

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 437 198**

51 Int. Cl.:

C12N 9/24 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **28.10.2004 E 04816951 (0)**

97 Fecha y número de publicación de la concesión europea: **18.09.2013 EP 1682656**

54 Título: **Polipéptidos con actividad de beta-glucosidasa y polinucleótidos aislados que codifican los polipéptidos**

30 Prioridad:

28.10.2003 US 515482 P

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

09.01.2014

73 Titular/es:

**NOVOZYMES INC. (100.0%)
1445 DREW AVENUE
DAVIS, CA 95616, US**

72 Inventor/es:

**HARRIS, PAUL y
GOLIGHTLY, ELIZABETH**

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 2 437 198 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

Polipéptidos con actividad de beta-glucosidasa y polinucleótidos que codifican dichos polipéptidos

5 **Antecedentes de la invención****Campo de la invención**

10 [0001] La presente invención se refiere a polipéptidos aislados con actividad de beta-glucosidasa y polinucleótidos aislados que codifican los polipéptidos. La invención también se refiere a constructos de ácidos nucleicos, vectores y células huésped que comprenden tanto los polinucleótidos como los métodos para producir y usar los polipéptidos.

Descripción de las técnicas relacionadas

15 [0002] La celulosa es un polímero de la glucosa de azúcar simple unida covalentemente por enlaces beta-1,4. Muchos microorganismos producen enzimas que hidrolizan beta-glucanos. Estas enzimas incluyen endoglucanasas, celobiohidrolasas y beta-glucosidasas. Las endoglucanasas digieren el polímero de celulosa en lugares aleatorios, exponiéndolo al ataque por parte de celobiohidrolasas. Las celobiohidrolasas liberan secuencialmente moléculas de celobiosa de los extremos del polímero de celulosa. La celobiosa es un dímero de glucosa de enlace beta 1,4 hidrosoluble. Las beta-glucosidasas hidrolizan celobiosa a glucosa.

20 [0003] La conversión de materias primas celulósicas en etanol tiene como ventajas la disponibilidad de grandes cantidades de materia prima, la conveniencia de evitar la quema o el vertido de los materiales y la limpieza del combustible de etanol. La madera, residuos agrícolas, los brotes herbáceos y los residuos sólidos municipales han sido considerados materias primas para la producción de etanol. Estos materiales consisten principalmente en celulosa, hemicelulosa y lignina. Una vez que la celulosa se convierte en glucosa, esta es fácilmente fermentada a etanol usando levadura. Ya que la glucosa se fermenta fácilmente a etanol usando una variedad de levaduras mientras que la celobiosa no, cualquier celobiosa restante al final de la hidrólisis representa una pérdida de rendimiento del etanol. Es más, la celobiosa es un inhibidor potente de endoglucanasas y celobiohidrolasas. La acumulación de celobiosa durante la hidrólisis es extremadamente indeseable en la producción de etanol.

25 [0004] La acumulación de celobiosa ha sido un problema importante en la hidrólisis enzimática porque los microorganismos productores de celulasa pueden producir poca beta-glucosidasa. La baja cantidad de beta-glucosidasa produce un déficit de la capacidad para hidrolizar la celobiosa a glucosa. Se han utilizado diferentes métodos para aumentar la cantidad de beta-glucosidasa en la conversión de celulosa a glucosa.

30 [0005] Un método es producir beta-glucosidasa usando microorganismos que producen poca celulasa y añadir la beta-glucosidasa exógenamente a la endoglucanasa y la celobiohidrolasa para mejorar la hidrólisis. No obstante, las cantidades requeridas son demasiado costosas para un uso comercial de la biomasa en una operación de etanol.

35 [0006] Un segundo método es llevar a cabo la hidrólisis de celulasa simultáneamente con la fermentación de la glucosa usando levadura. Este proceso es conocido como sacarificación y fermentación simultáneas (SSF). En un sistema SSF, la fermentación de la glucosa elimina esta de la solución. No obstante, los sistemas SSF están comercialmente disponibles debido a que la temperatura operativa para la levadura de 28 °C es demasiado baja para las condiciones requeridas de 50 °C.

40 [0007] Un tercer método para superar la escasez de beta-glucosidasa es sobreexpresar la beta-glucosidasa en un huésped, aumentando así el rendimiento de la beta-glucosidasa.

45 [0008] Ximenes et al., 1996, Current Microbiology 32: 119-123; Hang y Woodams, 1994, Lebensmittel-Wissenschaft and Technologie 27: 587-589; y Kitpreechavanich et al., Agricultural and Biological Chemistry 50: 1703- 1712, revelan la existencia de beta-glucosidasas de *Aspergillus fumigatus*.

50 [0009] Sería muy ventajoso para la técnica usar beta-glucosidasa termoestable para convertir materiales celulósicos en monosacáridos, disacáridos y polisacáridos y así mejorar la eficacia del proceso.

55 [0010] Es un objeto de la presente invención proporcionar nuevos polipéptidos con actividad de beta-glucosidasa y ácido nucleico que codifiquen los polipéptidos.

60 **Resumen de la invención**

[0011] La presente invención se refiere a polipéptidos aislados con actividad de beta-glucosidasa seleccionados de un polipéptido que consiste en una secuencia de aminoácidos que tiene al menos el 90% de identidad con los aminoácidos 20 a 863 de la SEC ID n°: 2.

65 [0012] La presente invención también se refiere a polinucleótidos aislados que codifican polipéptidos con actividad de

beta-glucosidasa.

[0013] La presente invención también se refiere a constructos de ácidos nucleicos, vectores de expresión recombinantes y células huésped recombinantes que comprenden los polinucleótidos.

[0014] La presente invención también se refiere a métodos para producir tales polipéptidos con actividad de beta-glucosidasa que comprenden (a) el cultivo de una célula huésped recombinante que comprende un constructo de ácidos nucleicos que, a su vez, comprende un polinucleótido que codifica el polipéptido en las condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido.

[0015] La presente invención se refiere además al uso de los polipéptidos con actividad de beta-glucosidasa para convertir materiales celulósicos en monosacáridos, disacáridos y polisacáridos.

Breve descripción de las figuras

[0016]

Las **Figuras 1A** y **1B** muestran la secuencia de ADN genómico y la secuencia de aminoácidos deducida de una beta-glucosidasa de *Aspergillus fumigatus* (SEC ID n°: 1 y 2, respectivamente). El péptido señal predicho aparece subrayado y los intrones en cursiva.

La **Figura 2** muestra un mapa de restricción de pAIIo1.

La **Figura 3** muestra un mapa de restricción de pBANE10.

La **Figura 4** muestra un mapa de restricción de pAIIo2.

La **Figura 5** muestra un mapa de restricción de pEJG97.

La **Figura 6** muestra un mapa de restricción de pMJ04.

La **Figura 7** muestra un mapa de restricción de pMJ06.

La **Figura 8** muestra un mapa de restricción de pMJ09.

La **Figura 9** muestra un mapa de restricción de pEJG107.

La **Figura 10** muestra la termoestabilidad de la beta-glucosidasa de *Aspergillus fumigatus* a 50 °C y 65 °C.

La **Figura 11** muestra la termoestabilidad de la beta-glucosidasa de *Aspergillus fumigatus* a 70 °C.

La **Figura 12** muestra la hidrólisis de la celobiosa por beta-glucosidasa de *Aspergillus fumigatus* a 65 °C.

Definiciones

[0017] **La actividad de beta-glucosidasa:** el término "beta-glucosidasa" es definido en la presente como una glucohidrolasa de beta-D-glucósido (E.C. 3.2.1.21) que cataliza la hidrólisis de residuos terminales no reductores de beta-D-glucosa con la liberación de beta-D-glucosa. Para fines de la presente invención, la actividad de beta-glucosidasa se determina según el procedimiento básico descrito por Venturi et al., 2002, J. Basic Microbial. 42: 55-66, a menos que se emplearan condiciones diferentes a las descritas aquí. Una unidad de actividad de beta-glucosidasa es definida como 1,0 μmol de p-nitrofenol producido por minuto a 50 °C usando pH 5 de 4 mM de p-nitrofenol-beta-D-glucopiranosido como sustrato en 100 mM de citrato sódico, Tween-20 al 0,01%.

[0018] **Beta-glucosidasa de la familia GH3AA:** el término "beta-glucosidasa de la familia GH3AA" es definido en la presente como una hidrolasa glucósida de la familia 3 según Coutinho, P.M. y Henrissat, B., 1999, Carbohydrate enzymes: an integrated database approach, en "Recent Advances in Carbohydrate Bioengineering", H.J. Gilberto, G. Davies, B. Henrissat y B. Svensson eds., The Royal Society of Chemistry, Cambridge, págs. 3-12.

[0019] Los polipéptidos de la presente invención tienen al menos el 20%, preferiblemente al menos el 40%, más preferiblemente al menos el 50%, más preferiblemente al menos el 60%, más preferiblemente al menos el 70%, más preferiblemente al menos el 80%, incluso más preferiblemente al menos el 90%, más preferiblemente al menos el 95%, e incluso más preferiblemente al menos el 100% de la actividad de beta-glucosidasa del polipéptido que consiste en la secuencia de aminoácidos mostrada como los aminoácidos 20 a 863 de la SEC ID n°: 2.

[0020] **Polipéptido aislado:** el término "polipéptido aislado" según se utiliza en este caso se refiere a un polipéptido que es al menos un 20% puro, preferiblemente al menos un 40% puro, más preferiblemente al menos un 60%, puro, incluso

más preferiblemente al menos un 80% puro, más preferiblemente al menos 90% puro, e incluso más preferiblemente al menos un 95% puro, como determinada la SDS-PAGE.

5 **[0021] Polipéptido substancialmente puro:** el término "polipéptido substancialmente puro" denota en la presente una preparación de polipéptidos que contiene como máximo el 10%, preferiblemente como máximo el 8%, más preferiblemente como máximo el 6%, más preferiblemente como máximo el 5%, más preferiblemente como máximo el 4%, como máximo el 3%, incluso más preferiblemente como máximo el 2%, más preferiblemente como máximo el 1% e incluso más preferiblemente como máximo el 0,5% en peso de otro material polipeptídico con el cual está asociado originalmente. Es, por lo tanto, preferible que el polipéptido substancialmente puro sea al menos un 92% puro, preferiblemente al menos un 94% puro, más preferiblemente al menos un 95% puro, más preferiblemente al menos un 96% puro, más preferiblemente al menos un 96% puro, más preferiblemente al menos un 97% puro, más preferiblemente al menos un 98% puro, incluso más preferiblemente al menos un 99%, más preferiblemente al menos un 99,5% puro e incluso más preferiblemente un 100% puro en peso del material de polipéptido total presente en la preparación.

15 **[0022]** Los polipéptidos de la presente invención están preferiblemente en una forma substancialmente pura. En particular, se prefiere que los polipéptidos estén en "una forma esencialmente pura", es decir, que la preparación de polipéptidos esté esencialmente libre de otro material polipeptídico con el cual esté asociado originalmente. Este se puede conseguir, por ejemplo, preparando el polipéptido mediante métodos recombinantes conocidos o mediante métodos de purificación tradicionales.

20 **[0023]** Aquí, el término "polipéptido substancialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada"

25 **[0024] Identidad:** la relación entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos está descrita por el parámetro "identidad".

30 **[0025]** Para fines 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 software LASERGENE™ MEGALIN™ (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 por parejas son Ktuple=1, penalización de espacio=3, ventanas=5, y diagonales=5.

35 **[0026]** Para fines de la presente invención, el grado de identidad entre dos secuencias de nucleótidos se determina por el método de Wilbur-Lipman (Wilbur y Lipman, 1983, Proceedings of the National Academy of Science USA 80: 726-730) usando el software LASERGENE™ MEGALIGN™ (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 por parejas son Ktuple=3, penalización de espacio=3, y ventanas=20.

40 **[0027] Fragmento polipeptídico:** el término "fragmento polipeptídico" es definido en la presente como un polipéptido que tiene uno o más aminoácidos delecionados del terminal amino y/o carboxilo de la SEC ID n°: 2 o una secuencia homóloga del mismo, donde el fragmento tiene actividad de beta-glucosidasa. Preferiblemente, un fragmento contiene al menos 770 residuos de aminoácidos, más preferiblemente al menos 800 residuos de aminoácidos y más preferiblemente al menos 830 residuos de aminoácidos.

45 **[0028] Subsecuencia:** el término "subsecuencia" es definido en la presente como una secuencia de nucleótidos que tiene uno o más nucleótidos delecionados del extremo 5' y/o 3' de la SEC ID n°: 1 o una secuencia homóloga de la misma, donde la subsecuencia codifica un fragmento polipeptídico con actividad de beta-glucosidasa. Preferiblemente, una subsecuencia contiene al menos 2310 nucleótidos, más preferiblemente al menos 2400 nucleótidos, y más preferiblemente al menos 2490 nucleótidos.

50 **[0029] Variante alélica:** el término "variante alélica" denota en la presente cualquiera de las 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 derivar en polimorfismo dentro de las poblaciones. Las mutaciones de genes pueden ser silenciosas (sin 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.

55 **[0030] Polinucleótido aislado:** el término "polinucleótido aislado" según se utiliza en este caso se refiere a un polinucleótido que es al menos un 20% puro, preferiblemente al menos un 40% puro, más preferiblemente al menos un 60% puro, incluso más preferiblemente al menos un 80% puro, más preferiblemente al menos un 90% puro, e incluso más preferiblemente al menos un 95% puro, como determina la electroforesis de agarosa.

60 **[0031] Polinucleótido substancialmente puro:** el término "polinucleótido substancialmente puro" según se utiliza en este caso se refiere a una preparación de polinucleótido libre de nucleótidos extraños o indeseados y en una forma adecuada para el uso dentro de los sistemas de producción de proteínas diseñadas genéticamente. Así, un polinucleótido substancialmente puro contiene como máximo el 10%, preferiblemente como máximo el 8%, más

65

preferiblemente como máximo el 6%, más preferiblemente como máximo el 5%, más preferiblemente como máximo el 4%, más preferiblemente como máximo el 3%, incluso más preferiblemente como máximo el 2%, más preferiblemente como máximo el 1%, e incluso más preferiblemente como máximo el 0,5% en peso de otro material polinucleótido con el cual está originalmente asociado. No obstante, un polinucleótido substancialmente puro puede incluir regiones no traducidas 5' y 3' de origen natural, tales como promotores y terminadores. Se prefiere que el polinucleótido substancialmente puro sea al menos un 90% puro, preferiblemente al menos un 92% puro, más preferiblemente al menos un 94% puro, más preferiblemente al menos un 95% puro, más preferiblemente al menos un 96% puro, más preferiblemente al menos un 97% puro, incluso más preferiblemente al menos un 98% puro, más preferiblemente al menos un 99% puro, e incluso más preferiblemente al menos un 99,5% puro en peso. Los polinucleótidos de la presente invención están preferiblemente en una forma substancialmente pura. En particular, se prefiere que los polinucleótidos descritos en la presente estén en "forma esencialmente pura", es decir, que la preparación de polinucleótidos esté esencialmente libre de otra materia polinucleótida con la cual está asociado originalmente. En la presente, el término "polinucleótido sustancialmente puro" es sinónimo de los términos "polinucleótico aislado" y "polinucleótido en forma aislada". Los polinucleótidos pueden ser de origen genómico ADNc, ARN, semisintético, sintético o cualquier combinación de los mismos.

[0032] ADNc: el término "ADNc" es definido en la presente como una molécula de ADN que puede ser preparada por transcripción inversa de una molécula de ARNm madura y empalmada obtenida de una célula eucariótica. El ADNc no tiene secuencias de intrón que normalmente están presentes en el ADN genómico correspondiente. La transcripción de ARN inicial primaria es un precursor de ARNm que se procesa a través de una serie de fases antes de aparecer como ARN maduro empalmado. Estas fases incluyen la eliminación de secuencias de intrón por un proceso llamado empalme. El ADNc derivado de ARNm, en consecuencia, carece de cualquier secuencia de intrones.

[0033] Constructo de ácidos nucleicos: el término "constructo de ácidos nucleicos" según se utiliza en este caso se refiere a una molécula de ácido nucleico, única o bicatenaria, que es aislada de un gen de origen natural o modificada para contener segmentos de ácidos nucleicos que de otra manera no existirían en la naturaleza. El término constructo de ácidos nucleicos es sinónimo del término "cassette de expresión" cuando el constructo de ácidos nucleicos contiene las secuencias de control requeridas para la expresión de una secuencia codificante de la presente invención.

[0034] Secuencia de control: el término "secuencia de control" se refiere en la presente a todos los componentes, que son necesarios o ventajosos para la expresión de un polinucleótido que codifica un polipéptido de la presente invención. Cada secuencia de control puede ser nativa o extraña a la secuencia de nucleótidos que codifica el polipéptido. Tales secuencias de control incluyen, entre otras, una guía, secuencia de poliadenilación, una secuencia de propéptido, un promotor, una secuencia de péptido señal y un terminador de la transcripción. Como mínimo, las secuencias de control incluyen un promotor y señales de terminación transcripcionales y traduccionales. Las secuencias de control pueden ser provistas de 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 un polipéptido.

[0035] Operativamente enlazado: el término "operativamente enlazado" denota en la presente una configuración en la cual una secuencia de control se coloca en una posición apropiada en relación a la secuencia codificante de la secuencia polinucleótida de tal manera que la secuencia de control dirige la expresión de la secuencia codificante de un polipéptido.

[0036] Secuencia codificante: cuando se usa en la presente, el término "secuencia codificante" se refiere a una secuencia de nucleótidos que directamente especifica la secuencia de aminoácidos de su producto proteico. Los límites de la secuencia codificante son determinados generalmente por un marco de lectura abierto, que normalmente se inicia con un codón de iniciación ATG o codones de iniciación alternativos tales como GTG y TTG. La secuencia codificante puede ser una secuencia de ADN, ADNc o de nucleótidos recombinantes.

[0037] Expresión: el término "expresión" incluye cualquier fase implicada en la producción del polipéptido, incluidas, entre otras, la transcripción, la modificación postranscripcional, la traducción, la modificación postraduccional y la secreción.

[0038] Vector de expresión: el término "vector de expresión" es definido en la presente como una molécula de ADN lineal o circular que comprende un polinucleótido que codifica un polipéptido de la invención y que está operativamente enlazado a nucleótidos adicionales que permiten su expresión.

[0039] Célula huésped: el término "célula huésped", según se utiliza en este caso, incluye cualquier tipo de célula susceptible de transformación, transfección, transducción y similares con un constructo de ácidos nucleicos o vector de expresión que comprenda un polinucleótido de la presente invención.

[0040] Modificación: el término "modificación" significa en la presente cualquier modificación química del polipéptido consistente en los aminoácidos 20 a 863 de la SEC ID nº: 2 o una secuencia homóloga de la misma, así como la manipulación genética del ADN que codifica ese polipéptido. La modificación puede consistir en sustituciones, deleciones y/o inserciones de uno o más aminoácidos al igual que sustituciones de una o más cadenas laterales del aminoácido.

[0041] Variante artificial: cuando se usa en la presente, el término "variante artificial" se refiere a un polipéptido con actividad de beta-glucosidasa producida por un organismo que expresa una secuencia de nucleótidos modificada de la SEC ID n°: 1 o una secuencia homóloga de la misma o la región de codificación madura de la misma. La secuencia de nucleótidos modificada se obtiene mediante la intervención humana a través de la modificación de la secuencia de nucleótidos descrita en la SEC ID n°: 1 o una secuencia homóloga de la misma o la región de codificación madura de la misma.

Descripción detallada de la invención

Polipéptidos con actividad de beta-glucosidasa

[0042] En un primer aspecto, la presente invención se refiere a polipéptidos aislados que comprenden una secuencia de aminoácidos con un grado de identidad a los aminoácidos 20 a 863 de la SEC ID n°: 2 (es decir, el polipéptido maduro) de al menos el 90%, más preferiblemente al menos el 95%, e incluso más preferiblemente al menos el 97%, que tienen actividad de beta-glucosidasa (en adelante, "polipéptidos homólogos"). En un aspecto preferido, los polipéptidos homólogos tienen una secuencia de aminoácidos que difiere por diez aminoácidos, preferiblemente por cinco aminoácidos, más preferiblemente por cuatro aminoácidos, incluso más preferiblemente por tres aminoácidos, más preferiblemente por dos aminoácidos e incluso más preferiblemente por un aminoácido, de los aminoácidos 20 a 863 de la SEC ID n°: 2.

[0043] Un polipéptido de la presente invención comprende preferiblemente la secuencia de aminoácidos de la SEC ID n°: 2 o una variante alélica de la misma; o un fragmento de la misma que tenga actividad de beta-glucosidasa. En un aspecto preferido, un polipéptido comprende la secuencia de aminoácidos de la SEC ID n°: 2. En otro aspecto preferido, un polipéptido comprende los aminoácidos 20 a 863 de la SEC ID n°: 2. En otro aspecto preferido, un polipéptido comprende los aminoácidos 20 a 863 de la SEC ID n°: 2. En otro aspecto preferido, un polipéptido consiste en la secuencia de aminoácidos de la SEC ID n°: 2. En otro aspecto preferido, un polipéptido consiste en la secuencia de aminoácidos de la SEC ID n°: 2. En otro aspecto preferido, un polipéptido consiste en los aminoácidos 20 a 863 de la SEC ID n°: 2.

[0044] La secuencia de nucleótidos de la SEC ID n°: 1 o una subsecuencia de la misma, al igual que la secuencia de aminoácidos de la SEC ID n°: 2 o un fragmento de la misma, puede ser utilizada para diseñar una sonda de ácido nucleico para identificar y clonar ADN que contiene polipéptidos con actividad de beta-glucosidasa de cepas de diferentes géneros o especies según métodos bien conocidos en la técnica. En particular, tales sondas se pueden usar para la hibridación con el genómico o ADNc del género o especies de interés, siguiendo procedimientos estándar de Southern blot, para identificar y aislar el gen correspondiente allí presente. Tales sondas pueden ser considerablemente más cortas que la secuencia entera, pero deben tener al menos 14, preferiblemente al menos 25, más preferiblemente al menos 35, y de la forma más preferible al menos 70 nucleótidos de longitud. No obstante, se prefiere que la sonda de ácido nucleico tenga al menos 100 nucleótidos de longitud. Por ejemplo, la sonda de ácido nucleico puede tener al menos 200 nucleótidos, preferiblemente al menos 300 nucleótidos, más preferiblemente al menos 400 nucleótidos, o más preferiblemente al menos 500 nucleótidos de longitud. Incluso se pueden utilizar sondas más largas, por ejemplo, sondas de ácido nucleico que tengan al menos 600 nucleótidos, preferiblemente al menos 700 nucleótidos, más preferiblemente al menos 800 nucleótidos, o más preferiblemente al menos 900 nucleótidos de longitud. Se pueden utilizar tanto sondas de ADN como de ARN. Las sondas normalmente están marcadas para detectar el gen correspondiente (por ejemplo, con ³²P, ³H, ³⁵S, biotina o avidina). Tales sondas se incluyen en la presente invención.

[0045] Un ADN genómico o genoteca de ADNc obtenidos a partir de estos otros organismos pueden, por lo tanto, ser seleccionados para ADN que se hibrida con las sondas anteriormente descritas y que codifica un polipéptido con actividad de beta-glucosidasa. El ADN genómico u otro ADN de estos otros organismos puede ser separado por electroforesis en gel de agarosa o poliacrilamida, u otras técnicas de separación. El ADN de las genotecas o el ADN separado puede ser transferido e inmovilizado en nitrocelulosa u otro material portador adecuado. Para identificar un clon o ADN homólogo con la SEC ID n°: 1 o una subsecuencia de la misma, se usa el material portador en una transferencia de Southern.

[0046] Para fines de la presente invención, la hibridación indica que la secuencia de nucleótidos se hibrida a una sonda de ácido nucleico marcada correspondiente a la secuencia de nucleótidos mostrada en la SEC ID n°: 1, la secuencia de ADNc contenida en la SEC ID n°: 1, su cadena complementaria, o una subsecuencia de la misma, en condiciones de astringencia desde muy bajas a muy altas. Las moléculas a las que se hibrida la sonda de ácido nucleico en estas condiciones pueden ser detectadas usando película radiográfica.

[0047] En un aspecto preferido, la sonda de ácido nucleico es de nucleótidos 58 a 2580 de la SEC ID n°: 1. En otro aspecto preferido, la sonda de ácido nucleico es una secuencia polinucleótida que codifica el polipéptido de la SEC ID n°: 2, o una subsecuencia de la misma. En otro aspecto preferido, la sonda de ácido nucleico es la SEC ID n°: 1 o su cadena complementaria. En otro aspecto preferido, la sonda de ácido nucleico es la región de codificación del polipéptido maduro de la SEC ID n°: 1. En otro aspecto preferido, la sonda de ácido nucleico es la secuencia de polinucleótidos contenida en plásmido pEJG113, contenido a su vez en *Escherichia coli* NRRL B-30695, donde la

secuencia polinucleótida de la misma codifica un polipéptido con actividad de beta-glucosidasa. En otro aspecto preferido, la sonda de ácido nucleico es la región de codificación del polipéptido maduro contenida en el plásmido pEJG113, contenido a su vez en *Escherichia coli* NRRL B-30695.

5 **[0048]** Para sondas largas de al menos 100 nucleótidos de longitud, las condiciones de astringencia muy altas o muy bajas son definidas como prehibridación e hibridación a 42 °C en 5X

10 **[0049]** SPE, 0,3% de SDS, 200 µg/ml de ADN de esperma de salmón cortado y desnaturalizado, y o bien 25% de formamida para astringencias muy bajas y bajas, 35% de formamida para astringencias medias y medias-altas, o 50% de formamida para astringencias altas y muy altas, siguiendo los procedimientos estándar de Southern blot durante 12 a 24 horas óptimamente.

15 **[0050]** Para sondas largas de al menos 100 nucleótidos de longitud, el material portador es lavado finalmente tres veces cada uno durante 15 minutos usando 2X SSC, 0,2% de SDS preferiblemente al menos a 45 °C (astringencia muy baja), más 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), incluso más preferiblemente al menos a 65 °C (astringencia alta), y más preferiblemente al menos a 70 °C (astringencia muy alta).

20 **[0051]** 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 de post-hibridación de aproximadamente 5 °C a aproximadamente 10 °C por debajo de la T_m calculada usando el cálculo según Bolton y McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) en 0,9 M de NaCl, 0,09 M de Tris-HCl pH 7,6, 6 mM de EDTA, 0,5% NP-40, 1X solución de Denhardt, 1 mM de pirofosfato de sodio, 1 mM fosfato monobásico de sodio, 0,1 mM ATP, y 0,2 mg de levadura ARN por ml siguiendo procedimientos estándar de Southern blot durante 25 12 a 24 horas óptimamente.

30 **[0052]** 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 SCC más 0,1% SDS durante 15 minutos y dos veces cada uno durante 15 minutos usando 6X SSC a 5 °C hasta 10 °C por debajo de la T_m calculada.

35 **[0053]** La presente divulgación se refiere a variantes artificiales que comprenden una sustitución, delección y/o inserción conservadora de uno o más aminoácidos de la SEC ID n°: 2 o una secuencia homóloga de la misma; o el polipéptido maduro de la misma. Preferiblemente, los cambios de los aminoácidos son de una naturaleza menor, es decir, sustituciones o inserciones conservadoras de aminoácidos que no afectan significativamente al pliegue y/o actividad de la proteína; delecciones pequeñas, normalmente de uno a aproximadamente 30 aminoácidos; pequeñas extensiones amino-terminales o carboxi-terminales, tales como un residuo de metionina aminoterminal; un péptido de enlace pequeño de hasta 20-25 residuos aproximadamente; o una pequeña extensión que facilita la purificación cambiando la carga neta u otra función, como un tracto de polihistidina, un epítipo antigénico o un dominio de unión.

40 **[0054]** Los ejemplos de sustituciones conservadoras están dentro del grupo de aminoácidos básicos (arginina, lisina e histidina), aminoácidos ácidos (ácido glutámico y ácido aspártico), aminoácidos polares (glutamina y asparagina), aminoácidos hidrofóbicos (leucina, isoleucina y valina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina), y aminoácidos pequeños (glicina, alanina, serina, treonina y metionina). En la técnica se conocen sustituciones de aminoácidos que generalmente no alteran la actividad específica y son descritas, por ejemplo, por H. Neurath y R.L. Hill, 45 1979, en *The Proteins*, Academic Press, Nueva York. Los cambios más frecuentes 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.

50 **[0055]** Además de los 20 aminoácidos estándar, los aminoácidos no estándar (tales como 4-hidroxiprolina, 6-N-metil lisina, ácido 2-aminoisobutírico, isovalina, y alfa-metil serina) pueden ser sustituidos por residuos de aminoácidos de un polipéptido de tipo salvaje. Un número limitado de aminoácidos no conservadores, los aminoácidos que no se codifican por el código genético y los aminoácidos no naturales pueden ser sustituidos por residuos de aminoácidos. Los "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(s) de los aminoácidos estándar. Los aminoácidos no naturales pueden ser 55 sintetizados químicamente y, preferiblemente, están disponibles comercialmente, e incluyen ácido pipercolico, ácido carboxílico de tiazolidina, dehidroprolina, 3- y 4-metilprolina, y 3,3-dimetilprolina.

60 **[0056]** De forma alternativa, los cambios de los aminoácidos son de una naturaleza tal que las propiedades fisicoquímicas de los polipéptidos son alteradas. Por ejemplo, cambios en los aminoácidos pueden mejorar la termoestabilidad del polipéptido, alterar la especificidad de sustrato, cambiar el pH óptimo y similares.

65 **[0057]** Los aminoácidos esenciales en el polipéptido progenitor pueden ser identificados según procedimientos conocidos en la técnica, tales como la mutagénesis dirigida o la mutagénesis por barrido de alanina (Cunningham y Wells, 1989, *Science* 244: 1081- 1085). En esta técnica, las mutaciones simples de alanina se introducen en cada residuo en la molécula, y las moléculas mutantes resultantes se evalúan para detectar la actividad biológica (es decir, actividad de beta-glucosidasa) con el fin de identificar los 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 mediante el análisis físico de la estructura, como determinan técnicas tales como la resonancia magnética nuclear, la cristalografía, la difracción electrónica o el marcaje por fotoafinidad, conjuntamente con la mutación de aminoácidos de sitio de contacto putativo. 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 los aminoácidos esenciales también pueden ser inferidas del análisis de identidades con polipéptidos que están relacionados con un polipéptido según la invención.

[0058] Las sustituciones de aminoácidos simples o múltiples pueden ser realizadas y evaluadas utilizando métodos conocidos de mutagénesis, recombinación y/o redistribución, seguidos de un procedimiento de selección pertinente, tal como los descritos por Reidhaar-Olson y Sauer, 1988, Science 241: 53-57; Bowie y Sauer, 1989, Proc. Natl. Acad. Sci. USA 86: 2152-2156; WO 95/17413; o WO 95/22625. Otros de los métodos que se pueden usar incluyen PCR propensa a error, exposición en el fago (p. ej., Lowman et al., 1991, Biochem. 30:10832-10837; Patente estadounidense n° 5,223,409; WO 92/06204), y mutagénesis dirigida (Derbyshire et al., 1986, Gene 46:145; Ner et al., 1988, ADN 7:127).

[0059] Los métodos de mutagénesis/redistribución se pueden combinar con métodos de selección automatizados de alto rendimiento para detectar actividad de polipéptidos mutagenizados clonados expresados por células huésped. Las moléculas de ADN mutagenizadas que codifican los polipéptidos activos pueden ser recuperadas de las células huésped y rápidamente ordenadas usando métodos estándar de la técnica. Estos métodos permiten la rápida determinación de la importancia de residuos de aminoácidos individuales en un polipéptido de interés y pueden ser aplicados a polipéptidos de estructura desconocida.

[0060] La cantidad total de sustituciones de aminoácidos, deleciones y/o inserciones de aminoácidos 20 a 863 de la SEC ID n°: 2 es 10, preferiblemente 9, más preferiblemente 8, más preferiblemente 7, más preferiblemente como máximo 6, más preferiblemente 5, más preferiblemente 4, incluso más preferiblemente 3, más preferiblemente 2, e incluso más preferiblemente 1.

Fuentes de polipéptidos con actividad de beta-glucosidasa

[0061] Un polipéptido de la presente invención se puede obtener de microorganismos de cualquier género. Para fines de la presente invención, el término "obtener de" según se utiliza en este caso en relación con una fuente dada significa que el polipéptido codificado por una secuencia de nucleótidos es producido por la fuente o por una cepa en la que ha sido insertada la secuencia de nucleótidos de la fuente. En un aspecto preferido, el polipéptido obtenido de una fuente dada es secretado extracelularmente.

[0062] Un polipéptido de la presente invención puede ser un polipéptido fúngico, y más preferiblemente un *Aspergillus*.

[0063] En un aspecto preferido, el polipéptido es un *Aspergillus fumigatus*.

[0064] En un aspecto más preferido, el polipéptido es un *Aspergillus fumigatus*, el polipéptido de la SEC ID n°: 2.

[0065] Se entenderá que para las especies ya mencionadas la invención abarca tanto los estados perfectos como los imperfectos y otros equivalentes taxonómicos, por ejemplo, anamorfos, independientemente del nombre por el que es conocida la especie. Aquellos expertos en la técnica reconocerán fácilmente la identidad de los equivalentes apropiados.

[0066] Las cepas de estas especies son fácilmente accesibles para el público a través de varias colecciones de cultivo, como la American Type Culture Collection (ATCC), la Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), la Centraalbureau voor Schimmelcultures (CBS) y la Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

[0067] Los polipéptidos de la presente invención también incluyen polipéptidos fusionados o polipéptidos fusión seleccionables en los que otro polipéptido se fusiona en el N-terminal o el C-terminal del polipéptido o fragmento del mismo. Un polipéptido fusionado se produce mediante la fusión de una secuencia de nucleótidos (o una parte de la misma) codificando otro polipéptido a una secuencia de nucleótidos (o una parte de la misma) de la presente invención. Las técnicas para producir polipéptidos de fusión son conocidas en la técnica e incluyen la ligadura de las secuencias de codificación que codifican los polipéptidos para que estén dentro del marco y que la expresión del polipéptido fusionado esté bajo control del/los mismo/s promotor/es y terminador.

Polinucleótidos

[0068] La presente invención también se refiere a polinucleótidos aislados que comprenden una secuencia de nucleótidos que codifica un polipéptido de la presente invención. En un aspecto preferido, la secuencia de nucleótidos se establece en la SEC ID n°: 1. En otro aspecto más preferido, la secuencia de nucleótidos es la secuencia contenida en el plásmido pEJG113, contenido a su vez en *Escherichia coli* NRRL B-30695. En otro aspecto preferido, la secuencia de nucleótidos es la región de codificación de polipéptido maduro de la SEC ID n°: 1. En otro aspecto más preferido, la

secuencia de nucleótidos es la región de codificación del polipéptido maduro contenida en el plásmido pEJG113, contenido a su vez en *Escherichia coli* NRRL B-30695. La presente invención también incluye secuencias de nucleótidos que codifican un polipéptido que comprende la secuencia de aminoácidos de la SEC ID nº: 2 o el polipéptido maduro de la misma, que difiere de la SEC ID nº: 1 en virtud de la degeneración del código genético.

5 **[0069]** Las técnicas usadas para aislar o clonar un polinucleótido que codifica un polipéptido son conocidas en la técnica e incluyen el aislamiento del ADN genómico, la preparación de ADNc o una combinación de los mismos. La clonación de los polinucleótidos de la presente invención a partir de ese ADN genómico puede ser efectuada, por ejemplo, usando la conocida reacción en cadena de la polimerasa (PCR) o una selección de anticuerpos de bibliotecas de expresión para detectar fragmentos de ADN clonados con características estructurales compartidas. Véase, por ejemplo, Innis et al., 10 1990, PCR: A Guide to Methods and Application, Academic Press, Nueva York. Se pueden utilizar procedimientos de amplificación de ácido nucleico tales como la reacción en cadena de la ligasa (LCR), la transcripción activada ligada (LAT) y la amplificación basada en la secuencia de nucleótidos (NASBA). Los polinucleótidos pueden ser clonados de una cepa de *Aspergillus* u otro organismo relacionado y de ese modo, por ejemplo, puede ser una variante alélica o de especie de la región de codificación del polipéptido de la secuencia de nucleótidos.

15 **[0070]** Puede ser necesaria la modificación de una secuencia de nucleótidos que codifica un polipéptido de la presente invención para la síntesis de polipéptidos sustancialmente similares al polipéptido. El término "sustancialmente similares" al polipeptídico se refiere a formas del polipéptido que no están presentes de forma natural. Estos polipéptidos pueden diferir en algún aspecto del diseño genético del polipéptido aislado de su fuente nativa, por ejemplo, variantes artificiales que difieren en actividad específica, termoestabilidad, pH óptimo o similares. La secuencia de la variante puede estar construida sobre la base de la secuencia de nucleótidos presentada como la región de codificación del polipéptido de la SEC ID nº: 1, por ejemplo, una subsecuencia de la misma, y/o por introducción de sustituciones de nucleótidos que no dan lugar a otra secuencia de aminoácidos del polipéptido codificado por la secuencia de 20 nucleótidos, pero que corresponde al uso del codón del organismo huésped destinado para la producción de la enzima, o por introducción de sustituciones de nucleótidos que pueden dar lugar a una secuencia de aminoácidos diferente. Para obtener una descripción general de sustitución de nucleótido, véase, por ejemplo, Ford et al., 1991, Protein Expression and Purification 2: 95-107.

25 **[0071]** Resultará evidente para los expertos en la técnica que tales sustituciones pueden ser realizadas fuera de las regiones críticas a la función de la molécula y aun así darán como resultado un polipéptido activo. Los residuos de aminoácidos esenciales para la actividad del polipéptido codificado por un polinucleótido aislado de la invención y, por lo tanto, preferiblemente no sometidos a la sustitución, pueden ser identificados según procedimientos conocidos en la técnica, tales como la mutagénesis dirigida o mutagénesis por barrido de alanina (véase, por ejemplo, Cunningham y Wells, 1989, Science 244: 1081-1085). En esta técnica, las mutaciones se introducen en cada residuo positivamente cargado en la molécula y las moléculas mutantes resultantes se evalúan para detectar actividad de beta-glucosidasa y así identificar residuos de aminoácidos que son críticos para la actividad de la molécula. Los sitios de interacción sustrato-enzima también pueden ser determinados por análisis de la estructura tridimensional como lo determinan técnicas tales como el análisis de resonancia magnética nuclear, la cristalografía o el marcaje por fotoafinidad (véase, por ejemplo, de Vos et al., 1992, Science 255: 306-312; Smith et al., 1992, Journal of Molecular Biology 224: 899-904; Wlodaver et al., 1992, FEBS Letters 309: 59-64).

Constructos de ácidos nucleicos

35 **[0072]** La presente invención también se refiere a constructos de ácidos nucleicos que comprenden un polinucleótido aislado de la presente invención operativamente enlazados a una o más secuencias de control que dirigen la expresión de la secuencia codificante en una célula huésped adecuada en condiciones compatibles con las secuencias de control.

40 **[0073]** Un polinucleótido aislado que codifica un polipéptido de la presente invención puede ser manipulado en una variedad de formas para proporcionar la expresión del polipéptido. La manipulación de la secuencia del polinucleótido antes de su inserción en un vector puede ser deseable o necesaria dependiendo del vector de expresión. Las técnicas para modificar las secuencias polinucleótidas utilizando métodos de ADN recombinante se conocen bien en la técnica.

45 **[0074]** La secuencia de control puede ser una secuencia promotora apropiada, una secuencia de nucleótidos que sea reconocida por una célula huésped para la expresión de un polinucleótido que codifica un polipéptido de la presente invención. La secuencia promotora contiene secuencias de control transcripcional que controlan la expresión del polipéptido. El promotor puede ser cualquier secuencia de nucleótidos que muestre actividad transcripcional en la célula huésped elegida, incluidos los promotores mutantes, truncados e híbridos, y se pueden obtener a partir de genes que codifican polipéptidos extracelulares o intracelulares homólogos o heterólogos a la célula huésped.

50 **[0075]** Algunos ejemplos de promotores adecuados para dirigir la transcripción de los constructos de ácidos nucleicos de la presente invención en una célula huésped filamentosa fúngica son los promotores obtenidos a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, proteínaasa aspártica de *Rhizomucor miehei*, alfa-amilasa neutra de *Aspergillus niger*, alfaamilasa estable en ácido de *Aspergillus niger*, glucoamilasa de *Aspergillus niger* o *Aspergillus awamori* (glaA), lipasa de *Rhizomucor miehei*, proteasa alcalina de *Aspergillus oryzae*, triosa fosfato isomerasa de *Aspergillus oryzae*, acetamidasa de *Aspergillus nidulans*, amiloglucosidasa de *Fusarium venenatum* (WO 00/56900),

- Fusarium venenatum Daria (WO 00/56900), *Fusarium venenatum* Quinn (WO 00/56900), proteasa 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 NA2-tpi (un híbrido de los promotores de los genes para alfa-amilasa neutra de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae*); y promotores mutantes, truncados e híbridos de los mismos.
- [0076]** En un huésped de levadura, se obtienen promotores útiles los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactoquinasa de *Saccharomyces cerevisiae* (GAL1), alcohol dehidrogenasa/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*. Otro promotores útiles para células huésped de levadura son descritos por Romanos et al., 1992, Yeast 8: 423-488.
- [0077]** La secuencia de control también puede ser una secuencia del terminador de la transcripción adecuada, una secuencia reconocida por una célula huésped para terminar la transcripción. La secuencia del terminador está operativamente enlazada al terminal 3' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier terminador que sea funcional en la célula huésped de elección puede ser usado en la presente invención.
- [0078]** Los terminadores preferidos para células huésped 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 tipo tripsina de *Fusarium oxysporum*.
- [0079]** Los terminadores preferidos para células huésped 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 las células huésped de levadura son descritos por Romanos et al., 1992, *supra*.
- [0080]** 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 líder está operativamente enlazada al terminal 5' de la secuencia de nucleótidos que codifica el polipéptido. En la presente invención se puede utilizar cualquier secuencia líder que sea funcional en la célula huésped elegida.
- [0081]** Los líderes preferidos para las células huésped fúngicas filamentosas se obtienen a partir de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.
- [0082]** Los líderes adecuados para células huésped de levadura se obtienen a partir de los genes para enolasa (ENO-1) de *Saccharomyces cerevisiae*, 3-fosfoglicerato-quinasa de *Saccharomyces cerevisiae*, factor alfa de *Saccharomyces cerevisiae* y alcohol dehidrogenasa/gliceraldehído-3-fosfato deshidrogenasa (ADH2/GAP) de *Saccharomyces cerevisiae*.
- [0083]** La secuencia de control también puede ser una secuencia de poliadenilación, una secuencia operativamente enlazada al terminal 3' de la secuencia de nucleótidos y que transcrita es reconocida por la célula huésped como una señal para añadir residuos de poliadenosina al ARNm transcrito. En la presente invención se puede utilizar cualquier secuencia de poliadenilación que sea funcional en la célula huésped elegida.
- [0084]** Las secuencias de poliadenilación preferidas para las células huésped fúngicas filamentosas 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*.
- [0085]** Secuencias de poliadenilación útiles para las células huésped de levadura son descritas por Guo y Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.
- [0086]** La secuencia de control también puede ser una región de codificación del péptido señal que codifica una secuencia de aminoácidos enlazada al término amino de un polipéptido y dirige el polipéptido codificado hacia la vía secretora de la célula. El extremo 5' de la secuencia codificante de la secuencia de nucleótidos puede contener intrínsecamente una región de codificación del péptido señal enlazado naturalmente en el marco de lectura de traducción con el segmento de la región de codificación que codifica el polipéptido segregado. De forma alternativa, el extremo 5' de la secuencia codificante puede contener una región de codificación del péptido señal que es extraña a la secuencia codificante. La región de codificación del péptido señal extraña 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 extraña puede simplemente reemplazar la región de codificación del péptido señal natural para incrementar la secreción del polipéptido. No obstante, cualquier región de codificación del péptido señal que dirige el polipéptido expresado en la vía secretora de una célula huésped elegida puede ser utilizada en la presente invención.

[0087] Las regiones de codificación del péptido señal eficaces para células huésped fúngicas filamentosas son las regiones codificantes del péptido señal obtenidas de los genes para TAKA amilasa de *Aspergillus oryzae*, amilasa neutra de *Aspergillus niger*, glucoamilasa de *Aspergillus niger*, proteinasa aspártica de *Rhizomucor miehei*, celulasa de *Humicola insolens*, endoglucanasa V de *Humicola insolens* y lipasa *Humicola lanuginosa*.

[0088] En un aspecto preferido, la región de codificación del péptido señal son los nucleótidos 1 a 58 de la SEC ID n°: 1 que codifican los aminoácidos 1 a 19 de la SEC ID n°: 2.

[0089] Péptidos señal útiles para células huésped de levadura se obtienen de los genes para factor alfa de *Saccharomyces cerevisiae* e invertasa de *Saccharomyces cerevisiae*. Otras regiones de codificación útiles del péptido señal útiles otras son descritas por Romanos et al., 1992, *supra*.

[0090] La secuencia de control también puede ser una región de codificación del propéptido que codifica para una secuencia de aminoácidos situada en el término amino de un polipéptido. El polipéptido resultante se conoce como una proenzima o propolipéptido (o un zimógeno en algunos casos). Un propolipéptido está generalmente inactivo y puede ser convertido a un polipéptido maduro activo por ruptura catalítica o autocatalítica del propéptido del propolipéptido. La región de codificación del propéptido se puede obtener a partir de los genes para proteasa alcalina (aprE) de *Bacillus subtilis*, proteasa neutra (nprT) de *Bacillus subtilis*, factor alfa de *Saccharomyces cerevisiae*, proteinasa aspártica de *Rhizomucor miehei* y lacasa de *Myceliophthora thermophila* (WO 95/33836).

[0091] Cuando tanto las regiones del péptido señal y como las del propéptido están presentes en el término amino de un polipéptido, la región del propéptido se sitúa junto al término amino de un polipéptido y la región del péptido señal junto al término amino de la región del propéptido.

[0092] También puede ser conveniente añadir secuencias reguladoras que permitan la regulación de la expresión del polipéptido relativo al crecimiento de la célula huésped. Ejemplos de sistemas reguladores son los que causan que la expresión del gen sea activada o desactivada en respuesta a un estímulo físico o químico, 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 TAKA de alfa-amilasa, el 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 las que permiten la amplificación del gen. En sistemas eucarióticos, estas incluyen el gen de dihidrofolato-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 codifica el polipéptido estaría operativamente enlazada con la secuencia reguladora.

Vectores de expresión

[0093] La presente invención también se refiere a vectores de expresión recombinante que comprenden un polinucleótido de la presente invención, un promotor y señales de parada transcripcional y traduccional. Los diferentes ácidos nucleicos y secuencias de control descritos en la presente 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 sustitución de la secuencia de nucleótidos que codifica el polipéptido en tales sitios. De forma alternativa, una secuencia de nucleótidos de la presente invención 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 enlazada con las secuencias de control apropiadas para la expresión.

[0094] El vector de expresión recombinante puede ser cualquier vector (p. ej., 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 dependerá normalmente de la compatibilidad del vector con la célula huésped en la que el vector será introducido. Los vectores pueden ser plásmidos circulares lineales o cerrados.

[0095] El vector puede ser un vector de replicación autónoma, es decir, un vector que existe como una entidad extracromosómica, cuya replicación es independiente 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, cuando se introduce en la célula huésped, está integrado en el genoma y se replica con el/los cromosoma(s) en el/los que ha sido integrado. Además, se puede usar un único vector o plásmido o dos o más vectores o plásmidos que juntos contienen el ADN total para ser introducido en el genoma de la célula huésped, o un transposón.

[0096] Los vectores de la presente invención contienen preferiblemente uno o más marcadores seleccionables que permiten la selección fácil de células transformadas, transfectadas, transducidas o similares. Un marcador seleccionable es un gen cuyo producto proporciona resistencia biocida o vírica, resistencia a metales pesados, prototrofia para auxótrofos y similares.

[0097] Marcadores adecuados para células huésped de levadura son ADE2, HIS3, LEU2, LYS2, MET3, TRP1, y URA3.

Marcadores seleccionables para el uso en una célula huésped fúngica filamentosa incluyen, entre otros, *amdS* (acetamidasa), *argB* (ornitincarbamoiltransferasa), *bar* (fosfonitrina acetiltransferasa), *hph* (higromicina fosfotransferasa), *niaD* (nitrato-reductasa), *pyrG* (orotidina-5'-fosfato-descarboxilasa), *sC* (sulfato adeniltransferasa) y *trpC* (antranilato sintasa), así como equivalentes de los mismos. Para el uso en una célula del *Aspergillus* se prefieren los genes *amdS* y *pyrG* de *Aspergillus nidulans* o *Aspergillus oryzae* y el gen *bar* de *Streptomyces hygroscopicus*.

[0098] Los vectores de la presente invención contienen preferiblemente (un) elemento(s) que permite(n) la integración del vector en el genoma de la célula huésped o la replicación autónoma del vector en la célula independiente del genoma.

[0099] Para la integración en el genoma de la célula huésped, el vector puede depender de la secuencia de polinucleótidos que codifica el polipéptido o de cualquier otro elemento del vector para la integración en el genoma por recombinación homóloga o no homóloga. De forma alternativa, el vector puede contener secuencias de nucleótidos adicionales para dirigir integración por recombinación homóloga en el genoma de la célula huésped a una ubicación o ubicaciones precisa(s) en el/los cromosoma(s). Para aumentar la probabilidad de integración a una ubicación precisa, los elementos integracionales deberían contener preferiblemente un número suficiente de ácidos nucleicos, como 100 a 10.000 pares de bases, preferiblemente 400 a 10.000 pares de bases, y más preferiblemente 800 a 10.000 pares de bases, que tienen un alto grado de identidad con la secuencia diana correspondiente para incrementar la probabilidad de recombinación homóloga. Los elementos integracionales pueden ser cualquier secuencia que sea homóloga con la secuencia diana en el genoma de la célula huésped. Además, los elementos integracionales pueden ser secuencias de nucleótidos no codificantes o codificantes. Por otro lado, el vector se puede integrar en el genoma de la célula huésped por recombinación no homóloga.

[0100] Para la replicación autónoma, el vector puede comprender además un origen de replicación que le permita al vector replicarse de manera autónoma en la célula huésped en cuestión. El origen de replicación puede ser cualquier replicador plásmido que controle la replicación autónoma que funciona en una célula. El término "origen de replicación" o "replicador plásmido" se define en este caso como una secuencia de nucleótidos que permite a un plásmido o vector replicarse en vivo.

[0101] Algunos ejemplos de orígenes de replicación para el uso en una célula huésped de levadura son los orígenes de replicación de 2 micrones, ARS1, ARS4, la combinación de ARS1 y CEN3 y la combinación de ARS4 y CEN6.

[0102] Algunos ejemplos de orígenes de replicación 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 el documento WO 00/24883.

[0103] Se puede insertar más de una copia de un polinucleótido de la presente invención en la célula huésped para aumentar la producción del producto génico. Se puede obtener un aumento en el número de copias del polinucleótido integrando al menos una copia adicional de la secuencia en el genoma de la célula huésped o incluyendo un gen marcador seleccionable amplificable con el polinucleótido donde las células contienen copias amplificadas del gen marcador seleccionable, y, así, las copias adicionales del polinucleótido pueden seleccionarse cultivando las células en presencia del agente seleccionable apropiado.

[0104] 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

[0105] La presente invención también se refiere a células huésped recombinantes, que comprenden un polinucleótido de la presente invención, que son ventajosamente usadas en la producción recombinante de los polipéptidos. Un vector que comprende un polinucleótido de la presente invención se introduce en una célula huésped de modo que el vector se mantiene como un integrante cromosómico o como un vector extracromosómico que se auto-duplica 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 replicación. La elección de una célula huésped depende en gran parte del gen que codifica el polipéptido y su fuente.

[0106] La célula huésped puede ser una célula eucariota, tal como una célula de mamífero, de insecto, de planta o una célula fúngica.

[0107] En un aspecto preferido, la célula huésped es una célula fúngica. "Hongo", según se utiliza en este caso, incluye el filo Ascomycota, Basidiomycota, Chytridiomycota y Zygomycota (según la definición de Hawkswort et al., en Ainswort and Bisby's Dictionary of the Fungi, Octava edición, 1995, CAB International, University Press, Cambridge, Reino Unido), al igual que la Oomycota (según se cita en Hawkswort et al., 1995, *supra*, página 171) y todos los hongos mitospóricos (Hawkswort et al., 1995, *supra*).

5 [0108] En un aspecto más preferido, la célula huésped fúngica es una célula de levadura. "Levadura", según se utiliza en este caso, incluye levadura ascoesporógena (Endomycetales), levadura basidioesporogénea y levadura de los Fungi Imperfecti (Blastomycetes). Ya que la clasificación de levadura puede cambiar en el futuro, para los fines de esta 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).

10 [0109] En un aspecto más preferido incluso, la célula huésped de levadura es una célula *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*.

15 [0110] En un aspecto más preferido, la célula huésped de levadura es un *Saccharomyces*, *carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, o *Saccharomyces oviformis*. En otro aspecto más preferido, la célula huésped de levadura es una célula de *Kluyveromyces lactis*. En otro aspecto más preferido, la célula huésped de levadura es una célula de *Yarrowia lipolytica*.

20 [0111] En otro aspecto más preferido, la célula huésped fúngica es una célula fúngica filamentosa. Los "hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión Eumycota y Oomycota (según la definición de Hawkswort et al., 1995, *supra*). Los hongos filamentosos generalmente están caracterizados por una pared micelial compuesta por quitina, celulosa, glucano, quitosano, manano y otros polisacáridos complejos. El crecimiento vegetativo se produce por elongación hifal, mientras que 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.

25 [0112] En un aspecto incluso más preferido, la célula huésped fúngica filamentosa es una célula de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Coprinus*, *Coriolus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Phlebia*, *Piromyces*, *Pleurotus*, *Schizophyllum*, *talaromyces*, *Teramoascus*, *Thielavia*, *Tolipocladium*, *Trametes*, *Trichoderma*.

30 [0113] En un aspecto más preferido, la célula huésped fúngica filamentosa es una célula de *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*. En otro aspecto más preferido, la célula huésped 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*, *Fusarium venenatum*. En otro aspecto más preferido, la célula huésped fúngica filamentosa es una célula de *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Ceriporiopsis cinereus*, *Coprinus subvermispora*, *Coriolus hirsutus*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Phlebia radiata*, *Pleurotus eryngii*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.

45 [0114] Las células fúngicas se pueden transformar por un proceso que implica la formación de protoplastos, la transformación de los protoplastos y la regeneración de la pared celular de una manera conocida. Los procedimientos adecuados para la transformación de las células huésped de *Aspergillus* y *Trichoderma* están descritos en el documento EP 238 023 y Yelton et al., 1984, *Proceedings of the National Academy of Sciences EEUU* 81: 1470-1474. Métodos adecuados para transformar *Fusarium* son descritos por Malardier et al., 1989, *Gene* 78: 147-156, y el documento 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 Ecymology*, Volumen 194, págs. 182-187, Academic Press, Inc., Nueva York; Ito et al., 1983, *Journal of Bacteriology* 153: 163; e Hinnen et al., 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920.

55 Métodos de producción

60 [0115] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que comprenden (a) el cultivo de una célula, que en su forma salvaje es capaz de producir el polipéptido, en las condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido. Preferiblemente, la célula es del género *Aspergillus*, y más preferiblemente *Aspergillus fumigatus*.

[0116] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que comprenden (a) el cultivo de una célula huésped en las condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido.

65 [0117] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que comprenden (a) el cultivo de una célula huésped en las condiciones propicias para la producción del polipéptido, donde

la célula huésped comprende una secuencia de nucleótidos mutante con al menos una mutación en la región de codificación del polipéptido maduro de la SEC ID n°: 1, donde la secuencia de nucleótidos mutante codifica un polipéptido que consiste en los aminoácidos 20 a 863 de SEC ID n°: 2, y (b) la recuperación del polipéptido.

5 **[0118]** En los métodos de producción de la presente invención, las células se cultivan en un medio nutritivo adecuado para la producción del polipéptido usando métodos conocidos en la técnica. Por ejemplo, la célula se puede cultivar por cultivo en matraz de agitación, y fermentación a pequeña o gran escala (incluidas las fermentaciones continuas, de lote, lote alimentado o en estado sólido) en fermentadores de laboratorio o industriales realizados en un medio adecuado y en 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 a través de proveedores comerciales o se pueden preparar según composiciones publicadas (p. ej., en catálogos de la American Type Culture Collection). Si el polipéptido se secreta en el medio nutritivo, el polipéptido se puede recuperar directamente del medio. Si el polipéptido no es segregado del medio, se puede recuperar de lisados celulares.

15 **[0119]** Los polipéptidos pueden ser detectados 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.

20 **[0120]** El polipéptido resultante se puede recuperar usando métodos conocidos en la técnica. Por ejemplo, el polipéptido puede ser recuperado del medio nutritivo por procedimientos convencionales, incluidos, entre otros, el centrifugado, la filtración, la extracción, el secado por pulverización, la evaporación o la precipitación.

25 **[0121]** Los polipéptidos de la presente invención se pueden purificar por una variedad de procedimientos conocidos en la técnica, incluidos, entre otros, la cromatografía (p. ej., intercambio iónico, por afinidad, hidrofóbica, de cromatoenfoco y de exclusión de tamaño), los procedimientos electroforéticos (p. ej., isoelectroenfoco preparatorio), solubilidad diferencial (p. ej., precipitación de sulfato amónico), SDS-PAGE o extracción (véase, por ejemplo, *Protein Purification*, J.-C. Janson y Lars Ryden, editores, VCH Publishers, Nueva York, 1989) para obtener polipéptidos substancialmente puros.

Composiciones

35 **[0122]** La presente divulgación también se refiere a composiciones que comprenden un polipéptido de la presente invención. Preferiblemente, las composiciones están enriquecidas en tal polipéptido. El término "enriquecido" indica que la actividad de beta-glucosidasa de la composición ha sido aumentada, por ejemplo, con un factor de enriquecimiento de al menos 1,1.

40 **[0123]** La composición puede comprender un polipéptido de la presente invención como el componente enzimático principal, por ejemplo, una composición monocomponente. De forma alternativa, la composición puede comprender actividades enzimáticas múltiples, tales como una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, ciclodextrina glicosiltransferasa, desoxiribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, glucosidasa alfa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolítica, peptidoglucaminasa, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa o xilanas. La enzima(s) adicional(es) puede(n) ser producida(s), por ejemplo, por un microorganismo del género *Aspergillus*, preferiblemente *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*; *Fusarium*, preferiblemente *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 sulphureum*, *Fusarium toruloseum*, *Fusarium trichothecioides* o *Fusarium venenatum*; *Humicola*, preferiblemente *Humicola insolens* o *Humicola lanuginosa*; o *Trichoderma*, preferiblemente *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.

55 **[0124]** Las composiciones de polipéptido se pueden preparar conforme a métodos conocidos en la técnica y pueden estar en la forma de un líquido o una composición seca. Por ejemplo, la composición de polipéptido puede estar en la forma de un granulado o un microgranulado. El polipéptido que será incluido en la composición se puede estabilizar conforme a métodos conocidos en la técnica.

60 **[0125]** Más abajo se proporcionan ejemplos de usos preferidos de las composiciones de polipéptido de la invención. La dosificación de la composición de polipéptido de la invención y otras condiciones en las que se usa la composición pueden determinarse basándose en métodos conocidos en la técnica.

Degradación de biomasa a monosacáridos, disacáridos y polisacáridos

65 **[0126]** Los polipéptidos con actividad de beta-glucosidasa 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, plásticos, u otros productos o productos intermedios. Los polipéptidos con actividad de beta-glucosidasa beta pueden estar en forma de caldo de fermentación crudo con o sin las células eliminadas o en forma de preparación enzimática semipurificada o purificada. De manera alternativa, una célula huésped de la presente invención se puede utilizar como una fuente del polipéptido con actividad de beta-glucosidasa en un proceso de fermentación con la biomasa.

[0127] La biomasa puede incluir, entre otros, recursos de madera, desperdicios municipales sólidos, papel usado y residuos de cosechas (véase, por ejemplo, Wiselogel et al., 1995, en Handbook on Bioethanol (Carlos E. Wiman, editor), pp.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 of Lignocellulosics, in Advances in Biochemical, T. Scheper, Editor jefe, Volumen 65; pp.23-40, Springer-Verlag, Nueva York).

[0128] El polisacárido predominante en la pared celular primaria de la biomasa es la celulosa, el segundo más abundante es la hemicelulosa y el tercero la pectina. La pared celular secundaria, producida una vez 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, por lo tanto, 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. Aunque en general es polimorfa, se encuentra celulosa 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 para otras hemicelulosas, lo que ayuda estabilizar la matriz de la pared celular.

[0129] Se utilizan tres clases principales de glicohidrolasas para descomponer la biomasa celulósica:

(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 beta-1,4-glucano solubles e insolubles.

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

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

[0130] Estas tres clases de enzimas trabajan juntas sinérgicamente dando lugar a la decristalización eficaz e hidrólisis de la celulosa nativa de biomasa para producir azúcares reductores.

[0131] Los polipéptidos con actividad de beta-glucosidasa de la presente invención pueden ser utilizados conjuntamente con las enzimas mencionadas anteriormente para lograr una mayor degradación del componente de celulosa del sustrato de biomasa, (véase, por ejemplo, Brigham et al., 1995, en Handbook on Biotechnology (Charles E. Wyman, editor), pp.119-141, Taylor & Francis, Washington D.C.; Lee, 1997, Journal of Biotechnology 56: 1-24).

[0132] El etanol se puede producir mediante la degradación enzimática de biomasa y la conversión de los sacáridos liberados a etanol. Este tipo de etanol es frecuentemente denominado bioetanol o biocombustible. Se puede usar como aditivo para combustible o suplemento en mezclas de menos de un 1% y hasta un 100% (un sustituto de combustible).

Composiciones detergentes

[0133] Los polipéptidos con actividad de beta-glucosidasa de la presente invención pueden ser añadidos a la composición detergente para así convertirse en un componente de la misma.

[0134] La composición de detergente de la presente divulgación puede, por ejemplo, ser formulada como una composición detergente de lavado a mano o a máquina que incluye una composición aditiva de lavado adecuada para el tratamiento previo de tejidos manchados y una composición de suavizante agregada al enjuague, o formulada como una composición de detergente para el uso en operaciones de limpieza de superficies duras del hogar en general, o ser formulada para operaciones de lavado de vajillas a mano o a máquina.

[0135] En un aspecto específico, la presente divulgación proporciona un aditivo detergente que comprende los polipéptidos que tienen actividad de beta-glucosidasa de la presente invención. El aditivo detergente, al igual que la composición detergente, puede comprender una o más enzimas, tales como una proteasa, lipasa, cutinasa, una amilasa, carbohidrasa, celulasa, pectinasa, mananasa, arabinasa, galactanasa, xilanas, oxidasa, por ejemplo, una lacasa y/o peroxidasa.

[0136] 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 deberían estar presentes en cantidades eficaces.

5 **[0137] Proteasas:** las proteasas adecuadas incluyen las de origen animal vegetal u origen microbiano. Se prefiere el origen microbiano. Se incluyen los mutantes de proteínas diseñados genéticamente o modificados químicamente. La proteasa puede ser una serina proteasa o una metaloproteasa, preferiblemente una proteasa alcalina microbiana o una proteasa de tipo tripsina. Algunos ejemplos de proteasas alcalinas son las subtilisinas, especialmente aquellas derivadas de *Bacillus*, por ejemplo, subtilisina Novo, subtilisina Carlsberg, subtilisina 309, subtilisina 147 y subtilisina 168 (descritas en WO 89/06279). Algunos ejemplos de proteasas de tipo tripsina son la tripsina (p. ej., de origen bovino
10 o porcino) y la proteasa de *Fusarium* descrita en los documentos WO 89/06270 y WO 94/25583.

15 **[0138]** Algunos ejemplos de proteasas útiles son las variantes descritas en los documentos 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.

[0139] Las enzimas proteásicas disponibles comercialmente que se prefieren incluyen Alcalase™, Savinase™, Primase™, Duralase™, Esperase™, y Kannase™ (Novozymes A/S), Maxatase™, Maxacal™, Maxapem™, Properase™, Purafect™, OxP™ Purafect, FN2™ y FN3™ (Genencor Internacional Inc.).

20 **[0140] Lipasas:** las lipasas adecuadas incluyen aquellas de origen fúngico o bacteriano. Se incluyen los mutantes de proteínas diseñados genéticamente o modificados químicamente. Entre los ejemplos de lipasas útiles se incluyen las lipasas de *Humicola* (sinónimo *Thermomyces*), por ejemplo, de *H. lanuginosa* (*T. Lanuginosus*), según se describe en los documentos EP 258 068 y EP 305 216 o de *H. insolens* como se describe en el documento WO 96/13580, una lipasa de *Pseudomonas*, por ejemplo, de *P. alcaligene* o *P. pseudoalcaligenes* (EP 218 272), *P. cepacia* (EP 331 376),
25 *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).

30 **[0141]** Otros ejemplos son variantes de lipasa tales como las descritas en los documentos 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.

35 **[0142]** Las enzimas de lipasas disponibles comercialmente preferidas incluyen Lipolase™ y Lipolase Ultra™ (Novozymes A/S).

[0143] Amilasas: las amilasas adecuadas (α y/o β) incluyen las de origen bacteriano o fúngico. Se incluyen los mutantes de proteínas diseñados genéticamente o modificados químicamente. Las amilasas incluyen, por ejemplo, α -amilasas obtenidas de *Bacillus*, por ejemplo, una cepa especial de *Bacillus licheniformis*, descrita con más detalle en el documento GB 1.296.839.

40 **[0144]** Algunos ejemplos de amilasas útiles son las variantes descritas en los documentos 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.

45 **[0145]** Las amilasas disponibles comercialmente son Duramyl™, Termamyl™, Fungamyl™ y BAN™ (Novozymes A/S), Rapidase™ y Purastar™ (de Genencor Internacional Inc.).

50 **[0146] Celulasas:** las celulasas adecuadas incluyen las de origen bacteriano o fúngico. Se incluyen los mutantes de proteínas diseñados genéticamente o modificados químicamente. 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 la patente estadounidense 4.435.307, la patente estadounidense 5.648.263, la patente estadounidense 5.691.178, la patente estadounidense 5.776.757 y WO 89/09259.

55 **[0147]** Las celulasas especialmente adecuadas son las celulasas alcalinas o neutras que tienen beneficios para el cuidado del color. Ejemplos de tales celulasas son las celulasas descritas en los documentos EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son las variantes de celulasa tales como las descritas en los documentos WO 94/07998, EP 0 531 315, en la patente estadounidense 5.457.046, US 5.686.593, la patente estadounidense 5.763.254, WO 95/24471, WO 98/12307 y PCT/DK98/00299.

60 **[0148]** Las celulasas disponibles comercialmente incluyen Celluzyme™, y Carezyme™ (Novozymes A/S), Clazinase™, y Puradax HA™ (Genencor Internacional Inc.) y KAC-500(B)™ (Kao Corporation).

65 **[0149] Peroxidasas/oxidases:** las peroxidasas/oxidases adecuadas incluyen las de origen vegetal, bacteriano o fúngico. Se incluyen los mutantes de proteínas diseñados genéticamente o modificados químicamente. Entre los ejemplos de

peroxidasas útiles se incluyen peroxidasas de *Coprinus*, por ejemplo, de *C. cinereus*, y variantes de las mismas como las descritas en los documentos WO 93/24618, WO 95/10602, y WO 98/15257.

[0150] Las peroxidasas disponibles comercialmente incluyen Guardzyme™ (Novozymes A/S).

[0151] 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 presente invención, es decir, un aditivo separado o un aditivo combinado, puede ser formulado, por ejemplo, como un granulado, líquido, lodo, etc. Las formulaciones de aditivo detergente preferidas son los granulados, en particular los granulados no pulverizados, líquidos, en particular líquidos estabilizados, o lodos.

[0152] Los granulados no pulverizados pueden ser producidos, por ejemplo, como se describe en las patentes estadounidenses 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), (polietilenoglicol; PEG) con pesos molares medios de 1000 a 20000; nonilfenoles etoxilados con de 16 a 50 unidades de óxido de etileno; alcoholes grasos etoxilados en los que el alcohol contiene de 12 a 20 átomos de carbono y en el cual hay de 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 dan en el documento GB 1483591. Las reparaciones enzimáticas líquidas pueden, por ejemplo, ser estabilizadas añadiendo un poliol, como el 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 el documento EP 238.216.

[0153] La composición detergente de la presente divulgación puede estar en cualquier forma conveniente, por ejemplo, una barra, una pastilla, en polvo, un gránulo, una pasta o en líquido. Un detergente líquido puede ser acuoso, por lo general con un 70% agua y 0-30% de solvente orgánico, o no acuoso.

[0154] La composición detergente comprende uno o más tensioactivos, que pueden ser no iónicos incluidos semipolares y/o aniónico y/o catiónico y/o zwitteriónicos. Los agentes tensioactivos normalmente están presentes a un nivel de 0,1% a 60% en peso.

[0155] Cuando se incluye en el mismo, el detergente normalmente contiene de aproximadamente el 1% a aproximadamente el 40% de un surfactante aniónico como el alquilbencenosulfonato lineal, alfa-olefinsulfonato, sulfato de alquilo (sulfato de alcohol graso), etoxisulfato alcohólico, alcanosulfonato secundario, éster metílico de ácido alfa-ulfó graso, ácido alquil o alquenilsuccínico o jabón.

[0156] Cuando se incluye en el mismo, el detergente normalmente contiene de aproximadamente el 0,2% a aproximadamente el 40% de un surfactante no iónico como el alcohol etoxilato, nonilfenol etoxilato alquilpoliglicósido, óxido de alquildimetilamina, monoetanolamida de ácido graso etoxilado, monoetanolamida de ácido graso, de polihidroxi aquil amida de ácido graso, o derivados de N-alquilo de N-acilo de glucosamina ("glucamidas").

[0157] El detergente puede contener el 0-65% de un constructor de detergente o agente complejante como la zeolita, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilendiaminatetraacético, ácido dietileno-triaminopentaacético, ácido alquil o alquenilsuccínico, silicatos solubles o silicatos estratificados (p. ej., SKS-6 de Hoechst).

[0158] El detergente puede comprender uno o más polímeros. Algunos ejemplos son la carboximetilcelulosa, poli(vinilpirrolidona), poli(etilenglicol), alcohol poli(vinílico), poli(vinilpiridina-N-óxido), poli(vinilimidazol), policarboxilatos como los poliacrilatos, copolímeros de ácido maleico/acrílico y copolímeros de ácido lauril metacrilato/ácido acrílico.

[0159] El detergente puede contener un sistema blanqueante que puede comprender una H₂O₂ como el perborato o percarbonato que se puede combinar con un activador blanqueante de formación de perácido como la tetraacetiletilenodiamina o el nonanoiloxibencenosulfonato. De forma alternativa, el sistema blanqueante puede comprender peroxiácidos, por ejemplo, de tipo amida, imida, o sulfona.

[0160] El/Los componente(s) enzimático(s) de la composición detergente de la presente divulgación puede(n) ser estabilizado(s) usando agentes estabilizantes convencionales de uso, por ejemplo, un poliol como el 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 borónico de fenilo como el ácido 4-formilfenil borónico y la composición puede ser formulada como se describe, por ejemplo, en los documentos WO 92/19709 y WO 92/19708.

[0161] El detergente también puede contener ingredientes de otros detergentes convencionales tales como acondicionadores de tejidos, incluyendo arcillas, reforzadores de espuma, supresores de espuma, agentes anticorrosivos, agentes de suspensión de suciedad, agentes anti redeposición de suciedad, tintes, bactericidas, blanqueadores ópticos, hidrótopos, inhibidores de decoloración o perfumes.

[0162] En las composiciones detergentes cualquier componente enzimático, en particular los polipéptidos con actividad

de beta-glucosidasa de la presente invención, puede ser añadido 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.

5 **[0163]** Los polipéptidos con actividad de beta-glucosidasa de la presente invención pueden ser incorporados adicionalmente en las formulaciones detergentes descritas en el documento WO 97/07202.

Planta

10 **[0164]** La presente divulgación también se refiere a una planta transgénica, parte de planta, o célula vegetal que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de beta-glucosidasa de la presente invención para expresar y producir el polipéptido en cantidades recuperables. El polipéptido se puede recuperar de la planta o parte de planta. De forma alternativa, la planta o parte de planta que contiene el polipéptido recombinante puede ser utilizado como tal para mejorar la calidad de un alimento o pienso, por ejemplo, mejorando el valor nutritivo, la palatabilidad y propiedades reológicas o para destruir un factor antinutritivo.

15 **[0165]** La planta transgénica puede ser dicotiledónea (un dicotiledóneo) o monocotiledónea (un monocotiledóneo). Algunos ejemplos de plantas monocotiledóneas son hierbas tales como la poa de prados (poa pratense, Poa), la hierba forrajera como la Festuca, el Lolium, el césped templado, como el Agrostis y los cereales, por ejemplo, el trigo, la avena, el centeno, la cebada, el arroz, el sorgo, y el maíz.

20 **[0166]** Algunos ejemplos de plantas dicotiledóneas son el tabaco, las leguminosas, como los altramuces, la patata, la remolacha azucarera, el guisante, la judía y la semilla de soja, y las plantas crucíferas (familia de Brassicaceae), tales como la coliflor, la semilla de colza y el organismo modelo íntimamente relacionado *Arabidopsis thaliana*.

25 **[0167]** Algunos ejemplos de partes de planta son el vástago, el callo, las hojas, la raíz, las frutas, las semillas y los tubérculos, al igual que los tejidos individuales que comprenden estas partes, por ejemplo, la epidermis, el mesofilo, la parénquima, los tejidos vasculares, los meristemas. Los compartimentos específicos de célula vegetal, como cloroplastos, apoplastos, mitocondria, vacuolas, peroxisomas y citoplasma se consideran también una parte de planta. Además, cualquier célula vegetal, sea cual sea el origen del tejido, se considera una parte de planta. Asimismo, las partes de planta tales como tejidos específicos y células aisladas para facilitar la utilización de la invención también se consideran partes de planta, por ejemplo, embriones, endospermas, aleurona y revestimientos de semillas.

30 **[0168]** La planta transgénica o célula vegetal que exprese un polipéptido de la presente invención se puede construir conforme a métodos conocidos en la técnica. En resumen, la planta o célula vegetal se construye incorporando uno o más constructos de expresión que codifica(n) un polipéptido 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 vegetal.

35 **[0169]** Convenientemente, el constructo de expresión es un constructo de ácidos nucleicos que comprende un polinucleótido que codifica un polipéptido de la presente invención operativamente enlazado con secuencias reguladoras apropiadas requeridas para la expresión de la secuencia de nucleótidos en la planta o parte de planta elegida. Además, el constructo de expresión puede comprender un marcador seleccionable útil para identificar las células huésped en las que el constructo de expresión ha sido integrado y secuencias de ADN necesarias para la introducción del constructo en la planta en cuestión (esto depende del método de introducción de ADN utilizado).

40 **[0170]** La elección de secuencias reguladoras, tales como las secuencias de terminador y promotor y opcionalmente las secuencias de tránsito o señal, está determinada, por ejemplo, en base a cuándo, dónde y cómo se desea expresar el polipéptido. Por ejemplo, la expresión del gen que codifica un polipéptido de la presente invención puede ser constitutiva o inducible, o desarrollable, específica de fase o tejido, y el producto génico puede ser dirigido hacia un tejido específico o parte de planta como las semillas u hojas. Las secuencias reguladoras son, por ejemplo, descritas por Tague et al., 1988, Plant Physiology 86: 506.

45 **[0171]** Para la expresión constitutiva pueden utilizarse el 35S-CaMV, la ubiquitina 1 de maíz y el promotor de la actina 1 de 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 las semillas, los tubérculos de patata y las frutas (Edwards & Coruzzi, 1990, Ana. Rev. Genet. 24: 275-303), o de tejidos sumideros metabólicos tales como los meristemas (Ito et al., 1994, Plant Mol. Biol. 24: 863-878), un promotor específico de semilla como el promotor de glutelina, prolamina, globulina o albúmina de arroz (Wu et al., 1998, Plant and Cell Physiology 39: 885-889), promotor de *Vicia faba* de la legúmina B4 y el gen de proteína de semilla desconocida de *Vicia faba* (Conrad et al., 1998, Journal of 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, como se describe, por ejemplo, en el documento WO 91/14772. Además, el promotor puede ser un promotor específico de hoja como el promotor de *rbcs* de arroz o tomate (Kyojuka et al., 1993, Plant Physiology 102: 991-1000, el promotor del gen de metiltransferasa de adenina de virus de chlorella (Mitra and 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 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 la temperatura, la sequía o alteraciones en 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.

[0172] También puede utilizarse un elemento intensificador del promotor para conseguir una mayor expresión de un polipéptido de la presente invención en la planta. Por ejemplo, el elemento intensificador del promotor puede ser un intrón que se coloca entre el promotor y la secuencia de nucleótidos que codifica un polipéptido de la presente invención. Por ejemplo, Xu et al., 1993, *supra*, revelan el uso del primer intrón del gen de actina 1 de arroz para mejorar la expresión.

[0173] El gen marcador seleccionable y cualquier otra parte del constructo de expresión se puede elegir de entre aquellos disponibles en la técnica.

[0174] El constructo de ácidos nucleicos se incorpora en el genoma de planta según técnicas convencionales conocidas en la técnica, incluida la transformación mediada por *Agrobacterium*, la transformación mediada de virus, la microinyección, el bombardeo de partícula, la transformación biolística y la electroporación (Gasser et al., 1990, Science 244: 1293; Potrykus, 1990, Bio/Technology 8: 535; Shimamoto et al., 1989, Nature 338: 274).

[0175] Actualmente, la transferencia de genes mediada por *Agrobacterium tumefaciens* es el método de elección para generar dicotiledóneas transgénicas (para una revisión, véase Hooykas y Schilperoort, 1992, Plant Molecular Biology 19:15-38) y también pueden usarse para transformar monocotiledóneas, aunque frecuentemente se utilizan otros métodos de transformación para estas plantas. Actualmente, el método de elección para generar monocotiledóneas transgénicas es 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 protoplastos como lo describe Omirulleh et al., 1993, Plant Molecular Biology 21: 415-428.

[0176] 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 bien durante la regeneración o en las siguientes generaciones usando, por ejemplo, la cotransformación con dos Constructos de T-ADN separados o la escisión específica de sitio del gen de selección por una recombinasa específica.

[0177] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que comprenden (a) el cultivo de una planta transgénica o una célula vegetal que comprende un polinucleótido que codifica un polipéptido con actividad de beta-glucosidasa de la presente invención en las condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido.

Péptido señal

[0178] La presente divulgación también se refiere a constructos de ácidos nucleicos que comprenden un gen que codifica una proteína operativamente enlazado a una secuencia de nucleótidos que consiste en los nucleótidos 1 a 58 de la SEC ID n°: 1 que codifica un péptido señal que consiste en los aminoácidos 1 a 19 de la SEC ID n°: 2, donde el gen es foráneo a la primera y segunda secuencia de nucleótidos.

[0179] La proteína puede ser nativa o heteróloga a una célula huésped. El término "proteína" no se utiliza aquí para referirse a una longitud específica del producto codificado y, por lo tanto, abarca péptidos, oligopéptidos y proteínas. El término "proteína" también abarca dos o más polipéptidos combinados para formar el producto codificado. Las proteínas también incluyen polipéptidos híbridos que comprenden una combinación de secuencias de polipéptidos completas o parciales obtenidas de al menos dos proteínas diferentes donde una o más pueden ser heterólogas o nativas a la célula huésped. Las proteínas además incluyen variantes de las proteínas mencionadas anteriormente diseñadas genéticamente y alélicas de origen natural y proteínas híbridas.

[0180] Preferiblemente, la proteína es una hormona o variante de la misma, enzima, receptor o parte del mismo, anticuerpo o parte del mismo, o indicador. En un aspecto más preferido, la proteína es una oxidoreductasa, transferasa, hidrolasa, liasa, isomerasa o ligasa. En un aspecto incluso más preferido, la proteína es una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, ciclodextrina glicosiltransferasa, desoxirribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, invertasa, lacasa, lipasa, manosidasa, mutanasa, oxidasa, enzima pectinolítica, peroxidasa, fitasa, polifenoloxidasas, enzima proteolítica, ribonucleasa, transglutaminasa o xilanasas.

[0181] El gen se puede obtener de cualquier procariótica, eucariótica u otra fuente.

[0182] La presente invención aparece descrita posteriormente a través de los siguientes ejemplos, que no deberían ser interpretados como una limitación del ámbito de la invención.

Ejemplos

Materiales

[0183] Los productos químicos usados como tampones y sustratos fueron productos comerciales de al menos calidad reactiva.

Cepas

[0184] La cepa *Aspergillus oryzae* Jal250 (WO 99/61651) fue usada para la expresión de la beta-glucosidasa de *Aspergillus fumigatus*. El *Aspergillus fumigatus* PaHa34 fue usado como fuente de la beta-glucosidasa de la familia GH3A.

Medios

[0185] El medio de dextrosa de patata fue compuesto por litro de 39 gramos de dextrosa de patata (Difco).

[0186] Las placas de PDA fueron compuestas por litro de 39 gramos de agar de dextrosa de patata.

[0187] El medio MDU2BP fue compuesto por litro de 45 g de maltosa, 1 g de $MgSO_4 \cdot 7H_2O$, 1 g de NaCl, 2 g de K_2SO_4 , 12 g de KH_2PO_4 , 7 g de extracto de levadura, 2 g de urea, y 0,5 ml de solución de metales traza AMG, pH a 5,0.

[0188] La solución de metales traza AMG fue compuesta por litro de 14,3 g de $ZnSO_4 \cdot 7H_2O$, 2,5 g de $CuSO_4 \cdot 5H_2O$, 0,5 g de $NiCl_2 \cdot 6H_2O$, 13,8 g de $FeSO_4 \cdot 7H_2O$, 8,5 g de $MnSO_4 \cdot H_2O$, y 3 g de ácido cítrico.

[0189] Las placas Cove fueron compuestas por litro de 342,3 g de sacarosa, 20 ml de solución salina Cove, 10 ml de 1 M acetamida, 10 ml de 1,5 M CsCl2 y 25 g de agar Noble.

[0190] La solución salina Cove fue compuesta por litro de 26 g de KCl, 26 g de $MgSO_4 \cdot 7H_2O$, 76 g de KH_2PO_4 y 50 ml de solución de metales traza COVE.

[0191] La solución de metales traza Cove fue compuesta por litro de 0,04 g de $Na_2B_4O_7 \cdot 10H_2O$, 0,4 g de $CuSO_4 \cdot 5H_2O$, 1,2 g de $FeSO_4 \cdot 7H_2O$, 0,7 g de $MnSO_4 \cdot H_2O$, 0,8 g de $Na_2MoO_4 \cdot 2H_2O$, y 10 g de $ZnSO_4 \cdot 7H_2O$.

[0192] El medio CIM fue compuesto por litro de 20 g de celulosa, 10 g de sólidos de maíz fermentados, 1,45 g de $(NH_4)_2SO_4$, 2,08 g de KH_2PO_4 , 0,28 g de $CaCl_2$, 0,42 g de $MgSO_4 \cdot 7H_2O$, y 0,42 ml de solución de metales traza, pH a 6,0.

[0193] La solución de metales traza fue compuesta por litro de 41,2 mg de $FeCl_3 \cdot 6H_2O$, 11,6 mg de $ZnSO_4 \cdot 7H_2O$, 5,4 mg de $MnSO_4 \cdot H_2O$, 2,0 mg de $CuSO_4 \cdot 5H_2O$, 0,48 mg de H_3BO_3 y 67,2 mg de ácido cítrico.

Ejemplo 1: identificación de un gen de la familia GH3A de la glicosil hidrolasa en la secuencia genómica del *Aspergillus fumigatus*

[0194] Una búsqueda tblastn (Altschul et al., 1997, Nucleic Acids Res. 25: 3389-3402) de la secuencia genómica parcial de *Aspergillus fumigatus* (The Institute of Genomic Research, Rockville, MD) se efectuó usando como criterio de búsqueda una secuencia de proteínas de beta-glucosidasa de *Aspergillus aculeatus* (Nº de acceso P48825). Diferentes genes fueron identificados como homólogos putativos de la familia GH3A en base a un grado alto de similitud con la secuencia del criterio de búsqueda en el nivel aminoácido. Una región genómica de aproximadamente 3000 pares de bases con más de un 70% de identidad con la secuencia del criterio de búsqueda en el nivel aminoácido fue elegido para un posterior estudio.

Ejemplo 2: Extracción de ADN genómico de *Aspergillus fumigatus*

[0195] Se cultivó *Aspergillus fumigatus* en 250 ml de medio de dextrosa de patata en un matraz de agitación disipado a 37 °C y 240 r.p.m. Los micelios se cosecharon por filtración, se lavaron dos veces en TE (10 mM de Tris-1 mM EDTA) y se congelaron en nitrógeno líquido. Los micelios congelados se cultivaron, por mortero y mano de mortero, hasta obtener un polvo fino, que se resuspendió en tampón a pH 8,0 con 10 mM Tris, 100 mM EDTA, 1% de Triton X-100, 0,5 a X M HCl de guanidina y 200 mM NaCl. Se añadió ribonucleasa pancreática A sin ADNsa a una concentración de 20 mg/litro y el lisado se incubó a 37 °C durante 30 minutos. El detrito celular se eliminó por centrifugado y el ADN se aisló usando una columna Qiagen Maxi 500 (QIAGEN Inc., Chatswort, CA). Las columnas se equilibraron en 10 ml de QBT, se lavaron con 30 ml de QC y se eluyeron con 15 ml de QF (todos los tampones de QIAGEN Inc., Chatswort, CA). El ADN se precipitó en isopropanol, se lavó en etanol al 70% y se recuperó por centrifugado. El ADN se resuspendió en tampón TE.

Ejemplo 3: construcción del vector de expresión pAIIo2

5 **[0196]** El vector de expresión pAIIo1 se construyó modificando pBANE6 (patente estadounidense 6.461.837), que comprende un híbrido de los promotores de los genes para alfa-amilasa neutra de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae* (promotor NA2-tpi), la secuencia del terminador de amiloglucosidasa de *Aspergillus niger* (terminador AMG) y el gen de acetamidasa de *Aspergillus nidulans* (amdS). La modificación de pBANE6 se realizó en primer lugar eliminando tres sitios de restricción *Nco* I en las posiciones 2051, 2722 y 3397 pares de bases del marcador de selección *amdS* por mutagénesis dirigida. Todos los cambios fueron diseñados para ser "silenciosos" dejando la secuencia de proteína real del producto genético de *amdS* invariada. La eliminación de estos tres sitios se realizó 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):

15 AMDS3NcoMut (2050): 5'-GTGCCCCATGATACGCCTCCGG-3' (SEC ID nº: 3)
 AMDS2NcoMut (2721): 5'-GAGTCGTATTTCCAAGGCTCCTGACC-3' (SEC ID nº: 4)
 AMDS1NcoMut (3396): 5'-GGAGGCCATGAAGTGGACCAACGG-3' (SEC ID nº: 5)

20 **[0197]** Un plásmido que comprende los tres cambios de secuencia previstos se sometió 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. Se usaron los siguientes cebadores (el nucleótido subrayado representa la base cambiada) para la mutagénesis:

25 Cebador superior para mutagenizar la secuencia del terminador AMG:
 5'-CACCGTGAAAGCCATGCTCTTTCCTTCGTGTAGAAGACCAGACAG-3' (SEC ID nº: 6)
 Cebador inferior para mutagenizar la secuencia del terminador el AMG:
 5'-CTGGTCTTCTACACGAAGGAAAGAGCATGGCTTTCACGGTGTCTG-3' (SEC ID nº: 7)

30 **[0198]** La última fase en la modificación de pBANE6 fue la adición de un nuevo sitio de restricción *Nco* I en el inicio del poliligador usando un equipo de mutagénesis QuickChange y los siguientes cebadores (los nucleótidos subrayados representan las bases cambiadas) para producir pAIIo1 (**Figura 2**).

35 Cebador superior para mutagenizar el promotor NA2-tpi:
 5'-CTATATACACAACCTGGATTTACCATGGGCCCGCGGCCGCGCAGATC-3' (SEC ID nº: 8)
 Cebador inferior para mutagenizar el promotor NA2-tpi:
 5'-GATCTGCGGCCGCGGCCCATGGTAAATCCAGTTGTGTATATAG-3' (SEC ID nº: 9)

40 **[0199]** El gen *amdS* de pAIIo1 fue trocado con el gen de *Aspergillus nidulans pyrG*. El plásmido pBANE10 (**Figura 3**) se usó como una fuente para el gen *pyrG*. El análisis de la secuencia de pBANE10 mostró que el marcador *pyrG* estaba contenido dentro de un fragmento de restricción *Nsi* I y no contiene ninguno de los sitios de restricción *Nco* I o *Pac* I. Puesto que el *amdS* también está flanqueado por los sitios de restricción *Nsi* I, la estrategia para cambiar el marcador de selección fue un simple cambio de los fragmentos de restricción de *Nsi* I. El ADN plásmido de pAIIo1 y pBANE10 fueron digeridos con la enzima de restricción *Nsi* I y los productos se purificaron por electroforesis en gel de agarosa usando procedimientos estándar. El fragmento *Nsi* I de pBANE10 que contenía el gen *pyrG* gen fue ligado al esqueleto de pAIIo1 para reemplazar el fragmento original de ADN *Nsi* I que contenía el gen *amdS*. Los clones recombinantes fueron analizados por digestión de la restricción para determinar si tenían el inserto y la orientación correctas. Se selecciona un clon con el gen *pyrG* transcrito en el sentido contrario a las agujas del reloj. El nuevo plásmido se ha designado como pAIIo2 (**Figura 4**).

Ejemplo 4: clonación del gen de beta-glucosidasa de la familia GH3A y construcción de un vector de expresión de *Aspergillus oryzae*

50 **[0200]** Los dos cebadores oligonucleótidos sintéticos mostrados abajo fueron diseñados para amplificar por PCR un gen de *Aspergillus fumigatus* que codifica una beta-glucosidasa putativa de la familia GH3A a partir del ADN genómico preparado en el Ejemplo 2. Un equipo de clonación InFusion (BD Biosciencias, Palo Alto, CA) fue usado para clonar el fragmento directamente en el vector de expresión, pAIIo2, sin necesidad de digeridos de restricción y ligadura.

60 Cebador directo: 5'-ACTGGATTTACCATGAGATTCGGTTGGCTCG-3' (SEC ID nº: 10)
 Cebador inverso: 5'-AGTCACCTCTAGTTACTAGTAGACACGGGGC-3' (SEC ID nº: 11)
 Las letras en negrita representan la secuencia codificante. La secuencia restante es homóloga a los sitios de inserción de pAIIo2.

65 **[0201]** 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 mezcla de dATP, dTTP, dGTP y dCTP, 2,5 unidades de polimerasa de ADN Platinum Pfx (Invitrogen, Carlsbad, CA), 1 µl de 50 mM 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 fue conducido más tarde a un ciclo de remojo de 4 °C.

[0202] Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando un tampón de 40 mM de Tris base-20 mM de acetato de sodio-1 mM de disodio EDTA (TAE) donde una banda de producto 3 kb fue excindida del gel y purificada usando un equipo de extracción de gel QIAquick (QIAGEN, Chatswort, CA) según las instrucciones del fabricante.

[0203] El fragmento fue clonado más tarde en el vector de expresión pALo2 usando un equipo de clonación InFusion. El vector fue digerido con endonucleasas de restricción *Nco* I y *Pac* I (bajo las condiciones de uso especificadas por el fabricante). El fragmento fue purificado por electroforesis de gel y purificación de gel QIAquick. El fragmento del gen y el vector de corte fueron ligados en una reacción dando como resultado el plásmido de expresión pEJG97 (**Figura 5**) en el que la transcripción del gen de beta-glucosidasa de la familia GH3A estuvo bajo el control del promotor NA2-tpi. La reacción de ligadura (50 µl) estaba compuesta por un tampón InFusion 1X (BD Biosciences, Palo Alto, CA), 1X BSA (BD Biosciences, Palo Alto, CA), 1 µl de enzima InFusion (diluida 1:10) (BD Biosciences, Palo Alto, CA), 150 ng de pALo2 digerido con *Nco* I y *Pac* I y 50 ng del producto PCR purificado de beta-glucosidasa de *Aspergillus fumigatus*. La reacción se incubó a temperatura ambiente durante 30 minutos. 1 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 se preparó ADN plásmido usando un Biorobot 9600 (QIAGEN, Inc., Chatswort, CA)

Ejemplo 5: caracterización secuencia genómica del *Aspergillus fumigatus* que codifica una beta-glucosidasa de la familia GH3A

[0204] La secuenciación del ADN del gen de beta-glucosidasa de *Aspergillus fumigatus* a partir de pEJG97 fue realizada con un secuenciador de ADN automático modelo 377 XL de Perkin-Elmer Applied Biosystem (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA) usando la química de terminador de coloración (Giesecke et al., 1992, Journal of Virology Methods 38: 47-60) y la estrategia de primer-walking. Los datos de la secuencia de nucleótidos fueron analizados en cuanto a la calidad y se compararon todas las secuencias entre sí con asistencia del software PHRED/PHRAP (Universidad de Washington, Seattle, WA).

[0205] Se construyó un modelo de gen para la secuencia *Aspergillus fumigatus* sobre la base de la similitud con genes homólogos de *Aspergillus aculeatus*, *Aspergillus niger* y *Aspergillus kawachii*. La secuencia de nucleótidos (SEC ID n°: 1) y la secuencia de aminoácidos deducida (SEC ID n°: 2) se muestran en las **Figuras 1A y 1B**. El fragmento genómico codifica un polipéptido de 863 aminoácidos, interrumpido por 8 intrones de 62, 55, 58, 63, 58, 58, 63 y 51 pares de bases. El contenido %G+C del gen es 54,3% y la región de codificación madura es 53,65%. Usando el 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.

[0206] Se determinó una alineación comparativa de secuencias de beta-glucosidasa usando el método de Clustal W (Higgins, 1989, CABIOS 5: 151-153) que usa el software LASERGENE™ MEGALIGN™ (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 un 78%, 76% y 76% de identidad con las secuencias de aminoácidos deducidas de beta-glucosidasas de *Aspergillus aculeatus* (número de acceso P48825), *Aspergillus niger* (000089) y *Aspergillus kawachii* (P87076).

Ejemplo 6: expresión del gen de beta-glucosidasa de *Aspergillus fumigatus* de la familia GH3A en *Aspergillus oryzae* JAL250

[0207] Los protoplastos de *Aspergillus oryzae* Jal250 fueron preparados según el método de Christensen et al., 1988, Bio/Technology 6: 1419-1422. Se utilizaron cinco µg de pEJG97 (y pALo2 como control de vector) para transformar *Aspergillus oryzae* JAL250.

[0208] La transformación de *Aspergillus oryzae* Jal250 con pEJG97 produjo aproximadamente 100 transformantes. Se aislaron diez transformantes en placas de PDA individuales.

[0209] Las placas de PDA confluentes de cinco de los diez transformantes fueron lavadas con 5 ml de Tween 20 al 0,01% e inoculadas por separado 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 Tris-glicina al 8-16% (Invitrogen, Carlsbad, CA) según las instrucciones del fabricante. Los perfiles SDS-PAGE de los cultivos mostraron que uno de los transformantes (designados transformante 1) tuvo una banda mayoritaria de aproximadamente 130 kDa.

[0210] Una placa confluyente de transformante 1 (cultivado 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 la caracterización de la enzima. El matraz fue recogido el día 5 y filtrado usando una membrana GP Express de 0,22 µm (Millipore, Bedford, MA).

5 Ejemplo 7: construcción de pMJ09

10 [0211] El vector pMJ04 fue construido amplificando por PCR el terminador de gen de celobiohidrolasa 1 *Trichoderma reesei* Cel7A (*cbh1*) del ADN genómico de *Trichoderma reesei* RutC30 usando los cebadores 993429 (antisentido) y 993428 (sentido) mostrados más abajo. El cebador antisentido fue diseñado genéticamente para tener un sitio *Pac* I en el extremo 5' y un sitio *Spe* I en el extremo 5' del cebador sentido.

Cebador 993429 (antisentido):
5'-AACGTTAATTAAGGAATCGTTTTGTGTTT-3' (SEC ID n°: 12)

15 Cebador 993428 (sentido):
5'-AGTACTAGTAGCTCCGTGGCGAAAGCCTG-3' (SEC ID n°: 13)

20 [0212] Las reacciones de amplificación (50 µl) fueron compuestas de un tampón de reacción 1X TernoPol (New England Biolabs, Beverly, MA), 0,3 mM dNTPs, 100 ng de ADN genómico de *Trichoderma reesei* RutC30 (que fue aislado usando un equipo DNeasi Plant Maxi de Qiagen, Chatswort, CA), 0,3 µM cebador 993429, 0,3 µM cebador 993428 y 2 unidades de Vent polimerasa (New England Biolabs, Beverly, MA). Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 (Hamburgo, Alemania) programado de la siguiente manera: 30 ciclos cada una durante 30 segundos a 94°C, 30 segundos a 55 °C y 30 segundos a 72 °C (extensión final de 15 minutos).

25 [0213] Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando un tampón de 40 mM de Tris base-20 mM de acetato de sodio-1 mM de disodio EDTA (TAE) donde una banda de producto de 229 pares de bases fue cortada del gel y purificada usando un equipo de extracción de gel QIAGEN QIAquick según las instrucciones del fabricante.

30 [0214] El fragmento de PCR resultante fue digerido con *Pac* I y *Spe* I y ligado en pAIIo1 digerido con las mismas enzimas de restricción usando un equipo de ligadura rápida (Roche, Indianapolis, IN) para generar pMJ04 (Figura 6).

35 [0215] El vector pMJ06 fue construido amplificando por PCR el promotor de gen de celobiohidrolasa 1 *Trichoderma reesei* Cel7A (*cbh1*) del ADN genómico de *Trichoderma reesei* RutC30 usando los cebadores 993696 (antisentido) y 993695 (sentido) mostrados más abajo. El cebador antisentido fue creado genéticamente para tener un sitio *Sal* I en el extremo 5' del cebador de sentido y un sitio *Nco* I en el extremo 5' del cebador antisentido.

Cebador 993695 (sentido):
5'-ACTAGTCGACCGAATGTAGGATTGTT-3' (SEC ID n°: 14)

40 Cebador 993696 (antisentido):
5'-TGACCATGGTGCAGTCC-3' (SEC ID n°: 15)

45 [0216] Las reacciones de amplificación (50 µl) fueron compuestas de un tampón de reacción 1X TernoPol, 0,3 mM dNTPs, 100 ng de ADN genómico *Trichoderma reesei* RutC30 (que fue preparado usando un equipo QIAGEN DNeasi Plant Maxi), 0,3 µM cebador 993696, 0,3 µM 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 una durante 30 segundos a 94 °C, 30 segundos a 55 °C, y 60 segundos a 72 °C (extensión final de 15 minutos).

50 [0217] Los productos reactivos fueron aislados en un gel de agarosa al 1,0% donde una banda de producto de 988 pares de bases fue cortada del gel y purificada usando un equipo de extracción de gel QIAGEN QIAquick según las instrucciones del fabricante.

[0218] El fragmento de PCR resultante fue digerido con *Nco* I y *Sal* I y ligado en pMJ04 digerido con las mismas enzimas de restricción usando de un equipo de ligadura rápida para generar pMJ06 (Figura 7).

55 [0219] El vector de expresión pMJ09 fue construido amplificando por PCR el terminador gen de celobiohidrolasa 1 *Trichoderma reesei* Cel7A (*cbh1*) de ADN genómico de *Trichoderma reesei* RutC30 usando los cebadores 993843 (antisentido) y 993844 (sentido) mostrados más abajo. El cebador antisentido fue creado genéticamente para tener un sitio *Pac* I y un sitio *Spe* I en el extremo 5' y un sitio *Pvu* I en el extremo 5' del cebador sentido.

60 Cebador 993844 (sentido):
5'-CGATCGTCTCCCTATGGGTCATTACC-3' (SEC ID n°: 16)

Cebador 993843 (antisentido):
5'-ACTAGTTAATTAAGCTCCGTGGCGAAAG-3' (SEC ID n°: 17)

65 [0220] Las reacciones de amplificación (50 µl) fueron compuestas de un tampón de reacción 1X TernoPol, 0,3 mM dNTPs, 100 ng de ADN genómico *Trichoderma reesei* RutC30 (que fue extraído usando un equipo QIAGEN DNeasi

Plant Maxi), 0,3 μ M cebador 993844, 0,3 μ M cebador 993843 y 2 unidades de Vent de polimerasa. Las reacciones fueron incubadas en un Eppendorf Mastercycler 5333 programado de la siguiente manera: 30 ciclos cada una durante 30 segundos a 94 °C, 30 segundos a 55 °C y 60 segundos a 72 °C (extensión final de 15 minutos).

5 **[0221]** Los productos reactivos fueron aislados en un gel de agarosa al 1% usando un 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 QIAGEN QIAquick según las instrucciones del fabricante.

10 **[0222]** El fragmento de PCR resultante fue digerido con *Pvu* I y *Spe* I y ligado en pMJ06 digerido con *Pac* I y *Spe* I usando un equipo de ligadura rápida para generar pMJ09 (**Figura 8**).

Ejemplo 8: expresión del gen de beta-glucosidasa de *Aspergillus fumigatus* de la familia GH3A en *Trichoderma reesei*

15 **[0223]** Los dos cebadores oligonucleótidos sintéticos mostrados abajo fueron diseñados para amplificar por PCR un gen de beta-glucosidasa de *Aspergillus fumigatus* a partir de pEJG97 como se describe en el Ejemplo 4. Se utilizó un equipo de clonación InFusion para clonar el fragmento directamente en el vector de expresión, pMJ09, sin necesidad de digeridos de restricción y ligadura.

20 Cebador directo: 5'-GGACTGCGCACCATGAGATTCCGGTTGGCTC-3' (SEC ID n°: 18)
Cebador inverso: 5'-TCGCCACGGAGCTTACTAGTAGACACGGGG-3' (SEC ID n°: 19)

25 **[0224]** Las letras en negrita representan la secuencia codificante. La secuencia restante es homóloga a los sitios de inserción de pMJ09.

30 **[0225]** Cincuenta picomoles de cada uno de los cebadores mencionados fueron usados en una reacción de PCR (50 μ l) con 100 ng de ADN pEJG97, tampón de amplificación 1X Pfx (Invitrogen, Carlsbad, CA), 1,5 μ l de 10 mM mezcla de dATP, dTTP dGTP, y dCTP, 2,5 unidades de platino polimerasa de ADN Platinum Pfx (Invitrogen, Carlsbad, CA), 1 μ l de 50 mM MgSO₄ y 2,5 μ l solución intensificadora 10X pCRx (Invitrogen, Carlsbad, CA). 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 fue conducido más tarde a un ciclo de remojo de 4 °C.

35 **[0226]** Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando un tampón TAE donde una banda de producto 3 kb fue cortada del gel y purificada usando un equipo de extracción de gel QIAGEN QIAquick según las instrucciones del fabricante.

40 **[0227]** El fragmento 3 kb purificado fue clonado más tarde en pMJ09 usando un equipo de clonación InFusion. El vector fue digerido con *Nco* I y *Pac* I. El fragmento fue purificado por electroforesis de gel purificación de gel QIAquick, como se ha descrito anteriormente. El fragmento del gen y el vector digerido fueron ligados en una reacción dando como resultado el plásmido de expresión pEJG107 (**Figura 9**) en el que la transcripción del gen de beta-glucosidasa de la familia GH3A estuvo bajo el control del promotor *Trichoderma reesei cbhl*. La ligadura (50 μ l) estaba compuesta de un tampón InFusion 1X, 1X BSA, 1 μ l de enzima InFusion (diluida 1:10), 100 ng de pMJ09 digerido con *Nco* I y *Pac* I y 100 ng del producto PCR purificado de beta-glucosidasa de *Aspergillus fumigatus*. La reacción se incubó a temperatura ambiente durante 30 minutos. 1 μ l de la reacción fue usado para transformar las células SURE de *E. coli* electropetente (Stratagene, La Jolla, CA). Un transformante de *E. coli* con el plásmido pEJG107 fue detectado por digestión de restricción y se preparó ADN plásmido usando un Biorobot 9600 (QIAGEN, Inc., Chatsworth, CA)

50 **[0228]** La secuenciación del ADN del gen de beta-glucosidasa de *Aspergillus fumigatus* a partir de pEJG107 fue realizada con un secuenciador de ADN automático modelo 377 XL de Perkin-Elmer Applied Biosystems (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA) usando la química de terminador de coloración (Giesecke et al., 1992, Journal of Virology Methods 38: 47-60) y estrategia de *primer-walking*. Los datos de la secuencia de nucleótidos fueron analizados en cuanto a la calidad y se compararon todas las secuencias entre sí con asistencia del software PHRED/PHRAP (Universidad de Washington, Seattle, WA).

55 **[0229]** Los protoplastos de *Trichoderma reesei* RutC30 fueron preparados según el método de Christensen et al., 1988, *supra*. Se utilizaron cinco μ g de pEJG107 para transformar *Trichoderma reesei* RutC30. La transformación produjo aproximadamente 100 transformantes. Se aislaron sesenta transformantes en placas COVE/10 mM de uridina e incubados a 28 °C.

60 **[0230]** Se recogieron esporas de las placas de 8 de los 60 transformantes pipeteando tres veces con un bucle estéril e inoculadas por separado en 25 ml de medio CIM en frascos de agitación de vidrio 125 ml e incubadas a 28 °C, 250 r.p.m. Cinco días después de la incubación, se analizaron 0,5 μ l de sobrenadante de cada cultivo usando geles Tris SDS-PAGE al 7,5% (Biorad, Hércules, CA) según las instrucciones del fabricante. Los perfiles SDS-PAGE de los cultivos mostraron que uno de los transformantes (designado transformante 1) tuvo una banda mayoritaria de aproximadamente 130 kDa.

Ejemplo 9: caracterización de la beta-glucosidasa de *Aspergillus fumigatus*

5 [0231] Una alícuota de 3 ml de la beta-glucosidasa de *Aspergillus fumigatus*, obtenida como se describe en el Ejemplo 6, fue desalinizado usando una columna de desalinización Econo-pAc 10DG de BioRad (BioRad, Hercules, CA), dando como resultado aproximadamente 4 ml de caldo desalinizado en 100 mM de citrato sódico pH 5,0. El caldo desalinizado fue concentrado más tarde a aproximadamente a 180 μ l usando un Amicon Centricon Plus-20 (filtro BioBiomax-5,5 kD, membrana PES) y diluido con 100 mM de citrato sódico pH 5,0 en un volumen final de aproximadamente 500 μ l. El ensayo 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, el cual mostró resultados similares.

10 [0232] El SDS-PAGE (Tris-HCl al 7,5%, según el criterio BioRad) de las muestras desalinizadas concentradas mostró una banda mayoritaria de aproximadamente 130 kD. Se recortó del gel la banda mayoritaria de 130 kD y se sometió a la secuenciación de aminoácidos N-terminal.

15 [0233] Se evaluó la termoestabilidad de la beta-glucosidasa de *Aspergillus fumigatus* del caldo desalinizado/concentrado a 50 °C, 65 °C y 70 °C. Las condiciones del ensayo a 50 °C y 65 °C fueron: 100 mM decitrato sódico pH 5,0, Tween-20 al 0,01%, 4 mM de p-nitrofenil-beta-D-glucopiranosido, [proteína]_{AfumGH3A} = 6.9×10^{-6} mg/ml, incubado a 50 °C y 65 °C. Las alícuotas se tomaron a las 0,5, 1, 2, 3, 3,75, y 24 horas. Para cada alícuota se añadió 1 M de carbonato de sodio pH 10,0 y se determinó que la absorbencia de la concentración de anión de p-nitrofenol era de 405 nm. A 70 °C, las condiciones del 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 alícuotas se tomaron a las 0,25, 0,5, 1, 2, 4 y 5 horas. Para cada alícuota se añadió 1 M de carbonato de sodio pH 10,0 y se determinó que la absorbencia de la concentración de anión de p-nitrofenol era de 405 nm. Nótese que cada una de las concentraciones de proteína mencionadas se refiere a la concentración total de proteína en el ensayo, puesto que todas fueron evaluadas como caldos en lugar de como enzimas purificadas.

20 [0234] Los resultados se muestran en las **Figuras 10** y **11**. La termoestabilidad fue definida como cinética de reacción lineal para un intervalo de tiempo determinado, dentro de una conversión a anión de p-nitrofenol de porcentaje razonable (<15%). La beta-glucosidasa de *Aspergillus fumigatus* pareció mantenerse estable durante al menos 24 horas a 50 °C, aunque no hubo puntos temporales entre las 4 y las 24 horas. A 65 °C pareció mantenerse estable durante al menos 3,75 horas, pero después la estabilidad disminuyó gradualmente con una conversión a anión de p-nitrofenol del 65% a las 24 horas. Esta conversión fue bastante alta, con lo que algunas de las reducciones observadas en el índice pueden ser debidas a una depleción del sustrato y/o inhibición del producto además de a una posible inactivación térmica. A 70 °C pareció mantenerse estable durante 1 hora, y razonable estable durante 2 horas.

25 [0235] El caldo que contiene la beta-glucosidasa de *Aspergillus fumigatus* de la familia GH3A, obtenido como se describe en el Ejemplo 6, fue primero desalinizado (columna BioRad Econo-pAc 10DG) y concentrado (Centricon Plus-20, filtro BioBiomax-5,5 kD), en una concentración de 0,92 mg/ml (ensayo BCA). Después, fue incubado en 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 alícuotas se tomaron a las 0,5, 1, 2, 3, 4, 6, y 19 horas. Las alícuotas fueron hervidas durante 6 minutos para terminar la reacción y se determinó la concentración de glucosa usando el ensayo Trinder (Sigma Chemical Co., St. Luis, MO) y niveles de glucosa externa. Los resultados se muestran en la **Figura 12**. La beta-glucosidasa pareció mantener el 90% de su actividad hasta 6 horas y el 65% hasta 19 horas a 65 °C con la menor carga de proteínas (0,0092 μ g/ml). La beta-glucosidasa pareció mantenerse razonablemente estable hasta 6 horas a 65 °C.

Depósito de Material biológico

30 [0236] El siguiente material biológico ha sido depositado según las condiciones del Tratado de Budapest con la colección de Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, 1815 University Street, Peoria, Illinois, 61604, y se le ha asignado el siguiente número de :

Depósito Número de acceso Fecha de depósito
 E. coli TOP10 (pEJG113) NRRL B-30695 17 de octubre de 2003

35 [0237] La cepa ha sido depositada en condiciones que aseguran que el acceso al cultivo estará disponible durante la pendencia de esta solicitud de patente para una persona determinada por el Comisario de Patentes y Marcas Registradas para tener derecho a la misma bajo 37 C.F.R. §1,14 y 35 U.S.C. §122. El depósito representa un cultivo substancialmente puro de la cepa depositada. El depósito está disponible según se requiera por las leyes de patentes extranjeras en países donde se presenten duplicados de la solicitud en cuestión, o sus descendientes. No obstante, se debe entender que la disponibilidad de un depósito no constituye una licencia para practicar la presente invención sujeta en derogación de los derechos de las patentes garantizados por la acción gubernamental.

LISTA DE SECUENCIAS

60 [0238]

ES 2 437 198 T3

<110> Novozymes Biotech, Inc. Harris, Paul Zaretsky, Elizabeth

<120> Polipéptidos que tienen actividad de beta-glucosidasa y ácidos nucleicos que los codifican

5 <130> 10545.204-WO

<150> 60/515,482

<151> 2003-10-28

10 <160> 19

<170> PatentIn Version 3.2

<210> 1

15 <211> 3060

<212> ADN

<213> Aspergillus fumigatus

<400> 1

```

atgagattcg gttggctcga ggtggccgct ctgacggccg cttctgtagc caatgccag      60
gtttgtgatg ctttcccgtc attgtttcgg atatagttga caatagtcac ggaaataatc     120
aggaattggc tttctctcca ccattctacc cttcgccttg ggctgatggc cagggagagt     180
gggcagatgc ccatcgacgc gccgtcgaga tcgtttctca gatgacactg gcggagaagg     240
ttaaccttac aacgggtact ggggtgggtg cgactttttt gttgacagtg agctttcttc     300
actgaccatc tacacagatg ggaaatggac cgatgcgctg gtcaaaccgg cagcgttccc     360
aggtaagctt gcaattctgc aacaacgtgc aagtgtagtt gctaaaacgc ggtggtgcag     420
acttggatc aactggggtc tttgtggcca ggattcccct ttgggtatcc gtttctgtga     480
gctatacccg cggagtcttt cagtccttgt attatgtgct gatgattgtc tctgtatagc     540
tgacctcaac tccgccttcc ctgctggtac taatgtcgcc gcgacatggg acaagacact     600
cgcctacctt cgtggcaagg ccatgggtga ggaattcaac gacaagggcg tggacatttt     660
gctggggcct gctgctggtc ctctcgcaa ataccggac ggcggcagaa tctgggaagg     720
cttctctcct gatccggttc tactggtgt acttttcgcc gaaactatca agggtatcca     780
agacgcgggt gtgattgcta ctgccaagca ttacattctg aatgaacagg agcatttccg     840
acaggttggc gaggcccagg gatatggtta caacatcacg gagacgatca gctccaacgt     900
ggatgacaag accatgcacg agttgtacct ttggtgagta gttgacactg caaatgagga     960
ccttgattga tttgactgac ctggaatgca ggccctttgc agatgctgtg cgcggtaaga    1020
ttttccgtag acttgacctc gcgacgaaga aatcgctgac gaaccatcgt agctggcggt    1080

```

20

ES 2 437 198 T3

ggcgctgtca tgtgttccta caatcaaate aacaacagct acggttgtca aaacagtcaa 1140
 actctcaaca agctcctcaa ggctgagctg ggcttccaag gcttcgcatc gagtgactgg 1200
 agcgctcacc acagcgggtg cggcgctgcc ctcgctgggt tggatatgtc gatgcctgga 1260
 gacatttctc tcgacgacgg actctccttc tggggcacga acctaactgt cagtgttctt 1320
 aacggcaccg ttccagcctg gcgtgtcgat gacatggctg ttcgtatcat gaccgcgtac 1380
 tacaaggttg gtcgtgaccg tcttctgatt cccctaactc tcagctcctg gaccgggat 1440
 ggtacggctt gggagcattc tgctgtctcc gagggagcct ggaccaaggt gaacgacttc 1500
 gtcaatgtgc agcgcagtca ctctcagatc atccgtgaga ttggtgccgc tagtacagtg 1560
 ctcttgaaga acacgggtgc tcttctttg accggcaagg aggttaaagt ggggtgtctc 1620
 ggtgaagacg ctggttccaa cccgtggggg gctaaccgct gccccgaccg cggctgtgat 1680
 aacggcactc ttgctatggc ctggggtagt ggtactgcca acttccctta ccttgtcacc 1740
 cccgagcagg ctatccagcg agaggctatc agcaacggcg gcaatgtctt tgctgtgact 1800
 gataacgggg ctctcagcca gatggcagat gttgcatctc aatccagggt agtgccggct 1860
 cttagaaaaa gaacgttctc tgaatgaagt ttttaacca ttgcgaacag cgtgtctttg 1920
 gtgtttgtca acgcccactc tggagagggg ttcacagctg tcgacggcaa cgagggtgac 1980
 cgcaaaaatc tcaactctgt gaagaacggc gaggcgctca ttgacactgt tgtcagccac 2040
 tgcaacaaca cgattgtggg tattcacagt gttgggcccg tcttgatcga ccgggtggat 2100
 gataacccca acgtcaactgc catcatctgg gccggcttgc ccggtcagga gagtggcaac 2160
 tccttggtcg acgtgctcta tggccgctc aaccccagcg ccaagacccc gttcacctgg 2220
 ggcaagactc gggagtctta cggggctccc ttgctcaccg agcctaacia tggcaatggg 2280
 gctcccagc atgatttcaa cgaggcgctc ttcattgact accgtcactt tgacaagcgc 2340
 aatgagacc ccatattatga gtttggccat ggcttgagct acaccacctt tggttactct 2400
 caccttcggg ttcaggccct caatagttcg agttcggcat atgtcccgc tagcgggagag 2460
 accaagcctg cgccaaccta tggtagatc ggtagtgcg ccgactacct gtatcccag 2520
 ggtctcaaaa gaattaccaa gtttatttac ccttggctca actcgaccga cctcagggat 2580
 tcttctgacg acccgaacta cggctgggag gactcggagt acattcccga aggcgctagg 2640
 gatgggtctc ctcaacccct cctgaaggct ggcggcctc ctggtggtaa ccctaccctt 2700
 tatcaggatc ttgttagggg gtcggccacc ataaccaaca ctggtaacgt cgccggttat 2760
 gaagtccctc aattgggtgag tgaccgcgat gttccttgcg ttgcaatttg gctaactcgc 2820
 ttctagtatg tttcaactggg cggaccgaac gagcctcggg tcgttctgcg caagttcgc 2880

ES 2 437 198 T3

cgaatcttcc tggctcctgg ggagcaaaag gtttggacca cgactcttaa ccgtcgtgat 2940

ctcgccaatt gggatgtgga ggctcaggac tgggtcatca caaagtaccc caagaaagtg 3000

cacgtcggca gctcctcgcg taagctgcct ctgagagcgc ctctgccccg tgtctactag 3060

<210> 2

<211> 863

5 <212> PRT

<213> *Aspergillus fumigatus*

<400> 2

ES 2 437 198 T3

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

ES 2 437 198 T3

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

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

ES 2 437 198 T3

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

ES 2 437 198 T3

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

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

<210> 3
 <211> 22
 <212> DNA
 <213> Aspergillus niger

<400> 3
 gtgccccatg atacgcctcc gg 22

<210> 4
 <211> 26
 <212> DNA
 <213> Aspergillus niger
 5
 <400> 4
 gagtcgtatt tccaaggctc ctgacc 26
 <210> 5
 <211> 24
 <212> DNA
 <213> Aspergillus niger
 10
 <400> 5
 ggaggccatg aagtggacca acgg 24
 <210> 6
 <211> 45
 <212> DNA
 <213> Aspergillus niger
 20
 <400> 6
 caccgtgaaa gccatgctct ttcttcgtg tagaagacca gacag 45
 <210> 7
 <211> 45
 <212> DNA
 <213> Aspergillus niger
 25
 <400> 7
 ctggtctct acacgaagga aagagcatgg cttcaccgt gtctg 45
 <210> 8
 <211> 44
 <212> DNA
 <213> Aspergillus niger
 35
 <400> 8
 ctatatacac aactggattt accatgggcc cgcggccgca gatc 44
 <210> 9
 <211> 44
 <212> DNA
 <213> Aspergillus niger
 40
 <400> 9
 gatctgccc cgcgggccc tggtaaatcc agttgtgtat atag 44
 <210> 10
 <211> 31
 <212> DNA
 <213> Aspergillus fumigatus
 45
 <400> 10
 actggattta ccatgagatt cggttgctc g 31
 <210> 11
 <211> 31
 <212> DNA
 <213> Aspergillus fumigatus
 60

<400> 11
 agtcacctct agtfactagt agacacgggg c 31
 5 <210> 12
 <211> 29
 <212> DNA
 <213> Trichoderma reesei
 10 <400> 12
 aacgtaatt aaggaatcgt ttgtgttt 29
 <210> 13
 <211> 29
 15 <212> DNA
 <213> Trichoderma reesei
 <400> 13
 agtactagta gctccgtggc gaaagcctg 29
 20 <210> 14
 <211> 26
 <212> DNA
 <213> Trichoderma reesei
 25 <400> 14
 actagtcgac cgaatgtagg attggt 26
 <210> 15
 <211> 19
 30 <212> DNA
 <213> Trichoderma reesei
 <400> 15
 35 tgaccatggt gcgagtc 19
 <210> 16
 <211> 26
 <212> DNA
 40 <213> Trichoderma reesei
 <400> 16
 cgatcgtctc ccatgggctc attacc 26
 45 <210> 17
 <211> 28
 <212> DNA
 <213> Trichoderma reesei
 50 <400> 17
 actagttaat taagctccgt ggcgaaag 28
 <210> 18
 <211> 30
 55 <212> DNA
 <213> Aspergillus fumigatus
 <400> 18
 60 ggactgcgca ccatgagatt cggttgctc 30

<210> 19
<211> 30
<212> DNA
<213> *Aspergillus fumigatus*

5

<400> 19
tcgccacgga gcttactagt agacacgggg 30

REIVINDICACIONES

- 5 1. Polipéptido aislado con actividad de beta-glucosidasa, seleccionado de un polipéptido que consiste en una secuencia de aminoácidos que tiene al menos un 90% de identidad con aminoácidos 20 a 863 de la SEC ID n°: 2.
2. Polipéptido según la reivindicación 1, que consiste en una secuencia de aminoácidos que tiene al menos un 95% de identidad con los aminoácidos 20 a 863 de la SEC ID n°: 2.
- 10 3. Polipéptido de cualquiera de las reivindicaciones 1 o 2, que consiste en los aminoácidos 20 a 863 de la SEC ID n°: 2.
4. El polipéptido de cualquiera de las reivindicaciones 1 a 3, que está codificado por el polinucleótido contenido en el plásmido pEJG113 que a su vez está contenido en *E. coli* NRRL B-30695.
- 15 5. Polinucleótido aislado que comprende una secuencia de nucleótidos que codifica el polipéptido de cualquiera de las reivindicaciones 1 a 4.
6. Constructo de ácidos nucleicos que comprende el polinucleótido según la reivindicación 5 operativamente enlazado a una o más secuencias de control que dirigen la producción del polipéptido en un huésped de expresión.
- 20 7. Vector de expresión recombinante que comprende el constructo de ácidos nucleicos según la reivindicación 6.
8. Célula huésped recombinante que comprende el constructo de ácidos nucleicos según la reivindicación 6.
- 25 9. Método para producir el polipéptido de cualquiera de las reivindicaciones 1 a 4 que comprende (a) el cultivo de una célula, que en su forma tipo salvaje es capaz de producir el polipéptido, bajo condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido.
- 30 10. Método para producir el polipéptido de cualquiera de las reivindicaciones 1 a 4 que comprende (a) el cultivo de una célula huésped que comprende un constructo de ácidos nucleicos que comprende una secuencia de nucleótidos que codifica el polipéptido bajo condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido.
- 35 11. Método para producir el polipéptido de cualquiera de las reivindicaciones 1 a 4, que comprende (a) el cultivo de una planta transgénica o una célula vegetal que comprende un polinucleótido que codifica un polipéptido con actividad de beta-glucosidasa de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido.
12. Método para degradar biomasa que contiene celulosa y semicelulosa, que comprende el tratado de la biomasa con una cantidad eficaz del polipéptido de cualquiera de las reivindicaciones 1 a 4 y la recuperación de la biomasa degradada.

ATGAGATTCCGGTTGGCTCGAGGTTGGCCCGCTCTGAGCCGCTTCTGTAGCCCAATGCCAGAGTTTGTGATGCTTTCCCGTCATTGTTTCCGGATAATAGTTGA 100
 M R F G W L E V A A L T A A S V A N A Q
 CAATAGTCAATGAAATAATCAGGAATTGGCTTTCTCTCCACCATTCTACCCCTTCGGCTTGGGTGATGGCCAGGAGAGTGGCAGANGCCCATCGACGC 200
 E L A F S P P F Y P S P W A D G Q G E W A D A H R R
 GCCGTCGAGATCGTTTTCTCAGATGACACTGGCCGGAGAAGTTAACTTACAACGGGTACTGGGTGGCTTGGACACTTTTGTGACAGTGGCTTTCTTC 300
 A V E I V S Q M T L A E K V N L T T G T G
 ACTGACCATCTACACAGATGGAAATGGACCGATCGTCGGTCAAAACCGGACGGCTTCCAGGTAAGCTTGCAAAATCTGCAACAACCGTGCAGGTGACTT 400
 W E M D R C V G Q T G S V P R
 GCTAAAACGGGTGGTGCAGACTTGGTATCAACTGGGTCTTTGTGGCCAGGATCCCTTTGGGTATCCGTTTCTGTGAGCTATACCCGGGAGTCTTTT 500
 L G I N W G L C G Q D S P L G I R F
 CAGTCCCTGTATTATGTGCTGATGATTGTCTCTGTATAGTAGCCTCAACTCCGCCCTTCCCTGCTGGTACTAAATGTCGCCCGACATGGGACAAGACACT 600
 S D L N S A F P A G T N V A A T W D K T L
 CGCCTACCTTCGGCAAGCCATGGGTGAGGAATTCACACGACAAGGGCGIGGACATTTGTCTGGGCCCTGCTGTGGTCTCTCGGCAAAATACCCGGAC 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
 GCGGCAGAACTCTGGAAAGCTTCTCTCCTGATCCGGTCTCACTGGTGTACTTTCCGCCGAAACTATCAAGGGTATCCAAGACGGGGTGTGATTTGCTA 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
 CTGCCAAGCATTTACATTTCTGAATGAACAGGAGCATTTCCGACAGGTTGGCGAGGCCCAAGGATATGGTTACAACATCACCGAGACGATCAGCTCCACCT 900
 T A K H Y I L N E Q E H F R Q V G E A Q G Y N I T E T I S S N V
 GGATGACAAGACCATGCACGAGTTGTACCTTTGGTGTAGTAGTTGACACTGCCAAAATGAGGACCTTGTGATTTGACTGACCTGGAAATGCAGGCCCTTTGC 1000
 D D K T M H E L Y L W P F A
 AGATGCTGTGGCGGTAAGATTTCCGTAGACTTGGACCTCGGACGGAAGAAATCGCTGAGCAACCAATCGTAGCTGGCGTTGGCGCTGTCAATGTTTCCCTA 1100
 A G V G A V M C S Y
 D A V R
 CAATCAAATCAAACAAGCTACGGTTGTCAAAAACAGTCAAACCTCAAACAAGCTCCCTCAAAGCTGAGCTGGGCTTCCAAAGCTTCGTCAATGAGTACTGG 1200
 N Q I N N S Y G C Q N S Q T L N K L L K A E L G F Q G F V M S D W
 AGCGCTCACCACAGCGGTGTGGCGCTGCCCTCGCTGGTTGGATATGTCGATCCCTGGAGACATTTCCCTTCGACGACGACTCCTCCTTCTGGGGCACGA 1300
 S A H S G V G A A L A G L D M S M P G D I S F D D G L S F W G T
 ACCTAACTGTGAGTGTCTTAACGGCACCGTTCCAGCTGGCGTGTGATGACATGGCTGTTCGTATCATGACCGGCTACTACAAGTTGGTCTGTGACCG 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
 TCTTCGTATTTCCCTTCACTTCCAGCTCCCTGGACCCGGATGAGTACGGCTGGGACATTTCTGCTGTCTCCGAGGGAGCTGGACCAAGTTGAACGACTTC 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
 GTCAAATGTGACGCGACTCACTCTCAGATCATCCGTGAGATTTGGTCCGCTAGTACTGCTCTTTGAAGAACAACGGGTGCTCTTCTTTGACCCGGCAAG 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
 AGGTTAAAGTGGGTGTTCTCGGTGAAGACCGCTGGTTCCAAACCCCGTGGGTTGTAACGGCTGCCCGACCCCGGCTGTGATAACGGCACTCTTGTCTATGGC 1700

Fig. 1A

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
 CTGGGTAGTGGTACTGCCAACTTCCCTTACCTTGTACCCCGAGCAGGCTATCCAGCGGAGAGGTTCATCAGCAACGGCGGCAATGTCTTTGCTGTGACT 1800
 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
 GATAACGGGGCTCTCAGCCAGATGGCAGATGTTGCATCTCAATCCAGGTGAGTCGGGGCTTTAGAAAAAAGAACGTTCTCTGAATGAAGTTTTTTAAACCA 1900
 D N G A L S Q M A D V A S Q S S
 TTGGAAACAGCGGTCTTTGGTGTGTTGTTCAACCGCCGACTCTGGAGAGGGTTTCATCAGTGTGACGGGCAACGAGGGTGACCGCAAAAAATCTCACTCTGTGG 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
 GAAGAACGGGAGGCGGTCATTTGACACTGTTGTGACAGCACTGCAACAACACAGANTGTGGTTATTACAGTGTGGCCCGGTCTTGATCGACGGGTGGTAT 2100
 K N G E A V I D T V V S H C N N T I V V I H S V G P V L I D R W Y
 GATAACCCCAAGTCACTGCCATCATCTGGCCGGCTTGGCCGGTCAGGAGAGTGGCAACTCCCTGGTGGTCGACGTCTCTATGGCCGGTCAACCCCGAGCG 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
 CCAAGACCCCGTTACCTGGCCAAAGACTCGGAGTCTTACGGGGCTCCCTTGTCTACCGAGCCTAAACAATGGCAATGGTGTCTCCCAGGATGATTTCAA 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 A P Q D D F N
 CGAGGGGTCTTCAATTGACTACCGTCACTTTGACAAGCGCAATGAGACCCCAATTTATGAGTTTGGCCATGGCTTGAGCTACACCACCTTTGGTTACTCT 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
 CACCTTCGGGTTCCAGGCCCTCAATAGTTCGAGTTCGGCAATGTCGCCGACTAGCCGAGACCAAGCCTGGCCCAACCTATGGTGAGATCGGTAGTGCCG 2500
 H L R V Q A L N S S A Y V P T S G E T K P A P T Y G E I G S A
 CCGACTACCTGTATCCCGAGGTTCTCAAAGAAATTAACCAAGTTTATACCTTGGCTCAACTCGACCCGACCTCGAGGATCTCTTGACGACCCCGAACTA 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
 CGGCTGGGAGGACTCGGAGTACATCCCGAAGCGCTAGGGTCTCCTCAACCCCTCCTGAAGGCTGGCGGCTCCTGGTGGTAAACCTTACCCCTT 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
 TATCAGGATCTTGTAGGGTGCGGCCACCATAACCAACTGTTAACCGTCCGGGTTATGAAGTCCCTCAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 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
 TTGCAATTTGGCTAACTCGCTTCTAGTATGTTTCACTGGGCGGACCGAACCAGCCTCGGTTCTGCGCAAGTTCGACCCGAATCTTCTGGCTCCTGG 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
 GGAGCAAAGGTTTGGACCCGACTCTTAACCGTCTGATCTCGCCAAATTTGGGATTTGGAGGCTCAGGACTGGGTTCATCACAAGTACCCCAAGAAAGTG 3000
 E Q K V W 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
 CACGTCCGCGAGCTCCCTCGGTAAGCTGCCTGTGAGAGCGCCCTCGCCCGTGTCTACTAG 3060
 H V G S S S R K L P L R A P L P R V Y .

Fig. 1B

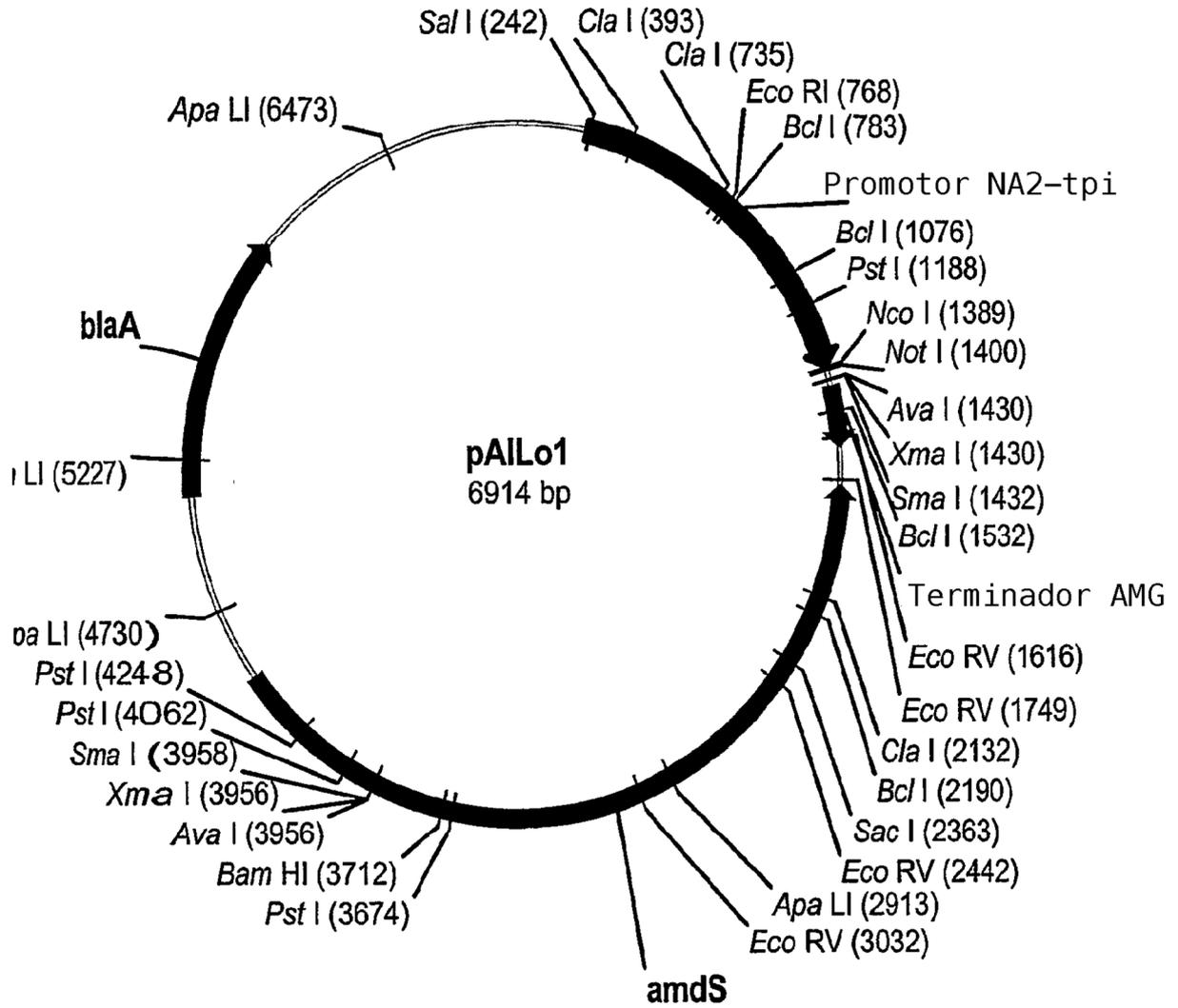


Fig. 2

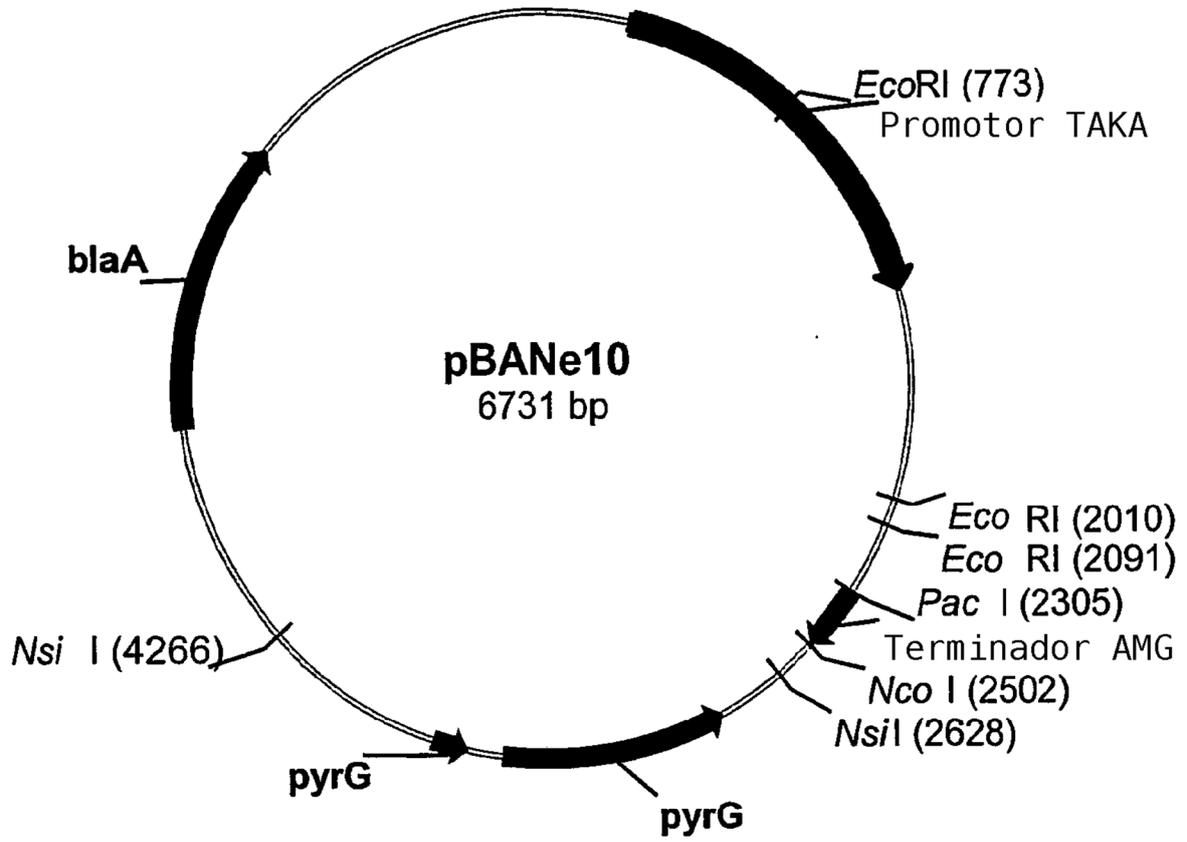


Fig. 3

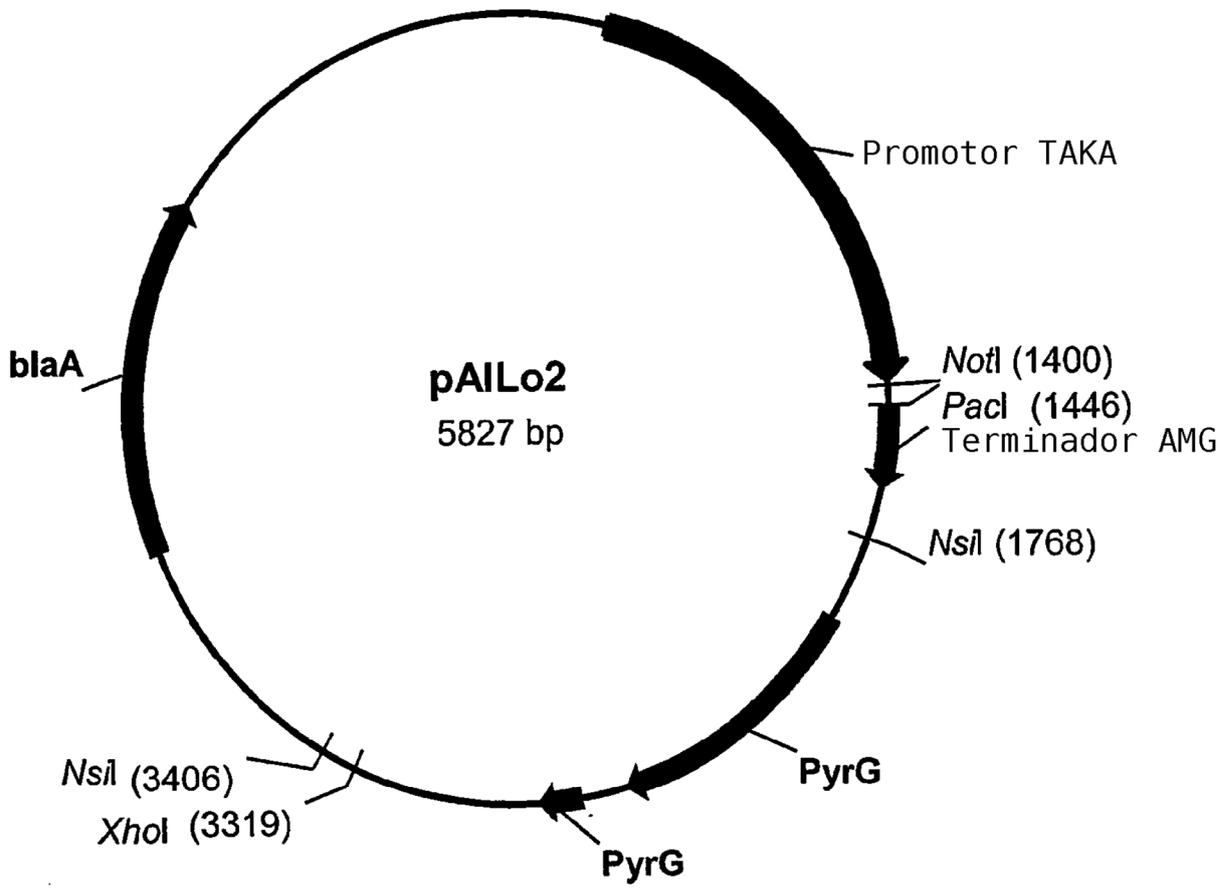


Fig. 4

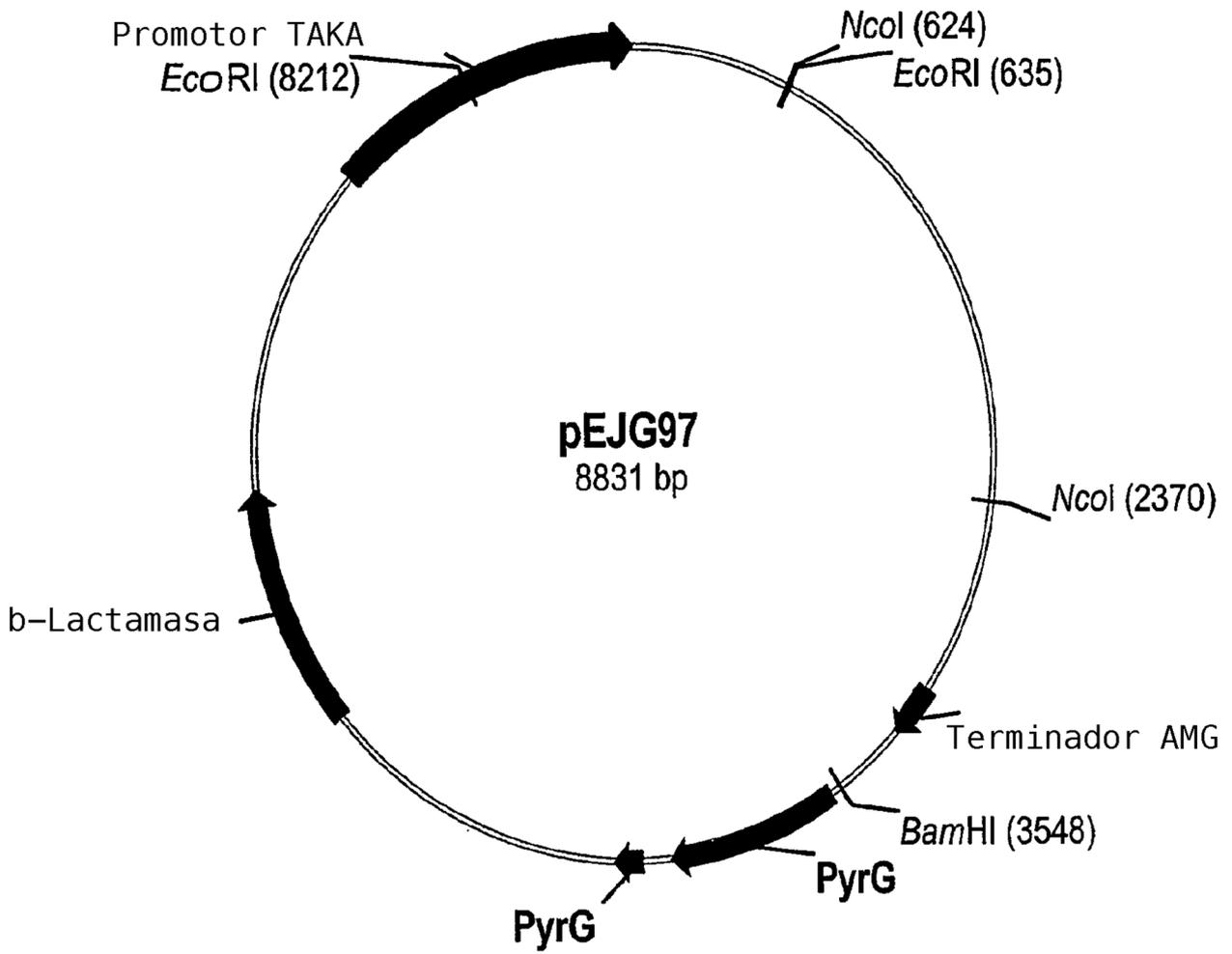


Fig. 5

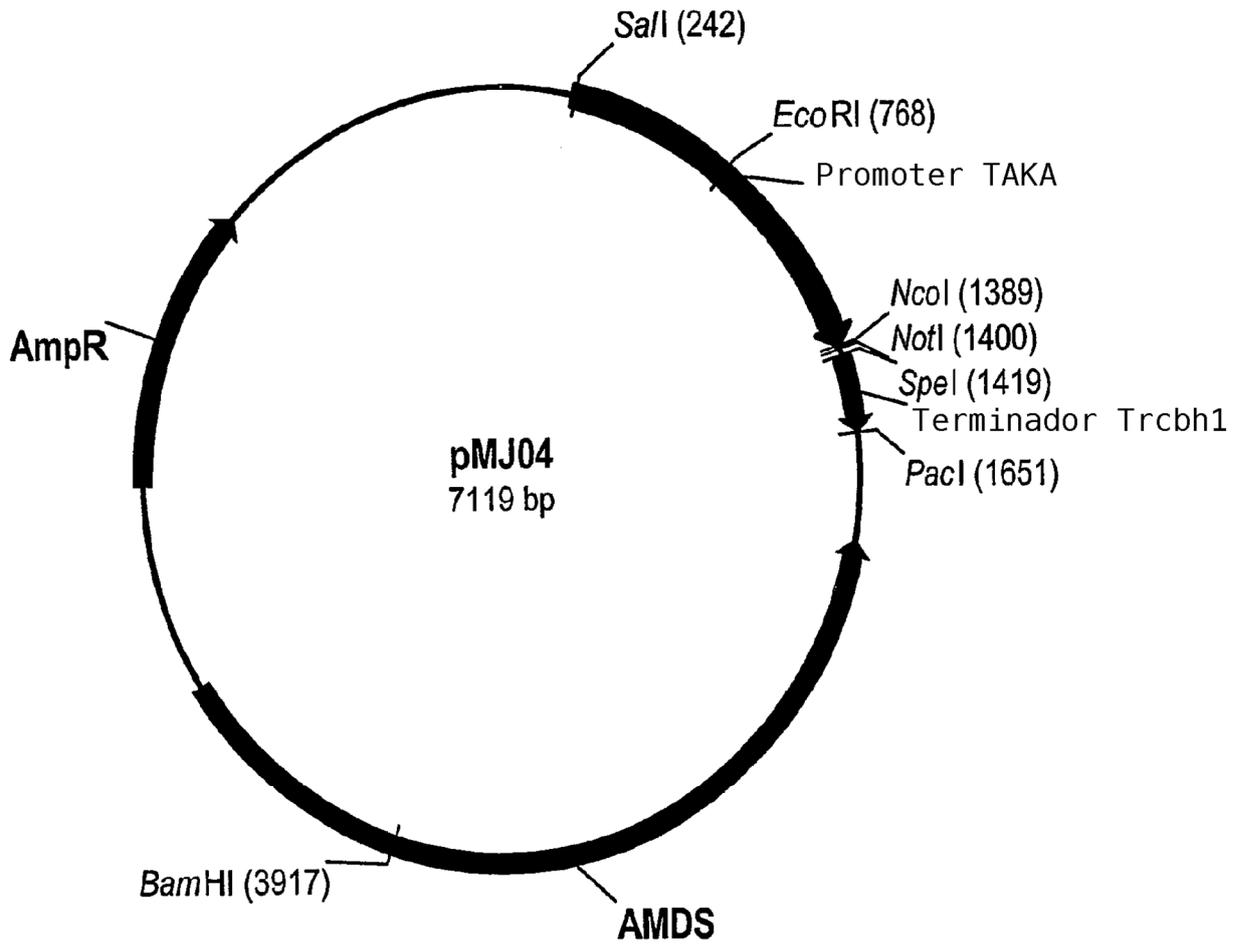


Fig. 6

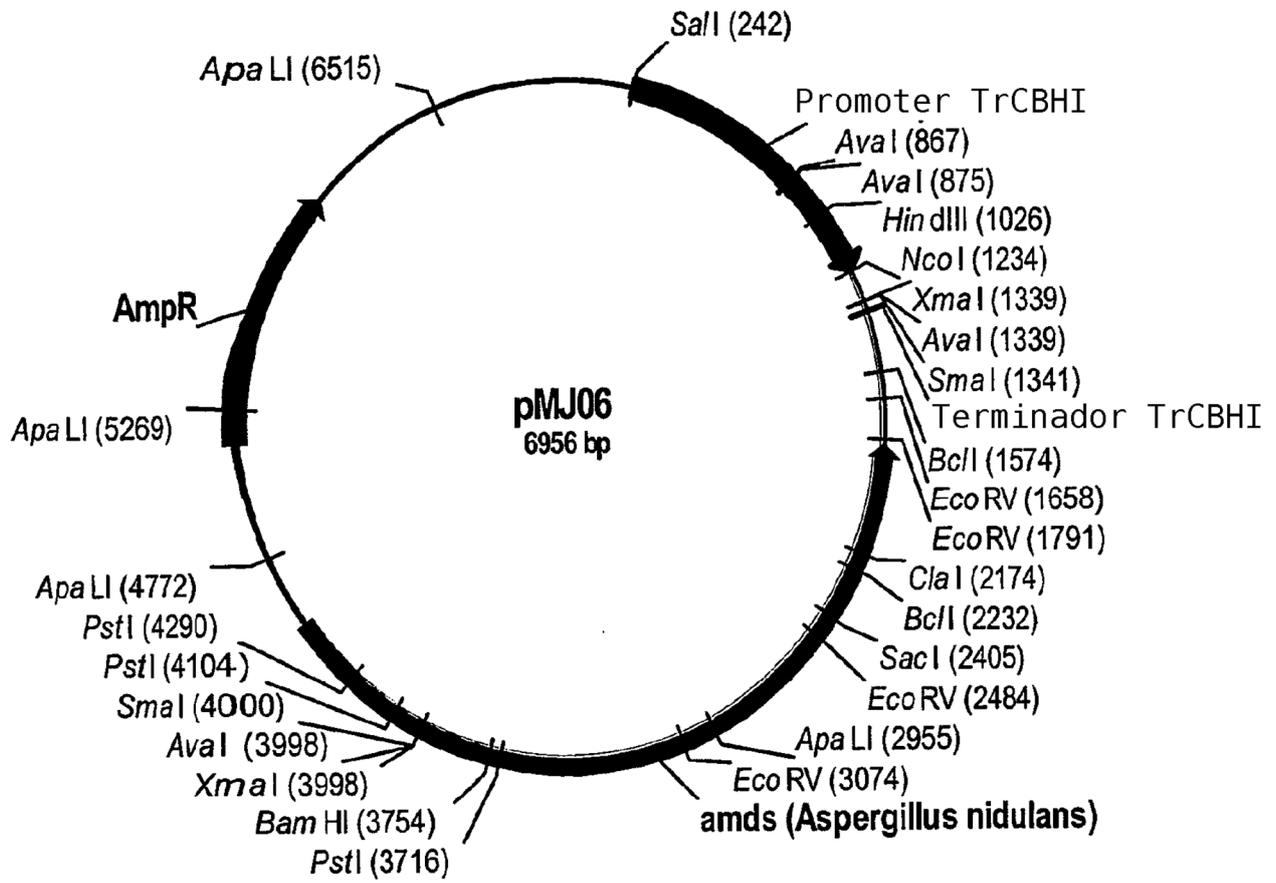


Fig. 7

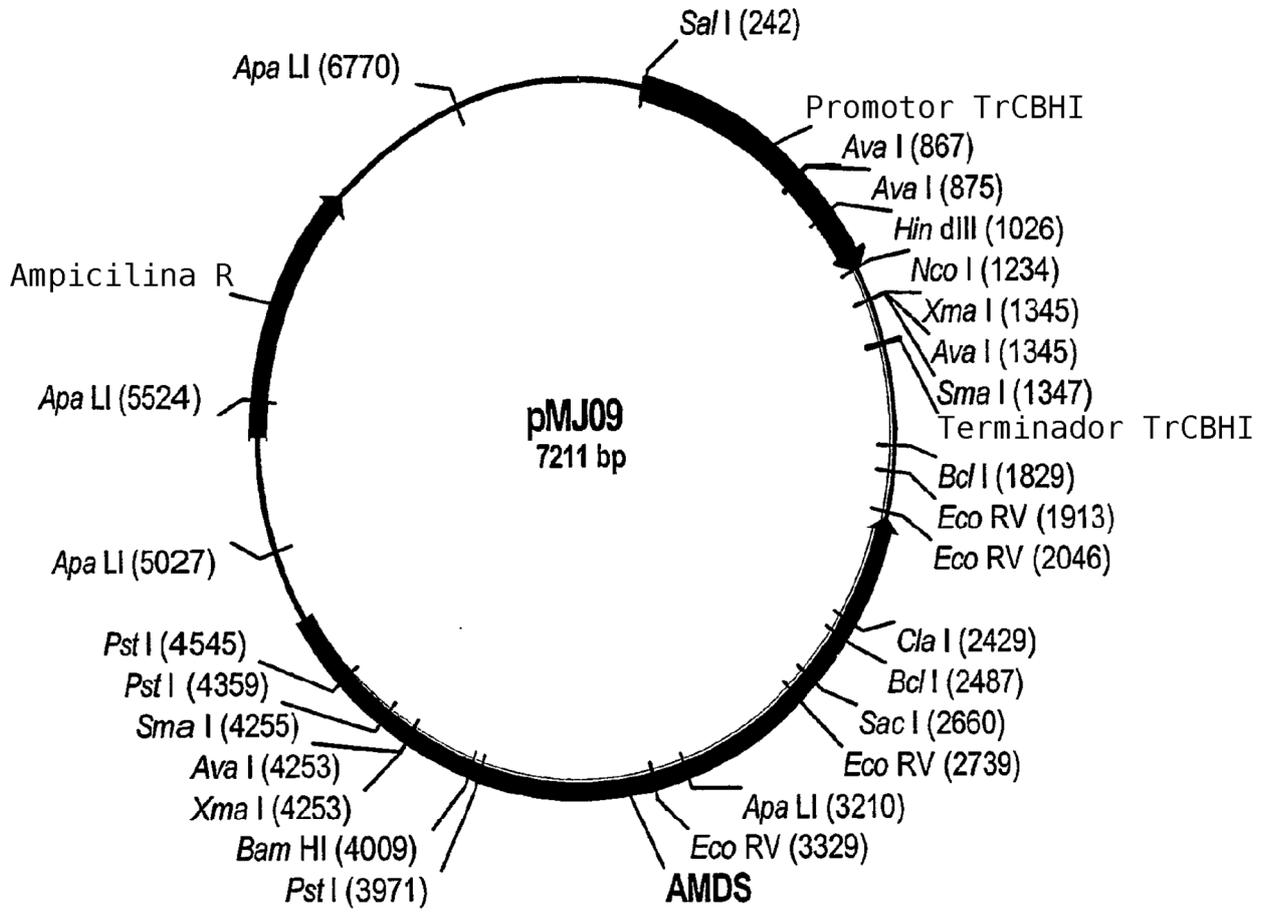


Fig. 8

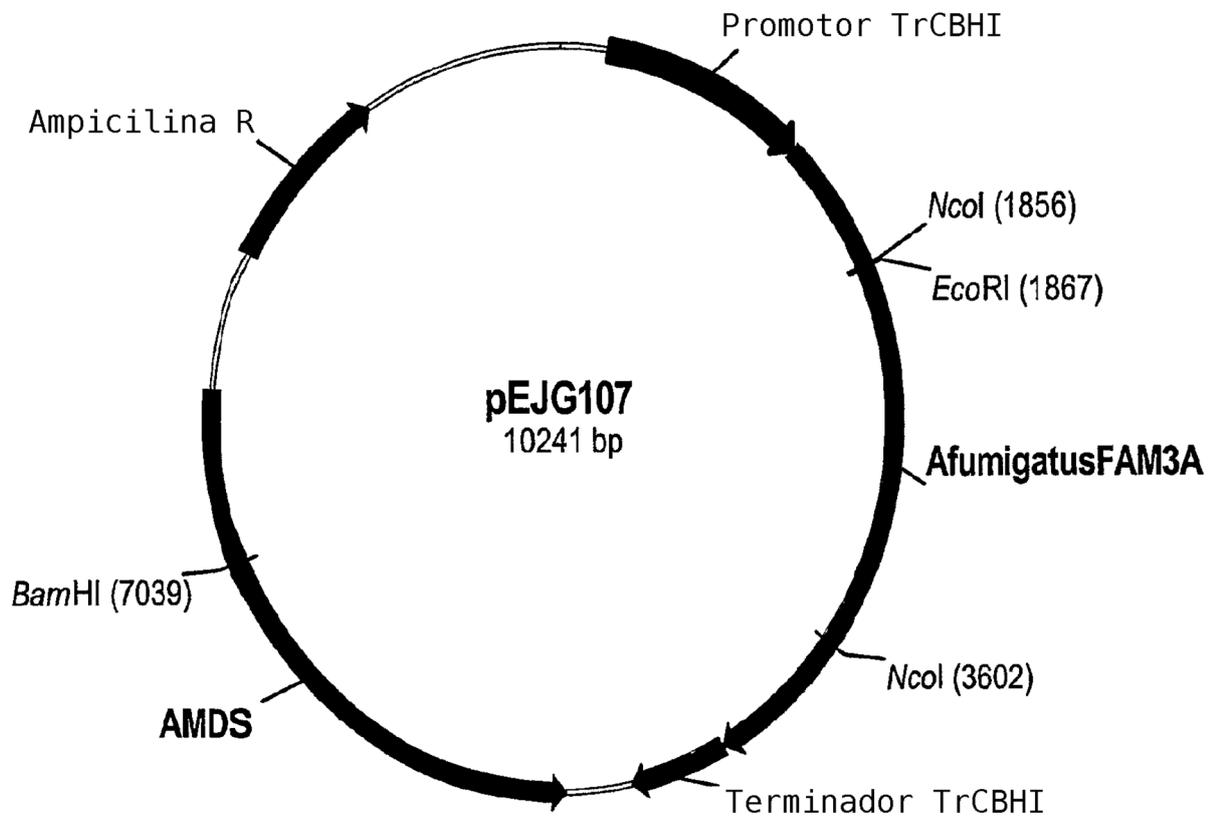


Fig. 9

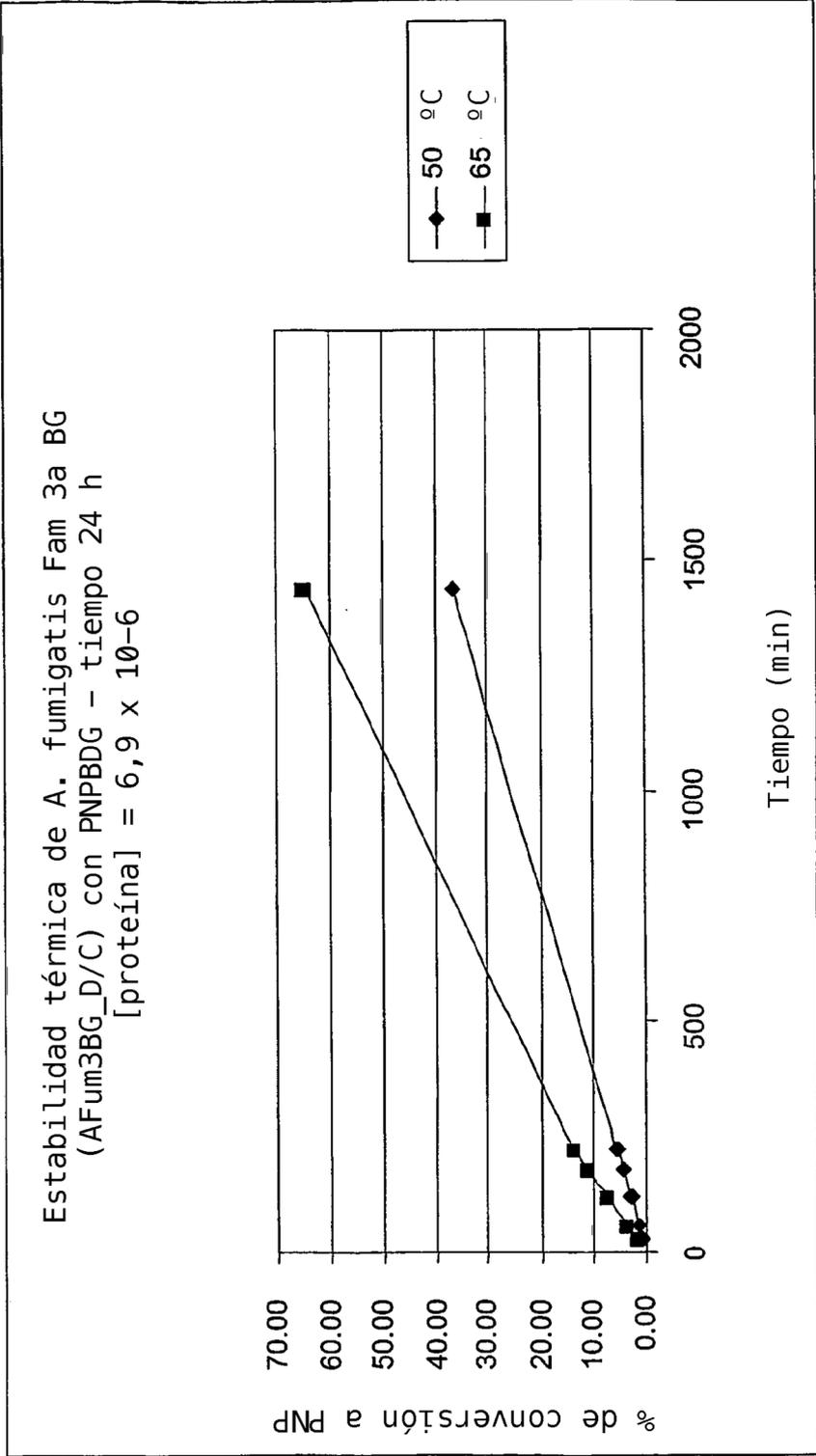


Fig. 10

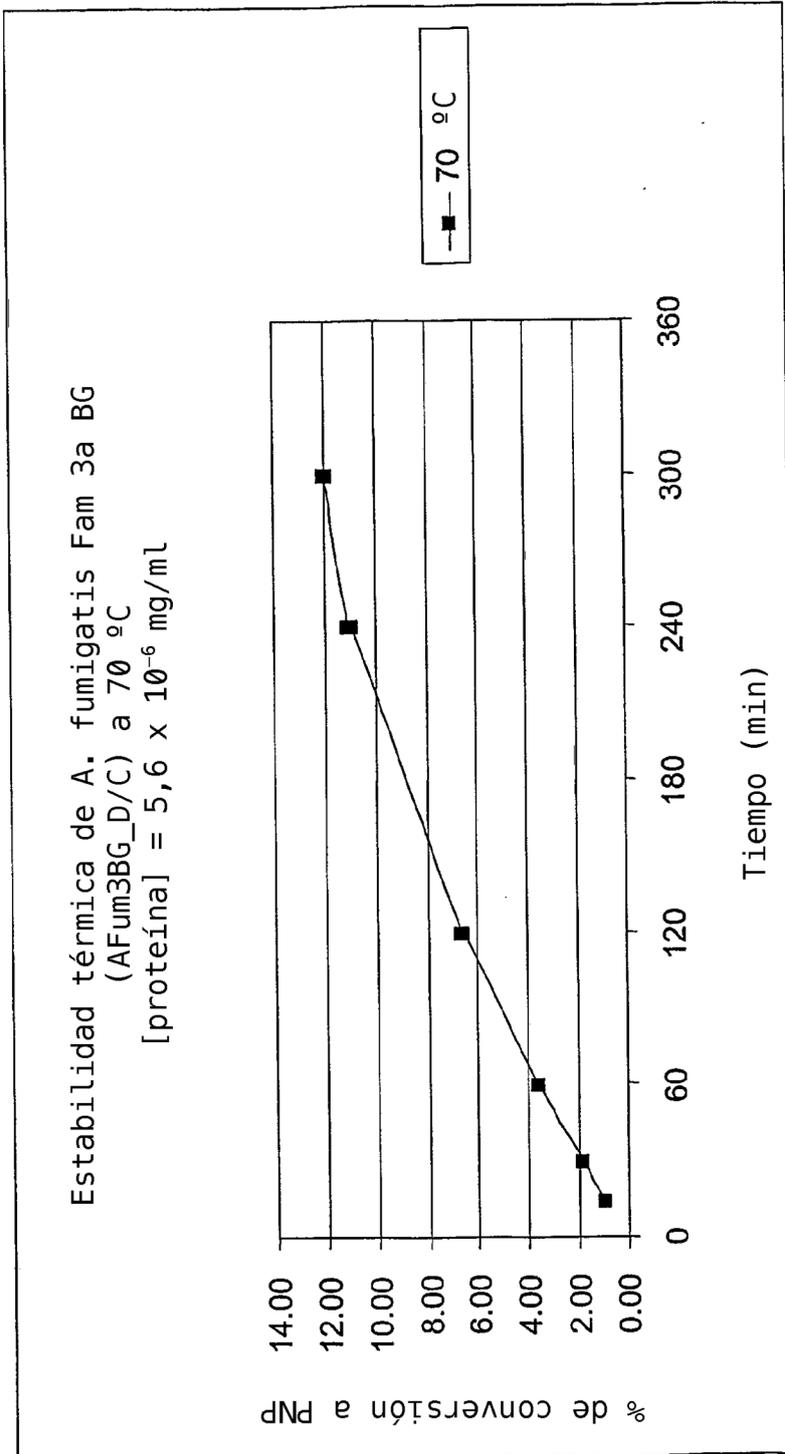


Fig. 11

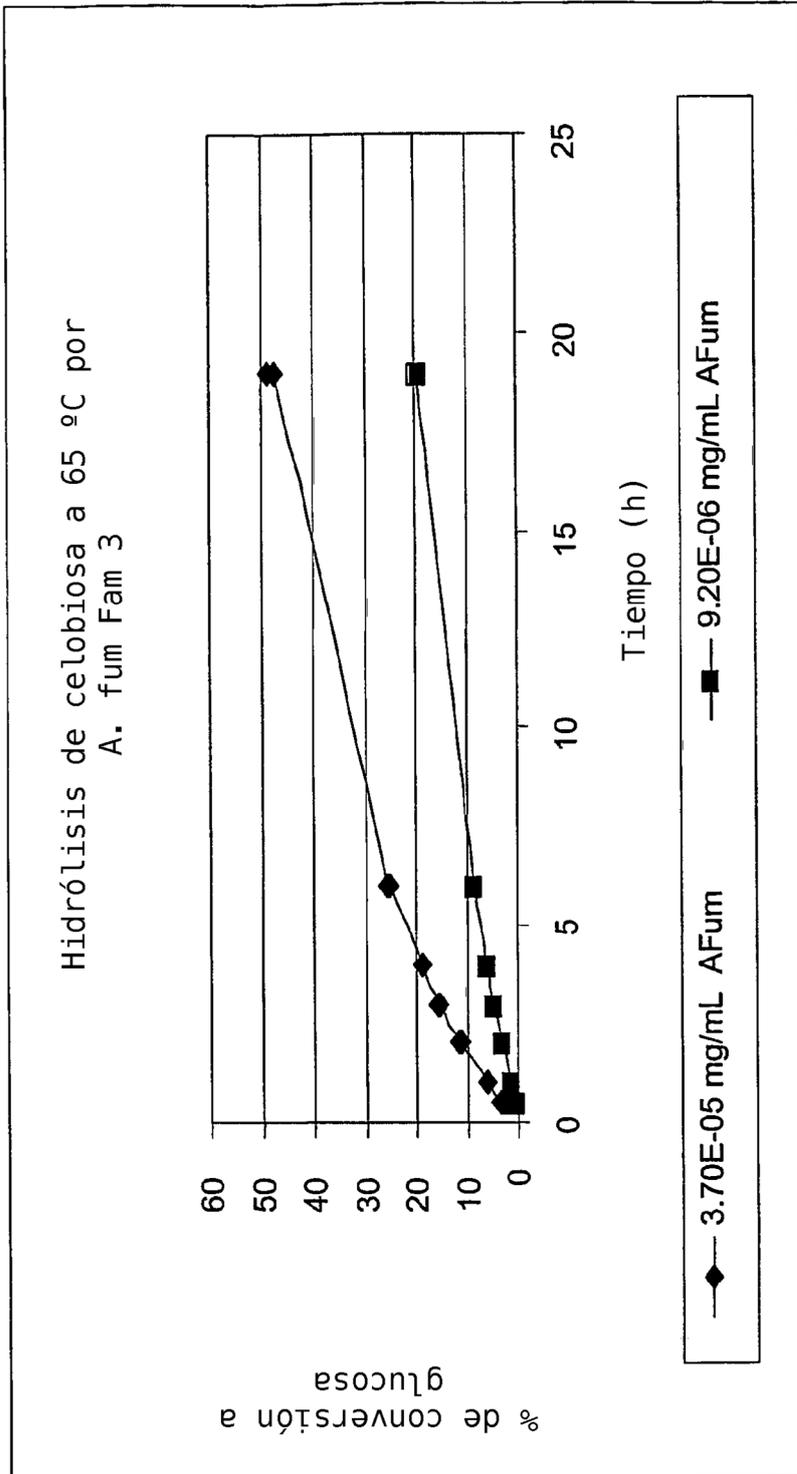


Fig. 12