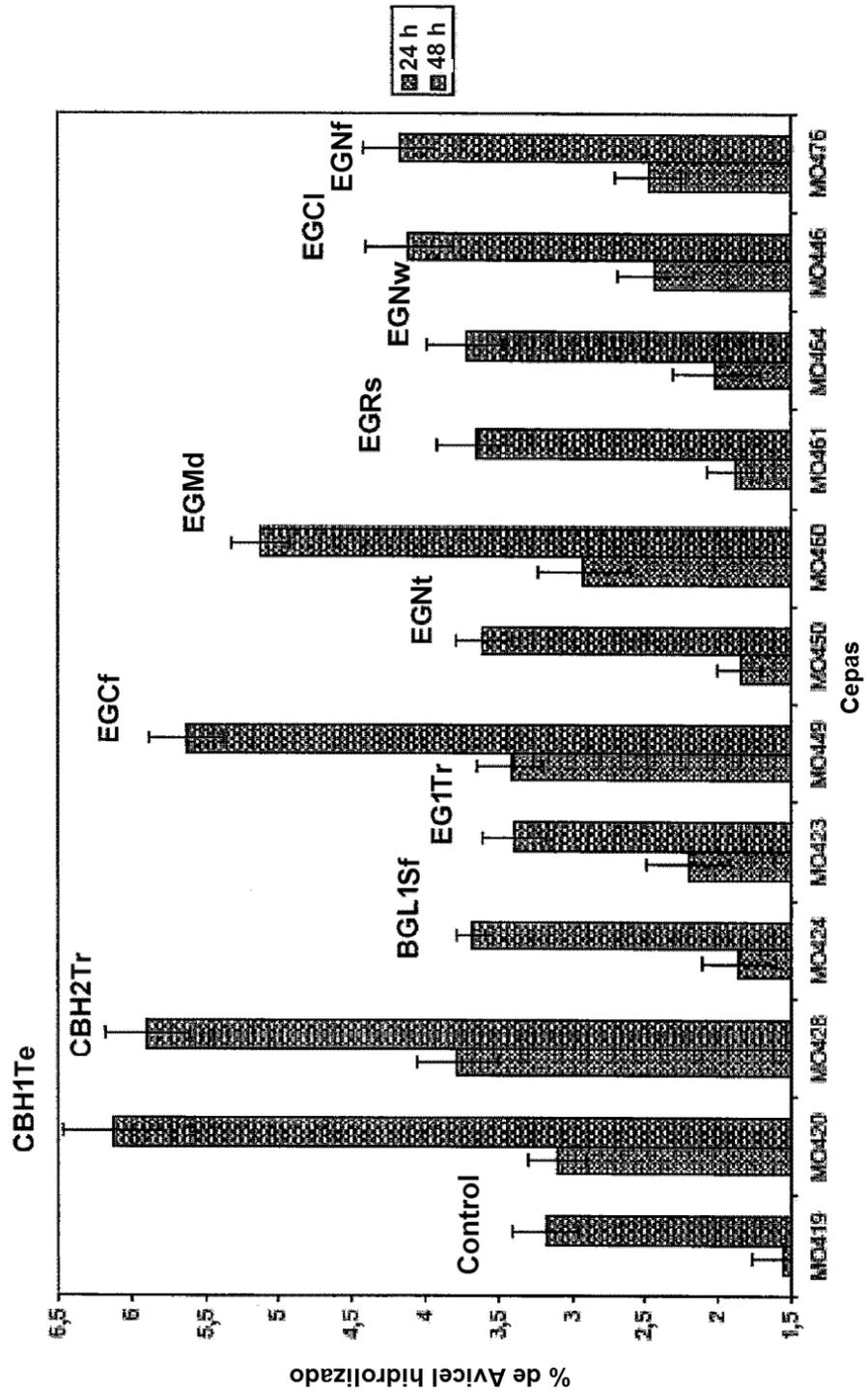


Figura 15

Figura 16

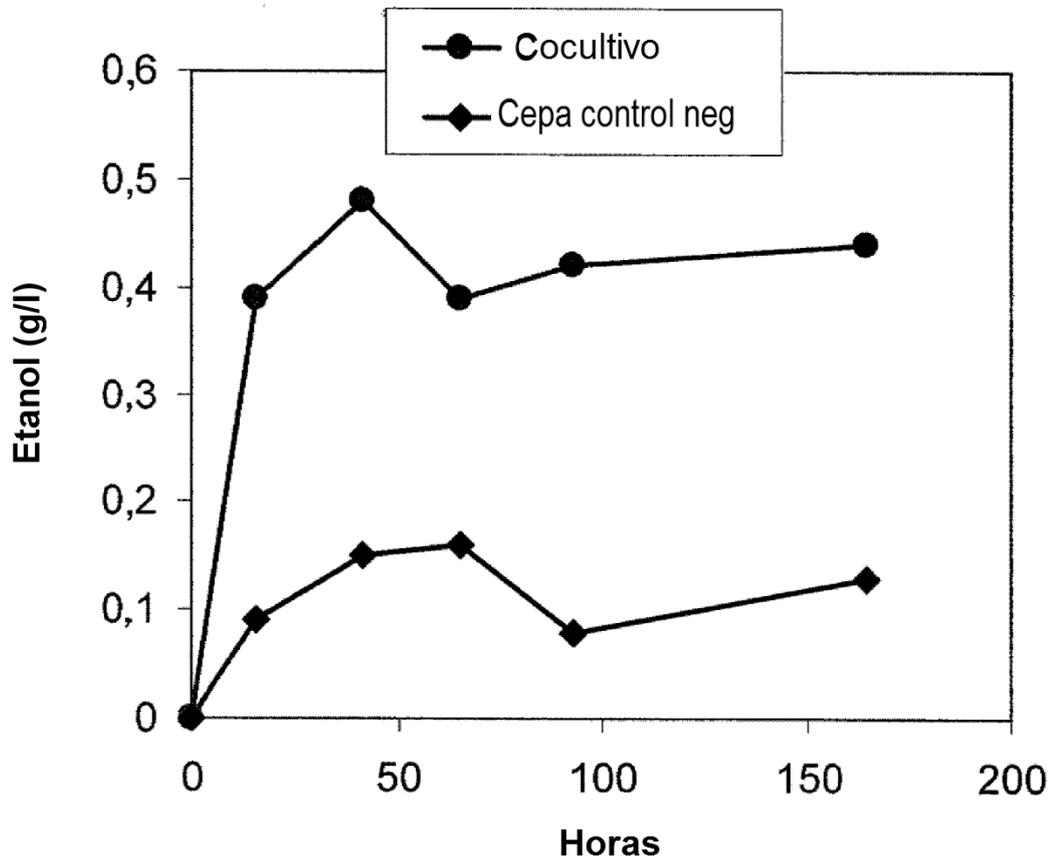


Figura 17

Producción de etanol con carga de celulosa cero en SFS

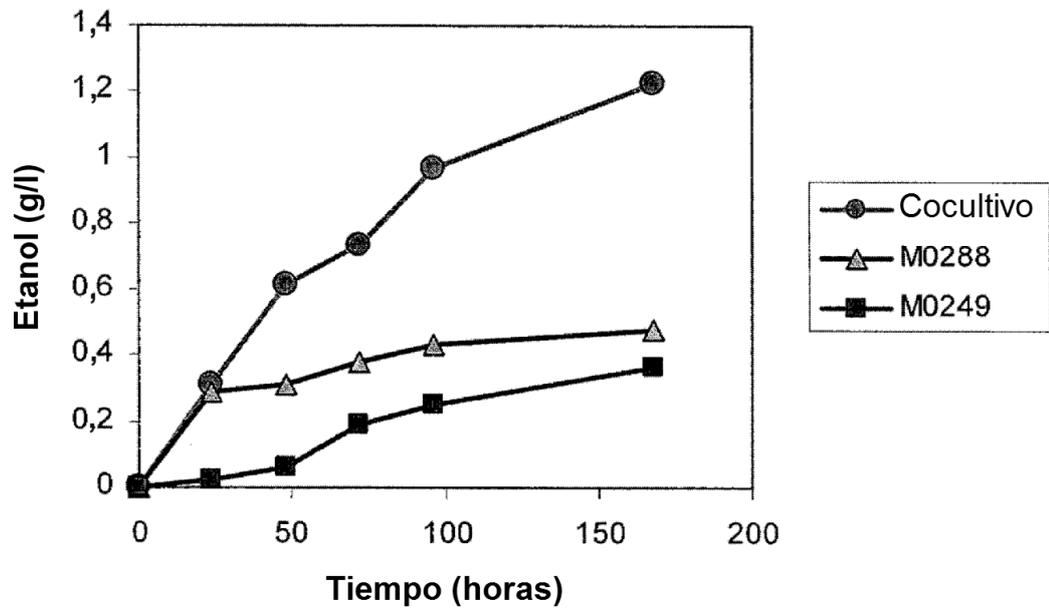
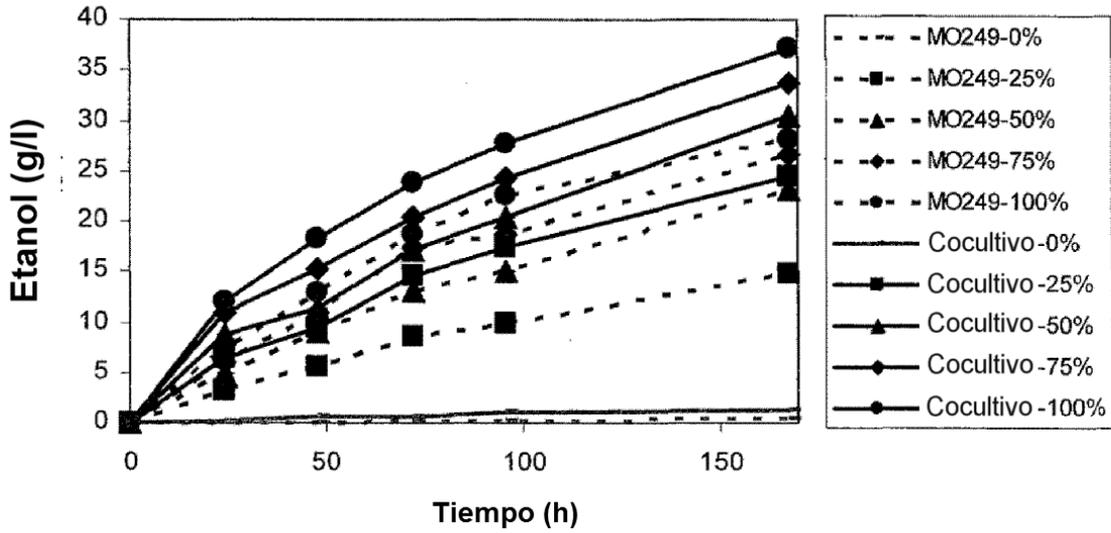


Figura 18

SFS 10 mg/g de celulasa = 100 %, sustrato Avicel 10 %



SFS 10 mg/g de celulasa = 100 %, sustrato Avicel 10 %

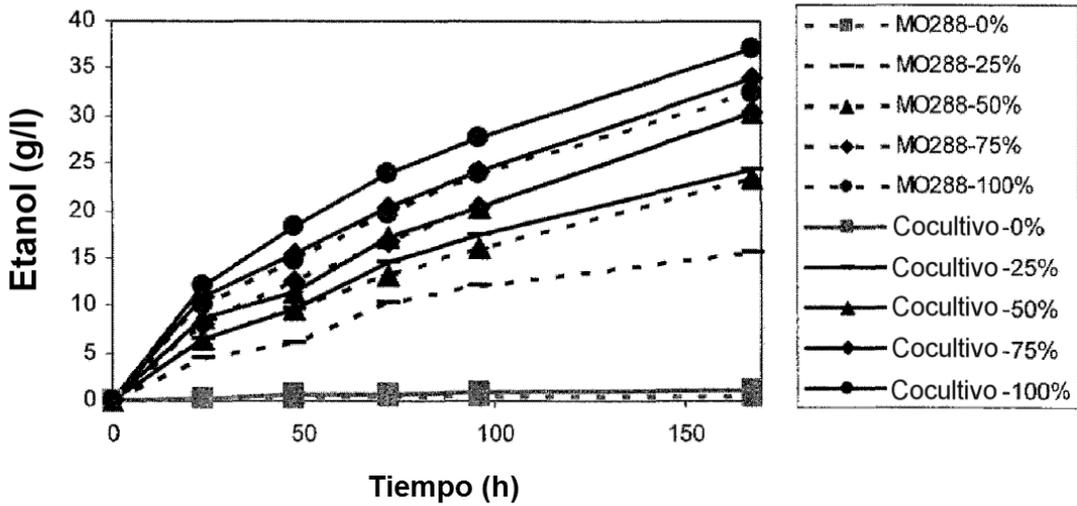


Figura 19

Ahorros enzimáticos - Avicel 10 % SSF, 100 % = 10 mg/g de celulasa

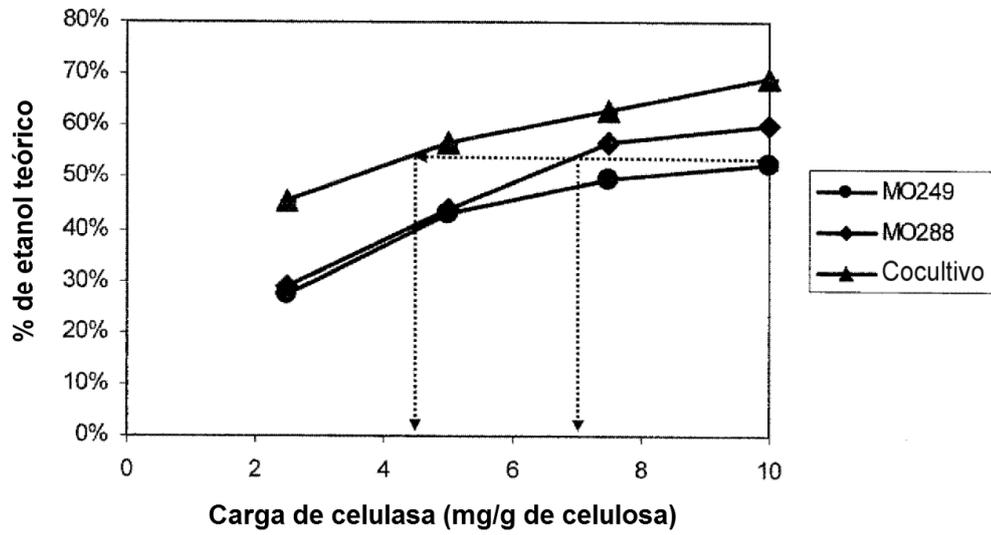


Figura 20

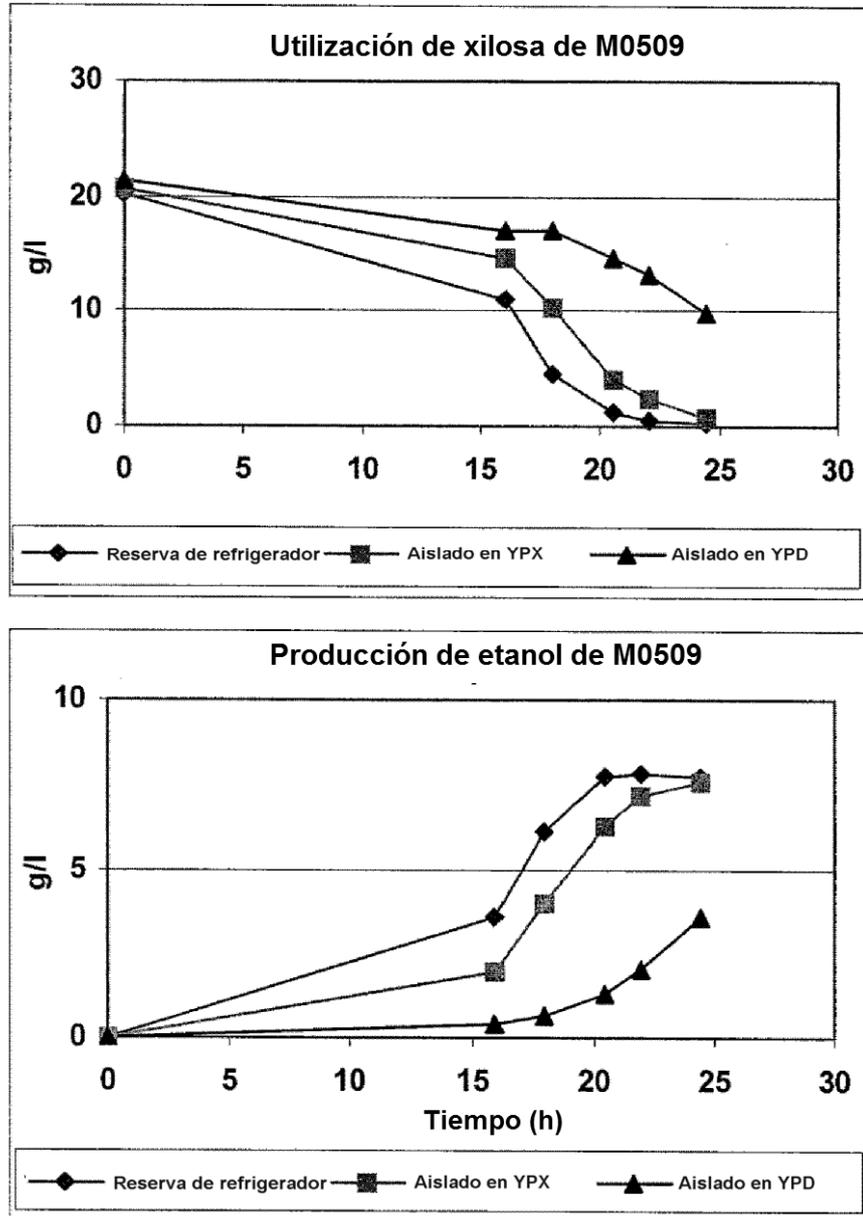


Figura 21

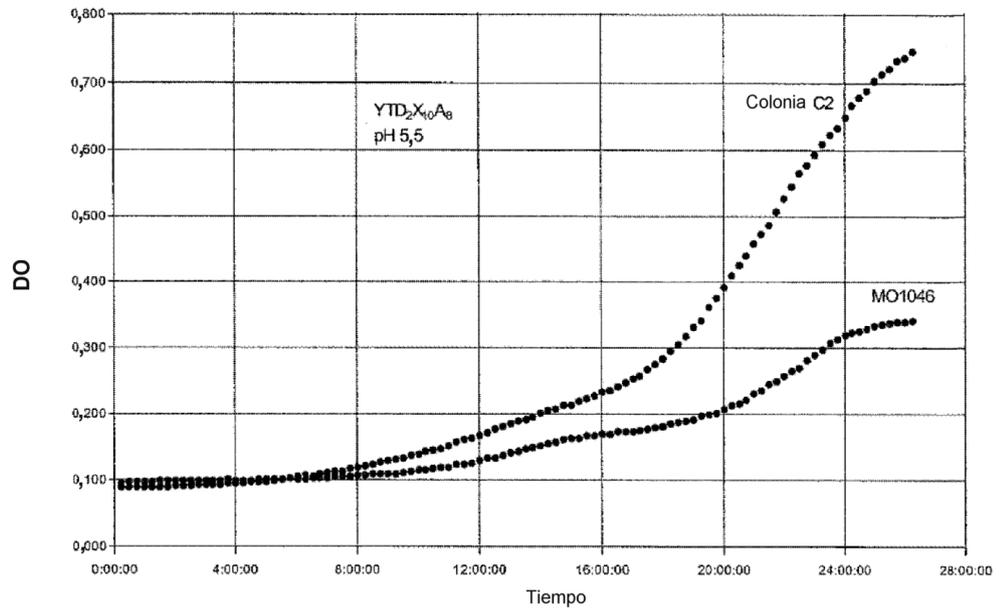


Figura 22

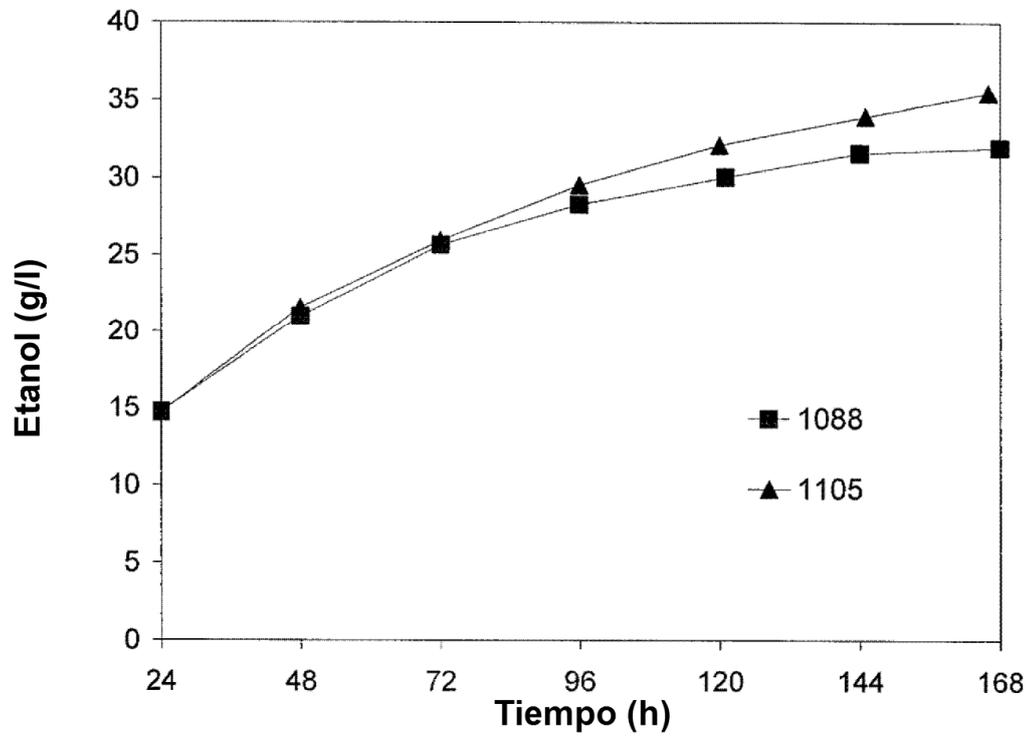


Figura 23

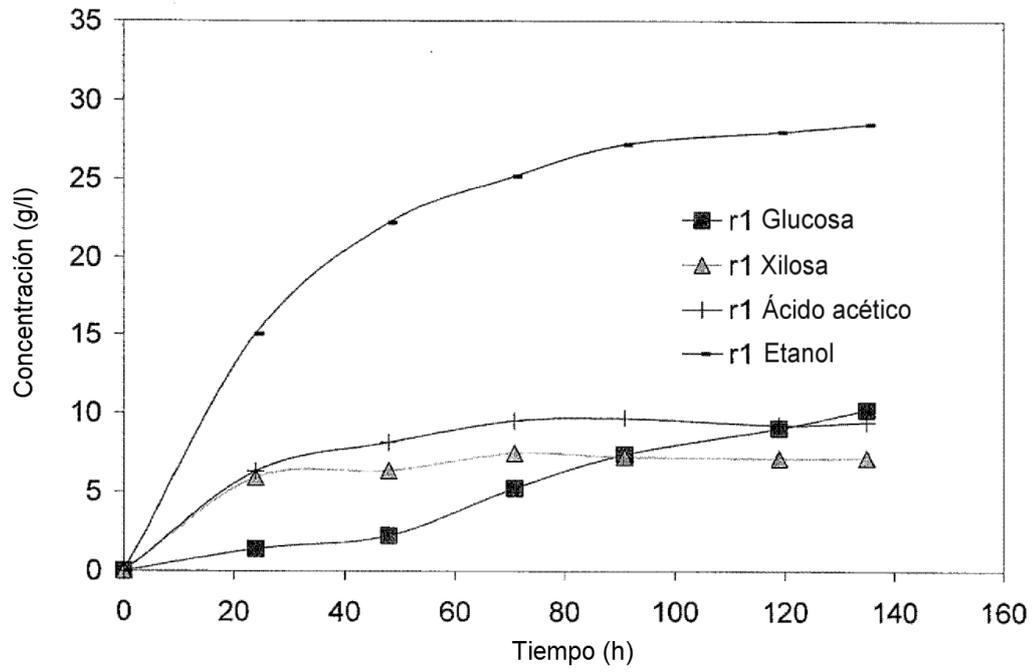


Figura 24

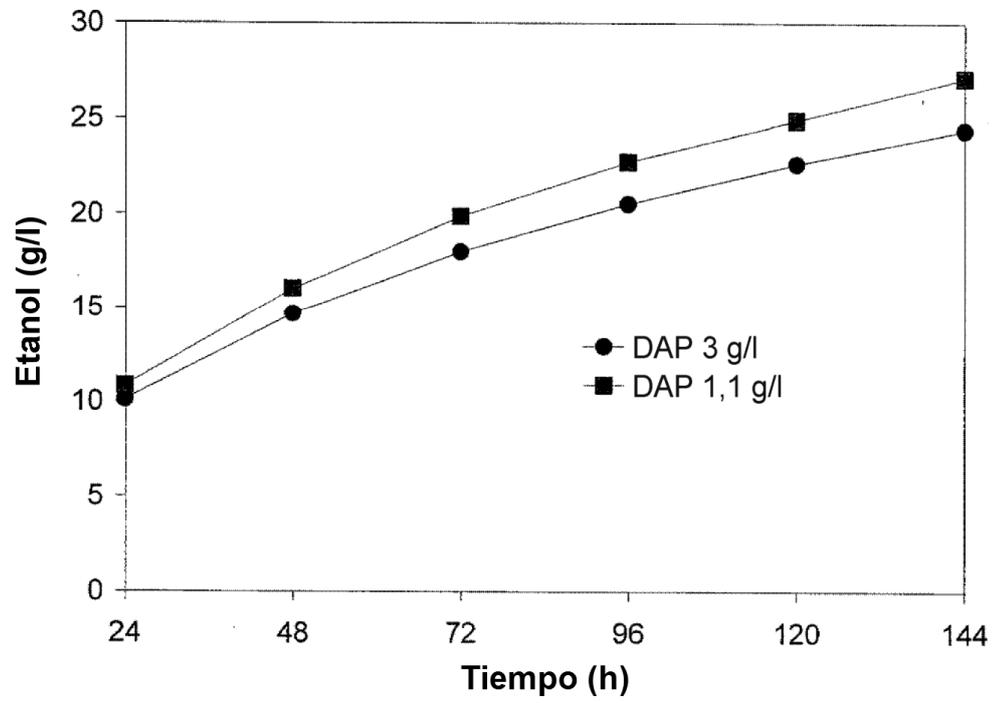


Figura 25

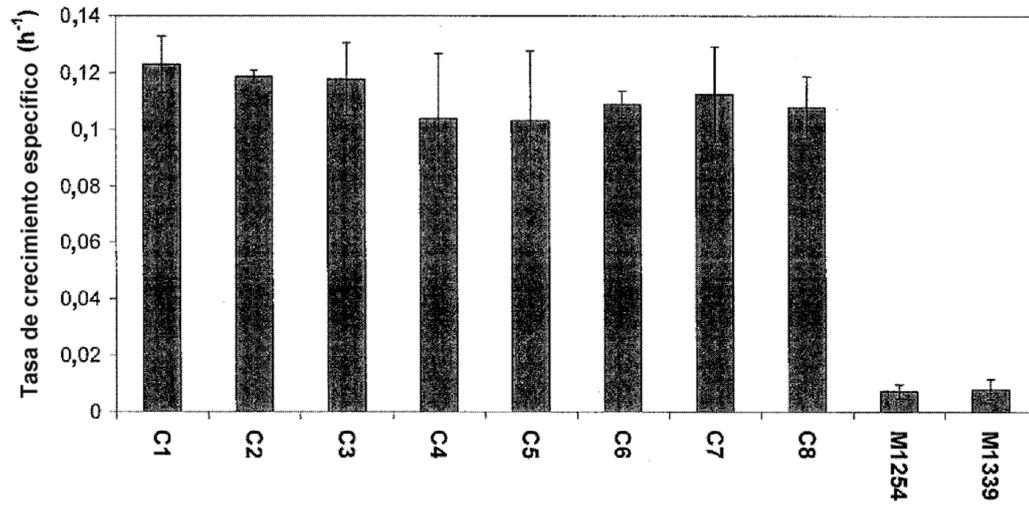


Figura 26

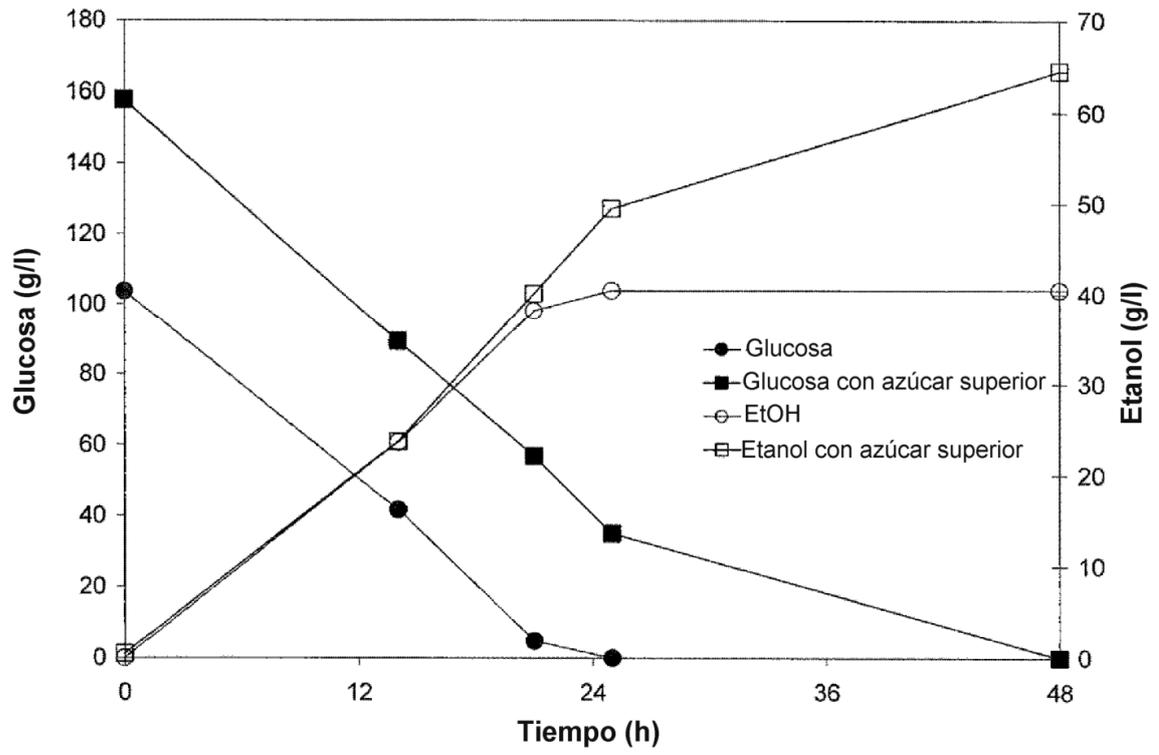


Figura 27

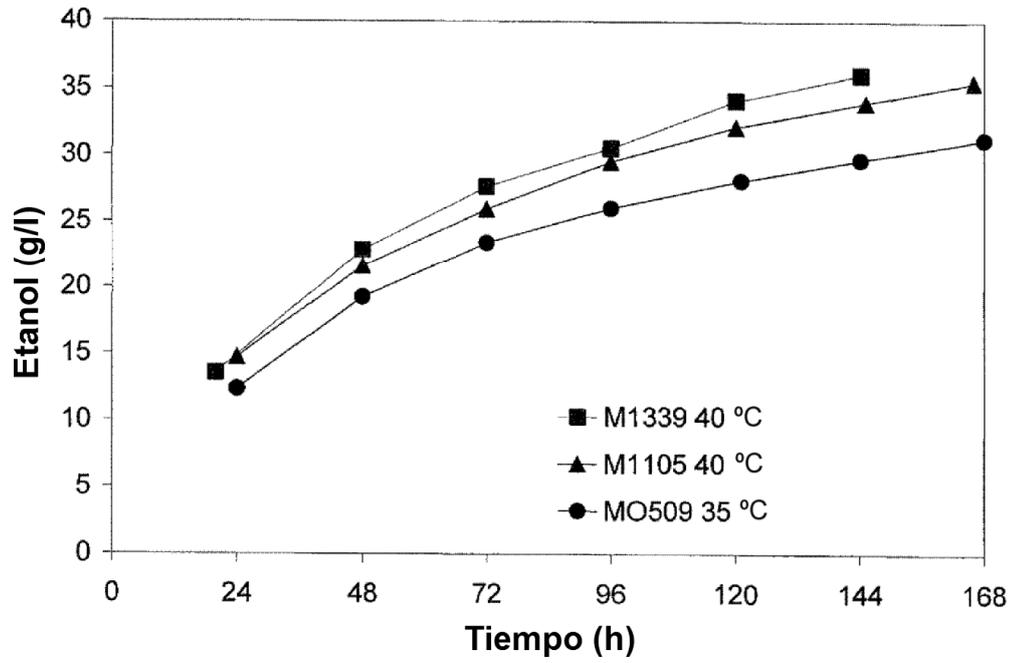


Figura 28

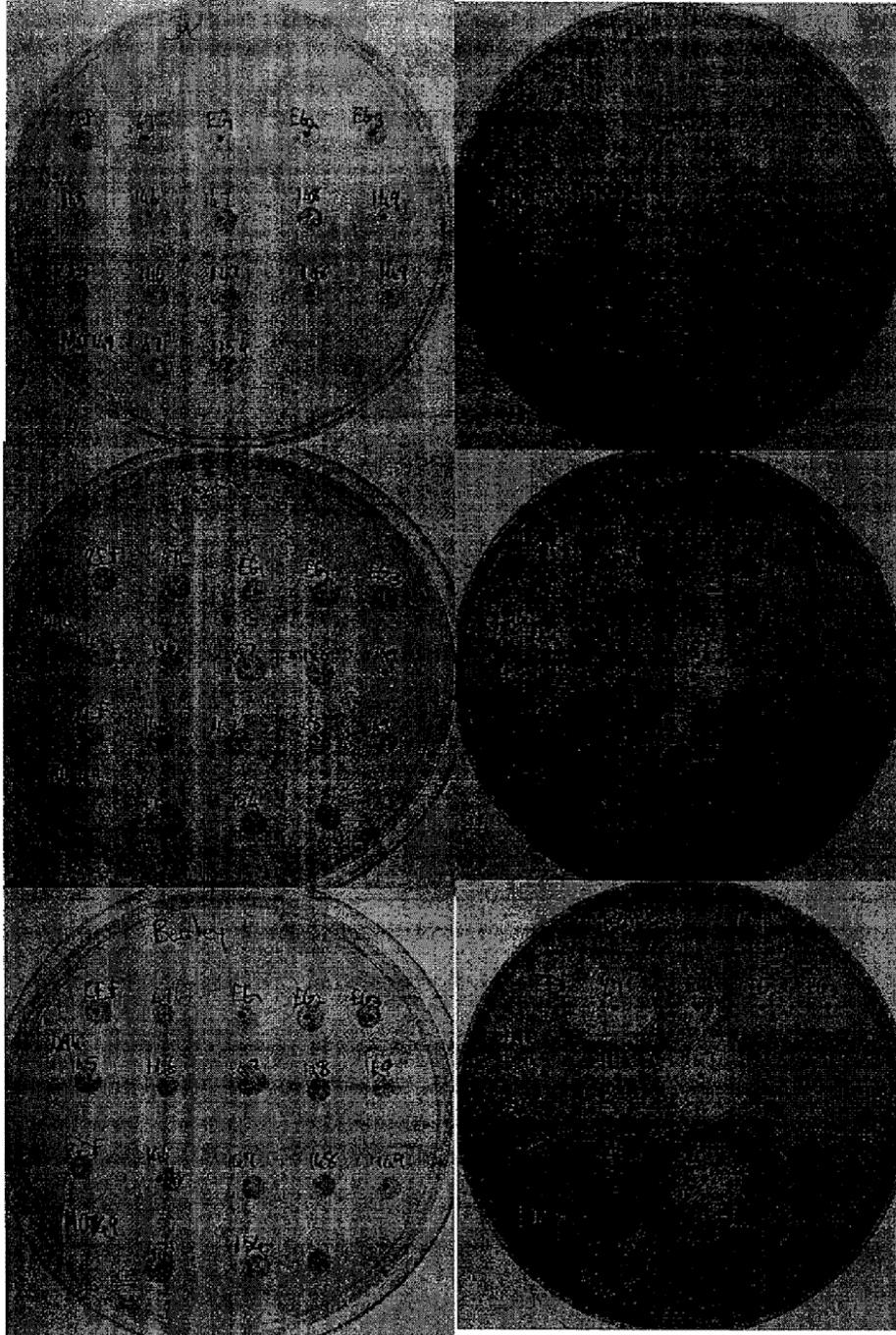


Figura 29

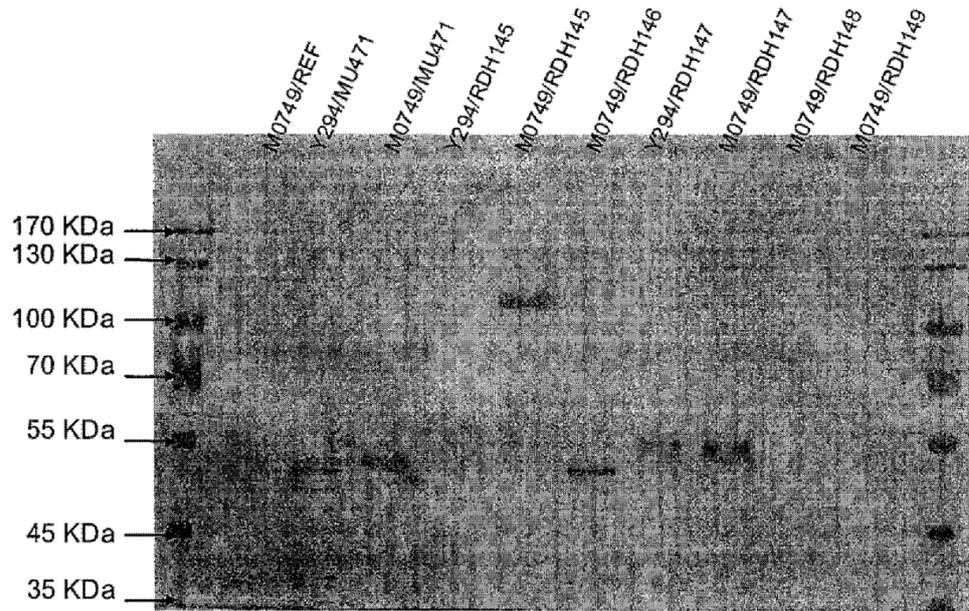


Figura 30

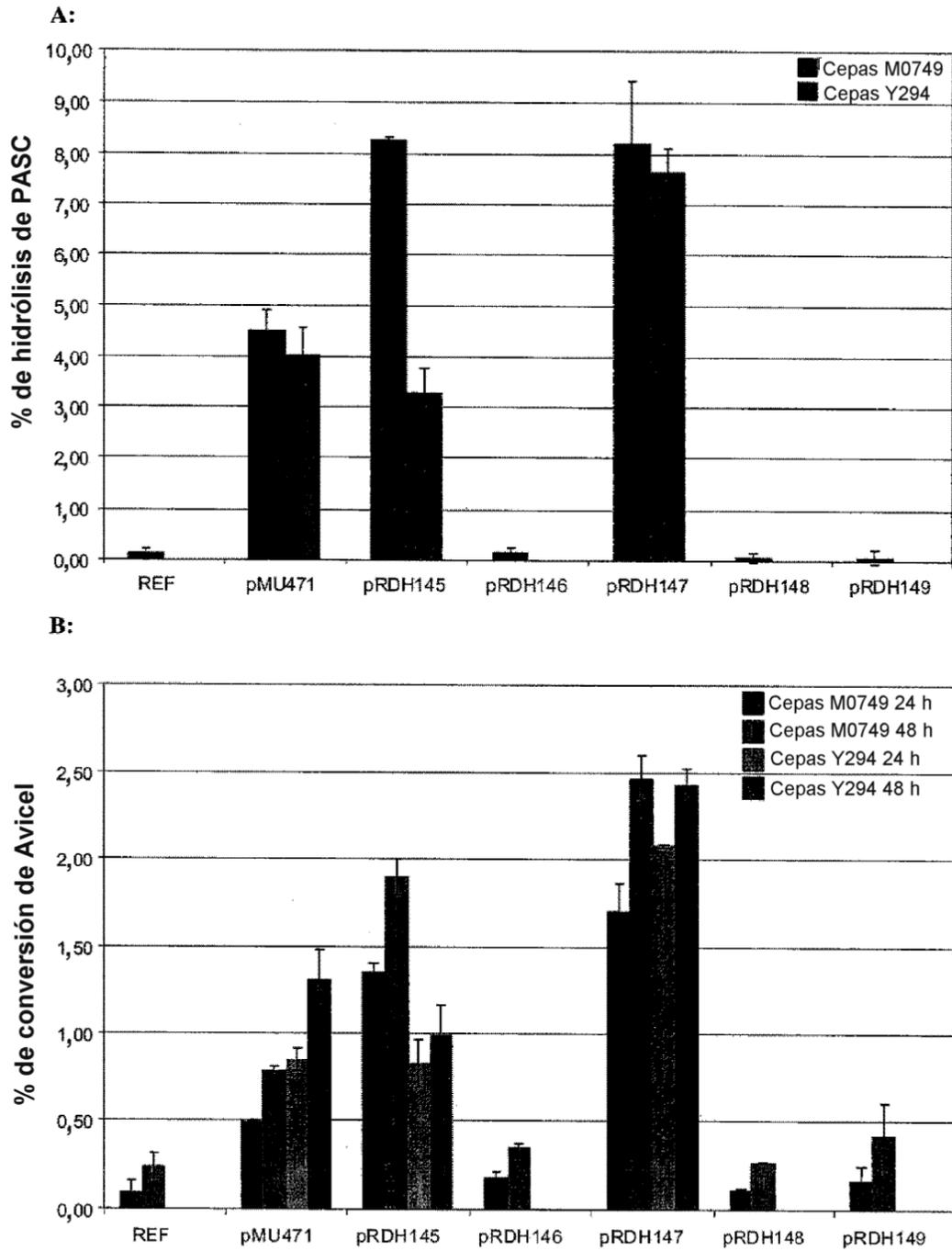


Figura 31

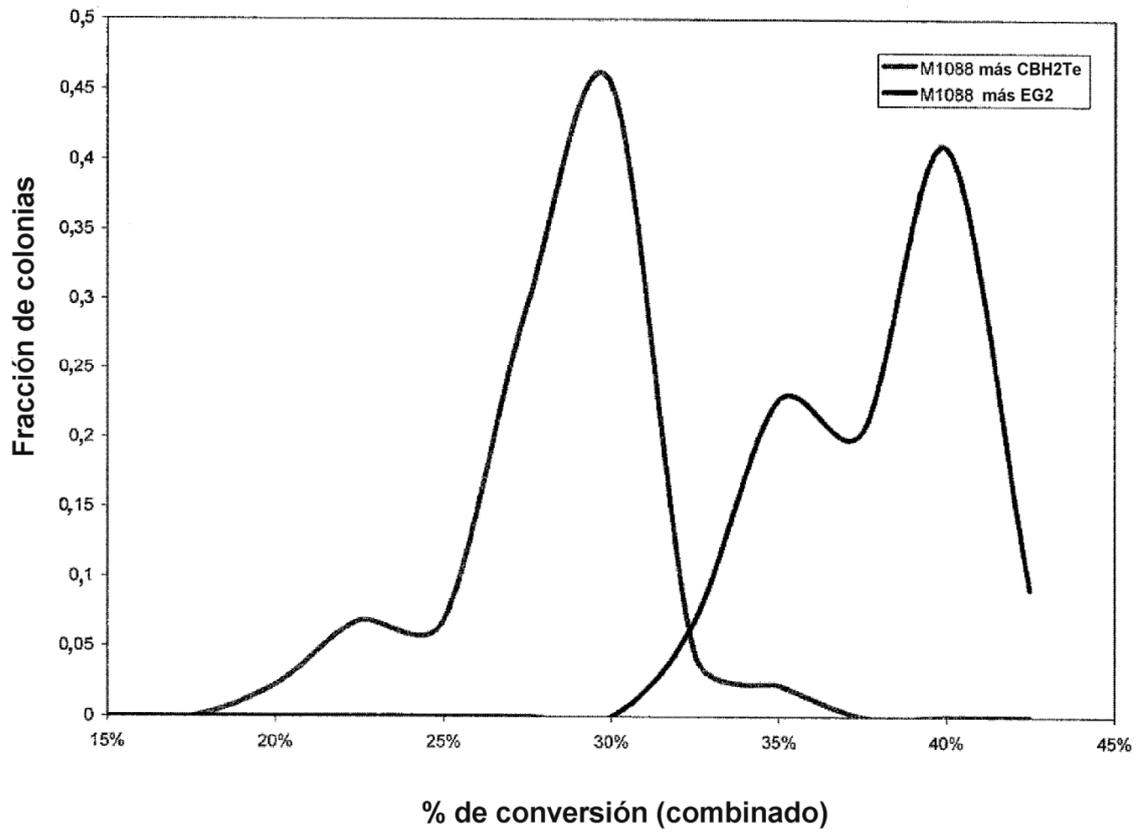


Figura 32

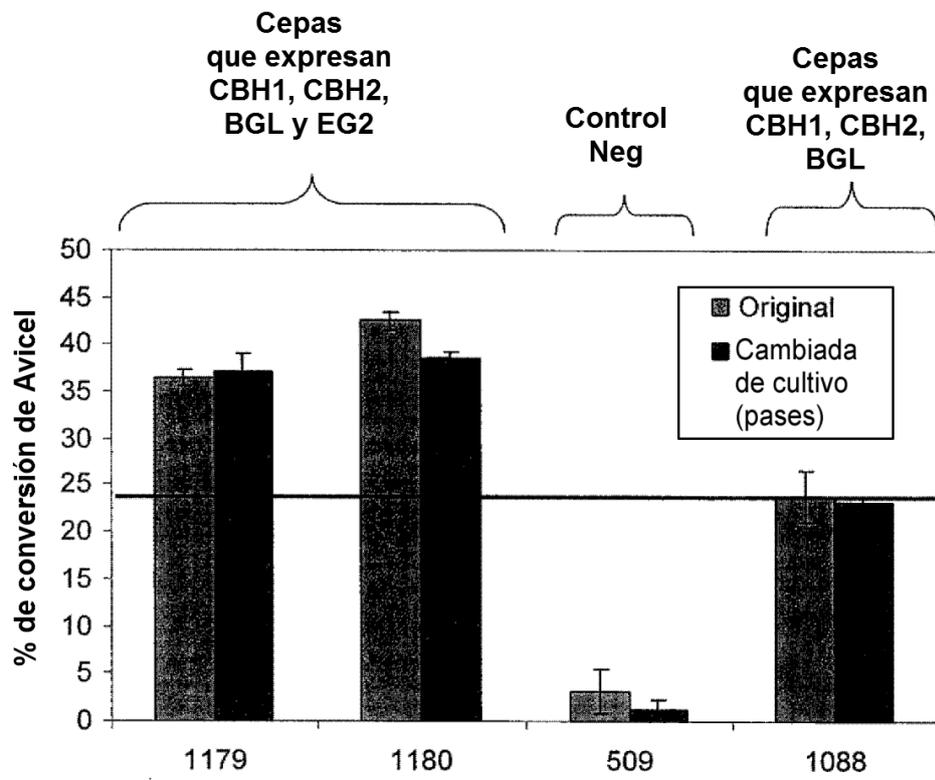


Figura 33

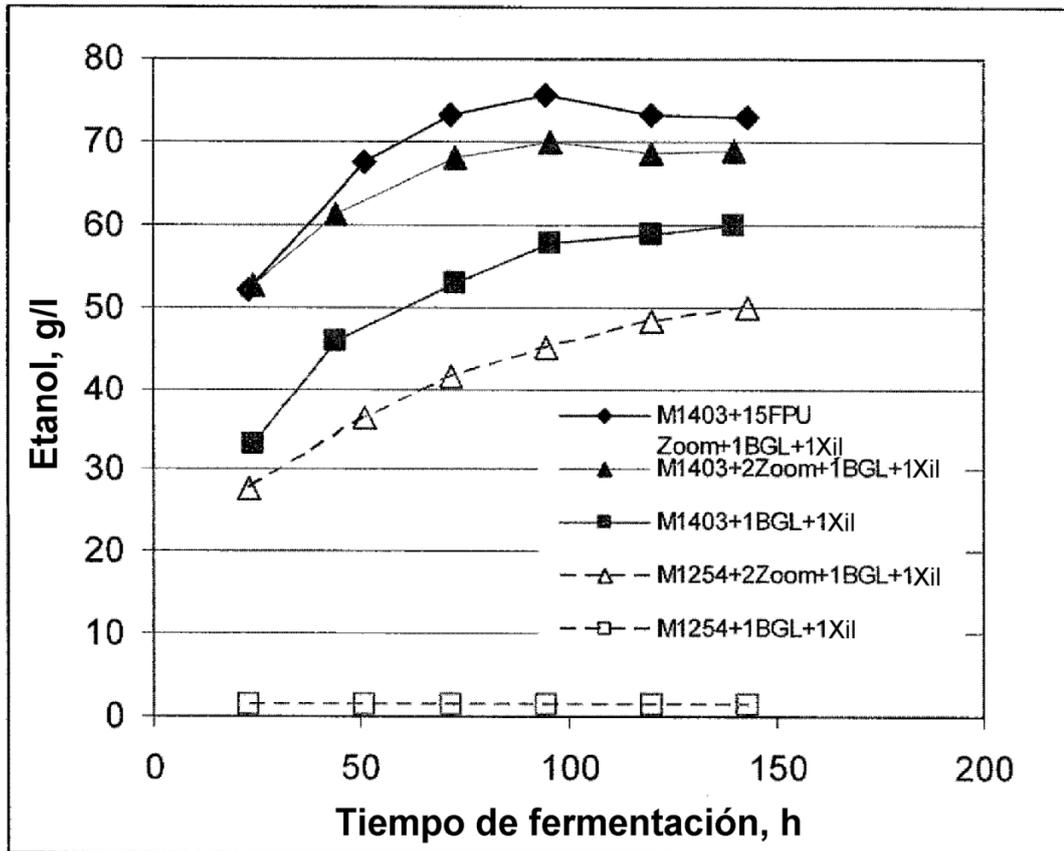


Figura 34

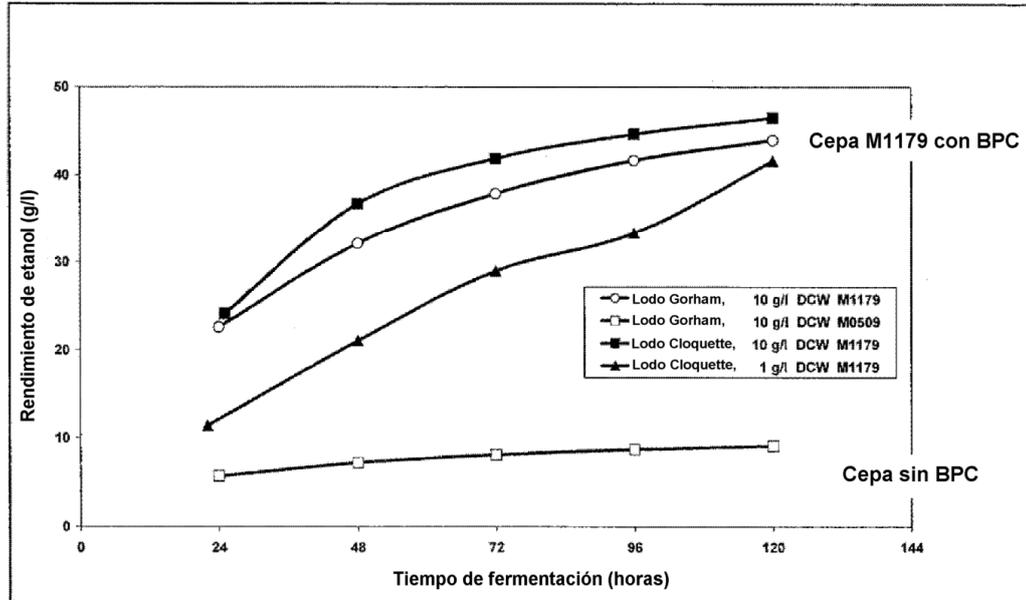


Figura 35

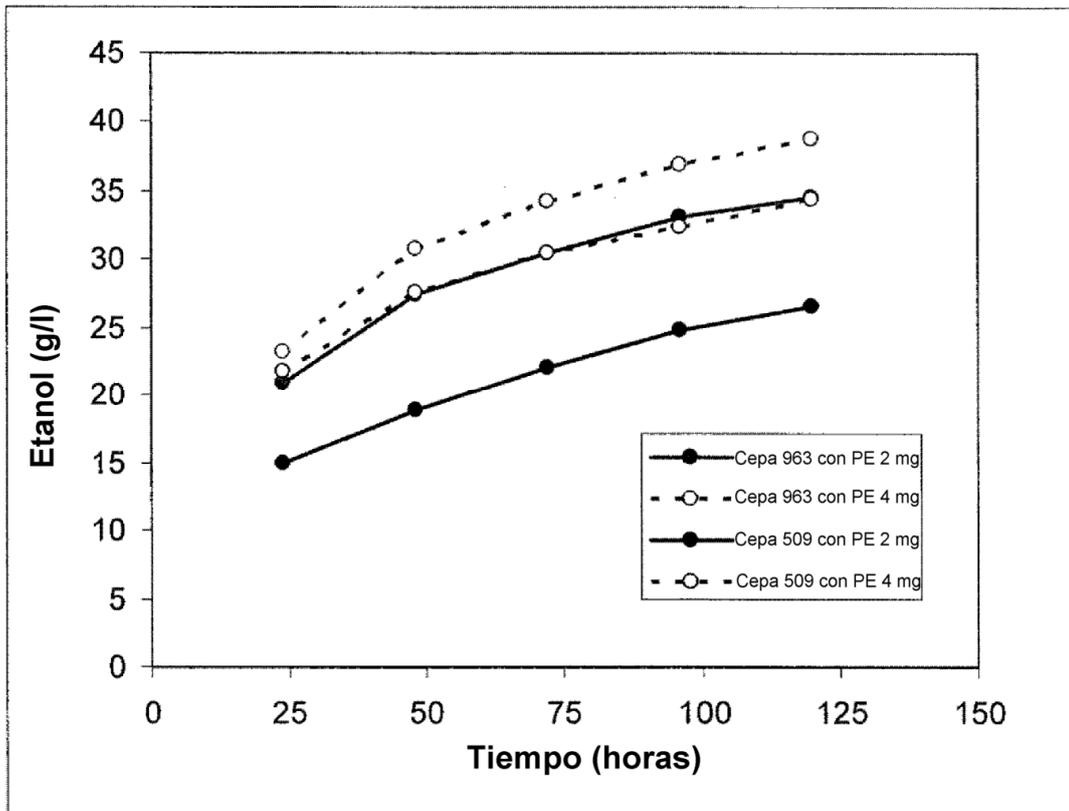


Figura 36

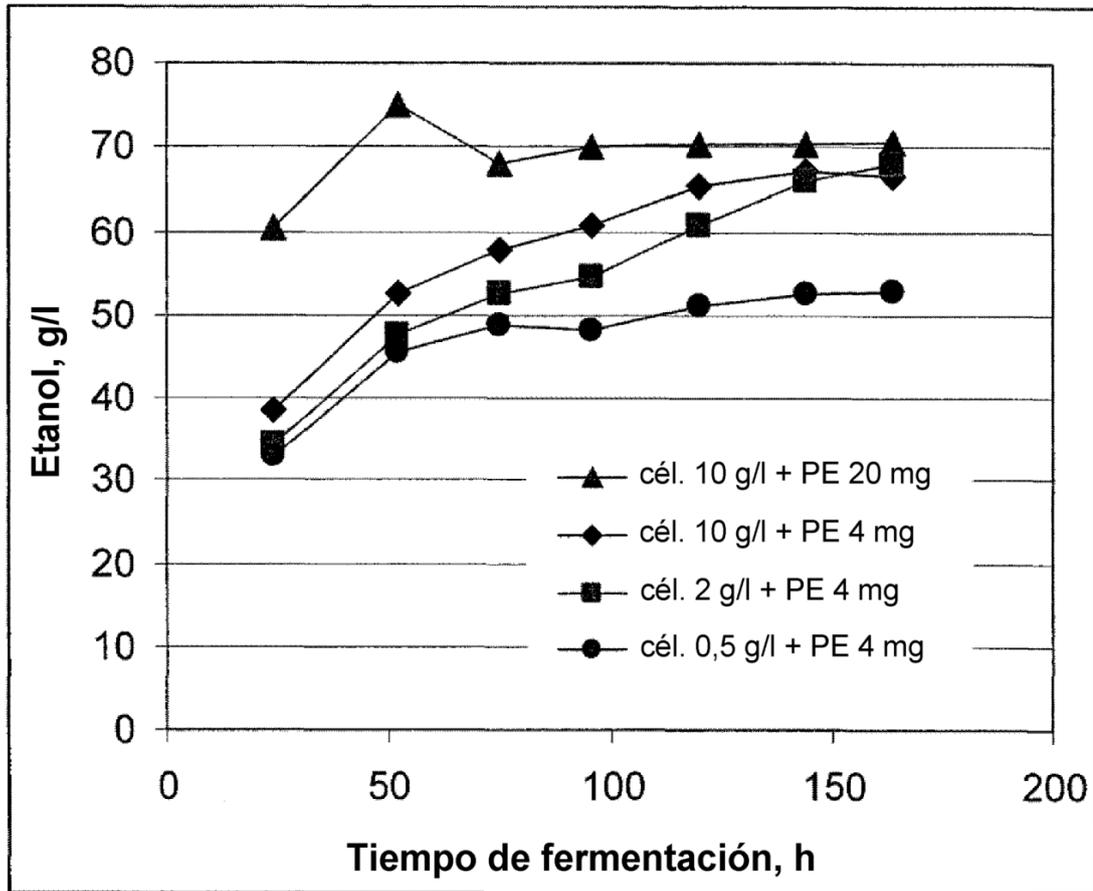


Figura 37

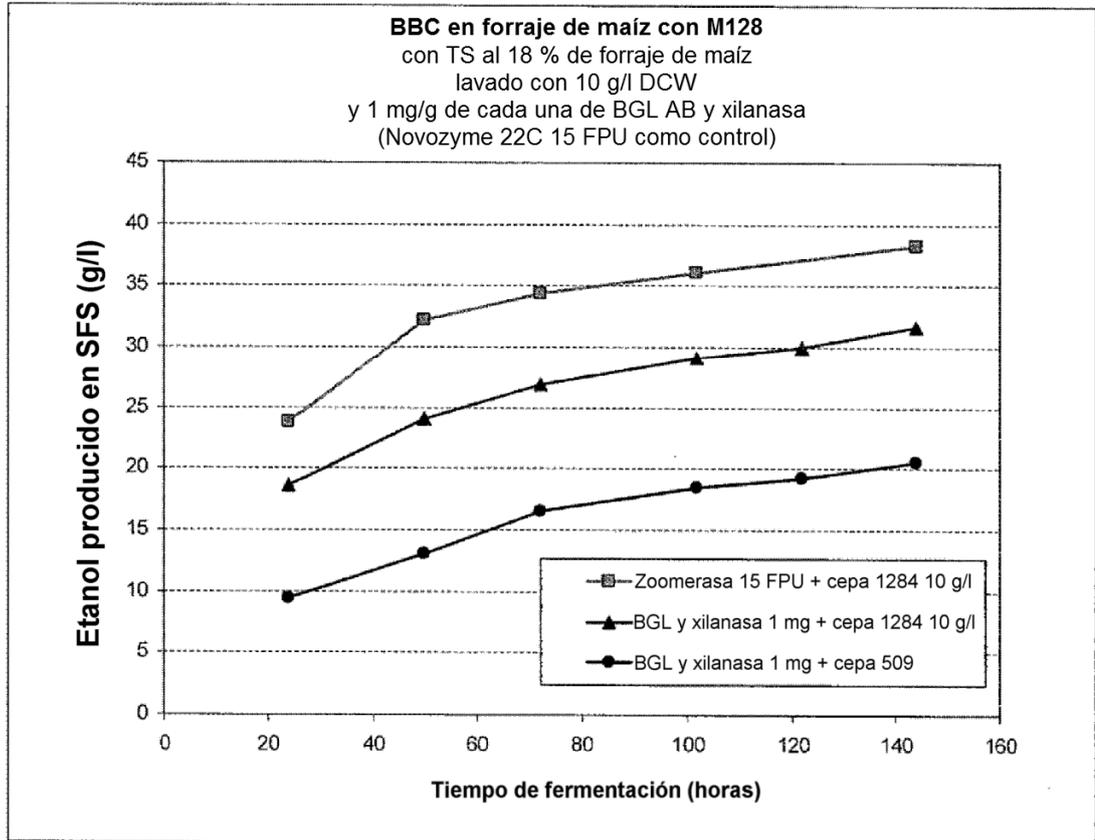


Figura 38A

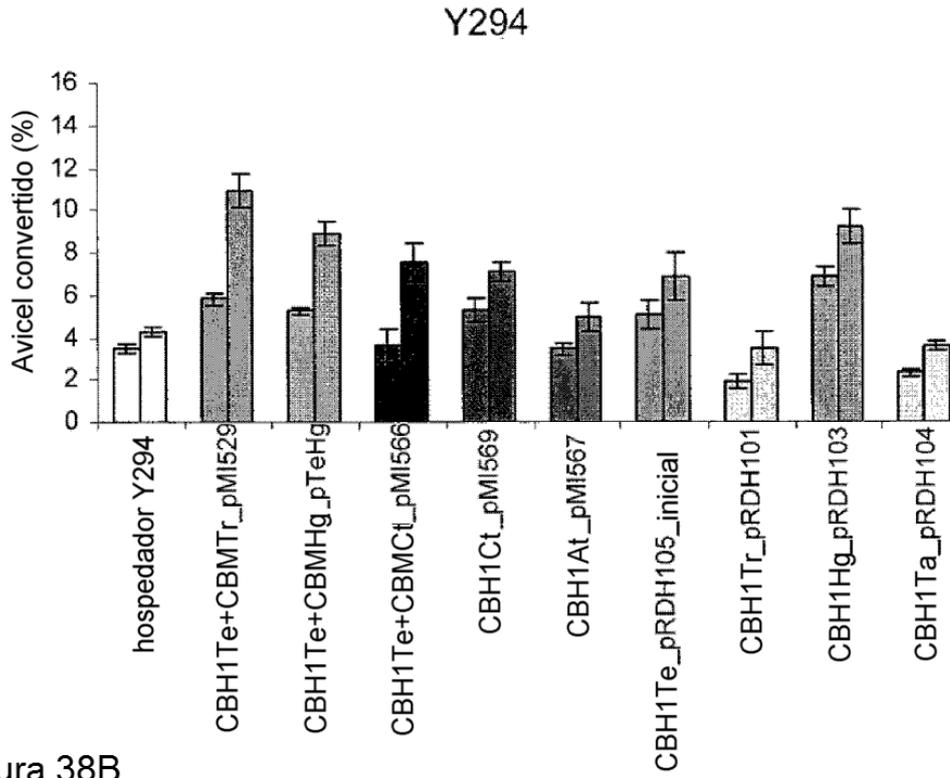


Figura 38B

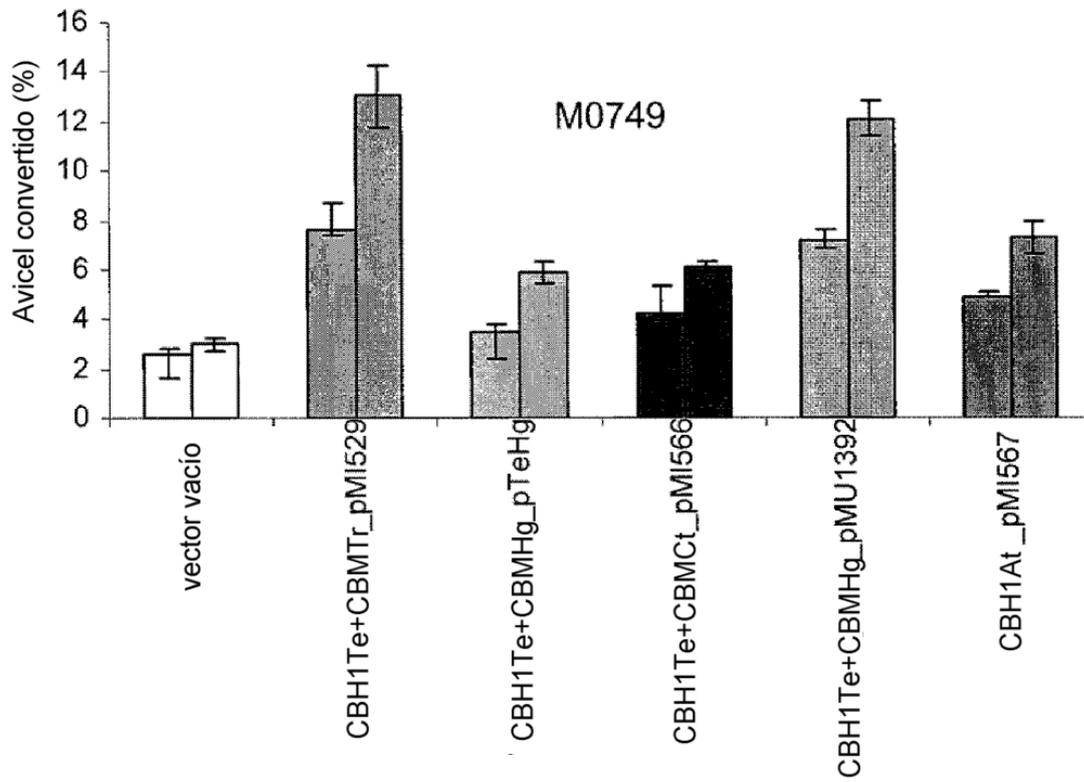


Figura 38C

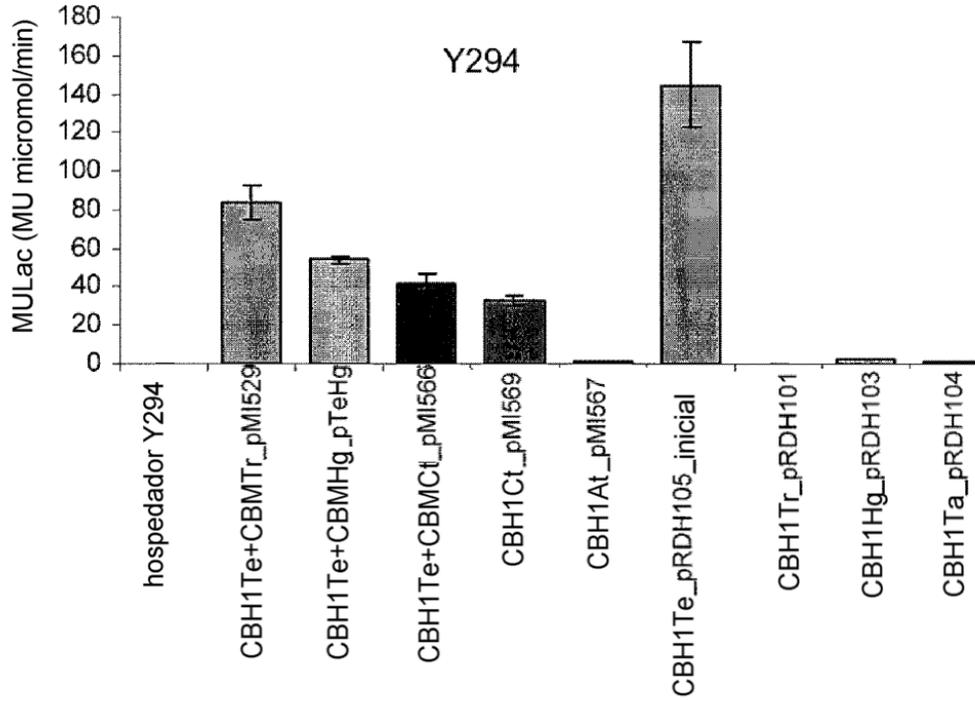


Figura 38D

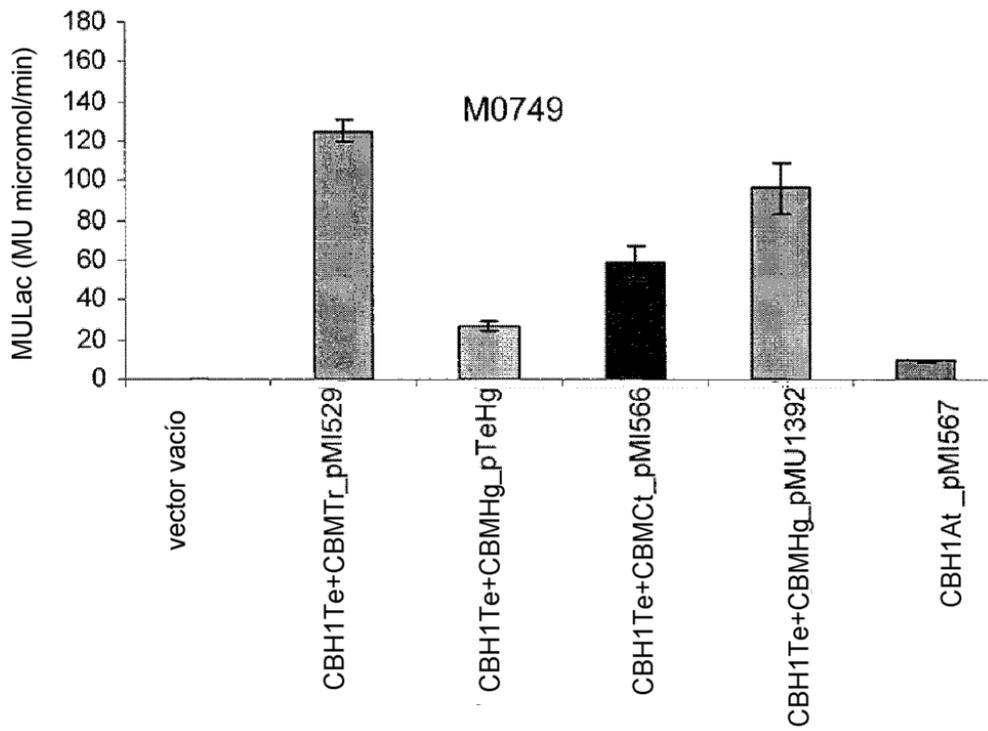


Figura 38E

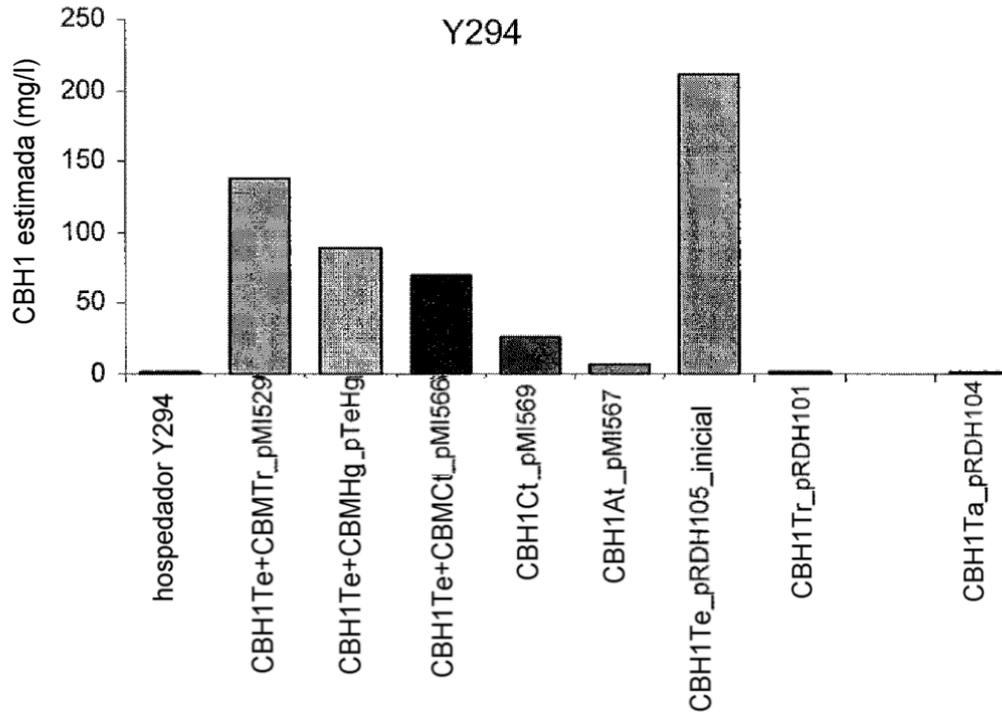


Figura 38F

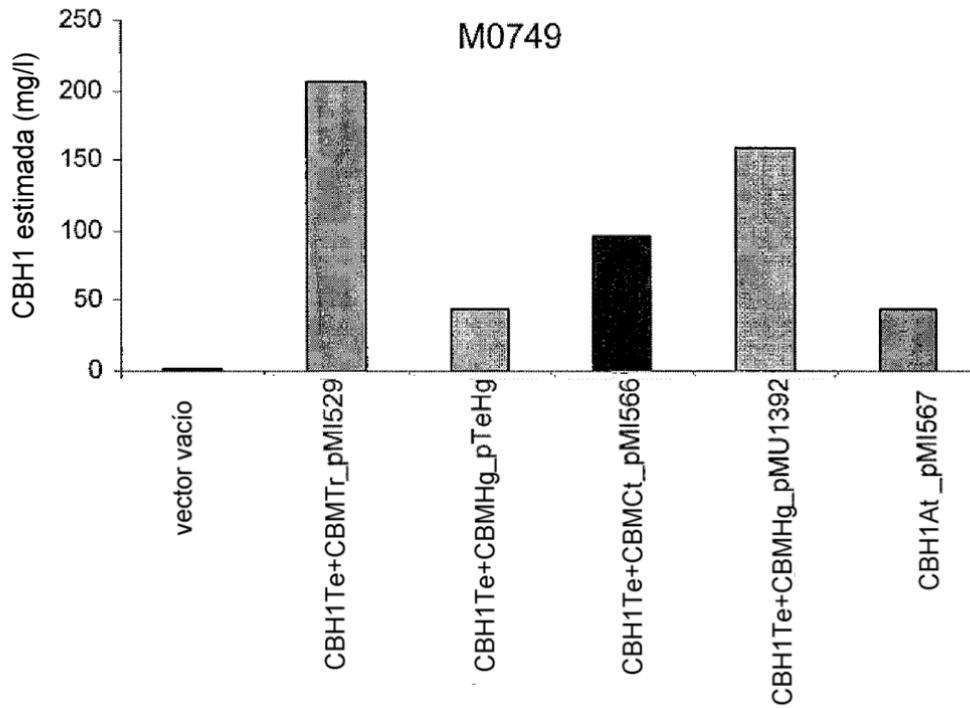


Figura 39

• Genes modificados en M0509.

Gen	Enzima	Número EC	Organismo	Modificación	Promotor	Terminador
XylA	xilosa isomerasa	5.3.1.5	<i>Piromyces</i> sp. E2	integración multicopia	TPI	CYC1
XKS1	xiluloquinasa	2.7.1.17	<i>S. cerevisiae</i>	integración multicopia	ADH1	CYC1
TAL1	transaldolasa	2.2.1.2	<i>S. cerevisiae</i>	expresión aumentada	TPI	-----
TKL1	transcetolasa	2.2.1.1	<i>S. cerevisiae</i>	expresión aumentada	TPI	-----
RPE1	ribulosa-5-fosfato epimerasa	5.1.3.1	<i>S. cerevisiae</i>	expresión aumentada	TPI	-----
RKI1	ribulosa-5-fosfato epimerasa	5.3.1.6	<i>S. cerevisiae</i>	expresión aumentada	TPI	-----
GRE3	aldosa reductasa	1.1.1.21	<i>S. cerevisiae</i>	delecionada	-----	-----

ADH1 = alcohol deshidrogenasa1 de *S. cerevisiae*; TPI = inosa fosfato isomerasa de *S. cerevisiae*; CYC1 = citocromo c, isoforma 1 de *S. cerevisiae*

Figura 40

• Cepas utilizadas para construir M0509 y modificaciones genéticas relevantes

Cepa	Loci genéticos	RPE1	TAL1	TKL1	XylA y XKS1
M0086	GRE3				
M0150	GRE3 ^{+/+}				
M0207	gre3::loxP-gre3::loxP-kan				
M0222	gre3::loxP-gre3::loxP				
M0227	gre3::loxP/gre3::loxP				
M0237	gre3::loxP/gre3::loxP	RPE1 ^{+/+} /loxP-kan-P _{TPI} -RKI1			
M0252	gre3::loxP/gre3::loxP	RPE1 ^{+/+} /loxP-kan-P _{TPI} -RKI1			
M0267	gre3::loxP/gre3::loxP	RPE1 ^{+/+} /loxP-P _{TPI} -RKI1			
M0277	gre3::loxP/gre3::loxP	RPE1 ^{+/+} /loxP-P _{TPI} -RKI1	TAL1 ^{+/+} /loxP-kan-P _{TPI} -TAL1		
M0280	gre3::loxP/gre3::loxP	RPE1 ^{+/+} /loxP-P _{TPI} -RKI1	TAL1 ^{+/+} /loxP-P _{TPI} -TAL1		
M0407	gre3::loxP/gre3::loxP	RPE1 ^{+/+} /loxP-P _{TPI} -RKI1	TAL1 ^{+/+} /loxP-P _{TPI} -TAL1		
M0509	gre3::loxP/gre3::loxP	RPE1 ^{+/+} /loxP-P _{TPI} -RKI1	TAL1 ^{+/+} /loxP-P _{TPI} -TAL1	TKL1 ^{+/+} /loxP-P _{TPI} -TKL	delta::P _{ADH1} -XKS P _{TPI} -XylA -10 copias delta::P _{ADH1} -XKS P _{TPI} -TKL -20 copias

Figura 41

Genealogía de la cepa M1105

Cepa	Selección	Precursor	Genotipo
M0509			<i>gre3::loxP/gre3::loxP TAL1+loxP-PTPI-TAL1 RK11+loxP-PTPI-RK11 RPE1+loxP-PTPI-RPE1 TKL+loxP-PTPI-TKL delta::PTPI-xyA PADH1-XKS</i>
M1017	YPX, 41°C	M0509	genotipo no confirmado después de evolución en citostato
M1046	YMX, 41°C	M1017	genotipo no confirmado después de evolución en citostato
M1080	YMX, 40°C	M1046	genotipo no confirmado después de evolución en citostato
M1105	YPDx+acetato, 39°C	M1080	genotipo no confirmado después de evolución en citostato

Figura 42

Genealogía de la cepa 1254

Cepa	Selección	Precursor	Genotipo
M0509			<i>gre3::loxP/gre3::loxP TAL1+loxP-PTPI-TAL1 RK11+loxP-PTPI-RK11 RPE1+loxP-PTPI-RPE1 TKL+loxP-PTPI-TKL delta::PTPI-xyA PADH1-XKS</i>
M1017	YPX, 41°C	M0509	genotipo no confirmado después de evolución en citostato
M1046	YMX, 41°C	M1017	genotipo no confirmado después de evolución en citostato
M1080	YMX, 40°C	M1046	genotipo no confirmado después de evolución en citostato
M1105	YPDx+acetato, 39°C	M1080	genotipo no confirmado después de evolución en citostato
M1155	Xilosa mínima + acetato 8 g/l, pH 5,5, 40°C	M1105	genotipo no confirmado después de evolución en citostato
M1202	Complejo glucosa + acetato 12 g/l, pH 5,4, 41°C	M1155	genotipo no confirmado después de evolución en citostato
M1254	Hidrolizado MS419 equivalente a 5 % de sólidos + base de nitrógeno de levadura sin aminoácidos, pH 4,8, 39°C	M1202+	genotipo no confirmado después de evolución en citostato
		M1155	