

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 424 346**

51 Int. Cl.:

C12N 9/42 (2006.01)

C12N 15/56 (2006.01)

C12P 19/14 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.08.2004 E 11167129 (3)**

97 Fecha y número de publicación de la concesión europea: **08.05.2013 EP 2377931**

54 Título: **Variantes de glucósido hidrolasas**

30 Prioridad:

25.08.2003 US 497809 P

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

01.10.2013

73 Titular/es:

**NOVOZYMES INC. (100.0%)
1445 Drew Avenue
Davis, CA 95616, US**

72 Inventor/es:

**TETER, SARAH;
CHERRY, JOEL;
WARD, CONNIE;
HARRIS, PAUL;
JONES, AUBREY y
YI, JUNG**

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 2 424 346 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

Variantes de glucósido hidrolasas.

5 [0001] Esta invención fue realizada con el apoyo del Gobierno bajo el Subcontrato de NREL No. ZCO-30017-02, Contrato principal DE-AC36-98GO10337 otorgado por el Departamento de Energía. El gobierno tiene ciertos derechos en esta invención.

10 **Antecedentes de la invención**

Campo de la invención

15 [0002] La presente invención se refiere a variantes de una glucósido hidrolasa con una o más propiedades mejoradas en relación a su enzima progenitora, ácidos nucleicos que codifican las variantes, métodos de producción de las variantes y métodos de uso de las variantes.

Descripción de las técnicas relacionadas

20 [0003] La celulosa es un polímero de la glucosa de azúcar simple unido covalentemente por enlaces beta-1,4. Muchos microorganismos producen enzimas que hidrolizan glucanos betaenlazados. Estas enzimas incluyen endoglucanasas, celobiohidrolasa y beta-glucosidasas. Las endoglucanasas digieren el polímero de celulosa en lugares al azar, abriéndolo al ataque por parte de celobiohidrolasas. Las celobiohidrolasas secuencialmente liberan moléculas de celobiosa de las extremidades del polímero de celulosa. La celobiosa es un dímero de glucosa hidrosoluble con enlace beta-1,4. Las beta-glucosidasas hidrolizan celobiosa a glucosa.

25 [0004] La conversión de materias primas celulósicas en etanol tiene las ventajas de la disponibilidad inmediata de grandes cantidades de materia prima, la conveniencia de evitar quemar o verter en terrenos los materiales y la limpieza del combustible de etanol. La madera, los residuos agrícolas, los brotes herbáceos y los desperdicios sólidos municipales han sido considerados materias primas para la producción de etanol. Estos materiales principalmente consisten en celulosa, hemicelulosa y lignina. Una vez la celulosa se convierte a glucosa, la glucosa es fácilmente fermentada por levadura en el etanol.

30 [0005] Ståhlberg *et al.*, 1996, J. Mol. Biol. 264: 337-349, describen estudios de actividad y estructuras de cristal de mutantes catalíticamente deficitarios de celobiohidrolasa I de *Trichoderma reesei*. Boer and Koivula, 2003, Eur. J. Biochem. 270: 841-848, revela la relación entre la termoestabilidad y el óptimo de pH estudiado con tipo salvaje y mutante de celobiohidrolasa de *Trichoderma reesei* Cel7A.

[0006] WO 2004/016760 divulga variantes de una celobiohidrolasa *Hypocrea jecorina*.

40 [0007] Sería una ventaja en la técnica proporcionar variantes de glucósido hidrolasa con propiedades mejoradas para convertir materiales celulósicos a monosacáridos, disacáridos y polisacáridos. Las propiedades mejoradas incluyen perfiles de actividad alterada dependiente de la temperatura, termoestabilidad, actividad de pH, estabilidad del pH, especificidad de sustrato, especificidad de producto y estabilidad química.

45 [0008] Es un objeto de la presente invención proporcionar variantes de glucósido hidrolasas con propiedades mejoradas en comparación con sus enzimas progenitoras.

Resumen de la invención

50 [0009] La presente invención también se refiere a polipéptidos aislados con actividad de glucósido hidrolasa, que tiene un grado de identidad para los aminoácidos 1 a 513 de SEC ID n°: 2 de al menos 85%, incluso más preferiblemente al menos 90%, de forma más preferible al menos 95%, e incluso de forma más preferible al menos 97%, donde las secuencias de aminoácidos 1 de los polipéptidos difieren de los aminoácidos 1 a 513 de SEC ID n°: 2 en al menos una posición correspondiente a las posiciones 21, 94, 157, 205, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n°: 2, donde la diferencia en al menos una posición es una sustitución en la posición 205 y la variante ha mejorado propiedades en comparación con la glucósido hidrolasa progenitora, donde las propiedades mejoradas son seleccionadas del grupo que consiste en perfil de actividad dependiente de la temperatura, termoestabilidad, pH actividad, pH estabilidad y especificidad de sustrato, especificidad de producto y estabilidad química.

60 [0010] La presente invención también se refiere a secuencias de nucleótidos aisladas que codifican las glucósido hidrolasas o polipéptidos que tienen actividad de glucósido hidrolasas y a constructos de ácidos nucleicos, vectores y células huéspedes que comprenden las secuencias de nucleótidos.

65 [0011] La presente invención también se refiere a métodos para producir variantes de una glucósido hidrolasa progenitora o polipéptidos con actividad de glucósido hidrolasa en una célula huésped.

[0012] La presente invención además se refiere a métodos de uso de las variantes de glucósido hidrolasa en detergentes y en la conversión de celulosa a glucosa.

5 Breve descripción de las figuras

[0013]

10 La Figura 1 muestra un mapa de restricción de pAJ052.
 La Figura 2 muestra un mapa de restricción de pJC106.
 La Figura 3 muestra un mapa de restricción de pAILo1.
 La Figura 4 muestra un mapa de restricción de pBANE10.
 La Figura 5 muestra un mapa de restricción de pAILo2.
 La Figura 6 muestra un mapa de restricción de pCW026.
 15 La Figura 7 muestra un mapa de restricción de pNP776G205R.
 La Figura 8 muestra un mapa de restricción de pMJ04.
 La Figura 9 muestra un mapa de restricción de pMJ06.
 La Figura 10 muestra un mapa de restricción de pMJ09.
 La Figura 11 muestra un mapa de restricción de pCW045.
 20 La Figura 12 muestra un mapa de restricción de pSTM01.
 La Figura 13 muestra un mapa de restricción de pSMK03.
 La Figura 14 muestra un mapa de restricción de pEJG97.
 La Figura 15 muestra la secuencia de ADN genómico y la secuencia de aminoácidos deducida de una beta-glucosidasa *Aspergillus fumigatus* SEC ID n.º: 56 y 57, respectivamente). El péptido señal predicho está subrayado y los intrones predichos están en cursiva.
 25 La Figura 16 muestra la termoestabilidad de la beta-glucosidasa de *Aspergillus fumigatus* a 50° y 65° C.
 La Figura 17 muestra la termoestabilidad de la beta-glucosidasa de *Aspergillus fumigatus* a 70° C.
 La Figura 18 muestra la hidrólisis de celobiosa por beta-glucosidasa de *Aspergillus fumigatus* a 65° C.
 La Figura 19 muestra los perfiles de curso temporal de hidrólisis PCS por la cepa de *Trichoderma reesei* parental RutC30 y la variante de expresión de la cepa 776-M57.

Descripción detallada de la invención

35 [0014] La presente invención se refiere a una variante polipeptídica aislada con actividad de glucósido hidrolasa, que tiene un grado de identidad para los aminoácidos 1 a 513 de SEC ID n.º: 2 de al menos 85%, incluso más preferiblemente al menos 90%, de forma más preferible al menos 95%, e incluso de forma más preferible al menos 97%, donde la secuencia de aminoácidos del polipeptídico difiere de la SEC ID n.º: 2 en al menos una posición correspondiente a las posiciones 21, 94, 157, 205, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2, donde la diferencia en al menos una posición es una sustitución en la posición 205 y la variante ha mejorado propiedades en comparación con la glucósido hidrolasa progenitora, donde las propiedades mejoradas son seleccionadas del grupo que consiste en perfil de actividad dependiente de la temperatura, termoestabilidad, pH actividad, pH estabilidad y especificidad de sustrato, especificidad de producto y estabilidad química.

45 Definiciones

50 [0015] El término "glucósido hidrolasa" es definida en la presente como hidrolasas descritas por Coutinho, P.M. y Henrissat, B., 1999, Carbohydrate-active enzymes: an integrated database approach, in "Recent Advances in Carbohydrate Bioengineering", H.J. Gilbert, G. Davies, B. Henrissat y B. Svensson eds., The Royal Society de Chemistry, Cambridge, págs. 3-12. Ejemplos de glucósido hidrolasas incluyen, de modo enunciativo y no limitativo, celobiohidrolasa, endoglucanasa y exoglucanasa. En una forma de realización preferida, las glucósido hidrolasas pertenecen a familia 7 tal y como se define por Coutinho, P.M. y Henrissat, B., 1999, *supra*.

55 [0016] El término "celobiohidrolasa" es definido aquí como un 1, 4-beta-D-glucano celobiohidrolasa (E.C. 3.2.1.91), que cataliza la hidrólisis de enlaces 1, 4-beta-D-glucosídicos en la celulosa, LA celotetrisa, o cualquier glucosa unida por enlace beta-1, 4 con polímero, que libera celobiosa de los extremos reductores o no reductores de la cadena.. Para fines de la presente invención, la actividad de celobiohidrolasa se determina según los procedimientos descritos por Lever *et al.*, 1972, Anal. Biochem. 47: 273-279 y por van Tilbeurgh *et al.*, 1982, FEBS Letters, 149: 152-156 ; van Tilbeurgh y Claeysens, 1985, FEBS Letters, 187: 283-288 . En la presente invención, se empleó el método de Lever *et al.* para analizar la hidrólisis de la celulosa en forraje de maíz, mientras que el método de van Tilbeurgh *et al.* se utilizó para determinar la actividad de celobiohidrolasa en un derivado de disacárido fluorescente

60 [0017] El término "endoglucanasa" es definido aquí como una endo-1,4-(1,3,1,4)-beta-D-glucan 4-glucanohidrolasa (E.C. n.º 3.2.1.4), que cataliza la endohidrólisis de enlaces 1,4-beta-D-glicosídicos en celulosa, derivados de celulosa (tal como carboximetilcelulosa e hidroxietil celulosa), liquenina, enlaces beta-1,4 en beta-1,3 glucanos mezclados tales como beta-D-glucanos de cereal o xiloglucanos y otro material vegetal que contenga componentes celulósicos. Para

objetivos de la presente invención, la actividad de la endoglucanasa es determinada usando la hidrólisis de carboximetilcelulosa (CMC) según el procedimiento de Ghose, 1987, Pure and Appl. Chem. 59: 257-268 .

5 [0018] El término "exoglucanasa" es definido aquí como un 1,4-beta-D-glucan glucohidrolasa (E.C. 3.2.1.74) que cataliza la hidrólisis de enlaces 1,4 (enlaces O-glicosil) en 1,4-beta-D-glucanos para eliminar unidades sucesivas de glucosa o celobiosa. Para fines de la presente invención, la actividad de exoglucanasa se determina según el procedimiento descrito por Himmel *et al.*, 1986, J. Biol. Chem. 261: 12948-12955 .

10 [0019] Variante: El término "variante" es definido en la presente como una glucósido hidrolasa que comprende una o más alteraciones, tales como sustituciones, inserciones, deleciones, y/o truncamientos de uno o más residuos de aminoácidos específicos en una o más posiciones específicas en el polipéptido.

15 [0020] Enzima tipo salvaje: El término glucósido hidrolasa de "tipo salvaje" denota una glucósido hidrolasa expresada por un microorganismo de origen natural, tal como una levadura u hongo filamentoso encontrado en la naturaleza.

20 [0021] Enzima progenitora: El término glucósido hidrolasa "progenitora" según se utiliza en la presente significa una glucósido hidrolasa a la cual se le realizan modificaciones, por ejemplo, sustitución (sustituciones), inserción (inserciones), deleción (deleciones) y/o truncamiento(s), para producir las variantes enzimáticas de la presente invención. Este término también se refiere al polipéptido con el cual una variante es comparada y alineada. El progenitor puede ser un polipéptido de origen natural (tipo salvaje), o este incluso puede ser una variante del mismo preparada por cualquiera de los medios adecuados. Por ejemplo, la proteína progenitora puede ser una variante de un polipéptido de origen natural que ha sido modificado o alterado en la secuencia de aminoácidos. Un progenitor también puede ser una variante alélica que es un polipéptido codificado por cualquiera de dos o más formas alternativas de un gen que ocupa el mismo locus cromosómico.

25 [0022] Redistribución: El término "redistribución" significa recombinación de secuencia/s de nucleótidos entre dos o más secuencias de nucleótidos homólogos que dan como resultado secuencias de nucleótidos recombinadas (es decir, secuencias de nucleótidos que han sido sometidas a un ciclo de redistribución) con un número de nucleótidos cambiado, en comparación con las secuencias de nucleótidos iniciales.

30 [0023] Biblioteca aleatorizada: El término, "biblioteca aleatoria", "biblioteca de variantes" o "biblioteca" es definido en la presente como una biblioteca de polipéptidos variantes. La diversidad en la biblioteca de variantes puede ser generada por medio de mutagénesis de los genes que codifican las variantes en el nivel de triplete del ADN, de manera que los codones individuales son diversificados, por ejemplo, usando cebadores de secuencias parcialmente aleatorizadas en una reacción de PCR. Se han descrito diferentes técnicas, por las que puede crearse una biblioteca combinatoria diversa diversificando diferentes posiciones de nucleótido en un gen y recombinándolas, por ejemplo, donde estas posiciones están demasiado alejadas para ser cubiertas por un solo cebador oligonucleótido (adicionado o dopado). Estas técnicas incluyen el uso de recombinación *in vivo* de los segmentos de gen individualmente diversificados como se describe en WO 97/07205 en la página 3, líneas 8 a 29. También incluyen el uso de técnicas de redistribución de ADN para crear una biblioteca de genes de longitud total, caracterizada por el hecho de que diferentes segmentos de gen son combinados y caracterizada por el hecho de que cada segmento puede ser diversificado, por ejemplo, por mutagénesis adicionada (Stemmer, 1994; Nature 370: 389-391; patente estadounidense n.º 5 811 238; patente estadounidense n.º 5 605 793 y patente estadounidense n.º 5 830 721). Puede usarse un gen que codifica una proteína "de estructura principal" (polipéptido progenitor de tipo salvaje) como un polinucleótido modelo y combinarlo con uno o más oligonucleótidos mono o bicatenarios como se describe en WO 98/41623 y WO 98/41622. Los oligonucleótidos monocatenarios pueden ser parcialmente aleatorizados durante la síntesis. Los oligonucleótidos bicatenarios pueden ser productos de PCR que incorporan diversidad en una zona específica. En ambos casos, se puede diluir la diversidad con segmentos correspondientes que codifican la secuencia de la proteína de estructura principal para limitar el número medio de cambios que se introducen.

50 [0024] Recombinación: El término "recombinación" es definido en la presente como un proceso caracterizado por el hecho de que los ácidos nucleicos se asocian entre sí en regiones de homología, conduciendo a intercambio de ADN intercatenario entre esas secuencias. Para fines de la presente invención, la recombinación homóloga se determina según los procedimientos resumidos por Paques y Haber, 1999, Microbiology and Molecular Biology Reviews 63 349-404 . La "recombinación homóloga" es definida en la presente como la recombinación en la cual no ocurre ningún cambio en las secuencias de nucleótidos en las regiones de homología relativas a las secuencias de nucleótidos de entrada. Para la recombinación homóloga perfecta, las regiones deberían contener un número suficiente de ácidos nucleicos, tales como 100 a 1500 pares de bases, preferiblemente 400 a 1500 pares de bases y de forma más preferible 800 a 1500 pares de bases, que son altamente homólogos a la secuencia de ácidos nucleicos correspondientes para mejorar la probabilidad de recombinación homóloga. La recombinación puede también ocurrir por recombinación no homóloga. La "recombinación no homóloga" es definida en la presente como recombinación caracterizada por el hecho de que cualquier modo de reparación del ADN que incorpora intercambio de cadenas da como resultado una secuencia de nucleótidos diferente de cualquiera de las secuencias recombinantes.

65 [0025] Propiedad mejorada: El término "propiedad mejorada" es definido en la presente como una característica asociada con una variante que se mejora en comparación con la glucósido hidrolasa progenitora. Tales propiedades

mejoradas incluyen, de modo enunciativo y no limitativo, perfil de actividad alterada dependiente de la temperatura, termoestabilidad, actividad de pH, estabilidad de pH, especificidad de sustrato, especificidad de producto y estabilidad química.

5 [0026] Actividad térmica mejorada: El término "actividad térmica mejorada" es definido en la presente como una alteración del perfil de actividad dependiente de la temperatura de una variante de glucósido hidrolasa a una temperatura específica relativa al perfil de actividad dependiente de la temperatura de la glucósido hidrolasa progenitora. El valor de la actividad térmica proporciona una medida de la eficiencia enzimática en la realización de catálisis de una reacción de hidrólisis sobre una gama de temperaturas. Una glucósido hidrolasa tiene una gama de temperatura específica caracterizada por el hecho de que la proteína es estable y retiene su actividad enzimática, pero se vuelve menos estable y así menos activa con temperatura en aumento. Además, el nivel inicial de una reacción catalizada por una glucósido hidrolasa se puede acelerar por un aumento en la temperatura que se mide determinando la actividad térmica de una variante. Una variante más termoactiva llevará a un aumento en el índice de hidrólisis disminuyendo el tiempo requerido y/o disminuyendo la concentración enzimática requerida para la hidrólisis. De forma alternativa, una variante con una actividad térmica reducida catalizará una reacción de hidrólisis a una temperatura inferior a la temperatura óptima de la enzima progenitora definida por el perfil de actividad dependiente de la temperatura del progenitor.

20 [0027] Termoestabilidad mejorada: El término "termoestabilidad mejorada" es definido en la presente como una enzima variante que muestra retención de actividad enzimática después de un periodo de incubación a temperatura elevada relativa a la enzima progenitora. Tal variante puede o no mostrar un perfil de actividad térmica alterada relativa al progenitor. Por ejemplo, una variante puede tener una capacidad mejorada para replegarse después de la incubación a temperatura elevada relativa al progenitor.

25 [0028] En una forma de realización preferida, la actividad térmica de la glucósido hidrolasa variante es al menos 1,5 veces, preferiblemente al menos 2 veces, más preferiblemente al menos 5 veces, de forma más preferible al menos 7 veces, e incluso de forma más preferible al menos 20 veces más activa térmicamente que la variante de tipo salvaje cuando la actividad en 4-metilumbeliferil beta-d-lactosida a 64° C o una temperatura más alta es comparada con actividad a 50° C, durante 45 minutos a pH 5.0.

30 [0029] Especificidad de producto mejorado: El término "especificidad de producto mejorada" es definido en la presente como una enzima variante que muestra un perfil de producto alterado relativo al progenitor en el cual el perfil de producto alterado mejora el rendimiento de la variante en una aplicación relativa al progenitor. El término "perfil del producto" es definido en la presente como la composición química de los productos reactivos producida por hidrólisis enzimática.

35 [0030] Estabilidad química mejorada: El término "estabilidad química mejorada" es definido en la presente como una enzima variante que muestra retención de actividad enzimática después de un periodo de incubación en presencia de un producto químico o productos químicos, de origen natural o bien sintético, que reducen la actividad enzimática de la enzima progenitora. La estabilidad química mejorada puede también suponer variantes más capaces de catalizar una reacción en presencia de tales productos químicos.

Convenciones para la designación de las variantes

45 [0031] En la presente invención, se emplea una numeración específica de las posiciones de residuos de aminoácido en las variantes de glucósido hidrolasa. Por ejemplo, alineando las secuencias de aminoácidos de glucósido hidrolasas conocidas, es posible designar un número de posición de aminoácido a cualquier residuo de aminoácido en cualquier enzima de glucósido hidrolasa.

50 [0032] Uso del sistema de numeración originado de la secuencia de aminoácidos de la glucósido hidrolasa descrito en SEC ID n°: 2, alineado con la secuencia de aminoácidos de una cantidad de otras glucósido hidrolasa, es posible indicar la posición de un residuo de aminoácido en una glucósido hidrolasa en regiones de homología estructural.

55 [0033] Pueden hacerse alineamientos múltiples de secuencias proteínicas, por ejemplo, usando "ClustalW" (Thompson, J.D., Higgins, D.G. y Gibson, T.J., 1994, CLUSTAL W: Improving the sensitivity de progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, Nucleic Acids Research 22: 4673-4680). Pueden realizarse alineamientos múltiples de secuencias de ADN usando la alineación de proteína como un molde, sustituyendo los aminoácidos con el codón correspondiente de la secuencia de ADN.

60 [0034] Los algoritmos de comparación de secuencias por parejas de uso común son adecuados para detectar similitudes entre secuencias proteínicas que no se han bifurcado más allá del punto de aproximadamente 20-30% de identidad de secuencia (Doolittle, 1992, Protein Sci. 1: 191-200; Brenner *et al.*, 1998, Proc. Natl. Acad. Sci. USA 95, 6073-6078). No obstante, las proteínas realmente homólogas con el mismo pliegue y función biológica similar frecuentemente se han bifurcado al punto en el cual las comparaciones basadas en la secuencia tradicional no detectan su relación (Lindahl y Eldesson, 2000, J. Mol. Biol. 295: 613-615). Puede lograrse una sensibilidad superior en la búsqueda basada en secuencia usando programas de búsqueda que utilizan representaciones probabilísticas de

familias de proteína (perfiles) para buscar bases de datos. Por ejemplo, el programa PSI-BLAST genera perfiles a través de un proceso de búsqueda de base de datos iterativo y es capaz de detectar homólogos remotos (Atschul *et al.*, 1997, Nucleic Acids Res. 25: 3389-3402). Puede lograrse una sensibilidad aún mayor si la familia o superfamilia para la proteína de interés tiene uno o más representantes en las bases de datos de estructura de proteína. Programas tales como GenTHREADER (Jones 1999, J. Mol. Biol. 287: . 797-815; McGuffin y Jones, 2003, Bioinformatics 19: 874-881) utilizan información de una variedad de fuentes (PSI- BLAST, predicción de estructura secundaria, perfiles de alineación estructural y potenciales de disolución) como entrada a una red neuronal que predice el pliegue estructural para una secuencia problema. De forma similar, el método de Gough *et al.*, 2000, J. Mol. Biol. 313: 903 919, puede utilizarse para alinear una secuencia de estructura desconocida con los modelos de superfamilia presentes en la base de datos SCOP. Estos alineamientos pueden ser usados sucesivamente para generar modelos de homología para la proteína de interés y tales modelos pueden evaluarse en cuanto a su exactitud usando una variedad de herramientas desarrolladas para tal fin.

[0035] Para las proteínas de estructura conocida, están disponibles diferentes herramientas y recursos para recuperar y generar alineamientos estructurales. Por ejemplo, las superfamilias de SCOP de proteínas han sido estructuralmente alineadas y esos alineamientos son accesibles y pueden descargarse. Estos alineamientos pueden utilizarse para predecir los residuos de aminoácidos correspondientes estructuralmente y funcionalmente en proteínas en la misma superfamilia estructural. Esta información, junto con información derivada de modelado por homología y búsquedas de perfiles, pueden utilizarse para predecir qué residuos van a mutar cuando se mueven las mutaciones de interés de una proteína a un homólogo remoto o cercano.

[0036] Al describir las diferentes variantes de glucósido hidrolasa de la presente invención, la nomenclatura descrita a continuación se adapta para facilidad de referencia. En cualquier caso, se utilizan las abreviaturas de aminoácidos aceptadas de IUPAC de una letra o tres letras.

[0037] Sustituciones. Para una sustitución de aminoácido, se usa la siguiente nomenclatura: aminoácido original, posición, aminoácido sustituido. Por consiguiente, la sustitución de treonina con alanina en la posición 226 es designada como "Thr226Ala" o "T226A". Las mutaciones múltiples se separan por marcas de adición ("+"), por ejemplo, "Gly205Arg + Ser411Phe" o "G205R + S411F", representando mutaciones en las posiciones 205 y 411 sustituyendo la glicina (G) con arginina (R) y la serina (S) con fenilalanina (F), respectivamente.

[0038] Deleciones. Para una deleción de aminoácido, se usa la siguiente nomenclatura: aminoácido original, posición*. Por consiguiente, la deleción de glicina en la posición 195 es designada como "Gly195*" o "G195*". Las deleciones múltiples se separan por marcas de adición ("+"), por ejemplo, "Gly195* + Ser411*" o "G195* + S411*".

[0039] Inserciones. Para una inserción de aminoácido, se usa la siguiente nomenclatura: Aminoácido original, posición, aminoácido original, aminoácido nuevo insertado. Por consiguiente la inserción de lisina después de la glicina en la posición 195 es designada "Gly195GlyLys" o "G195GK".

[0040] Modificaciones múltiples. Las variantes que comprenden modificaciones múltiples se separan por marcas de adición ("+"), por ejemplo, "Arg170Tyr+Gly195Glu" o "R170Y+G195E" que representan modificaciones en las posiciones en 170 y 195 substituyendo ácido glutámico y tirosina por arginina y glicina, respectivamente.

Glucósido hidrolasas progenitoras

[0041] En la presente invención, la glucósido hidrolasa progenitora es (a) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 70% de identidad con aminoácidos 1 a 513 de SEC ID n.º: 2 o (b) un polipéptido codificado por una secuencia de nucleótidos que se hibridiza bajo condiciones de astringencia al menos bajas con los nucleótidos 52 a 1539 de SEC ID .nº: 1, o su hebra complementaria.

[0042] La glucósido hidrolasa progenitora comprende una secuencia de aminoácidos que tiene un grado de identidad de aminoácidos 1 a 513 de SEC. ID n.º: 2 de al menos 70%, preferiblemente al menos 75%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, incluso más preferiblemente al menos 90%, de forma más preferible al menos 95%, e incluso de forma más preferible al menos 97%, que tiene actividad de glucósido hidrolasa (en adelante, "polipéptidos homólogos"). Para objetivos de la presente invención, el grado de identidad entre dos secuencias de aminoácidos se determina por el método de Clustal (Higgins, 1989, CABIOS 5: 151-153) usando el Sdeware de MEGALIGN™ de LASERGENE™ (DNASTAR, Inc., Madison, WI) con una tabla de identidad y los siguientes parámetros de alineación múltiple: Penalización por espacio de 10 y penalización por longitud de espacio de 10. Los parámetros de alineación en pareja fueron Ktuple=1, penalización por espacio =3, ventanas =5 y diagonales=5.

[0043] Las glucósido hidrolasas progenitoras sustancialmente homólogas pueden tener una o más sustituciones, deleciones o adiciones de aminoácidos. Estos cambios son preferiblemente de una naturaleza menor, es decir, sustituciones de aminoácido conservador como se ha descrito anteriormente y otras sustituciones que no afectan significativamente al pliegue tridimensional o a la actividad de la proteína o el polipéptido; deleciones pequeñas, típicamente de uno a aproximadamente 30 aminoácidos; y pequeñas extensiones del amino o carboxilo-terminal, tales como un residuo de metionina aminoterminal, un pequeño péptido de enlace de hasta 20-25 residuos aproximadamente, o una pequeña extensión que facilita la purificación (una marca de afinidad), tal como un tracto de

polihistidina, o proteína A (Nilsson *et al.*, 1985, EMBO J. 4: 1075 ; Nilsson *et al.*, 1991, Methods Enzymol. 198: 3 . Véase, también, en general, Ford *et al.*, 1991, Protein Expression and Purification 2: 95-107 . Ejemplos de modificaciones conservadoras están en el grupo de aminoácidos básicos (arginina, lisina e histidina), aminoácidos
 5 acídicos (ácido glutámico y ácido aspártico), aminoácidos polares (glutamina y asparagina), aminoácidos hidrófobos (leucina, isoleucina y valina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina) y pequeños aminoácidos (glicina, alanina, serina, treonina y metionina). Las modificaciones de aminoácidos, que generalmente no alteran la actividad específica son conocidas en la técnica y son descritas, por ejemplo, por H. Neurath y R.L. Hill, 1979, en The
 10 Proteins, Academic Press, New York. Los intercambios que ocurren más frecuentemente son Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu y Asp/Gly al igual que lo inverso (Taylor, 1986, Journal de Theoretical Biology 119: 205-218 .

[0044] Aunque los cambios anteriormente descritos son preferiblemente de una naturaleza menor, tales cambios también pueden ser de una naturaleza considerable tal como fusión de polipéptidos más grandes de hasta 300 aminoácidos o más tanto como extensiones amino-terminales como carboxilo-terminales.

[0045] Además de los 20 aminoácidos estándares, los aminoácidos no estándares (como 4-hidroxiprolina, 6-*N*-metil lisina, ácido 2-aminoisobutírico, isovalina y alfa-metil serina) pueden sustituirse por residuos de aminoácidos de una glucósido hidrolasa de tipo salvaje. Un número limitado de aminoácidos no conservadores, aminoácidos que no son
 20 "aminoácidos no naturales" han sido modificados después de la síntesis de proteína y/o tienen una estructura química en su/s cadena/s lateral/es diferente de los aminoácidos estándares. Los aminoácidos no naturales pueden ser sintetizados químicamente y preferiblemente están disponibles comercialmente, e incluyen ácido pipécólico, ácido carboxílico de tiazolidina, dehidroprolina, 3- y 4-metilprolina y 3,3-dimetilprolina.

[0046] Preferiblemente, la glucósido hidrolasa progenitora comprende la secuencia de aminoácidos de SEC ID n.º: 2, o una variante alélica de la misma; o un fragmento de la misma que tiene actividad de glucósido hidrolasa. En una forma de realización preferida, el polipéptido progenitor comprende la secuencia de aminoácidos de SEC ID n.º: 2. En otra forma de realización preferida, el polipéptido progenitor comprende los aminoácidos 1 a 513 de SEC ID n.º: 2, o una
 30 variante alélica de la misma; o un fragmento de la misma que tiene actividad de glucósido hidrolasa. En otra forma de realización preferida, el polipéptido progenitor comprende los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, el polipéptido progenitor consiste en la secuencia de aminoácidos de SEC ID n.º: 2, o una variante alélica de la misma; o un fragmento de la misma que tiene actividad de glucósido hidrolasa. En otra forma de realización preferida, el polipéptido progenitor consiste en la secuencia de aminoácidos de SEC ID n.º: 2. En otra forma de realización preferida, el polipéptido progenitor consiste en los aminoácidos 1 a 513 de SEC ID n.º: 2 o una variante alélica de la misma; o un fragmento de la misma que tiene actividad de glucósido hidrolasa. En otra forma de realización preferida, el polipéptido progenitor es codificado por la secuencia de nucleótidos contenida en el plásmido pAJ052 que está contenido en *Escherichia coli* NRRL B- 30683, caracterizado por el hecho de que la secuencia de ácidos nucleicos codifica un polipéptido que tiene actividad de glucósido hidrolasa. En otra forma de realización preferida, el polipéptido progenitor es codificado por la región de codificación de polipéptido maduro contenida en el plásmido pAJ052 que está
 40 contenido en *Escherichia coli* NRRL B-30683.

[0047] Un fragmento de SEC ID n.º: 2 es un polipéptido que tiene uno o más aminoácidos delecionados del término amino y/o carboxilo de esta secuencia de aminoácidos. Preferiblemente, un fragmento contiene al menos 450 residuos de aminoácidos, más preferiblemente al menos 470 residuos de aminoácidos y de forma más preferible al menos 490
 45 residuos de aminoácidos .

[0048] La glucósido hidrolasa progenitora es codificada por una secuencia de nucleótidos que hibridiza bajo condiciones de astringencia baja, preferiblemente condiciones de astringencia media, más preferiblemente, condiciones de astringencia media-alta, incluso más preferiblemente, condiciones de astringencia alta, y de forma más preferible, condiciones de astringencia muy alta con una sonda de nucleótidos que se hibridiza bajo las mismas condiciones con (i)
 50 (i) los nucleótidos 52 a 1539 de SEC ID n.º: 1, (ii) la secuencia genómica de nucleótidos que comprende los nucleótidos 52 a 1539 de SEC ID n.º: 1, (iii) una subsecuencia de (i) o (ii), o (iv) una cadena complementaria de (i); (ii), o (iii) (J. Sambrook, E.F. Fritsch y T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 2ª edición, Cold Spring Harbor, New York). La subsecuencia de la SEC ID n.º: 1 puede ser al menos 100 nucleótidos contiguos o preferiblemente al menos 200 nucleótidos contiguos. Por otra parte, la subsecuencia puede codificar un fragmento polipeptídico que tiene actividad de glucósido hidrolasa.

[0049] Una subsecuencia de SEC ID n.º: 1, u homólogo de la misma, es una secuencia de nucleótidos en la cual uno o más nucleótidos han sido delecionados del extremo 5' y/o 3'. Preferiblemente, una subsecuencia contiene al menos 1350 nucleótidos, más preferiblemente al menos 1410 nucleótidos y de forma más preferible al menos 1470 nucleótidos.

[0050] El polipéptido progenitor también puede ser una variante alélica de un polipéptido que tiene actividad de glucósido hidrolasa. Una variante alélica denota cualquiera de dos o más formas alternativas de un gen que ocupa el mismo locus cromosómico. La variación alélica surge naturalmente a través de la mutación y puede suponer
 65 polimorfismo dentro de las poblaciones. Las mutaciones de gen pueden ser silenciosas (ningún cambio en el polipéptido

codificado) o pueden codificar polipéptidos que tienen secuencias de aminoácidos alteradas. Una variante alélica de un polipéptido es un polipéptido codificado por una variante alélica de un gen.

5 [0051] La secuencia de nucleótidos de SEC ID n.º: 1 o una subsecuencia de la misma, al igual que la secuencia de aminoácidos de SEC ID n.º: 2, o un fragmento de la misma, se puede utilizar para diseñar sondas de nucleótido a identificar y clonar polipéptidos progenitores de codificación de ADN que tienen actividad de glucósido hidrolasa de cepas de diferentes géneros o especies según métodos bien conocidos en la técnica. En particular, tales sondas se pueden usar para hibridación con el ADN genómico o ADNc del género o la especie de interés, siguiendo procedimientos estándares de hibridación de Southern, para identificar y aislar el gen correspondiente. Tales sondas pueden ser considerablemente más cortas que la secuencia entera, pero deberían ser de al menos 15, preferiblemente al menos 25 y más preferiblemente al menos 35 nucleótidos de longitud. También pueden utilizarse sondas más largas. Pueden utilizarse tanto la sonda de ADN como la de ARN. Las sondas típicamente son marcadas para detectar el gen correspondiente (por ejemplo, con ³²P, ³H, ³⁵S, biotina o avidina).

15 [0052] Una genoteca de ADN genómico o ADNc obtenido a partir de esos otros organismos se pueden seleccionar para ADN que se hibrida con las sondas anteriormente descritas y que codifica un polipéptido progenitor que tiene actividad de glucósido hidrolasa.

El ADN genómico u otro ADN de esos otros organismos se pueden separar por agarosa o electroforesis en gel de poliacrilamida u otras técnicas de separación. El ADN de las bibliotecas o el ADN separado se puede transferir y ser inmovilizado en nitrocelulosa u otro material portador adecuado. Para identificar un clon o ADN que sea homólogo a SEC ID n.º: 1, o una subsecuencia de la misma, el material portador se usa en una hibridación de Southern. Para objetivos de la presente invención, la hibridación indica que la secuencia de nucleótidos hibridiza a una sonda de nucleótidos marcada correspondiente a la secuencia de nucleótidos mostrada en SEC ID n.º: 1, o su cadena complementaria, o una subsecuencia de la misma, bajo condiciones de astringencia de baja a muy alta. Las moléculas a las cuales la sonda se hibrida pueden ser detectadas usando, por ejemplo, película radiográfica o cualquier medio de detección conocido en la técnica.

[0053] En una forma de realización preferida, la sonda de nucleótidos es una secuencia de nucleótidos que codifica el polipéptido de SEC ID n.º: 2, o una subsecuencia de de la mismo. En otra forma de realización preferida, la sonda de nucleótidos es SEC ID n.º: 1. En otra forma de realización preferida, la sonda de nucleótidos son los nucleótidos 52 a 1539 de SEC ID n.º: 1. En otra forma de realización preferida, la sonda de nucleótidos es la secuencia de ácidos nucleicos contenida en el plásmido pAJ052 que está contenido en *Escherichia coli* NRRL B-30683, caracterizada por el hecho de que la secuencia de ácidos nucleicos codifica un polipéptido que tiene actividad de glucósido hidrolasa. En otra forma de realización preferida, la sonda de nucleótidos es la región de codificación de polipéptido maduro contenida en plásmido pAJ052 que está contenido en *Escherichia coli* NRRL B-30683.

[0054] Para las sondas largas de al menos 100 nucleótidos de longitud, las condiciones de astringencia baja a muy alta son definidas como prehibridación e hibridación a 42° C en 5X SSPE, SDS al 0,3%, 200 µg/ml de ADN de esperma de salmón cortado y desnaturalizado y bien formamida al 25% para astringencias bajas, formamida al 35% para astringencias medias y medias-altas o formamida al 50% para astringencias altas y muy altas, siguiendo procedimientos estándares de hibridación de Southern durante 12 a 24 horas óptimamente.

[0055] Para las sondas largas de al menos 100 nucleótidos de longitud, el material portador es finalmente lavado tres veces cada uno durante 15 minutos usando 2X SSC, SDS al 0,2% preferiblemente al menos a 50° C (astringencia baja), más preferiblemente al menos a 55° C (astringencia media), más preferiblemente al menos a 60° C (astringencia media-alta), de forma más preferible al menos a 65° C (astringencia alta), e incluso de forma más preferible al menos a 70° C (astringencia muy alta).

[0056] Para las sondas cortas que son de aproximadamente 15 nucleótidos a aproximadamente 70 nucleótidos de longitud, las condiciones de astringencia son definidas como prehibridación, hibridación y lavado post-hibridación a aproximadamente 5° C hasta aproximadamente 10° C por debajo de la T_m calculada usando el cálculo según Bolton y McCarthy (1962, Proceedings de the National Academy de Sciences USA 48:1390) en 0,9 M de NaCl, 0,09 M de tris-HCl pH 7,6, 6 mM de EDTA, NP-40 al 0,5%, 1X Solución de Denhardt, 1 mM de pirdeosfato de sodio, 1 mM de fosfato monobásico de sodio, 0,1 mM de ATP y 0,2 mg de ARN de levadura por ml siguiendo procedimientos estándares de hibridación de Southern durante 12 a 24 horas óptimamente.

[0057] Para las sondas cortas que son de aproximadamente 15 nucleótidos a aproximadamente 70 nucleótidos de longitud, el material portador se lava una vez en 6X SSC más SDS al 0,1% durante 15 minutos y dos veces cada una durante 15 minutos usando 6X SSC a 5° C a 10° C por debajo de la T_m calculada.

[0058] La glucósido hidrolasa progenitora se puede obtener a partir de microorganismos de cualquier género. Para objetivos de la presente invención, el término "obtenido a partir de" según se utiliza en este caso en relación con una fuente dada significa que la glucósido hidrolasa progenitora codificada por una secuencia de nucleótidos es producida por la fuente o por una célula en la cual la secuencia de nucleótidos de la fuente ha sido insertada. En una forma de realización preferida, la glucósido hidrolasa progenitora es segregada extracelularmente.

- [0059] La glucósido hidrolasa progenitora puede ser una glucósido hidrolasa fúngica. En una forma de realización más preferida, la glucósido hidrolasa fúngica es una glucósido hidrolasa de levadura tal como una glucósido hidrolasa de *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*. En otra forma de realización más preferida, la glucósido hidrolasa fúngica es una glucósido hidrolasa filamentosa fúngica tal como una glucósido hidrolasa de *Acremonium*, *Agaricus*, *Alternaria*, *Aspergillus*, *Botryosphaeria*, *Ceriporiopsis*, *Chaetomidium*, *Claviceps*, *Cochliobolus*, *Coprinopsis*, *Coptotermes*, *Corynascus*, *Cryphonectria*, *Diplodia*, *Exidia*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*, *Leptosphaeria*, *Magnaporthe*, *Melanocarpus*, *Meripilus*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Phanerochaete*, *Poitrasia*, *Pseudoplectania*, *Pseudotrichonympha*, *Rhizomucor*, *Scytalidium*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Trichoderma*, *Trichophaea*, *Verticillium*, *Volvariella* o *Xylaria*.
- [0060] En una forma de realización más preferida, la glucósido hidrolasa progenitora es una glucósido hidrolasa de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* o *Saccharomyces oviformis*.
- [0061] En otra forma de realización más preferida, la glucósido hidrolasa progenitora es una glucósido hidrolasa de *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus fumigatus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcocroum*, *Fusarium solani*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola grisea*, *Humicola insolens*, *Humicola lanuginosa*, *Irpex lacteus*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium funiculosum*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Schizophyllum commune*, *Sclerotium rolfsii*, *Sporotrichum cellulophilum*, *Talaromyces emersonii*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.
- [0062] En una forma de realización aún más preferida, la glucósido hidrolasa progenitora es una glucósido hidrolasa de *Trichoderma reesei* y de forma más preferible la celobiohidrolasa I de *Trichoderma reesei* de SEC ID n.º: 2 o el polipéptido maduro de la misma. En otra forma de realización más preferida, la glucósido hidrolasa progenitora es codificada por la secuencia de nucleótidos contenida en el plásmido pAJ052 que está contenida en *Escherichia coli* NRRL B-30683, caracterizado por el hecho de que la secuencia de nucleótidos codifica un polipéptido que tiene actividad de glucósido hidrolasa. En otra forma de realización preferida, la glucósido hidrolasa progenitora es codificada por la región de codificación del polipéptido maduro contenida en el plásmido pAJ052 que está contenida in *Escherichia coli* NRRL B-30683.
- [0063] Se entenderá que para las especies mencionadas, la invención incluye los estados perfectos e imperfectos y otros equivalentes taxonómicos, por ejemplo, anamorfos, independientemente del nombre de la especie por el cual son conocidos. Los expertos en la técnica fácilmente reconocerán la identidad de los equivalentes apropiados.
- [0064] Las cepas de estas especies son fácilmente accesibles al público en una cantidad de colecciones de cultivos, tales como el American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS) y Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).
- [0065] La glucósido hidrolasa también puede ser identificada y obtenida a partir de otras fuentes incluidos los microorganismos aislados de la naturaleza (p. ej., tierra, abonos, agua, etc.) o muestras de ADN obtenidas directamente de materiales naturales (p. ej., tierra, abonos, agua, etc.) usando las sondas mencionadas anteriormente. Las técnicas para aislar microorganismos y ADN directamente de hábitats naturales son conocidas en la técnica. La secuencia de nucleótidos que codifica una glucósido hidrolasa luego puede ser derivada seleccionando de manera similar una genoteca de ADN genómico o ADNc de otro microorganismo o muestra de ADN mezclado. Una vez que una secuencia de nucleótidos que codifica una glucósido hidrolasa ha sido detectada con sonda/s adecuada/s como se describe en la presente, la secuencia se puede aislar o clonar utilizando técnicas conocidas por los técnicos en la materia (véase, por ejemplo, J. Sambrook, E.F. Fritsch y T. Maniatus, 1989, *Molecular Cloning, A Laboratory Manual*, 2ª edición, Cold Spring Harbor, New York).
- [0066] Tal y como se define en la presente, una glucósido hidrolasa "aislada" es un polipéptido que está esencialmente libre de otros polipéptidos no de glucósido hidrolasa, por ejemplo, al menos aproximadamente 20% puro, preferiblemente al menos 40% puro, más preferiblemente 60% puro, incluso más preferiblemente 80% puro, de forma más preferible 90% puro, e incluso de forma más preferible aproximadamente 95% puro, según se determina por SDS-PAGE.
- [0067] La glucósido hidrolasa progenitora también puede incluir polipéptidos fusionados o polipéptidos de fusión divisibles en los cuales otro polipéptido se fusiona en el N-término o el C-término del polipéptido o fragmento del mismo. Un polipéptido fusionado se produce por fusión de una secuencia de nucleótidos (o una parte de la misma) que codifica otro polipéptido a una secuencia de nucleótidos (o una parte de la misma) de la presente invención. Las técnicas para producir polipéptidos de fusión se conocen en la técnica, e incluyen ligar las secuencias de codificación que codifican los polipéptidos de modo que estas estén dentro del marco y que la expresión del polipéptido fusionado esté bajo

control del/ de los mismo/s promotor/es y terminador. Las proteínas de fusión también pueden ser construidas usando tecnología de intein donde las fusiones se crean post-traduccionalmente (Cooper *et al.*, 1993, EMBO J. 12: 2575-2583 ; Dawson *et al.*, 1994, Science 266: 776-779).

5 [0068] Los aminoácidos esenciales en la glucósido hidrolasa progenitora pueden identificarse de acuerdo con procedimientos conocidos en la técnica, tales como mutagénesis dirigida o mutagénesis de barrido de alanina (Cunningham and Wells, 1989, Science 244: 1081-1085). En esta última técnica, se introducen mutaciones de alanina
10 únicas en cada residuo en la molécula y las moléculas mutantes resultantes se prueban para analizar su actividad biológica (es decir, actividad de glucósido hidrolasa) para identificar residuos de aminoácidos que son críticos para la actividad de la molécula. Véase también, Hilton *et al.*, 1996, J. Biol. Chem. 271: 4699-4708 . El sitio activo de la enzima u otra interacción biológica también puede ser determinado por análisis físico de estructura, como se determina por técnicas tales como resonancia magnética nuclear, cristalografía, difracción electrónica o marcaje por fotoafinidad, junto con mutación de aminoácidos de sitio putativo de contacto. Véase, por ejemplo, de Vos *et al.*, 1992, Science 255: 306-312 ; Smith *et al.*, 1992, J. Mol. Biol. 224: 899-904 ; Wlodaver *et al.*, 1992, FEBS Lett. 309: 59-64. Las identidades de aminoácidos esenciales también pueden ser inferidas a partir del análisis de identidades con polipéptidos que se relacionan con un polipéptido según la invención.

[0069] Pueden realizarse y evaluarse sustituciones de aminoácidos únicas o múltiples usando métodos conocidos de mutagénesis, recombinación y/o redistribución, seguido de un procedimiento de selección pertinente, como los descritos por Reidhaar-Olson y Sauer, 1988, Science 241: 53-57 ; Bowie and Sauer, 1989, Proc. Natl. Acad. Sci. USA 86: 2152-2156 ; WO 95/17413 ; o WO 95/22625. Otros métodos que pueden ser usados incluyen PCR con tendencia al error, exposición en fago (por ejemplo, Lowman *et al.*, 1991, Biochem. 30:10832-10837; patente de EE. UU. N.º 5 223 409; WO 92/06204) y mutagénesis dirigida a una región (Derbyshire *et al.*, 1986, Gene 46:145; Ner *et al.*, 1988, ADN 7:127).

25 Plásmidos

[0070] El plásmido o los plásmidos usados para preparar variantes de glucósido hidrolasa pueden ser cualquier plásmido o vector que puede someterse a procedimientos de ADN recombinante. El plásmido que comprende una secuencia de nucleótidos que codifica una glucósido hidrolasa puede ser preparado ligando la secuencia de nucleótidos
30 en un plásmido adecuado, o por cualquier otro método adecuado. El plásmido contiene preferiblemente uno o más marcadores seleccionables descritos en la presente que permiten la fácil selección de células transformadas. La elección del plásmido frecuentemente dependerá de la célula huésped en la cual se va a introducir.

[0071] En la presente invención, el plásmido puede ser un plásmido que se replica de manera autónoma, es decir, un plásmido que existe como una entidad extracromosómica, cuya replicación es diferente de la replicación cromosómica.

[0072] El replicador de plásmido puede ser cualquier replicador de plásmido que medie la replicación autónoma que funciona en una célula. El término "replicador de plásmido" es definido en la presente como una secuencia que le permite a un plásmido o vector replicarse *in vivo*. Ejemplos de un replicador de plásmido útil en una célula de levadura son el origen de replicación de 2 micra, ARS1, ARS4, la combinación de ARS1 y CEN3 y la combinación de ARS4 y CEN6. Ejemplos de un replicador de plásmido útiles en una célula fúngica filamentosa son AMA1 y ANS1 (Gems *et al.*, 1991, Gene 98:61-67; Cullen *et al.*, 1987, Nucleic Acids Research 15: 9163-9175; WO 00/24883). El aislamiento del gen AMA1 y la construcción de plásmidos o vectores que comprenden el gen se puede realizar según los métodos descritos en WO 00/24883.

[0073] La linealización del/los plásmido/s puede ser dirigida hacia cualquier sitio en el plásmido. El/los plásmido/s se puede/n linealizar por cualquier método adecuado conocido en la técnica, por ejemplo, digestión con una o más enzimas de restricción. Las extremidades linealizadas del plásmido pueden ser rellenadas con nucleótidos según se describe por Pompon *et al.*, 1989, Gene 83: 15-24. No obstante, se prefiere no llenar en los extremos linealizados puesto que esto puede crear un desplazamiento del marco de lectura.

[0074] Para facilitar el proceso de selección, el plásmido es preferiblemente un vector de expresión en el que la secuencia de nucleótidos en cuestión está operativamente vinculada a segmentos adicionales requeridos para la transcripción del ADN. En general, el vector de expresión se deriva de un plásmido, un cósmido o un bacteriófago, o puede contener elementos de cualquiera de estos o todos estos. Para objetivos de la presente invención, los términos "plásmido" y "vector" se usan de forma intercambiable.

Fragmentos de ADN

60 [0075] La biblioteca de fragmentos de ADN para ser combinada de forma aleatoria (o "redistribuidos") con regiones homólogas en el/los plásmido/s linealizado/s por recombinación *in vivo* se puede preparar por medio de cualquier método adecuado. Por ejemplo, el fragmento de ADN se puede preparar por amplificación por PCR (por ejemplo, PCR con tendencia al error) de un plásmido que comprende la secuencia de nucleótidos, usando cebadores específicos, por ejemplo, como se describe en la patente estadounidense n.º 4 683 202 o Saiki *et al.*, 1988, Science 239: 487-491. El

fragmento de ADN también puede ser aislado de un plásmido que comprende la secuencia de nucleótidos deseada por digestión con enzimas de restricción, seguida de aislamiento usando, por ejemplo, electroforesis.

5 [0076] El fragmento de ADN puede ser preparado de forma alternativa sintéticamente por métodos estándares establecidos, por ejemplo, el método de fosfoamidita descrito por Beaucage y Caruthers, 1981, Tetrahedron Letters 22: 1859-1869, o el método descrito por Matthes *et al.*, 1984, EMBO Journal 3: 801-805. Según el método de fosfoamidita, los oligonucleótidos son sintetizados en un sintetizador de ADN automático, purificados, anillados, ligados y clonados en plásmidos adecuados.

10 [0077] El fragmento de ADN también puede ser de orígenes mezclados sintético y genómico, mezclado sintético y de ADNc o mezclado genómico y de ADNc preparados ligando fragmentos de origen de ADNc, genómico o sintético, los fragmentos correspondientes a varias partes de la secuencia de nucleótidos entera, conforme a técnicas estándares.

15 [0078] La biblioteca de fragmentos de ADN comprende una o más mutaciones de la secuencia de nucleótidos, caracterizados por el hecho de que los fragmentos comprenden al menos dos regiones, una o más regiones que son homólogas a la región 5' o la región 3' del espacio en la secuencia de nucleótidos linealizada y/o la secuencia de plásmidos y una o más segundas regiones que son homólogas a la región 5' o la región 3' de los fragmentos de ADN de la biblioteca.

20 [0079] Las regiones del fragmento de ADN pueden ser cualquier secuencia que sea homóloga a la secuencia de nucleótidos y/o secuencia de plásmidos.

[0080] En una forma de realización preferida, las regiones del fragmento de ADN son una región 5' y/o una región 3' que flanquea un gen que codifica una glucósido hidrolasa, o una región 5' y/o una región 3' de un gen que codifica una glucósido hidrolasa.

25 [0081] En otra forma de realización preferida, el fragmento o los fragmentos de ADN son preparados bajo condiciones que conducen a una frecuencia de mutagénesis aleatoria baja, media o alta. Para obtener la frecuencia de mutagénesis baja la/s secuencia/s de nucleótidos (que comprenden el o los fragmentos de ADN) se pueden preparar por un método de amplificación por PCR estándar (US 4 683 202 o Saiki *et al.*, 1988, Science 239: 487-491). Una frecuencia de mutagénesis media o alta se puede obtener mediante la realización de la amplificación por PCR bajo condiciones que reduzcan la fidelidad de la replicación por la polimerasa termoestable y aumenten la incorporación incorrecta de nucleótidos, por ejemplo como se describe por Deshler, 1992, GATA 9: 103-106; Leung *et al.*, 1989; BioTechniques1: 11-15 .

30 [0082] La amplificación por PCR se puede combinar con una fase de mutagénesis usando un agente mutagenizante físico o químico adecuado, por ejemplo, uno que induzca transiciones, transversiones, inversiones, aleatorización, deleciones y/o inserciones.

35 [0083] En una forma de realización preferida, el/los fragmento/s de ADN que se debe/n redistribuir preferiblemente tiene/n una longitud de aproximadamente 15 bp a 8 kb, más preferiblemente aproximadamente 30 bp a 6 kb, incluso más preferiblemente aproximadamente 40 bp a 6 kb, incluso más preferiblemente aproximadamente 80 bp a 4 kb yde forma más preferible aproximadamente 100 bp a 2 kb, para ser capaces de interactuar óptimamente con el plásmido linealizado.

45 **Células fúngicas**

[0084] La célula fúngica, en la cual la mezcla de secuencias de nucleótidos de plásmido/fragmento deben ser introducidas, puede ser cualquier célula fúngica útil en la presente invención. Una "célula fúngica de recombinación" es definida en la presente como una célula capaz de mediar la recombinación de varias secuencias de nucleótidos homólogas.

50 [0085] En una forma de realización preferida, la célula de recombinación fúngica es una célula de levadura. En una forma de realización más preferida, la célula de recombinación de levadura es una célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*..

55 [0086] En una forma de realización más preferida, la célula de recombinación de levadura es una célula de *Kluyveromyces lactis*, *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, *Saccharomyces oviiformis* o *Yarrowia lipolytica*.

60 [0087] En otra forma de realización preferida, la célula de recombinación fúngica es una célula micótica filamentosa. En una forma de realización más preferida, la recombinación fúngica filamentosa es una célula de *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium* o *Trichoderma*.

65

[0088] En una forma de realización más preferida, la célula de recombinación fúngica filamentosa es una célula de *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*. En otra forma de realización más preferida, la célula de recombinación fúngica filamentosa es una célula de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochromum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides* o *Fusarium venenatum*. En otra forma de realización más preferida, la célula de recombinación fúngica filamentosa es una célula de *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.

[0089] En otra forma de realización más preferida, la célula de *Aspergillus* es una célula de *Aspergillus oryzae*.

[0090] En otra forma de realización más preferida, la célula de *Aspergillus* es una célula de *Aspergillus niger*.

[0091] En otra forma de realización más preferida, la célula de *Fusarium venenatum* es *Fusarium venenatum* A3/5, que fue originalmente depositada como *Fusarium graminearum* ATCC 20334 y recientemente reclasificada como *Fusarium venenatum* por Yoder y Christianson, 1998, Fungal Genetics and Biology 23: 62-80 y O'Donnell *et al.*, 1998, Fungal Genetics and Biology 23: 57-67; al igual que equivalentes taxonómicos de *Fusarium venenatum* independientemente del nombre de especie por el cual son conocidos actualmente. En otra forma de realización más preferida, la célula de *Fusarium venenatum* es un mutante morfológico de *Fusarium venenatum* A3/5 o *Fusarium venenatum* ATCC 20334 como se describe en WO 97/26330.

[0092] Las células fúngicas se pueden transformar por un proceso que implica formación de protoplasto, transformación de los protoplastos y regeneración de la pared celular en un modo conocido *per se*. Los procedimientos adecuados para la transformación de las células huéspedes de *Aspergillus* y *Trichoderma* están descritos en EP 238 023 y Yelton *et al.*, 1984, Proceedings of the National Academy of Sciences USA 81: 1470-1474. Métodos adecuados para transformar especies de *Fusarium* son descritos por Malardier *et al.*, 1989; Gene 78: 147-156, y WO 96/00787. La levadura puede ser transformada usando los procedimientos descritos por Becker y Guarente, en Abelson, J.N. y Simon, M.I., editores, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, págs.182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, Journal de Bacteriology 153: y Hinnen *et al.*, 1978, Proceedings of the National Academy of Sciences USA 75: 1920.

Recombinación *in vivo*

[0093] Un gran número de variantes o genes homólogos se pueden combinar en una transformación para crear eficazmente genes quimeras de los genes homólogos. La redistribución de estos genes, que codifican variantes mejoradas de genes de tipo salvaje, o una combinación de las mismas, da como resultado quimeras que se pueden expresar y seguir por selección para identificar aquellas quimeras con la combinación óptima de mutaciones provechosas. El proceso aumenta de manera múltiple el número de variantes mejoradas adicionales que se puede obtener en comparación con un proceso que solo utiliza mutagénesis aleatoria (para obtener un resumen, véase Kuchner y Arnold, 1997, TIBTech 15: 523-530). La mutagénesis aleatoria introduce mutaciones en una secuencia de nucleótidos objetivo, creando mutaciones deletéreas con mucha más frecuencia que las provechosas. En ciclos iterativos de esa mutagénesis, las mutaciones deletéreas se acumulan más rápidamente que las provechosas, enmascarando eficazmente la identificación de mutaciones provechosas durante la selección. La recombinación aleatoria entre dos o más secuencias de nucleótidos homólogos que contienen múltiples cambios de nucleótidos únicos en sus secuencias de nucleótidos permite potencialmente que todos aquellos cambios de nucleótidos contenidos en una variante se separen uno del otro y sean combinados de forma aleatoria, en cambio, con cualquier mutación presente en otras variantes. Esta redistribución de mutaciones proporciona un medio por el cual las mutaciones de distintas secuencias progenitoras pueden ser combinadas entre sí de forma aleatoria para aumentar la probabilidad de combinación de los cambios de nucleótidos en una secuencia única de nucleótidos.

[0094] Se puede combinar los métodos de mutagénesis/redistribución con métodos de selección de alto rendimiento automatizados para detectar actividad de polipéptidos mutagenizados clonados expresado por células huésped. Las moléculas de ADN mutagenizadas que codifican polipéptidos activos se pueden recuperar de las células huéspedes y ser ordenadas rápidamente usando métodos estándar de la técnica. Estos métodos permiten la determinación rápida de la importancia de los residuos de aminoácidos individuales en un polipéptido de interés y se pueden aplicar a polipéptidos de estructura desconocida.

[0095] La recombinación eficaz de fragmentos de superposición múltiple usando el método de recombinación *in vivo* es un medio para generar quimeras a partir de variantes o genes homólogos. Una superposición tan pequeña como 15 bp es suficiente para la recombinación y se puede utilizar para redistribución de dominios muy sencilla incluso de genes remotamente relacionados. En la redistribución de dominios, los bloques más grandes de ADN no homólogos son clasificados de forma aleatoria mediante extensiones de homología en sus terminales.

[0096] Se prefiere que al menos un ciclo de redistribución sea un ciclo de retrocruce con el fragmento o fragmentos de ADN inicialmente usados, que pueden ser el fragmento de ADN de tipo salvaje. Esto elimina las mutaciones no

esenciales. Las mutaciones no esenciales también pueden ser eliminadas usando fragmentos de ADN de tipo salvaje como el material de ADN de entrada usado inicialmente.

5 [0097] Más de dos secuencias de nucleótidos pueden redistribuirse al mismo tiempo y pueden ser ventajosas puesto que un gran número de variantes bastante diferentes pueden ser hechos rápidamente sin una abundancia de procedimientos iterativos. Al recombinar muchos fragmentos de la misma región, la superposición múltiple de los fragmentos aumentará la frecuencia de intercambio de ADN por sí sola, pero también es importante tener un número relativamente alto de cruces aleatorios en las regiones de superposición para recombinar las variantes/diferencias cercanamente localizadas.

10 [0098] Una superposición tan pequeña como 15 bp entre dos fragmentos es suficiente para obtener una recombinación eficaz. Por lo tanto, en la presente invención, es conveniente la superposición en el rango de 15 a 5000 bp, preferiblemente de 30 bp a 500 bp, especialmente de 30 bp a 100 bp.

15 [0099] En la presente invención, preferiblemente 2 o más fragmentos superpuestos, más preferiblemente 2 a 50 fragmentos superpuestos y de forma más preferible 2 a 10 fragmentos superpuestos, pueden ser usados ventajosamente como fragmentos de ADN en un ciclo de redistribución.

20 [0100] Además de permitir la creación de genes quiméricos, la utilización de fragmentos superpuestos es un método útil para la redistribución de dominios mediante la creación de pequeñas superposiciones entre fragmentos de ADN de diferentes dominios y la selección para lograr la mejor combinación. Por ejemplo, en el caso de tres fragmentos de ADN, las regiones de superposición pueden ser de la siguiente manera: el primer extremo del primer fragmento se superpone al primer extremo del plásmido linealizado, el primer extremo del segundo fragmento se superpone con el segundo extremo del primer fragmento y el segundo extremo del segundo fragmento se superpone con el primer extremo del tercer fragmento, el primer extremo del tercer fragmento se superpone (como se indica más arriba) con el segundo extremo del segundo fragmento y el segundo extremo del tercer fragmento se superpone con el segundo extremo del plásmido linealizado.

30 [0101] Se entiende que cuando se usan dos o más fragmentos de ADN como la materia prima, se prefiere tener superposiciones continuas entre los extremos del plásmido y los fragmentos de ADN.

35 [0102] Aunque se prefiere redistribuir secuencias de nucleótidos homólogas en forma de fragmento/s de ADN y plásmido/s linealizado/s, también es posible redistribuir dos o más plásmidos linealizados que comprendan secuencias de nucleótidos homólogas que codifican polipéptidos. No obstante, en tal caso, es importante linearizar los plásmidos en diferentes sitios.

40 [0103] En la presente invención, dos o plásmidos linealizados y uno o más fragmentos de ADN homólogos pueden ser usados como la materia prima que se debe redistribuir. La proporción entre el/los plásmido/s linealizado/s y el/los fragmento/s de ADN homólogo preferiblemente se encuentran en el rango de 20:1 a 1:50, preferiblemente de 2:1 a 1:10 (moles de plásmido:moles de fragmentos) con las concentraciones específicas que van de 1 pM a 10 M del ADN.

45 [0104] Los plásmidos linealizados se pueden distanciar de manera que la superposición entre los fragmentos se delecione en el plásmido. La reparación del espacio en el plásmido luego requiere que los fragmentos se recombinen uno con el otro además de recombinarse con los extremos del plásmido escindido para reconstruir un plásmido que se replica de manera autónoma y circular. En una forma de realización preferida, la linearización del plásmido o vector crea un espacio suficiente en la secuencia codificante de la secuencia de nucleótidos para forzar la recombinación homóloga de los fragmentos de ADN con las regiones correspondientes de la secuencia de nucleótidos, recreando un plásmido de replicación circular.

50 Variantes

[0105] En la presente invención, las variantes aisladas de una glucósido hidrolasa progenitora comprende una sustitución en una o más posiciones correspondientes a las posiciones 21, 94, 157, 205, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2, caracterizados por el hecho de que las variantes, que tienen actividad de glucósido hidrolasa, comprenden secuencias de aminoácidos que tienen un grado de identidad de al menos 85%, incluso más preferiblemente al menos 90%, de forma más preferible al menos 95%, e incluso de forma más preferible al menos 97% a la secuencia de aminoácidos de la glucósido hidrolasa progenitora, caracterizado por el hecho de que una posición o más posiciones comprenden al menos una sustitución correspondiente a la posición 205 de los aminoácidos 1 a 513 de SEC ID n.º: 2. Para objetivos de la presente invención, el grado de identidad entre dos secuencias de aminoácidos se determina por el método de Clustal (Higgins, 1989, CABIOS 5: 151-153) usando el Sdetware de MEGALIGN™ de LASERGENE™ (DNASTAR, Inc., Madison, WI) con una tabla de identidad y los siguientes parámetros de alineación múltiple: Penalización por espacio de 10 y penalización por longitud de espacio de 10. Los parámetros de alineación en pareja fueron Ktuple=1, penalizaciones por espacio =3, ventanas =5 y diagonales=5.

5 [0106] Tal y como se define en la presente, una "variante aislada" de una glucósido hidrolasa progenitora es un polipéptido que está esencialmente libre de otros polipéptidos de hidrolasa no glucósida, por ejemplo, al menos 20% puro, preferiblemente al menos 40% puro, más preferiblemente al menos 60% puro, incluso más preferiblemente al menos 80% puro, de forma más preferible al menos 90% puro, e incluso de forma más preferible, al menos 95% puro, según se determina por SDS-PAGE.

10 [0107] En una forma de realización preferida, el número de sustituciones de aminoácidos en las variantes de la presente invención comprenden preferiblemente 33, más preferiblemente 32, incluso más preferiblemente 31, incluso más preferiblemente 30, incluso más preferiblemente 29, incluso más preferiblemente 28, incluso más preferiblemente 27, incluso más preferiblemente 26, incluso más preferiblemente 25, incluso más preferiblemente 24, incluso más preferiblemente 23, incluso más preferiblemente 22, incluso más preferiblemente 21, incluso más preferiblemente 20, incluso más preferiblemente 19, incluso más preferiblemente 18, incluso más preferiblemente 17, incluso más preferiblemente 16, incluso más preferiblemente 15, incluso más preferiblemente 14, incluso más preferiblemente 13, incluso más preferiblemente 12, incluso más preferiblemente 11, incluso más preferiblemente 10, incluso más preferiblemente 9, incluso más preferiblemente 8, incluso más preferiblemente 7, incluso más preferiblemente 6, incluso más preferiblemente 5, incluso más preferiblemente 4, incluso más preferiblemente 3, incluso más preferiblemente 2 y de forma más preferible 1.

20 [0108] En una forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además una sustitución en una o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en dos o varias posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en tres o varias posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en cuatro o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en cinco o más o varias posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en seis o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en siete o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en ocho o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en nueve o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en diez o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la variante de una glucósido hidrolasa progenitora comprende además sustituciones en once o más posiciones correspondientes a las posiciones 21, 94, 157,

513 de SEC ID n.º: 2 con Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la variante comprende Arg y Pro como sustituciones en posiciones correspondientes a las posiciones 205 e 255, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende las sustituciones G205R + T255P de los aminoácidos 1 a 513 de SEC ID n.º: 2.

[0150] En otra forma de realización preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 157, 205 y 255 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 157, 205 y 255 de los aminoácidos 1 a 513 de SEC ID n.º: 2 con Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la variante comprende Arg, Arg y Pro como sustituciones en posiciones correspondientes a las posiciones 157, 205 y 255, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende las sustituciones K157R + G205R + T255P de los aminoácidos 1 a 513 de SEC ID n.º: 2.

[0151] En otra forma de realización preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 205 e 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 205 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2 con Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la variante comprende Arg y Phe como sustituciones en posiciones correspondientes a las posiciones 205 e 411, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende las sustituciones G205R + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2.

[0152] En otra forma de realización preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 205, 255 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 205, 255 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2 con Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la variante comprende Arg, Pro y Phe como sustituciones en posiciones correspondientes a las posiciones 205, 255 y 411, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende las sustituciones G205R + T255P + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2.

[0153] En otra forma de realización preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 157, 205 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 157, 205 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2 con Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la variante comprende Arg, Arg y Phe como sustituciones en posiciones correspondientes a las posiciones 157, 205 y 411, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende las sustituciones K157R + G205R + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2.

[0154] En otra forma de realización preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 157, 205, 255 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende sustituciones en posiciones correspondientes a las posiciones 157, 205, 255 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2 con Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la variante comprende Arg, Arg, Pro y Phe como sustituciones en posiciones correspondientes a las posiciones 157, 205, 255 y 411, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la variante comprende las sustituciones K157R, G205R + T255P + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2.

[0155] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos una sustitución seleccionada del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

[0156] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos dos sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

- 5 [0166] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos doce sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 10 [0167] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos trece sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 15 [0168] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos catorce sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 20 [0169] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos quince sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 25 [0170] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos dieciséis sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 30 [0171] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos diecisiete sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 35 [0172] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos dieciocho sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 45 [0173] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos diecinueve sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 50 [0174] En otra forma de realización preferida, la variante comprende la sustitución G205R y al menos veinte sustituciones seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.
- 60 [0175] En otra forma de realización preferida, la variante comprende la sustitución G205R y sustituciones que consisten en al menos S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C,
- 65

D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S, y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente comprende además una o más sustituciones seleccionadas del grupo que consiste en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

5 [0176] Las variantes de la presente invención puede comprender además una o más deleciones y/o inserciones de la secuencia de aminoácidos.

10 [0177] En una forma de realización preferida, una variante de la presente divulgación consiste en 341 a 350,351 a 360,361 a 370,371 a 380,381 a 390,391 a 400,401 a 420,421 a 440,441 a 460,461 a 480,481 a 500, o 501 a 513 aminoácidos.

15 [0178] La variante que comprende la sustitución T226A de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30657. La variante que comprende las sustituciones S113N + S196T + T462A de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30658. La variante que comprende las sustituciones G22D + G467S de los aminoácidos 1 a 513 de SEC ID n.º: 2 se codifica por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30659. La variante comprende las sustituciones S21P + S57N de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30661. La variante comprende las sustituciones K157R + G205R + T255P de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30662. La variante comprende las sustituciones S196P + G205R de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30663. La variante que comprende las sustituciones S113N + S196T + P227A + T462A de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30664. La variante comprende las sustituciones T41I + E193K + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30665. La variante comprende las sustituciones N49S + S113N + P227A + P438L de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30666. La variante comprende las sustituciones N301S + E337V de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30674. La variante comprende las sustituciones S196P + T350S de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30675. La variante comprende las sustituciones G205R + H206Y de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30676. La variante comprende las sustituciones S8P + G205R de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30677. La variante comprende las sustituciones G94S + G205R de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30678. La variante comprende las sustituciones T383A + T455A de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30679. La variante comprende la sustitución N373H de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30680. La variante comprende las sustituciones Y247C + Y371C + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30681. La variante comprende las sustituciones S21P + S57N + T246I + R251K + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30682. La variante comprende las sustituciones P227G + D259N de los aminoácidos 1 a 513 de SEC ID n.º: 2 es codificada por la secuencia de nucleótidos contenida en *E. coli* NRRL B-30762.

45 [0179] La presente divulgación también se refiere a métodos para obtención de una variante de una glucósido hidrolasa progenitora, que comprende: (a) introducción en la glucósido hidrolasa progenitora una sustitución en una o más posiciones correspondientes a las posiciones 21, 94, 157, 205, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente la introducción además de una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2, caracterizada por el hecho de que la variante tiene actividad de glucósido hidrolasa; y (b) recuperación de la variante.

Secuencias de nucleótidos

55 [0180] La presente invención también se refiere a secuencias de nucleótidos aisladas que codifican variantes de una glucósido hidrolasa progenitora, caracterizada por el hecho de que las secuencias de nucleótidos han sido modificadas para codificar las variantes descritas en la presente.

60 [0181] El término "secuencia de nucleótidos aislada" según se utiliza en este caso se refiere a una secuencia de nucleótidos que está esencialmente libre de otras secuencias de nucleótidos, por ejemplo, al menos aproximadamente 20% puro, preferiblemente al menos aproximadamente 40% puro, más preferiblemente al menos aproximadamente 60% puro, incluso más preferiblemente al menos aproximadamente 80% puro, y de forma más preferible, al menos aproximadamente 90% puro, según se determina por electroforesis de agarosa.

Constructos de ácidos nucleicos

[0182] La presente invención también se refiere a constructos de ácidos nucleicos que comprenden una secuencia de nucleótidos que codifica una variante de glucósido hidrolasa de la presente invención operativamente vinculada a una o más secuencias de control que dirigen la expresión de la secuencia codificante en una célula huésped adecuada bajo condiciones compatibles con las secuencias de control. Se entenderá que la expresión incluye cualquier fase implicada en la producción de las variante incluidas, de modo enunciativo y no limitativo, transcripción, modificación postranscripcional, traducción, modificación postraduccional y secreción.

[0183] El "constructo de ácido nucleico" es definido en la presente como una molécula de ácido nucleico, monocatenaria o bicatenaria, que se aísla de un gen de origen natural o que ha sido modificada para contener segmentos de ácido nucleico combinados y superpuestos en un modo que, de lo contrario, no existiría en la naturaleza. El término constructo de ácidos nucleicos es sinónimo del término cassette de expresión cuando el constructo de ácidos nucleicos contiene todas las secuencias de control requeridas para la expresión de una secuencia codificante de una variante de la presente invención. El término "secuencia codificante" es definido en la presente como una secuencia de nucleótidos que especifica directamente la secuencia de aminoácidos de su producto proteínico. Los bordes de una secuencia codificante genómica generalmente son determinados por el codón de iniciación ATG o codones de inicio alternativo tales como GTG y TTG, localizados justo arriba del marco de lectura abierto en el extremo 5' del ARNm y una secuencia del terminador de transcripción localizado justo abajo del marco de lectura abierto en el extremo 3' del ARNm. Una secuencia codificante puede incluir, de modo enunciativo y no limitativo, ADN, ADNc y secuencias de nucleótidos recombinantes.

[0184] Una secuencia de nucleótidos aislada que codifica una variante de glucósido hidrolasa de la presente invención puede ser manipulada en una variedad de maneras para proporcionar la expresión de la variante. La manipulación de la secuencia de nucleótidos antes de su inserción en un vector puede ser conveniente o necesaria dependiendo del vector de expresión. Las técnicas para modificar las secuencias de nucleótidos utilizando métodos de ADN recombinante son bien conocidas en la técnica.

[0185] El término "secuencias de control" es definido en la presente para incluir todos los componentes que son necesarios o ventajosos para la expresión de una variante de glucósido hidrolasa de la presente invención. Cada secuencia de control puede ser nativa o extranjera a la secuencia de nucleótidos que codifica la variante. Tales secuencias de control incluyen, de manera enunciativa y no limitativa, una secuencia líder, de poliadenilación, secuencia de propéptido, promotor, secuencia de péptido señal y de terminador de transcripción. Como mínimo, las secuencias de control incluyen un promotor y señales de parada transcripcional y traduccional. Las secuencias de control pueden ser provistas con enlaces con el fin de introducir sitios de restricción específicos que faciliten la ligadura de las secuencias de control con la región de codificación de la secuencia de nucleótidos que codifica una variante de glucósido hidrolasa de la presente invención. El término "operativamente vinculado" es definido en la presente como una configuración en la cual una secuencia de control es apropiadamente colocada en una posición relativa a la secuencia codificante de la secuencia de nucleótidos de manera que la secuencia de control dirige la expresión de una variante de glucósido hidrolasa.

[0186] La secuencia de control puede ser una secuencia promotora apropiada, que es reconocida por una célula huésped para la expresión de la secuencia de nucleótidos. La secuencia promotora contiene secuencias de control transcripcionales que median la expresión de la variante de glucósido hidrolasa. El promotor puede ser cualquier secuencia de ácidos nucleicos que muestre actividad transcripcional en la célula huésped de elección incluidos promotores mutantes, truncados e híbridos, y pueden obtenerse a partir de genes que codifican polipéptidos extracelulares o intracelulares homólogos o heterólogos a la célula huésped.

[0187] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de ácido nucleico de la presente invención en una célula huésped fúngica filamentosa son promotores obtenidos a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, proteinasa aspártica de *Rhizomucor miehei*, alfa-amilasa neutra de *Aspergillus niger*, alfa-amilasa estable en ácido de *Aspergillus niger*, glucoamilasa (glaA) de *Aspergillus niger* o *Aspergillus awamori*, lipasa de *Rhizomucor miehei*, proteasa alcalina de *Aspergillus oryzae*, triosafosfato isomerasa de *Aspergillus oryzae*, acetamidasa de *Aspergillus nidulans*, amiloglucosidasa de *Fusarium oxysporum*, proteasa de tipo tripsina de *Fusarium oxysporum* (WO 96/00787), beta-glucosidasa de *Trichoderma reesei*, celobiohidrolasa I de *Trichoderma reesei*, endoglucanasa I de *Trichoderma reesei*, endoglucanasa II de *Trichoderma reesei*, endoglucanasa III de *Trichoderma reesei*, endoglucanasa IV de *Trichoderma reesei*, endoglucanasa V de *Trichoderma reesei*, xilanasa I de *Trichoderma reesei*, xilanasa II de *Trichoderma reesei*, beta xilosidasa de *Trichoderma reesei*, al igual que el promotor de NA2-tpi (un híbrido de los promotores de los genes para alfa amilasa neutra de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae*); equivalentes de los mismos; y promotores mutantes, truncados e híbridos de los mismos.

[0188] En un huésped de levadura, se obtienen promotores útiles de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactocinasa de *Saccharomyces cerevisiae* (GAL1), alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH1,ADH2/GAP), triosa fosfato isomerasa de *Saccharomyces cerevisiae* (TPI), metalotionina de *Saccharomyces cerevisiae* (CUP1) y 3-fosfoglicerato quinasa de *Saccharomyces*

cerevisiae. Otros promotores útiles para células huéspedes de levadura son descritas por Romanos *et al.*, 1992, Yeast 8: 423-488.

5 [0189] La secuencia de control también puede ser una secuencia del terminador de transcripción adecuada, que se reconoce por una célula huésped para terminar la transcripción. La secuencia del terminador está operativamente vinculada al término 3' de la secuencia de nucleótidos que codifica la variante de glucósido hidrolasa. En la presente invención, se puede utilizar cualquier terminador que sea funcional en la célula huésped de elección.

10 [0190] Los terminadores preferidos para células huéspedes fúngicas filamentosas se obtienen a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, antranilato sintasa de *Aspergillus nidulans*, alfa-glucosidasa de *Aspergillus niger* y proteasa de tipo tripsina de *Fusarium oxysporum*.

15 [0191] Los terminadores preferidos para células huéspedes de levadura se obtienen a partir de los genes para enolasa de *Saccharomyces cerevisiae*, citocromo C (CYC1) de *Saccharomyces cerevisiae* y gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae*. Otros terminadores útiles para células huéspedes de levadura son descritos por Romanos *et al.*, 1992, *supra*.

20 [0192] La secuencia de control también puede ser una secuencia líder adecuada, una región no traducida de un ARNm que es importante para la traducción por la célula huésped. La secuencia guía está operativamente vinculada al término 5' de la secuencia de nucleótidos que codifica la variante de glucósido hidrolasa. En la presente invención se puede utilizar cualquier secuencia líder que sea funcional en la célula huésped de elección.

[0193] Los líderes preferidos para las células fúngicas filamentosas huéspedes se obtienen de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.

25 [0194] Los líderes adecuados para células huéspedes de levadura se obtienen a partir de los genes para enolasa (ENO-1) de *Saccharomyces cerevisiae*, fosfogliceratocinasa de *Saccharomyces cerevisiae*, factor alfa de *Saccharomyces cerevisiae* y alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa (ADH2/GAP) de *Saccharomyces cerevisiae*.

30 [0195] La secuencia de control puede también ser una secuencia de poliadenilación, una secuencia operativamente vinculada al término 3' de la secuencia de codificación del polipéptido y que, al ser transcrita, es reconocida por la célula huésped como una señal para añadir residuos de poliadenosina a ARNm transcrito. En la presente invención, se puede utilizar cualquier secuencia de poliadenilación que sea funcional en la célula huésped de elección.

35 [0196] Las secuencias de poliadenilación preferidas para células fúngicas filamentosas huéspedes se obtienen a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, antranilato sintasa de *Aspergillus nidulans*, proteasa de tipo tripsina de *Fusarium oxysporum* y alfa-glucosidasa de *Aspergillus niger*.

40 [0197] Las secuencias de poliadenilación útiles para células huéspedes de levadura son descritas por Guo y Sherman, 1995, Molecular Cellular Biology 15: 5983-5990 .

45 [0198] La secuencia de control puede también ser una región de codificación del péptido señal que codifica una secuencia de aminoácidos vinculada al término amino de una glucósido hidrolasa variante y dirige el polipéptido codificado en la vía secretora de la célula. El extremo 5' de la secuencia codificante de la secuencia de nucleótidos puede intrínsecamente contener una región de codificación del péptido señal naturalmente vinculada en el marco de lectura de traducción con el segmento de la región de codificación que codifica la glucósido hidrolasa variante segregada. De forma alternativa, el extremo 5' de la secuencia codificante puede contener una región de codificación del péptido señal que sea extranjera a la secuencia codificante. La región de codificación del péptido señal extranjero puede ser requerida donde la secuencia codificante no contiene naturalmente una región de codificación del péptido señal. De forma alternativa, la región de codificación del péptido señal extranjero puede simplemente reemplazar la región de codificación del péptido señal natural para mejorar la secreción de la glucósido hidrolasa variante. No obstante, en la presente invención, se puede utilizar cualquier región de codificación del péptido señal que dirige el polipéptido expresado en la vía secretora de una célula huésped de elección.

55 [0199] Las regiones de codificación del péptido señal eficaces para células fúngicas filamentosas huéspedes son las regiones de codificación del péptido señal obtenidas a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, amilasa neutra de *Aspergillus niger*, glucoamilasa de *Aspergillus niger*, proteinasa aspártica de *Rhizomucor miehei*, celulasa Cel45A de *Humicola insolens* y lipasa de *Humicola lanuginosa*.

60 [0200] Los péptidos señal útiles para células huéspedes de levadura se obtienen a partir de los genes para factor alfa de *Saccharomyces cerevisiae* e invertasa de *Saccharomyces cerevisiae*. Otras regiones de codificación del péptido señal útiles son descritas por Romanos *et al.*, 1992, *supra*.

65 [0201] La secuencia de control también puede ser una región de codificación del propéptido que codifica una secuencia de aminoácidos situada en el término amino de una variante de glucósido hidrolasa. El polipéptido resultante es conocido como una proenzima o un propolipéptido (o un zimógeno en algunos casos). Un propolipéptido generalmente

es inactivo y puede convertirse en un polipéptido maduro activo por escisión catalítica o autocatalítica del propéptido del propolipéptido. La región de codificación del propéptido se puede obtener a partir de los genes para factor alfa de *Saccharomyces cerevisiae*, proteinasa aspártica de *Rhizomucor miehei* y lacasa de *Myceliophthora thermophila* (WO 95/33836).

5

[0202] En los casos en que las regiones de propéptido y péptido señal están presentes en el término amino de un polipéptido, la región del propéptido está situada junto al término amino de un polipéptido y la región del péptido señal está situada junto al término amino de la región del propéptido.

10

[0203] Puede también ser conveniente añadir secuencias reguladoras que permiten la regulación de la expresión de la glucósido hidrolasa variante relativa al crecimiento de la célula huésped. Ejemplos de sistemas reguladores son los que causan que la expresión del gen se active o no en respuesta a un estímulo químico o físico, incluida la presencia de un compuesto regulador. En la levadura, se puede utilizar el sistema ADH2 o el sistema GAL1. En hongos filamentosos, el promotor de TAKA alfa-amilasa, promotor de glucoamilasa de *Aspergillus niger*, y el promotor de glucoamilasa de *Aspergillus oryzae* pueden ser utilizados como secuencias reguladoras. Otros ejemplos de secuencias reguladoras son los que permiten la amplificación génica. En sistemas eucarióticos, estos incluyen el gen de dihidrodeolato-reductasa que se amplifica en presencia de metotrexato, y los genes de metalotioneína que se amplifican con metales pesados. En estos casos, la secuencia de nucleótidos que codifican la glucósido hidrolasa variante estaría operativamente vinculada con la secuencia reguladora.

15

20

Vectores de expresión

[0204] La presente invención también se refiere a vectores de expresión recombinante que comprenden una secuencia de nucleótidos que codifica una glucósido hidrolasa variante de la presente invención, un promotor y señales de parada transcripcional y traduccional. Las diferentes secuencias de control y de nucleótidos anteriormente descritas se pueden unir para producir un vector de expresión recombinante que puede incluir uno o más sitios de restricción convenientes para permitir la inserción o la sustitución de la secuencia de nucleótidos que codifica la variante en tales sitios. De forma alternativa, la secuencia de nucleótidos se puede expresar por inserción de la secuencia de nucleótidos o un constructo de ácidos nucleicos que comprende la secuencia en un vector apropiado para la expresión. En la creación del vector de expresión, la secuencia codificante se localiza en el vector de modo que la secuencia codificante está operativamente vinculada con las secuencias de control apropiadas para la expresión.

25

30

[0205] El vector de expresión recombinante puede ser cualquier vector (por ejemplo, un plásmido o virus) que pueda ser convenientemente sometido a procedimientos de ADN recombinante y pueda provocar la expresión de la secuencia de nucleótidos. La elección del vector típicamente dependerá de la compatibilidad del vector con la célula huésped en la cual el vector será introducido. Los vectores pueden ser plásmidos circulares lineales o cerrados.

35

[0206] Los vectores de la presente invención contienen preferiblemente uno o más marcadores seleccionables que permiten la fácil selección de células transformadas. Un marcador seleccionable es un gen cuyo producto proporciona resistencia biocida o vírica, resistencia a metales pesados, prototrofia a auxótrofos y similares. Los marcadores adecuados para las células huéspedes de levadura son ADE2, HIS3, LEU2, LYS2, MET3, TRP1 y URA3. Los marcadores seleccionables para el uso en una célula fúngica filamentosa huésped incluyen, de modo enunciativo y no limitativo, *amdS* (acetamidasa), *argB* (ornitina-carbamoiltransferasa), *bar* (fosfotricina acetiltransferasa), *hph* (higromicina fosfotransferasa), *niaD* (nitrato reductasa), *pyrG* (orotidina-5'-fosfato descarboxilasa), *sC* (sulfato adeniltransferasa) y *trpC* (antranilato sintasa), al igual que equivalentes de los mismos. Para el uso en una célula de *Aspergillus* se prefieren los genes *amdS* y *pyrG* de *Aspergillus nidulans* o *Aspergillus oryzae* y el gen *bar* de *Streptomyces hygroscopicus*.

40

45

[0207] El vector puede ser un vector de replicación autónoma, es decir, un vector que existe como una entidad extracromosómica, cuya replicación es diferente de la replicación cromosómica, por ejemplo, un plásmido, un elemento extracromosómico, un minicromosoma o un cromosoma artificial. El vector puede contener cualquier medio para asegurar la autorreplicación. De forma alternativa, el vector puede ser uno que, al ser introducido en la célula huésped, se integra en el genoma y se replica con el/los cromosoma/s en que ha sido integrado. Además, puede utilizarse un único vector o plásmido o dos o más vectores o plásmidos que juntos contengan el ADN total que se introducirá en el genoma de la célula huésped, o un transposón.

50

55

[0208] Los vectores de la presente invención preferiblemente contienen un elemento o elementos que permitan la integración del vector en el genoma de célula huésped o la replicación autónoma del vector en la célula independiente del genoma.

60

[0209] Para la integración en el genoma de la célula huésped, el vector puede depender de la secuencia de nucleótidos que codifica la variante o cualquier otro elemento del vector para la integración del vector en el genoma por recombinación homóloga o no homóloga. De forma alternativa, el vector puede contener secuencias de ácidos nucleicos adicionales para dirigir la integración por recombinación homóloga en el genoma de la célula huésped. Las secuencias de ácido nucleico adicionales permiten que el vector se integre en el genoma de la célula huésped en una ubicación o ubicaciones precisas en el/los cromosoma/s. Para aumentar la probabilidad de integración a una ubicación precisa, los

65

elementos integracionales deberían contener preferiblemente un número suficiente de ácidos nucleicos, tal como 100 a 10 000 pares de bases, preferiblemente 400 a 10 000 pares de bases, y de forma más preferible 800 a 10 000 pares de bases, que son altamente homólogos a la secuencia meta correspondiente para mejorar la probabilidad de recombinación homóloga. Los elementos integracionales pueden ser cualquier secuencia que sea homóloga a la secuencia meta en el genoma de la célula huésped. Además, los elementos integracionales pueden ser no codificantes o codificantes de secuencias de ácidos nucleicos. En cambio, el vector se puede integrar en el genoma de la célula huésped por recombinación no homóloga.

[0210] Para la replicación autónoma, el vector puede comprender además un origen de que le permita al vector replicarse de manera autónoma en la célula huésped en cuestión. Ejemplos de orígenes de replicación para el uso en una célula huésped de levadura son el origen de replicación de 2 micra, ARS1, ARS4, la combinación de ARS1 y CEN3 y la combinación de ARS4 y CEN6. El origen de la replicación puede ser uno que tenga una mutación que haga que el funcionamiento sea termosensible en la célula huésped (véase, por ejemplo, Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75: 1433). Ejemplos de un replicador de plásmido útil en una célula fúngica filamentosa son AMA1 y ANS1 (Gems *et al.*, 1991, Gene 98:61-67; Cullen *et al.*, 1987, Nucleic Acids Research 15: 9163-9175 ; WO 00/24883). El aislamiento del gen AMA1 y la construcción de plásmidos o vectores que comprenden el gen se puede realizar según los métodos descritos en WO 00/24883.

[0211] Se puede insertar más de una copia de una secuencia de nucleótidos de la presente invención en la célula huésped para aumentar la producción de una variante de glucósido hidrolasa. Se puede obtener un aumento en el número de copias de la secuencia de nucleótidos integrando al menos una copia adicional de la secuencia en el genoma de la célula huésped o incluyendo un gen marcador seleccionable amplificable con la secuencia de nucleótidos donde las células que contienen copias amplificadas del gen marcador seleccionable, y por tanto las copias adicionales de la secuencia de nucleótidos, pueden seleccionarse cultivando las células en presencia del agente seleccionable apropiado.

[0212] Los procedimientos usados para enlazar los elementos anteriormente descritos para construir los vectores de expresión recombinante de la presente invención son conocidos por los expertos en la técnica (véase, por ejemplo, Sambrook *et al.*, 1989, *supra*).

Células huésped

[0213] La presente invención glucósido hidrolasa variante, que es ventajosamente usada en la producción recombinante de la variante. Un vector que comprende una secuencia de nucleótidos de la presente invención se introduce en una célula huésped de modo que el vector es mantenido como un integrante cromosómico o como un vector extracromosómico que se auto-replica como se describe anteriormente. El término "célula huésped" comprende cualquier progenie de una célula madre que no es idéntica a la célula madre debido a mutaciones que ocurren durante la replicación. La elección de una célula huésped depende en gran medida del gen que codifica el polipéptido y su fuente.

[0214] La célula huésped puede ser cualquier eucariota, como un mamífero, un insecto, una planta o una célula fúngica.

[0215] La célula huésped puede ser cualquier célula fúngica. "Hongos" según se utiliza en este caso incluye el filo Ascomycota, Basidiomycota, Chytridiomycota y Zygomycota (según la definición de Hawksworth *et al.*, en Ainsworth and Bisby's Dictionary of The Fungi, 8.^a edición, 1995, CAB International, University Press, Cambridge, Reino Unido) al igual que el Oomycota (según se cita en Hawksworth *et al.*, 1995, *supra*, página 171) y todos los hongos mitospóricos (Hawksworth *et al.*, 1995, *supra*).

[0216] En una forma de realización preferida, la célula fúngica huésped es una célula de levadura. "Levadura", según se utiliza en este caso, incluye levadura ascosporigénea (Endomycetales), levadura basidiosporogénea y levadura de los Fungi Imperfecti (Blastomycetes). Ya que la clasificación de levadura puede cambiar en el futuro, para los objetivos de la presente invención, la levadura será definida como se describe en Biology and Activities of Yeast (Skinner, F.A., Passmore, S.M. y Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

[0217] En una forma de realización más preferida, la célula huésped de levadura es una célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*.

[0218] En una forma de realización más preferida, la célula huésped de levadura es una célula *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* o *Saccharomyces oviformis*. En otra forma de realización más preferida, la célula huésped de levadura es una célula de *Kluyveromyces lactis*. En otra forma de realización más preferida, la célula huésped de levadura es una célula de *Yarrowia lipolytica*.

[0219] En otra forma de realización preferida, la célula huésped fúngica es una célula fúngica filamentosa. Los "hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión Eumycota y Oomycota (según se define por Hawksworth *et al.*, 1995, *supra*). Los hongos filamentosos generalmente son caracterizados por una pared micelial

compuesta por quitina, celulosa, glucano, quitosano, manano y otros polisacáridos complejos. El crecimiento vegetativo es por elongación hifal y el catabolismo de carbono es estrictamente aeróbico. En cambio, el crecimiento vegetativo por levaduras tales como *Saccharomyces cerevisiae* es por injerto de un talo unicelular y el catabolismo de carbono puede ser fermentativo.

5

[0220] En una forma de realización más preferida, la célula fúngica filamentosa huésped es, de modo enunciativo y no limitativo, una célula de *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolyposcladium* o *Trichoderma*.

10

[0221] En una forma de realización más preferida, la célula fúngica filamentosa huésped es una célula de *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*. En otra forma de realización más preferida, la célula fúngica filamentosa huésped es una célula de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, o *Fusarium venenatum*. En una forma de realización incluso más preferida, la célula fúngica filamentosa huésped es una célula de *Fusarium venenatum* (Nirenberg sp. nov.). En otra forma de realización más preferida, la célula fúngica filamentosa huésped es una célula de *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*. En otra forma de realización incluso más preferida, la célula fúngica filamentosa huésped es *Trichoderma reesei* RutC30.

15

20

[0222] Las células fúngicas se pueden transformar según los procedimientos descritos en la presente.

Métodos de producción

25

[0223] La presente invención también se refiere a métodos para producir una variante de glucósido hidrolasa, que comprende:

30

1. (a) cultivo de una célula huésped bajo condiciones adecuadas para la expresión de la variante, caracterizado por el hecho de que la célula huésped comprende una secuencia de nucleótidos de la invención que ha sido modificada para codificar la variante que comprende una sustitución en una o más posiciones correspondientes a las posiciones 21, 94, 157, 205, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a la 513 de SEC ID n.º: 2 y opcionalmente comprende además una sustitución en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2, como se describe en la presente; y

35

2. (b) recuperación de la variante del medio de cultivo.

40

[0224] En los métodos de producción de la presente invención, las células huéspedes se cultivan en un medio nutritivo adecuado para la producción de la glucósido hidrolasa variante usando métodos conocidos en la técnica. Por ejemplo, la célula se puede cultivar por cultivo en matraz de agitación o fermentación en pequeña escala o gran escala (incluyendo fermentaciones continuas, de lote, en flujo discontinuo, o estado sólido) en fermentadores de laboratorio o industriales realizados en un medio adecuado y bajo condiciones que permitan que el polipéptido sea expresado y/o aislado. El cultivo se desarrolla en un medio nutritivo adecuado que comprende fuentes de nitrógeno y carbono y sales inorgánicas, usando procedimientos conocidos en la técnica. Los medios adecuados están disponibles de proveedores comerciales o se pueden preparar según composiciones publicadas (por ejemplo, en catálogos de la American Type Culture Collection). Si el polipéptido se segrega en el medio nutritivo, el polipéptido se puede recuperar directamente del medio. Si el polipéptido no es segregado, se puede recuperar de lisatos celulares.

45

50

[0225] En una forma de realización alternativa, la variante de glucósido hidrolasa no es recuperada, sino que se usa como una fuente de la variante una célula huésped de la presente invención que exprese una variante.

55

[0226] La variante de glucósido hidrolasa puede ser detectadas usando métodos conocidos en la técnica que son específicos para los polipéptidos. Estos métodos de detección pueden incluir el uso de anticuerpos específicos, la formación de un producto enzimático o la desaparición de un sustrato enzimático. Por ejemplo, se puede utilizar un ensayo enzimático para determinar la actividad del polipéptido como se describe en este caso en los ejemplos.

60

[0227] La variante de glucósido hidrolasa resultante se puede recuperar por métodos conocidos en la técnica. Por ejemplo, el polipéptido puede ser recuperado del medio nutritivo por procedimientos convencionales incluidos, de modo enunciativo y no limitativo, recogida, centrifugado, filtración, extracción, secado por pulverización, evaporación o precipitación.

65

[0228] Una variante de glucósido hidrolasa de la presente invención se puede purificar por una variedad de procedimientos conocidos en la técnica incluidos, de modo enunciativo y no limitativo, cromatografía (por ejemplo, de intercambio iónico, por afinidad, hidrofóbica, de cromatoenfoco y de exclusión por tamaño), procedimientos electrodéoréticos (por ejemplo, isoelectroenfoco preparatorio), solubilidad diferencial (por ejemplo, precipitación de

sulfato amónico), SDS-PAGE o extracción (véase, por ejemplo, Protein Purification, J.-C. Janson y Lars Ryden, editores, VCH Publishers, New York, 1989) para obtener variantes de glucósido hidrolasas sustancialmente puras.

Otros polipéptidos con actividad de glucósido hidrolasa

5

[0229] La presente invención también se refiere a polipéptidos aislados con actividad de glucósido hidrolasa, caracterizados por el hecho de que la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en al menos la posición 205 y opcionalmente a una o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere a una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2.

10

15

20

25

[0230] En una forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de los aminoácidos 1 a 513 de SEC ID n.º: 2 por preferiblemente 33 aminoácidos, más preferiblemente 32 aminoácidos, incluso más preferiblemente 31 aminoácidos, incluso más preferiblemente 30 aminoácidos, incluso más preferiblemente 29 aminoácidos, incluso más preferiblemente 28 aminoácidos, incluso más preferiblemente 27 aminoácidos, incluso más preferiblemente 26 aminoácidos, incluso más preferiblemente 25 aminoácidos, incluso más preferiblemente 24 aminoácidos, incluso más preferiblemente 23 aminoácidos, incluso más preferiblemente 22 aminoácidos, incluso más preferiblemente 21 aminoácidos, incluso más preferiblemente 20 aminoácidos, incluso más preferiblemente 19 aminoácidos, incluso más preferiblemente 18 aminoácidos, incluso más preferiblemente 17 aminoácidos, incluso más preferiblemente 16 aminoácidos, incluso más preferiblemente 15 aminoácidos, incluso más preferiblemente 14 aminoácidos, incluso más preferiblemente 13 aminoácidos, incluso más preferiblemente 12 aminoácidos, incluso más preferiblemente 11 aminoácidos, incluso más preferiblemente 10 aminoácidos, incluso más preferiblemente 9 aminoácidos, incluso más preferiblemente 8 aminoácidos, incluso más preferiblemente 7 aminoácidos, incluso más preferiblemente 6 aminoácidos, incluso más preferiblemente 5 aminoácidos, incluso más preferiblemente 4 aminoácidos, incluso más preferiblemente 3 aminoácidos, incluso más preferiblemente 2 aminoácidos y de forma más preferible 1 aminoácido.

30

35

40

45

50

55

60

65

[0231] En una forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 al menos en la posición 205 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en una o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en dos o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en tres o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en cuatro o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en cinco o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en seis o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en siete o más posiciones correspondientes a las posiciones 21, 94, 157, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difiere de SEC ID n.º: 2 en una o más posiciones correspondientes a las posiciones 8, 22, 41, 49, 57, 113, 193, 196, 226, 227, 246, 251, 255, 259, 301, 356, 371, 411 y 462 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en la posición 205 y en ocho o más posiciones correspondientes a las posiciones 21,

otra forma de realización más preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por G205R + T255P + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2.

5 [0263] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en posiciones correspondientes a las posiciones 157, 205 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En una forma de realización más preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en posiciones correspondientes a las posiciones 157, 205 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2 por Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por Arg, Arg y Phe en posiciones correspondientes a las posiciones 157, 205 y 411, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por K157R + G205R + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2.

15 [0264] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en posiciones correspondientes a las posiciones 157, 205, 255 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2. En una forma de realización más preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 en posiciones correspondientes a las posiciones 157, 205, 255 y 411 de los aminoácidos 1 a 513 de SEC ID n.º: 2 por Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr o Val. En una forma de realización aún más preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por Arg, Arg, Pro y Phe en posiciones correspondientes a las posiciones 157, 205, 255 y 411, respectivamente, de los aminoácidos 1 a 513 de SEC ID n.º: 2. En otra forma de realización más preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por K157R, G205R + T255P + S411F de los aminoácidos 1 a 513 de SEC ID n.º: 2.

25 [0265] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos una diferencia seleccionada del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

30 [0266] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos dos diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

35 [0267] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos tres diferencias seleccionadas del grupo que consisten en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

45 [0268] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos cuatro diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A. En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos cinco diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

55 [0269] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos seis diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

60 [0270] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos siete diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S,

5 [0280] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos diecisiete diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

10 [0281] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos dieciocho diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

15 [0282] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos diecinueve diferencias seleccionada del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

20 [0283] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por al menos veinte diferencias seleccionadas del grupo que consiste en S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consisten en S8P, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

25 [0284] En otra forma de realización preferida, la secuencia de aminoácidos del polipéptido difiere de SEC ID n.º: 2 por la diferencia G205R y por las diferencias que consisten en al menos S21P, G94S, G94A, G94R, G94Q, K157R, E193K, S196P, S196F, G205R, H206Y, P227L, P227G, T246I, Y247C, D259N, R251K, N301S, E337V, T350S, N373H, T383A, P438L, T455A, G467S y C486W de los aminoácidos 1 a 513 de SEC ID n.º: 2 y opcionalmente además difieren por una o más diferencias seleccionadas del grupo que consiste en SBP, G22D, T41I, N49S, S57N, S113N, S196T, T226A, P227A, T255P, T356I, Y371C, S411F y T462A.

30 [0285] Los polipéptidos aislados tienen una o más propiedades mejoradas en comparación con el polipéptido de SEC ID n.º: 2, caracterizado por el hecho de que las propiedades mejoradas son seleccionadas del grupo que consiste en actividad térmica, termoestabilidad, actividad de pH, estabilidad de pH, especificidad de sustrato, especificidad de producto y estabilidad química, como se describe en la presente.

35 [0286] La presente invención también se refiere a secuencias de nucleótidos aisladas que codifican tales polipéptidos, constructos de ácidos nucleicos, vectores de expresión y células huésped que comprenden las secuencias de nucleótidos y métodos de producción de los polipéptidos con actividad de glucósido hidrolasa, según la misma descripción en la presente para las variantes de glucósido hidrolasa.

40 **Degradación de biomasa a monosacáridos, disacáridos y polisacáridos**

45 [0287] Las variantes de glucósido hidrolasa, los polipéptidos con actividad de glucósido hidrolasa y las células huésped de la presente invención se pueden utilizar en la producción de monosacáridos, disacáridos y polisacáridos como materias primas químicas o de fermentación de biomasa para la producción de etanol, plástico u otros productos o productos intermedios. Las variantes de glucósido y los polipéptidos con actividad de glucósido hidrolasa pueden ser en forma de un caldo de fermentación crudo con o sin las células eliminadas o en forma de una preparación enzimática semi-purificada o purificada. De forma alternativa, una célula huésped de la presente invención puede ser utilizada como una fuente de la variante o el polipéptido con actividad de glucósido hidrolasa en un proceso de fermentación con la biomasa.

50 [0288] La biomasa puede incluir, de modo enunciativo y no limitativo, recursos de madera, desperdicios municipales sólidos, papel usado y residuos de cosecha (véase, por ejemplo, Wiseloge *et al.*, 1995, en Handbook on Bioethanol (Charles E. Wyman, editor), páginas 105-118, Taylor & Francis, Washington D.C.; Wyman, 1994, Bioresource Technology 50: 3-16; Lynd, 1990, Applied Biochemistry and Biotechnology 24/25: 695-719; Mosier *et al.*, 1999, Recent Progress In Bioconversion de Lignocellulosics, in Advances in Biochemical Engineering/Blotechnology, T. Scheper, Director Edtorial, Volumen 65, págs. 23-40, Springer-Verlag, New York).

55 [0289] El polisacárido predominante en la pared celular primaria de la biomasa es la celulosa, el segundo más abundante es la hemicelulosa, y el tercero es la pectina. La pared celular secundaria, producida después de que la célula ha dejado de crecer, también contiene polisacáridos y se refuerza a través de lignina polimérica reticulada de manera covalente a la hemicelulosa. La celulosa es un homopolímero de anhidrocelobiosa y, de ese modo, es un beta-

(1-4)-D-glucano lineal, mientras que las hemicelulosas incluyen una variedad de compuestos, tales como xilanos, xiloglucanos, arabinoxilanos y mananos en estructuras ramificadas complejas con un espectro de sustituyentes. La celulosa se encuentra, aunque generalmente polimorfa, en el tejido vegetal principalmente como una matriz cristalina insoluble de cadenas paralelas de glucano. Las hemicelulosas normalmente tienen enlace de hidrógeno a celulosa, al igual que a otras hemicelulosas, lo cual ayuda a estabilizar la matriz de la pared celular.

[0290] Se utilizan tres clases importantes de glicohidrolasas para descomponer la biomasa celulósica:

1. (1) Las "endo-1,4-beta-glucanasas" o 1,4-beta-D-glucano-4-glucanohidrolasas (EC 3.2.1.4), que actúan de forma aleatoria en sustratos de 1,4-beta-glucano solubles e insolubles.

2. (2) Las "exo-1,4-beta-D-glucanasas" que incluyen las 1,4-beta-D-glucano glucohidrolasas (EC 3.2.1.74), que liberan D-glucosa de 1,4-beta-D-glucanos e hidrolizan D-celobiosa lentamente, y celobiohidrolasas (1,4-beta-D-glucano celobiohidrolasas, EC 3.2.1.91), que liberan D-celobiosa de 1,4-beta-glucanos.

3. (3) Las "beta-D-glucosidasas" o beta-D-glucósido glucohidrolasas (EC 3.2.1.21), que actúan para liberar unidades de D-glucosa de celobiosa y celodextrinas solubles, al igual que un conjunto de glucósidos.

[0291] Estas tres clases de enzimas trabajan juntas sinérgicamente dando como resultado descomposición e hidrólisis eficaces de celulosa nativa de biomasa para producir azúcares reducidos.

[0292] Las variantes de glucósido hidrolasa y los polipéptidos con actividad de glucósido hidrolasa de la presente invención se pueden utilizar conjuntamente con las enzimas mencionadas anteriormente para degradar aún más el componente de celulosa del sustrato de biomasa, (véase, por ejemplo, Brigham *et al.*, 1995, en Handbook on Bioethanol (Charles E. Wyman, editor), págs. 119-141, Taylor & Francis, Washington D.C.; Lee, 1997, Journal de Biotechnology 56: 1-24).

[0293] El etanol se puede producir por degradación enzimática de biomasa y conversión de los sacáridos liberados a etanol. Esta especie de etanol frecuentemente es denominada bioetanol o biocombustible. Esto puede ser usado como un aditivo o suplemento de combustible en mezclas de menos de 1% y hasta 100% (un sustituto de combustible).

Composiciones detergentes

[0294] Las variantes y los polipéptidos con actividad de glucósido hidrolasa de la presente invención pueden ser añadidas a la composición detergente y así convertirse en un componente de la misma.

[0295] La composición detergente de la presente invención puede, por ejemplo, ser formulada como una composición detergente de lavado a mano o a máquina incluida una composición de aditivo de lavado adecuada para el tratamiento previo de tejidos manchados y una composición de suavizante agregado al enjuague, o ser formulada como una composición detergente para el uso en operaciones de limpieza de superficies duras del hogar en general, o ser formulada para operaciones de lavado de vajilla a mano o máquina.

[0296] En un aspecto específico, la presente invención proporciona un aditivo detergente que comprende las variantes y los polipéptidos con actividad de glucósido hidrolasa de la presente invención. El aditivo detergente al igual que la composición detergente puede comprender una o más enzimas adicionales tales como una proteasa, lipasa, cutinasa, una amilasa, carbohidrasa, celulasa, pectinasa, mananasa, arabinasa, galactanasa, xilanasas, oxidasas, por ejemplo, una lacasa y/o peroxidasa.

[0297] En general, las propiedades de los componentes enzimáticos deberían ser compatibles con el detergente seleccionado, (es decir, pH óptimo, compatibilidad con otros ingredientes enzimáticos y no enzimáticos, etc.) y los componentes enzimáticos deberían estar presentes en cantidades eficaces.

[0298] Proteasas: Las proteasas adecuadas incluyen las de origen animal, vegetal o microbiano. Se prefiere el origen microbiano. Están incluidos mutantes creados genéticamente o modificados químicamente de proteínas. La proteasa puede ser una serina proteasa o una metaloproteasa, preferiblemente una proteasa alcalina microbiana o una proteasa de tipo tripsina. Ejemplos de proteasas alcalinas son subtilisinas, especialmente aquellas derivadas de *Bacillus*, por ejemplo, subtilisina Novo, subtilisina Carisberg, subtilisina 309, subtilisina 147 y subtilisina 168 (descritas en WO 89/06279). Ejemplos de proteasas de tipo tripsina son tripsina (por ejemplo, de origen porcino o bovino) y la proteasa de *Fusarium* descrita en WO 89/06270 y WO 94/25583.

[0299] Ejemplos de proteasas útiles son las variantes descritas en WO 92/19729, WO 98/20115, WO 98/20116 y WO 98/34946, especialmente las variantes con sustituciones en una o más de las siguientes posiciones: 27, 36, 57, 76, 87, 97, 101, 104, 120, 123, 167, 170, 194, 206, 218, 222, 224, 235 y 274.

[0300] Las enzimas proteásicas preferidas disponibles comercialmente incluyen AlcalaseTM, SavinaseTM, PrimaseTM, DuralaseTM, EsperaseTM y KannaseTM (Novozymes A/S), MaxataseTM, MaxacalTM, MaxapemTM, ProperaseTM, PurafectTM, Purafect OxFTM, FN2TM y FN3TM (Genencor International Inc.).

- [0301] Lipasas: Las lipasas adecuadas incluyen las de origen bacteriano o fúngico. Están incluidos mutantes creados genéticamente o modificados químicamente de proteínas. Ejemplos de lipasas útiles incluyen lipasas de *Humicola* (sinónimo, *Thermomyces*), por ejemplo, de *H. lanuginosa* (*T. lanuginosus*) según se describe en EP 258 068 y EP 305 216 o de *H. insolens* según se describe en WO 96/13580, una lipasa de *Pseudomonas*, por ejemplo, de *P. alcaligenes* o *P. pseudoalcaligenes* (EP 218 272), *P. cepacia* (EP 331376); *P. stutzeri* (GB 1,372,034), *P. fluorescens*, *Pseudomonas sp*, cepa SD 705 (WO 95/06720 y WO 96/27002), *P. wisconsinensis* (WO 96/12012), una lipasa de *Bacillus*, por ejemplo, de *B. subtilis* (Dartois *et al.*, 1993, *Biochemica et Biophysica Acta*, 1131:: 253-360), *B. stearothermophilus* (JP 64/744992) o *B. pumilus* (WO 91/16422).
- [0302] Otros ejemplos son variantes de lipasa tales como las descritas en WO 92/05249, WO 94/01541, EP 407 225, EP 260 105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 y WO 97/07202.
- [0303] Las enzimas de lipasa disponibles comercialmente preferidas incluyen Lipolase™ y Lipolase Ultra™ (Novozymes A/S).
- [0304] Amilasas: Las amilasas adecuadas (α y/o β) incluyen las de origen bacteriano o fúngico. Están incluidos mutantes creados genéticamente o modificados químicamente de proteínas. Las amilasas incluyen, por ejemplo, α -amilasas obtenidas de *Bacillus*, por ejemplo, una cepa especial de *Bacillus licheniformis*, descrita en más detalle en GB 1.296.839.
- [0305] Ejemplos de amilasas útiles son las variantes descritas en WO 94/02597, WO 94/18314, WO 96/23873 y WO 97/43424, especialmente las variantes con sustituciones en una o más de las siguientes posiciones: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408 y 444.
- [0306] Las amilasas disponibles comercialmente son Duramyl™, Termamyl™, Fungamyl™ y BAN™ (Novozymes A/S), Rapidase™ y Purastar™ (de Genencor internacional Inc.).
- [0307] Celulasas: Las celulasas adecuadas incluyen las de origen bacteriano o fúngico. Están incluidos mutantes creados genéticamente o modificados químicamente de proteínas. Las celulasas adecuadas incluyen celulasas de los géneros *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium*, por ejemplo, las celulasas fúngicas producidas a partir de *Humicola Insolens*, *Myceliophthora thermophila* y *Fusarium oxysporum* descritas en US 4.435.307, US 5.648.263, US 5.691.178, US 5.776.757 y WO 89/09259.
- [0308] Las celulasas especialmente adecuadas son las celulasas alcalinas o neutras que tienen beneficios de cuidados del color. Ejemplos de tales celulasas son celulasas descritas en EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son variantes de celulasa tales como las descritas en WO 94/07998, EP 0 531315, US 5.457.046, US 5.686.593, 5.763.254, WO 95/24471, WO 98/12307 y PCT/DK98/00299.
- [0309] Las celulasas disponibles comercialmente incluyen Celluzyme™ y Carezyme™ (Novo Nordisk A/S), Clazinase™ y Puradax HA™ (Genencor International Inc.) y KAC-500(B)™ (Kao Corporation).
- [0310] Peroxidasas/oxidasas: Las peroxidasas/oxidasas adecuadas incluyen las de origen vegetal, bacteriano o fúngico. Están incluidos mutantes creados genéticamente o modificados químicamente de proteínas. Ejemplos de peroxidasas útiles incluyen peroxidasas de *Coprinus*, por ejemplo, de *C. cinereus* y variantes de las mismas como las descritas en WO 93/24618, WO 95/10602 y WO 98/15257.
- [0311] Las peroxidasas disponibles comercialmente incluyen Guardzyme™ (Novozymes A/S).
- [0312] El/Los componente/s enzimático/s se puede/n incluir en una composición detergente añadiendo aditivos separados que contengan una o más enzimas, o añadiendo un aditivo combinado que comprenda todas estas enzimas. Un aditivo detergente de la invención, es decir, un aditivo separado o un aditivo combinado, puede ser formulado, por ejemplo, como un granulado, un líquido, un compuesto acuoso, etc. Las formulaciones de aditivo detergente preferidas son granulados, en particular granulados no pulverizados, líquidos, en particular líquidos estabilizados, o lodos.
- [0313] Los granulados no pulverizados pueden ser producidos, por ejemplo, como se describe en US 4.106.991 y 4.661.452 y pueden ser revestidos opcionalmente por métodos conocidos en la técnica. Ejemplos de materiales de revestimiento ceroso son productos de poli(etileno óxido) (polietilenglicol; PEG) con pesos molares medios de 1000 a 20.000; nonilfenoles etoxilados que tienen de 16 a 50 unidades de óxido de etileno; alcoholes etoxilados grasos en los que el alcohol contiene de 12 a 20 átomos de carbono y en el cual hay 15 a 80 unidades de óxido de etileno; alcoholes grasos; ácidos grasos; y mono- y di- y triglicéridos de ácidos grasos. Ejemplos de materiales de revestimiento que forman películas adecuadas para la aplicación por técnicas de lecho fluidificado se ofrecen en GB 1483591. Las preparaciones enzimáticas líquidas pueden, por ejemplo, ser estabilizadas añadiendo un poliol tal como propilenglicol, un azúcar o alcohol de azúcar, ácido láctico o ácido bórico según los métodos establecidos. Las enzimas protegidas se pueden preparar según el método descrito en EP 238.216.

[0314] La composición detergente de la invención puede ser en cualquier forma conveniente, por ejemplo, una barra, una pastilla, un polvo, un gránulo, una pasta o un líquido. Un detergente líquido puede ser acuoso, típicamente con 70% de agua y 0-30% de solvente orgánico, o no acuoso.

5 [0315] La composición detergente comprende uno o más agentes tensioactivos, que pueden ser no iónicos incluidos semipolares y/o aniónicos y/o catiónicos y/o zwitteriónicos. Los agentes tensioactivos están típicamente presentes a un nivel de 0,1% a 60% en peso.

10 [0316] Cuando se incluye en el mismo, el detergente normalmente contiene de aproximadamente 1% a aproximadamente 40% de un tensioactivo aniónico tal como alquilbencenosulfonato lineal, alfa-olefinsulfonato, sulfato de alquilo (sulfato de alcohol graso), etoxisulfato alcohólico, alcanosulfonato secundario, éster metílico de ácido alfa-sulfo graso, ácido alquil o alquenilsuccínico, o jabón.

15 [0317] Cuando se incluye en el mismo, el detergente normalmente contiene de aproximadamente 0,2% a aproximadamente 40% de un tensioactivo no iónico como alcohol etoxilato, nonilfenol etoxilato, alquilpoliglucósido, óxido de alquildimetilamina, monoetanolamida de ácido graso etoxilado, monoetanolamida de ácido graso, polihidroxi alquil amida de ácido graso, o derivados N-acilo N-alquilo de glucosamina ("glugamidas").

20 [0318] El detergente puede contener 0-65% de un constructor de detergente o agente complejante tal como zeolita, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilendiaminatetraacético, ácido dietilendiaminopentaacético, ácido alquil o alquenilsuccínico, silicatos solubles o silicatos estratificados (por ejemplo, SKS-6 de Hoechst).

25 [0319] El detergente puede comprender uno o más polímeros. Ejemplos son carboximetilcelulosa, poli(vinilpirrolidona), poli (etilenglicol), alcohol (polivinílico), poli(vinilpiridina-N-óxido), poli(vinilimidazol), policarboxilatos tales como poliacrilatos, copolímeros de ácido maleico/acrílico y copolímeros de lauril metacrilato/ácido acrílico.

30 [0320] El detergente puede contener un sistema blanqueante que puede comprender una fuente de H₂O₂ tal como perborato o percarbonato que se puede combinar con un activador blanqueante de formación de perácido tal como tetraacetiltilendiamina o nonanoiloxibencenosulfonato. De forma alternativa, el sistema blanqueante puede comprender peroxiácidos de, por ejemplo, tipo amida, imida, o sulfona.

35 [0321] El/Los componente/s enzimático/s de la composición detergente de la invención puede/n ser estabilizado/s usando agentes estabilizantes convencionales, por ejemplo, un poliol tal como propilenglicol o glicerol, un azúcar o alcohol de azúcar, ácido láctico, ácido bórico o un derivado de ácido bórico, por ejemplo, un éster de borato aromático, o un derivado de ácido fenilo borónico tal como ácido 4-formilfenil borónico y la composición puede ser formulada como se describe en, por ejemplo, WO 92/19709 y WO 92/19708.

40 [0322] El detergente también puede contener otros ingredientes de detergentes convencionales tales como, por ejemplo, acondicionadores de tejido incluidos arcillas, reforzadores de espuma, supresores de espuma, agentes anticorrosivos, agentes de suspensión de suciedad, agentes de anti redeposición de suciedad, tintes, bactericidas, blanqueadores ópticos, hidrótrópos, inhibidores de decoloración o perfumes.

45 [0323] En las composiciones detergentes cualquier componente enzimático, en particular, las variantes y los polipéptidos con actividad de glucósido hidrolasa de de la invención, se pueden adicionar en una cantidad correspondiente a 0,01-100 mg de proteína enzimática por litro de solución de lavado, preferiblemente 0,05-5 mg de proteína enzimática por litro de solución de lavado, en particular 0,1-1 mg de proteína enzimática por litro de solución de lavado.

50 [0324] Las variantes y los polipéptidos con actividad de glucósido hidrolasa de la invención pueden adicionalmente ser incorporados en las formulaciones detergentes descritas en WO 97/07202, que está incorporada en la presente como referencia.

Plantas

55 [0325] La presente invención también se refiere a una planta transgénica, parte de planta, o célula vegetal que ha sido transformada con una secuencia de nucleótidos que codifica una variante o polipéptidos con actividad de glucósido hidrolasa de la presente invención para expresar y producir la variante o el polipéptido en cantidades recuperables. La variante se puede recuperar de la planta o parte de planta. De forma alternativa, la planta o parte de planta que contiene la variante o el polipéptido recombinante puede ser utilizada como tal para mejorar la calidad de un alimento o pienso, por ejemplo, mejorando el valor nutritivo, de apetencia, y propiedades reológicas, o para destruir un factor antinutritivo.

60 [0326] La planta transgénica puede ser dicotiledónea o monocotiledónea. Ejemplos de plantas monocotiledóneas son hierbas, tales como poa de prados (poa pratense, Poa), hierba forrajera tal como Festuca, Lolium, césped templado, tal como Agrostis y cereales, por ejemplo, trigo, avena, centeno, cebada, arroz, sorgo, y maíz.

65

[0327] Ejemplos de plantas dicotiledóneas son tabaco, leguminosas, tales como altramuces, patata, remolacha azucarera, guisante, moldura y semilla de soja, y plantas crucíferas (familia de Brassicaceae), tales como coliflor, semilla de colza y el organismo modelo cercanamente relacionado *Arabidopsis thaliana*.

5 [0328] Ejemplos de partes de planta son tallo, callo, hojas, raíz, frutas, semillas y tubérculos, al igual que los tejidos individuales que comprenden estas partes, por ejemplo, epidermis, mesófilo, parénquima, tejidos vasculares, meristemas. Compartimentos específicos de la célula vegetal, como cloroplastos, apoplastos, mitocondria, vacuolas, peroxisomas y citoplasma también son considerados parte de una planta. Además, cualquier célula vegetal, cualquiera sea el origen del tejido, se considera una parte de una planta. De igual modo, las partes de planta tales como tejidos
10 específicos y células aisladas para facilitar la utilización de la invención también son consideradas partes de la planta, por ejemplo, embriones, endospermas, aleurona y revestimientos de semillas.

[0329] También está incluida dentro del campo de la presente invención la progenie de tales plantas, partes de planta y
15 células vegetales.

[0330] La planta transgénica o célula vegetal que exprese una variante o un polipéptido de la presente invención se puede construir conforme a métodos conocidos en la técnica. En breve, la planta o la célula vegetal se construye incorporando uno o más constructos de expresión que codifican una variante de la presente invención en el genoma huésped de la planta y propaga la planta modificada o célula vegetal resultante en una planta transgénica o célula
20 vegetal.

[0331] Convenientemente, el constructo de expresión es un constructo de ácidos nucleicos que comprende una secuencia de ácidos nucleicos que codifica una variante o un polipéptido de la presente invención operativamente vinculada con secuencias reguladoras apropiadas requeridas para la expresión de la secuencia de ácidos nucleicos en la planta o parte de planta de elección. Además, el constructo de expresión puede comprender un marcador seleccionable útil para identificar células huéspedes en las cuales el constructo de expresión ha sido integrado y secuencias de ADN necesarias para la introducción del constructo en la planta en cuestión (esto último depende del método de introducción de ADN que se utilizará).
25

[0332] La elección de secuencias reguladoras, tales como secuencias de terminador y promotor y opcionalmente secuencias de tránsito o señal, está determinada, por ejemplo, basándose en cuándo, dónde y cómo se desea expresar la variante o el polipéptido. Por ejemplo, la expresión del gen que codifica una variante de la presente invención puede ser constitutiva o inducible, o puede ser desarrollable, específica de fase o tejido, y el producto genético puede ser dirigido a un tejido o parte de planta específica tal como semillas u hojas. Las secuencias reguladoras son, por ejemplo,
30 descritas por Tague *et al.*, 1988, *Plant Physiology* 86: 506 .

[0333] Para la expresión constitutiva pueden utilizarse el 355-CaMV, la ubiquitina 1 de maíz y el promotor de la actina 1 del arroz (Franck *et al.*, 1980, *Cell* 21: 285-294, Christensen *et al.*, 1992, *Plant Mo. Biol.* 18: 675-689 ; Zhang *et al.*, 1991, *Plant Cell* 3: 1155-1165). Los promotores específicos de un órgano pueden ser, por ejemplo, un promotor de tejidos sumideros de almacenamiento tales como semillas, tubérculos de patata y frutas (Edwards y Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303) o de tejidos sumideros metabólicos tales como meristemas (Ito *et al.*, 1994, *Plant Mol. Biol.* 24: 863-878), un promotor específico de semilla tal como el promotor de glutelina, prolamina, globulina o albúmina de arroz (Wu *et al.*, 1998, *Plant and Cell Physiology* 39: 885-889), un promotor de *Vicia faba* de la legúmina B4 y el gen de proteína de semilla desconocido de *Vicia faba* (Conrad *et al.*, 1998, *Journal de Plant Physiology* 152: 708-711), un promotor de una proteína de cuerpo de aceite de semilla (Chen *et al.*, 1998, *Plant and Cell Physiology* 39: 935 941), el promotor napA de proteína de almacenamiento de *Brassica napus*, o cualquier otro promotor específico de semilla conocido en la técnica, por ejemplo, como se describe en WO 91/14772. Además, el promotor puede ser un promotor específico de hoja tal como el promotor de eritrocitos de arroz o tomate (Kyojuka *et al.*, 1993, *Plant Physiology* 102: 991-1000, el promotor del gen de metiltransferasa de adenina de virus chlorella (Mitra y Higgins, 1994, *Plant Molecular Biology* 26: 85-93), o el promotor de gen *aldP* de arroz (Kagaya *et al.*, 1995, *Molecular and General Genetics* 248: 668-674), o un promotor inducible por lesiones tal como el promotor pin2 de la patata (Xu *et al.*, 1993, *Plant Molecular Biology* 22: 573-588). Asimismo, el promotor puede ser inducible por tratamientos abióticos tales como temperatura, sequía o alteraciones en la salinidad o inducido por sustancias aplicadas exógenamente que activan el promotor, por ejemplo, etanol, estrógenos, hormonas de planta tales como etileno, ácido abscísico y ácido giberélico y metales pesados.
40
45
50
55

[0334] También puede utilizarse un elemento intensificador del promotor para conseguir mayor expresión de un polipéptido de la presente invención en la planta. Por ejemplo, el elemento intensificador del promotor puede ser un intrón que es colocado entre el promotor y la secuencia de nucleótidos que codifica un polipéptido de la presente invención. Xu *et al.*, 1993, *supra*, describen el uso del primer intrón del gen de actina 1 de arroz para mejorar la expresión.
60

[0335] El gen marcador seleccionable y cualquier otra parte del constructo de expresión se pueden elegir de aquellos disponibles en la técnica.
65

[0336] El constructo de ácido nucleico se incorpora en el genoma de la planta según técnicas convencionales conocidas en la técnica, incluida la transformación mediada por *Agrobacterium*, transformación mediada por virus, microinyección, bombardeo de partículas, transformación biolística y electroporación (Gasser *et al.*, 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535 ; Shimamoto *et al.*, 1989, *Nature* 338: 274).

5 [0337] Actualmente, la transferencia de genes mediada por *Agrobacterium tumefaciens* es el método de elección para generar dicotiledóneas transgénicas (para obtener un resumen, véase Hooykas y Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38) y también puede ser usada para transformar monocotiledóneas, aunque frecuentemente se utilizan otros métodos de transformación para estas plantas. Actualmente, el método de elección para generar monocotiledóneas transgénicas es el bombardeo de partículas (partículas de tungsteno u oro microscópico revestidas con el ADN transformante) de callos embrionarios o embriones en desarrollo (Christou, 1992, *Plant Journal* 2: 275-281 ; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162 ; Vasil *et al.*, 1992, *Bio/Technology* 10: 667-674). Un método alternativo para la transformación de monocotiledóneas se basa en la transformación de protoplasto como se describe por Omirulleh *et al.*,
10 1993, *Plant Molecular Biology* 21: 415-428.

15 [0338] Después de la transformación, los transformantes que hayan incorporado el constructo de expresión se seleccionan y regeneran en plantas enteras según métodos bien conocidos en la técnica. Frecuentemente, el procedimiento de transformación se diseña para la eliminación selectiva de genes de selección durante la regeneración o en las siguientes generaciones usando, por ejemplo, cotransformación con dos constructos de T-ADN separados o escisión específica de sitio del gen de selección por una recombinasa específica.

20 [0339] La presente invención también se refiere a métodos para producir una variante o un polipéptido de la presente invención que comprende (a) cultivo de una planta transgénica o una célula vegetal que comprende una secuencia de ácidos nucleicos que codifica una variante o un polipéptido que tiene actividad de glucósido hidrolasa de la presente invención bajo condiciones propicias para la producción de la variante; y (b) recuperación de la variante o el polipéptido.

Otros usos

30 [0340] Las variantes de glucósido hidrolasa o los polipéptidos de la presente invención también pueden ser usadas en el tratamiento de tejidos como agentes de biopulido y para reducción de pelusa, frisado, modificación de textura y lavado a la piedra (N.K. Lange, en P. Suominen, T. Reinikainen (Eds.), *Trichoderma reesei Cellulases and Other Hydrolases*, Foundation for Biotechnical and Industrial Fermentation Research, Helsinki, 1993, págs. 263-272). Además, las variantes o los polipéptidos con actividad de glucósido hidrolasa descritos también pueden ser usados en el tratamiento de la madera para biorreducción a pasta o descortezado, fabricación de papel para modificación de fibra, blanqueamiento y reducción de refinación de costes de energía, tratamiento de aguas blancas, importante para reciclaje de aguas residuales, reciclaje de fibra lignocelulósica tal como desentintado y tratamiento de fibras secundarias y utilización de residuos de madera (S.D. Mansfield y A.R. Esteghlalian en S.D. Mansfield y J.N. Saddler (Eds.), *Applications of Enzymes to Lignocellulosics*, ACS Symposium Series 855, Washington, D.C., 2003, págs. 2-29).

35 [0341] La presente invención es descrita adicionalmente por los siguientes ejemplos, los cuales no limitan el alcance de la invención.

Ejemplos

Cepas

45 [0342] *Trichoderma reesei* RutC30 (ATCC 56765; Montencourt y Eveleigh, 1979, *Adv. Chem. Ser.* 181: 289- 301) fue derivado de *Trichoderma reesei* Qm6A (ATCC 13631; Mandels and Reese, 1957, *J. Bacteriol.* 73: 269-278).

50 [0343] Cepa Jal250 de *Aspergillus oryzae* (WO 99/61651) fue usado para la expresión de la beta-glucosidasa termoestable.

[0344] Se utilizó *Saccharomyces cerevisiae* YNG 344 (MAT α , ura3-52, leu-2 Δ 2, pep4 Δ 1, his4-539, cir⁰) para generar bibliotecas de glucósido hidrolasa mutagenizada (Cel7A).

55 [0345] Cepas bacterianas usadas para generar plásmidos fueron células ultracompetentes en *E. coli* XL-10 Gold (Stratagene, Inc., la Jolla it, CA).

Medios y soluciones

60 [0346] El medio de selección de levadura fue compuesto por litro de 6,7 g de base nitrogenada de levadura, 0,8 g de mezcla de suplemento completo (CSM, Qbiogene, Inc., Carlsbad, CA; sin uracilo y con 40 mg/ml de adenina), 5 g de ácido de casamino y 20 g de agar Noble. El medio también contenía 50 mM de succinato pH 5,0, glucosa al 2% y 25 μ g de cloranfenicol por ml.

ES 2 424 346 T3

- [0347] El medio de selección de levadura fue compuesto por litro de 6,7 g de base nitrogenada de levadura, 0,8 g de CSM-URA, 5 g de ácido de casamino y 20 g de agar Noble. El medio también contenía 50 mM de succinato pH 5,0, galactosa al 0,2%, glucosa al 0,1% y 25 µg de cloranfenicol por ml.
- 5 [0348] El medio YPD fue compuesto por litro de 10 g de extracto de levadura, 20 g de bacto peptona y 40 ml de glucosa al 50%.
- [0349] El medio inductor de la celulasa fue compuesto por litro de 20 g de fibras celulósicas naturales de Arbocel B800 (J. Rettenmaier USA LP, Schoolcraft, Michigan), 10 g de sólidos de maíz fermentado (Sigma Chemical Co., St. Louis, MO), 1,45 g de $(\text{NH}_4)_2 \text{SO}_4$, 2,08 g de $\text{KH}_2 \text{PO}_4$, 0,28 g de CaCl_2 , 0,42 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0,42 ml de solución de metales de *Trichoderma reesei* y 2 gotas de ácido plurónico; pH a 6,0 con NaOH 10 N.
- 10
- [0350] La solución de metales traza de *Trichoderma reesei* fue compuesta por litro de 216 g de $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 58 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 27 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 10 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 2,4 g de H_3BO_3 y 336 g de ácido cítrico.
- 15
- [0351] Las placas de selección COVE fueron compuestas por litro de 342,3 g de sacarosa, 20 ml de solución salina COVE, 10 mM de acetamida, 15 mM de mM CsCl_2 y 25 g de agar Noble.
- [0352] Las placas de uridina COVE2 plus fueron compuestas por litro de 30 g de sacarosa, 20 ml solución salina COVE, 10 mM de acetamida, 10 mM de uridina y 25 g de agar Noble.
- 20
- [0353] La solución salina COVE fue compuesta por litro de 26 g de KCl, 26 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 76 g de $\text{KH}_2 \text{PO}_4$ y 50 ml de metales traza COVE.
- [0354] La solución de metales traza COVE fue compuesta por litro de 0,04 g de $\text{NaB}_4 \text{O}_7 \cdot 10\text{H}_2\text{O}$, 0,4 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 1,2 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0,7 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0,8 g de $\text{Na}_2 \text{MoO}_4 \cdot 2\text{H}_2\text{O}$ y 10 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$.
- 25
- [0355] El medio PDA fue compuesto por litro de 39 g agar de dextrosa de patata.
- [0356] 1X SSC fue compuesto por litro de 8,765 g de cloruro sódico y 4,41 g citrato sódico.
- 30
- [0357] El tampón PEG fue compuesto por litro de 500 g de PEG 4000 (BDH, Poole, Inglaterra), 10 mM de CaCl_2 y 10 mM de Tris-HCl pH 7,5 (filtro esterilizado).
- [0358] El STC fue compuesto por litro de 1 M de sorbitol, 10 mM de CaCl_2 y 10 mM de Tris-HCl pH 7,5 y fue esterilizado con filtro.
- 35
- [0359] El medio inóculo de *Trichoderma reesei* fue compuesto por litro de 20 g de glucosa, 10 g de sólidos de maíz fermentado (Sigma Chemical Co., St. Louis, MO), 1,45 g de $(\text{NH}_4)_2 \text{SO}_4$, 2,08 g de $\text{KH}_2 \text{PO}_4$, 0,28 g de CaCl_2 , 0,42 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0,42 ml ml de solución de minerales traza de *Trichoderma reesei* y 2 gotas de ácido plurónico; pH final 5,0.
- 40
- [0360] El medio de fermentación de *Trichoderma reesei* fue compuesto por litro de 4 g de glucosa, 10 g de sólidos de maíz fermentado, 30 g de fibras celulósicas naturales de Arbocel B800 (J. Rettenmaier USA LP, Schoolcraft, Michigan), $3(\text{NH}_4)_2 \text{SO}_4$, 2,8 g de $\text{KH}_2 \text{PO}_4$, 2,08 g de CaCl_2 , 1,63 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0,75 ml de solución de metales traza de *Trichoderma reesei* y 1,8 ml de ácido plurónico.
- 45
- [0361] El medio de alimentación de *Trichoderma reesei* fue compuesto por litro de 600 g de glucosa, 20 g de celulosa B800, 35,5 g de $\text{H}_3 \text{PO}_4$ y 5 ml de ácido plurónico.
- 50
- [0362] El medio de inóculo de *Aspergillus oryzae* fue compuesto por litro de 50 g de glucosa, 2 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 g de $\text{KH}_2 \text{PO}_4$, 2 g de $\text{K}_2 \text{SO}_4$, 2,08 g de $\text{CaCl}_2 \cdot 2 \text{H}_2\text{O}$, 2 g de ácido cítrico, 10 g de extracto de levadura, 0,5 g de metales traza AMG y 2 g de urea, pH final 6,0.
- 55
- [0363] Los metales traza AMG están compuestos por litro de 14,3 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 2,5 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0,5 g de $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$, 13,8 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 8,5 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ y 3 g de ácido cítrico.
- [0364] El medio de fermentación de *Aspergillus oryzae* fue compuesto por litro de 2 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 2 g de $\text{KH}_2 \text{PO}_4$, 3 g de $\text{K}_2 \text{SO}_4$, 9 g de $(\text{NH}_4)_2 \text{HPO}_4$, 1 g de ácido cítrico $\cdot \text{H}_2\text{O}$, 10 g de extracto de levadura, 0,5 ml de metales traza AMG, 25 g de sacarosa y 0,55 ml de plurónico.

[0365] El medio de alimentación de *Aspergillus oryzae* fue compuesto por litro de 1 g de ácido cítrico. H₂O, 320 g de SatinSweet 65 y 5 g de plurónico.

[0366] El medio MDU2BP está compuesto por litro de 45 g de maltosa, 1 g de MgSO₄ · 7H₂O, 1 g de NaCl, 2 g de K₂SO₄, 12 g de KH₂PO₄, 7 g de extracto de levadura, 2 g de urea, 0,5 ml de solución de metales traza AMG.

Ejemplo 1: Fermentación y tejido micelial

[0367] RutC30 de *Trichoderma reesei* fue cultivado bajo condiciones estándares de inducción de celulasa como se describe en la técnica (Mandels and Weber, 1969, Adv. Chem. Ser. 95: 391-413). Muestras miceliales fueron cosechadas por filtración a través de papel de Whatman y congeladas rápido en nitrógeno líquido. Las muestras fueron almacenadas a -80° C hasta que fueron interrumpidas para extracción de ARN.

Ejemplo 2: Construcción de genoteca de ADNc de etiquetas de secuencia expresada (EST)

[0368] El ARN celular total fue extraído de las muestras miceliales descritas en Ejemplo 1 según el método de Timberlake y Barnard (1981, Cell 26: 29-37) y las muestras de ARN fueron analizadas por hibridación de Northern después de transferencia por adsorción de 1% geles de agarosa de formaldehído (Davis *et al.*, Basic Methods in Molecular Biology, Elsevier Science Publishing Co., Inc., New York). Las fracciones de ARNm poliadenilatado fueron aisladas de ARN total con un mRNA Separator Kit™ (Clontech Laboratories, Inc., Palo Alto, CA) según las instrucciones del fabricante. El ADNc bicatenario fue sintetizado usando aproximadamente 5 µg de poli(A)+ ARNm según el método de Gubler y Hoffman (1983, Gene 25: 263-269), excepto que un cebador Not I-(dT)₁₈ (Pharmacia Biotech, Inc., Piscataway, NJ) fue usado para iniciar la síntesis de la primera cadena. El ADNc fue tratado con nucleasa de judía mungo (Boehringer Mannheim Corporation, Indianapolis, IN) y los extremos se hicieron romos con T4 ADN polimerasa (New England Biolabs, Beverly, MA).

[0369] Los adaptadores *BamH I/EcoR I* fueron ligados a los extremos romo del ADNc. Después de la digestión con NotI, el ADNc fue seleccionado por tamaño (aprox. 0.7-4.5 kb) por 0,7% de electroforesis en gel de agarosa usando tampón TAE (4,84 g de Tris Base, 1,14 ml de ácido acético glacial y 2 ml de 0,5 M EDTA pH 8,0 por litro) y ligado con pYES2 (Invitrogen Corporation, Carlsbad, CA) que ha sido dividido con Not I más *BamH I* defosforilado con fosfatasa alcalina de intestino de ternero (Boehringer Mannheim Corporation, Indianapolis, IN). La mezcla de ligadura fue usada para transformar células *E. coli* TOP10 competentes (Invitrogen Corporation, Carlsbad, CA). Transformantes fueron seleccionados en placas de agar 2YT (Miller, 1992, A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria, Cold Spring Harbor Press, Cold Spring Harbor, New York) suplementadas con ampicilina a una concentración final de 50 µg por ml.

Ejemplo 3: Preparación de modelo y secuenciación de nucleótidos de clones de ADNc

[0370] De la genoteca de ADNc descrita en Ejemplo 2, aproximadamente 7000 colonias transformantes fueron escogidas directamente de las placas de transformación en platos de microtitulación de 96 pocillos que contenían 100 µl de caldo 2YT suplementado con 50 µg de ampicilina por ml. Las placas fueron incubadas durante toda la noche a 37° C con agitación a 200 r.p.m. Tras la incubación, se añadieron 100 µl de glicerol al 50% estéril cada pocillo. Los transformantes fueron replicados en, placas secundarias para microcultivo de 96 pocillos de plato profundo (Advanced Genetic Technologies Corporation, Gaithersburg, MD) con 1 ml de Magnificent Broth™ (MacConnell Research, San Diego, CA) suplementado con 50 µg de ampicilina por ml en cada pocillo. Las placas de microtitulación primarias fueron almacenadas congeladas a -80° C. Las placas secundarias de plato profundo fueron incubadas a 37° C durante toda la noche con agitación vigorosa (300 r.p.m.) en un agitador giratorio. Para prevenir derrames y contaminación cruzada y para permitir aireación suficiente, cada placa secundaria de cultivo fue cubierta con una almohadilla de polipropileno (Advanced Genetic Technologies Corporation, Gaithersburg, MD) y una cobertura de plato de microtitulación plástica.

[0371] El ADN fue aislado de cada pocillo usando un protocolo de equipo de minipreparación de 96 pocillos de Advanced Genetic Technologies Corporation (Gaithersburg, MD) según modificación de Utterback *et al.* (1995, Genome Sci. Technol. 1: 1-8). La secuenciación del ADN monopaso (EST) fue realizada con un secuenciador de ADN automatizado Modelo 377 XL de Perkin-Elmer Applied Biosystems (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA) usando química de terminador de coloración (Giesecke *et al.*, 1992, Journal de Virology Methods 38: 47-60) y un cebador de secuenciación T7: 5'-TAATACGACTCACTATAGGG-3' (SEQ ID NO: 3)

Ejemplo 4: Análisis de datos de secuencia de ADN de clones de ADNc

[0372] Los datos de la secuencia de nucleótidos fueron analizados en detalle en cuanto a la calidad y las secuencias de vector y las llamadas de base ambiguas en los extremos de las secuencia de ADN fueron cortadas y todas las secuencias fueron comparadas entre sí con asistencia del software PHRED/PHRAP (University de Washington, Seattle, WA). Los contigs y los singletons fueron traducidos en seis marcos y analizados en comparación con bases de datos de proteínas públicamente disponibles usando el software GeneMatcher™ (Paracel, Inc., Pasadena, CA) con un algoritmo de Smith-Waterman modificado que usa la matriz BLOSUM 62.

Ejemplo 5: Identificación de clones de ADNc que codifican una celobiohidrolasa I de Familia 7 (Cel7A)

5 [0373] Los clones de ADNc putativos que codifican una celobiohidrolasa de familia 7 (Cel7A) fueron identificados por comparación de la secuencia de aminoácidos deducida del ensamblado ESTs con secuencias proteínicas depositadas en bases de datos públicamente disponibles tal como Swissprot, Genpept y PIR. Un clon, EST Tr0221 de *Trichoderma reesei*, fue seleccionado para análisis de secuencia de nucleótidos que reveló un inserto pYES2 de 1821 pares de bases que contenía un marco de lectura abierto de 1452 de pares de bases como se muestra en SEC ID n.º: 1 y una secuencia de aminoácidos deducida como se muestra en SEC ID n.º: 2. El plásmido con celobiohidrolasa I Cel7A de *Trichoderma reesei* fue designado pTr0221.

Ejemplo 6: Construcción de vectores de *Saccharomyces cerevisiae* para la generación de bibliotecas de celobiohidrolasa I Cel7A de *Trichoderma reesei* primaria y redistribuida

15 [0374] Dos vectores fueron utilizados en la generación de bibliotecas redistribuidas y primarias, pJC106 (WO9510602) y pAJ052 (Figura 1). El plásmido pJC106 es un derivado de pYES2 (Invitrogen Inc., Carlsbad, CA) pero difiere en que pJC106 tiene un replicón de 2 micras de longitud completa, que reemplaza las 2 micras parciales en pYES2 y contiene la secuencia codificante de peroxidasa de *Coprinus cinereus* (CIP) (Cherry *et al.*, 1999, Nat. Biotechnol. 4: 379-384), que es regulada por el promotor *GAL1*.

20 [0375] Para pAJ052, un fragmento de ADN de 1452 bp que se extiende desde el codón de inicio ATG al codón de terminación TAA de la secuencia codificante de Cel 7A de *Trichoderma reesei* fue amplificado por PCR de pTr0221 (Ejemplo 5) usando los cebadores aGal_776.1 (sentido) y aGal_776.1A (antisentido) mostrados a continuación:

25 Cebador
aGal_776.1:
5'-

TATACCTCTATACTTTAACGTC AAGGAGAAAAAACTATAGGATCCACCATGTATCGGAAGTTGCCG-

3' (SEQ ID NO: 4)

Cebador aGal_776.1A
5'-**CATAACTAATTACATGATGCGGCOOCTCTAGATGCACATGACTCGAGTTACAGGCACTGAGAGTAG-**

30 **3' (SEQ ID NO: 5)**

[0376] Los cebadores aGal_776.1 y aGal_776.1A fueron diseñados para contener un promotor *GAL1* homólogo (Giniger and Ptashne, 1988, Proc. Natl. Acad. Sci., USA 85: 382-386) y la secuencia del terminador *CYC1* (Osbourne and Guarente, 1988, Genes Dev. 2: 766-772) (subrayado; respectivamente) de pJC106 para la recombinación homóloga *in vivo* del producto de PCR y pJC106. Los cebadores aGal_776.1 y aGal_776.1A también fueron diseñados para contener sitios de restricción *BamH I* y *Xho I*, respectivamente. El cebador aGal_776.1 además fue diseñado para contener la secuencia Kozak de levadura (ACC, -3 a -1 par de bases; Kozak, 1984, Nature 308: 241-246) inmediatamente arriba del gen ATG de celobiohidrolasa I (Cel7A) de *Trichoderma reesei*. La reacción de amplificación (50 µl) fue compuesta por tampón 1X PCR (Applied Biosystems Inc., Foster City, CA), 0,2 mM de dNTPs, 3,2 pM de cebador aGal_776.1, 3,2 pM de cebador aGal_776.1A, aproximadamente 100 ng de pTr0221 y 2,5 unidades de *Taq* ADN polimerasa (Roche Applied Science, Mannheim, Alemania). Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 (Eppendorf EG, Hamburgo, Alemania) programado durante 1 ciclo a 94° C durante 3 minutos seguido de 30 ciclos cada uno a 94° C durante 30 segundos, 55° C durante 30 segundos y 72° C durante 90 segundos seguido de 1 ciclo a 72° C durante 5 minutos. El producto de PCR luego fue purificado usando un equipo de PCR QIAquick (QIAGEN Inc., Valencia, CA), según las instrucciones del fabricante y luego fue introducido en *Saccharomyces cerevisiae* por recombinación *in vivo*. Para realizar la recombinación *in vivo*, aproximadamente 100 ng de pJC016, digeridos con *BamH I* y *Xho I*, y aproximadamente 500 ng del fragmento de PCR purificado, fueron cotransformados en YNG 344 de *Saccharomyces cerevisiae* después del protocolo de transformación de levadura YEASTMAKER (Clontech Laboratories Inc., Palo Alto, CA). La transformación fue colocada en placas sobre medio de selección de levadura para crecimiento de colonias a 30° C durante 4 días.

50 [0377] Una única colonia fue seleccionada y se aisló ADN plásmido según el protocolo descrito por Kaiser y Auer, 1993, BioTechniques 14: 552, que fue posteriormente transformado en la cepa de *E. coli* XL-10 (Stratagene Inc., La Jolla, CA) según las instrucciones del fabricante. El plásmido derivado de la cepa de *E. coli* transformada fue ordenado para verificar la fidelidad del PCR y el evento de recombinación.

Ejemplo 7: Generación de bibliotecas primarias de celobiohidrolasa I Cel7A mutagenizada en *Saccharomyces cerevisiae*

5 [0378] En un esfuerzo para identificar regiones de la celobiohidrolasa I Cel7A de *Trichoderma reesei* que son críticos para la termoestabilidad de proteína y actividad de alta temperatura mejorada, el gen de celobiohidrolasa I Cel7A de *Trichoderma reesei* entero de tipo salvaje fue mutagenizado usando PCR con tendencia al error con secuencias homólogas al vector de expresión de levadura pJC106 (Figura 2), que puede sufrir recombinación *in vivo* entre dominios homólogos de fragmentos diferentes, generando plásmidos de replicación circular de una combinación de vector linealizado y productos de PCR.

10 [0379] Los productos de PCR para reparación del espacio fueron generados usando una de las siguientes combinaciones de modelo/cebador:

1) El cebador aGal_776.1 y el cebador aGal_776.1a, mostrados más abajo, fueron usados en la Amplificación de PCR con tendencia al error del gen de celobiohidrolasa I Cel7A de pTR0221 para generar secuencias mutagenizadas.

15 2) El cebador yes2term y el cebador CiPpcrdwn, mostrados más abajo, fueron usados en la Amplificación de pCR con tendencia al error del gen de celobiohidrolasa I Cel7A de pAJ052 para generar secuencias mutagenizadas.

3) El cebador cJC106.1a y el cebador CiPpcrdwn, mostrados más abajo, también fueron usados en la Amplificación de PCR con tendencia al error del gen de celobiohidrolasa I Cel7A de pAJ052 para generar secuencias mutagenizadas.

20 [0380] Los fragmentos fueron clonados en pJC106 para la expresión de variantes de celobiohidrolasa I Cel7A en la levadura.

Cebador
aGal_776.1:
5'-

TATACCTCTATACTTTAACGTCAAGGAGAAAAACTATAGGATCCACCATGTATCGGAAGTTGCCG-

25 **3' (SEQ ID NO: 6)**

Cebador
776.1a:

aGal-

5'-CATAACTAATTACATGATGCGGCCCTCTAGATGCACATGACTCGAGTTACAGGCCACTGAGAGTAG-

30 **3' (SEQ ID NO: 7)**

Cebador yes2term:
5'-GGCGTGAATGTAAGCGTGAC-3' (SEC ID N.º: 8)

Cebador CiPpcrdwn:
5'-CTGGGGTAATTAATCAGCGAAGCGATGA-3' (SEC ID N.º: 9)

35 Cebador cJC106.1a:
5'-GCGTACACGCGTCTGTACA-3' (SEC ID N.º: 10)

[0381] Las Amplificaciones de PCR con tendencia al error (50 µl) fueron compuestas de tampón 1X PCR con MgCl₂, 0,2 mM de dATP, 0,2 mM de dGTP, 0,1 mM de dCTP y 0,1 mM de dTTP, 50 pmol de cebador sentido y antisentido, 0,05 mM a 0,6 mM de MnCl₂ y 10- 50 ng de ADN plásmido (en algunos casos pTR0221, en otros casos pAJ052). Las reacciones fueron incubadas usando un termociclador de MJ Research (MJ Research, Inc. Boston, MA) programado para un ciclo a 95° C durante 3 minutos después de lo cual se añadieron 2,5 unidades de Amplitaq (Perkin Elmer, Foster City, CA) seguidos de 30 ciclos cada a 95° C durante 60 segundos, 55° C durante 60 segundos y 72° C durante 90 segundos. Las reacciones fueron luego incubadas a 72° C para un extensión de 5 minutos. Una alícuota de cada producto de PCR fue realizada en un gel de agarosa al 0,7% usando tampón TAE, tal y como se describe anteriormente, generando bandas previstas de aproximadamente 1680 a 2030 pares de bases. Las reacciones de PCR fueron purificadas usando un equipo de purificación de PCR MiniElute (QIAGEN, Valencia, CA) eluido en 50 µl de tampón EB (QIAGEN Valencia, CA).

50 [0382] El plásmido pJC106 fue escindido por digestión con *Bam*H I y *Xho* I y luego fue purificado en gel usando una columna QIAquick Minielute (QIAGEN, Inc., Valencia, CA). La digestión fue verificada por fraccionamiento de una alícuota de la digestión en un gel de agarosa al 0,8% usando tampón TAE y coloreando con bromuro de etidio donde los fragmentos previstos de 10771 pares de bases (escindidos) y 1030 pares de bases (del gen de peroxidasa de *Coprinus cinereus*) fueron obtenidos.

55 [0383] Las reacciones de la PCR fueron mezcladas a aproximadamente una proporción de 3 a 1 con el vector pJC106 escindido para cotransformación en las células competentes de *Saccharomyces cerevisiae* YNG344. Los fragmentos cotransformados, amplificados usando los cebadores aGal_776.1 y aGal_776.1a, contuvieron al menos 67 pares de

bases de 5' y 66 pares de bases de 3' ADN homólogo, amplificados yes2term y CiPpcrdwn contuvieron al menos 293 pares de bases de 5' y 41 pares de bases de 3' ADN homólogo y amplificado usando cJC106.1a y CiPpcrdwn contuvieron al menos 293 pares de bases de 5' y 190 pares de bases de 3' ADN homólogo en los extremos para facilitar la reparación del espacio del plásmido expresado. Las células competentes de *Saccharomyces cerevisiae* YNG 344 fueron preparadas antes de cada transformación siguiendo el Protocolo de transformación de levadura YEASTMAKER con las siguientes modificaciones: (1) el volumen de cultivo de levadura usado para inocular la incubación durante toda la noche (16-20 horas) fue entre 100-1000 µl; (2) la recuperación de células después de la transformación fue realizada en medio YPD durante 45 minutos a 30° C; (3) la mezcla de transformación fue dividida en partes alícuotas para colocar en placas en medio de selección de levadura y fueron congeladas a -80° C en un congelador de índice controlado (Nalge Nunc International, Rochester, NY).

Ejemplo 8: Construcción de vectores de expresión *Aspergillus oryzae* pAILo1 y pAILo2

[0384] Como un vector estructural para clonar varias de las variantes de celobiohidrolasa Cel7A, se construyeron dos vectores de expresión de *Aspergillus oryzae*. El vector pAILo1 fue construido modificando pBANE6 (patente U.S. 6.461.837), que comprende el promotor alfa-amilasa de *Aspergillus oryzae*, la secuencia del terminador de amiloglucosidasa de *Aspergillus niger* (terminador AMG) y gen de acetamidasa de *Aspergillus nidulans* (*amdS*). La modificación de pBANE6 fue realizada eliminando primero tres sitios de restricción *Nco* I en las posiciones 2051, 2722 y 3397 bp del marcador de selección *amdS* por mutagénesis dirigida al sitio. Todos los cambios fueron diseñados para que fueran "silenciosos" dejando sin cambios la secuencia de la proteína real del producto genético *amdS*. La eliminación de estos tres sitios fue realizada simultáneamente con un equipo de mutagénesis dirigida GeneEditor (Promega, Madison, WI) según las instrucciones del fabricante usando los siguientes cebadores (el nucleótido subrayado representa la base cambiada):

25 Cebador AMDS3NcoMut (2050):
5'-GTGCCCCATGATACGCCTCCGG-3' (SEC ID N.º: 11)
Cebador AMDS2NcoMut (2721):
5'-GAGTCGTATTTCCAAGGCTCCTGACC-3' (SEC ID N.º: 12)
30 Cebador AMDS1NcoMut (3396):
5'-GGAGGCCATGAAGTGGACCAACGG-3' (SEC ID N.º: 13)

[0385] Un plásmido que comprende los tres cambios de secuencia previstos fue luego sometido a mutagénesis dirigida, usando un equipo de mutagénesis QuickChange (Stratagene, La Jolla, CA), para eliminar el sitio de restricción *Nco* I al final del terminador AMG en la posición 1643. Los siguientes cebadores (el nucleótido subrayado representa la base cambiada) fueron usados para la mutagénesis:

35 Cebador superior para mutagenizar la secuencia del terminador AMG de *Aspergillus niger*:
5'-CACCGTGAAAGCCATGCTCTTTCTTCGTGTAGAAGACCAGACAG-3' (SEC ID N.º: 14)
Cebador antisentido para mutagenizar la secuencia del terminador de AMG de *Aspergillus niger*.
40 5'-CTGGTCTTCTACACGAAGGAAAGAGCATGGCTTTCACGGTGTCTG-3' (SEC ID N.º: 15)

[0386] El último paso en la modificación de pBANE6 fue la adición de un nuevo sitio de restricción *Nco* I en el principio del polilíggador usando un equipo de mutagénesis QuickChange y los siguientes cebadores (los nucleótidos subrayados representan las bases cambiadas) para producir pAILo1 (Figura 3).

45 Cebador de sentido para mutagenizar el promotor de TAKA de *Aspergillus oryzae*:
5'-CTATATACACAACCTGGATTTACCATGGGCCCGCGCCGCAGATC-3' (SEC ID N.º: 16)

50 Cebador antisentido para mutagenizar el promotor de TAKA de *A. oryzae*:
5'-GATCTGCGCCCGCGGCCCATGGTAAATCCAGTTGTGTATATAG-3' (SEC ID N.º: 17)

[0387] El gen *amdS* de pAILo1 fue cambiado con el gen *pyrG* de *Aspergillus nidulans*. El plásmido pBANE10 (Figura 4) fue usado como una fuente para el gen *pyrG* como un marcador de selección. El análisis de la secuencia de pBANE10 mostró que el marcador *pyrG* fue contenido dentro de un fragmento de restricción *Nsi* I y no contiene sitios de restricción *Nco* I ni *Pac* I. Ya que el *amdS* también es flanqueado por sitios de restricción *Nsi* I, la estrategia para alternar el marcador de selección fue un simple cambio de fragmentos de restricción *Nsi* I.

[0388] El ADN plásmido de pAILo1 y pBANE10 fue digerido con la enzima de restricción *Nsi* I y los productos fueron purificados por electroforesis en gel de agarosa al 0,07% usando tampón de TAE. El fragmento de *Nsi* I de pBANE10 con el gen *pyrG* fue ligado a la estructura de pAILo1 para reemplazar el fragmento de ADN *Nsi* I original con el gen *amdS*. Los clones recombinantes fueron analizados por digestión de enzima de restricción para determinar la orientación del inserto. Se seleccionó un clon con el gen *pyrG* transcrito en el sentido contrario a las agujas del reloj. El plásmido nuevo fue designado pAILo2 (Figura 5).

Ejemplo 9: Diseño racional de G205R variante de celobiohidrolasa I Cel7A y generación de bibliotecas primarias de G205R en *Saccharomyces cerevisiae*

[0389] La secuencia de proteína celobiohidrolasa I Cel7A de *Trichoderma reesei* (SEC ID n°: 2) fue comparada con otras proteínas de la misma familia de enzimas. Las secuencias incluyeron una celobiohidrolasa I Cel7A de *Chaetomium thermophilum* (WO 03/000941), *Hemicola insolens* (WO 95/02675) y *Neurospora crassa* (SWISSPROT: pariente filogenético cercano P38676a de *Chaetomium termofilum*). Se hicieron alineamientos múltiples de secuencias proteínicas de celobiohidrolasa I Cel7A de *Chaetomium termofilum*, *Hemicola insolens*, *Trichoderma reesei* y *Neurospora crassa*, usando el software ClustalX versión 1.81, (National Center for Biotechnology Information, NIH Bethesda, MD) (Thompson *et al*, 1994, Nucleic Acids Res 22: 4673-4680; Thompson *et al*, 1997, Nucleic Acids Res 25: 4876-4882), usando la matriz Gonnet con parámetros de penalización de espacio predeterminados. Las regiones que aparecieron mal alineadas fueron realineadas iterativamente, a veces usando una matriz alternativa (Blosum) y/o parámetros de espacio variables. Se generaron modelos de homología para secuencias de celobiohidrolasa I Cel7A públicamente disponibles usando el servicio automatizado SwissModel (Biozentrum, Basilea, Suiza). El modelo de homología para celobiohidrolasa I Cel7A de *Hemicola insolens* fue generado usando los programas Insight II (Accelrys, San Diego CA). El programa DeepView (Guex y Peitsch, 1997, Electrophoresis 18: 2714-2723) fue usado para introducir mutaciones virtuales y para todas las otras manipulaciones de estructura y minimización de energía. La estructura de referencia usada para la celobiohidrolasa I Cel7A de *Trichoderma reesei* fue PDB: 7CEL (Divne *et al*, 1994, Science 265: 524-528 ; Stahlberg *et al*, 1996, 264:337-349). Sobre la base de las comparaciones, se priorizaron las mutaciones potenciales para celobiohidrolasa I de *Trichoderma reesei* según la probabilidad que éstas crearían una interacción estabilizante nueva (par iónico y/o enlace H) y también la probabilidad de incidencia en las secuencias y las estructuras de celobiohidrolasa I Cel7A de los hongos termofílicos *Chaetomium termofilum* y *Hemicola insolens* y su ausencia en el hongo mesofílico *Neurospora crassa*.

[0390] Una de las sustituciones de aminoácidos sugeridas fue un cambio de glicina en la posición 205 a arginina. La variante G205R fue racionalmente diseñada para introducir par de ión con E190 y E239. Para generar la sustitución G205R, el gen de celobiohidrolasa I Cel7A de *Trichoderma reesei* fue subclonado en el vector de *Aspergillus oryzae* pAILo2 digerido con *Nco I* y *Pac I* para formar una conjunción perfecta con el ATG del gen y el promotor de alfa-amilasa de *Aspergillus oryzae* y la secuencia del terminador de amiloglucosidasa de *Aspergillus niger*. La subclonación del gen de celobiohidrolasa I Cel7A en pAILo2 fue realizada diseñando dos cebadores, mostrados más abajo, que permitieron la clonación los sitios *Nco I* y *Pac I*. El cebador TR0221.7: 5'-GCAACATGTATCGGAAGTTGGC-3' (SEC ID N.º: 18) incorporó un sitio *BspLU II*, que fue compatible con el sitio *Nco I* en pAILo2, al extremo 5' de la celobiohidrolasa I Cel7A y el cebador cTR0221.7a: 5'-AATTAATTTTACAGGCACTGAG-3' (SEC ID N.º: 19) incorporó un sitio *BspLU II* sitio en el extremo 3'.

[0391] La amplificación del gen de celobiohidrolasa I Cel7A fue realizada usando tampón tampón de reacción de polimerasa 1X Tgo (Boehringer Mannheim Co, Indianapolis, IN), 25 ng de pTR0221, 0,2 mM cada uno de dATP, dGTP, dCTP y dTTP, 50 pmol de cada cebador (cTR0221.7 y cTR0221.7a) y 1 unidad de polimerasa Tgo (Boehringer Mannheim Co, Indianapolis, IN). Las reacciones fueron incubadas usando un termociclador MJ Research programado para un ciclo a 95° C durante 5 minutos, seguido de 35 ciclos cada uno a 94° C durante 60 segundos, 55° C durante 45 segundos y 72° C durante 2 minutos. Las reacciones fueron luego incubadas a 72° C para un extensión de 5 minutos. Una alícuota de cada producto de PCR fue realizada en un gel de agarosa al 0,7% usando tampón TAE generando bandas previstas de aproximadamente 1545 pares de bases. El producto de PCR de 1545 pares de bases fue subclonado usando un equipo de clonación PCR4 blunt TOPO (Invitrogen, Carlsbad, CA). El plásmido resultante fue digerido con *BspLU II* y *Pac I* y fraccionado en un gel de agarosa al 0,7% usando tampón TAE generando una secuencia codificante prevista de 1,5 kb, que fue cortado y purificado en gel usando una columna Ultra-free DA Amicon (Millipore, Billerica, MA). El fragmento resultante fue posteriormente ligado en pAILo2, que fue digerido de forma similar, para generar el vector de expresión designado pCW026 (Figura 6) con el gen de celobiohidrolasa I Cel7A de *Trichoderma reesei*.

[0392] Trabajando desde plásmido inicial pCW026, la variante G205R se obtuvo mediante mutación de una guanosina a citidina en la base 664 de la secuencia codificante de celobiohidrolasa I Cel7A, usando un Equipo de mutagénesis dirigida de cambio rápido (Stratagene, La Jolla, CA) con pCW026 como el modelo. Los cebadores usados para incorporar esta mutación fueron el cebador G205R.1 y el cebador G205R.1a que se muestran a continuación:

G205R. 1:
5'-GAACACGGGCATTGGACGACACGGAAGCTGCTG-3' (SEC ID N.º: 20)
G205R.1a:
5'-CAGCAGCTTCCGTGTCGTCCAATGCCCGTGTTTC-3' (SEC ID N.º: 21)

El vector de expresión resultante fue designado pNP776G205R (Figura 7).

Ejemplo 10: Selección de bibliotecas de celobiohidrolasa I Cel7A

[0393] Las bibliotecas de celobiohidrolasa I Cel7A primarias fueron extendidas en medio de selección de levadura de agar en Genetix Qtrays (placas de Petri de 22 x 22 cm, Genetics Ltd., Hampshire, Reino Unido) e incubadas durante 5

días a 30° C. Usando un Genetix QPix (Genetix Ltd., Hampshire, Reino Unido), las colonias fueron escogidas en placas de 96 pocillos con medio de selección de levadura. Las placas fueron incubadas durante 7 días a 30° C. Usando un robot ORCA (Beckman Coulter, Fullerton, CA), las placas de crecimiento fueron transportadas a un Biomek Fx (Beckman Coulter, Fullerton, CA) y las muestras de caldo fueron quitadas de la placa de crecimiento y divididas en partes alícuotas en dos placas de 96 pocillos de policarbonato de fondo en V. El sustrato 4-metilumbeliferil-beta-D-lactosida de celobiohidrolasa I (MUL, Marker Gene Tech. Inc., Eugene, OR) se añadió a cada placa de 96 pocillos de fondo en V a una concentración final de 0,2, 0,1 o 0,05 mg de 4-metilumbeliferil-beta-D-lactosida por ml, 0,1 M de succinato pH 5,0 y Tween-20 al 0,01%. Las placas de ensayo fueron transferidas a un incubador de temperatura controlada, donde una placa fue incubada a 50° C durante 45 minutos y otra fue incubada a una temperatura predeterminada, entre 62° C y 65° C durante 45 minutos. Después esta incubación, las placas fueron enfriadas a 4° C durante 1 minuto y luego fueron transferidas al Biomek Fx donde los ensayos fueron detenidos por adición de Tris-Cl, pH 9,5 a una concentración final de 0,75 M. Las muestras de reacción detenida fueron diluidas en agua y la fluorescencia de 4-metilumbeliferil liberada por hidrólisis de celobiohidrolasa I Cel7A de 4-metilumbeliferil-beta-D-lactosida se midió usando un fluorómetro BMG FLUOStar Galaxy (Defenburg; Alemania) (excitación a 360 nm, emisión a 460 nm). El índice de la fluorescencia de la placa tratada a alta temperatura ("actividad a alta temperatura") fue comparada con la fluorescencia de las mismas muestras incubadas a 50° C ("actividad a baja temperatura"), usando Microsoft Excel (Microsoft Corporation, Redmond, WA) para determinar el índice de actividad térmica relativo para cada variante. Sobre la base de los índices de actividad térmica, la selección de bibliotecas construidas en el Ejemplo 7 y el Ejemplo 9 generó las variantes presentadas en la Tabla 1. La Tabla 1 muestra el grado de mejora para variantes nuevas de celobiohidrolasa I Cel7A según la medición por evaluación del índice de actividad térmica de la actividad a 64° C en relación con la actividad a 50° C. Para los mutantes obtenidos en la pantalla primaria, las mejoras varían de mejoras 2,60 veces superiores a 10,20 veces superiores en relación a la enzima de tipo salvaje. Para los mutantes obtenidos de redistribución, la mejora observada fue actividad térmica 12,40 veces a 19,20 veces mejor que la enzima de tipo salvaje.

Tabla 1. Variantes de Cel7A con termoestabilidad y actividad térmica mejoradas. "Relación de mejora" indica la mejora relativa en el índice de actividad térmica, medida a 64° C/50° C.

Variante de Cel7A	Sustituciones de aminoácidos*	Relación de mejora
Tipo salvaje	ninguna	1
G205R	G205R	2,60
776-M1	T226A	6,20
776-M3	P227A, C486W	7,80
776-M4	S113N, S196T, T462A	6,00
776-M21	N301S, E337V	4,00
776-M22	S196P, T350S	1,60
776-M23	G22D, G467S	1,40
776-M26	S21P, S57N	3,80
776-M27	S411F	7,60
776-M30	T41I	2,60
776-M32	K157R, G205R, T255P	5,20
776-M35	G205R, S411F	8,40
776-M40	G205R, P227A	10,20
776-M41	G205R, H206Y	3,40
776-M42	S8P, G205R	4,60
776-M52	G94S, G205R	4,00
776-M53	S196P, G205R	4,80
776-M57	S113N, S196T, P227A, T462A	14,60
776-M65	S57N	5,00
776-M71	T383A, T455A	2,60
776-M73	N373H	4,80
776-M101	S113N, S411F	12,40
776-M108	T41I, E193K, S411F	15,60
776-M109	N49S, S113N, P227A, P438L	15,00
776-M124	Y247C, Y371C, S411F	17,40
776-M125	S21P, S57N, T246I, R251K, S411F	18,00
776-M192	K157R, G205R, T255P, S411F	15,00
776-M216	S113N, S196T, P227A, S411F	17,80
776-M252	S113N, S196T, P227A, T356I, T462A	19,20

*La numeración utiliza el primer residuo de la enzima Cel7A madura como posición 1.

Ejemplo 11: Secuenciación de ADN de las variantes

[0394] Para determinar la secuencia de las variantes de celobiohidrolasa I Cel7A derivadas de las bibliotecas de los Ejemplos 7 y 9, se aisló ADN plásmido. Cada variante fue estriada sobre medio de selección de levadura de agar e incubada durante 3-5 días a 30° C. Ocho colonias fueron aisladas e inoculadas en 1 ml de medio de selección de levadura, y cultivadas durante 5-8 días a 30° C. Las placas de agar fueron recultivadas a 30° C durante 3-5 días. El caldo de cultivo de las colonias individuales de cada variante fue evaluada en relación a la actividad térmica mejorada de la variante de celobiohidrolasa I Cel7A como se describe en el Ejemplo 10 para determinar cuáles tuvieron índices de actividad térmica mejorada en relación a la celobiohidrolasa I Cel7A de tipo salvaje (ts). El plásmido fue rescatado de las colonias que produjeron variantes de celobiohidrolasa I Cel7A con actividad térmica mejorada como describen Kaiser y Auer, 1993, BioTechniques 14: 552, usando la cepa de *E. coli* XL-10 (Stratagene Inc., La Jolla, CA).

[0395] La secuenciación del ADN fue realizada usando un secuenciador ABI 3700 (Applied Biosystems, Foster City, CA) usando química de terminador de tinte (Giesecke *et al.*, 1992, Journal de Virol. Methods 38: 47-60). El ADN plásmido para secuenciación fue preparado usando un Biorobot 9604, (QIAGEN Valencia, CA). La región de codificación entera para variante de celobiohidrolasa I Cel7A de *Trichoderma reesei* fue ordenada usando 0,5 µl de ADN plásmido y 3,2 pmol de los siguientes cebadores:

cTr0221.1: 5'-CTTCTTGGCCACAGCTCGTG-3' (SEC ID N.º: 22)
 cTr0221.2: 5'-GGCTTTGTCACCCAGTCTGC-3' (SEC ID N.º: 23)
 cTr0221.3: 5'-CGTCATCCAACAACGCGAAC-3' (SEC ID N.º: 24)
 cTr0221.4: 5'-TTCGAGACGTCGGGTGCCAT-3' (SEC ID N.º: 25)
 cTr0221.4: 5'-CGCGGAAGCTGCTCCACCAG-3' (SEC ID N.º: 26)
 cTr0221.1A: 5'-AATGGAGAGGCTGTTACCGC-3' (SEC ID N.º: 27)

Los ficheros de registro de secuencia fueron editados usando el Vector NTI Contig Express (Informax, Inc., Bethesda, MD).

Ejemplo 12: Generación de bibliotecas con tendencia a error usando genes de celobiohidrolasa I Cel7A variante como modelos

[0396] Se realizó la mutagénesis aleatoria de varias de las variantes mejoradas. Se utilizó ADN plásmido que había sido rescatado de la levadura como se describe en el Ejemplo 10, y se realizó PCR con tendencia al error como se describe en el ejemplo 7. Además, la variante G205R fue mutagenizada usando pNP776G205R como un modelo en PCR con tendencia al error. Los siguientes cebadores fueron usados para generar secuencias mutagenizadas:

Cebador

aGal_776.1:

5'TATACCTCTATACTTTAACGTCAAGGAGAAAAACTATAGGATCCACCATGTATCGGAAGTTGGCC
 G-3' (SEQ ID NO: 28)

Cebador

aGal_776.1a:

5'CATAACTAATTACATGATGCGGCCCTCTAGATGCACATGACTCGAGTTACAGGCACTGAGAGTAG-
 3' (SEQ ID NO: 29)

[0397] Las amplificaciones de PCR con tendencia al error fueron analizadas en geles de agarosa al 0,7% usando tampón de TAE y transformadas con vector pJC106 escindido en *Saccharomyces cerevisiae* como se describe en el Ejemplo 7.

[0398] Las variantes 76-M35,776-M40,776-M41,776-M42,776-M52, y 776-M53 fueron obtenidas de bibliotecas de selección creadas por amplificación con tendencia al error del modelo G205R. La variante 776-252 fue derivada de una biblioteca creada por amplificación mutagénica de la variante 776-M57. El ADN de estas variantes fue rescatado y ordenado como se describe en el Ejemplo 10.

Ejemplo 13: Bibliotecas redistribuidas de celobiohidrolasa I Cel7A

[0399] Para redistribuir las variantes de celobiohidrolasa I Cel7A derivadas de mutagénesis de celobiohidrolasa I Cel7A de tipo salvaje y modelos G205R, el ADN plásmido fue aislado de las variantes como se describe en el ejemplo 10. Los genes de celobiohidrolasa I Cel7A fueron amplificados de las variantes usando los siguientes cebadores:

Cebador CiPpcrdwn:

5'-CTGGGGTAATTAATCAGCGAAGCGATGA-3' (SEC ID N.º: 30)

Cebador cJC106.1A:

5'-GCGTACACGCGTCTGTACA-3' (SEC ID N.º: 31)

5 [0400] Cada reacción de amplificación (50 µl) fue compuesta por tampón 1X PCR, 0,2 mM de dNTPs, 3,2 pM de cebador aGal_776.1, 3,2 pM de cebador aGal_776.1A, aproximadamente 100 ng de pTr0221, y 2,5 unidades de Taq ADN polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 programado durante 1 ciclo a 94° C durante 3 minutos seguidos de 30 ciclos cada uno a 94° C durante 30 segundos, 55° C durante 30 segundos, y 72° C durante 90 segundos (extensión final de 5 minutos). Los productos de PCR luego fueron purificados usando un equipo de PCR QIAquick según las instrucciones del fabricante.

10 [0401] Para redistribución de las variantes de celobiohidrolasa I Cel7A, los productos de PCR de los genes de las variantes amplificadas fueron combinados con aproximadamente 50 ng de pJC016 digeridos con *Bam*H I y *Xho* I, y 100-200 ng de fragmentos surtidos de PCR fueron cotransformados en *Saccharomyces cerevisiae* YNG 344 después del protocolo de transformación de levadura YEASTMAKER para generar bibliotecas redistribuidas, como se describe en el ejemplo 7. Se generó un gran número de bibliotecas redistribuidas. El número total de variantes incluidas en una biblioteca única fue de dos a diez. El porcentaje de variantes activas en estas bibliotecas varió de 87% a 94%, según la actividad a 50° C determinada en el ensayo de 4-metilumbeliferil-beta-D-lactosidase descrito en ejemplo 9. Ya que la Taq ADN polimerasa usada tiene baja actividad de corrección, el proceso de redistribución puede tener nuevas mutaciones introducidas en la secuencia codificante de celobiohidrolasa I Cel7A.

20 [0402] Las colonias de las bibliotecas redistribuidas fueron escogidas y seleccionadas como se describe en el Ejemplo 9, usando una comparación de actividad a 64° C con la actividad a 50° C usando 4-metilumbeliferil-beta-D-lactosida como sustrato para valorar el grado de termoestabilidad y actividad térmica mejoradas de cada variante.

25 [0403] La selección de las bibliotecas redistribuidas dio como resultado el aislamiento de variantes mejoradas, designadas 776-M57,776-M101,776-M108,776-M109,776-M124,776-M125,776-M192 y 776-M216. El ADN de estas variantes fue rescatado y ordenado como se describe en el Ejemplo 10. Algunas de las sustituciones identificadas por secuenciación del ADN de las variantes 776-M57,776-M101,776-M108,776-M109,776-M124,776-M125,776-M192 y 776-M216 contienen sustituciones previamente identificadas por secuenciación de los mutantes de bibliotecas con tendencia al error. Por ejemplo, 776-M57 contuvo la mutación S113N que también fue encontrada en un progenitor (776-M4) de la biblioteca redistribuida de la cual fue derivada Asimismo, 776-M101 contuvo S411F, una sustitución encontrada en uno de los progenitores (776-M27) de la biblioteca redistribuida de la cual fue era obtenida. Otras sustituciones encontradas en las variantes derivadas de bibliotecas redistribuidas, incluidas N49S, E193K, R251K, T246I, Y247C, Y371C y P438L, no fueron observadas en las variantes ordenadas que comprendieron las variantes progenitoras para las bibliotecas redistribuidas correspondientes. Estas mutaciones pueden haber sido introducidas por un evento de mutagénesis mediada por PCR durante la amplificación del ADN variante durante la construcción de las bibliotecas redistribuidas.

Ejemplo 14: Mutagénesis de saturación específica de sitio

40 [0404] Para seleccionar la sustitución óptima para actividad térmica mejorada en posiciones específicas en celobiohidrolasa I Cel7A, se realizó aleatorización específica de sitio. Los aminoácidos G94; S196 y P227 fueron aleatorizados por sustitución del codón de tipo salvaje con NN (G/C) usando PCR megacebador (Landt, *et al.*, 1990 Gene 96: 125-128). Se realizó aleatorización en las tres posiciones usando los siguientes cebadores:

45 Aleatorización en la posición 94:
Cebador aTrCBHI.2:
5'-GCGGTAACAGCCTCTCCATTNNSTTTGTACAC-3' (SEC ID N.º: 32)
Cebador aTrCBHI.2a:
5'-CTGCGCAGACTGGGTGACAAASNNAATGGAGAG-3' (SEC ID N.º: 33)
50 Aleatorización en la posición 196:
Cebador aTrCBHI.3:
5'-CCATCTCCGAGGCTCTTACCNNSCACCCTTGC-3' (SEC ID N.º: 34)
Cebador aTrCBHI.3a:
5'-GGCCGACAGTCGTGCAAGGGTGSNNGGTAAGAG-3' (SEC ID N.º: 35)
55 Aleatorización en la posición 227:
Cebador aTrCBHIR.1:
5'-GAGGGCTGGGAGCCGTCANNSAACAACGCG-3' (SEC ID N.º: 36)
Cebador aTrCBHI.1bA:
5'-CCAATGCCCGTGTTCGCGTTGTTSNNTGACGGC-3' (SEC ID N.º: 37)

60 [0405] Cada reacción de amplificación (50 µl) fue compuesta por tampón 1X PCR, 0,2 mM de dNTPs, 3,2 pM de cebador de sentido, 3,2 pM de cebador antisentido, aproximadamente 100 ng de pTr0221 y 2,5 unidades de Taq ADN polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 programado durante 1 ciclo a 95° C durante 3 minutos seguidos de 30 ciclos cada uno a 95° C durante 30 segundos, 55° C durante 60 segundos, y 72° C durante 90 segundos (extensión final de 5 minutos). Los productos de PCR luego fueron purificados usando un equipo de PCR QIAquick según las instrucciones del fabricante. Cada fragmento de PCR generado para las mutaciones de sitio

individuales sirvieron como megacebadores para la reacción de amplificación de extensión de superposición para generar un fragmento de gen completo de celobiohidrolasa I Cel7A, con mutaciones específicas de sitio indicadas, para generación de biblioteca Los primeros 5 ciclos de la reacción de amplificación fueron compuestos por tampón 1X PCR, 0,2 mM de dNTPs, 100 ng de megacebador de sentido (5'), 100 ng de megacebador antisentido (3') y 2,5 unidades de Taq ADN polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 programadas durante 1 ciclo a 95° C durante 3 minutos seguidos de 5 ciclos cada uno a 95° C durante 30 segundos, 55° C durante 60 segundos y 72° C durante 90 segundos. A esta reacción, se agregaron 50 pmol de cebador aGal776.1 y aGal776.1A para amplificar el fragmento aleatorizado de sitio recién generado en la reacción de 5 ciclos precedente. Las reacciones después de la adición de los nuevos cebadores fueron incubadas en un Eppendorf Mastercycler 5333 programado durante 30 ciclos cada uno a 95° C durante 30 segundos, 55° C durante 60 segundos y 72° C durante 90 segundos (extensión final de 5 minutos). Los fragmentos de la PCR resultantes luego fueron purificados usando un equipo de PCR QIAquick según las instrucciones del fabricante. El fragmento de PCR final luego fue transformado directamente en la levadura junto con el plásmido pJC106 escindido con *Bam*HI y *Xho*I (Ejemplo 6).

[0406] Ya que la Taq ADN polimerasa usada tiene baja actividad de corrección, el proceso de mutagénesis de saturación puede haber introducido mutaciones nuevas en la secuencia codificante además de los cambios diseñados en las posiciones 94, 196 y 227. Las bibliotecas que contienen aminoácidos aleatorizados específicos de sitio fueron seleccionadas como se describe en el ejemplo 10. Como se muestra en la tabla 2, varias sustituciones que mejoraron el índice de actividad térmica a 63° C/50° C en relación con el tipo salvaje fueron descubiertos además de las sustituciones identificadas mediante selección de bibliotecas mutagenizadas aleatoriamente. Por ejemplo, la variante 776-M3 se obtuvo de la selección de una biblioteca primaria, donde una prolina en la posición 227 fue convertida a una alanina. En la selección de la biblioteca aleatorizada específica de sitio en esta posición, se identificaron sustituciones a leucina y a glicina al igual que alanina.

[00407] Además de descubrir las sustituciones de aminoácidos en las posiciones 227, 94, y 196, que confieren termoestabilidad y actividad mejoradas, se identificó una segunda mutación de sitio que podría mejorar la actividad térmica. Específicamente, la sustitución D259N fue identificada en la variante 776-M273 como se muestra en tabla 2.

Tabla 2, variantes de Cel7A con termoestabilidad mejorada que contiene sustituciones en las posiciones 227, 94 y 196. "Relación de mejora" indica la mejora relativa en el índice de actividad térmica, medida a 63° C/50° C.

Variante de Cel7A	Sustitución aminoácidos*	de	Relación de mejora	Tipo de librería
Tipo salvaje	ninguna		1	ninguna
776-M3	P227A, C486W		2,67	primaria
776-M259	P227L		1,50	Aleatorización específica de sitio en 227
776-M273	P227G, D259N		3,00	Aleatorización específica de sitio en 227
776-M274	P227A		2,67	Aleatorización específica de sitio en 227
776-M275	P227L		2,00	Aleatorización específica de sitio en 227
776-M52	G94S, G205R		2,00	primaria
776-M268	G94A, T226A		2,67	Aleatorización específica de sitio 94
776-M264	G94R		2,07	Aleatorización específica de sitio en 94
776-M266	G94Q		2,17	Aleatorización específica de sitio en 94
776-M269	G94A		2,03	Aleatorización específica de sitio en 94
776-M53	S196P, G205R		1,50	primaria
776-M4	S113N, S196T, T462A		n.a.	primaria
776-M261	S196P		1,33	Aleatorización específica de sitio en 196
776-M263	T41I, S196F		1,67	Aleatorización específica de sitio en 196

*La numeración utiliza el primer residuo de la enzima Cel7A madura como posición 1.

Ejemplo 15: Construcción de los vectores de expresión pMJ04, pMJ06 y pMJ09

[0408] El vector de expresión pMJ04 fue construido por PCR amplificando el terminador del gen de celobiohidrolasa 1 (cbh1) de *Trichoderma reesei* de ADN genómico de *Trichoderma reesei* RutC30 usando cebadores 993429 (antisentido)

y 993428 (sentido) mostrados más abajo. El cebador de antisentido fue creado genéticamente para que tuviera un sitio *Pac I* en el extremo 5' y un sitio *Spe I* en el extremo 3' del cebador de sentido.

Cebador 993429 (antisentido): 5'-AACGTTAATTAAGGAATCGTTTTGTGTTT-3' (SEC ID N.º: 38)
 Cebador 993428 (sentido): 5'-AGTACTAGTAGCTCCGTGGCGAAAGCCTG-3' (SEC ID N.º: 39)

[0409] El ADN genómico de *Trichoderma reesei* RutC30 fue aislado usando un equipo DNeasy Plant Maxi (QIAGEN, Valencia, CA). Las reacciones de amplificación (50 µl) fueron compuestas por tampón de reacción 1X ThermoPol (New England Biolabs, Beverly, MA), 0,3 mM de dNTPs, 100 ng de ADN genómico de *Trichoderma reesei* RutC30, 0,3 µM de cebador 993429, 0,3 µM de cebador 993428 y 2 unidades de Vent polimerasa (New England Biolabs, Beverly, MA). Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 programado de la siguiente manera: 5 ciclos cada uno durante 30 segundos a 94° C, 30 segundos a 50° C, y 60 segundos a 72° C, seguidos de 25 ciclos cada uno durante 30 segundos a 94 ° C, 30 segundos a 65° C y 120 segundos a 72° C (final extensión de 5 minutos). Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 229 pares de bases fue cortada del gel y purificada usando un equipo de extracción de gel QIAquick (QIAGEN, Valencia, CA) según las instrucciones del fabricante.

[0410] El fragmento de PCR resultante fue digerido con *PacI* y *SpeI* y fue ligado en pAllLo1 (Ejemplo 8) digerido con las mismas enzimas de restricción usando un equipo Rapid Ligation Kit (Roche, Indianapolis, IN), para generar pMJ04 (Figura 8).

[0411] El vector de expresión pMJ06 fue construido por PCR amplificando el terminador del gen de celobiohidrolasa 1 (*cbh1*) *Cel7A* de *Trichoderma reesei* de ADN genómico de *Trichoderma reesei* RutC30 usando cebadores 993696 (antisentido) y 993695 (sentido) mostrados más abajo. El cebador de antisentido fue creado genéticamente para que tuviera un sitio *Sal* en el extremo 5' del cebador de sentido y un sitio *Nco I* en el extremo 5' del cebador de antisentido.

Cebador 993695 (sentido):
 5'- ACTAGTCGACCGAATGTAGGATTGTT-3' (SEC ID N.º: 40)
 Cebador 993696 (antisentido):
 5'- TGACCATGGTGCGCAGTCC-3' (SEC ID N.º: 41)

[0412] Las reacciones de amplificación (50 µl) fueron compuestas por tampón de reacción 1X ThermoPol, 0,3 mM de dNTPs, 100 ng de ADN genómico de *Trichoderma reesei* RutC30 (que se preparó usando un equipo DNeasy Plant Maxi), 0,3 µM de cebador 993696, 0,3 µM de cebador 993695 y 2 unidades de Vent polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 programado de la siguiente manera: 30 ciclos cada uno durante 30 segundos a 94° C, 30 segundos a 55° C y 60 segundos a 72° C (extensión final de 15 minutos). Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 988 pares de bases fue cortada del gel y purificada usando un equipo de extracción de gel QIAquick según las instrucciones del fabricante.

[0413] El fragmento de PCR resultante fue digerido con *Nco I* y *Sal I* y fue ligado en pMJ04 digerido con las mismas enzimas de restricción usando un equipo Rapid Ligation Kit para generar pMJ06 (Figura 9).

[0414] El vector de expresión pMJ09 fue construido por PCR amplificando el terminador del gen de celobiohidrolasa 1 (*cbh1*) *Cel7A* de *Trichoderma reesei* de ADN genómico de *Trichoderma reesei* RutC30 usando cebadores 993843 (antisentido) y 993844 (sentido) mostrados más abajo. El cebador de antisentido fue creado genéticamente para que tuviera un sitio *Pac I* y un sitio *Spe I* en el extremo 5' y un sitio *Pvu I* en el extremo 5' del cebador de sentido.

Cebador 993844 (sentido):
 5'-CGATCGTCTCCCTATGGGTCATTACC-3' (SEC ID N.º: 42)
 Cebador 993843 (antisentido):
 5'-ACTAGTTAATTAAGCTCCGTGGCGAAAG-3' (SEC ID N.º: 43)

[0415] Las reacciones de amplificación (50 µl) fueron compuestas por tampón de reacción 1X ThermoPol, 0,3 mM de dNTPs, 100 ng de ADN genómico de *Trichoderma reesei* RutC30 (que se preparó usando un equipo DNeasy Plant Maxi), 0,3 µM de cebador 993844, 0,3 µM de cebador 993843 y 2 unidades de Vent polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 programado de la siguiente manera: 30 ciclos cada uno durante 30 segundos a 94° C, 30 segundos a 55° C y 60 segundos a 72° C (extensión final de 15 minutos). Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 473 pares de bases fue cortada del gel y purificada usando un equipo de extracción de gel QIAquick según las instrucciones del fabricante.

[0416] El fragmento de PCR resultante fue digerido con *Pvu I* y *Spe I* y fue ligado en pMJ06 digerido con *IPac I* y *Spe I* usando un equipo Rapid Ligation Kit para generar pMJ09 (Figura 10).

Ejemplo 16: Construcción de pCW045 para la expresión de la variante 776-M57 en *Trichoderma reesei*

[0417] La variante 776-M57 fue subclonada en el vector de expresión de *Trichoderma reesei* pMJ09 usando un equipo de clonación de PCR In-Fusion (BD Biosciences, Clontech, Palo Alto, CA). La amplificación de PCR de las variantes fue realizada usando ADN polimerasa Platinum Pfx (Invitrogen, Carlsbad, CA). Las amplificaciones estuvieron compuestas por tampón de amplificación 1X Pfx, 0.3 mM cada una de dCTP, dATP, dTTP y dGTP; 1 mM de MgSO₄, 1 pmol de cebador IF-F1, 1 pmol de cebador IF-R1, 50-100 ng de DNA modelo 776-57 y 1 unidad de polimerasa Pfx. Las reacciones fueron incubadas usando un termociclador MJ Research programado para un ciclo a 95° C durante 3 minutos, seguido de 30 ciclos cada uno a 95° C durante 15 segundos, 55° C durante 30 segundos y 68° C durante 2 minutos 30 segundos. Las reacciones fueron luego incubadas a 72° C para una extensión de 7 minutos.

Cebador IF-F1:

5'-CGCGGACTGCGCACCATGTATCGGAAGTTG-3' (SEC ID N.º: 44)

Cebador IF-R1:

5'-CGCCACGGAGCTTAATTACAGGCACTGAGA-3' (SEC ID N.º: 45)

[0418] Una alícuota de cada producto de PCR fue realizada en un gel de agarosa al 0,7% usando tampón TAE, tal y como se describe anteriormente, generando bandas previstas de aproximadamente 1.575 pares de bases. Las reacciones de PCR fueron purificadas usando un equipo de purificación de PCR MiniElute y con eluido del ADN en 50 µl de tampón EB. Se estimó que el rendimiento de cada producto de PCR fue 125 ng por microlitros por visualización en un gel de agarosa al 0,07% usando tampón de TAE.

[0419] El plásmido pMJ09 digerido con *Pac I* y despuntado en el sitio *Nco I* fue purificado usando una columna QIAquick Minielute (QIAGEN, Inc., Valencia, CA). El plásmido tuvo una concentración de 100 ng por microlitro. La concentración fue verificada por visualización en un gel de agarosa al 0,07% usando tampón de TAE. La clonación del producto de PCR 776-M57 anteriormente descrito y el vector pMJ09 digerido fue realizada usando un equipo de clonación de PCR In-Fusion (BD Biosciences, Clontech, Palo Alto, CA). El vector de expresión *Trichoderma reesei* con la variante 776-M57 fue designado pCW045 (Figura 11).

Ejemplo 17: Expresión de celobiohidrolasa I Cel7A en *Trichoderma reesei* como la única celobiohidrolasa I

[0420] Para evaluar la variante 776-M57 como la única celobiohidrolasa I después de la expresión en el huésped nativo, una cepa de *Trichoderma reesei* fue construida caracterizada por el hecho de que el gen de celobiohidrolasa I Cel7A, *cbh1*, fue interrumpido. Una cassette de interrupción fue construida usando el gen de resistencia de higromicina (*hph*) de *Escherichia coli* como un marcador seleccionable. El marcador de *hph* fue flanqueado por secuencia homóloga de *cbh1* para concentrarse en el gen *cbh1* nativo.

[0421] La amplificación del gen *cbh1* genómico fue hecha usando el método de reacción en cadena de polimerasa con PWO polimerasa (Roche Applied Science, Mannheim, Alemania). Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333. Las amplificaciones fueron compuestas por tampón de amplificación 1X PWO, 0,3 mM cada uno de dCTP, dATP, dTTP y dGTP, 1 mM de MgSO₄, 1 pmol cada uno de cebador *cbh1* N-término y *cbh1* C-término descritos más abajo, y 50-100 ng de molde de ADN. Las amplificaciones han utilizado una desnaturalización inicial de 2 minutos a 95° C seguidas de 35 ciclos de una desnaturalización de 1 minuto, fijado de 2 minutos a 55° C y una extensión de 2 minutos a 68° C.

Cebador *cbh1* N-término:

5'-GCC TTC GGC CTT TGG GTG TA-3' (SEC ID N.º: 46)

Cebador *cbh1* C-término:

5'-GAG CGG CGA TTC TAC GGG TT-3' (SEC ID N.º: 47)

[0422] El fragmento de PCR de 2,3 kb fue clonado usando el equipo de clonación ZeroBlunt TOPO PCR (Invitrogen Corporation, Carlsbad, CA) y propagado en las células *E. coli* Top10 (Invitrogen Corporation, Carlsbad, CA). El constructo resultante fue un plásmido de 5,8 kb designado pSTM01 (Figura 12).

[0423] El gen de *hph* fue amplificado del plásmido del PHT1 (Cummings, *et al.*, 1999, Curr. Genet. 36: 371-382) usando el método de reacción en cadena de polimerasa con PWO polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 y fueron compuestas por tampón de amplificación 1X PWO, 0,3 mM cada uno de dCTP, dATP, dTTP y dGTP, 1 mM de MgSO₄, 1 pmol cada uno de cebador *hph* N-término y *hph* C-término descritos más abajo, y 50-100 ng de molde de ADN. Las amplificaciones utilizaron una desnaturalización inicial de 2 minutos a 95° C seguidas de 35 ciclos de una desnaturalización de 1 minuto, fijado de 2 minutos a 55° C y una extensión de 2 minutos a 68° C.

Cebador *hph* N-término:

5'-GCC GCG GCA CGC GCC ACA CGG AAA AT-3' (SEC ID N.º: 48)

Cebador *hph* C-término

5'-GAC CGG TCG CAA AAT GAC AAA TAG AAG-3' (SEC ID N.º: 49)

5 [0424] Para generar una cassette de interrupción de *cbh1*, pSTM01 fue digerido con *Mlu* I y *BstE* II para crear un espacio de 1,1 kb en la secuencia codificadora de *cbh1*. Luego se utilizó T4 ADN polimerasa (New England Biolabs, Beverly, MA) para llenar y generar extremos romos en el plásmido digerido. La reacción de T4 fue incubada durante 20 minutos a 15° C seguida de una inactivación por calor durante 15 minutos a 65° C. La mezcla de reacción entera fue fraccionada en un gel de agarosa al 1% usando tampón de TAE, y la banda mayor fue cortada y eluida del gel que usando un equipo de extracción de gel Qiaquick. El gen de *hph* de 1,7 kb luego fue insertado en el vector pSTM01 escindido usando un Equipo de ligamiento de ADN rápido y transformado en células competentes One Shot *E. coli* (Invitrogen Corporation, Carlsbad, CA). El plásmido resultante fue designado pSMK03 (Figura 13).

10 [0425] Usando pSMK03 como un modelo, el ADN *cbh1* fue amplificado usando cebadores de PCR flanqueando el gen de *hph*. Un fragmento de 2,7 kb fue amplificado de pSMK03 usando los siguientes cebadores:

15 Cebador pSMK03-F1:

5'-GCTCCGGGCAAATGCAAA GTG TG-3' (SEC ID N.º: 50)

Cebador pSMK03-R1:

20 5'-AGCAGGCCGCATCTCCAGTGAAAG-3' (SEC ID N.º: 51)

[0426] La amplificación fue realizada usando el método de reacción en cadena de polimerasa con PWO polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333, y fueron compuestas por tampón de amplificación 1X PWO, 0,3 mM cada uno de dCTP, dATP, dTTP y dGTP, 1 mM de MgSO₄, 1 pmol cada uno de cebador KO N-término y KO C-término anteriormente descrito, y 50-100 ng de molde de ADN. Las amplificaciones han utilizado una desnaturalización inicial de 2 minutos a 95° C seguidas de 35 ciclos de una desnaturalización de 1 minuto, fijado de 2 minutos a 55° C y una extensión de 2 minutos a 68° C. El fragmento de ADN resultante ha contuvo 639 kb del secuencia del promotor/gen de *cbh1*, los 1,7 kb de la región de codificación de *hph*, y 659 bases de la secuencia del gen/terminador de *cbh1*.

30 [0427] Los protoplastos de *Trichoderma reesei* *RutC30* fueron preparados para transformación por el siguiente método. Los frascos de agitación con 25 ml de medio YPD fueron inoculados con 5X10⁷ conidia. Después de una incubación durante toda la noche (aproximadamente 18 horas) a 34° C (150 r.p.m.), los micelios fueron recogidos por filtración a través de Miracloth™ estéril (Calbiochem, San Diego, CA) y transferidos a 20 mg/ml Glucanex (Novozymes A/S, Bagsvaerd, Dinamarca) y 1,6 mg/ml de quitinasa (Sigma-Aldrich, St. Luis, MO) en 25 ml de 1 M de sorbitol. Las digestiones fueron típicamente 25-40 minutos con agitación suave (80- 100 r.p.m.) a 34° C. Los protoplastos luego fueron filtrados a través de un filtro de gasa y lavados con 1 M de sorbitol helado y luego centrifugados a 400 x g durante 8 minutos. Los protoplastos fueron lavados dos veces con 25 ml de 1 M de sorbitol y dos veces con 25 ml de 1 M de sorbitol, 50 mM de Tris-HCl pH 8,0, 50 mM de CaCl₂. Los protoplastos luego fueron contados usando un haemacitómetro y resuspendidos en una solución compuesta por 7 partes de 1 M de sorbitol, 50 mM de Tris-HCl pH 8,0, 50 mM de CaCl₂ y 2 partes de PEG-4000 al 50%, 50 mM de Tris-HCl pH 8,0, 50 mM de CaCl₂ a una concentración de 1 X 10⁸ protoplastos/ml. Los protoplastos fueron usados inmediatamente o almacenados a -80° C en un congelador de índice controlado (Nalge Nunc international, Rochester, NY) hasta la transformación.

45 [0428] La transformación de los protoplastos fue realizada usando 5 µg de ADN lineal y 5 µl de heparina (5 mg/ml), que fueron mezclados e incubados en hielo durante 5 minutos. Cien microlitros de protoplastos luego fueron adicionados y la mezcla fue incubada durante 15 minutos en hielo. Quinientos microlitros de PEG-4000 al 50%, 50 mM de Tris-HCl pH 8,0, 50 mM de CaCl₂ luego fue adicionado e incubado a temperatura ambiente durante 15 minutos. Un mililitro de de PEG-4000 al 50%, 50 mM de Tris- HCl pH 8,0, 50 mM de CaCl₂ se añadió a la mezcla y se incubó a 34° C durante 20 minutos. Después de la incubación a 34° C, se añadieron 2 ml de 1 M de sorbitol, 50 mM de Tris-HCl pH 8,0, 50 mM de CaCl₂. El contenido fue suavemente mezclado antes de extender 350 µl de la mezcla de protoplasto en placas de PDA. Las placas fueron incubadas a 29° C durante aproximadamente 18 horas antes de cubrir con 100 µg/ml de higromicina B (Roche Applied Science, Manheim, Alemania) en el medio PDA: Los transformantes aparecidos dentro de 3 días de la cubierta de higromicina. Los transformantes luego fueron subcultivados a placas de PDA con 100 µg/ml de higromicina B. Los transformantes que crecieron en las placas secundarias luego fueron usados para inoculación en la medio de inducción de celulasa (CIM).

55 [0429] Las cepas fueron seleccionadas para interrupción de celobiohidrolasa I por el siguiente método. Los transformantes fueron cultivados en el medio de inducción de celulasa en placas de cultivo de 24 pocillos (Coming, Acton, MA) durante 3 días a 34° C. Se añadieron diez microlitros de caldo de cultivo a 10 µl de tampón de muestra 2X más beta-mercaptoetanol al 1% y ejecutadas en un gel de acrilamida de tris-glicina al 8-16% (NUPAGE Novex Gels Invitrogen Corporation, Carlsbad, CA) usando tampón de migración de SDS NuPAGE® MES (NUPAGE Novex Gels Invitrogen Corporation, Carlsbad, CA). Inactivaciones potenciales fueron seleccionadas basadas en ausencia de una banda de proteína celobiohidrolasa Cel7A en comparación con la cepa progenitora

[0430] Para confirmar la eliminación del gen *cbh1* en cepas candidatas que fueron seleccionadas por análisis SDS-PAGE, se realizó el análisis Southern para evaluar si el gen *cbh1* había sido interrumpido. Los micelios fueron cultivados durante toda la noche en frascos de agitación con 25 ml de medio YPD. Los micelios fueron luego cosechados, filtrados y molido usando un mortero y mano de mortero con nitrógeno líquido. El ADN genómico fue aislado usando un equipo DNeasy Plant Maxi. Un microgramo de ADN fue digerido con *Hind* III o *Nde* I y ejecutado en un gel de agarosa al 0,8% gel de agarosa en tampón de TAE. El ADN fue fragmentado en el gel por tratamiento con 0,25 M de HCl, desnaturalizado con 0.5 M de NaOH, 1,5 M de NaCl, y neutralizado con 1 M de Tris, pH 8,0, 1,5 M de NaCl para transferencia posterior en 10X SSC para membrana Nitran Plus (Schleicher y Schuell BioScience, GmbH, Dassel, Alemania). El ADN fue entrecruzado en UV a la membrana y prehibridizado durante 1 hora a 60° C en 20 ml de DIG Easy Hyb (Roche Applied Science, Mannheim, Alemania).

[0431] Dos sondas fueron preparadas para análisis de Southern. Una sonda de higromicina de 1,7 kb fue amplificada de pSMK03, usando cadena de reacción de polimerasa y los siguientes cebadores:

Cebador cpht-p:
5'-GCACGCGCCACACGGAAAAT-3' (SEC ID N.º: 52)
Cebador cpht-t:
5'-CGCAAATGACAAATAGAAG-3' (SEC ID N.º: 53)

[0432] Además, una sonda de celobiohidrolasa I de 2,1 kb fue amplificada de pSTM01 por PCR usando los cebadores pSMK03-F1 y pSMK03-R1. Las sondas fueron preparadas con un equipo de síntesis de sonda PCR DIG (Roche Applied Science, Mannheim, Alemania). La mezcla de síntesis de sonda PCR DIG y tampón PCR con magnesio fueron usados en 1X. Polimerasa Expand High Fidelity (Roche Applied Science, Mannheim, Alemania) fue añadida a 0,75 µl por reacción. Se usaron un pmol cada uno de los cebadores y 50-100 ng de molde de ADN por reacción. Antes de su uso, la sonda fue desnaturalizada por de ebullición durante 5 minutos y luego añadida al tampón de hibridación.

[433] La sonda desnaturalizada fue añadida directamente al tampón DIG Easy Hyb y la hibridación fue durante toda la noche a 65° C. Después de los lavados de post hibridación (dos veces en 2X SSC, una vez en 0,4X SSC, 60° C, 10 minutos cada uno), se realizó detección quimioluminiscente usando el sistema de detección DIG y CPD-Star (Roche Applied Science, Mannheim, Alemania). El Equipo de marcador de peso molecular de ADN III etiquetado con DIG (Roche Applied Science, Mannheim, Alemania) fue usado como la fuente de marcadores estándar.

[0434] El análisis Southern indicó la presencia de una única integración de *hph* en el gen *cbh1* para un transformante específico. Esta cepa de delección génica de celobiohidrolasa I fue designada *Trichoderma reesei* SaMe013.

[0435] El plásmido pCW045, con la variante de celobiohidrolasa I 776-M57 detrás del promotor *cbh1*, fue introducido en *Trichoderma reesei* SaMe013 por transformación mediada por PEG (Penttila *et al.*, 1987, Gene 61: 155-164). El plásmido contiene el gen *Aspergillus nidulans amdS* para permitir que los transformantes crezcan en acetamida como la única fuente de nitrógeno. *Trichoderma reesei* RutC30 fue cultivada a 27° C y 90 r.p.m. en 25 ml de medio YPD suplementado con 10 mM de uridina durante 17 horas. Los micelios fueron recogidos por filtración usando el Sistema de filtración desechable impulsado por vacío de Millipore (Millipore, Bedford, MA) y lavados dos veces con agua desionizada y dos veces con 1,2 M de sorbitol. Los protoplastos fueron generados suspendiendo los micelios lavados en 20 ml de 1,2 M de sorbitol con 15 mg de Glucanex por ml y 0,36 unidades de quitinasa por ml e incubando durante 15-25 minutos a 34° C con agitación suave a 90 r.p.m. Los protoplastos fueron recogidos por centrifugado durante 7 minutos a 400 x g y lavados dos veces con 1,2 M de sorbitol helado. Los protoplastos fueron contados usando un hemacitómetro y resuspendidos en el STC a una concentración final de 1 X 10⁸ protoplastos por ml. Los protoplastos de exceso fueron almacenados en un recipiente refrigerante Cryo 1° C (Nalgene, Rochester, NY) a -80° C.

[0436] Aproximadamente 1 µg de pCW045 digerido con *Pme* I se añadió a 100 µl de solución de protoplasto y se mezcló suavemente, seguido de 260 µl de tampón PEG, se mezcló e incubó a temperatura ambiente durante 30 minutos. Luego STC (3 ml) fue añadido y mezclado y la solución de transformación fue colocada en placas COVE usando selección *Aspergillus nidulans amdS*. Las placas fueron incubadas a 28° C durante 5-7 días. Los transformantes fueron sub-cultivados sobre placas de uridina COVE2 plus y cultivados a 28° C durante 7 días.

[0437] Los transformantes de *Trichoderma reesei* fueron cultivados en frascos de agitación de 125 ml disipados con 25 ml de medios inductores de celulasa a pH 6,0 inoculados con esporas de los transformantes e incubada a 28° C y 200 r.p.m. durante 7 días. *Trichoderma reesei* RutC30 fue ejecutada como un control. Las muestras de caldo de cultivo fueron quitadas el día 7. SDS-PAGE se efectuó usando el geles de Criterion Tris-HCl (poliacrilamida al 8-16%) (BioRad, Hercules, CA) con el Sistema Criterion SDS-PAGE (BioRad, Hercules, CA). Cinco µl de sobrenadantes de día 7 fueron suspendidos en 5 µl de tampón de muestra 2X (BioRad, Hercules, CA) y calentados en presencia de beta-mercaptoetanol al 1% durante 5 minutos. Las muestras sobrenadantes fueron cargadas sobre una gel de poliacrilamida y sometidas a electroforesis con 1X tris/Glicina/SDS como tampón de desplazamiento (BioRad, Hercules, CA). El gel resultante fue manchado con Bio-Safe Coomassie Stain (BioRad, Hercules, CA). Los transformantes candidatos fueron evaluados en cuanto a sus capacidad para expresar la celobiohidrolasa I Cel7A variante 776-M57 en niveles que fueron aparentemente equivalentes a niveles observados en la cepa progenitora de *Trichoderma reesei*. Un único transformante fue seleccionado sobre esta base para producción en mayor escala en la fermentación.

[0438] La fermentación fue realizada usando la cepa que expresa la variante de celobiohidrolasa I 776-M57, y la *Trichoderma reesei* RutC30 (cepa huésped) se ejecutó como un control. Las esporas de *Trichoderma reesei* RutC30 fueron inoculadas en matraces de agitación de 500 ml, con 100 ml de Medio de inóculo de *Trichoderma*. Los matraces fueron colocados en un agitador orbital a 28° C por aproximadamente 48 horas tiempo en el cual 50 ml del cultivo fue inoculado en 1,8 litros de medio de fermentación de *Trichoderma* en un vaso de fermentación de 2 litros. Las fermentaciones fueron ejecutadas a pH 5, 0,28° C, con oxígeno disuelto mínimo a 25% a una corriente de aire 10 WM y una agitación de 1100. El medio de alimentación de *Trichoderma* fue administrado en el vaso de fermentación a 18 horas con un índice de alimentación de 3,6 g/hora durante 33 horas y luego 7,2 g/hora. Las fermentaciones se ejecutaron durante 165 horas tiempo en el que los caldos de fermentación final fueron centrifugados y los sobrenadantes fueron almacenados a -20° C.

Ejemplo 18: Identificación de un gen de la familia GH3A de glicosil hidrolasa en la secuencia genómica de *Aspergillus fumigatus*

[0439] Una búsqueda de TblastN (Altschul *et al.*, 1997, Nucleic Acids Res. 25: 3389-3402) de la secuencia de genoma parcial de *Aspergillus fumigatus* (The Institute for Genomic Research, Rockville, MD) se efectuó usando como secuencia de búsqueda una secuencia de proteína de beta-glucosidasa de *Aspergillus aculeatus* (N.º de muestreo P48825). Diferentes genes fueron identificados como homólogos putativos de familia GH3A basado en un alto grado de similitud con la secuencia de búsqueda en el nivel de aminoácidos. Una región genómica de aproximadamente 3000 pares de bases con identidad mayor que 70% a la secuencia de búsqueda en el nivel de aminoácidos fue elegida para estudio posterior.

Ejemplo 19: Extracción de ADN genómico de *Aspergillus fumigatus*

[0440] *Aspergillus fumigatus* fue cultivado en 250 ml de medio de dextrosa de patata en un matraz oscilante disipado a 37° C y 240 r.p.m. Las micelas fueron recogidas por filtración, lavadas dos veces en tampón TE (10 mM de Tris-1 mM de EDTA), y congeladas bajo nitrógeno líquido. Los micelios congelados fueron molidos por mortero y mano de mortero a un polvo fino, que fue resuspendido en tampón de pH 8,0 con 10 mM de Tris, 100 mM de EDTA, Tritón X-100 al 1%, 0,5 M de HCl de guanidina y 200 mM de NaCl. Se agregó RNasa libre de ADNsa a una concentración de 20 mg/litro y el lisado fue incubado a 37° C durante 30 minutos. El detrito celular fue eliminado por centrifugado, y el ADN fue aislado usando una columna Qiagen Maxi 500 (QIAGEN Inc., Valencia, CA). Las columnas fueron equilibradas en 10 ml de QBT lavadas con 30 ml de QC, y eluidas con 15 ml de QF (todos los tampones de QIAGEN Inc., Valencia, CA). El ADN fue precipitado en isopropanol, lavado en etanol al 70% y recuperado por centrifugado. El ADN fue resuspendido en tampón TE).

Ejemplo 20: Clonación del gen de beta-glucosidasa de la Familia GH3A y construcción de un vector de expresión de *Aspergillus oryzae*

[0441] Dos cebadores oligonucleótidos sintéticos mostrados más abajo fueron diseñados para amplificar por PCR un gen de *Aspergillus fumigatus* que codifica una beta-glucosidasa putativa de la familia GH3A del ADN genómico preparado en el Ejemplo 19. Un equipo de clonación InFusion (BD Biosciences, Palo Alto, CA) fue usado para clonar el fragmento directamente en el vector de expresión, pAILo2, (Ejemplo8; Figura 5).

Cebador directo 5'-ACTGGATTTACCAT**GAGATTGGTTGGCTCG**-3' (SEC ID N.º: 54)

Cebador inverso: 5'-AGTCACCTCTAGTT**ACTAGTAGACACGGGGC**-3' (SEC ID n.º: 55)

Las letras en negrita representan la secuencia codificante. La secuencia restante contiene la identidad de secuencia en comparación con los sitios de inserción de pAILo2 (Figura 5).

[0442] Cincuenta picomoles de cada uno de los cebadores mencionados fueron usados en una reacción de PCR con 100 ng de ADN genómico de *Aspergillus fumigatus*, tampón de amplificación 1X Pfx (Invitrogen, Carlsbad, CA), 1,5 µl de 10 mM de mezcla de dATP, dTTP, dGTP y dCTP, 2,5 unidades de polimerasa de ADN Platinum Pfx (Invitrogen, Carlsbad, CA), 1 µl de 50 mM de MgSO₄ y 2,5 µl de solución intensificadora 10X pCRx (Invitrogen, Carlsbad, CA) en un volumen final de 50 µl. Las condiciones de amplificación fueron un ciclo a 94° C durante 2 minutos; y 30 ciclos cada uno a 94° C durante 15 segundos, 55° C durante 30 segundos y 68° C durante 3 minutos. El bloque de calor luego fue conducido a un ciclo de remojo de 4° C.

[0443] Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando tampón de 40 mM de Tris base-20 mM de sodio acetato-1 mM de EDTA disodio (TAE) donde una banda de 3 kb fue cortada del gel y purificada usando un equipo de extracción de gel QIAquick (QIAGEN, Valencia, CA) según las instrucciones del fabricante.

[0444] El fragmento luego fue clonado en el vector de expresión pAILo2 usando un equipo de clonación Infusion El vector fue digerido con las endonucleasas de restricción *Nco I* y *Pac I* (utilizando las condiciones especificadas por el fabricante) El fragmento fue purificado por electrodéresis de gel y purificación de gel QIAquick. El fragmento del gen y el vector digerido fueron ligados en una reacción que dio como resultado el plásmido de expresión pEJG97(Figura 14) en

la cual la transcripción del gen de beta-glucosidasa de la familia GH3A estuvo bajo el control del promotor NA2-tpi. La reacción (50 µl) fue compuesta por tampón de InFusion 1X (BD Biosciences, Palo Alto, CA), 1X BSA (BD Biosciences, Palo Alto, CA), 1 µl de enzima de InFusion (diluido 1:10) (BD Biosciences, Palo Alto, CA), 150 ng de pAILo2 digerido con *NcoI* y *PacI*, y 50 ng del producto de PCR purificado de beta-glucosidasa de *Aspergillus fumigatus*. La reacción fue incubada a temperatura ambiente durante 30 minutos. Un µl de la reacción fue usado para transformar las células de *E. coli* XL10 Solopac Gold (Stratagene, La Jolla, CA). Un transformante de *E. coli* con el plásmido pEJG97 fue detectado por digestión de restricción y el ADN plásmido fue preparado usando un Biorobot 9600 (QIAGEN, Inc., Valencia, CA)

Ejemplo 21: Caracterización de la secuencia genómica de *Aspergillus fumigatus* que codifica una beta-glucosidasa de la familia GH3A

[0445] La secuenciación del ADN del gen de beta-glucosidasa de *Aspergillus fumigatus* de pEJG97 fue realizada con un secuenciador de ADN automatizado Modelo 377 XL de Perkin-Elmer Applied Biosystems (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA) usando una estrategia de química de terminador de coloración (Giesecke *et al.*, 1992, Journal of Virology Methods 38: 47-60) y estrategia de desplazamiento de cebador. Los datos de la secuencia de nucleótidos fueron analizados en detalle en cuanto a la calidad y todas las secuencias de fueron comparadas entre sí con asistencia del software PHRED/PHRAP (University of Washington, Seattle, WA).

[0446] Un modelo de gen para la secuencia de *Aspergillus fumigatus* fue construida según la similitud con genes homólogos de *Aspergillus aculeatus*, *Aspergillus niger* y *Aspergillus kawachii*. La secuencia de nucleótidos (SEC ID n.º: 56) y la secuencia de aminoácidos deducida (SEC ID N.º: 57) se muestran en la Figura 15. El fragmento genómico codifica un polipéptido de 863 aminoácidos, interrumpido por 8 intrones de 62, 55, 58, 63, 58, 63 y 51 pares de bases. El contenido %G+C del gen es 54,3%. Usando el programa de software SignalP (Nielsen *et al.*, 1997, Protein Engineering 10: 1-6), se predijo un péptido señal de 19 residuos. La proteína madura predicha contiene 844 aminoácidos con una masa molecular de 91,7 kDa.

[0447] Una alineación comparativa de las secuencias de beta-glucosidasa fue determinada usando el Método de Clustal W (Higgins, 1989, CABIOS 5: 151-153) usando el Software de MEGALIGN™ LASERGENE™ (DNASTAR, Inc., Madison, WI) con una tabla de identidad y los siguientes parámetros de alineación múltiple: Penalización de espacio de 10 y penalización de longitud de espacio de 10. Los parámetros de alineación en pareja fueron Ktuple=1, penalización de espacio=3, ventanas=5, y diagonales=5. La alineación mostró que la secuencia de aminoácidos deducida del gen de beta-glucosidasa de *Aspergillus fumigatus* compartió 78%, 76% y 76% de identidad de las secuencias de aminoácidos deducidas de beta-glucosidasa de *Aspergillus aculeatus* (número de registro P48825), *Aspergillus niger* (número de registro 000089) y *Aspergillus kawachii* (número de registro P87076).

Ejemplo 22: Expresión del gen de beta-glucosidasa de *Aspergillus fumigatus* de la familia GH3A en *Aspergillus oryzae* JAL250

[0448] Los protoplastos de *Aspergillus oryzae* Jal250 fueron preparados según el método de Christensen *et al.*, 1988, Bio/Technology 6: 1419-1422 . Cinco µg de pEJG97 (así como pAILo2 como un control de vector) fue usado para transformar *Aspergillus oryzae* JAL250.

[0449] La transformación de *Aspergillus oryzae* Jal250 con pEJG97 produjo aproximadamente 100 transformantes. Diez transformantes fueron aislados en placas de PDA individuales.

[0450] La placas de PDA confluentes de cinco de los diez transformantes fueron lavadas con 5 ml de Tween 20 al 0,01% e inoculadas separadamente en 25 ml de medio MDU2BP en frascos de agitación de vidrio de 125 ml e incubadas a 34° C, 250 r.p.m. Cinco días después de la incubación, se analizaron 0,5 µl de sobrenadante de cada cultivo usando geles de SDS-PAGE de Tris-glicina al 8-16% (Invitrogen, Carlsbad, CA) según las instrucciones del fabricante. Los perfiles de SDS-PAGE de los cultivos mostraron que uno de los transformantes (designado transformante 1) tuvo una banda mayoritaria de aproximadamente 130 kDa.

[0451] Una placa confluyente de transformante 1 (cultivada en PDA) fue lavada con 10 ml de Tween 20 al 0,01% e inoculada en un Fernbach de 2 litros con 400 ml de medio MDU2BP para generar caldo para caracterización de la enzima. El matraz fue cosechado en el día 5 y filtrado usando una membrana GP Express plus de 0,22 µm (Millipore, Bedford, MA).

[0452] Un único transformante de pEJG97 en *Aspergillus oryzae* fue cultivado en un tanque de fermentación. Las esporas de *Aspergillus oryzae* fueron inoculadas en matraces de agitación de 500 ml, con 100 ml de Medio de inóculo. Los matraces fueron colocados en un agitador orbital a 34° C a 200 r.p.m. por aproximadamente 24 horas tiempo en el cual 50 ml del cultivo fue inoculado en 1,8 litros de medio de fermentación en un vaso de fermentación de 2 litros. Las fermentaciones fueron ejecutadas a pH 7,0, 34° C, con oxígeno disuelto mínimo a 25% a una corriente de aire 10 WM y una agitación de 1100. El medio de alimentación fue administrado en el vaso de fermentación con una adición de oxígeno disuelto con un índice de alimentación de 4 g/hora. Las fermentaciones se ejecutaron durante 164 horas tiempo en el que los caldos de fermentación final fueron centrifugados y los sobrenadantes fueron almacenados a -20° C.

Ejemplo 23: Caracterización de beta-glucosidasa de *Aspergillus fumigatus*

[0453] Una alícuota de 3 ml de la beta-glucosidasa de *Aspergillus fumigatus*, obtenida como se describe en el Ejemplo 22, fue desalinizada usando una columna de desalinización BioRad Econo-pAc 10DG (BioRad, Hercules, CA), dando como resultado aprox. 4 ml de caldo desalinizado en 100 mM de citrato sódico pH 5,0. El caldo desalinizado fue luego concentrado a aproximadamente 180 μ l usando un Amicon Centricon Plus-20 (membrana PES de corte de 5 kDa Biomax-5) y diluido con 100 mM de citrato sódico pH 5,0 a un volumen final de aproximadamente 500 μ l. El ensayo de BCA (Smith *et al.*, 1985, Anal. Biochem. 150: 76-85) del caldo desalinizado concentrado mostró una concentración de 1,00 mg de proteína por ml. Una segunda alícuota también fue desalinizada para obtener más material, que produjo resultados similares.

[0454] SDS-PAGE (criterio BioRad Tris-HCl al 7,5%) de las muestras desalinizadas concentradas mostró una banda mayor a aprox. 130 kD. La banda mayor a 130 kD fue recortada del gel y sometida a secuenciación de aminoácido N-terminal.

[0455] La beta-glucosidasa de *Aspergillus fumigatus* del caldo desalinizado/concentrado fue evaluado en relación a la termoestabilidad a 50°, 65°, y 70° C. Las condiciones de ensayo a 50° y 65° C fueron: 100 mM de citrato sódico pH 5,0, Tween-20 al 0,01%, 4 mM de P-nitrofenil-beta-D-glucopiranosido, [proteína]_{AfumGH3A} = $6,9 \times 10^{-6}$ mg/ml, incubado a 50° y 65° C. Las partes alícuotas fueron tomadas a 0,5, 1, 2, 3, 3,75, y 24 horas. A cada alícuota se añadió 1 M de carbonato de sodio pH 10,0, y la concentración de aniones de p- nitrofenil se determinó partir de la absorbancia a 405 nm. A 70° C, las condiciones de ensayo fueron: 100 mM de citrato sódico (pH 5,0), Tween-20 al 0,01%, 4 mM de p-nitrofenil-beta-D-glucopiranosido, [proteína] = $5,6 \times 10^{-6}$ mg/ml. Las partes alícuotas fueron tomadas a 0,25, 0,5, 1, 2, 4, y 5 horas. A cada alícuota se añadió 1 M de carbonato de sodio pH 10,0, y la concentración de aniones de p- nitrofenil se determinó partir de la absorbancia a 405 nm. Obsérvese que cada una de las concentraciones de proteína anteriores se refiere a la concentración total de proteína en el ensayo, ya que todas fueron evaluadas como caldos en lugar de enzimas purificadas.

[0456] Los resultados se muestran en las Figuras 16 y 17. La termoestabilidad fue definida como cinética de reacción lineal para un intervalo de tiempo dado, dentro de una conversión de porcentaje razonable (<15%) a anión de p-nitrofenilo. La beta-glucosidasa de *Aspergillus fumigatus* pareció ser estable al menos 24 horas a 50° C, aunque no había ningún punto temporal entre 4 horas y 24 horas. A 65 °C pareció ser estable al menos 3,75 horas, pero después la estabilidad gradualmente disminuyó con una conversión del 65% a anión de p-nitrofenilo a las 24 horas. Esta conversión fue bastante alta, de modo que cierta parte de la reducción observada en el índice puede haberse debido a depleción de sustrato y/o inhibición de producto además de posible inactivación térmica. A 70° C pareció ser estable durante 1 hora, y razonablemente estable a 2 horas.

[0457] El caldo con familia beta-glucosidasa GH3A de *Aspergillus fumigatus*, obtenido como se describe en el Ejemplo 22, fue primero desalinizado (columna Econo-Pac 10DG de Bio-Rad) y concentrado (Centricon Plus-20, BioBiomax-5, corte 5 kD), a una concentración de 0,92 mg/ml (ensayo BCA). Luego, fue incubado a 0,037 y 0,0092 μ g/ml de proteína total con 10 mM de celobiosa en 100 mM de citrato sódico pH 5,0 más Tween-20 al 0,01% a 65°C. Las partes alícuotas fueron tomadas a 0,5, 1, 2, 3, 4, 6, y 19 horas. Las partes alícuotas fueron hervidas 6 minutos para terminar la reacción, y la concentración de glucosa se determinó usando el ensayo Trinder (Sigma Chemical Co., St. Louis, MO) y estándares de glucosa externos. Los resultados se muestran en la Figura 18. La beta-glucosidasa pareció mantener 90% de su actividad a 6 horas, y 65% a 19 horas a 65° C a la carga de proteína inferior (0,0092 μ g/ml). La beta-glucosidasa pareció ser razonablemente estable hasta 6 horas a 65° C.

Ejemplo 24: Ensayo de variante de Cel7A 776-M57 en combinación con celulasas de *Trichoderma*

[0458] Se obtuvo forraje de maíz pretratado (PCS) del Laboratorio de Energía Nacional Renovable (NREL) del Departamento de Energía de EE. UU. Los sólidos insolubles en agua en PCS incluyen: 56,5% de celulosa, 4,6% de hemicelulosa y 28,4% de lignina. Las condiciones de pretratamiento fueron: forraje de maíz, 1,4% (p/vol) de ácido sulfúrico, 165° C, 107 p.s.i., durante 8 minutos. Antes de ensayo, el PCS fue lavado con un volumen grande de agua desionizada destilada en un filtro de vidrio. El PCS fue luego molido usando un desfibrador de café para reducir el tamaño de partícula, luego fue lavado con agua en un filtro Millipore de 22 μ m (6P Express Membrane, Stericup, Millipore, Billerica, MA). El PCS lavado fue resuspendido en agua desionizada para hacer una suspensión de 20 mg/ml, y fue almacenado a 4° C.

[0459] Los caldos de fermentación de *Trichoderma reesei* fueron filtrados (membrana de 0,4 μ m, Stericup, Millipore, Billerica, MA), y el contenido de proteína fue evaluado por ensayo BCA (Pierce Biotech., Inc., Rockford, IL). La beta-glucosidasa termoestable de *Aspergillus fumigatus* fue usada en los ensayos de PCS. La beta-glucosidasa de *Aspergillus fumigatus* fue preparada como se describe en el Ejemplo 22. Los caldos de fermentación de la beta-glucosidasa de *Aspergillus fumigatus* heterológicamente expresada fueron filtrados (membrana de 0,4 μ m, Millipore, Billerica, MA), y el contenido de proteína fue evaluado por ensayo de BCA (Pierce, Biotech., Inc., Rockford, IL). Los ensayos de PCS fueron realizados en un volumen de 1,0 ml. Los ensayos incluyeron 10 mg de PCS por ml, 50 mM de tampón de acetato sódico (pH 5,0), 8,5 mg de proteína de caldo de fermentación total por gramo de celulosa y 0,27 mg

de caldo de fermentación de beta-glucosidasa de *Aspergillus fumigatus*. Los ensayos fueron mantenidos a 55° C y a 60° C en tubos de ensayo sellados (ImmunoWare Microtubes, Pierce Biotech., Inc., Rockford, IL) durante 5 días con inversión intermitente de tubos de ensayo.

- 5 [0460] Aproximadamente cada 24 horas, se tomaron puntos temporales de las reacciones de PCS. Partes alícuotas de diez microlitros fueron quitadas de los tubos de ensayo en 90 µl de mezcla de temple alcalina (0,102 M Na₂ CO₃ más 0,058 M NaHCO₃). Se realizó la dilución de las muestras apagadas para un octavo o un decimosexto de la concentración de inicio. La concentración de azúcares reductores en muestras producida por degradación de PCS fue determinada por ensayo de ácido 4-hidroxibenzoico hidrazida (PHBAH) (Lever, 1972, Anal Biochem 47: 273-279)
 10 usando 2 partes de PHBAH al 1,25% en la mezcla de temple alcalina añadida a 3 partes de la solución de ensayo detenida. Esta solución fue luego calentada durante 10 minutos a 95° C, luego se diluyeron muestras en agua, y se midió la absorbancia a 405 nm usando un Lector de placa UltraMark (Bio-Rad, Hercules, CA). La absorbancia a 405 nm fue convertida en equivalentes de glucosa usando una curva de estándar de glucosa. Finalmente, se calculó el grado de conversión de celulosa usando la concentración inicial de celulosa y un factor de aumento de peso en la conversión de celulosa a glucosa. El grado de conversión de la celulosa a azúcar reductora (rendimiento RS, %) fue calculado usando la siguiente ecuación:

$$\begin{aligned} \text{Rendimiento RS (\%)} &= \text{RS}_{(\text{mg/ml})} * 100 * 162 / (5.65_{(\text{mg/ml})} * 180) \\ &= \text{RS}_{(\text{mg/ml})} * 100 / (5.65_{(\text{mg/ml})} * 1.111) \end{aligned}$$

- 20 [0461] La Figura 19 muestra los perfiles temporales de la hidrólisis de PCS por la cepa *Trichoderma reesei* progenitora, y la cepa que expresa la variante 776-M57 en lugar de la cellobiohidrolasa I Cel7A de tipo salvaje. Es evidente que la cepa que contiene la variante supera a la progenitora bajo estas condiciones.

Depósito de material biológico

- 25 [0462] El siguiente material biológico ha sido depositado según las condiciones del Tratado de Budapest con la Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, 1815 University Street, Peoria, Illinois, 61604, y se les ha dado los siguientes números de registro

Depósito	Número de acceso	Fecha de depósito
pAJ052	NRRL B-30683	29 de julio de 2003
776-M1	NRRL B-30657	22 de mayo de 2003
776-M4	NRRL B-30658	22 de mayo de 2003
776-M23	NRRL B-30659	22 de mayo de 2003
776-M26	NRRL B-30661	22 de mayo de 2003
776-M32	NRRL B-30662	22 de mayo de 2003
776-M53	NRRL B-30663	22 de mayo de 2003
776-M57	NRRL B-30664	22 de mayo de 2003
776-M108	NRRL B-30665	22 de mayo de 2003
776-M109	NRRL B-30666	22 de mayo de 2003
776-M21	NRRL B-30674	28 de julio de 2003
776-M22	NRRL B-30675	28 de julio de 2003
776-M41	NRRL B-30676	28 de julio de 2003
776-M42	NRRL B-30677	28 de julio de 2003
776-M52	NRRL B-30678	28 de julio de 2003
776-M71	NRRL B-30679	28 de julio de 2003
776-M73	NRRL B-30680	28 de julio de 2003

		2003
776-M124	NRRL B-30681	28 de julio de
		2003
776-M125	NRRL B-30682	28 de julio de
		2003
776-M273	NRRL B-30762	25 de agosto de
		2004

[0463] Las cepas han sido depositadas bajo condiciones que aseguran que el acceso a los cultivos estará disponible durante la pendency de la presente solicitud de patente a una persona designada por el Comisario de Patentes y Marcas Registradas que tendrá derecho a ello bajo 37 C.F.R. §.14 y 35 U.S.C. §122. Los depósitos representan cultivos substancialmente puros de las cepas depositadas. Los depósitos están disponibles según sea necesario por las leyes de patentes extranjeras en los países donde se presentaron duplicados de la solicitud en cuestión, o su progenie. No obstante, debería entenderse que la disponibilidad de un depósito no constituye una licencia para poner en práctica la invención en cuestión en derogación de los derechos de las patentes concedidas por acción gubernamental.

LISTADO DE SECUENCIAS

[0464]

<110> Novozymes Biotech, Inc.

<120> Variants de Glycoside Hydrolases

<130> 10479.204-WO

<150> 60/497,809

<151> 2003-08-25

<160> 57

<170> PatentIn version 3.2

<210> 1

<211> 1545

<212> DNA

<213> Trichoderma reesei

<220>

<221> CDS

<222> (1) .. (1539)

<220>

<221> sig_peptide

<222> (1) .. (51)

<220>

<221> mat_peptide

<222> (52)..()

<400>

1

ES 2 424 346 T3

atg	tat	egg	aag	ttg	gcc	gtc	atc	tgg	gcc	ttc	ttg	gcc	aca	gct	cgt	48
Met	Tyr	Arg	Lys	Leu	Ala	Val	Ile	Ser	Ala	Phe	Leu	Ala	Thr	Ala	Arg	
		-15					-10					-5				
gct	cag	tgg	gcc	tgc	act	ctc	caa	tgg	gag	act	cac	ccg	cct	ctg	aca	96
Ala	Gln	Ser	Ala	Cys	Thr	Leu	Gln	Ser	Glu	Thr	His	Pro	Pro	Leu	Thr	
-1	1			5				10						15		
tgg	cag	aaa	tgc	tgg	tct	ggc	ggc	acg	tgc	act	caa	cag	aca	ggc	tcc	144
Trp	Gln	Lys	Cys	Ser	Ser	Gly	Gly	Thr	Cys	Thr	Gln	Gln	Thr	Gly	Ser	
			20					25						30		
gtg	gtc	atc	gac	gcc	aac	tgg	cgc	tgg	act	cac	gct	acg	aac	agc	agc	192
Val	Val	Ile	Asp	Ala	Asn	Trp	Arg	Trp	Thr	His	Ala	Thr	Asn	Ser	Ser	
			35					40					45			
acg	aac	tgc	tac	gat	ggc	aac	act	tgg	agc	tgg	acc	cta	tgt	cct	gac	240
Thr	Asn	Cys	Tyr	Asp	Gly	Asn	Thr	Trp	Ser	Ser	Thr	Leu	Cys	Pro	Asp	
		50					55					60				
aac	gag	acc	tgc	gcg	aag	aac	tgc	tgt	ctg	gac	ggc	gcc	gcc	tac	gcg	288
Asn	Glu	Thr	Cys	Ala	Lys	Asn	Cys	Cys	Leu	Asp	Gly	Ala	Ala	Tyr	Ala	
	65					70					75					
tcc	acg	tac	gga	gtt	acc	acg	agc	ggc	aac	agc	ctc	tcc	att	ggc	ttt	336
Ser	Thr	Tyr	Gly	Val	Thr	Thr	Ser	Gly	Asn	Ser	Leu	Ser	Ile	Gly	Phe	

80				85				90			95					
gtc	acc	cag	tct	gcg	cag	aag	aac	gtt	ggc	gct	cgc	ctt	tac	ctt	atg	384
Val	Thr	Gln	Ser	Ala	Gln	Lys	Asn	Val	Gly	Ala	Arg	Leu	Tyr	Leu	Met	
				100					105					110		
gcg	agc	gac	acg	acc	tac	cag	gaa	ttc	acc	ctg	ctt	ggc	aac	gag	ttc	432
Ala	Ser	Asp	Thr	Thr	Tyr	Gln	Glu	Phe	Thr	Leu	Leu	Gly	Asn	Glu	Phe	
			115					120					125			
tct	ttc	gat	gtt	gat	gtt	tcg	cag	ctg	ccg	tgc	ggc	ttg	aac	gga	gct	480
Ser	Phe	Asp	Val	Asp	Val	Ser	Gln	Leu	Pro	Cys	Gly	Leu	Asn	Gly	Ala	
		130					135					140				
ctc	tac	ttc	gtg	tcc	atg	gac	gcg	gat	ggt	ggc	gtg	agc	aag	tat	ccc	528
Leu	Tyr	Phe	Val	Ser	Met	Asp	Ala	Asp	Gly	Gly	Val	Ser	Lys	Tyr	Pro	
	145					150					155					
acc	aac	acc	gct	ggc	gcc	aag	tac	ggc	acg	ggg	tac	tgt	gac	agc	cag	576
Thr	Asn	Thr	Ala	Gly	Ala	Lys	Tyr	Gly	Thr	Gly	Tyr	Cys	Asp	Ser	Gln	
	160					165				170					175	
tgt	ccc	cgc	gat	ctg	aag	ttc	atc	aat	ggc	cag	gcc	aac	gtt	gag	ggc	624
Cys	Pro	Arg	Asp	Leu	Lys	Phe	Ile	Asn	Gly	Gln	Ala	Asn	Val	Glu	Gly	
				180					185					190		
tgg	gag	ccg	tca	tcc	aac	aac	gcg	aac	acg	ggc	att	gga	gga	cac	gga	672
Trp	Glu	Pro	Ser	Ser	Asn	Asn	Ala	Asn	Thr	Gly	Ile	Gly	Gly	His	Gly	
			195					200					205			
agc	tgc	tgc	tct	gag	atg	gat	atc	tgg	gag	gcc	aac	tcc	atc	tcc	gag	720
Ser	Cys	Cys	Ser	Glu	Met	Asp	Ile	Trp	Glu	Ala	Asn	Ser	Ile	Ser	Glu	
		210					215					220				
gct	ctt	acc	ccc	cac	cct	tgc	acg	act	gtc	ggc	cag	gag	atc	tgc	gag	768
Ala	Leu	Thr	Pro	His	Pro	Cys	Thr	Thr	Val	Gly	Gln	Glu	Ile	Cys	Glu	
	225					230					235					
ggt	gat	ggg	tgc	ggc	gga	act	tac	tcc	gat	aac	aga	tat	ggc	ggc	act	816
Gly	Asp	Gly	Cys	Gly	Gly	Thr	Tyr	Ser	Asp	Asn	Arg	Tyr	Gly	Gly	Thr	
	240				245					250					255	
tgc	gat	ccc	gat	ggc	tgc	gac	tgg	aac	cca	tac	cgc	ctg	ggc	aac	acc	864
Cys	Asp	Pro	Asp	Gly	Cys	Asp	Trp	Asn	Pro	Tyr	Arg	Leu	Gly	Asn	Thr	
				260					265					270		
agc	ttc	tac	ggc	cct	ggc	tca	agc	ttt	acc	ctc	gat	acc	acc	aag	aaa	912
Ser	Phe	Tyr	Gly	Pro	Gly	Ser	Ser	Phe	Thr	Leu	Asp	Thr	Thr	Lys	Lys	
			275					280					285			
ttg	acc	gtc	gtc	acc	cag	ttc	gag	acg	tcg	ggt	gcc	atc	aac	cga	tac	960
Leu	Thr	Val	Val	Thr	Gln	Phe	Glu	Thr	Ser	Gly	Ala	Ile	Asn	Arg	Tyr	
		290					295					300				
tat	gtc	cag	aat	ggc	gtc	act	ttc	cag	cag	ccc	aac	gcc	gag	ctt	ggt	1008
Tyr	Val	Gln	Asn	Gly	Val	Thr	Phe	Gln	Gln	Pro	Asn	Ala	Glu	Leu	Gly	
	305					310					315					
agt	tac	tct	ggc	aac	gag	ctc	aac	gat	gat	tac	tgc	aca	gct	gag	gag	1056
Ser	Tyr	Ser	Gly	Asn	Glu	Leu	Asn	Asp	Asp	Tyr	Cys	Thr	Ala	Glu	Glu	

320		325		330		335	
gca gaa ttc ggc gga tcc tct ttc tca gac aag ggc ggc ctg act cag							1104
Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln							
		340		345		350	
ttc aag aag gct acc tct ggc ggc atg gtt ctg gtc atg agt ctg tgg							1152
Phe Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp							
		355		360		365	
gat gat tac tac gcc aac atg ctg tgg ctg gac tcc acc tac ccg aca							1200
Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr							
		370		375		380	
aac gag acc tcc tcc aca ccc ggt gcc gtg cgc gga agc tgc tcc acc							1248
Asn Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr							
		385		390		395	
agc tcc ggt gtc cct gct cag gtc gaa tct cag tct ccc aac gcc aag							1296
Ser Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys							
		400		405		410	
gtc acc ttc tcc aac atc aag ttc gga ccc att ggc agc acc ggc aac							1344
Val Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn							
		420		425		430	
cct agc ggc ggc aac cct ccc ggc gga aac ccg cct ggc acc acc acc							1392
Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr							
		435		440		445	
acc cgc cgc cca gcc act acc act gga agc tct ccc gga cct acc cag							1440
Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln							
		450		455		460	
tct cac tac ggc cag tgc ggc ggt att ggc tac agc ggc ccc acg gtc							1488
Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val							
		465		470		475	
tgc gcc agc ggc aca act tgc cag gtc ctg aac cct tac tac tct cag							1536
Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln							
		480		485		490	
tgc ctgtaa							1545
Cys							
<210> 2							
<211> 513							
<212> PRT							
<213> Trichoderma reesei							
<400>							
Met Tyr Arg Lys Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Arg							
		-15		-10		-5	
Ala Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr							
2 -1 1		5		10		15	

ES 2 424 346 T3

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

Cys Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr
 260 265 270
 Ser Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys
 275 280 285
 Leu Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr
 290 295 300
 Tyr Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly
 305 310 315
 Ser Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu
 320 325 330 335
 Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln
 340 345 350
 Phe Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp
 355 360 365
 Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr
 370 375 380
 Asn Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr
 385 390 395
 Ser Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys
 400 405 410 415
 Val Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn
 420 425 430
 Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr
 435 440 445
 Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln
 450 455 460
 Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val
 465 470 475
 Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln
 480 485 490 495 Cys

ES 2 424 346 T3

<211> 20
 <212> DNA
 <213> E. coli

5 <400> 3
 taatacgact cactataggg 20

<210> 4
 <211> 67
 <212> DNA
 <213> Trichoderma reesei

<400>
 4
tatacctcta tactttaacg tcaaggagaa aaactatag gatccaccat gtatcggaag 60
ttggccg 67

<210> 5
 <211> 65
 <212> DNA
 <213> Trichoderma reesei

20 <400>
 5
cataactaat tacatgatgc ggcctctag atgcacatga ctcgagttac aggcactgag 60
agtag 65

25 <210> 6
 <211> 67
 <212> DNA
 <213> Saccharomyces cerevisiae

30 <400>
 6
tatacctcta tactttaacg tcaaggagaa aaactatag gatccaccat gtatcggaag 60
ttggccg 67

35 <210> 7
 <211> 65
 <212> DNA
 <213> Saccharomyces cerevisiae

40 <400>
 7
cataactaat tacatgatgc ggcctctag atgcacatga ctcgagttac aggcactgag 60
agtag 65

45 <210> 8
 <211> 20
 <212> DNA
 <213> Saccharomyces cerevisiae

50 <400> 8
 ggcgtgaatg taagcgtgac 20
 <210> 9

<211> 28
 <212> DNA
 <213> *Saccharomyces cerevisiae*
 5 <400> 9
 ctggggaat taatcagcga agcgatga 28
 <210> 10
 <211> 19
 10 <212> DNA
 <213> *Saccharomyces cerevisiae*
 <400> 10
 15 gcgtacacgc gtctgtaca 19
 <210> 11
 <211> 22
 <212> DNA
 20 <213> *Aspergillus nidulans*
 <400> 11
 gtgccccatg atacgcctcc gg 22
 <210> 12
 25 <211> 26
 <212> DNA
 <213> *Aspergillus nidulans*
 <400> 12
 30 gagtcgtatt tccaaggctc ctgacc 26
 <210> 13
 <211> 24
 <212> DNA
 35 <213> *Aspergillus nidulans*
 <400> 13
 ggaggccatg aagtggacca acgg 24
 40 <210> 14
 <211> 45
 <212> DNA
 <213> *Aspergillus niger*
 45 <400> 14
 caccgtgaaa gccatgctct ttcttcgtg tagaagacca gacag 45
 <210> 15
 <211> 45
 50 <212> DNA
 <213> *Aspergillus niger*
 <400> 15
 55 ctggtcttct acacgaagga aagagcatgg ctttcacggt gtctg 45
 <210> 16
 <211> 44
 <212> DNA
 60 <213> *Aspergillus oryzae*

ES 2 424 346 T3

<400> 16
 ctatatacac aactggattt accatgggcc cgcgccgca gatc 44
 5
 <210> 17
 <211> 44
 <212> DNA
 <213> *Aspergillus oryzae*
 <400> 17
 10 gatctgcgcc cgcgggccca tggtaaacc agttgtgat atag 44
 <210> 18
 <211> 22
 <212> DNA
 15 <213> *Trichoderma reesei*
 <400> 18
 20 gcaacatgta tcggaagttg gc 22
 <210> 19
 <211> 22
 <212> DNA
 <213> *Trichoderma reesei*
 25 <400> 19
 aattaattt acaggcactg ag 22
 <210> 20
 <211> 33
 <212> DNA
 30 <213> *Trichoderma reesei*
 <400> 20
 35 gaacacgggc attggacgac acggaagctg ctg 33
 <210> 21
 <211> 33
 <212> DNA
 <213> *Trichoderma reesei*
 40 <400> 21
 cagcagcttc cgtgctgcc aatgccctg ttc 33
 <210> 22
 <211> 20
 <212> DNA
 <213> *Trichoderma reesei*
 45 <400> 22
 50 ctcttgcc acagctcgtg 20
 <210> 23
 <211> 20
 <212> DNA
 55 <213> *Trichoderma reesei*
 <400> 23
 ggctttgtca cccagtctgc 20
 60 <210> 24

5 <210> 31
 <211> 19
 <212> DNA
 <213> Saccharomyces cerevisiae

<400> 31
 gcgtacacgc gtctgtaca 19

10 <210> 32
 <211> 31
 <212> DNA
 <213> Trichoderma reesei

15 <220>
 <221> misc_feature
 <222> (21)..(21)
 <223> N= A,C, G,OR T

20 <220>
 <221> misc_feature
 <222> (22)..(22)
 <223> N= A,C, G,OR T

25 <400> 32
 gcggtaacag cctctccatt nnstttgtca c 31

30 <210> 33
 <211> 33
 <212> DNA
 <213> Trichoderma reesei

35 <220>
 <221> misc_feature
 <222> (23)..(24)
 <223> N= A, C, G, OR T

40 <400> 33
 ctgcgacagac tgggtgacaa asnnaatgga gag 33

45 <210> 34
 <211> 32
 <212> DNA
 <213> Trichoderma reesei

50 <220>
 <221> misc_feature
 <222> (21)..(22)
 <223> N= A, C, G, OR T

<400> 34
 ccatctccga ggctcttacc nnsaccctt gc 32

55 <210> 35
 <211> 33
 <212> DNA
 <213> Trichoderma reesei

60 <220>
 <221> misc_feature

ES 2 424 346 T3

<222> (24)..(25)
 <223> N= A, C, G, OR T

5 <400> 35
 ggccgacagt cgtgcaaggg tgsnnggtaa gag 33

<210> 36
 <211> 30
 <212> DNA
 10 <213> Trichoderma reesei

<220>
 <221> misc_feature
 <222> (19)..(20)
 15 <223> N= A, C, G, OR T

<400> 36
 gagggctggg agccgtcann saacaacgcg 30

20 <210> 37
 <211> 33
 <212> DNA
 <213> Trichoderma reesei

25 <220>
 <221> misc_feature
 <222> (25)..(26)
 <223> N= A, C, G, OR T

30 <400> 37
 ccaatgcccg tgttcgcgtt gtsnntgac ggc 33

35 <210> 38
 <211> 29
 <212> DNA
 <213> Trichoderma reesei

40 <400> 38
 aacgtaatt aaggaatcgt tttgtgtt 29

45 <210> 39
 <211> 29
 <212> DNA
 <213> Trichoderma reesei

<400> 39
 agtactagta gctccgtggc gaaagcctg 29

50 <210> 40
 <211> 26
 <212> DNA
 <213> Trichoderma reesei

55 <400> 40
 actagtcgac cgaatgtagg attggt 26

60 <210> 41
 <211> 19
 <212> DNA
 <213> Trichoderma reesei

ES 2 424 346 T3

<400> 41
 tgaccatggt gcgcagtcc 19
 5
 <210> 42
 <211> 26
 <212> DNA
 <213> Trichoderma reesei
 10
 <400> 42
 cgatcgtctc cctatgggtc attacc 26
 <210> 43
 <211> 28
 <212> DNA
 <213> Trichoderma reesei
 15
 <400> 43
 actagttaat taagctccgt ggcgaaag 28
 20
 <210> 44
 <211> 30
 <212> DNA
 <213> *Trichoderma reesei*
 25
 <400> 44
 cgcggactgc gcaccatgta tcggaagttg 30
 <210> 45
 <211> 30
 <212> DNA
 <213> Trichoderma reesei
 30
 <400> 45
 cgccacggag ctaattaca ggcactgaga 30
 <210> 46
 <211> 20
 <212> DNA
 <213> Trichoderma reesei
 40
 <400> 46
 gccttcggcc tftgggtgta 20
 45
 <210> 47
 <211> 20
 <212> DNA
 <213> Trichoderma reesei
 50
 <400> 47
 gagcggcgat tctacgggtt 20
 <210> 48
 <211> 26
 <212> DNA
 <213> Escherichia coli
 55
 <400> 48
 gcccgggcac gcgccacacg gaaaat 26
 60
 <210> 49
 <211> 27
 <212> DNA
 <213> Escherichia coli
 65
 <400> 49

gaccggtcgc aaaatgacaa atagaag 27
 <210> 50
 <211> 23
 <212> DNA
 <213> Trichoderma reesei
 5
 <400> 50
 gctccgggca aatgcaaagt gtg 23
 10
 <210> 51
 <211> 24
 <212> DNA
 <213> Trichoderma reesei
 15
 <400> 51
 agcaggccgc atctccagtg aaag 24
 20
 <210> 52
 <211> 20
 <212> DNA
 <213> Escherichia coli
 25
 <400> 52
 gcacgcgcca cacgaaaat 20
 30
 <210> 53
 <211> 20
 <212> DNA
 <213> Escherichia coli
 <400> 53
 cgcaaatga caatagaag 20
 35
 <210> 54
 <211> 31
 <212> DNA
 <213> Aspergillus fumigatus
 40
 <400> 54
 actggattta ccatgagatt cggttggctc g 31
 45
 <210> 55
 <211> 31
 <212> DNA
 <213> Aspergillus fumigatus
 <400> 55
 agtcacctct agttactagt agacacgggg c 31
 50
 <210> 56
 <211> 3060
 <212> DNA
 <213> Aspergillus fumigatus
 55
 <400>
 56

atgagattcg gttggctoga ggtggccgct ctgaaggccg ettctgtago caatgccag 60
gtttgtgatg ctttcccgtc attgtttcgg atatagttga caatagtcot ggaaataato 120
aggaattggc tttctctcca ccattctacc cttcgccctg ggctgatggc cagggagagt 180
gggcagatgc ccctcgacgc gccgtcgaga t cgtttctca gatgacactg gcggagaagg 240
ttaaccttac aacgggtaot ggggtgggtg cgaacttttt gttgacagtg agctttcttc 300
actgaccato tacacagatg ggaaatggac cgatgcgtcg gtcaaaccgg cagcgttccc 360
aggtaagctt gcaattotgc aacaacgtgc a.agtgtagt gctaaaacgc ggtggtgcag 420
acttggatc aactggggtc tttgtggcca ggattocct ttgggtatcc gtttctgtga 480
gctatacccg cggagtcttt cagtccctgt attatgtgct gatgattgtc totgtatago 540
tgacctcaac tcggccttc ctgctggtac tmaatgtcgcc gcgacatggg acaagacact 600
cgctacott cgtggcaagg ccctgggtga ggaattcaac gacaaggcg tggacatttt 660
gctggggcct gctgctggtc ctctcggcaa ataccocggac ggccggcagaa tctgggaagg 720
cttctctcct gatccggctc tcactggtgt actttctcgc gaaactatca agggtatcca 780
agacgggggt gtgattgcta ctgccaagca ttacattctg aatgaacagg agcatttccg 840
acaggttggc gaggcccagg gatatggtta caacatcacg gagacgatca gctccaacgt 900
ggatgacaag accatgcaog agttgtacct ttgggtgagta gttgacactg caaatgagga 960
ccttgattga tttgactgac ctggaatgca ggccctttgc agatgctgtg ccgggtaaga 1020
ttttccgtag acttgacctc gggacgaaga aatcgctgac gaacctcgt agctggcgtt 1080
ggcgtgtca tgtgttcta caatcaaato aa caacagct acggttgtca aaacagtcaa 1140
actotcaaca agtctctcaa ggctgagctg ggcttccaag gcttcgtcat gagtgactgg 1200
agcgtcacc acagcgggtg cggcgtgccc ct cgtgggt tggatattgtc gatgcctgga 1260
gacatttctc tcgacgacgg actctcttc tggggcacga acctaaactgt cagtgttctt 1320
aacggcaocg ttccagcctg gctgtcgat gacatggctg ttctgtatcat gaccgcgtac 1380
tacaaggttg gtctgaccg tcttctatt cc ccttaact tcagctcctg gaccgggat 1440
gagtacggct gggagcattc tgctgtctcc gagggagcct ggaaccaagg gaacgacttc 1500
gtcaatgtgc agcgcagtca ctctcagatc at cgtgaga ttggtgccc tagtacagt 1560
ctcttgaaga acacgggtgc tottctctg ac cggcaagg aggttaaagt ggggttctc 1620
ggtgaagaog ctggttcaa cccgtgggt gc taacggct gcccgcacog cggctgtgat 1680
aacggcaotc ttgctatggc ctgggtagt ggtactgcca aottcotta ccttgtcacc 1740

ccgagcagg ctatccagcg agaggteate agcaacggcg gcaatgtott tgctgtgact 1800
 gataacgggg ctctcagcca gatggcagat gttgcctctc aatccagggtg agtgccgggt 1860
 cttagaaaaa gaacgttctc tgeatgaagt tttttaacca ttgcgaacag cgtgtctttg 1920
 gtgtttgtca acgcccactc tggagagggt ttcctcagtg tcgacggcaa cgagggtgac 1980
 cgcaaaaate tcaactctgtg gaagaacggc gagggcgtca ttgacactgt tgtcagccac 2040
 tgcaacaaca cgattgtggt tattcacagt gttgggcccg tcttgatoga ccggtggtat 2100
 gataacccc aagtcactgc catcatctgg gccggcttgc ccggtcagga gagtggcaac 2160
 tccctggctg acgtgtctca tggccggctc aaccccagcg ccaagaaccc gttcaactgg 2220
 ggcaagactc gggagtctta cggggctccc ttgctcacgg agcctaacaa tggcaatggt 2280
 gctcccaggg atgatttcaa cgaggggctc ttcattgact accgtcaact tgacaagcgc 2340
 aatgagaccc ccatttatga gtttggccat ggcttgagct acaccacact tggttactct 2400
 caccttcggg ttcaggccct caatagtctg agttcggcat atgtcccagc tagcggagag 2460
 accaagcctg cgcacaacta tggtgagatc ggtagtgcgg ccgactacct gtatcccag 2520
 ggtctcaaaa gaattacaa gtttatttac ccttggctca actcgaccga cctcgaggat 2580
 tctttctgac accogaacta cggctggggag gactcggagt acattcccga aggcgttagg 2640
 gatgggtctc ctcaacccct cctgaaggct ggcggcgctc ctgggtggtaa cctaccctt 2700
 tatcaggatc ttgttagggt gtccggccacc ataaccaaca ctggtaacgt cgcgggttat 2760
 gaagtcctc aattgggtgag tgaccocgat gttccttggc ttgcaatttg gctaaactgc 2820
 ttctagtatg tttcaactgg cggaccgaac gagcctcggg togttctgoc caagtctgac 2880
 cgaatcttcc tggctcctgg ggagcaaaag gtttggacca cgaactctta ccgtcgtgat 2940
 ctcgccaatt gggatgtgga ggctcaggac tgggtcatca caaagtaacc caagaagtg 3000
 cacgtcggca gctcctcggc taagctgcct ctgagagcgc ctctgcccgg tgtctactag 3060

<210> 57
 <211> 863
 <212> PRT
 <213> *Aspergillus fumigatus*

5

<400> 57
 Met Arg Phe Gly Trp Leu Glu Val Ala Ala Leu Thr Ala Ala Ser Val
 1 5 10 15

Ala Asn Ala Gln Glu Leu Ala Phe Ser Pro Pro Phe Tyr Pro Ser Pro
 20 25 30

Trp Ala Asp Gly Gln Gly Glu Trp Ala Asp Ala His Arg Arg Ala Val
 35 40 45

Glu Ile Val Ser Gln Met Thr Leu Ala Glu Lys Val Asn Leu Thr Thr
 50 55 60

Gly Thr Gly Trp Glu Met Asp Arg Cys Val Gly Gln Thr Gly Ser Val
 65 70 75 80

Pro Arg Leu Gly Ile Asn Trp Gly Leu Cys Gly Gln Asp Ser Pro Leu
 85 90 95

Gly Ile Arg Phe Ser Asp Leu Asn Ser Ala Phe Pro Ala Gly Thr Asn
 100 105 110

Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Lys Ala
 115 120 125

Met Gly Glu Glu Phe Asn Asp Lys Gly Val Asp Ile Leu Leu Gly Pro
 130 135 140

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

Gly Phe Ser Pro Asp Pro Val Leu Thr Gly Val Leu Phe Ala Glu Thr
 165 170 175

Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
 180 185 190

Ile Leu Asn Glu Gln Glu His Phe Arg Gln Val Gly Glu Ala Gln Gly
 195 200 205

Tyr Gly Tyr Asn Ile Thr Glu Thr Ile Ser Ser Asn Val Asp Asp Lys
 210 215 220

Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala
 225 230 235 240

Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr
 245 250 255

Gly Cys Gln Asn Ser Gln Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu
 260 265 270

ES 2 424 346 T3

Gly Phe Gln Gly Phe Val Met Ser Asp Trp Ser Ala His His Ser Gly
 275 280 285

Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Ile
 290 295 300

Ser Phe Asp Asp Gly Leu Ser Phe Trp Gly Thr Asn Leu Thr Val Ser
 305 310 315 320

Val Leu Asn Gly Thr Val Pro Ala Trp Arg Val Asp Asp Met Ala Val
 325 330 335

Arg Ile Met Thr Ala Tyr Tyr Lys Val Gly Arg Asp Arg Leu Arg Ile
 340 345 350

Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Trp Glu His
 355 360 365

Ser Ala Val Ser Glu Gly Ala Trp Thr Lys Val Asn Asp Phe Val Asn
 370 375 380

Val Gln Arg Ser His Ser Gln Ile Ile Arg Glu Ile Gly Ala Ala Ser
 385 390 395 400

Thr Val Leu Leu Lys Asn Thr Gly Ala Leu Pro Leu Thr Gly Lys Glu
 405 410 415

Val Lys Val Gly Val Leu Gly Glu Asp Ala Gly Ser Asn Pro Trp Gly
 420 425 430

Ala Asn Gly Cys Pro Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met
 435 440 445

Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu
 450 455 460

Gln Ala Ile Gln Arg Glu Val Ile Ser Asn Gly Gly Asn Val Phe Ala
 465 470 475 480

Val Thr Asp Asn Gly Ala Leu Ser Gln Met Ala Asp Val Ala Ser Gln
 485 490 495

Ser Ser Val Ser Leu Val Phe Val Asn Ala Asp Ser Gly Glu Gly Phe
 500 505 510

Ile Ser Val Asp Gly Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp
515 520 525

Lys Asn Gly Glu Ala Val Ile Asp Thr Val Val Ser His Cys Asn Asn
530 535 540

Thr Ile Val Val Ile His Ser Val Gly Pro Val Leu Ile Asp Arg Trp
545 550 555 560

Tyr Asp Asn Pro Asn Val Thr Ala Ile Ile Trp Ala Gly Leu Pro Gly
565 570 575

Gln Glu Ser Gly Asn Ser Leu Val Asp Val Leu Tyr Gly Arg Val Asn
580 585 590

Pro Ser Ala Lys Thr Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr
595 600 605

Gly Ala Pro Leu Leu Thr Glu Pro Asn Asn Gly Asn Gly Ala Pro Gln
610 615 620

Asp Asp Phe Asn Glu Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys
625 630 635 640

Arg Asn Glu Thr Pro Ile Tyr Glu Phe Gly His Gly Leu Ser Tyr Thr
645 650 655

Thr Phe Gly Tyr Ser His Leu Arg Val Gln Ala Leu Asn Ser Ser Ser
660 665 670

Ser Ala Tyr Val Pro Thr Ser Gly Glu Thr Lys Pro Ala Pro Thr Tyr
675 680 685

Gly Glu Ile Gly Ser Ala Ala Asp Tyr Leu Tyr Pro Glu Gly Leu Lys
690 695 700

Arg Ile Thr Lys Phe Ile Tyr Pro Trp Leu Asn Ser Thr Asp Leu Glu
705 710 715 720

Asp Ser Ser Asp Asp Pro Asn Tyr Gly Trp Glu Asp Ser Glu Tyr Ile
725 730 735

Pro Glu Gly Ala Arg Asp Gly Ser Pro Gln Pro Leu Leu Lys Ala Gly
740 745 750

ES 2 424 346 T3

Gly Ala Pro Gly Gly Asn Pro Thr Leu Tyr Gln Asp Leu Val Arg Val
 755 760 765

Ser Ala Thr Ile Thr Asn Thr Gly Asn Val Ala Gly Tyr Glu Val Pro
 770 775 780

Gln Leu Tyr Val Ser Leu Gly Gly Pro Asn Glu Pro Arg Val Val Leu
 785 790 795 800

Arg Lys Phe Asp Arg Ile Phe Leu Ala Pro Gly Glu Gln Lys Val Trp
 805 810 815

Thr Thr Thr Leu Asn Arg Arg Asp Leu Ala Asn Trp Asp Val Glu Ala
 820 825 830

Gln Asp Trp Val Ile Thr Lys Tyr Pro Lys Lys Val His Val Gly Ser
 835 840 845

Ser Ser Arg Lys Leu Pro Leu Arg Ala Pro Leu Pro Arg Val Tyr
 850 855 860

REIVINDICACIONES

- 5 1. Polipéptido de variante aislada con actividad de glucósido hidrolasa, que tiene un grado de identidad con los aminoácidos 1 a 513 de SEC ID n.º: 2 de al menos 85%, incluso más preferiblemente al menos 90%, de forma más preferible al menos 95%, e incluso de forma más preferible al menos 97%, caracterizado por el hecho de que la secuencia de aminoácidos del polipéptido difiere de la SEC ID n.º: 2 en al menos una posición correspondiente a las posiciones 21, 94, 157, 205, 206, 247, 337, 350, 373, 383, 438, 455, 467 y 486 de los aminoácidos 1 a 513 de SEC ID n.º: 2, caracterizado por el hecho de que la diferencia en al menos una posición es una sustitución en la posición 205 y la variante tiene propiedades mejoradas en comparación con la glucósido hidrolasa progenitora, caracterizado por el hecho de que las propiedades mejoradas son seleccionadas del grupo que consiste en perfil de actividad dependiente de la temperatura, termoestabilidad, actividad del pH, estabilidad del pH y especificidad de sustrato, especificidad de producto y estabilidad química.
- 10 2. Polipéptido variante según la reivindicación 1, caracterizado por el hecho de que la sustitución en la secuencia de aminoácidos del polipéptido de SEC ID n.º: 2 en una posición correspondiente a la posición 205 es Arg.
- 15 3. Polipéptido variante según la reivindicación 1, caracterizado por el hecho de que la sustitución en la secuencia de aminoácidos del polipéptido de SEC ID n.º: 2 en una posición correspondiente a la posición 205 es G205R.
- 20 4. Polipéptido variante de cualquiera de las reivindicaciones 1-3, caracterizado por el hecho de que el polipéptido variante comprende sustituciones en posiciones correspondientes a las posiciones 94 e 205 de los aminoácidos 1 a 513 de SEQ ID n.º: 2.
- 25 5. Polipéptido variante según la reivindicación 4, caracterizado por el hecho de que la variante comprende las sustituciones G94S + G205R.
6. Secuencia de nucleótidos aislada que codifica el polipéptido de cualquiera de las reivindicaciones 1-5.
- 30 7. Un constructo de ácidos nucleicos que comprende la secuencia de nucleótidos según la reivindicación 6.
8. Un vector de expresión que comprende la secuencia de nucleótidos según la reivindicación 7.
- 35 9. Una célula huésped que comprende la secuencia de nucleótidos según la reivindicación 6.
10. Un método para producir un polipéptido variante con actividad de glucósido hidrolasa, que comprende (a) cultivo de la célula huésped según la reivindicación 9 bajo condiciones adecuadas para la producción de la variante; y (b) recuperación de la variante del medio de cultivo.
- 40 11. Una composición detergente que comprende un polipéptido de cualquiera de las reivindicaciones 1-5 y un surfactante.
- 45 12. Un método para degradar biomasa que contiene celulosa y hemicelulosa, que comprende tratar la biomasa con una cantidad eficaz de un polipéptido de cualquiera de las reivindicaciones 1-5 y recuperación de la biomasa degradada.

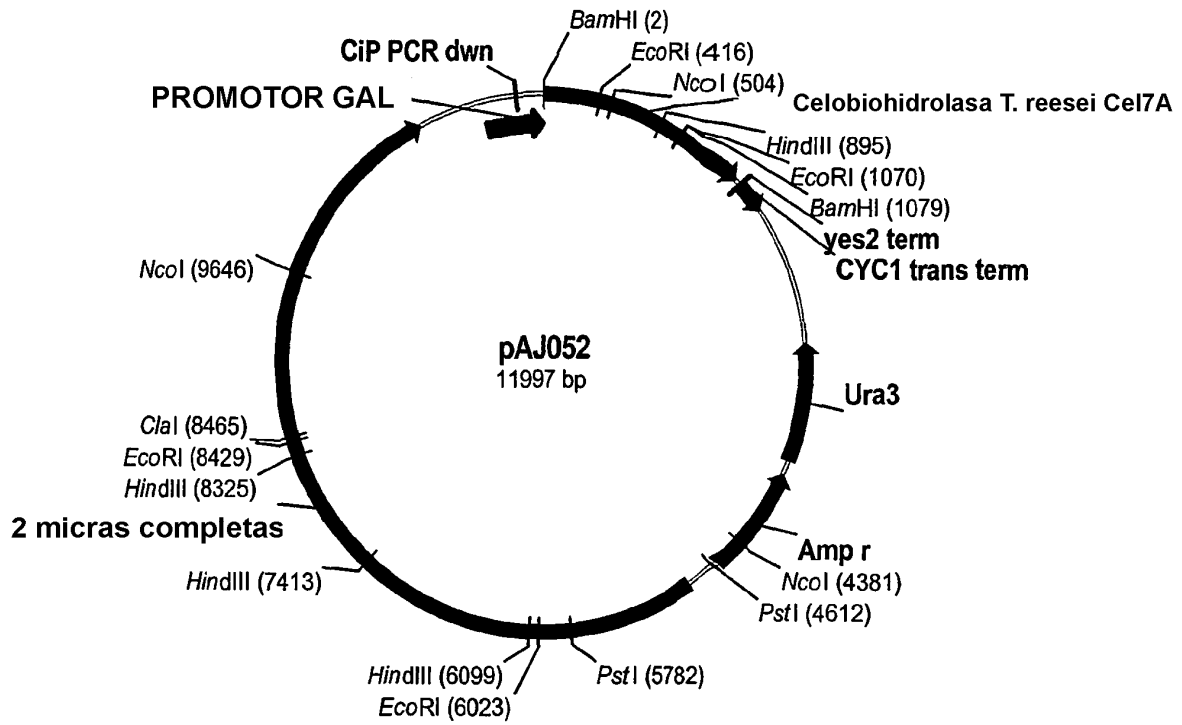


Fig. 1

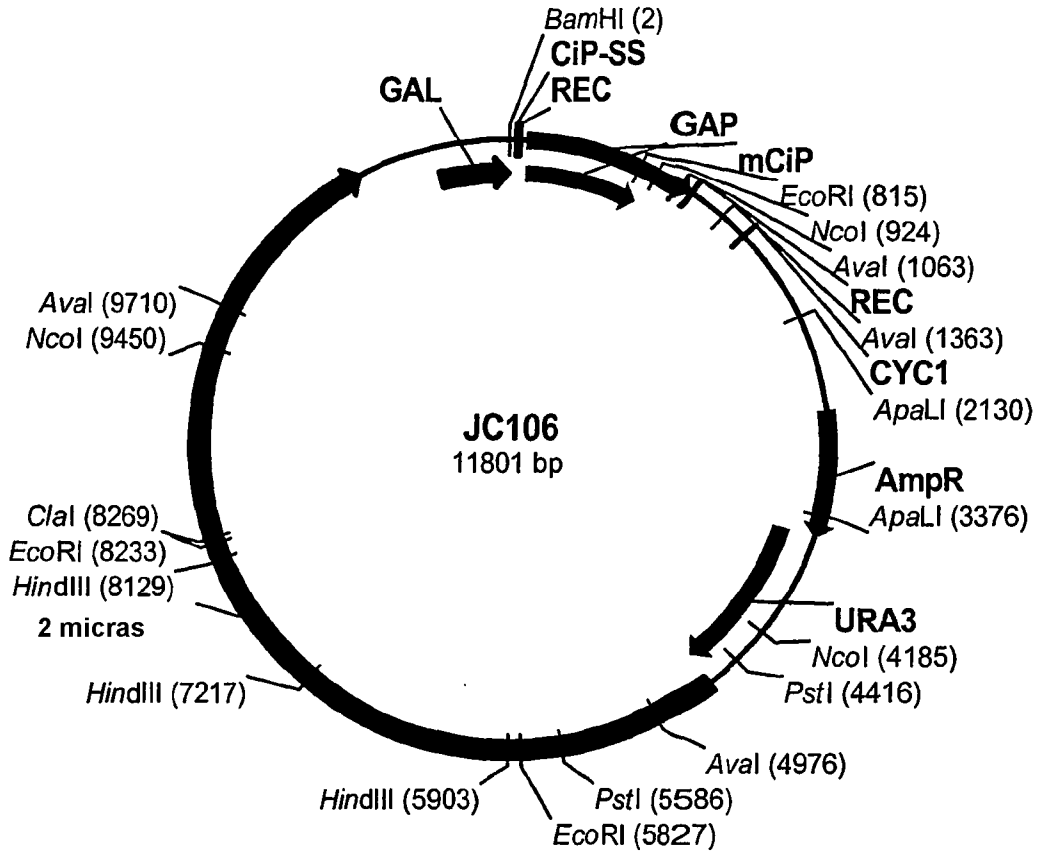


Fig. 2

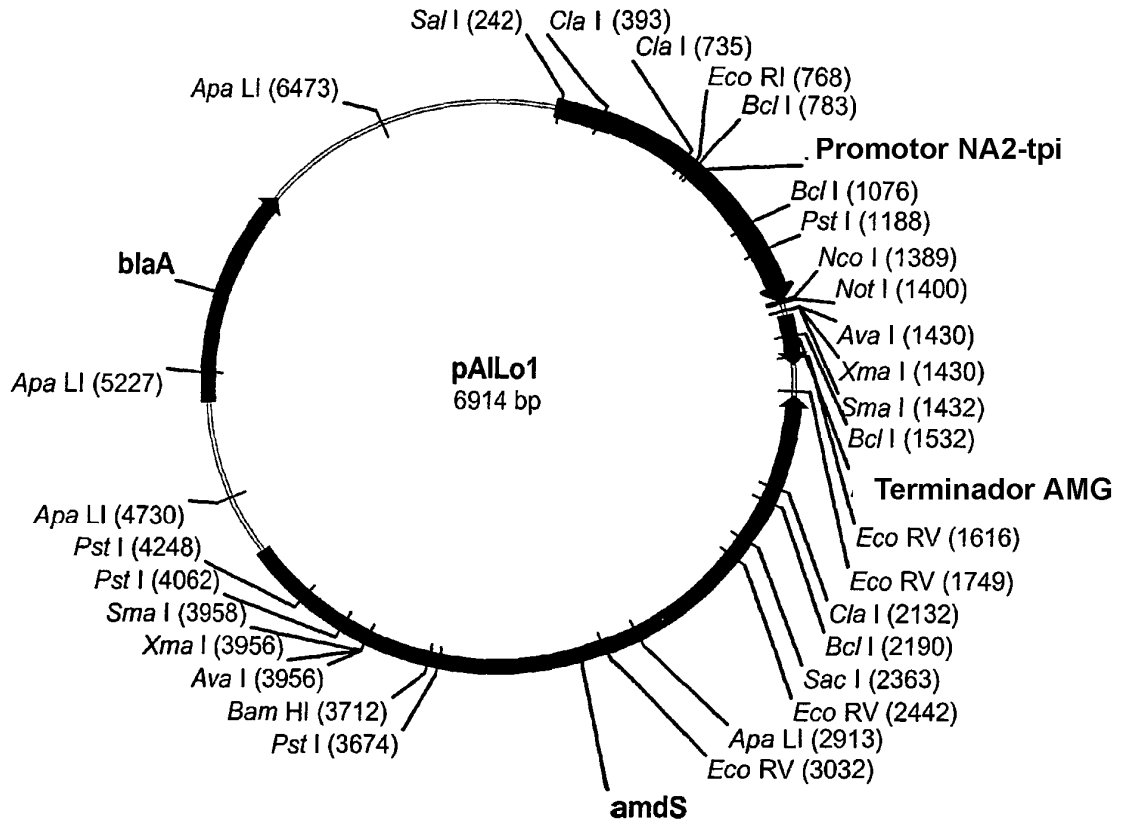


Fig. 3

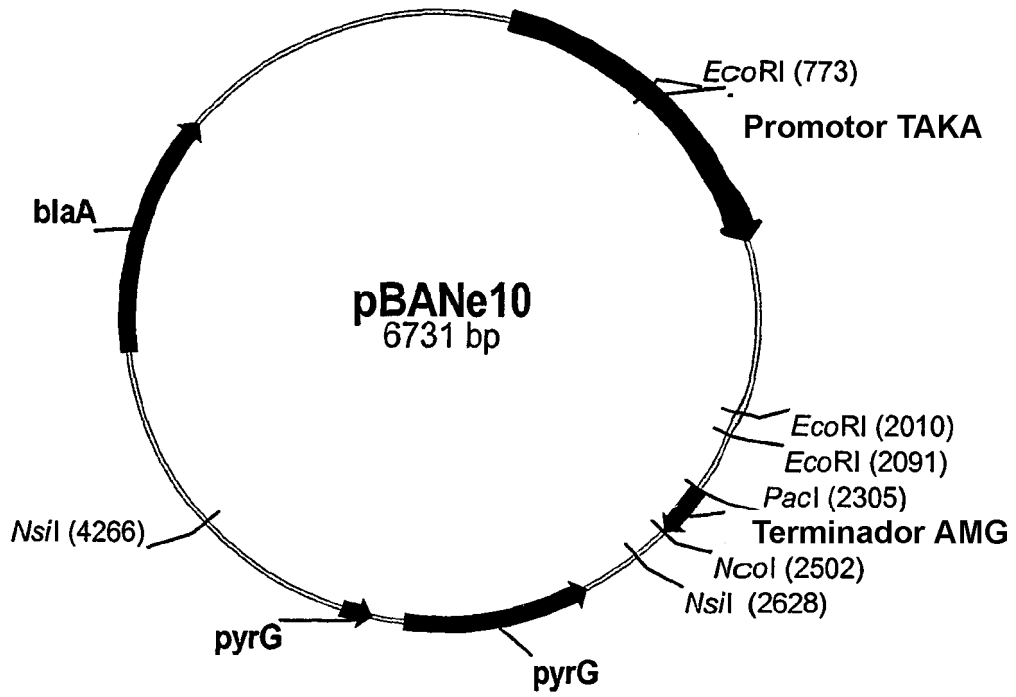


Fig. 4

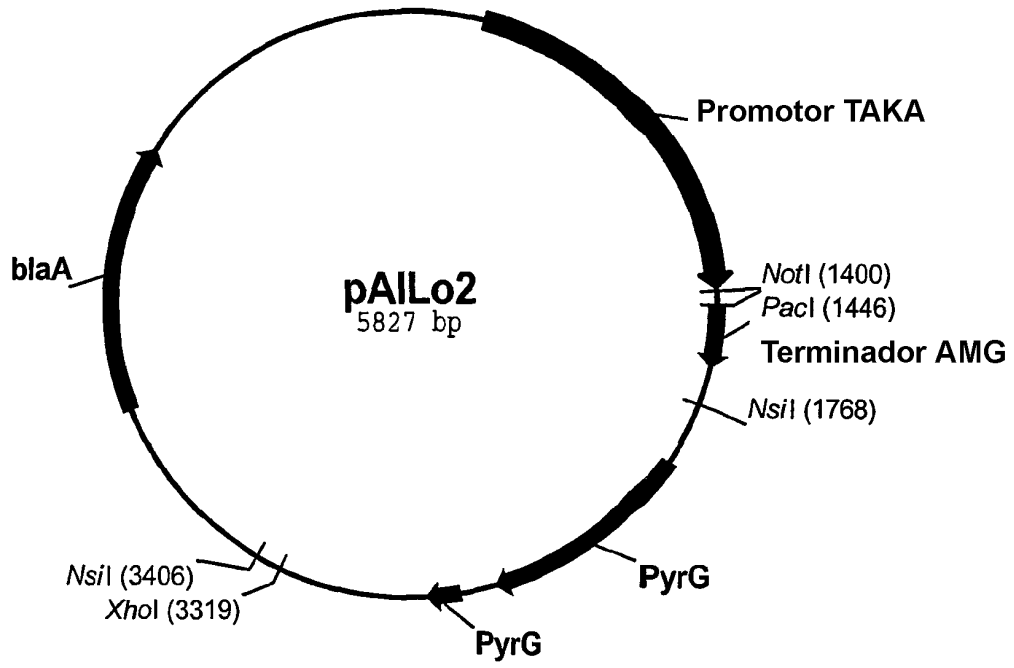


Fig. 5

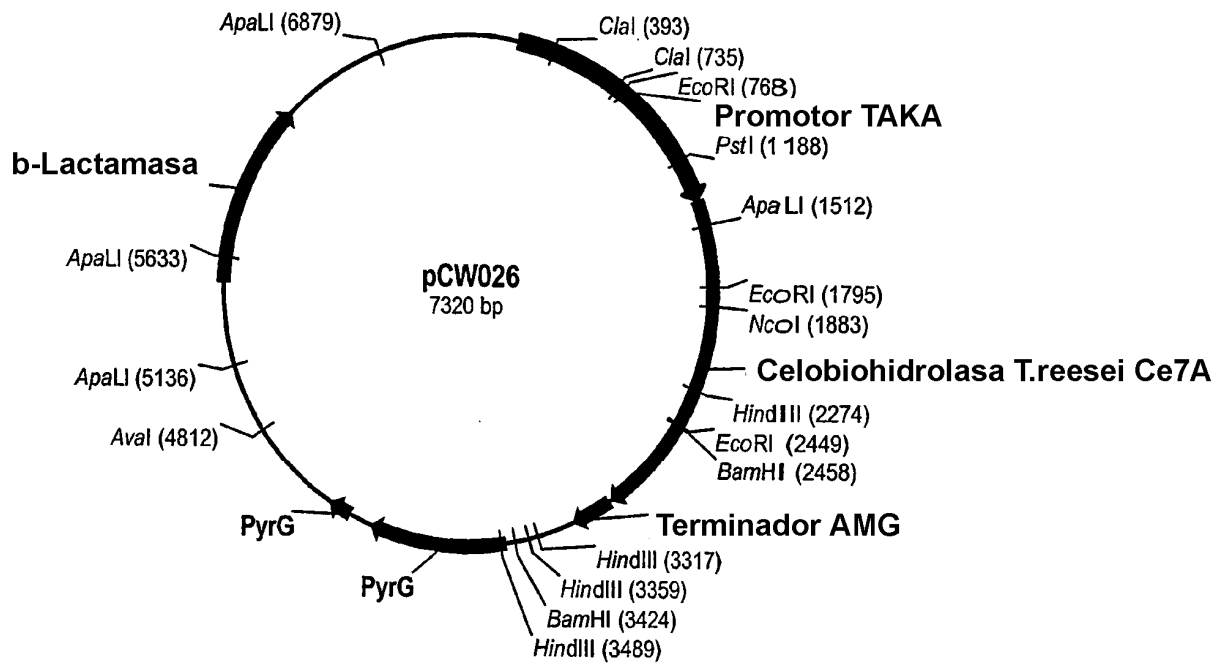


Fig. 6

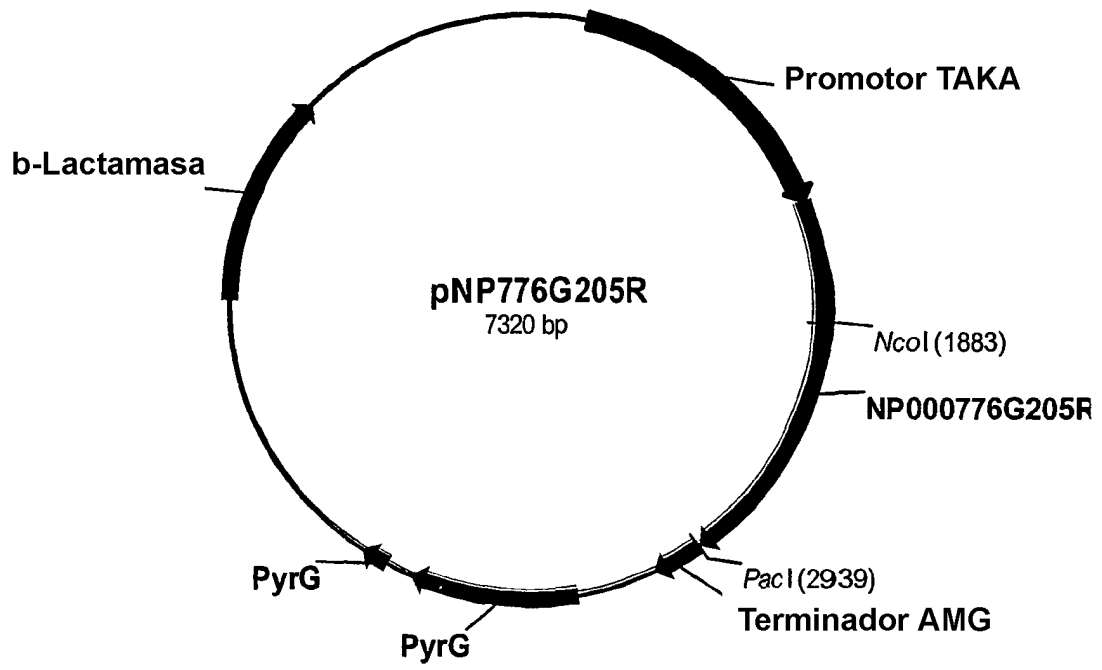


Fig. 7

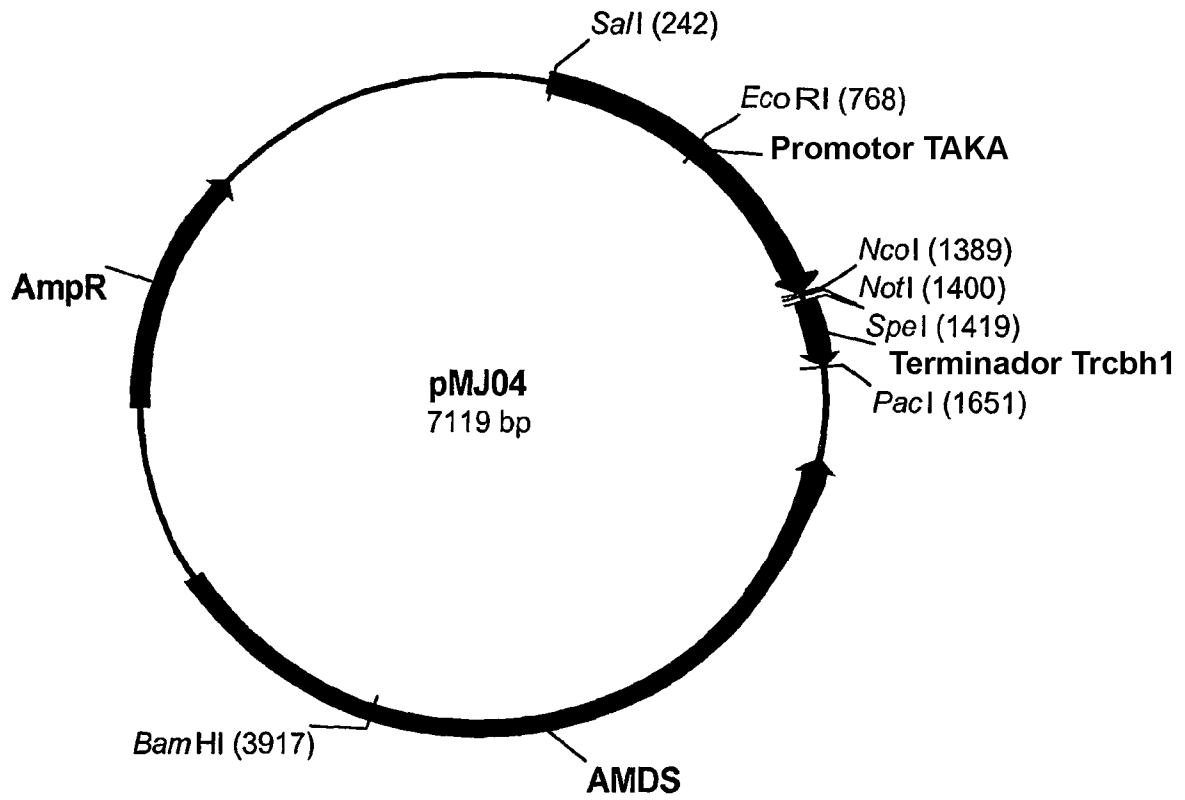


Fig. 8

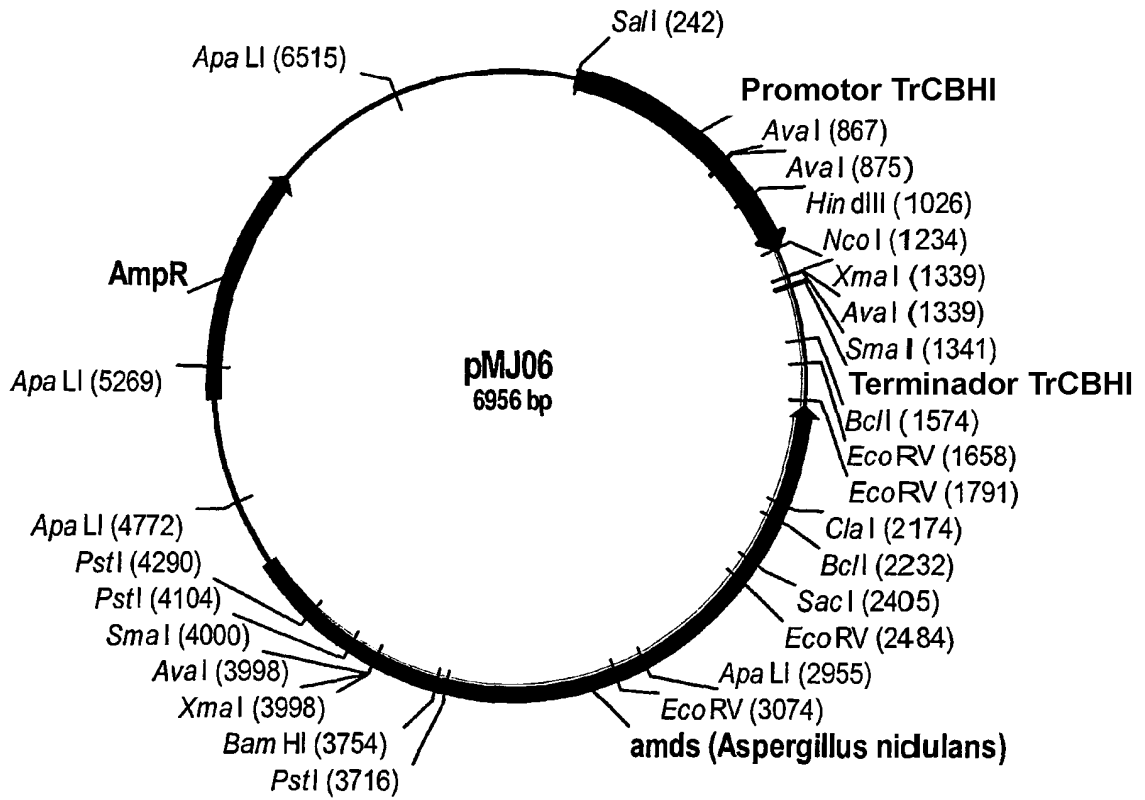


Fig. 9

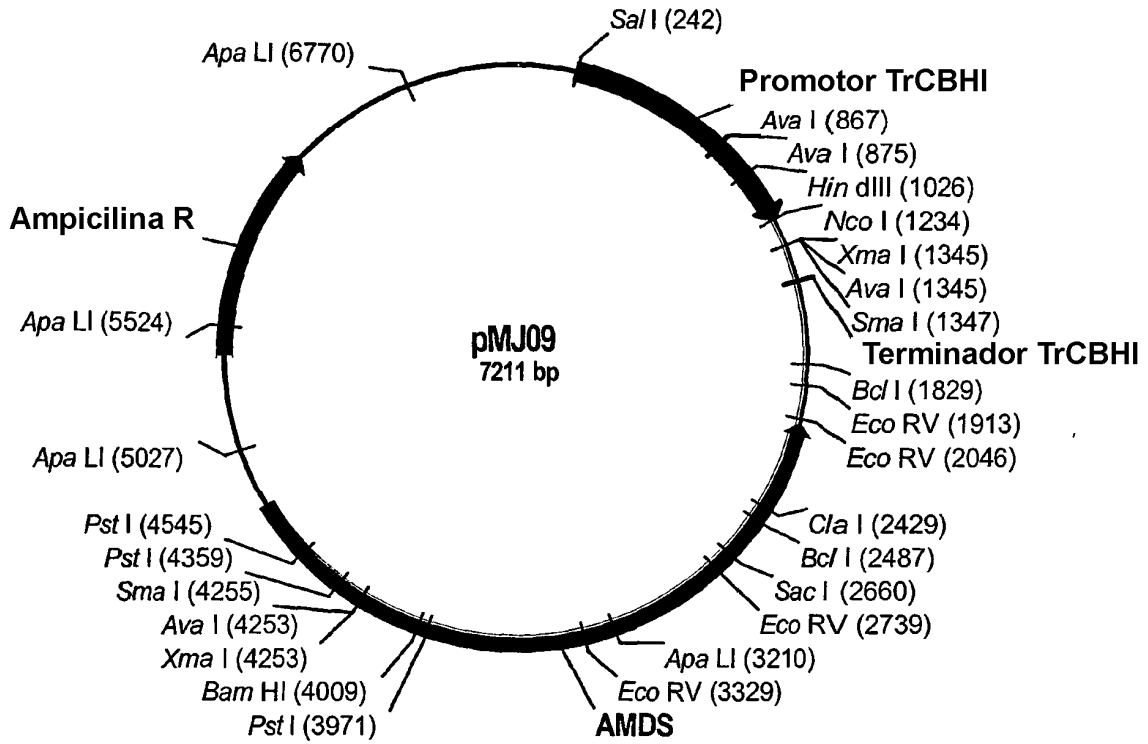


Fig. 10

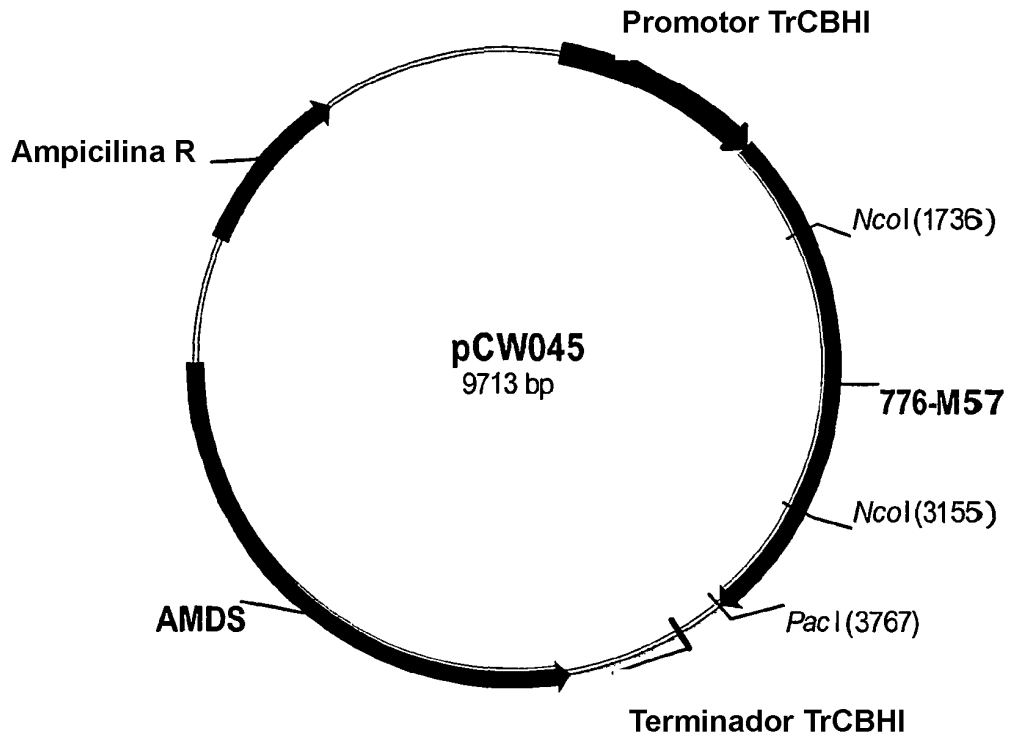


Fig. 11

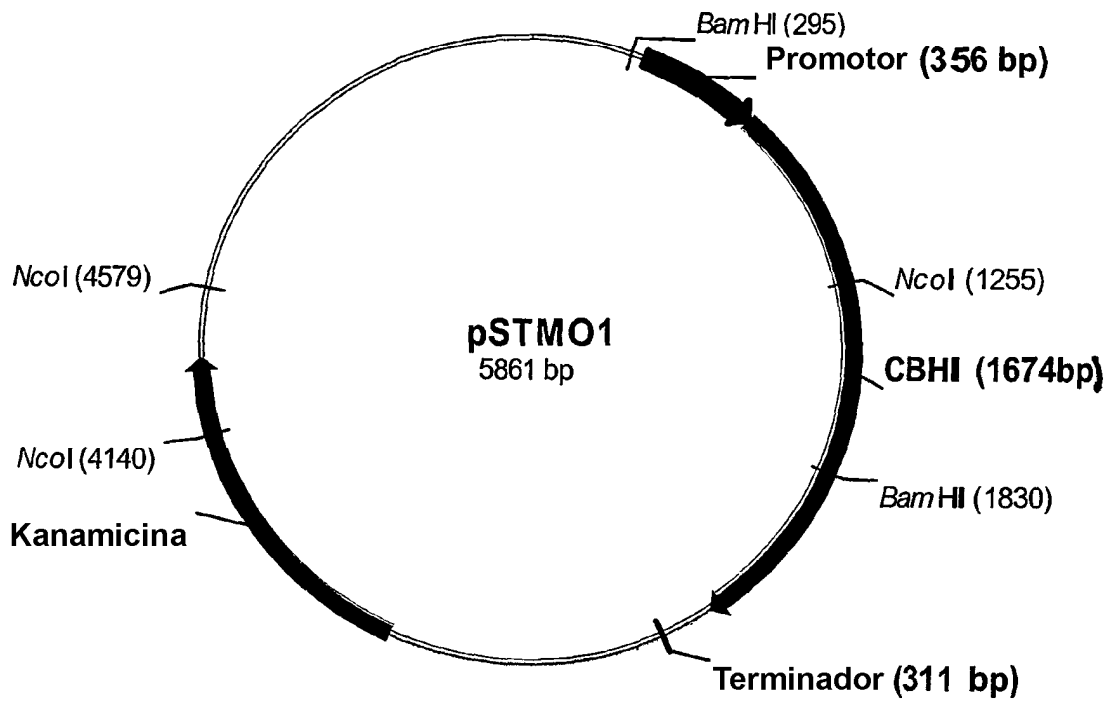


Fig. 12

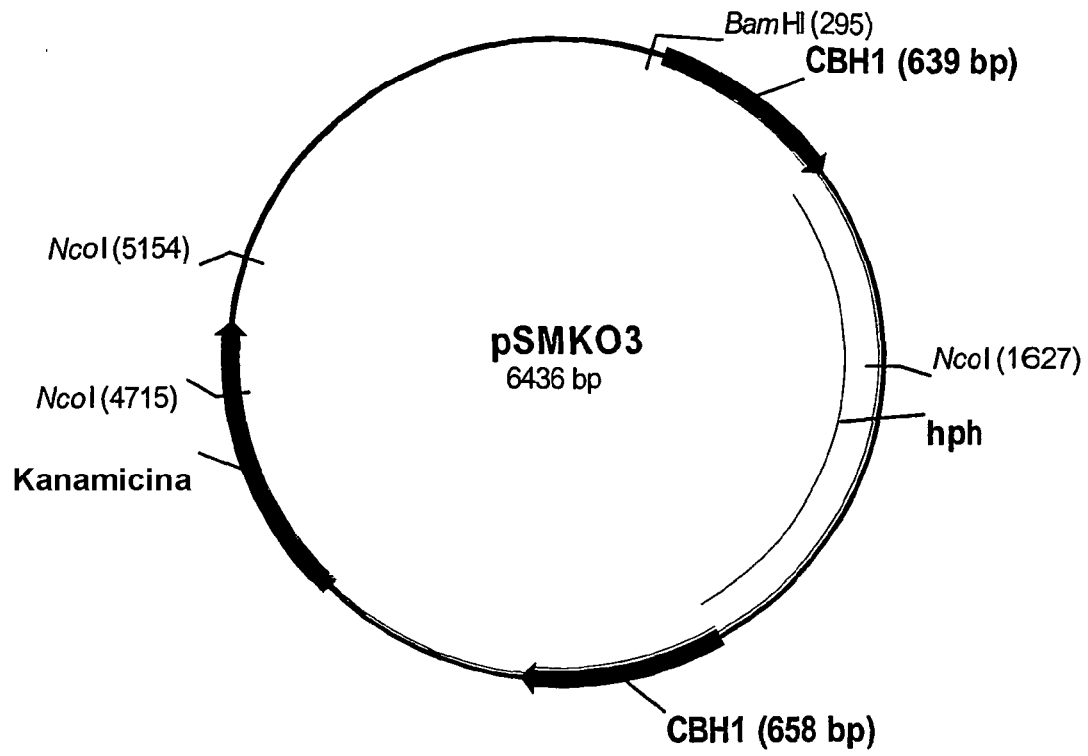


Fig. 13

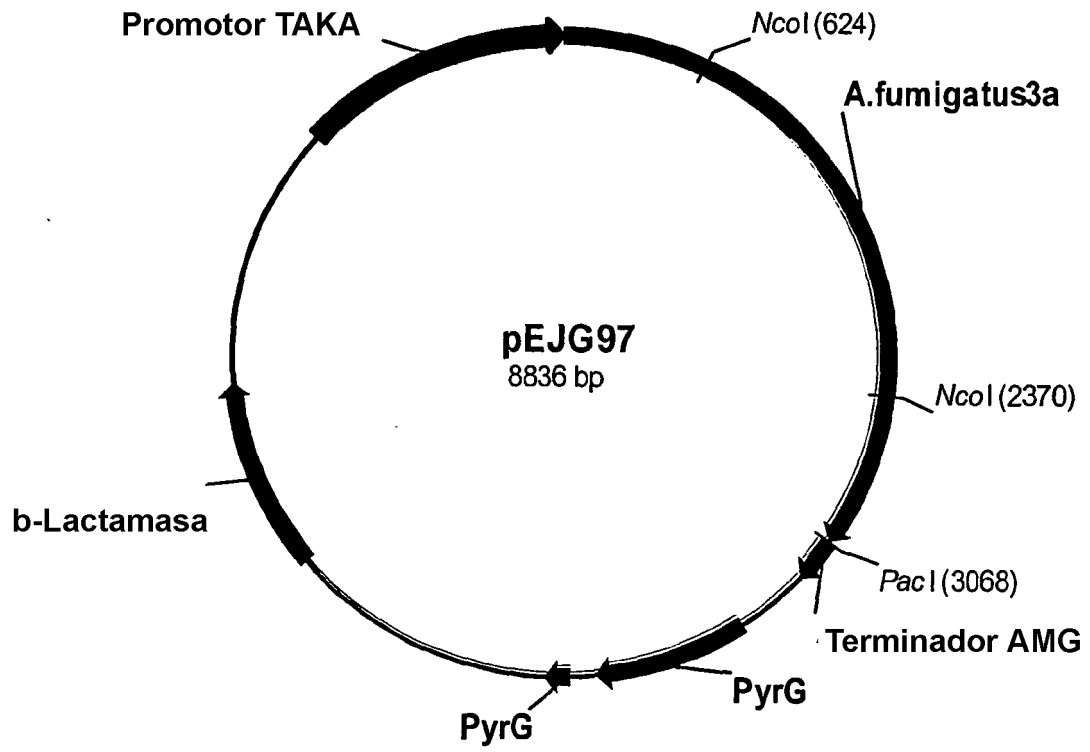


Fig. 14

ATGAGATTCCGTTGGCTCGAGGTGGCCGCTTGTAGCCCAATGCCCCAGGTTTGTGATGCTTTCCCGTCAATTGTTCCGGAATATAGTTGA 100
M R F G W L E V A A L T A A S V A N A Q
 CAATAGTCAATGGAAATAATCAGAAATGGCTTTCTCCACCATTCTACCTTTGCGCTTCCGCTTGGGCTGATGCCCAGGGAGAGTGGCAGATGCCCATCGACGC 200
 E L A F S P P F Y P S P W A D G G Q G E W A D A H R R
 GCCGTCGAGATCGTTTCTCAGATGACACTGGCGGAGAGTTAACCTTACAACAGGGTACTGGGTGGGACTTTTTTGTGTGACAGTGAAGCTTCTTC 300
 A V E I V S Q M T L A E K V N L T G T G
 ACTGACCAATCAGATGGGAAATGGACCGATCGTCCGTTCAAACCGCAGCGTTCCCCAGGTAAGCTTGCAATCTCGCAACAACCGTGCAAAGTGTAGTT 400
 W E M D R C V G Q T G S V P R
 GCTAAAACCGCGTGGTCAGACTGGTATCAACTGGGTTTGTGGCCAGGATCCCTTTGGTATCCGTTTCTGTGAGCTATACCCTGGGAGTCTTT 500
 L G I N W G L C G Q D S P L G I R F
 CAGTCTTGTATTGCTGATGATTGCTCTGTATAGCTGACCTCAAACCTCCGCTTCCCTGTGTTGTTACTAATGTCGCCCGGACATGGGACAAGACT 600
 S D L N S A F P A G T N V A A T W D K T L
 CGCTACCTTCGTGGCAAGGCCATGGGTGAGGAATTC AACGAC AAGGGCTGGACATTTTGTCTGGGCTGTCTGTCTCGGCAATACCCCGGAC 700
 A Y L R G K A M G E F N D K G V D I L L G P A A G P L G K Y P D
 GGCGGCAGAAATCGGGAAGGCTTCTCCTGATCCGTTTCTCAGTGGTACTTTTTCCGCGAAACTATCAAGGGTATCCAAGACGGGGTGTGATTGCTA 800
 G G R I W E G F S P D P V L T G V L F A E T I K G I Q D A G V I A
 CTGCCAAGCATTACATTTGAAATGAACAGGAGCATTTCGACAGGTTGGCGAGGCCAGGGATATGGTTACAACATCACGGAGACGATCAGCTCCAACGT 900
 T A K H Y I L N E Q E H F R Q V G E A Q G Y G Y N I T E T I S S N V
 GGATGACAAGACCATGCACGAGTTGATCCTTTGGTGTGATGACTGCAACTGCAAAATGAGGACCTTGAATTGATTTGACTGACCTGGAAATGCAGGCCCTTTGC 1000
 D D K T M H E L Y L W
 AGATGCTGTGCGGGTAAGATTTCCGTTAGACTTGACCTCGGACGGAAGAAATCGCTGACGAAACCATCGTAGCTGGCGTGGCGCTGTCATGTTGTTCCCTA 1100
 D A V R A G V G A V M C S Y
 CAATCAATCAACAACAGCTACGGTTGTCAAAAACAGTCAAACTCTCAACAGCTCTCAAGGCTGAGCTGGGCTTCCAAGGCTTCGTCATGAGTACTGG 1200
 N Q I N N S Y G C Q N S Q T L N X L L K A E L G F Q G F V M S D W
 AGCGCTCACACAGCGGTGTGGCGCTGCCCTCGTGGTGGATATGTCGATGCCCTGGAGACATTTCCCTCGACGACGACTCTCCTTCTGGGGCACGA 1300
 S A H H S G V G A A L A G L D M S M P G D I S F D D D L S F W G T
 ACCTAACTGTGAGTGTCTTAAACGGCACCGTCCAGCTGGGTGTCGATGACATGGCTTTCGTTATCATGACCGCGTACTACAAGGTTGGTCGTGACCG 1400
 N L T V S V L N G T V P A W R V D D M A V R I M T A Y Y K V G R D R
 TCTTCCGTAATCCCTAACCTCCTGACCTCCGACCCGGGATGAGTACGGCTGGGAGCATTTGCTGTCTCCGAGGGAGCCTGGACCAAGGTGAACGACTTC 1500
 L R I P P N F S S W T R D E Y G W E H S A V S E G A W T K V N D F
 GTCAATGTGACGCGCAGTCACTCAGATCAATCCGTTGAGTGGTCCGCTAGTACAGTCTCTTGAAGAACACCGGGTGTCTTCTTCCCTTACCGGCAAGG 1600
 V N V Q R S H S Q I I R E I G A A S T V L L K N T G A L P L T G K
 AGGTTAAAGTGGTGTCTCCGTTGAGACCGCTGGTTCCAAACCGTGGGTGCTAACCGCTCCCGGCTGATGATAACGGCCTCTTGTGTAIGGC 1700
 E V K V G V L G E D A G S N P W G A N G C P D R G C D N G T L A M A
 CTGGGGTAGTGGTACTGCCAACTTCCCTTACCTTGTCAACCCCGGACGAGGCTATCCAGCGAGAGGTCATCAGCAACCGCGGCAATGTTCTTGTGACT 1800

Fig. 15A

W G S G T A N F P Y L V T P E Q A I Q R E V I S N G G N V F A V T
 GATAACGGGCTCTCAGCCAGATGGCAGATGTTGCATCTCAATCCAGGTGAGTGGGGCTCTTAGAAAAAGAACCGTTCTCTGAATGAAGTTTTTTAAACCA 1900
 D N G A L S Q M A D V A S Q S S
 TTGCGAACAGCGGTGCTTTGGTGTGTTGTCAACGCGGACTCTGGAGAGGTTTCATCAGTGTGACGGCAACGAGGGTACCGCAAAAATCTCACTCTGTG 2000
 V S L V F V N A D S G E G F I S V D G N E G D R K N L T L W
 GAAGAACGGGAGCGGTCATTGACACTGTTGTCAGCCACTGCAACAAACACAGATTGGTTAFTACAGTGTGGCCCGCTTGTGATCGACCCGGTGGTAT 2100
 K N G E A V I D T V S H C N N T I V I H S V G P V L I D R W Y
 GATAACCCCAACGTACTGCCATCATCTGGGCCGGCTTGCCTGGTACGAGAGTGGCAACTCCCTGGTGCAGCTCTATGGCCCGCTCAACCCACCCAGCG 2200
 D N P N V T A I I W A G L P G Q E S G N S L V D V L Y G R V N P S
 CCAAGACCCCGTTTCCCTGGGCAAGACTCGGAGTCTTACGGGGCTCCCTGTCTCACCGAGCCTAACAAATGGCAATGGTGTCTCCCCAGGATGATTTCAA 2300
 A K T P F T W G K T R E S Y G A P L L T E P N N G N G A P Q D D F N
 CGAGGGCGTCTTCAATTGACTACCGTCACTTTGACAAGCGCAATGAGACCCCATTTAAGTGTGGCCATGGCTTGAGCTACACCCACCTTTGGTFACTCT 2400
 E G V F I D Y R H F D K R N E T P I Y E F G H G L S Y T T F G Y S
 CACCTTCGGGTTCCAGCCCTCAATAGTTCGAGTTCGGCATAATGTCCCGACTAGCGGAGAGACCAAGCCTGCGCCCAACCTATGGTGAGATCGGTAGTGCCG 2500
 H L R V Q A L N S S S A Y V P T S G E T K P A P T Y G E I G S A
 CCGACTACCTGTATCCGGAGGGTCTCAAAAGAATTACCAAGTTTTATTACCCCTTGGCTCAACTCGACCGACCTCGAGGATTTCTTGACGACCCCGAACTA 2600
 A D Y L Y P E G L K R I T K F I Y P W L N S T D L E D S S D D P N Y
 CGGTGGGAGGACTCGGAGTACATTCGCCAAGGCGCTAGGGATGGGTCTCCCTCAACCCCTCTGAAGGCTGGCCGGCTCCTGGTGGTAAACCCCTACCCCTT 2700
 G W E D S E Y I P E G A R D G S P Q P L L K A G G A P G G N P T L
 TATCAGGATCTTGTAGGGTGTGGCCACCATAACCAACACTGGTAAACGTCGCCGGTTATGAAGTCCCTCAATTTGGTGAAGTACCCCGCATGTTCTTGGG 2800
 Y Q D L V R V S A T I T N T G N V A G Y E V P Q L
 TTGCAATTTGGCTAACTCGCTTCTAGTATGTTTCACTGGGCGGACCGAACGAGCCCTCGGGTGTCTTGGCAAGTTTCGACCCGAATCTTCTGGCTCTGG 2900
 Y V S L G G P N E P R V V L R K F D R I F L A P G
 GGAGCAAAAGGTTTTGGACCACGACTCTTAACCGTCTGTGATCTCGCCAAATGGGATGTGGAGGCTCAGGACTGGGTCAACAAAGTACCCCAAGAAAGTG 3000
 E Q K V W T T T L N R R D L A N W D V E A Q D W V I T K Y P K K V
 CAGTCCGGCAGCTCCTCGCGTAAGCTGCCTCTGAGAGCGCCCTCTGCCCGTGTCTACTAG 3060
 H V G S S S R K L P L R A P L P R R V Y .

Fig. 15B

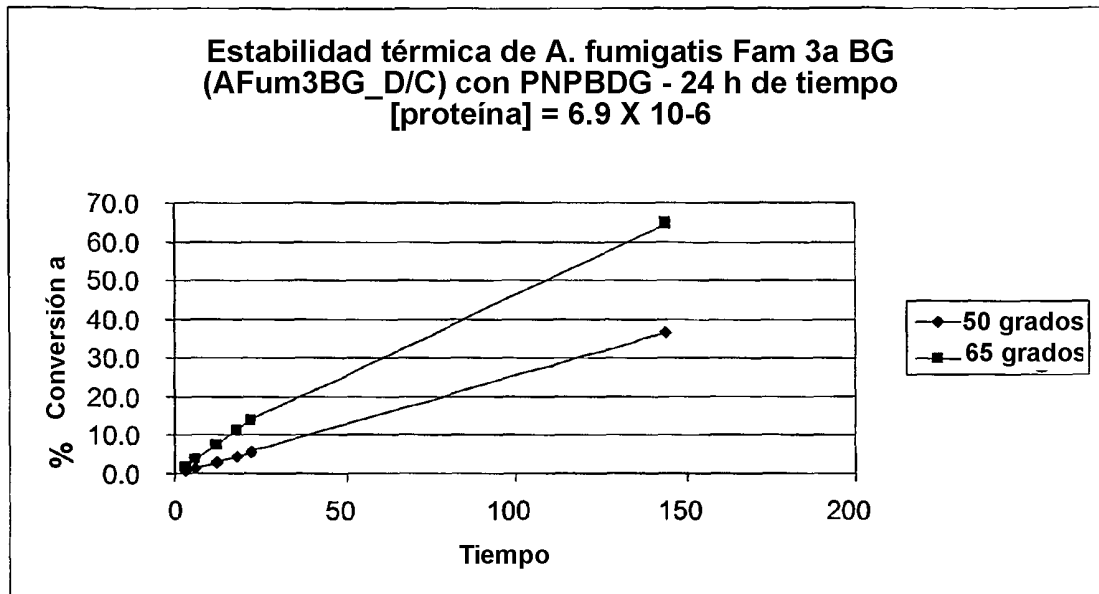


Fig. 16

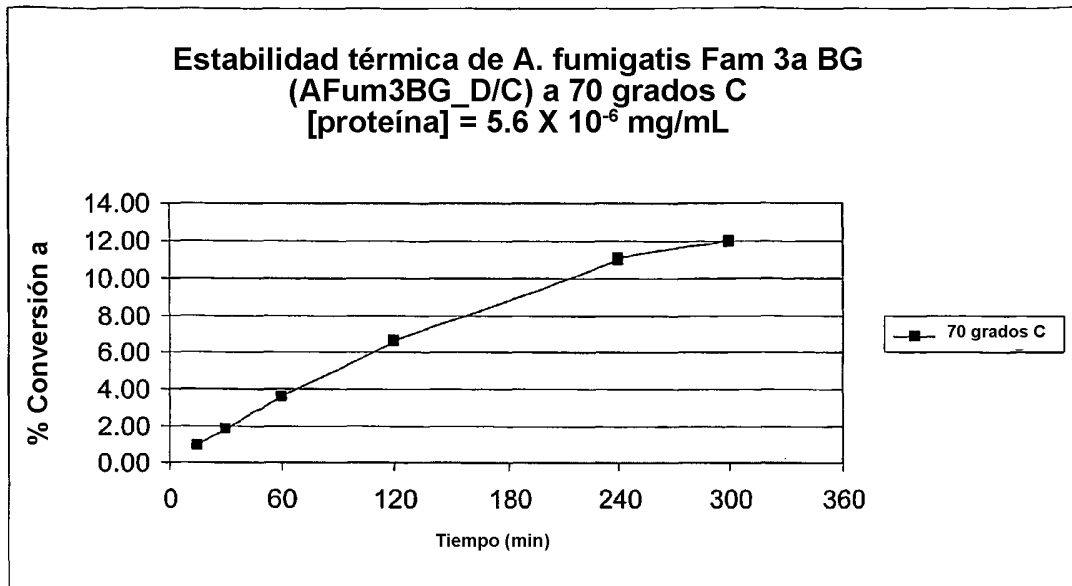


Fig. 17

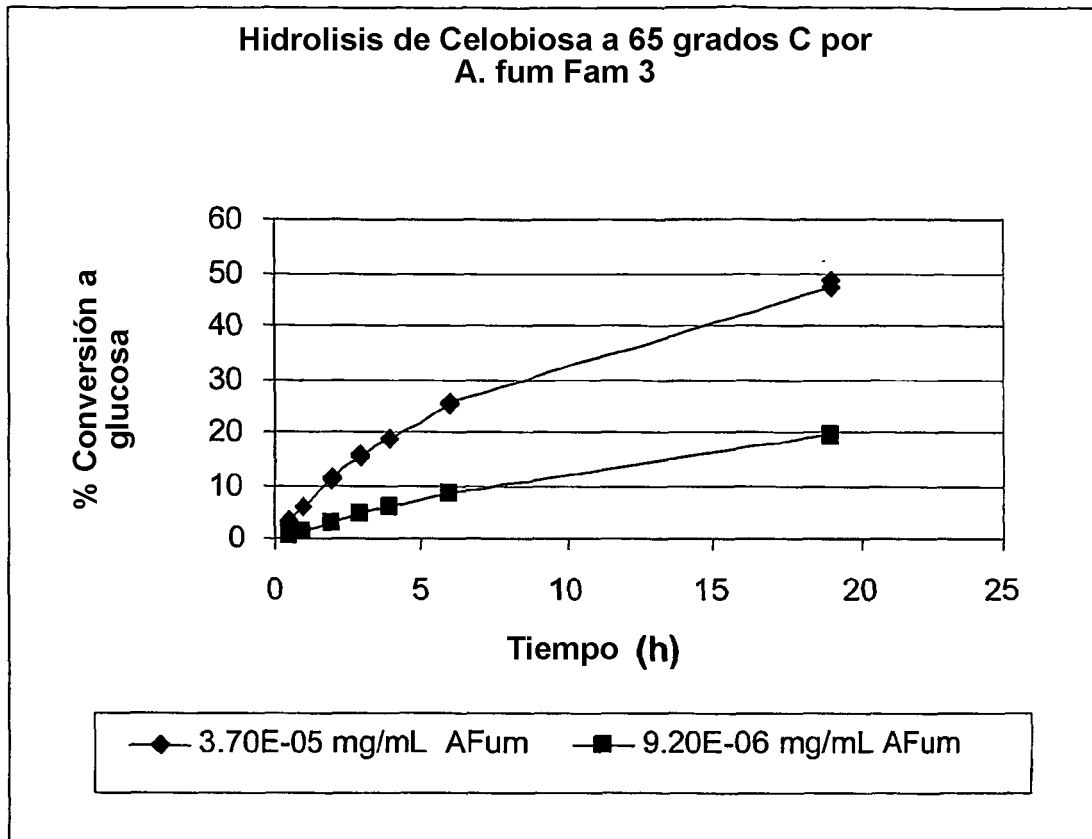


Fig. 18

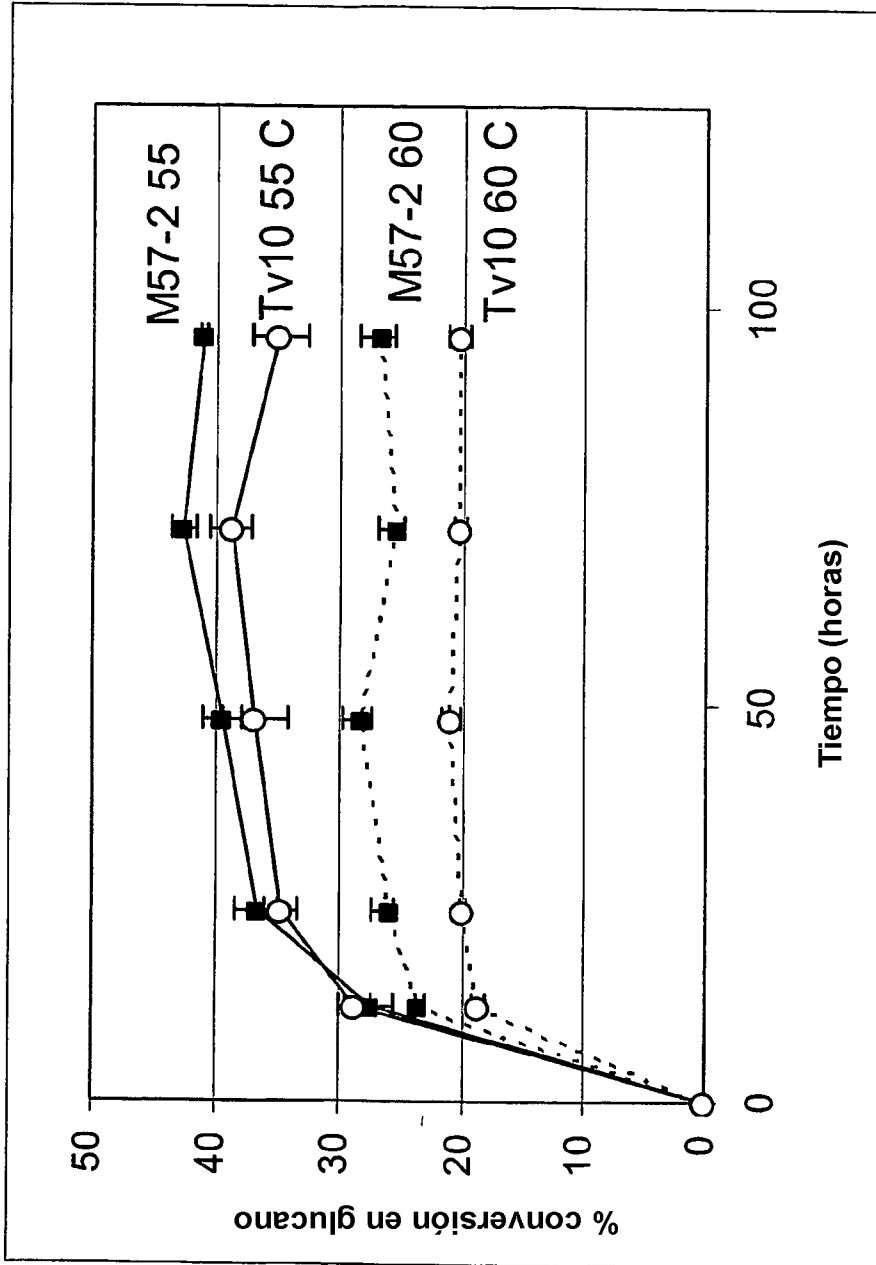


Fig. 19