

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 336 026**

51 Int. Cl.:

**C12N 9/42** (2006.01)

**C12N 9/24** (2006.01)

**C12N 15/00** (2006.01)

**C11D 3/386** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA MODIFICADA  
TRAS OPOSICIÓN

T5

96 Fecha de presentación y número de la solicitud europea: **27.04.2006 E 06751769 (8)**

97 Fecha y número de publicación de la concesión europea modificada tras oposición: **26.02.2014 EP 1877551**

54

Título: **Polipéptidos con actividad endoglucanasa y polinucleótidos que codifican los mismos**

30

Prioridad:

**27.04.2005 US 675601 P**

45

Fecha de publicación y mención en BOPI de la  
traducción de la patente modificada:  
**20.05.2014**

73

Titular/es:

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

72

Inventor/es:

**HARRIS, PAUL;  
LOPEZ DE LEON, ALFREDO;  
REY, MICHAEL;  
DING, HANSHU y  
VLASENKO, ELENA**

74

Agente/Representante:

**TOMAS GIL, Tesifonte Enrique**

ES 2 336 026 T5

## DESCRIPCIÓN

Polipéptidos con actividad endoglucanasa y polinucleótidos que codifican los mismos

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

10 [0001] La presente invención se refiere a polipéptidos aislados que tienen actividad endoglucanasa y a polinucleótidos aislados que codifican los polipéptidos. La invención también se refiere a constructos de ácido nucleico, vectores y células huéspedes que comprenden los polinucleótidos así como métodos para producir y usar los polipéptidos.

**Descripción de la técnica relacionada**

15 [0002] La celulosa es un polímero de la glucosa de azúcar simple unido covalentemente por enlaces beta-1,4. Muchos microorganismos producen enzimas que hidrolizan los betaglucanos. Estas enzimas incluyen endoglucanasas, celobiohidrolasas y beta-glucosidasas. Las endoglucanasas digieren el polímero de celulosa en lugares aleatorios, abriéndolo al ataque por parte de celobiohidrolasas. Las celobiohidrolasas liberan secuencialmente moléculas de celobiosa de los extremos del polímero de celulosa. La celobiohidrolasa I es una actividad 1,4-D-glucano  
20 celobiohidrolasa (E.C. 3, 2, 1, 91) que cataliza la hidrólisis de los enlaces 1,4-beta-D-glucosídicos en celulosa, celotetrosa o cualquier glucosa beta-1,4-enlazada que contenga polímero, liberando celobiosa de las extremidades reductoras de la cadena. La celobiohidrolasa II es una actividad 1,4-D-glucano celobiohidrolasa (E.C. 3,2,1,91) que cataliza la hidrólisis de los enlaces 1,4-beta-D-glucosídicos en celulosa, celotetrosa o cualquier glucosa beta 1,4-enlazada que contenga polímero, liberando celobiosa de las extremidades no reductoras de la cadena. La celobiosa es  
25 un dímero hidrosoluble beta-1,4-enlazado de glucosa. Las beta-glucosidasas hidrolizan la celobiosa a glucosa.

[0003] La conversión de materias primas celulósicas en etanol tiene las ventajas de la disponibilidad de grandes cantidades de materia prima, la conveniencia de evitar la quema o vertido de los materiales y la limpieza del combustible de etanol. La madera, los residuos agrícolas, los brotes herbáceos y los residuos sólidos municipales han  
30 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, la glucosa es fácilmente fermentada por levadura en etanol.

[0004] Kvesitadze et al., 1995, *Applied Biochemistry and Biotechnology* 50: 137-143, describen el aislamiento y las propiedades de una endoglucanasa termoestable de una cepa mutante termofílica de *Thielavia terrestris*. Gilbert et al., 1992, *Bioresource Technology* 39: 147-154, describen la caracterización de las enzimas presentes en el sistema de la celulosa de 255B de *Thielavia terrestris*. Breuil et al., 1986, *Biotechnology Letters* 8: 673-676, describen la producción y la localización de celulasas y beta-glucosidasas de cepas de *Thielavia terrestris* C464 y NRRL 8126.

40 [0005] Sería una ventaja en la técnica identificar las endoglucanasas nuevas que tienen propiedades mejoradas, tales como índices de hidrólisis mejorados, mejor termoestabilidad, adsorción reducida para lignina y la capacidad para hidrolizar componentes no celulósicos de biomasa, tales como hemicelulosa, además de hidrolizar celulosa. Las endoglucanasas con una amplia gama de actividades laterales en la hemicelulosa pueden ser especialmente provechosas para mejorar el rendimiento general de la hidrólisis de sustratos de biomasa ricos en hemicelulosa  
45 complejos.

[0006] Es un objeto de la presente invención proporcionar polipéptidos mejorados que tengan actividad de endoglucanasa y polinucleótidos que codifiquen los polipéptidos.

50 **Resumen de la invención**

[0007] La presente invención se refiere en un primer aspecto a polipéptidos aislados que tienen actividad de endoglucanasa y que comprende una secuencia de aminoácidos que tiene al menos el 80% de identidad con la secuencia codificante del polipéptido maduro de SEC ID n.º: 2. La presente invención también se refiere a  
55 polinucleótidos aislados que comprenden una secuencia de nucleótidos que codifica el polipéptido del primer aspecto.

[0008] En un aspecto preferido, el polipéptido maduro son los aminoácidos 18 a 336 de la SEC ID n.º: 2. En otro aspecto preferido, la secuencia codificante del polipéptido maduro son los nucleótidos 52 a 1008 de la SEC ID n.º: 1.

60 [0009] La presente invención también se refiere a constructos de ácidos nucleicos, vectores de expresión recombinantes y células huéspedes recombinantes que comprenden los polinucleótidos.

[0010] La presente invención también se refiere a métodos para producir tal polipéptido que tiene actividad de endoglucanasa que comprende: (a) cultivar una célula huésped recombinante que comprende un constructo de ácidos  
65 nucleicos que comprende un polinucleótido que codifica el polipéptido bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido.

[0011] La presente invención también se refiere a métodos de uso de los polipéptidos que tienen actividad de endoglucanasa en la degradación de biomasa con celulosa y hemicelulosa.

## 5 Breve descripción de las figuras

[0012]

Las Figuras 1A y 1B muestran la secuencia de ADNc y la secuencia de aminoácidos deducida de una endoglucanasa (CEL7F) de *Thielavia terrestris* NRRL 8126 (SEC ID n.º: 1 y 2, respectivamente).

10 La Figura 2 muestra un mapa de restricción de pTter7F.

La Figura 3 muestra un mapa de restricción de pAIlO1.

La Figura 4 muestra un mapa de restricción de pBAnE10.

La Figura 5 muestra un mapa de restricción de pAIlO2.

La Figura 6 muestra un mapa de restricción de pAIlO22.

15 La Figura 7 muestra la conversión relativa de betaglucano (1% p/v) después de 2 horas de hidrólisis a pH 5,5 y 60°C.

La Figura 8 muestra la conversión relativa de betaglucano (1% p/v) después de 24 horas de hidrólisis a pH 5,5 y 60°C.

## Definiciones

20 [0013] **Actividad endoglucanasa:** El término "actividad endoglucanasa" es definido en la presente como una endo-1,4-beta-D- glucano 4-glucanohidrolasa (E.C. N.º. 3,2,1,4) que cataliza la endohidrólisis de enlaces 1,4-beta-D-glicosídicos en celulosa, derivados de celulosa (tales como carboximetilcelulosa e hidroxietil celulosa), liquenina, enlaces beta-1,4 en beta-1,3 glucanos mezclados tales como beta-D-glucanos de cereales o xiloglucanos y otra materia vegetal que  
25 contiene componentes celulósicos. Para fines de la presente invención, la actividad endoglucanasa es determinada utilizando hidrólisis de carboximetilcelulosa (CMC) según el procedimiento de Ghose, 1987, *Pure and Appl. Chem.* 59: 257-268. Una unidad de actividad endoglucanasa es definida como 1,0 µmol de azúcares reducidos producidos por minuto a 50°C, pH 4,8.

[0014] En un aspecto preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además  
30 tienen actividad enzimática hacia uno o más sustratos seleccionados del grupo que consiste en xilano, xiloglucano, arabinoxilano, galactano, galactomanano, dextrano y quitina. La actividad de los polipéptidos que tienen actividad endoglucanasa en estos sustratos polisacáridos es determinada como la cantidad relativa de colorante liberado desde distintos sustratos teñidos con AZCL después de incubar los sustratos (5 g por litro) con un polipéptido que tiene actividad endoglucanasa de la presente invención (1 mg de proteína por g de sustrato) durante 1 y 92 horas sin agitar a  
35 pH 5,0 (50 mM de acetato sódico) y 50°C. La liberación de colorante fue determinada midiendo la absorbencia a 590 nm.

[0015] En un aspecto más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia el xilano. En otro aspecto más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia el xiloglucano. En otro aspecto  
40 más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia el arabinoxilano. En otro aspecto más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia el galactano. En otro aspecto más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia el galactomanano. En otro aspecto más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia el dextrano. En otro aspecto más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia la quitina. En otro aspecto más preferido, los polipéptidos de la presente invención que tienen actividad endoglucanasa además tienen actividad enzimática hacia el xilano, el xiloglucano, el arabinoxilano, el galactano, el galactomanano, el dextrano y la quitina.

50 [0016] Los polipéptidos de la presente invención tienen como mínimo 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%, de la forma más preferible al menos el 95% e incluso de la forma más preferible al menos 100% de la actividad endoglucanasa del polipéptido que consiste en la secuencia de aminoácidos mostrada como los aminoácidos 18 a 336 de la SEC ID n.º: 2.

[0017] **Familia 7 de glicósido hidrolasa o familia GH7:** El término "Familia 7 de glucósido hidrolasa" o "Familia GH7" o "CEL7F" es definida en la presente como un polipéptido que cae dentro de la familia 7 de glucósido hidrolasa según  
60 Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acid sequence similarities, *Biochem. J.* 280: 309-316, y Henrissat B., y Bairoch A., 1996, Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J.* 316: 695-696.

[0018] **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  
65 más preferiblemente al menos un 80% puro, más preferiblemente al menos un 90% puro, e incluso más preferiblemente al menos un 95% puro, según lo determina la SDS-PAGE.

[0019] **Polipéptido sustancialmente puro:** El término "polipéptido sustancialmente puro" denota en la presente una preparación de polipéptido 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%, 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 de polipéptido con el cual está originalmente asociado. Es preferible, en consecuencia, que el polipéptido sustancialmente 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%, 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.

[0020] Los polipéptidos de la presente invención están preferiblemente en una forma sustancialmente pura. En particular, se prefiere que los polipéptidos estén en "forma esencialmente pura", es decir, que la preparación de polipéptido esté esencialmente libre de otra materia polipeptídica con la cual esté originalmente asociado. Este se puede conseguir, por ejemplo, preparando el polipéptido mediante métodos recombinantes conocidos o por métodos de purificación tradicionales.

[0021] En la presente, el término "polipéptido sustancialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada".

[0022] **Polipéptido maduro:** el término "polipéptido maduro" es definido en la presente como un polipéptido que tiene actividad endoglucanasa que está en su forma final después de la traducción y cualquier modificación postraduccional, tal como tratamiento del N-terminal, truncamiento del C-terminal, glicosilación, etc.

[0023] **Identidad:** la relación entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos es descrita por el parámetro "identidad".

[0024] Para objetivos de la presente invención, el grado de identidad entre dos secuencias de aminoácidos es determinado por el método Clustal (Higgins, 1989, CABIOS 5: 151-153) usando el software LASERGENE™ MEGALIGN™ DNASTAR, Inc., Madison, WI) con una tabla de identidad y los siguientes parámetros de alineación múltiple: Penalización por gaps de 10 y penalización por longitud del gaps de 10. Los parámetros de alineación por parejas son Ktuple=1, penalización por gaps=3, ventanas=5 y diagonales=5.

[0025] Para objetivos de la presente invención, el grado de identidad entre dos secuencias de nucleótidos se determina por el método de Wilbur-Lipman (Wilbur and 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 por gaps de 10 y penalización por longitud de gaps de 10. Los parámetros de alineación por parejas son Ktuple=3, penalización por gaps=3 y ventanas=20.

[0026] **Fragmento polipeptídico:** El término "fragmento polipeptídico" es definido en la presente como un polipéptido que tiene uno o más aminoácidos delecionados del extremo amino y/o carboxilo de la SEC ID n.º: 2 o una secuencia homóloga del mismo, caracterizado por el hecho de que el fragmento tiene actividad endoglucanasa. Preferiblemente, un fragmento contiene al menos 270 residuos de aminoácidos, más preferiblemente al menos 285 residuos de aminoácidos y más preferiblemente al menos 300 residuos de aminoácidos, p. ej., aminoácidos 18 a 336 de la SEC ID n.º: 2.

[0027] **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.º: 1 o una secuencia homóloga de la misma, donde la subsecuencia codifica un fragmento de polipéptido que tiene actividad endoglucanasa. Preferiblemente, una subsecuencia contiene al menos 810 nucleótidos, más preferiblemente al menos 855 nucleótidos y más preferiblemente al menos 900 nucleótidos.

[0028] **Variante alélica:** El término "variante alélica" denota en la presente cualquiera de dos o más formas alternativas de un gen que ocupa el mismo locus cromosómico. La variación alélica surge naturalmente a través de la mutación y puede suponer 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.

[0029] **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, de la forma más preferible al menos un 90% puro e incluso de la forma más preferible al menos un 95% puro, según lo determina la electroforesis de agarosa.

[0030] **Polinucleótido substancialmente puro:** El término "polinucleótido substancialmente puro" según se utiliza en este caso se refiere a una preparación de polinucleótido libre de otros nucleótidos extraños o indeseados y en una forma adecuada para el uso dentro de los sistemas de producción de proteínas creadas genéticamente. De este modo, un polinucleótido substancialmente puro 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%, 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 otra materia polinucleótida con la cual está originalmente asociado. Un polinucleótido substancialmente puro puede, no obstante, 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, de la forma más preferible al menos un 99% e incluso de la forma más preferible 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 polinucleótido esté esencialmente libre de otra materia polinucleótida con la cual está originalmente asociado. En la presente, el término "polinucleótido substancialmente puro" es sinónimo de los términos "polinucleótido aislado" y "polinucleótido en forma aislada". Los polinucleótidos pueden ser de origen genómico, ADNc, ARN, semisintético, sintético, o cualquier combinación de los mismos

[0031] **Secuencia codificante del polipéptido maduro:** El término "secuencia codificante del polipéptido maduro" se define en la presente como una secuencia de nucleótidos que codifica un polipéptido maduro que tiene actividad endoglucanasa.

[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 maduro dividida 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 ARNm 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 intrón.

[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 que es modificada para contener segmentos de ácidos nucleicos que de otra manera no existiría en la naturaleza. El término constructo de ácidos nucleicos es sinónimo del término "cassette de expresión" cuando el constructo de ácidos nucleicos contiene las secuencias de control requeridas para la expresión de una secuencia codificante de la presente invención.

[0034] **Secuencia de control:** El término "secuencias de control" es definido en la presente para incluir 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 extranjera a la secuencia de nucleótidos que codifica el polipéptido o nativa o extranjera entre sí. Las secuencias de control de este tipo incluyen, pero sin limitarse a ello, la secuencia de poliadenilación líder, secuencia de propéptido, promotor, secuencia de péptido señal y 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 codificante 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 de polinucleótido de 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", significa una secuencia de nucleótidos que directamente especifica la secuencia de aminoácidos de su producto de proteína. Los bordes de la secuencia codificante son determinados generalmente por un marco de lectura abierto, que normalmente se inicia con el codón de iniciación ATG o codones de iniciación alternativos tales como GTG y TTG y extremidades con un codón de detención tal como TAA, TAG y TGA. La secuencia codificante puede ser una secuencia de ADN, de ADNc, o de nucleótidos recombinantes.

[0037] **Expresión:** El término "expresión" incluye cualquier fase implicada en la producción del polipéptido incluyendo, pero sin limitarse a ello, 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.

5 [0039] **Célula huésped:** El término "célula huésped", según se utiliza en este caso, incluye cualquier tipo de célula susceptible a transformación, transfección, transducción y similares con un constructo de ácidos nucleicos o vector de expresión que comprenda un polinucleótido de la presente invención.

10 [0040] **Modificación:** El término "modificación" significa en la presente cualquier modificación química del polipéptido que consiste en el polipéptido maduro de la SEC ID n.º: 2 o una secuencia homóloga de la misma al igual que la manipulación genética del ADN que codifica ese polipéptido. La modificación puede ser sustituciones, deleciones y/o inserciones de uno o más aminoácidos al igual que sustituciones de una o más cadenas laterales del aminoácido.

15 [0041] **Variante artificial:** Cuando se usa en la presente, el término "variante artificial" significa un polipéptido que tiene actividad de endoglucanasa producida por un organismo que expresa una secuencia de nucleótidos modificada de la SEC ID n.º: 1 o una secuencia homóloga de la misma, o la región codificante madura de la misma. La secuencia de nucleótidos modificada se obtiene a través de intervención humana por modificación de la secuencia de nucleótidos descrita en la SEC ID n.º: 1 o una secuencia homóloga de la misma, o la región codificante madura de la misma.

## 20 Descripción detallada de la invención

### Polipéptidos que tienen actividad de endoglucanasa

25 [0042] En un primer aspecto, la presente invención se refiere a polipéptidos aislados que comprenden una secuencia de aminoácidos con un grado de identidad al polipéptido maduro de la SEC ID n.º: 2 de al menos el 80%, preferiblemente al menos el 85%, más preferiblemente al menos el 90%, incluso más preferiblemente al menos el 95% e incluso de la forma más preferible al menos el 97%, 98%, o 99%, que tiene actividad endoglucanasa (de ahora 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 del polipéptido maduro de la SEC ID n.º: 2.

35 [0043] Un polipéptido de la presente invención preferiblemente comprende la secuencia de aminoácidos de la SEC ID n.º: 2 o una variante alélica de la misma; o un fragmento de la misma que tenga actividad endoglucanasa. En un aspecto preferido, un polipéptido comprende la secuencia de aminoácidos de la SEC ID n.º: 2. En otro aspecto preferido, un polipéptido comprende el polipéptido maduro de la SEC ID n.º: 2. En otro aspecto preferido, un polipéptido comprende los aminoácidos 18 a 336 de la SEC ID n.º: 2, o una variante alélica de la misma; o un fragmento de la misma que tiene actividad endoglucanasa. En otro aspecto preferido, un polipéptido comprende los aminoácidos 18 a 336 de la SEC ID n.º: 2. En otro aspecto preferido, un polipéptido consiste en la secuencia de aminoácidos de la SEC ID n.º: 2 o una variante alélica de la misma; o un fragmento de la misma que tenga actividad endoglucanasa. En otro aspecto preferido, un polipéptido consiste en la secuencia de aminoácidos de la SEC ID n.º: 2. En otro aspecto preferido, un polipéptido consiste en el polipéptido maduro de la SEC ID n.º: 2. En otro aspecto preferido, un polipéptido consiste en los aminoácidos 18 a 336 de la SEC ID n.º: 2 o una variante alélica de la misma; o un fragmento de la misma que tenga actividad endoglucanasa. En otro aspecto preferido, un polipéptido consiste en los aminoácidos 18 a 336 de la SEC ID n.º: 2.

45 [0044] Una subsecuencia de la SEC ID n.º: 1 contiene al menos 100 nucleótidos contiguos o preferiblemente al menos 200 nucleótidos contiguos. Además, la subsecuencia puede codificar un fragmento de polipéptido que tiene actividad endoglucanasa. En un aspecto preferido, la secuencia que codifica el polipéptido maduro es los nucleótidos 52 a 1008 de la SEC ID n.º: 1.

50 [0045] La secuencia de nucleótidos de la SEC ID n.º: 1 o una subsecuencia de la misma, al igual que la secuencia de aminoácidos de la SEC ID n.º: 2 o un fragmento de la misma, puede ser usado para diseñar una sonda de ácido nucleico para identificar y clonar ADN que codifican polipéptidos con actividad endoglucanasa de cepas de diferentes géneros o especies según métodos bien conocidos en la técnica. En particular, las sondas de este tipo pueden ser usadas para la hibridación con el ADN genómico o ADNc del género o especies de interés, siguiendo procedimientos de transferencia estándar de Southern, para identificar y aislar el gen correspondiente allí. Las sondas de este tipo pueden ser considerablemente más cortas que la secuencia entera, pero deben ser al menos 14, preferiblemente al menos 25, más preferiblemente al menos 35, y más preferiblemente al menos 70 nucleótidos en longitud. No obstante, se prefiere que la sonda de ácido nucleico sea de al menos 100 nucleótidos en longitud. Por ejemplo, la sonda de ácido nucleico puede ser de al menos 200 nucleótidos, preferiblemente al menos 300 nucleótidos, más preferiblemente al menos 400 nucleótidos, o más preferiblemente al menos 500 nucleótidos en longitud. Incluso se pueden utilizar sondas más largas, p. ej., sondas de ácido nucleico que sean de al menos 600 nucleótidos, preferiblemente al menos 700 nucleótidos, más preferiblemente al menos 800 nucleótidos, o más preferiblemente al menos 900 nucleótidos en longitud. Se pueden utilizar tanto sondas de ADN como de ARN. Las sondas normalmente están marcadas para detectar el gen

correspondiente (por ejemplo, con  $^{32}\text{P}$ ,  $^3\text{H}$ ,  $^{35}\text{S}$ , biotina o avidina). Las sondas de este tipo están incluidas en la presente invención.

[0046] Un ADN genómico o biblioteca de ADNc obtenido a partir de estos otros organismos puede, en consecuencia, ser seleccionado para ADN que se hibrida con las sondas anteriormente descritas y que codifica un polipéptido que tiene actividad endoglucanasa. 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 bibliotecas 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.º: 1 o una subsecuencia de la misma, el material portador se usa en una transferencia de Southern.

[0047] La hibridación indica que la secuencia de nucleótidos hibrida a una sonda de ácido nucleico marcada correspondiente a la secuencia de nucleótidos mostrada en la SEC ID n.º: 1, la secuencia de ADN genómico que comprende la SEC ID n.º: 1, su cadena complementaria, o una subsecuencia de la misma, condiciones de astringencia muy baja a muy alta. Las moléculas a las cuales la sonda de ácido nucleico se hibrida bajo estas condiciones pueden ser detectadas usando, por ejemplo, película radiográfica.

[0048] En un aspecto preferido, la sonda de ácidos nucleicos es la secuencia codificante de polipéptido maduro de la SEC ID n.º: 1. En otro aspecto preferido, la sonda de ácidos nucleicos es de los nucleótidos 52 a 1008 de la SEC ID n.º: 1. En otro aspecto preferido, la sonda de ácidos nucleicos es una secuencia de polinucleótidos que codifica el polipéptido de la SEC ID n.º: 2, o una subsecuencia de la misma. En otro aspecto preferido, la sonda de ácidos nucleicos es la SEC ID n.º: 1. En otro aspecto preferido, la sonda de ácidos nucleicos es la secuencia que codifica el polipéptido maduro de la SEC ID n.º: 1. En otro aspecto preferido, la sonda de ácidos nucleicos es la secuencia de polinucleótidos contenida en el plásmido pTter7F que está contenido en *E. coli* NRRL B-30837, caracterizado por el hecho de que la secuencia de polinucleótidos de la misma codifica un polipéptido que tiene actividad endoglucanasa. En otro aspecto preferido, la sonda de ácidos nucleicos es la secuencia que codifica el polipéptido maduro contenida en plásmido pTter7F que está contenido en *E. coli* NRRL B-30837.

[0049] Para sondas largas de al menos 100 nucleótidos de longitud, las condiciones de astringencia media-alta a muy alta son definidas como prehibridación e hibridación a 42°C en 5X SSPE, 0,3% de SDS, 200 µg/ml de ADN de esperma de salmón cortado y desnaturalizado y 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 procedimientos de transferencia de Southern estándar durante 12 a 24 horas óptimamente.

[0050] Para sondas largas de al menos 100 nucleótidos de longitud, el material portador finalmente es lavado tres veces cada una durante 15 minutos usando 2X SSC, 0,2% de SDS 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).

[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 post-hibridación a 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 de sodio fosfato monobásico, 0,1 mM de ATP y 0,2 mg de ARN de levadura por ml siguiendo procedimientos de transferencia de Southern estándar durante 12 a 24 horas óptimamente.

[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 SSC más 0,1% SDS durante 15 minutos y dos veces cada una durante 15 minutos usando 6X SSC a 5°C a 10°C por debajo de la  $T_m$  calculada.

[0053] En un tercer aspecto, la presente invención se refiere a variantes artificiales que comprenden una sustitución, delección y/o inserción conservadora de uno o más aminoácidos de la SEC ID n.º: 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 a la actividad de la proteína; delecciones pequeñas, normalmente de uno a aproximadamente 30 aminoácidos; pequeñas extensiones amino-terminales o carboxi-terminales, tal como un residuo de metionina aminoterminal; un pequeño péptido de enlace de hasta 20-25 residuos aproximadamente; o una pequeña extensión que facilita la purificación cambiando la carga neta u otra función, tal como un tracto de polihistidina, un epítipo antigénico o un dominio de unión.

[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 pequeños aminoácidos (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,

1979, In, The Proteins, Academic Press, New York. Los cambios de aparición más frecuente 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.

5 [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 son  
10 codificados 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 de la de los aminoácidos estándar. Los aminoácidos no naturales pueden ser sintetizados químicamente y preferiblemente, están disponibles comercialmente, e incluyen ácido pipecólico, ácido carboxílico de tiazolidina, dehidroprolina, 3 y 4-metilprolina, y 3,3-dimetilprolina.

15 [0056] De forma alternativa, los cambios de los aminoácidos son de una naturaleza tal que las propiedades físico-químicas de los polipéptidos son alteradas. Por ejemplo, los cambios de los aminoácidos pueden mejorar la termoestabilidad del polipéptido, alterar la especificidad de sustrato, cambiar el pH óptimo y similares.

[0057] Los aminoácidos esenciales en el polipéptido progenitor pueden ser identificados según procedimientos conocidos en la técnica, tales como mutagénesis dirigida o mutagénesis por barrido de alanina (Cunningham y Wells, 20 1989, Science 244: 1081-1085). En ésta 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 endoglucanasa) para identificar los residuos de aminoácidos que son críticos para la actividad de la molécula. Ver también, Hilton et al., 1996, J. Biol. Chem. 271: 4699-4708. El sitio activo de la enzima u otra interacción biológica también pueden ser determinados por análisis físico de la estructura, como lo determinan las técnicas como resonancia  
25 magnética nuclear, cristalografía, difracción electrónica, o marcaje por fotoafinidad, conjuntamente con la mutación de aminoácidos de sitio de contacto putativo. Ver, 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 de análisis de identidades con polipéptidos que están relacionados con un polipéptido según la invención.

30 [0058] Las sustituciones de aminoácidos simples o múltiples pueden ser realizadas y evaluadas utilizando métodos de mutagénesis, recombinación, y/o redistribución, seguidos de un procedimiento de selección pertinente, tales como los descritos por Reidhaar-Olson y Sauer, 1988 Science 241: 53-57 ; Bowie and Sauer, 1989, Proc. Natl. Acad. Sci. USA 86: 2152-2156; WO 95/17413; o WO 95/22625. Otros métodos que pueden ser usados incluyen PCR propensa al error, exposición en el fago (p. ej., Lowman et al., 1991, Biochem. 30: 10832-10837; Patente U.S. n.º. 5 223 409; WO 35 92/06204) y mutagénesis dirigida (Derbyshire et al., 1986, Gene 46: 145 ; Ner et al., 1988, DNA 7: 127).

[0059] Los métodos de mutagénesis/redistribución pueden ser combinados con métodos de selección automatizados de alto rendimiento para detectar actividad de polipéptidos mutagenizados clonados expresados por células huéspedes (Ness et al., 1999, Nature Biotechnology 17: 893-896). Las moléculas de ADN mutagenizadas que codifican los  
40 polipéptidos activos pueden ser recuperadas de las células huéspedes y ser rápidamente ordenadas usando métodos estándares 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.

45 [0060] La cantidad total de sustituciones de aminoácidos, delecciones y/o inserciones del polipéptido maduro de la SEC ID n.º: 2, tales como aminoácidos 18 a 336 de la SEC ID n.º: 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, de la forma más preferible 2 e incluso de la forma más preferible 1.

## 50 Fuentes de polipéptidos que tienen actividad endoglucanasa

[0061] Un polipéptido de la presente invención puede ser obtenido a partir de microorganismos de cualquier género. Para objetivos de la presente invención, el término "obtenidos a partir de" según se utiliza en este caso en relación a una fuente dada significará que el polipéptido codificado por una secuencia de nucleótidos es producido por la fuente o  
55 por una cepa en la cual la secuencia de nucleótidos de la fuente ha sido insertada. En un aspecto preferido, el polipéptido obtenido de una fuente dada es segregado extracelularmente.

[0062] Un polipéptido de la presente invención puede ser un polipéptido bacteriano. Por ejemplo, el polipéptido puede ser un polipéptido de bacterias gram positivas tales como un polipéptido de *Bacillus*, p. ej., un polipéptido de *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, o de *Bacillus thuringiensis* con actividad endoglucanasa; o un polipéptido de *Streptomyces* con actividad endoglucanasa, p. ej., un polipéptido de *Streptomyces lividans* o *Streptomyces murinus* con actividad endoglucanasa; o un polipéptido de bacterias gram negativas, p. ej., un polipéptido de *E. coli* o *Pseudomona sp.* con actividad endoglucanasa.

60



[0063] Un polipéptido de la presente invención puede también ser un polipéptido fúngico y más preferiblemente un polipéptido de levadura tal como un polipéptido de *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, o *Yarrowia* con actividad endoglucanasa; o más preferiblemente un polipéptido filamentoso fúngico tal como un polipéptido de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium* o *Trichoderma* con actividad endoglucanasa.

[0064] En un aspecto preferido, el polipéptido es un polipéptido de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, o *Saccharomyces oviformis* con actividad endoglucanasa.

[0065] En otro aspecto preferido, el polipéptido es un polipéptido de *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride* con actividad endoglucanasa.

[0066] En otro aspecto preferido, el polipéptido es un polipéptido de *Thielavia achromatica*, *Thielavia albomyces*, *Thielavia albopilosa*, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovispora*, *Thielavia peruviana*, *Thielavia spededonium*, *Thielavia setosa*, *Thielavia subthermophila*, *Thielavia terrestris*, *Thielavia terricola*, *Thielavia thermophila*, *Thielavia variospora* o *Thielavia wareingii* con actividad endoglucanasa.

[0067] En un aspecto más preferido, el polipéptido es un polipéptido de *Thielavia terrestris* con actividad endoglucanasa, y más preferiblemente un polipéptido de *Thielavia terrestris* NRRL 8126 con actividad endoglucanasa, p. ej., el polipéptido de la SEC ID n.º: 2 o el polipéptido maduro de la misma.

[0068] Se entenderá que para las especies mencionadas la invención incluye los estados perfecto e imperfecto y otros equivalentes taxonómicos, p. ej., anamorfos, sin importar el nombre de la especie por el cual son conocidos. Los expertos en la técnica fácilmente reconocerán la identidad de los equivalentes apropiados.

[0069] Las cepas de estas especies son fácilmente accesibles al público en un número de colecciones de cultivo, tales como American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS) y Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

[0070] Además, los polipéptidos de este tipo pueden ser identificados y obtenidos a partir de otras fuentes incluidos los microorganismos aislados de la naturaleza (p. ej., tierra, abonos, agua, etc.) usando las sondas mencionadas más arriba. Las técnicas para aislar microorganismos de hábitats naturales son bien conocidas en la técnica. El polinucleótido entonces puede ser obtenido seleccionando de forma similar un ADN genómico o una biblioteca de ADNc de ese microorganismo. Una vez que se ha detectado una secuencia de polinucleótidos que codifica un polipéptido con la/s sonda/s, el polinucleótido puede ser aislado o clonado utilizando técnicas que son bien conocidas por los expertos en la técnica (ver, p. ej., Sambrook et al., 1989, supra).

[0071] Los polipéptidos de la presente invención también incluyen polipéptidos fusionados o polipéptidos de fusión seccionables en los cuales otro polipéptido se fusiona en el N-terminal o el C-terminal del polipéptido o fragmento del mismo. Un polipéptido fusionado es producido fusionando 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 ligar las secuencias codificantes que codifican los polipéptidos de modo que estén dentro del marco y que la expresión del polipéptido fusionado esté bajo control del/los mismo/s promotor/es y terminador.

## Polinucleótidos

[0072] La presente invención también se refiere a polinucleótidos aislados que comprenden o que consisten en una secuencia de nucleótidos que codifica un polipéptido de la presente invención con actividad endoglucanasa.

[0073] En un aspecto preferido, la secuencia de nucleótidos comprende o consiste en la SEC ID n.º: 1. En otro aspecto más preferido, la secuencia de nucleótidos comprende o consiste en la secuencia contenida en el plásmido pTter7F que está contenido en *E. coli* NRRL B-30837. En otro aspecto preferido, la secuencia de nucleótidos comprende o consiste en la región codificante del polipéptido maduro de la SEC ID n.º: 1. En otro aspecto preferido, la secuencia de nucleótidos comprende o consiste en los nucleótidos 52 a 1008 de la SEC ID n.º: 1. En otro aspecto más preferido, la secuencia de nucleótidos comprende o consiste en la región codificante del polipéptido maduro contenida en el

plásmido pTter7F que está contenido en *E. coli* NRRL B 30837. La presente invención también incluye secuencias de nucleótidos que codifican un polipéptido que comprende o que consiste en la secuencia de aminoácidos de la SEC ID n.º: 2 o el polipéptido maduro de la misma, que difiere de la SEC ID n.º: 1 o la secuencia codificante del polipéptido maduro de la misma en virtud de la degeneración del código genético. La presente invención también se refiere a subsecuencias de la SEC ID n.º: 1 que codifican fragmentos de la SEC ID n.º: 2 que tiene actividad endoglucanasa.

[0074] La presente invención también se refiere a polinucleótidos mutantes que comprenden al menos una mutación en la secuencia codificante del polipéptido maduro de la SEC ID n.º: 1, en la cual la secuencia de nucleótidos mutante codifica el polipéptido maduro de la SEC ID n.º: 2. En un aspecto preferido, el polipéptido maduro son los aminoácidos 18 a 336 de la SEC ID n.º: 2.

[0075] 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 de ese ADN genómico puede ser efectuada, p. ej., usando la reacción en cadena de la polimerasa (PCR) bien conocida o selección de anticuerpos de bibliotecas de expresión para detectar fragmentos de ADN clonados con características estructurales compartidas. Ver, por ej., Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Se pueden utilizar otros 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 a partir de una cepa de *Thielavia*, u otro organismo o un organismo relacionado y de ese modo, por ejemplo, puede ser una variante alélica o de especie de la región codificante del polipéptido de la secuencia de nucleótidos.

[0076] La presente invención también se refiere a polinucleótidos que comprenden secuencias de nucleótidos que tienen un grado de identidad con la secuencia codificante del polipéptido maduro de la SEC ID n.º: 1 de al menos el 60%, preferiblemente al menos el 65%, más preferiblemente al menos el 70%, más preferiblemente al menos el 75%, más preferiblemente al menos el 80%, más preferiblemente al menos el 85%, más preferiblemente al menos el 90%, incluso más preferiblemente al menos el 95% y de la forma más preferible al menos el 97% de identidad, que codifica un polipéptido activo. En un aspecto preferido, la secuencia que codifica el polipéptido maduro es los nucleótidos 52 a 1008 de la SEC ID n.º: 1.

[0077] 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 polipéptido se refiere a formas del polipéptido que no están presentes de manera natural. Estos polipéptidos pueden diferir de alguna manera creada genéticamente del polipéptido aislado de su fuente nativa, p. ej., variantes artificiales que difieren en actividad específica, termostabilidad, pH óptimo o similares. La secuencia de la variante puede estar construida sobre la base de la secuencia de nucleótidos presentada como la región codificante del polipéptido de la SEC ID n.º: 1, p. ej., 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 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 nucleótidos, ver, p. ej., Ford et al., 1991, Protein Expression and Purification 2: 95-107.

[0078] Resultará evidente para los expertos en la técnica que las sustituciones de este tipo pueden ser realizadas fuera de las regiones críticas a la función de la molécula y aún 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 en consecuencia preferiblemente no sometidos a la sustitución, pueden ser identificados según procedimientos conocidos en la técnica, tales como mutagénesis dirigida o mutagénesis por barrido de alanina (ver, p. ej., Cunningham y Wells, 1989, Science 244: 1081-1085). En esta última técnica, las mutaciones se introducen en cada residuo positivamente cargado en la molécula, y las moléculas mutantes resultantes se evalúan para detectar actividad endoglucanasa para identificar residuos de aminoácidos que son críticos para la actividad de la molécula. Los sitios de interacción de sustrato-enzima también pueden ser determinados por análisis de la estructura tridimensional como lo determinan técnicas tales como análisis por resonancia magnética nuclear, cristalografía o marcaje por fotoafinidad (ver, p. ej., 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).

[0079] En un aspecto preferido, la secuencia que codifica el polipéptido maduro son los nucleótidos 52 a 1008 de la SEC ID n.º: 1.

## **Constructos de ácidos nucleicos**

[0080] 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 bajo condiciones compatibles con las secuencias de control.

[0081] Un polinucleótido aislado que codifica un polipéptido de la presente invención puede ser manipulado en una variedad de maneras 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 de polinucleótidos utilizando métodos de ADN recombinante son bien conocidas en la técnica.

[0082] 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 del promotor contiene secuencias de control transcripcional que median la expresión del polipéptido. El promotor puede ser cualquier secuencia de nucleótidos que muestre actividad transcripcional en la célula huésped de elección incluidos promotores mutantes, truncados e híbridos, y pueden ser obtenidos a partir de genes que codifican polipéptidos extracelulares o intracelulares homólogos o heterólogos a la célula huésped.

[0083] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de ácidos nucleicos de la presente invención, especialmente en una célula huésped bacteriana, son los promotores obtenidos a partir del lac operon de *E. coli*, gen de agarasa de *Streptomyces coelicolor* (*dagA*), gen de levansucrasa de *Bacillus subtilis* (*sacB*), gen de alfa-amilasa de *Bacillus licheniformis* (*amyL*), gen de amilasa maltogénica de *Bacillus stearothermophilus* (*amyM*), gen de alfa-amilasa de *Bacillus amyloliquefaciens* (*amyQ*), gen de penicilinas de *Bacillus licheniformis* (*penP*), genes *xylA* y *xylB* de *Bacillus subtilis* y gen procariótico de beta lactamasa (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75: 3727-3731), al igual que el promotor *tac* (DeBoer et al., 1983, Proceedings of the National Academy of Sciences USA 80: 21-25). Otros promotores son descritos en "Useful proteins from recombinant bacteria" en Scientific American, 1980, 242: 74-94; y en Sambrook et al., 1989, *supra*.

[0084] 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 promotores obtenidos a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, proteinasa aspártica de *Rhizomucor miehei*, alfa-amilasa estable en ácido de *Aspergillus niger*, alfa-amilasa estable en ácido de *Aspergillus niger*, glucoamilasa (*glaA*) de *Aspergillus niger* o *Aspergillus awamori*, lipasa de *Rhizomucor miehei*, proteasa alcalina de *Aspergillus oryzae*, 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*, celobiohidrolasa II 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* e triosa fosfato isomerasa de *Aspergillus oryzae*); y promotores mutantes, truncados e híbridos de los mismos.

[0085] En un huésped de levadura, promotores útiles se obtienen de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactokinasa de *Saccharomyces cerevisiae* (GAL1), alcohol deshidrogenasa/glicer aldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH1, ADH2/GAP), triosa fosfato isomerasa de *Saccharomyces cerevisiae* (TPI), metalotionina de *Saccharomyces cerevisiae* (CUP1) y 3-fosfoglicerato quinasa de *Saccharomyces cerevisiae*. Otros promotores útiles para células huéspedes de levadura son descritos por Romanos et al., 1992, Yeast 8: 423-488.

[0086] La secuencia de control también puede ser una secuencia del terminador de la transcripción adecuado, una secuencia reconocida por una célula huésped para terminar la transcripción. La secuencia del terminador está operativamente enlazada al 3' término 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.

[0087] Los terminadores preferidos para células huéspedes filamentosas fúngicas 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*.

[0088] Los terminadores preferidos para células huéspedes de levadura se obtienen a partir de los genes para enolasa de *Saccharomyces cerevisiae*, citocromo C (CYC1) de *Saccharomyces cerevisiae* y glicer aldehído-3- deshidrogenasa fosfato de *Saccharomyces cerevisiae*. Otros terminadores útiles para las células huéspedes de levadura son descritos por Romanos et al., 1992, *supra*.

[0089] 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 5' término 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 de elección.

[0090] Líderes preferidos para células huéspedes fúngicas filamentosas se obtienen de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.

- [0091] Los líderes adecuados para células huéspedes de levadura se obtienen a partir de los genes para enolasa (ENO-1) de *Saccharomyces cerevisiae*, fosfoglicerato-quinasa de *Saccharomyces cerevisiae*, factor alfa de *Saccharomyces cerevisiae* y alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa (ADH2/GAP) de *Saccharomyces cerevisiae*.
- [0092] La secuencia de control también puede ser una secuencia de poliadenilación, una secuencia operativamente enlazada al 3' término 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 de elección.
- [0093] Las secuencias de poliadenilación preferidas para las células huéspedes fúngicas filamentosas se obtienen a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, antranilato sintasa de *Aspergillus nidulans*, proteasa tipo tripsina de *Fusarium oxysporum* y alfa-glucosidasa de *Aspergillus niger*.
- [0094] Secuencias de poliadenilación útiles para células huéspedes de levadura son descritas por Guo y Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.
- [0095] La secuencia de control también puede ser una región codificante del péptido señal que codifica para una secuencia de aminoácidos enlazada al amino término de un polipéptido y dirige el polipéptido codificado en la vía secretora de la célula. El extremo 5' de la secuencia codificante de la secuencia de nucleótidos puede contener intrínsecamente una región codificante del péptido señal naturalmente enlazada en el marco de lectura de traducción con el segmento de la región codificante que codifica el polipéptido segregado. De forma alternativa, el extremo 5' de la secuencia codificante puede contener una región codificante del péptido señal que sea externa a la secuencia codificante. La región codificante del péptido señal extranjero puede ser requerida donde la secuencia codificante naturalmente no contiene una región codificante del péptido señal. De forma alternativa, la región codificante del péptido señal extranjero puede simplemente reemplazar la región codificante del péptido señal natural para aumentar la secreción del polipéptido. No obstante, en la presente invención se puede utilizar cualquier región codificante del péptido señal que dirige el polipéptido expresado a la vía secretora de una célula huésped de elección, es decir, segregada en un medio de cultivo.
- [0096] Las regiones codificantes del péptido señal eficaces para células huéspedes bacterianas son las regiones codificantes del péptido señal obtenidas a partir de los genes para amilasa maltogénica de *Bacillus* NCIB 11837, alfa-amilasa de *Bacillus stearothermophilus*, subtilisina de *Bacillus licheniformis*, beta-lactamasa de *Bacillus licheniformis*, proteasas neutras de *Bacillus stearothermophilus* (nprT, nprS, nprM), y prsA de *Bacillus subtilis*. Otros péptidos señal son descritos por Simonen y Palva, 1993, Microbiological Reviews 57: 109-137.
- [0097] Las regiones codificantes del péptido señal eficaces para células huéspedes fúngicas filamentosas son las regiones codificantes del péptido señal obtenidas a partir de los genes para TAKA amilasa de *Aspergillus oryzae*, amilasa neutra de *Aspergillus niger*, glucoamilasa de *Aspergillus niger*, proteinasa aspártica de *Rhizomucor miehei*, celulasa de *Humicola insolens*, endoglucanasa V de *Humicola insolens*, y lipasa de *Humicola lanuginosa*.
- [0098] En un aspecto preferido, el péptido señal son los aminoácidos 1 a 17 de la SEC ID n.º: 2. En otro aspecto preferido, la región codificante del péptido señal son los nucleótidos 1 a 51 de la SEC ID n.º: 1 que codifica los aminoácidos 1 a 17 de la SEC ID n.º: 2.
- [0099] Los péptidos señal útiles para las células huéspedes de levadura se obtienen a partir de los genes para factor alfa de *Saccharomyces cerevisiae* e invertasa de *Saccharomyces cerevisiae*. Otras regiones codificantes del péptido señal útiles son descritas por Romanos et al., 1992, *supra*.
- [0100] La secuencia de control también puede ser una región codificante del propéptido que codifica una secuencia de aminoácidos situada en el amino término de un polipéptido. El polipéptido resultante es conocido como una proenzima o propolipéptido (o un zimógeno en algunos casos). Un propolipéptido generalmente está inactivo y puede ser convertido a un polipéptido activo maduro por corte catalítico o autocatalítico del propéptido del propolipéptido. La región codificante del propéptido puede ser obtenida a partir de los genes para proteasa alcalina de *Bacillus subtilis* (aprE), proteasa neutra de *Bacillus subtilis* (nprT), factor alfa de *Saccharomyces cerevisiae*, proteinasa aspártica de *Rhizomucor miehei* y lacasa de *Myceliophthora thermophila* (WO 95/33836).
- [0101] Donde tanto las regiones del péptido señal como del propéptido están presentes en el amino término de un polipéptido, la región del propéptido está situada junto al amino término de un polipéptido y la región del péptido señal está situada junto al amino término de la región del propéptido.
- [0102] También puede ser conveniente añadir secuencias reguladoras que permitan la regulación de la expresión del polipéptido relativa al crecimiento de la célula huésped. Ejemplos de sistemas reguladores son los que causan que la expresión del gen se active o desactive en respuesta a un estímulo químico o físico, incluida la presencia de un compuesto regulador. Los sistemas reguladores en sistemas procarióticos incluyen los sistemas de operadores *lac*, *tac*

y *trp*. En la levadura, se puede utilizar el sistema ADH2 o el sistema GAL1. En hongos filamentosos, se pueden utilizar el promotor de TAKA alfa-amilasa, el promotor de glucoamilasa de *Aspergillus niger* y el promotor de glucoamilasa de *Aspergillus oryzae* como secuencias reguladoras. Otros ejemplos de secuencias reguladoras son los que permiten la amplificación génica. En sistemas eucarióticos, estos incluyen el gen de la 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

[0103] La presente invención también se refiere a vectores de expresión recombinantes que comprenden un polinucleótido de la presente invención, un promotor y señales de terminación transcripcionales y traduccionales. Los diferentes ácidos nucleicos y secuencias de control descritos en la presente pueden ser unidos para producir un vector de expresión recombinante que puede incluir uno o más sitios de restricción convenientes para permitir la inserción o la sustitución de la secuencia de nucleótidos que codifica el polipéptido en esos sitios. De forma alternativa, una secuencia de nucleótidos de la presente invención puede ser expresada insertando 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 esté operativamente enlazada con las secuencias de control apropiadas para la expresión.

[0104] El vector de expresión recombinante puede ser cualquier vector (p. ej., un plásmido o virus) que puede ser convenientemente sometido a procedimientos de ADN recombinante y puede provocar la expresión de la secuencia de nucleótidos. La elección del vector normalmente dependerá de la compatibilidad del vector con la célula huésped en la cual se introducirá el vector. Los vectores pueden ser plásmidos circulares lineales o cerrados.

[0105] 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, p. ej., 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, introducido en la célula huésped, es integrado en el genoma y replicado con el o los cromosomas en los cuales ha sido integrado. Además, se puede utilizar un único vector o plásmido o dos o más vectores o plásmidos que juntos contienen el ADN total que se introducirá en el genoma de la célula huésped o un transposón.

[0106] Los vectores de la presente invención preferiblemente contienen uno o más marcadores seleccionables que permiten la fácil selección de células transformadas, modificadas, transducidas o similares. Un marcador seleccionable es un gen cuyo producto proporciona resistencia biocida o vírica, resistencia a metales pesados, prototrofia a auxótrofos y similares.

[0107] Ejemplos de marcadores seleccionables bacterianos son los genes *dal* de *Bacillus subtilis* o *Bacillus licheniformis*, o marcadores que confieren resistencia antibiótica tales como ampicilina, canamicina, cloranfenicol o resistencia a la tetraciclina. Marcadores adecuados para células huéspedes de levadura son ADE2, HIS3, LEU2, LYS2, MET3, TRP1 y URA3. Marcadores seleccionables para el uso en una célula huésped fúngica filamentosa incluyen, pero sin limitarse a ellos, *amdS* (acetamidasa), *argB* (ornitina-carbamoyltransferasa), *bar* (fosfinotricina acetiltransferasa), *hph* (higromicina fosfotransferasa), *niaD* (nitrato-reductasa), *pyrG* (orotidina-5'-fosfato-descarboxilasa), *sC* (sulfato adeniltransferasa), y *trpC* (antranilato sintasa) y equivalentes de los mismos. Para el uso en una célula de *Aspergillus* se prefieren los genes *amdS* y *pyrG* de *Aspergillus nidulans* o *Aspergillus oryzae* y el gen *bar* de *Streptomyces hygroscopicus*.

[0108] Los vectores de la presente invención preferiblemente contienen un elemento o elementos que permiten la integración del vector en el genoma de la célula huésped o la replicación autónoma del vector en la célula independiente del genoma.

[0109] Para la integración en el genoma de la célula huésped, el vector puede depender de la secuencia del polinucleótido que codifica el polipéptido o 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 la integración por recombinación homóloga en el genoma de la célula huésped a una ubicación o ubicaciones precisas en el o los cromosomas. Para aumentar la probabilidad de integración en una ubicación precisa, los elementos integracionales preferiblemente deberían contener un número suficiente de ácidos nucleicos, tal como 100 a 10000 pares de bases, preferiblemente 400 a 10000 pares de bases y más preferiblemente 800 a 10000 pares de bases, que tienen un alto grado de identidad con la secuencia objetivo correspondiente para aumentar la probabilidad de recombinación homóloga. Los elementos integracionales pueden ser cualquier secuencia que sea homóloga a la secuencia objetivo en el genoma de la célula huésped. Además, los elementos integracionales pueden ser secuencias de nucleótidos no codificantes o codificantes. En cambio, el vector puede ser integrado en el genoma de la célula huésped por recombinación no homóloga.

[0110] Para la replicación autónoma, el vector además puede comprender un origen de replicación que permita que el vector se replique de manera autónoma en la célula huésped en cuestión. El origen de la replicación puede ser cualquier plásmido replicador que medie la replicación autónoma que funciona en una célula. El término "origen de la

replicación" o "replicador del plásmido" es definido en la presente como una secuencia de nucleótidos que permite que un plásmido o vector se replique *in vivo*.

[0111] Ejemplos de orígenes bacterianos de replicación son los orígenes de replicación de los plásmidos pBR322; pUC19; pACC177, y pACYC184 que permiten la replicación en *E. coli* y pUB110; pE194; pTA1060 y pAM $\beta$ 1 que permiten la replicación en *Bacillus*.

[0112] Ejemplos de orígenes de replicación para el uso en una célula huésped de la levadura son los orígenes de replicación de 2 micras, ARS1, ARS4, la combinación de ARS1 y CEN3 y la combinación de ARS4 y CEN6.

[0113] 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 lograr según los métodos descritos en WO 00/24883.

[0114] Se puede insertar más de una copia de un polinucleótido de la presente invención en la célula huésped para aumentar la producción del producto genético. 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 que contienen copias amplificadas del gen marcador seleccionable, y de ese modo copias adicionales del polinucleótido, pueden ser seleccionadas cultivando las células en presencia del agente seleccionable apropiado.

[0115] Los procedimientos usados para enlazar los elementos descritos anteriormente para construir los vectores de expresión recombinantes de la presente invención son conocidos por el experto en la técnica (ver, p. ej., Sambrook et al., 1989, supra).

#### **Células huéspedes**

[0116] La presente invención también se refiere a células huéspedes recombinantes que comprenden un polinucleótido de la presente invención, que son usadas ventajosamente 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 es mantenido como un integrante cromosómico o como un vector extracromosómico autorreplicante como se describe anteriormente. El término "célula huésped" incluye cualquier progenie de una célula madre que no es idéntica a la célula madre debido a mutaciones que ocurren durante la replicación. La elección de una célula huésped dependerán en gran parte del gen que codifica el polipéptido y su fuente.

[0117] La célula huésped puede ser un microorganismo unicelular, p. ej., un procariota, o un microorganismo no unicelular, p. ej., un eucariota.

[0118] Microorganismos unicelulares útiles son las células bacterianas tales como bacterias gram positivas incluyendo, pero sin limitarse a ellas, una célula de *Bacillus*, p. ej., *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis* y *Bacillus thuringiensis*; o una célula de *Streptomyces*, por ej., *Streptomyces lividans* y *Streptomyces murinus* o bacterias gram negativas tales como *E. coli* y *Pseudomonas sp.* En un aspecto preferido, la célula huésped bacteriana es una célula de *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus* o *Bacillus subtilis*. En otro aspecto preferido, la célula de *Bacillus* es un *Bacillus alcalofílico*.

[0119] La introducción de un vector en una célula huésped bacteriana puede, por ejemplo, ser efectuada por transformación de protoplasto (ver, p. ej., Chang y Cohen, 1979, Molecular General Genetics 168: 111-115), usando células competentes (ver, p. ej., Young y Spizizen, 1961, Journal of Bacteriology 81: 823-829, o Dubnau y Davidoff-Abelson, 1971, Journal of Molecular Biology 56: 209-221), electroporación (ver, por ej., Shigekawa y Dower, 1988, Biotechniques 6: 742-751), o conjugación (ver, p. ej., Koehler y Thorne, 1987, Journal of Bacteriology 169: 5771-5278).

[0120] La célula huésped también puede ser un eucariota, tal como un mamífero, un insecto, una planta o una célula fúngica.

[0121] En un aspecto preferido, la célula huésped es una célula fúngica. "Hongos" según se utiliza en este caso incluye los filos Ascomycota, Basidiomycota, Chytridiomycota y Zygomycota (según lo definen Hawksworth et al., en, Ainsworth and Bisby's Dictionary of The Fungi, 8ª edición, 1995, CAB International, University Press, Cambridge, Reino Unido) al igual que el Oomycota (según lo citan Hawksworth et al., 1995, supra, página 171) y todos los hongos mitospóricos (Hawksworth et al., 1995, supra).

[0122] En un aspecto más preferido, la célula huésped fúngica es una célula de levadura. "Levadura" según se utiliza en este caso incluye levadura ascosporógena (Endomycetales), levadura basidiosporógena y levadura de los Fungi Imperfecti (Blastomycetes). Puesto que la clasificación de la levadura puede cambiar en el futuro, para los objetivos 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).

[0123] En un aspecto incluso más preferido, la célula huésped de la levadura es una célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*.

[0124] En un aspecto más preferido, la célula huésped de la levadura es una célula de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* o *Saccharomyces oviformis*. En otro aspecto preferido, la célula huésped de la levadura es una célula de *Kluyveromyces lactis*. En otro aspecto preferido, la célula huésped de la levadura es una célula de *Yarrowia lipolytica*.

[0125] En otro aspecto preferido, la célula huésped fúngica es una célula fúngica filamentosa. "Hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión Eumycota y Oomycota (según la definen Hawksworth 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 es por alargamiento hifal y el catabolismo de carbono es estrictamente aeróbico. Por el contrario, el crecimiento vegetativo por levaduras tales como *Saccharomyces cerevisiae* es por injerto de un talo unicelular y el catabolismo de carbono puede ser fermentativo.

[0126] En un aspecto 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*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trametes* o *Trichoderma*.

[0127] En un aspecto 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 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* o *Fusarium venenatum*. En otro aspecto preferido, la célula huésped fúngica filamentosa es una célula *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Ceriporiopsis subvermispora*, *Coprinus cinereus*, *Coriolus hirsutus*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Phlebia radiata*, *Pleurotus eryngii*, *Thielavia achromatica*, *Thielavia albomyces*, *Thielavia albopilosa*, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovispora*, *Thielavia peruviana*, *Thielavia spededonium*, *Thielavia setosa*, *Thielavia subthermophila*, *Thielavia terrestris*, *Thielavia terricola*, *Thielavia thermophila*, *Thielavia variopora*, *Thielavia wareingii*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.

[0128] Las células fúngicas pueden ser transformadas por un proceso que implica la formación de protoplastos, la transformación de los protoplastos y la regeneración de la pared celular de manera conocida *per se*. Los procedimientos adecuados para la transformación de células huéspedes de *Aspergillus* y *Trichoderma* están descritos en EP 238 023 y Yelton et al., 1984, Proceedings of the National Academy of Sciences USA 81: 1470-1474. Los métodos adecuados para transformar las especies de *Fusarium* son descritas por Malardier et al., 1989, Gene 78: 147-156 y WO 96/00787. La levadura puede ser transformada usando los procedimientos descritos por Becker y Guarente, en Abelson, J.N. and Simon, M.I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito et al., 1983, Journal of Bacteriology 153: 163; y Hinnen et al., 1978, Proceedings of the National Academy of Sciences USA 75: 1920.

#### Métodos de producción

[0129] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención, que comprenden: (a) cultivar una célula, que en su tipo de forma salvaje es capaz de producir el polipéptido, bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido. En un aspecto preferido, la célula es del género *Thielavia*. En un aspecto preferido, la célula es *Thielavia terrestris* NRRL 8126.

[0130] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención, que comprende: (a) cultivar una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido.

[0131] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención, que comprende: (a) cultivar una célula huésped bajo condiciones propicias para la producción del polipéptido, donde la

célula huésped comprende una secuencia de nucleótidos mutantes que comprenden al menos una mutación en la secuencia codificadora de polipéptido maduro de la SEC ID n.º: 1, donde la secuencia de nucleótidos mutante codifica un polipéptido que consiste en el polipéptido maduro de la SEC ID n.º: 2, y (b) recuperar el polipéptido. En un aspecto preferido, el polipéptido maduro son los aminoácidos 18 a 336 de la SEC ID n.º: 2.

[0132] En los métodos de producción de la presente invención, las células son cultivadas en un medio nutritivo adecuado para la producción del polipéptido usando métodos bien conocidos en la técnica. Por ejemplo, la célula puede ser cultivada por cultivo de matraz vibrante y fermentación a pequeña escala o gran escala (incluidas fermentaciones continuas, por lote, por lote alimentado o en estado sólido) en fermentadores de laboratorio o industriales realizados en un medio adecuado y bajo condiciones que permitan que el polipéptido sea expresado y/o aislado. El cultivo se desarrolla en un medio nutritivo adecuado que comprende fuentes de nitrógeno y carbono y sales inorgánicas, usando procedimientos conocidos en la técnica. Se encuentran disponibles medios adecuados de proveedores comerciales o pueden ser preparados según composiciones publicadas (p. ej., en catálogos de la American Type Culture Collection). Si el polipéptido es segregado en el medio nutritivo, el polipéptido puede ser recuperado directamente del medio. Si el polipéptido no es segregado en el medio, puede ser recuperado a partir de lisados celulares.

[0133] 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 la presente.

[0134] El polipéptido resultante puede ser recuperado usando métodos conocidos en la técnica. Por ejemplo, el polipéptido puede ser recuperado del medio nutritivo por procedimientos convencionales incluyendo, pero sin limitarse a ello, centrifugado, filtración, extracción, secado por pulverización, evaporación o precipitación.

[0135] Los polipéptidos de la presente invención pueden ser purificados por una variedad de procedimientos conocidos en la técnica incluyendo, pero sin limitarse a ello, cromatografía (p. ej., de intercambio iónico, de afinidad, hidrofóbico, cromatografía de exclusión de tamaños), procedimientos electroforéticos (p. ej., isoelectroenfoque preparatorio), solubilidad diferencial (p. ej., precipitación de sulfato de amonio), SDS-PAGE, o extracción (ver, p. ej., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989) para obtener polipéptidos sustancialmente puros.

## Plantas

[0136] La presente invención también se refiere a una planta transgénica, parte de planta, o célula vegetal que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad endoglucanasa de la presente invención para expresar y producir el polipéptido en cantidades recuperables. El polipéptido puede ser recuperado de la planta o parte de la planta. De forma alternativa, la planta o parte de la planta que contiene el polipéptido recombinante puede ser usado como tal para mejorar la calidad de un alimento o alimentos, p. ej., mejorar el valor nutritivo, la palatabilidad y las propiedades reológicas o para destruir un factor antinutritivo.

[0137] La planta transgénica puede ser dicotiledónea (una dicot) o monocotiledónea (una monocot). Ejemplos de plantas monocotiledóneas son las hierbas, tales como poa de prados (poa pratense, Poa), hierba forrajera tal como Festuca, Lolium, hierba templada, tal como Agrostis, y cereales, p. ej., trigo, avena, centeno, cebada, arroz, sorgo y mazorca (maíz).

[0138] Ejemplos de plantas dicotiledóneas son tabaco, legumbres, tales como altramuces, patata, remolacha azucarera, guisante, judía y semilla de soja, y plantas crucíferas (familia Brassicaceae) tales como coliflor, semilla de colza, y el organismo modelo estrechamente relacionado *Arabidopsis thaliana*.

[0139] Ejemplos de partes de planta son tallo, callo, hojas, raíz, frutas, semillas y tubérculos al igual que los tejidos individuales que comprenden estas partes, p. ej., epidermis, mesófilo, parenquima, tejidos vasculares, meristemas. Compartimentos de célula vegetal específicos, tales como cloroplastos, apoplastos, mitocondria, vacuolas, peroxisomas y citoplasma son también considerados parte de una planta. Además, cualquier célula vegetal, cualquiera que sea el origen del tejido, se considera una parte de la planta. De igual modo, las partes de la planta tales como tejidos específicos y células aisladas para facilitar la utilización de la invención también son consideradas partes de la planta, p. ej., embriones, endospermas, aleurona y revestimientos de semillas.

[0140] También están incluidas dentro del campo de la presente invención la descendencia de tales plantas, partes de planta y células vegetales.

[0141] La planta transgénica o célula vegetal que expresa un polipéptido de la presente invención puede ser construida conforme a métodos conocidos en la técnica. En breve, la planta o célula vegetal es construida incorporando uno o más constructos de expresión que codifican un polipéptido de la presente invención en el genoma huésped de la planta o genoma de cloroplasto y propagando la planta modificada o célula vegetal resultante en una planta transgénica o célula vegetal.



[0142] El constructo de expresión es convenientemente 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 de elección. Además, el constructo de expresión puede comprender un marcador seleccionable útil para identificar las células huéspedes en las cuales el constructo de expresión ha sido integrado y las secuencias de ADN necesarias para la introducción del constructo en la planta en cuestión (éste depende del método de introducción de ADN que se utilizará).

[0143] La elección de secuencias reguladoras, tales como secuencias de promotor y terminador y opcionalmente secuencias de señal o de tránsito es determinada, por ejemplo, basándose en 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 puede ser desarrollable, específica de la fase o del tejido, y el producto genético puede ser dirigido a un tejido específico o parte de la planta tal como las semillas u hojas. Las secuencias reguladoras son, por ejemplo, descritas por Tague et al., 1988, *Plant Physiology* 86: 506.

[0144] Para la expresión constitutiva, se pueden utilizar el promotor 35S-CaMV, de la ubiquitina de maíz 1 y de la actina de arroz 1 (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 sumidero de almacenamiento tales como semillas, tubérculos de patata y frutas (Edwards & Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), o de tejidos sumideros metabólicos tales como meristemas (Ito et al., 1994, *Plant Mol. Biol.* 24: 863-878), un promotor específico de las semillas tal como el promotor de glutelina, prolamina, globulina o albúmina del arroz (Wu et al., 1998, *Plant and Cell Physiology* 39: 885-889), un promotor de *Vicia faba* de la legúmina B4 y el gen de proteína de semilla desconocida de *Vicia faba* (Conrad et al., 1998, revista de *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 la proteína de almacenamiento de *Brassica napus*, o cualquier otro promotor específico de semilla conocido en la técnica, por ej., como se describe en WO 91/14772. Además, el promotor puede ser un promotor específico de una hoja tal como el promotor *rbcs* de arroz o tomate (Kyoizuka et al., 1993, *Plant Physiology* 102: 991-1000, el promotor del gen de adenina metiltransferasa del virus de *chlorella* (Mitra y Higgins, 1994, *Plant Molecular Biology* 26: 85-93), o el promotor del gen *aldP* de arroz (Kagaya et al., 1995, *Molecular and General Genetics* 248: 668-674), o un promotor inducible por herida tal como el promotor *pin2* de la patata (Xu et al., 1993, *Plant Molecular Biology* 22: 573-588). Asimismo, el promotor puede ser inducible por tratamientos abióticos tales como temperatura, sequía, o alteraciones en la salinidad o inducidas por sustancias aplicadas exógenamente que activan el promotor, p. ej., etanol, estrógenos, hormonas de planta tales como etileno, ácido abscísico y ácido giberélico y metales pesados.

[0145] También se puede utilizar un elemento intensificador del promotor para conseguir expresión más alta de un polipéptido de la presente invención en la planta. Por ejemplo, el elemento intensificador del promotor puede ser un intrón que esté colocado 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 la actina de arroz 1 para aumentar la expresión.

[0146] El gen marcador seleccionable y cualquier otra parte del constructo de expresión puede ser elegido de aquellos disponibles en la técnica.

[0147] El constructo de ácidos nucleicos se incorpora en el genoma de la planta según técnicas convencionales conocidas en la técnica, incluidas transformación mediada por *Agrobacterium*, transformación mediada por virus, microinyección, bombardeo de partículas, transformación biolística y electroporación (Gasser et al., 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto et al., 1989, *Nature* 338: 274).

[0148] Actualmente, la transferencia de genes mediada por *Agrobacterium tumefaciens* es el método de elección para generar dicotiledóneas transgénicas (para obtener una revisión, ver Hoooykas y Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38) y también puede ser usado para transformar monocotiledóneas, aunque frecuentemente se utilizan otros métodos de transformación para estas plantas. Actualmente, el método de elección para generar monocotiledóneas transgénicas es el bombardeo de partículas (oro microscópico o partículas de tungsteno revestidas con 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 del protoplasto como lo describe Omirulleh et al., 1993, *Plant Molecular Biology* 21: 415-428.

[0149] Después de la transformación, los transformantes que tienen incorporados el constructo de expresión son seleccionados y regenerados en plantas enteras según métodos bien conocidos en la técnica. Frecuentemente el procedimiento de transformación se diseña para la eliminación selectiva de genes de selección durante la regeneración o bien en las siguientes generaciones usando, por ejemplo, cotransformación con dos constructos de ADN-T separados o escisión específica del sitio del gen de selección por una recombinasa específica.

[0150] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que comprende: (a) cultivar una planta transgénica o una célula vegetal que comprende un polinucleótido que codifica un polipéptido que tiene actividad endoglucanasa de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido.

## Composiciones

[0151] La presente invención también se refiere a composiciones que comprenden un polipéptido de la presente invención. Preferiblemente, las composiciones se enriquecen en ese polipéptido. El término "enriquecen" indica que la actividad de la endoglucanasa de la composición ha sido aumentada, p. ej., con un factor de enriquecimiento de al menos 1.1.

[0152] La composición puede comprender un polipéptido de la presente invención como el componente enzimático más importante, p. ej., una composición monocomponente. De forma alternativa, la composición puede comprender actividades enzimáticas múltiples, tales como una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, glicosiltransferasa de ciclodextrina, desoxiribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolítica, peptidoglutaminasa, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa o xilanas. Las enzimas adicionales pueden ser producidas, 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 sarcochroum*, *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*.

[0153] Las composiciones de polipéptido pueden ser preparadas conforme a métodos conocidos en la técnica y pueden ser en forma de un líquido o una composición seca. Por ejemplo, la composición de polipéptido puede ser en forma de un granulado o un microgranulado. El polipéptido que se incluirá en la composición puede ser estabilizado conforme a métodos conocidos en la técnica.

[0154] A continuación se presentan 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 bajo las cuales se usa la composición pueden ser determinadas basándose en métodos conocidos en la técnica.

## Usos

[0155] La presente invención también está dirigida a métodos para usar polipéptidos que tienen actividad endoglucanasa, o composiciones de los mismos.

## Degradación de biomasa para monosacáridos, disacáridos y polisacáridos

[0156] Los polipéptidos que tienen actividad endoglucanasa, y las células huéspedes de la presente invención pueden ser usados en la producción de monosacáridos, disacáridos y polisacáridos como materias primas químicas o de fermentación de biomasa para la producción de etanol, plástico u otros productos o productos intermedios. Los polipéptidos que tienen actividad endoglucanasa pueden ser en forma de un caldo de fermentación crudo con o sin las células eliminadas o en forma de una preparación enzimática semi-purificada o purificada. De forma alternativa, una célula huésped de la presente invención puede ser usada como una fuente del polipéptido que tiene actividad endoglucanasa en un proceso de fermentación con la biomasa.

[0157] La biomasa puede incluir, de modo enunciativo y no limitativo, recursos de madera, desperdicios sólidos municipales, papel usado y residuos de cosecha (ver, por ejemplo, Wiseloge et al., 1995, en Handbook on Bioethanol (Charles E. Wyman, 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 Engineering/Biotechnology, T. Scheper, managing editor, Volume 65, pp.23-40, Springer-Verlag, New York).

[0158] El polisacárido predominante en la pared celular primaria de la biomasa es la celulosa, el segundo más abundante es la hemicelulosa y el tercero es la pectina. La pared celular secundaria, producida después de que la célula ha dejado de crecer, también contiene polisacáridos y se refuerza a través de lignina polimérica reticulada de manera covalente a la hemicelulosa. La celulosa es un homopolímero de anhidrocelobiosa y, por 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 generalmente es polimorfa, la celulosa se encuentra en el tejido vegetal principalmente como una matriz cristalina

insoluble de cadenas de glucano paralelas. Las hemicelulosas normalmente tienen un enlace de hidrógeno a la celulosa, al igual que otras hemicelulosas, lo cual ayuda a estabilizar la matriz de la pared celular.

[0159] 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 1,4-beta-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 1,4-beta-D-glucanos e hidrolizan D-celobiosa lentamente, y celobiohidrolasas (1-beta-D-glucano celobiohidrolasas, EC 3,2,1,91), que liberan D-celobiosa de 1,4-beta-glucanos.

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

[0160] Estas tres clases de enzimas trabajan juntas sinérgicamente dando como resultado decristalización e hidrólisis eficaces de la celulosa nativa de la biomasa para producir azúcares reducidos.

[0161] Los polipéptidos que tienen actividad endoglucanasa de la presente invención pueden ser usados conjuntamente con las enzimas mencionadas más arriba para degradar más aún el componente de celulosa del sustrato de biomasa, (ver, por ejemplo, Brigham et al., 1995, en Handbook on Bioethanol (Charles E. Wyman, editor), pp.119-141, Taylor & Francis, Washington D.C.; Lee, 1997, Journal of Biotechnology 56: 1-24).

[0162] El etanol puede ser producido por degradación enzimática de biomasa y conversión de los sacáridos liberados a etanol. Este tipo de etanol frecuentemente es denominado bioetanol o biocombustible. Puede ser usado como un aditivo de combustible o suplemento en mezclas de menos de 1% y hasta 100% (un sustituto de combustible).

## **Composiciones de detergentes**

[0163] Los polipéptidos que tienen actividad endoglucanasa de la presente invención pueden ser añadidos a una composición de detergente y así convertirse en un componente de la misma.

[0164] La composición de detergente de la presente invención puede ser, por ejemplo, formulada como una composición de detergente de lavado a mano o a máquina incluida una composición de aditivo de lavandería adecuada para el pretratamiento de las telas manchadas y una composición de suavizante añadida al enjuague, o formulada como una composición de detergente para el uso en tareas limpieza de superficies duras del hogar en general, o formuladas para operaciones de lavado a mano o de lavavajillas.

[0165] En un aspecto específico, la presente invención proporciona un aditivo de detergente que comprende los polipéptidos que tienen actividad endoglucanasa de la presente invención. El aditivo de detergente al igual que la composición de detergente puede comprender una o más enzimas adicionales tales como una proteasa, lipasa, cutinasa, una amilasa, carbohidrasa, celulasa, pectinasa, mananasa, arabinasa, galactanasa, xilanasa, oxidasa, p. ej., una lacasa, y/o peroxidasa.

[0166] En general las propiedades de los componentes enzimáticos deberían ser compatibles con el detergente seleccionado, (es decir, pH óptimo, compatibilidad con otros ingredientes enzimáticos y no enzimáticos, etc.), y los componentes enzimáticos deberían estar presentes en cantidades eficaces.

[0167] Proteasas: Las proteasas adecuadas incluyen las de origen animal, vegetal o microbiano. Se prefiere el origen microbiano. Están incluidos los mutantes modificados químicamente o creados genéticamente de proteínas. La proteasa puede ser una serina proteasa o una metaloproteasa, preferiblemente una proteasa alcalina microbiana o una proteasa tipo tripsina. Ejemplos de proteasas alcalinas son subtilisinas, especialmente las derivadas de *Bacillus*, p. ej., subtilisina Novo, subtilisina Carlsberg, subtilisina 309, subtilisina 147 y subtilisina 168 (descrita en WO 89/06279). Ejemplos de proteasas tipo tripsina son tripsina (p. ej., de origen porcino o bovino) y la proteasa de *Fusarium* descrita en WO 89/06270 y WO 94/25583.

[0168] Ejemplos de proteasas útiles son las variantes descritas en WO 92/19729, WO 98/20115, WO 98/20116 y WO 98/34946, especialmente las variantes con sustituciones en una o más de las siguientes posiciones: 27, 36, 57, 76, 87, 97, 101, 104, 120, 123, 167, 170, 194, 206, 218, 222, 224, 235 y 274.

[0169] Las enzimas proteásicas disponibles comercialmente preferidas incluyen Alcalase™, Savinase™, Primase™, Duralase™, Esperase™ y Kannase™ (Novozymes A/S), Maxatase™, Maxacal™, Maxapem™, Properase™, Purafect™, Purafect OxP™, FN2™ y FN3™ (Genencor International Inc.).

[0170] Lipasas: Las lipasas adecuadas incluyen aquellas de origen bacteriano o fúngico. Están incluidos los mutantes modificados químicamente o creados genéticamente de proteínas. Ejemplos de lipasas útiles incluyen lipasas de *Humicola* (sinónimo *Thermomyces*), por ej., de *H. lanuginosa* (*T. lanuginosus*) como se describe en EP 258 068 y EP 305 216 o de *H. insolens* como se describe en WO 96/13580, una lipasa de *Pseudomonas*, por ej. de *P. alcaligenes* o *P. pseudoalcaligenes* (EP 218 272), *P. cepacia* (EP 331 376), *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 ej., 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).

[0171] Otros ejemplos son variantes de lipasa tales como aquellas descritas en WO 92/05249, WO 94/01541, EP 407 225, EP 260 105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 y WO 97/07202.

[0172] Las lipasas preferidas disponibles comercialmente incluyen Lipolase™, Lipex™ y Lipolase Ultra™ (Novozymes A/S).

[0173] Amilasas: Las amilasas adecuadas ( $\alpha$  y/o  $\beta$ ) incluyen aquellas de origen bacteriano o fúngico. Están incluidos los mutantes modificados químicamente o creados genéticamente de proteínas. Las amilasas incluyen, por ejemplo,  $\alpha$ -amilasas obtenidas a partir de *Bacillus*, p. ej., una cepa especial de *Bacillus licheniformis*, descrita en más detalle en GB 1 296 839.

[0174] Ejemplos de amilasas útiles son las variantes descritas en WO 94/02597, WO 94/18314, WO 96/23873, y WO 97/43424, especialmente las variantes con sustituciones en una o más de las siguientes posiciones: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408 y 444.

[0175] Las amilasas disponibles comercialmente son Duramyl™, Termamyl™, Fungamyl™ y BAN™ (Novozymes A/S), Rapidase™ y Purastar™ (de Genencor International Inc.).

[0176] Celulasas: Las celulasas adecuadas incluyen aquellas de origen bacteriano o fúngico. Están incluidos los mutantes modificados químicamente o creados genéticamente de proteínas. Las celulasas adecuadas incluyen celulasas de los géneros *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium* o *Trichoderma*, por ej., las celulasas fúngicas producidas de *Humicola insolens*, *Myceliophthora thermophila* y *Fusarium oxysporum* descritas en la patente U.S. n.º 4 435 307, la patente U.S. n.º 5 648 263, la patente U.S. n.º 5 691 178, la patente U.S. n.º 5 776 757 y WO 89/09259.

[0177] Celulasas especialmente adecuadas son las celulasas alcalinas o neutrales que tienen beneficios de cuidado del color. Ejemplos de celulasas de este tipo son las celulasas descritas en EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son variantes de celulasa tales como las descritas en WO 94/07998, EP 0 531 315, patente U.S. n.º 5,457,046, patente U.S. n.º 5,686,593, patente U.S. n.º 5,763,254, WO 95/24471, WO 98/12307 y PCT/DK98/00299.

[0178] Las celulasas disponibles comercialmente incluyen Celluclast®, Celluzyme™, y Carezyme™ (Novozymes A/S), Clazinase™ y Puradax HA™ (Genencor International Inc.) y KAC-500(B)™ (Kao Corporation).

[0179] Peroxidasas/oxidasas: Las peroxidasas/oxidasas adecuadas incluyen las de origen vegetal, bacteriano o fúngico. Están incluidos los mutantes modificados químicamente o creados genéticamente de proteínas. Los ejemplos de peroxidasas útiles incluyen peroxidasas de *Coprinus*, p. ej., de *C. cinereus*, y variantes de las mismas como las descritas en WO 93/24618, WO 95/10602 y WO 98/15257.

[0180] Las peroxidasas disponibles comercialmente incluyen Guardzyme™ (Novozymes A/S).

[0181] El o los componente(s) enzimático(s) puede(n) ser incluido(s) en una composición de detergente añadiendo aditivos separados que contienen una o más enzimas, o añadiendo un aditivo combinado que comprende todas estas enzimas. Un aditivo de detergente de la presente invención, es decir, un aditivo separado o un aditivo combinado, puede ser formulado, por ejemplo, como un granulado, líquido, mezcla, etc. Las formulaciones de aditivo de detergente preferidas son granulados, en particular granulados no en polvo, líquidos, en particular líquidos estabilizados, o pastas.

[0182] Los granulados no en polvo pueden ser producidos, p. ej., como se describe en la patente U.S. n.º 4 106 991 y 4 661 452 y opcionalmente pueden ser revestidos por métodos conocidos en la técnica. Ejemplos de materiales de revestimiento ceroso son productos de poli(óxido de etileno) (polietilenglicol, PEG) con pesos molares medios de 1000 a 20000; nonilfenoles etoxilados que tienen de 16 a 50 unidades de óxido de etileno; alcoholes etoxilados grasos en los cuales el alcohol contiene de 12 a 20 átomos de carbono y en los cuales hay 15 a 80 unidades de óxido de etileno; alcoholes grasos; ácidos grasos; y mono- y di- y triglicéridos de ácidos grasos. En GB 1483591 se presentan ejemplos de materiales de revestimientos formadores de película adecuados para la aplicación por técnicas de lecho fluidizado. Las preparaciones enzimáticas líquidas pueden, por ejemplo, ser estabilizadas añadiendo un poliol tal como propilenglicol, un azúcar o alcohol de azúcar, ácido láctico o ácido bórico según métodos establecidos. Las enzimas protegidas pueden ser preparadas según el método descrito en EP 238 216.

[0183] La composición de detergente de la presente invención puede estar en cualquier forma conveniente, p. ej., una barra, una pastilla, un polvo, un gránulo, una pasta o un líquido. Un detergente líquido puede ser acuoso, normalmente con un 70% de agua como máximo y un 0-30% de solvente orgánico, o no acuoso.

[0184] La composición de detergente comprende uno o más agentes tensioactivos, que pueden ser no iónicos incluyendo semipolares y/o aniónicos y/o catiónicos y/o zwitteriónicos. Los agentes tensioactivos normalmente están presentes en un nivel del 0,1% al 60% en peso.

[0185] Cuando está incluido, el detergente normalmente contiene del 1% aproximadamente al 40% aproximadamente de un tensioactivo aniónico tal como alquilbenceno sulfonato lineal, alfa olefinsulfonato, sulfato de alquilo (sulfato de alcohol graso), etoxisulfato de alcohol, alcanosulfonato secundario, éster alfa-sulfo metílico de ácido graso, ácido alquilo- o alqueniilsuccínico o jabón.

[0186] Cuando está incluido, el detergente normalmente contiene del 0,2% aproximadamente al 40% aproximadamente de un tensioactivo no iónico tal como alcohol etoxilato, nonilfenol, etoxilato, alquilpoliglicósido, óxido de alquildimetilamina, monoetanolamida etoxilada de ácidos grasos, monoetanolamida de ácidos grasos, polihidroxi alquilo amidas de ácidos grasos o derivados de N-acilo N-alquilo de glucosamina ("glucamidas").

[0187] El detergente puede contener un 0-65 % de un constructor de detergente o agente complejante tal como zeolita, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilenodiaminatetraacético, ácido dietilenotriaminopentaacético, ácido alquilo o alqueniilsuccínico, silicatos solubles o silicatos estratificados (p. ej., SKS-6 de Hoechst).

[0188] El detergente puede comprender uno o más polímeros. Ejemplos son carboximetilcelulosa, poli(vinilpirrolidona), poli(etileno glicol), poli(vinil alcohol), poli(N-óxido de vinilpiridina), poli(vinilimidazol), policarboxilatos tales como poliacrilatos, copolímeros de ácido maléico/acrílico, y copolímeros de lauril metacrilato/ ácido acrílico.

[0189] El detergente puede contener un sistema blanqueante que puede comprender una fuente de H<sub>2</sub>O<sub>2</sub> tal como perborato o percarbonato que puede ser combinada con un activador blanqueante formador de perácido tal como tetraacetililenodiamina o nonanoiloxibencenosulfonato. De forma alternativa, el sistema blanqueante puede comprender peroxiácidos, por ejemplo, de tipo amida, imida o sulfona.

[0190] El o los componente(s) enzimático(s) de la composición de detergente de la presente invención pueden ser estabilizados usando agentes estabilizantes convencionales, p. ej., un poliol tal como propilenoglicol o glicerol, un azúcar o alcohol de azúcar, ácido láctico, ácido bórico, o un derivado de ácido bórico, p. ej., un éster de borato aromático, o un derivado de ácido fenil borónico tal como ácido 4-formilfenil borónico y la composición puede ser formulada como se describe, por ejemplo, en WO 92/19709 y WO 92/19708.

[0191] El detergente también puede contener otros ingredientes de detergentes convencionales tales como acondicionadores de telas incluidas arcillas, reforzadores de espuma, supresores de espuma, agentes anticorrosivos, agentes de suspensión de suciedad, agentes contra la redeposición de suciedad, tintes, bactericidas, blanqueadores ópticos, hidrótopos, inhibidores de decoloración o perfumes.

[0192] En las composiciones detergentes cualquier componente enzimático, en particular los polipéptidos que tienen actividad endoglucanasa 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.

[0193] Los polipéptidos que tienen actividad endoglucanasa de la presente invención adicionalmente pueden ser incorporados en las formulaciones de detergente descritas en WO 97/07202 que están incorporadas en la presente a modo de referencia.

#### **Péptido señal**

[0194] La presente invención también se refiere a constructos de ácidos nucleicos que comprenden un gen que codifica una proteína operativamente enlazada a una secuencia de nucleótidos que codifica un péptido señal que comprende o que consiste en los aminoácidos 1 a 17 de la SEC ID n.º: 2, lo cual permite la secreción de la proteína en un medio de cultivo, en el cual el gen es externo a la secuencia de nucleótidos.

[0195] En un aspecto preferido, la secuencia de nucleótidos comprende los nucleótidos 1 a 51 de la SEC ID n.º: 1. En otro aspecto preferido, la secuencia de nucleótidos consiste en los nucleótidos 1 a 51 de la SEC ID n.º: 1.

[0196] La presente invención también se refiere a vectores de expresión recombinantes y células huéspedes recombinantes que comprenden constructos de ácidos nucleicos de este tipo.

[0197] La presente invención también se refiere a métodos para producir una proteína que comprende: (a) cultivar esa célula huésped recombinante bajo condiciones adecuadas para la producción de la proteína; y (b) recuperar la proteína.

[0198] La proteína puede ser nativa o heteróloga a una célula huésped. El término "proteína" no tiene como objetivo referirse en la presente a una longitud específica del producto codificado y, en consecuencia, comprende péptidos, oligopéptidos y proteínas. El término "proteína" también comprende dos o más polipéptidos combinados para formar el producto codificado. Las proteínas también incluyen polipéptidos híbridos que comprenden una combinación de secuencias de polipéptidos parciales o completas obtenidas de al menos dos proteínas diferentes en las cuales una o más pueden ser heterólogas o nativas a la célula huésped. Las proteínas además incluyen variaciones alélicas de origen natural y creadas genéticamente de las proteínas mencionadas anteriormente y las proteínas híbridas.

[0199] 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 preferido, la proteína es una oxidoreductasa, transferasa, hidrolasa, liasa, isomerasa o ligasa. En un aspecto preferido, la proteína es una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, glicosiltransferasa ciclodextrina, desoxirribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, invertasa, lacasa, lipasa, manosidasa, mutanasa, oxidasa, enzima pectinolítica, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa o xilanasas.

[0200] El gen puede ser obtenido a partir de cualquier fuente procariótica, eucariótica u otra fuente.

[0201] La presente invención está descrita además por los siguientes ejemplos que no deberían ser interpretados como limitativos del alcance de la invención.

## Ejemplos

### Materiales

[0202] Los productos químicos usados como tampones y sustratos fueron productos comerciales de al menos grado reactivo.

[0203] Los geles SDS-PAGE, el tampón de carga y el tampón de desplazamiento fueron obtenidos de Invitrogen/Novex (Carlsbad, CA). La tripsina de grado de secuenciación modificado fue de Princeton Separations (Aldelphia, NJ). Las manchas de proteína BioSafe Commassie Blue G250 fueron obtenidas de BioRad Laboratories (Hercules, CA).

### Cepas

[0204] La cepa Jal250 de *Aspergillus oryzae* (WO 99/61651) fue usada para la expresión de un polipéptido de *Thielavia terrestris* con actividad endoglucanasa. La cepa NRRL 8126 de *Thielavia terrestris* fue usada como la fuente de un gen para un polipéptido de la familia 7F con actividad endoglucanasa.

### Medios

[0205] Las placas PDA fueron compuestas por litro de 39 gramos de agar dextrosa de patata.

[0206] El medio NNCYP fue compuesto por litro de 5,0 g de  $\text{NH}_4\text{NO}_3$ , 0,5 g de  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0,3 g de  $\text{CaCl}_2$ , 2,5 g de ácido cítrico, 1,0 g de Bacto-peptona, 5,0 g de extracto de levadura, 1 ml de metales traza de COVE, y suficiente  $\text{K}_2\text{HPO}_4$  para conseguir un pH final de aproximadamente 5,4.

[0207] El medio NNCYPmod fue compuesto por litro de 1,0 g de  $\text{NaCl}$ , 5,0 g de  $\text{NH}_4\text{NO}_3$ , 0,2 g de  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 0,2 g de  $\text{CaCl}_2$ , 2,0 g de ácido cítrico, 1,0 g de Bacto peptona, 5,0 g de extracto de levadura, 1 ml de solución de metales traza de COVE y suficiente  $\text{K}_2\text{HPO}_4$  para conseguir el pH final de aproximadamente 5,4.

[0208] La solución de metales traza de Cove fue compuesta por litro de 0,04 g de  $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$ , 0,4 g de  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 1,2 g de  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 0,7 g de  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ , 0,8 g de  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  y 10 g de  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ .

[0209] Las placas LB fueron compuestas por litro de 10 g de triptona, 5 g de extracto de levadura, 5 g de cloruro sódico y 15 g de Bacto agar.

[0210] El medio MDU2BP fue compuesto por litro de 45 g de maltosa, 1 g de  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 1 g de  $\text{NaCl}$ , 2 g de  $\text{K}_2\text{HSO}_4$ , 12 g de  $\text{KH}_2\text{PO}_4$ , 2 g de urea y 500  $\mu\text{l}$  de metales de traza AMG, el pH fue ajustado a 5,0 y luego esterilizado por filtro con una unidad de filtración de 0,22  $\mu\text{m}$ .

[0211] Los metales traza de AMG fueron compuestos por litro de 14,3 g de  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 2,5 g de  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 0,5 g de  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ , 13,8 g de  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ , 8,5 g de  $\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$  y 3 g de ácido cítrico.

[0212] El medio Soc fue compuesto por 2% de triptona, 0,5% de extracto de levadura, 10 mM de  $\text{NaCl}$ , 2,5 mM de  $\text{KCl}$ , 10 mM de  $\text{MgCl}_2$  y 10 mM de  $\text{MgSO}_4$ , esterilizado por autoclave y luego se añadió glucosa esterilizada por filtro a 20 mM.

[0213] El medio de congelación fue compuesto por 60% de SOC y 40% de glicerol.

[0214] El medio 2X YT fue compuesto por litro de 16 g de triptona, 10 g de extracto de levadura, 5 g de NaCl y 15 g de Bacto agar, esterilizado por autoclave.

### Ejemplo 1: Construcción de biblioteca de ADNc de marcadores de secuencia expresada (EST)

[0215] *Thielavia terrestris* NRRL 8126 fue cultivada en 50 ml de medio NNCYPmod suplementado con 1% de glucosa en un matraz de 250 ml a 45°C, 200 r.p.m. durante 24 horas. Se utilizó una alícuota de dos ml del cultivo líquido de 24 horas para cultivar un matraz de 500 ml con 100 ml de medio NNCYPmod suplementado con 2% de Sigmacell-20 (Sigma Chemical Co., St. Louis, MO). El cultivo fue incubado a 45°C, 200 r.p.m. durante 3 días. Los micelios fueron recolectados por filtración a través de un embudo de Buchner con un prefiltro de fibra de vidrio (Nalgene, Rochester, NY), lavados dos veces con 10 mM de Tris-HCl-1 mM de EDTA pH 8 (TE), y rápidamente congelados en nitrógeno líquido.

[0216] El ARN total fue aislado usando el siguiente método. Los micelios congelados de *Thielavia terrestris* NRRL 8126 fueron triturados en un molinillo de café eléctrico. El material triturado fue mezclado 1:1 v/v con 20 ml de Fenazol (Ambion, Inc., Austin, TX) en un tubo tipo Falcon de 50 ml. Una vez que los micelios fueron suspendidos, fueron extraídos con cloroformo y tres veces con una mezcla de fenol-cloroformo-álcool isoamílico 25:24:1 v/v/v. De la fase acuosa resultante, el ARN fue precipitado añadiendo 1/10 volumen de 3 M de acetato sódico pH 5,2 y 1,25 volúmenes de isopropanol. El ARN precipitado fue recuperado por centrifugado a 12 000 x g durante 30 minutos a 4°C. El granulado final fue lavado con 70% de etanol frío, secado al aire y resuspendido en 500 ml de agua tratada con dietilpirocarbonato (agua DEPC).

[0217] La calidad y la cantidad del ARN purificado fueron evaluadas con un Agilent Bioanalyzer 2100 (Agilent Technologies, Inc., Palo Alto, CA). El ARNm poliadenilado fue aislado de 360 µg de ARN total con la ayuda de un Poli (A) Purist Magnetic Kit (Ambion, Inc., Austin, TX) según las instrucciones del fabricante.

[0218] Para crear la biblioteca de ADNc, un kit CloneMiner™ (Invitrogen, Carlsbad, CA) fue empleado para construir una biblioteca direccional que no requiere el uso de clonación de enzimas de restricción, reduciendo así el número de clones quiméricos y el sesgo de tamaño.

[0219] Para asegurar la síntesis exitosa de la primera cadena de ADNc, se realizaron dos reacciones en paralelo con dos concentraciones diferentes de ARNm (2,2 y 4,4 µg de poli(A)<sup>+</sup> ARNm). Las muestras de ARNm fueron mezcladas con un cebador Biotin-attB2- Oligo(dt) (CloneMiner™ Kit, Invitrogen, Carlsbad, CA), 1X tampón de primera cadena (Invitrogen, Carlsbad, CA), 2 µl de 0,1 M de ditiotretiol (DTT), 10 mM de cada dNTP y agua a un volumen final de 18 y 16 µl, respectivamente. Las mezclas de reacción fueron mezcladas cuidadosamente y luego 2 y 4 µl de transcriptasa inversa SuperScript™ (Invitrogen, Carlsbad, CA) fueron añadidos e incubados a 45°C durante 60 minutos para sintetizar la primera cadena complementaria.

[0220] Para la síntesis de la segunda cadena, a cada reacción de primera cadena se añadieron 30 µl de 5X tampón de segunda cadena (Invitrogen, Carlsbad, CA), 3 µl de 10 mM de cada dNTP, 10 unidades de ADN ligasa de *E. coli* (Invitrogen, Carlsbad, CA), 40 unidades de ADN polimerasa I de *E. coli* (Invitrogen, Carlsbad, CA), y 2 unidades de ribonucleasa H de *E. coli* (Invitrogen, Carlsbad, CA) en un volumen total de 150 µl. Las mezclas luego fueron incubadas a 16°C durante dos horas. Después de la incubación de dos horas, se añadieron 2 µl de T4 ADN polimerasa (Invitrogen, Carlsbad, CA) a cada reacción y se incubaron a 16°C durante 5 minutos para crear un ADNc de extremos romos. Las reacciones de ADNc fueron extraídas con una mezcla de fenol-cloroformo-álcool isoamílico 25:24:1 v/v/v y precipitadas en presencia de 20 µg de glicógeno, 120 µl de 5 M de acetato amónico y 660 µl de etanol. Después del centrifugado a 12 000 x g durante 30 minutos a 4°C, los granulosos de ADNc fueron lavados con 70% de etanol frío, secados al vacío durante 2-3 minutos y resuspendidos en 18 µl de agua DEPC. A cada muestra de ADNc resuspendida se añadieron 10 µl de 5X tampón adaptado (Invitrogen, Carlsbad, CA), 10 µg de adaptador attB1 (Invitrogen, Carlsbad, CA) mostrado más abajo, 7 µl de 0,1 M de DTT y 5 unidades de T4 ADN ligasa (Invitrogen, Carlsbad, CA).

Cadena superior de adaptador attB1:

5'-TCGTCGGGGACAACCTTTGTACAAAAAAGTTGG-3' (SEQ ID N.º: 3)

Cadena inferior de adaptador attB1:

3'-CCCCTGTTGAAACATGTTTTTCAACCP-5' (SEQ ID N.º: 4)

[0221] Las reacciones de ligadura fueron incubadas durante toda la noche a 16°C. Los adaptadores de exceso fueron eliminados por cromatografía por exclusión de tamaños en 1 ml de resina Sephacryl™ S-500 HR (Amersham Biosciences, Piscataway, NJ). Las fracciones de columna fueron recogidas según las instrucciones del Kit CloneMiner™ y las fracciones 3 a 14 fueron analizadas con un Agilent Bioanalyzer para determinar la fracción en las cuales los adaptadores attB1 comenzaron a eluir. Este análisis mostró que los adaptadores comenzaron a eluir alrededor de la fracción 10 u 11. Para la primera biblioteca se agruparon las fracciones 6 a 11 y para la segunda biblioteca se agruparon las fracciones 4-11.

[0222] La clonación del ADNc fue realizada por recombinación homóloga de ADN según el Protocolo de Gateway System (Invitrogen, Carlsbad, CA) usando BP Clonase™ (Invitrogen, Carlsbad, CA) como la recombinasa. Cada reacción de recombinación de BP Clonase™ contuvo aproximadamente 70 ng de ADNc flanqueado por *attB*, 250 ng de pDONR™222, 2 µl de 5X de tampón BP Clonase™, 2 µl de tampón TE y 3 µl de BP Clonase™. Todos los reactivos fueron obtenidos de Invitrogen, Carlsbad, CA. Las reacciones de recombinación fueron incubadas a 25°C durante toda la noche.

[0223] Las reacciones de recombinación BP inactivadas por calor luego fueron divididas en 6 alícuotas y sometidas a electroporación en células electrocompetentes ElectroMax™ DH10B (Invitrogen, Carlsbad, CA) BioRad Gene Pulser II (BioRad, Hercules, CA) con los siguientes parámetros: voltaje: 2.0 kV; resistencia: 200 Ω; y Capacidad: 25 µF. Las células electroporadas fueron resuspendidas en 1 ml de medio SOC e incubadas a 37°C durante 60 minutos con agitación constante a 200 r.p.m. Después del periodo de incubación, las células transformadas fueron agrupadas y mezcladas 1:1 con medio de congelación. Una alícuota de 200 µl fue eliminada para titulación de la biblioteca y luego el resto de cada biblioteca fue dividido en alícuotas en crioviales de 1,8 ml (Wheaton Science Products, Millville, NJ) y almacenado congelado a -80°C.

[0224] Se prepararon cuatro diluciones en serie de cada biblioteca: 1/100, 1/1000, 1/104, 1/105. De cada dilución 100 µl fueron colocados sobre placas LB de 150 mM suplementadas con 50 µg de canamicina por ml e incubados a 37°C durante toda la noche. El número de colonias en cada placa de dilución fue contado y usado para calcular el número total de transformantes en cada biblioteca.

[0225] Se demostró que la primera biblioteca tenía 5,4 millones de clones independientes y se demostró que la segunda biblioteca tenía 9 millones de clones independientes.

## **Ejemplo 2: Preparación de molde y secuenciación nucleótida de clones de ADNc**

[0226] Alícuotas de ambas bibliotecas fueron mezcladas y colocadas sobre placas LB de 25 x 25 cm suplementadas con 50 µg de canamicina por ml. Las colonias individuales fueron seleccionadas sobre placas de 96 pocillos con 100 µl de LB suplementado con 50 µg de canamicina por ml con la ayuda de un Genetix QPix Robot (Genetix Inc., Boston, MA). Cuarenta y cinco placas de 96 pocillos fueron obtenidas para un total de 4320 clones individuales. Las placas fueron incubadas durante toda la noche a 37°C con agitación a 200 r.p.m. Tras la incubación, se añadieron 100 µl de glicerol estéril al 50% a cada pocillo. Los transformantes fueron replicados con la ayuda de una herramienta de 96 espinas (Boekel, Feasterville, PA) en placas de microcultivo de plato profundo de 96 pocillos secundarios (Advanced Genetic Technologies Corporation, Gaithersburg, MD) con 1 ml de Magnificent Broth™ (MacConnell Research, San Diego, CA) suplementado con 50 µg de canamicina por ml en cada pocillo. Las placas de microtitulación primarias fueron almacenadas congeladas a -80°C. Las placas de plato profundo secundarias fueron incubadas a 37°C durante toda la noche con agitación vigorosa a 300 r.p.m. en un agitador giratorio. Para prevenir el derrame y la contaminación cruzada y para permitir aireación suficiente, cada placa de cultivo secundario fue cubierta con una almohadilla de polipropileno (Advanced Genetic Technologies Corporation, Gaithersburg, MD) y una cobertura plástica de plato de microtitulación. El ADN plásmido fue preparado con un MWG Robot-Smart 384 (MWG Biotech Inc., High Point, NC) y Montage Plasmid Miniprep Kits (Millipore, Billerica, MA).

[0227] Las reacciones de secuenciación fueron realizadas usando química de terminador Big-Dye™ (Applied Biosystems, Inc., Foster City, CA) (Giesecke et al., 1992, Journal of Virology Methods 38: 47-60) y un cebador de secuenciación M13 Forward (-20) mostrado más abajo.  
5'-GTAAACGACGGCCAG-3' (SEQ ID N.º: 5)

[0228] Las reacciones de secuenciación fueron realizadas en un formato de 384 pocillos con un Robot-Smart 384 (MWG Biotech Inc., High Point, NC) y la eliminación de terminador con Millipore MultiScreen Seq384 Sequencing Clean-up Kits (Millipore, Billerica, MA). Las reacciones contuvieron 6 µl de ADN plásmido y 4 µl de mezcla maestra de secuenciación que contiene 2 µl de 5x de tampón de secuenciación (Millipore, Billerica, MA), 1 µl de terminador Big-Dye™ (Applied Biosystems, Inc., Foster City, CA), 1,6 pmoles de cebador M13 Forward y 1 µl de agua. La secuenciación del ADN de paso simple fue realizada con un Secuenciador de ADN automatizado ABI PRISM Modelo 3700 (Applied Biosystems, Foster City, CA).

## **Ejemplo 3: Análisis de datos de secuencia de ADN de clones de ADNc**

[0229] La lectura automática de nucleótidos, la atribución de valor de calidad y el ajuste de vector fueron realizados con la asistencia del software PHRED/PHRAP (University of Washington, Seattle, WA). El análisis de agrupamiento del ESTs fue realizado con un Parcel Transcript Assembler v. 2.6.2. (Paracel, Inc., Pasadena, CA). El análisis del agrupamiento EST indicó 395 clusters independientes.

[0230] El análisis de homología secuencial de las secuencias EST ensambladas contra la base de datos PIR fue realizado con el programa Blastx (Altschul et. al., 1990, J. Mol. Biol. 215: 403-410) en un cluster Linux de 32 nodos (Paracel, Inc., Pasadena, CA) usando la matriz BLOSUM 62 (Henikoff, 1992, Proc. Natl. Acad. Sci. USA 89: 10915-10919). De los 395 clusters, 246 tuvieron hits de blast para genes conocidos en bases de datos de proteínas de dominio



público y 149 no tuvieron hits significativos frente a estas bases de datos. Entre estos 246 genes, 13 tuvieron hits frente a homólogos caracterizados de genes de glicosil hidrolasa.

#### Ejemplo 4: Identificación de clones de ADNc que codifican una endoglucanasa de Familia 7 (CEL7F)

[0231] Un clon de ADNc que codifica una endoglucanasa de Familia 7 (CEL7F) fue inicialmente identificado por su identidad a la proteína de endoglucanasa EG-1 de Familia 7 de *Trichoderma longibrachiatum* (NREF NF00756647). Este análisis indicó que las dos proteínas fueron 44% idénticas en el nivel de proteína sobre una extensión de 113 aminoácidos (339 pares de bases). Después de que este clon de identificación inicial Tter08C4 fuera recuperado de la placa de materia prima congelada original y mantenido sobre una placa LB suplementada con 50 µg de canamicina por ml. La placa fue incubada durante toda la noche a 37°C y al día siguiente una colonia individual de la placa fue usada para inocular 3 ml de LB suplementado con 50 µg de canamicina por ml. El cultivo líquido fue incubado durante toda la noche a 37°C y el ADN plásmido fue preparado con un BioRobot 9600 (QIAGEN, Inc., Valencia, CA). El ADN plásmido de Clon Tter08C4 fue secuenciado otra vez con química de terminador Big-Dye™ como se ha descrito anteriormente, usando el M13 forward y un cebador Poly-T mostrado más abajo para secuenciar el extremo 3' del clon.

5'-TTTTTTTTTTTTTTTTTTTTTTTTTVN-3' (SEC ID n.º: 6)

Donde V=G, A, C y N= G, A, C, T

[0232] El análisis de homología Blastx de la información de secuencia indicó que la proteína codificada por el clon Tter08C4 fue similar a la proteína EG1 de *Trichoderma reesei* (NREF NF00494331). Estas proteínas fueron un 46% idénticas sobre una extensión de 365 aminoácidos.

[0233] El análisis de la secuencia de proteína deducida del clon Tter08C4 con el programa Interproscan (Zdobnov y Apweiler, 2001 Bioinformatics 17: 847-8) mostró que el gen contenía la distinción de secuencia de las proteínas de Familia 7. En esta distinción de la secuencia conocida como el modelo Pfam PF00840 (Bateman et. al., 2002, Nucleic Acids Research 30: 276-280) se descubrió 18 aminoácidos desde el aminoácido metionina inicial confirmando que el clon Tter08C4 codifica una endoglucanasa de Familia 7.

[0234] La secuencia de ADNc (SEC ID n.º: 1) y la secuencia de aminoácidos deducida (SEC ID n.º: 2) de la endoglucanasa de *Thielavia terrestris* se muestran en las Figuras 1A y 1B. El clon de ADNc codifica un polipéptido de 336 aminoácidos. El contenido %G+C del clon de ADNc del gen es del 67,5% y de la región codificante de la proteína madura (nucleótidos 55 a 1011 de la SEC ID n.º: 1) es del 67,5%. Usando el programa de software SignalP (Nielsen et al., 1997, Protein Engineering 10: 1-6), se predijo un péptido señal de 17 residuos. La proteína madura predicha contiene 319 aminoácidos con una masa molecular de 33,3 kDa.

[0235] Una alineación comparativa de secuencias de endoglucanasa de Familia 7 fue determinada por el método Clustal W (Higgins, 1989, supra) usando el software LASERGENE™ MEGALIGN™ (DNASTAR, Inc., Madison, WI) con una tabla de identidad y los siguientes parámetros de alineación múltiple: la penalización por gaps de 10 y penalización de longitud de gaps de 10. Los parámetros de alineación en pareja fueron Ktuple=1, penalización por gaps=3, ventanas=5 y diagonales=5. La alineación mostró que la secuencia de aminoácidos deducida del gen maduro de CEL7F de *Thielavia terrestris* comparte un 46,7% de identidad con la secuencia de aminoácidos deducida de la región catalítica del gen de endoglucanasa I de *Trichoderma reesei* (NREF NF00494331, Uniprot Q5BMS5) y el 44,9% de identidad con la secuencia de aminoácidos deducida del gen de endoglucanasa I de longitud completa de *Trichoderma reesei* (NREF NF00494331, Uniprot Q5BMS5). El análisis de la alineación de las regiones catalíticas de estas proteínas demostró que a la endoglucanasa CEL7F de *Thielavia terrestris* le faltan al menos tres motivos de secuencia distintos que se conservan en todos los otros elementos conocidos de la familia 7 de glicosil hidrolasa de los hongos. Dos de estos consisten en más de diez residuos de aminoácidos y contienen residuos altamente conservados que no están presentes en la endoglucanasa CEL7F de *Thielavia terrestris*.

[0236] Una vez que la identidad del clon Tter08C4 fue confirmada, una alícuota de 0,5 µl de ADN plásmido de este clon, designada pTter7F (Figura 2), fue transferida en un frasco de células TOP10 de *E. coli* (Invitrogen, Carlsbad, CA), fue suavemente mezclada e incubada en hielo durante 10 minutos. Las células luego recibieron un choque térmico a 42°C durante 30 segundos y fueron incubadas nuevamente en hielo durante 2 minutos. Las células fueron resuspendidas en 250 µl de medio SOC y fueron incubadas a 37°C durante 60 minutos con agitación constante (200 r.p.m.). Después del periodo de incubación, dos alícuotas de 30 µl fueron colocadas sobre placas LB suplementadas con 50 µg de canamicina por ml y fueron incubadas durante toda la noche a 37°C. Al día siguiente, una colonia individual fue escogida y sembrada sobre tres crioviales de 1,8 ml con aproximadamente 1,5 ml de agarosa LB suplementada con 50 µg de canamicina por ml. Los frascos fueron sellados con PetriSeal™ (Diversified Biotech, Boston MA) y depositados en Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, 1815 University Street, Peoria, Illinois, 61604, como NRRL B 30802, con fecha de depósito del 11 de abril 11 de 2005.

#### Ejemplo 5: Construcción del vector de expresión pAILo2

[0237] El vector de expresión pAILo1 fue construido modificando pBANe6 (patente U.S. n.º 6461837), 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), secuencia de terminador de amiloglucosidasa de *Aspergillus niger* (terminador

AMG) y gen de acetamidasa de *Aspergillus nidulans* (*amdS*). Todas las fases de mutagénesis fueron verificadas por secuenciación usando química de terminador Big-Dye™ según se describe. La modificación de pBANE6 fue realizada eliminando primero tres sitios de restricción Nco I en las posiciones 2051, 2722 y 3397 bp del marcador de selección *amdS* por mutagénesis dirigida. Todos los cambios fueron diseñados para ser "silenciosos" dejando la secuencia de proteína real del producto genético *amdS* sin variación. La eliminación de estos tres sitios fue realizada simultáneamente con un Kit de mutagénesis dirigida GeneEditor™ *in vitro* (Promega, Madison, WI) según las instrucciones del fabricante usando los siguientes cebadores (el nucleótido subrayado representa la base cambiada):

AMDS3NcoMut (2050): 5'-GTGCCCCATGATACGCCTCCGG-3' (SEC ID N.º: 7)

AMDS2NcoMut (2721): 5'-GAGTCGTATTTCCAAGGCTCCTGACC-3' (SEC ID N.º: 8)

AMDS1NcoMut (3396): 5'-GGAGGCCATGAAGTGGACCAACGG-3' (SEC ID N.º: 9)

[0238] Un plásmido que comprende los tres cambios de secuencia previstos fue después sometido a mutagénesis dirigida, usando un kit de mutagénesis dirigida QuickChange™ (Stratagene, La Jolla, CA), para eliminar el sitio de restricción Nco I al final del terminador AMG en la posición 1643. Los siguientes cebadores (el nucleótido subrayado representa la base cambiada) fueron usados para la mutagénesis:

Cebador superior para mutagenizar la secuencia del terminador AMG:

5'-CACCGTGAAAGCCATGCTCTTTCCTTCGTGTAGAAGACCAGACAG-3' (SEC ID N.º: 10)

Cebador inferior para mutagenizar la secuencia del terminador AMG:

5'-CTGGTCTTCTACACGAAGGAAAGAGCATGGCTTTCACGGTGTCTG-3' (SEC ID N.º: 11)

[0239] La última fase en la modificación de pBANE6 fue la adición de un nuevo sitio de restricción Nco I en el principio del polilígar usando un kit de mutagénesis dirigida QuickChange™ y los siguientes cebadores (los nucleótidos subrayados representan las bases cambiadas) para producir pAILo1 (Figura 3).

Cebador superior para mutagenizar el promotor NA2-tpi:

5'-CTATATACACAACCTGGATTACCATGGGCCCGCGGCCGACATC-3' (SEC ID N.º: 12)

Cebador inferior para mutagenizar el promotor NA2-tpi:

5'-GATCTGCGGCCGCGGGCCCATGGTAAATCCAGTTGTGTATATAG-3' (SEC ID N.º: 13)

[0240] El gen *amdS* de pAILo1 fue cambiado con el gen *pyrG* de *Aspergillus nidulans*. El plásmido pBANE10 (Figura 4) fue usado como una fuente para el gen *pyrG* como un marcador de selección. El análisis de la secuencia de pBANE10 mostró que el marcador *pyrG* estaba contenido dentro de un fragmento de restricción Nsi I y no contiene sitios de restricción Nco I ni Pac I. Puesto que el *amdS* también es flanqueado por sitios de restricción Nsi I, la estrategia para cambiar el marcador de selección fue un simple cambio de los fragmentos de restricción Nsi I. El ADN plásmido de pAILo1 y pBANE10 fue digerido con la enzima de restricción Nsi I y los productos purificados por electroforesis en gel de agarosa. El fragmento Nsi I de pBANE10 conteniendo el gen *pyrG* fue ligado a la estructura principal de pAILo1 para reemplazar el fragmento de ADN Nsi I original conteniendo el gen *amdS*. Los clones recombinantes fueron analizados por digestión de la restricción para determinar que tenían el inserto correcto y también su orientación. Se seleccionó un clon con el gen *pyrG* transcrito en el sentido contrario a las agujas del reloj. El plásmido nuevo fue designado pAILo2 (Figura 5).

#### **Ejemplo 6: Clonación del gen de endoglucanasa de familia CEL7F en un vector de expresión de *Aspergillus oryzae***

[0241] Dos cebadores oligonucleótidos sintéticos, mostrados más abajo, fueron diseñados para amplificar por PCR el marco de lectura abierto en toda su longitud de EST Tter08C4 de *Thielavia terrestris* que codifica una endoglucanasa de Familia CEL7F. Un Kit de clonación In-Fusion (BD Biosciences, Palo Alto, CA) fue usado para clonar el fragmento directamente en pAILo2.

Cebador directo In-Fusion:

5'-ACTGGATTACCATGACCTACGGCTCCCTGTCATCA-3' (SEC ID N.º: 14)

Cebador inverso In-Fusion:

5'-TCACCTCTAGTTAATTAAGTAGTTCTTCGTGGTAGACC-3' (SEC ID N.º: 15)

Las letras en negrita representan la secuencia codificante. La secuencia restante contiene la identidad de secuencia comparada con los sitios de inserción de pAILo2.

[0242] Cincuenta picomoles de cada uno de los cebadores mencionados fueron usados en una reacción de PCR que contiene 50 ng de ADN pTter11C9, 1X de tampón de amplificación Pfx (Invitrogen, Carlsbad, CA), 6 µl de 10 mM de mezcla de dATP, dTTP, dGTP y dCTP, 2,5 unidades de *Pfx* ADN polimerasa Platinum (Invitrogen, Carlsbad, CA), 1 µl de 50 mM de MgSO<sub>4</sub> y 5 µl de 10X de solución intensificadora pCRx (Invitrogen, Carlsbad, CA) en un volumen final de 50 µl. Un Eppendorf Mastercycler 5333 (Eppendorf Scientific, Inc., Westbury, NY) fue usado para amplificar el fragmento programado para un ciclo a 98°C durante 2 minutos; y 35 ciclos cada uno a 94°C durante 30 segundos, 65°C durante 30 segundos y 68°C durante 1,5 minutos. Después del 35 ciclos, la reacción fue incubada a 68°C durante 10 minutos y luego enfriada a 10°C hasta ser procesada adicionalmente. Un producto de reacción de PCR de 1,4 kb fue aislado en un gel de agarosa GTG al 0,8% (Cambrex Bioproducts One Meadowlands Plaza East Rutherford, New Jersey 07073) usando 40 mM de base tris -20 mM de acetato de sodio -1 mM de tampón de disodio EDTA (TAE) y 0,1 µg de bromuro de etidio por ml. La banda de ADN fue visualizada con la ayuda de un Dark Reader™ (Clare Chemical Research, Dolores, CO) para evitar mutaciones inducidas por UV. La banda de ADN de 1,4 kb fue cortada con una hoja de afeitar

desechable y purificada con una unidad centrífuga Ultrafree-DA (Millipore, Billerica, MA) según las instrucciones del fabricante.

[0243] El vector pAIIo2 fue linealizado por digestión con Nco I y Pac I. El fragmento fue purificado por electroforesis en gel y ultrafiltración como se ha descrito anteriormente. La clonación del fragmento de PCR purificado en el vector pAIIo2 linealizado y purificado fue realizado con un kit de clonación In-Fusion (BD Biosciences, Palo Alto, CA). La reacción (20 µl) contuvo 1X de tampón In-Fusion (BD Biosciences, Palo Alto, CA), 1X de BSA (BD Biosciences, Palo Alto, CA), 1 µl de enzima In-Fusion (diluida 1:10) (BD Biosciences, Palo Alto, CA), 100 ng de pAIIo2 digerido con Nco I y Pac I y 50 ng del producto de PCR purificado de CEL7F de *Thielavia terrestris*. La reacción fue incubada a la temperatura ambiente durante 30 minutos. Se utilizó una muestra de 2 µl de la reacción para transformar las células XL10 SoloPac® Gold de *E. coli* (Stratagene, La Jolla, CA) según las instrucciones del fabricante. Después del periodo de recuperación, dos alícuotas de 100 µl de la reacción de transformación fueron colocadas sobre placas 2X YT de 150 mM suplementadas con 100 µg de ampicilina por ml. Las placas fueron incubadas durante toda la noche a 37°C. Un set de ocho clones recombinantes putativos fue seleccionado al azar de las placas de selección y el ADN plásmido fue obtenido a partir de cada una usando un BioRobot 9600. Los clones fueron analizados por digestión de restricción *Xho* I. Dos clones que tienen el modelo de digestión de restricción previsto luego fueron secuenciados para confirmar que no había mutaciones en el inserto clonado. El clon n.º 1 fue seleccionado y designado pAIIo22 (Figura 6).

#### **Ejemplo 7: Expresión del gen de endoglucanasa de Familia CEL7F de *Thielavia terrestris* en *Aspergillus oryzae* JAL250**

[0244] Los protoplastos de *Aspergillus oryzae* JAL250 (WO 99/61651) fueron preparados según el método de Christensen et al., 1988, Bio/Technology 6: 1419-1422. Cinco microgramos de pAIIo22 (y pAIIo2 como un control del vector) fueron usados para transformar protoplastos de *Aspergillus oryzae* JAL250.

[0245] La transformación de *Aspergillus oryzae* JAL250 con pAIIo2 produce aproximadamente 50 transformantes. Ocho transformantes fueron aislados en placas PDA individuales e incubados durante cinco días a 34°C.

[0246] Placas de esporas confluentes fueron lavadas con 5 ml de Tween 80 al 0,01% y la suspensión de esporas fue usada para inocular 25 ml de medio MDU2BP en frascos de agitación de vidrio de 125 ml. Los cultivos transformantes fueron incubados a 34°C con agitación constante a 200 r.p.m. El día cinco después de la inoculación, los cultivos fueron centrifugados a 6000 x g y sus sobrenadantes recogidos. Cinco µl de cada sobrenadante fueron mezclados con un volumen igual de 2X tampón de carga (10% de β-mercaptoetanol) y cargados sobre un 1,5 mm de gel SDS-PAGE de Tris-glicina al 8 %-16 % y manchados con Simply Blue SafeStain (Invitrogen, Carlsbad, CA). Los perfiles de SDS-PAGE de los caldos de cultivo mostraron que seis de ocho transformantes tienen una banda proteínica nueva de aproximadamente 40 kDa. El transformante número 7 fue seleccionado para estudios adicionales y fue designado JAL250AIIo22 de *Aspergillus oryzae*.

#### **Ejemplo 8: Grandes cultivos en frasco de agitación de *Aspergillus oryzae* JAL250AIIo22**

[0247] Las esporas JAL250AIIo22 de *Aspergillus oryzae* fueron extendidas sobre una placa PDA e incubadas durante cinco días a 34°C. La placa de esporas confluentes fue lavada dos veces con 5 ml de Tween 80 al 0,01% para maximizar el número de esporas recogidas. La suspensión de esporas luego fue usada para inocular 500 ml de medio MDU2BP en un matraz Fernbach de dos litros. El cultivo de transformante fue incubado a 34°C con agitación constante (200 r.p.m.). El día cinco después de la inoculación, el caldo de cultivo fue recogido por filtración en una unidad de filtro de nilón de 75 mm de 500 ml con un tamaño de poro de 0,45 µm con un prefiltro de fibra de vidrio. Una muestra de 5 µl del caldo fue analizada por SDS-PAGE como se ha descrito anteriormente para confirmar que el modelo de proteína fue el mismo que el obtenido antes. Una vez que el caldo mostró que contenía la banda proteínica de 40 kDa, el caldo fue sometido a caracterización enzimática.

#### **Ejemplo 9: Caracterización de la endoglucanasa CEL7F de *Thielavia terrestris***

[0248] El caldo de *Aspergillus oryzae* JAL250AIIo22 descrito en el ejemplo 8 fue filtrado a través de un filtro de tamaño de poro de 0,22 µm (Millipore, Billerica, MA), concentrado usando una célula Amicon agitada equipada con una membrana PM10 (Millipore, Billerica, MA) y desalado usando una columna Econo-Pac 10DG (BioRad Laboratories, Hercules, CA).

[0249] El caldo de *Aspergillus oryzae* JAL250 (vector solo), como un control negativo, fue tratado del mismo modo que anteriormente.

[0250] Los sustratos teñidos usados para evaluar la especificidad de sustrato de la endoglucanasa CEL7F de *Thielavia terrestris* incluyeron: AZCL-arabinoxilano (trigo), AZCL-β-glucano, AZCL-dextrano, AZCL-HE-celulosa, AZCL-galactano de patata, AZCL-galactomanano (Carob), AZCL-xilano (Birchwood), AZCL-xiloglucano (Megazyme, Bray, Irlanda) y Chitin Azure (Sigma, St Louis, MO).

- [0251] Los ensayos de actividad fueron realizados en placas de 96 pocillos profundos (Axygen Scientific, Union City, CA) selladas por un sellador de placa (ALPS-300, Abgene, Epsom, UK). Ochocientos  $\mu$ l de los sustratos mencionados (6,25 g por litro de 50 mM de acetato de sodio pH 5,0) fueron transferidos en cada pocillo de la placa de 96 pocillos profundos, seguidos de 180  $\mu$ l de 50 mM de acetato de sodio pH 5,0 y 20  $\mu$ l de solución de endoglucanasa CEL7F de *Thielavia terrestris* (0,25 g/l) para iniciar las reacciones. La concentración de sustrato y carga enzimática en la mezcla reactiva final fue 5 g por litro y 1 mg de enzima por g de sustrato, respectivamente. El caldo de *Aspergillus oryzae* Jal250 fue evaluado junto con el caldo de endoglucanasa CEL7F de *Thielavia terrestris* bajo la misma condición, sirviendo como un control negativo. Las reacciones fueron incubadas a 50°C sin mezcla. Antes del muestreo, las placas de pocillos profundos fueron centrifugadas en un centrifugador de placas (Sorvall RT7, Global Medical Instrumentation, Ramsey, MN) a 3000 r.p.m. durante 5 minutos. Una muestra de 150  $\mu$ l del sobrenadante fue transferida en una placa de filtración de 96 pocillos (tamaño de poro de 0,45  $\mu$ m, Millipore, Billerica, MA), fue aspirada y se recogió el filtrado. Una muestra de 100  $\mu$ l del filtrado fue transferida a otra placa de 96 pocillos y se midió la absorbencia a 590 nm usando un Spectra MAX340 (Molecular Devices, Sunnyvale, CA).
- [0252] Después de incubaciones de 1 hora y 92 horas, el colorante liberado de los diferentes sustratos teñidos por endoglucanasa CEL7F de *Thielavia terrestris* (después de sustraer el colorante liberado por *Aspergillus oryzae* Jal250) se muestra en la tabla 1 como valores de 590 nm relativos.

Tabla 1.  $A_{590nm}$  relativa después de la incubación de endoglucanasa de *Thielavia terrestris* con diferentes sustratos teñidos durante 1 hora y 92 horas a 50°C, pH 5,0.

Tiempo hr	AX	$\beta$ G	Dex	HEC	Gal	GM	Xly	XG	Chitin Azure
1	0,00	0,22	0,00	0,16	0,00	0,00	0,01	0,01	0,00
92	0,58	0,78	0,00	1,00	0,00	0,08	0,32	0,75	0,00

AX: AZCL-arabinosilano  $\beta$ G: AZCL- $\beta$ -glucano  
Dex: AZCL-dextrano HEC: AZCL-HE-celulosa  
Gal: AZCL-galactano de patata GM: AZCL-galactomanano  
Xly: AZCL-xilano XG: AZCL-xiloglucano

- [0253] La endoglucanasa CEL7F de *Thielavia terrestris* tuvo actividad hacia AZCL- $\beta$ -glucano y AZCL-HE-celulosa después de 1 hora. La endoglucanasa CEL7F de *Thielavia terrestris* también mostró actividad en sustratos teñidos de arabinosilano, xilano y xiloglucano y actividad baja en el sustrato teñido de galactomanano después de una incubación de 92 horas.

#### Ejemplo 10: Hidrólisis de rastrojo de maíz pretratado con la endoglucanasa CEL7F de *Thielavia terrestris*.

- [0254] El rastrojo de maíz fue pretratado en el Department of Energy National Renewable Energy Laboratory (NREL) de Estados Unidos usando ácido sulfúrico diluido. Se utilizaron las siguientes condiciones para la pretratamiento: 0.048 g de ácido sulfúrico / g de biomasa seca a 190°C y 25% p/p de sólidos secos durante alrededor de 1 minuto. Los sólidos insolubles en agua en el rastrojo de maíz pretratado (PCS) contenían 52% de celulosa, 3,6% de hemicelulosa y 29,8% de lignina. La celulosa y la hemicelulosa fueron determinadas por una hidrólisis de ácido sulfúrico de dos etapas con análisis posterior de azúcares por cromatografía en fase líquida de alto rendimiento usando el procedimiento analítico estándar n.º 002 de NREL. La lignina fue determinada gravimétricamente después de hidrolizar la celulosa y fracciones de hemicelulosa con ácido sulfúrico usando el procedimiento analítico estándar n.º 003 de NREL. Antes de la hidrólisis enzimática, el PCS fue lavado con un gran volumen de agua destilada doble hasta que el pH fue superior a 4,0, y luego fue tamizado a través de una criba de malla 100 y sometido a autoclave a 121 °C durante 30 minutos.
- [0255] La hidrólisis de PCS fue conducida en placas de 96 pocillos profundos, (Axygen Scientific, Union City, CA) sellada por un sellador de placa (ALPS-300, Abgene, Epsom, Reino Unido), con un volumen de reacción total de 1,0 ml. La hidrólisis de PCS (10 mg/ml en 50 mM de tampón de acetato de sodio pH 5,0) fue realizada usando 1,25 mg de endoglucanasa CEL7F de *Thielavia terrestris* (preparada como se describe en el ejemplo 9) por gramo de PCS. El caldo de *Aspergillus oryzae* Jal250 (preparado como se describe en el ejemplo 9) fue realizado como un control. Se realizó la hidrólisis de PCS a 50°C, pH 5,0. Las reacciones fueron realizadas en duplicados y se tomaron alícuotas durante el curso de la hidrólisis. Las reacciones de la hidrólisis de PCS fueron detenidas mediante la mezcla de una alícuota de 20  $\mu$ l de cada hidrolizado con 180  $\mu$ l de 0,11 M de NaOH (reactivo de detención). Diluciones en serie apropiadas fueron generadas para cada muestra y el contenido de azúcar reductor determinado usando un ensayo de hidracida de ácido para-hidroxibenzoico (PHBAH, Sigma, St. Louis, MO) adaptado a un formato de microplaca de 96 pocillos como se describe más abajo. Brevemente, una alícuota de 90  $\mu$ l de una muestra apropiadamente diluida fue colocada en una microplaca de fondo cónico de 96 pocillos. Las reacciones fueron iniciadas añadiendo 60  $\mu$ l de 1,5% (p/v) de PHBAH en NaOH al 2% a cada pocillo. Las placas fueron calentadas descubiertas a 95°C durante 10 minutos. Se permitió que las placas se enfriaran a temperatura ambiente (RT) y se añadieron 50  $\mu$ l de H<sub>2</sub>O destilado a cada pocillo. Una alícuota de 100  $\mu$ l de cada pocillo fue transferida a una placa de 96 pocillos de fondo plano y se midió la absorbencia en A<sub>410nm</sub> usando un SpectraMax Microplate Reader (Molecular Devices, Sunnyvale, CA). Se utilizaron los estándares de glucosa (0,1-0,0125 mg/ml diluidos con hidróxido de sodio al 0,4%) para preparar una curva estándar para traducir los valores de A<sub>410nm</sub> obtenidos en equivalentes de glucosa. Los equivalentes resultantes fueron usados para calcular el porcentaje

de conversión de celulosa PCS para cada reacción. El grado de conversión de celulosa a azúcar reductor (conversión, %) fue calculado usando la siguiente ecuación:

$$\text{Conversión (\%)} = \text{RS (mg/ml)} * 100 * 162 / (\text{Celulosa (mg/ml)} * 180) \\ = \text{RS (mg/ml)} * 100 / (\text{Celulosa (mg/ml)} * 1,111)$$

En esta ecuación, RS es la concentración de azúcar reductor en la solución medida en equivalentes de glucosa (mg/ml) y el factor 1.111 refleja el aumento de peso en la conversión de celulosa a glucosa.

[0256] La hidrólisis de PCS por la endoglucanasa CEL7F de *Thielavia terrestris* (1,25 mg/g de PCS) produjo una conversión de celulosa de 2,1% después de 120 horas. *Aspergillus oryzae* Jal250 (1,25 mg/g de PCS) produjo menos del 1% de conversión después de 120 horas.

#### Ejemplo 11: Hidrólisis de betaglucono soluble de cebada por endoglucanasa de *Thielavia terrestris*

[0257] La endoglucanasa de *Thielavia terrestris* Cel7F fue evaluada en forma del caldo de *Aspergillus oryzae* Jal250AIL022 descrito en el ejemplo 8. El caldo fue concentrado y cambiado a 50 mM de acetato de sodio pH 5,0 usando el filtro centrífugo Centricon Plus-20 con membrana de polietersulfona Biomax-5 (5000 NMWL) de Millipore (Bedford, MA). El caldo de *Aspergillus oryzae* Jal250 (vector solo) fue tratado igual que más arriba.

[0258] La concentración de proteínas en las soluciones enzimáticas fue determinada usando el ensayo de microplaca de ácido bicinonínico (BCA) según las instrucciones del fabricante para un kit de reactivo de ensayo de proteína de BCA (Pierce Chemical Co., Rockford, IL).

[0259] Las diluciones enzimáticas fueron recién preparadas antes cada experimento de soluciones de enzima de materia prima, que fueron almacenadas a -20°C.

[0260] La actividad de la endoglucanasa de *Thielavia terrestris* Cel7F en beta-glucano soluble de cebada (viscosidad media, 230 kDa, Megazyme International Ireland Ltd., Bray, Irlanda) fue determinada a pH 5,5 (50 mM de acetato de sodio con azida sódica al 0,02%) y 60°C. Los resultados fueron comparados con los de la endoglucanasa *Trichoderma reesei* Cel7Bde (EGI). La endoglucanasa de *Trichoderma reesei* Cel7B recombinante (EGI) puede ser preparada según Takashima et al., 1998, Journal of Biotechnology 65: 163-171.

[0261] La concentración inicial de beta-glucano en las reacciones de hidrólisis fue 1,0% (p/v). Se realizaron reacciones de un ml sin agitar en placas de 96 pocillos profundos de Eppendorf (1,2 ml, VWR Scientific, West Chester, PA). Las enzimas fueron usadas en tres cargas de proteínas, 0,05, 0,1 y 0,2 mg por g de glucano. En las reacciones de control, las endoglucanasas fueron sustituidas con 50 mM de acetato de sodio pH 5,5 con azida sódica al 0,02% (control de tampón) o con caldo de *Aspergillus oryzae* Jal250 concentrado y con tampón cambiado con enzimas no expresadas recombinantemente (control de Jal250).

[0262] Las alícuotas fueron eliminadas de las reacciones de hidrólisis en 2 horas y 24 horas, diluidas con agua desionizada y analizadas para detectar azúcares reducidos usando el ensayo de hidracida de ácido p-hidroxibenzoico (PHBAH) como se describe en el ejemplo 10. La conversión relativa del beta-glucano como función de carga de proteína en dos tiempos de incubación, 2 horas y 24 horas, se muestra en las Figuras 7 y 8, respectivamente. La conversión relativa se muestra como un porcentaje de conversión obtenido después de hidrólisis de 24 horas de beta-glucano por endoglucanasa de *Thielavia terrestris* Cel7F (0,2 mg de proteína por g de glucano).

[0263] La endoglucanasa de *Thielavia terrestris* Cel7F mostró mayor conversión de beta-glucano que la endoglucanasa de *Trichoderma reesei* Cel7B y continuó produciendo nuevos grupos terminales reducidos más allá del período de incubación de 2 horas. Por el contrario, la endoglucanasa de *Trichoderma reesei* Cel7B no mostró casi ningún aumento adicional en la concentración de azúcar reductor después de 2 horas de hidrólisis.

#### Depósito de material biológico

[0264] El siguiente material biológico ha sido depositado según las condiciones del Tratado de Budapest con Agricultural Research Service Patent Culture Collection, Northern Regional Research Center, 1815 University Street, Peoria, Illinois, 61604 y recibió el siguiente número de acceso:

Depósito	Número de acceso	Fecha de depósito
<i>E. coli</i> pTter7F	NRRL B-30837	11 de abril de 2005

[0265] La cepa ha sido depositada bajo condiciones que aseguran que el acceso al cultivo estará disponible durante la pendencia de esta solicitud de patente a una determinada por el Comisario de Patentes y marcas registradas para tener derecho a ellas 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 lo requerido por las leyes extranjeras de patentes en países donde se presentan duplicados de la aplicación de referencia o su progenie. No obstante, debe entenderse que la disponibilidad de un depósito no constituye una licencia para poner en práctica la invención de referencia en derogación de derechos de patentes concedidos por acción gubernativa.

## LISTADO DE SECUENCIAS

[0266]

&lt;110&gt; Novozymes, Inc.

&lt;120&gt; Polipéptidos con actividad endoglucanasa y polinucleótidos que codifican los mismos

&lt;130&gt; 10802.204-WO

&lt;150&gt; 60/675,601

&lt;151&gt; 2005-04-27

&lt;160&gt; 15

&lt;170&gt; PatentIn versión 3.2

&lt;210&gt; 1

&lt;211&gt; 1011

&lt;212&gt; ADN

&lt;213&gt; Thielavia terrestris

&lt;400&gt; 1

```

atgacctac ggcctccctgt catcagcctg ctggcctcgc tggcagcagg cgcgctcgtc      60
gtcccacggg cggagtttca ccccccctctc ccgacttggg aatgcacgac ctccggggggc      120
tgcgctgcagc agaacaccag cgtcgtctctg gacgctgact cgaagtacgc cgcacacagc      180
gccggctcgc ggaagggaatc ggattacgcg gcaatgggag tgtccacttc gggcaatgcc      240
gtgacgctgt accactacgt caagaccacc ggcaccctcg tcccgccttc gcgcgcgcatc      300
tacctctctg ggcgggacgg caagtacgtg cttatggacc tctcaacca ggagctgtcg      360
gtggacgtcg acttctcggc gctgcgctgc ggcgagaacg gggccttcta cctgtccgag      420
atggcgggcg acggcgggcg cgcgcggggg ggcggcgacg ggtactgcga cgcgcagtcg      480
cagggctact gctgcacaga gatggacatc ctcgaggcca aotcgatggc gacggccatg      540
acgcgcgacc cgtgcacagg caacaaactgc gacgcgagcg gctgcggcta caaccgctac      600
gcagcgggcc agcgcggtt ctacggggcc ggcaagacgg tcgacacgag caagcccttc      660
accgtcgtca cgcagttcgc cgcgcggggc ggcaagctga ccagatcac ccgcaagtac      720
atccagaacg gccgggagat cggcgggggc ggcaccatct ccagctgagg ctccgagttc      780
tcgacgggcg gctgacggc catggggcag gcgctggggc gcggaatggg gctggccatg      840
agcatctgga acgacgggc ccaggagatg gcatggctcg atgcgggcaa caacggccct      900
tgcgccagtg gccagggcag ccgctccgtc attcagtcgc agcatccga caccacgtc      960
gttttttcca acatcaggtg gggcgacatc gggctctacca cgaagaacta g      1011

```

&lt;210&gt; 2

&lt;211&gt; 336

&lt;212&gt; PRT

&lt;213&gt; Thielavia terrestris

&lt;400&gt; 2

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

Gln Phe Ala Ala Ser Gly Gly Lys Leu Thr Gln Ile Thr Arg Lys Tyr  
225 230 235 240

Ile Gln Asn Gly Arg Glu Ile Gly Gly Gly Gly Thr Ile Ser Ser Cys  
245 250 255

Gly Ser Glu Ser Ser Thr Gly Gly Leu Thr Gly Met Gly Glu Ala Leu  
260 265 270

Gly Arg Gly Met Val Leu Ala Met Ser Ile Trp Asn Asp Ala Ala Gln  
275 280 285

Glu Met Ala Trp Leu Asp Ala Gly Asn Asn Gly Pro Cys Ala Ser Gly  
290 295 300

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

Val Phe Ser Asn Ile Arg Trp Gly Asp Ile Gly Ser Thr Thr Lys Asn  
325 330 335

5 <210> 3  
<211> 32  
<212> ADN  
<213> Escherichia coli  
  
10 <400> 3  
tcgtcgggga caactttgta caaaaaagtt gg 32  
  
15 <210> 4  
<211> 27  
<212> ADN  
<213> Escherichia coli  
  
20 <400> 4  
cccctgttga aacatgtttt ttcaacc 27  
<210> 5  
<211> 16  
<212> ADN  
<213> Escherichia coli  
  
25 <400> 5  
gtaaaacgac ggccag 16  
  
30 <210> 6  
<211> 25  
<212> ADN  
<213> Escherichia coli  
  
35 <220>  
<221> misc\_feature  
<222> (25)..(25)  
<223> N = A, C, G o T  
  
<400> 6  
ttttttttt ttttttttt tttvn 25



<210> 7  
 <211> 22  
 <212> ADN  
 <213> Aspergillus nidulans  
 5  
 <400> 7  
 gtgccccatg atacgcctcc gg 22  
 <210> 8  
 10 <211> 26  
 <212> ADN  
 <213> Aspergillus nidulans  
 <400> 8  
 15 gagtcgtatt tccaaggctc ctgacc 26  
 <210> 9  
 <211> 24  
 <212> ADN  
 20 <213> Aspergillus nidulans  
 <400> 9  
 ggaggccatg aagtggacca acgg 24  
 25 <210> 10  
 <211> 45  
 <212> ADN  
 <213> Aspergillus niger  
 30 <400> 10  
 caccgtgaaa gccatgctct ttccttcgtg tagaagacca gacag 45  
 <210> 11  
 <211> 45  
 35 <212> ADN  
 <213> Aspergillus niger  
 <400> 11  
 ctggtctct acacgaagga aagagcatgg ctttcacggt gtctg 45  
 40 <210> 12  
 <211> 44  
 <212> > ADN  
 <213> Aspergillus niger  
 45 <400> 12  
 ctatatacac aactggatt accatgggcc cgcggccgca gatc 44  
 <210> 13  
 50 <211> 44  
 <212> ADN  
 <213> Aspergillus niger  
 <400> 13  
 55 gatctgcggc cgcgggccca tggtaaacc agttgtgtat atag 44  
 <210> 14  
 <211> 36  
 <212> ADN  
 60 <213> Thielavia terrestris  
 <400> 14  
 actggattac catgacccta cggctccctg tcatca 36  
 65 <210> 15  
 <211> 38

## ES 2 336 026 T5

<212> ADN

<213> Thielavia terrestris

<400> 15

5    tcacctctag ttaattaact agttcttcgt ggtagacc    38

# REIVINDICACIONES

1. Polipéptido aislado que tiene actividad endoglucanasa y que comprende una secuencia de aminoácidos que tiene al menos un 80% de identidad con el polipéptido maduro de la SEC ID n.º: 2.
2. Polipéptido según la reivindicación 1, que comprende o que consiste en el polipéptido maduro de la SEC ID n.º: 2; que tiene actividad endoglucanasa.
3. Polipéptido según la reivindicación 1, que es codificado por el polinucleótido contenido en el plásmido pTter7F que está contenido en *E. coli* NRRL B-30837.
4. Polipéptido según cualquiera de las reivindicaciones 1-3, en donde el polipéptido maduro es los aminoácidos 18 a 336 de la SEC ID n.º: 2 y la secuencia que codifica el polipéptido maduro es los nucleótidos 52 a 1008 de la SEC ID n.º 1.
5. Polinucleótido aislado que comprende una secuencia de nucleótidos que codifica el polipéptido según cualquiera de las reivindicaciones 1-4.
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.
7. Célula huésped recombinante que comprende el constructo de ácidos nucleicos según la reivindicación 6.
8. Método para producir el polipéptido según cualquiera de las reivindicaciones 1-4, que comprende: (a) cultivar 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) recuperar el polipéptido.
9. Planta transgénica, parte de planta o célula vegetal, que ha sido transformada con un polinucleótido que codifica el polipéptido según cualquiera de las reivindicaciones 1-4.
10. Método para degradar biomasa con celulosa y hemicelulosa, que comprende tratar la biomasa con una cantidad eficaz del polipéptido según cualquiera de las reivindicaciones 1-4 o la célula huésped según la reivindicación 7 y recuperar la biomasa degradada.

```

      M T L R L P V I S L L A S L A A G
1  ATGACCCTACGGCTCCCTGTGTCATCAGCCTGCTGGCCTCGCTGGCAGCAGGC
   A V V V P R A E F H P P L P T W K
52 GCCGTGTCGTCCCACGGGCGGAGTTTCACCCCCCTCTCCCGACTTGGA
   C T T S G G C V Q Q N T S V V L D
103 TGCACGACCTCCGGGGGCTGCGTGCAGCAGAACACCAGCGTCGTCTTGGAC
   R D S K Y A A H S A G S R T E S D
154 CGTGACTCGAAGTACGCCGCACACAGCGCCGGCTCGCGGACGGAATCGGAT
   Y A A M G V S T S G N A V T L Y H
205 TACGCGGCAATGGGAGTGTCCACTTCGGGCAATGCCGTGACGCTGTACCAC
   Y V K T N G T L V P A S P R I Y L
256 TACGTCAAGACCAACGGCACCCTCGTCCCCGCTTCGCCGCGCATCTACCTC
   L G A D G K Y V L M D L L N Q E L
307 CTGGGCGCGGACGGCAAGTACGTGCTTATGGACCTCCTCAACCAGGAGCTG
   S V D V D F S A L P C G E N G A F
358 TCGGTGGACGTGCACTTCTCGGCGCTGCCGTGCGGCGAGAACGGGGCCTTC
   Y L S E M A A D G R G D A G A G D
409 TACCTGTCCGAGATGGCGGCGGACGGGCGGGCGACGCGGGGGCGGGCGAC
   G Y C D A Q C Q G Y C C N E M D I
460 GGGTACTGCGACGCGCAGTGCCAGGGCTACTGCTGCAACGAGATGGACATC
   L E A N S M A T A M T P H P C K G
511 CTCGAGGCCAACTCGATGGCGACGGCCATGACGCCCGACCCGTGCAAGGGC
   N N C D R S G C G Y N P Y A S G Q
562 AACAACTGCGACCGCAGCGGCTGCGGCTACAACCCGTACGCCAGCGGGCCAG
   R G F Y G P G K T V D T S K P F T
613 CGCGGCTTCTACGGGCGCGCAAGACGGTCGACACGAGCAAGCCCTTCACC
   V V T Q F A A S G G K L T Q I T R
664 GTCGTACGCGAGTTCGCCGCCAGCGGCGGCAAGCTGACCCAGATCACCCGC
   K Y I Q N G R E I G G G G T I S S
715 AAGTACATCCAGAACGGCCGGGAGATCGGCGGCGGCGGCACCATCTCCAGC
   C G S E S S T G G L T G M G E A L
766 TGCGGCTCCGAGTCTTCGACGGGCGGCTGACCGGCATGGGCGAGGCGCTG
   G R G M V L A M S I W N D A A Q E
817 GGGCGCGGAATGGTGCTGGCCATGAGCATCTGGAACGACGCGGCCAGGAG
   M A W L D A G N N G P C A S G Q G

```

Fig. 1A

```

868  ATGGCATGGCTCGATGCCGGCAACAACGGCCCTTGCGCCAGTGGCCAGGGC
      S P S V I Q S Q H P D T H V V F S
919  AGCCCGTCCGTCATTCAGTCGCAGCATCCCGACACCCACGTCGTCTTCTCC
      N I R W G D I G S T T K N *
970  AACATCAGGTGGGGCGACATCGGGTCTACCACGAAGAAGTAG
    
```

**Fig. 1B**

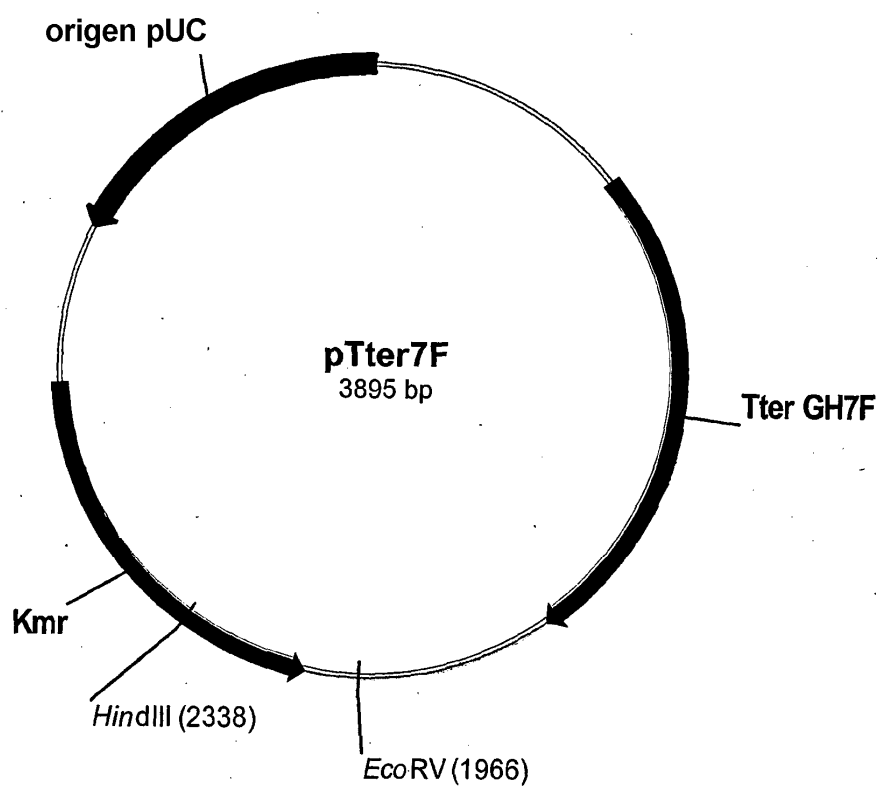


Fig. 2

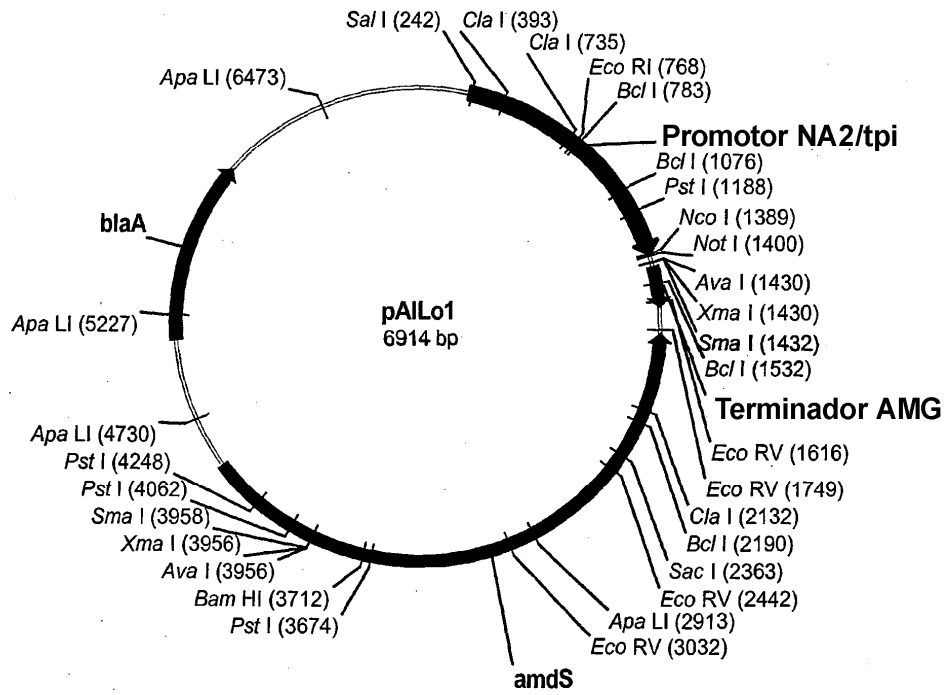


Fig. 3

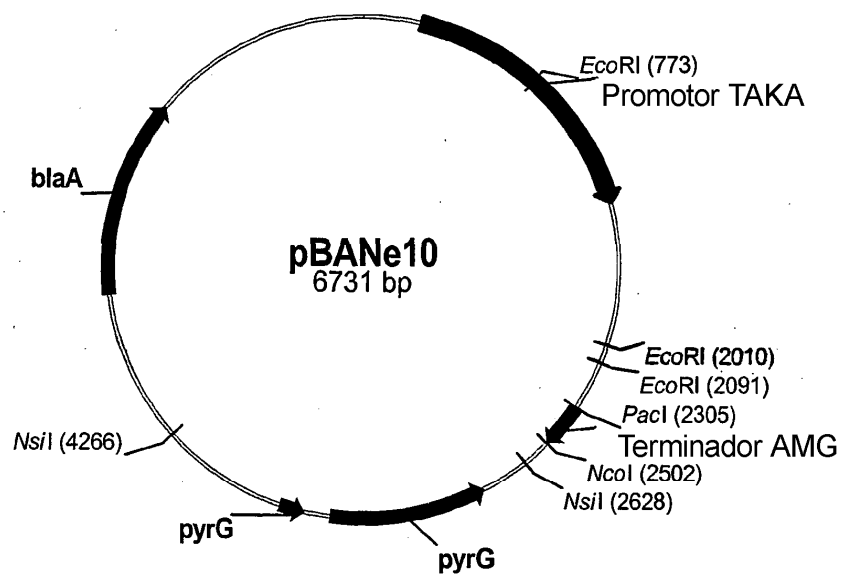


Fig. 4



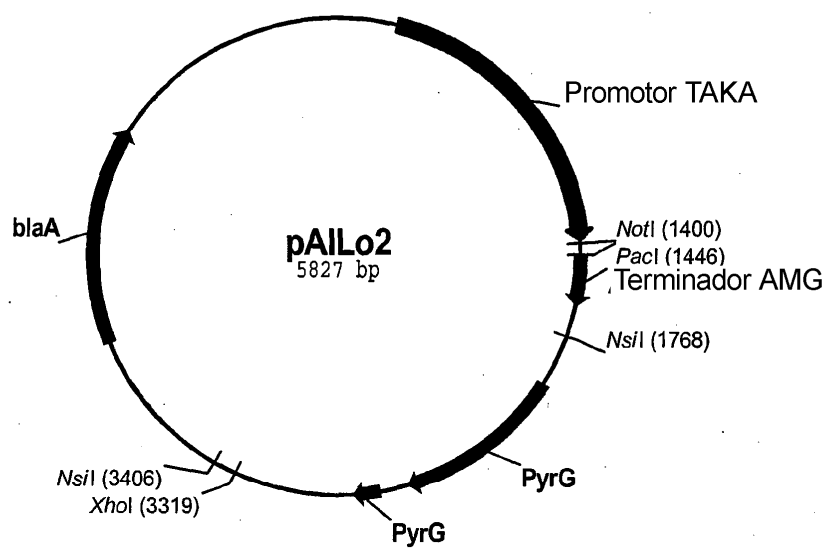


Fig. 5

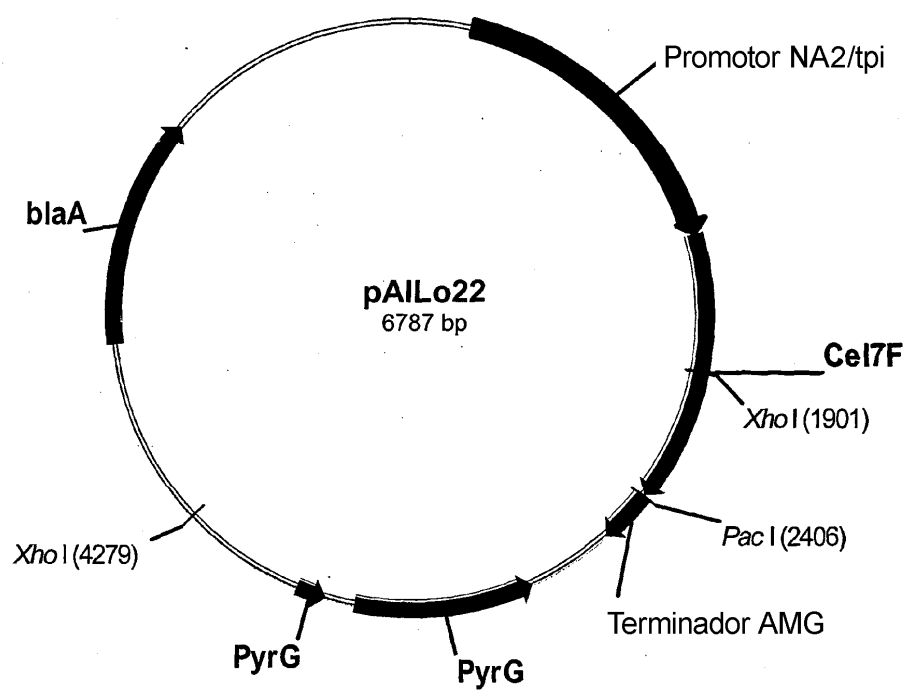


Fig. 6

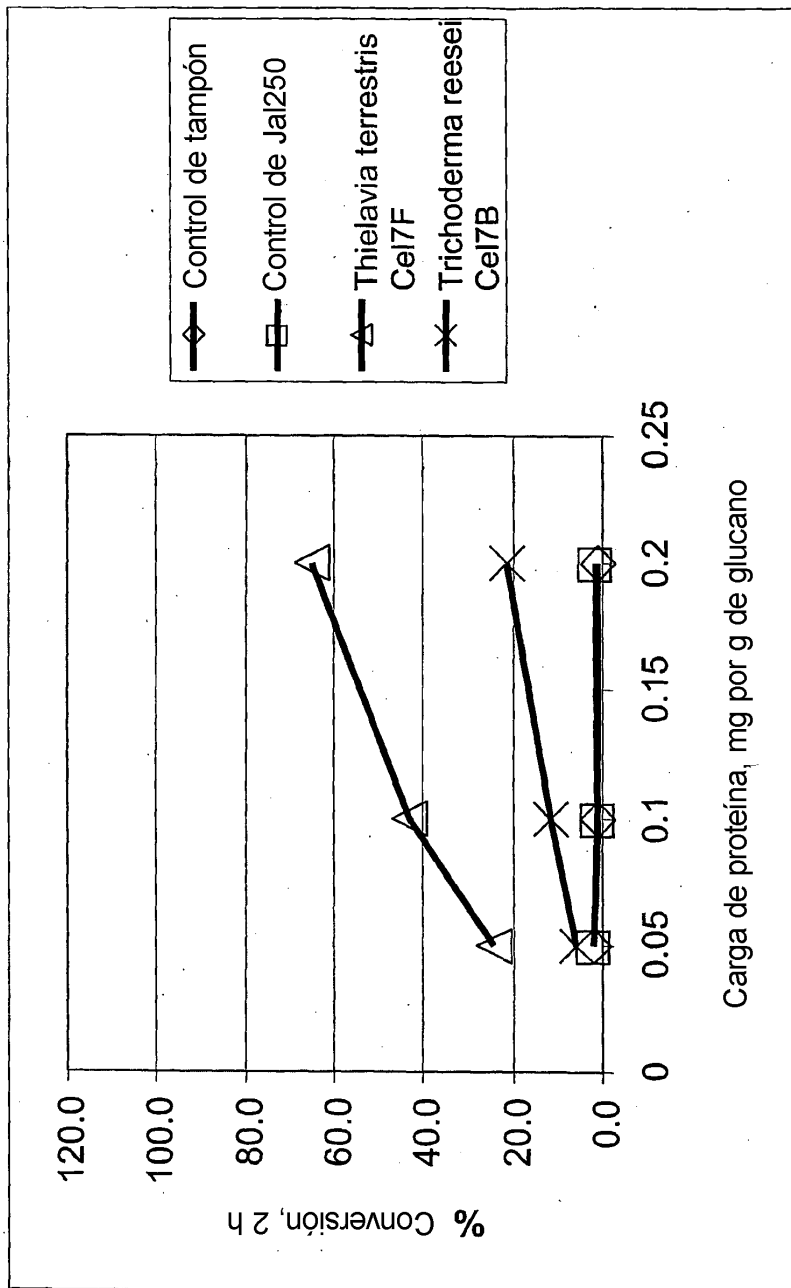


Fig. 7

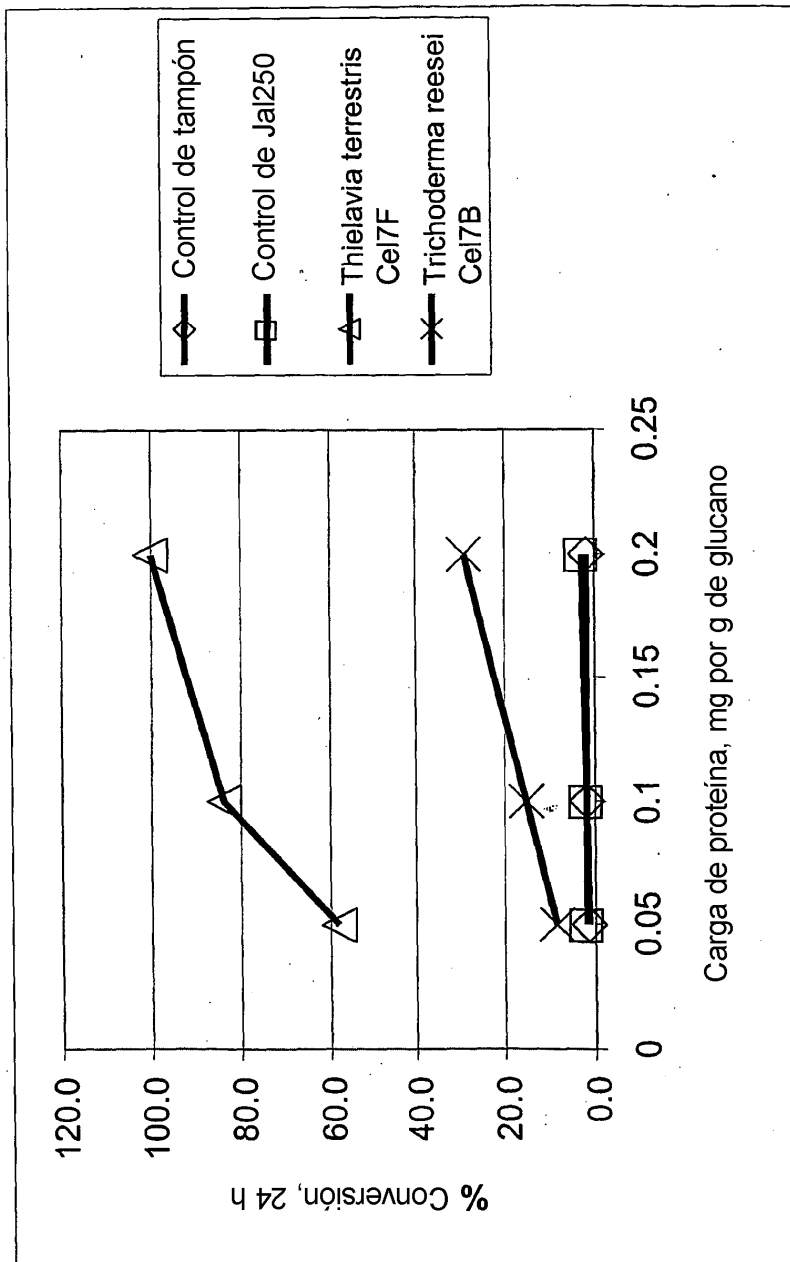


Fig. 8