

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 428 028**

51 Int. Cl.:

C07K 1/00 (2006.01)
C07H 21/04 (2006.01)
C12P 21/06 (2006.01)
C12N 9/00 (2006.01)
C12N 9/24 (2006.01)
C12N 1/20 (2006.01)
C12N 15/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **14.11.2003 E 03786718 (1)**

97 Fecha y número de publicación de la concesión europea: **19.06.2013 EP 1567543**

54 Título: **Beta-glucosidasa BGL7 y ácidos nucleicos que codifican la misma**

30 Prioridad:

21.11.2002 US 301015

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

05.11.2013

73 Titular/es:

**DANISCO US INC. (100.0%)
925 Page Mill Road
Palo Alto, CA 94304, US**

72 Inventor/es:

**DUNN-COLEMAN, NIGEL y
WARD, MICHAEL**

74 Agente/Representante:

RIZZO, Sergio

ES 2 428 028 T3

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

BETA-GLUCOSIDASA BGL7 Y ÁCIDOS NUCLEICOS QUE CODIFICAN LA MISMA

Descripción

Apoyo gubernamental

[0001] Partes de este trabajo fueron financiadas por el subcontrato N°: ZCO-30017-01 con el National
5 Renewable Energy Laboratory con el contrato principal N°: DE-AC36-99GO10337 con el
Departamento de energía de los Estados Unidos. En consecuencia, el gobierno de los Estados
Unidos puede tener determinados derechos sobre esta invención.

Campo de la invención

[0002] La presente invención se refiere a secuencias de ácido nucleico de *bgl7* aislado que codifican
10 polipéptidos que tienen actividad de β -glucosidasa, La invención también hace referencia a células
huésped, vectores y construcciones de ácido nucleico que constan de secuencias de ácido nucleico y
de métodos para la producción de polipéptidos BGL7 recombinantes.

Referencias

[0003]

- 15 Altschul, S. F., *et al. J. Mol. Biol.* 215:403-410, 1990.
Altschul, S. F., *et al., Nucleic Acids Res.* 25:3389-3402, 1997.
Aro, N., *et al., J. Biol. Chem.*, 10.1074/ M003624200, 13 de abril, 2001.
Aubert, *et al.*, Ed., pág. 11 y ss., Academic Press, 1988.
Ausubel G. M., *et al. CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, John Wiley & Sons,
20 Nueva York, N.Y., 1993.
Baldwin, D., *et al., Curr. Opin. Plant Biol.* 2(2):96-103, 1999.
Baulcombe, D., *Arch. Virol. Suppl.* 15:189-201, 1999.
Bhikhabhai, R. *et al., J. Appl. Biochem.* 6:336, 1984.
Brumbauer, A. *et al., Bioseparation* 7:287-295, 1999.
25 Carter *et al., Nucl. Acids Res.* 13:4331, 1986.
Chen *et al., Biochem. Biophys. Acta.* 1121:54-60, 1992.
Coligan, J. E. *et al.*, eds., *CURRENT PROTOCOLS IN IMMUNOLOGY*, 1991.
Collen, A., *et al., Journal of Chromatography A* 910:275-284, 2001.
Coughlan, *et al., BIOCHEMISTRY AND GENETICS OF CELLULOSE DEGRADATION.*
30 Cummings and Fowler, *Curr. Genet.* 29:227-233, 1996.
Dayhoff *et al. en Atlas of Protein Sequence and Structure*, Volumen 5, Suplemento 3, Capítulo
22, pp. 345-352, 1978.
Deutscher, M.P., *Methods Enzymol.* 182:779-80, 1990.
Doolittle, R. F., *OF URFs AND ORFs*, University Science Books, CA, 1986.

- Ellouz, S. *et al.*, *J. Chromatography* 396:307, 1987.
- Fields y Song, *Nature* 340:245-246, 1989.
- Filho, *et al.* *Can. J. Microbiol.* 42:1-5, 1996.
- Fliess, A., *et al.*, *Eur. J. Appl. Microbiol. Biotechnol.* 17:314, 1983.
- 5 Freer, *et al.* *J. Biol. Chem.* 268:9337-9342, 1993.
- Freshney, R. I., ed., *ANIMAL CELL CULTURE*, 1987.
- Goyal, A. *et al.* *Bioresource Technol.* 36:37, 1991.
- Halldorsdottir, S *et al.*, *Appl Microbiol Biotechnol.* 49(3):277-84, 1998.
- Hu *et al.*, *Mol Cell Biol.* 11:5792-9, 1991.
- 10 Hemmpel, W.H. *ITB Dyeing/Printing/Finishing* 3:5-14, 1991.
- Herr *et al.*, *Appl. Microbiol. Biotechnol.* 5:29-36, 1978.
- Jakobovits, A, *et al.*, *Ann N Y Acad Sci* 764:525-35, 1995.
- Jakobovits, A, *Curr Opin Biotechnol* 6(5):561-6, 1995.
- Jones *et al.*, *Nature* 321:522-525, 1986.
- 15 Kawaguchi, T *et al.*, *Gene* 173(2):287-8, 1996.
- Knowles, J. *et al.*, *TIBTECH* 5, 255-261, 1987.
- Kohler y Milstein, *Nature* 256:495, 1975.
- Krishna, S. *et al.*, *Bioresource Tech.* 77:193-196, 2001.
- Kumar, A., *et al.*, *Textile Chemist and Colorist* 29:37-42, 1997.
- 20 Lehtio, J. *et al.*, *FEMS Microbiology Letters* 195:197-204, 2001.
- Li y Ljungdahl *Appl. Environ. Microbiol.* 62:209-213, 1996.
- Linder, M. y Teeri, T.T., *Biotechnol.* 57:15-28, 1997.
- Medve, J. *et al.*, *J. Chromatography A* 808:153, 1998.
- Ohmiya *et al.*, *Biotechnol. Gen. Engineer. Rev.* 14:365-414, 1997.
- 25 Ooi *et al.*, *Nucleic Acids Res.* 18(19):5884, 1990.
- Ortega *et al.*, *International Biodeterioration and Biodegradation* 47:7-14, 2001.
- Penttila *et al.*, *Yeast* 3:175-185, 1987.
- Penttila *et al.*, *Gene* 63: 103-112, 1988.
- Pere, J., *et al.*, *In Proc. Tappi Pulping Conf.*, Nashville, TN, 27-31, pp. 693-696, 1996.
- 30 Riechmann *et al.*, *Nature* 332:323-327, 1988.
- Rothstein *et al.*, *Gene* 55:353-356, 1987.
- Saarilahti *et al.*, *Gene* 90:9-14, 1990.
- Sakamoto *et al.*, *Curr. Genet.* 27:435-439, 1995.
- Saloheimo M, *et al.*, *Gene* 63:11-22, 1988.
- 35 Sambrook *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL* (Segunda Edición), Cold Spring Harbor Press, Plainview, N.Y., 1989.

- Schulein, *Methods Enzymol.*, 160, 25, página 234 y ss, 1988.
- Scopes, *Methods Enzymol.* 90 Pt E:479-90, 1982.
- Spilliaert R, *et al.*, *Eur J Biochem.* 224(3):923-30, 1994.
- Stahlberg, J. *et al.*, *Bio/Technol.* 9:286-290, 1991.
- 5 Strathern *et al.*, eds. (1981) *The Molecular Biology of the Yeast Saccharomyces*.
- Suurnakki, A. *et al.*, *Cellulose* 7:189-209, 2000.
- Te'o, J. *et al.*, *FEMS Microbiology Letters* 190:13-19, 2000.
- Tilbeurgh, H. *et al.*, *FEBS Lett.* 16:215, 1984.
- Timberlake *et al.*, *Cell* 1:29-37, 1981.
- 10 Tomaz, C. y Queiroz, J., *J. Chromatography A* 865:123-128, 1999.
- Tomme, P. *et al.*, *Eur. J. Biochem.* 170:575-581, 1988.
- Tormo, J. *et al.*, *EMBO J.* 15:5739-5751, 1996.
- Tyndall, R.M., *Textile Chemist and Colorist* 24:23-26, 1992.
- Van Rensburg *et al.*, *Yeast* 14:67-76, 1998.
- 15 Van Tilbeurgh, H. *et al.*, *FEBS Lett.* 204:223-227, 1986.
- Verhoeyen *et al.*, *Science* 239:1534-1536, 1988.
- Warrington, *et al.*, *Genomics* 13:803-808, 1992.
- Wells *et al.*, *Gene* 34:315, 1985.
- Wells *et al.*, *Philos. Trans. R. Soc. London SerA* 317:415, 1986.
- 20 Wood, *Biochem. Soc. Trans.*, 13, pp. 407-410, 1985.
- Wood *et al.*, *METHODS IN ENZYMOLOGY*, 160, 25, p. 87 y ss., Academic Press, Nueva York, 1988.
- Zoller *et al.*, *Nucl. Acids Res.* 10:6487, 1987.

Antecedentes de la invención

- 25 **[0004]** La celulosa y la hemicelulosa son los materiales vegetales más abundantes producidos por fotosíntesis. Pueden degradarse y utilizarse como fuente de energía por numerosos microorganismos, entre los que se incluyen, bacterias, levadura y hongos, que producen enzimas extracelulares capaces de hidrólisis de los sustratos poliméricos en azúcares monoméricos (Aro *et al.*, 2001). Como se acercan los límites de los recursos no renovables, el potencial de la celulosa para convertirse en
- 30 un gran recurso de energía renovable es enorme (Krishna *et al.*, 2001). La utilización eficaz de la celulosa a través de procesos biológicos es una estrategia para superar la escasez de alimentos, piensos y combustibles (Ohmiya *et al.*, 1997).
- [0005]** Las celulasas son enzimas que hidrolizan la celulosa (enlaces β -1,4-glucán o β D-glucósidos), dando lugar a la formación de glucosa, celobiosa, celooligosacáridos, y similares. Las celulasas se
- 35 han dividido tradicionalmente en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 3.2.1.9.1) ("CBH") y β -glucosidasas (β -D-glucósido glucohidrolasa; EC 3.2.1.21) ("BG"). (Knowles *et al.*, 1987; Shulein, 1988). Las endoglucanasas

actúan principalmente en las partes amorfas de la fibra de celulosa, mientras que las celobiohidrolasas también son capaces de degradar la celulosa cristalina (Nevalainen y Penttila, 1995). De este modo, la presencia de una celobiohidrolasa en un sistema de celulasas es necesaria para la solubilización eficiente de la celulosa cristalina (Suurnakki, *et al.*, 2000). La β -glucosidasa actúa para liberar unidades de D-glucosidasa a partir de celobiosa, celooligosacáridos y otros glucósidos (Freer, 1993).

[0006] Se sabe que las celulasas son producidas por un gran número de bacterias, levaduras y hongos. Determinados hongos producen un sistema de celulasas completo capaz de degradar las formas cristalinas de la celulosa, de tal modo que las celulasas se producen fácilmente en grandes cantidades a través de la fermentación. Los hongos filamentosos desempeñan un papel especial, puesto que muchas levaduras, como *Saccharomyces cerevisiae*, no tienen la capacidad de hidrolizar la celulosa. Véanse, por ejemplo: Aro *et al.*, 2001; Aubert *et al.*, 1998; Wood *et al.*, 1988 y Coughlan, *et al.*

[0007] Las clasificaciones de la celulasa fúngica de CBH, EG y BG pueden extenderse aún más para incluir múltiples componentes en cada clasificación. Por ejemplo, se han aislado múltiples CBH, EG y BG de una diversidad de fuentes fúngicas entre las que se incluye *Trichoderma reesei* que contiene genes conocidos por 2CBH, es decir, CBH I y CBH II, al menos 5 EG, es decir, EG I, EG II, EG III, EG IV y EG V, y al menos 2 BG, es decir, BG1 y BG2.

[0008] Con el fin de convertir de modo eficaz la celulosa cristalina en glucosa se necesita el sistema de celulasas completo que comprende los componentes de cada una de las clasificaciones de CBH, EG y BG, con componentes aislados menos eficaces en la hidrólisis de la celulosa cristalina (Filho *et al.*, 1996). Se ha observado una relación sinérgica entre los componentes de la celulasa de distintas clasificaciones. En concreto, las celulasas de tipo EG y las celulasas tipo CBH interaccionan de modo sinérgico para degradar más eficientemente la celulosa. Véase, p.ej., Wood, 1985.

[0009] Las celulasas se conocen en la técnica por ser útiles en el tratamiento de tejidos con el fin de mejorar la capacidad de limpieza de las composiciones detergentes, para utilizar como un agente suavizante, y para mejorar el tacto y el aspecto de los tejidos de algodón y similares (Kumar *et al.*, 1997).

[0010] Se han descrito composiciones detergentes que contienen celulosa con mayor eficiencia de limpieza (Patente de EE. UU. Nº 4.435.307; Solicitudes de Reino Unido núms. 2.095.275 y 2.094.826) y para el uso en el tratamiento de la tela para mejorar el tacto y el aspecto del tejido (Patentes de EE. UU. núms. 5.648.263, 5.691.178 y 5.776.757; Solicitud de Reino Unido Nº 1.358.599; informe de The Shizuoka Prefectural Hammamatsu Textile Industrial Research Institute, Vol. 24, págs. 54-61, 1986).

[0011] Por tanto, las celulasas producidas en hongos y bacterias han recibido una atención significativa. En concreto, se ha demostrado que la fermentación de *Trichoderma spp.* (p.ej.: *Trichoderma longibrachiatum* o *Trichoderma reesei*) produce un sistema completo de celulasa capaz de degradar formas cristalinas de la celulosa. La Patente de EE. UU. Nº 5.475.101 divulga la purificación y clonación molecular de una enzima particularmente útil denominada EGIII se deriva de *Trichoderma longibrachiatum*.

[0012] Foreman PK *et al.* (*J. Biol. Chem.*, vol. 278, agosto 2003, páginas 31988-31997) describe la secuenciación de más de 5100 clones de ADNc de *T. Reesei*. Se evaluaron los niveles de expresión de estas secuencias en condiciones diferentes.

[0013] Takashima S *et al.*, (*J. Biochem.*, vol. 125, 1999, páginas 728-736) describe un gen de β -glucosidasa de origen fúngico (*bgl4*) y su homólogo (*bgl2*) clonados a partir del hongo celulolítico *Humicola grisea* y *Trichoderma reesei*, respectivamente.

[0014] EP-A-1 225227 describe un método para expresar b-glucosidasa extracelular en un hongo filamentoso mediante la expresión de una secuencia de ADN fúngico que codifica β -glucosidasa potenciada, eliminada o alterada en un microorganismo huésped recombinante. También se divulgan las composiciones de celulasa de origen fúngico recombinantes que contienen la expresión mejorada, eliminada o alterada de b-glucosidasa.

[0015] EP-A-O 338787 describe comestibles con bajo contenido de humedad que han reducido la recuperación de agua o aumentado la tolerancia a la humedad producidos por el tratamiento de modo enzimático de un material farináceo con una composición enzimática que está compuesta de pentosanasa o β -glucanasa, o mezclas de las mismas.

[0016] Aunque las composiciones de celulasa se han descrito anteriormente, aún existe la necesidad de nuevas y mejoradas composiciones de celulastas para su uso en detergentes domésticos, composiciones para el lavado a la piedra o detergentes para ropa, etc. Revisten especial interés las celulastas que presentan resistencia a los surfactantes (p.ej., sulfatos de alquilo lineal, LAS), rendimiento mejorado en condiciones de presión térmica, capacidad celulolítica aumentada o reducida y/o alto nivel de expresión *in vitro*.

Resumen de la invención

[0017] Esta invención proporciona una proteína celulasa aislada, identificada en el presente documento como BGL7, y ácidos nucleicos que codifican BGL7.

[0018] En un aspecto, los polipéptidos o proteínas BGL7 están compuestos de una secuencia que tiene al menos un 85%, 90%, 95%, 98% o más de identidad de secuencia con la secuencia presentada como SEQ ID N°:2.

[0019] Se describen en el presente documento fragmentos biológicamente activos purificados de BGL7 de al menos 20-100 aminoácidos de longitud aproximadamente, preferentemente sobre 100-200 aminoácidos de longitud de actividad de β -glucosidasa. El fragmento puede corresponder al dominio de extremo N-terminal de BGL7 o al dominio del extremo C-terminal de BGL7.

[0020] Se describe aquí donde el fragmento tiene un polinucleótido aislado que tiene una secuencia que codifica BGL7, una secuencia complementaria a la secuencia que codifica *bgl7*, y una composición que consta del polinucleótido. El polinucleótido puede ser ARNm, ADN, ADNc, o ADN genómico o un análogo antisentido del mismo.

[0021] Un polinucleótido *bgl7* puede constar de una molécula de ácido nucleico aislado de origen fúngico que se hibrida al complemento del ácido nucleico presentado como SEQ ID N°:4 en condiciones de alta astringencia, en las que la molécula de ácido nucleico codifica un polipéptido BGL7 que muestra actividad de β -glucosidasa.

[0022] El polinucleótido puede codificar una proteína BGL7 que tiene al menos un 85%, 90%, 95%, 98% o más de identidad de secuencia con la secuencia presentada como SEQ ID N°:2. El polinucleótido puede comprender la secuencia SEQ ID N°:4. En el presente documento también se describen fragmentos del polinucleótido, preferentemente de al menos aproximadamente 15-30 nucleótidos de longitud.

[0023] Además la invención también proporciona vectores de expresión recombinante que contienen una secuencia de ácidos nucleicos que codifica un polipéptido de la invención tal y como se describe en las reivindicaciones, unidos de modo funcional a elementos reguladores eficaces para la expresión de la proteína en un huésped seleccionado. En un aspecto relacionado, la invención incluye una célula huésped que contiene el vector.

[0024] La invención incluye además un método para producir un polipéptido de la invención como se define en las reivindicaciones con técnicas recombinantes, mediante el cultivo de células huésped procariontas o eucariotas que constan de una secuencia de ácido nucleico que codifica un polipéptido de la invención como se describe en las reivindicaciones en condiciones eficaces para promover la expresión de la proteína y la recuperación posterior de la proteína a partir de una célula huésped o del medio de cultivo de la célula.

[0025] En el presente documento se describe una composición enzimática útil para la conversión de celulosa en etanol. La composición enzimática puede además comprender un polipéptido de la invención como se define en las reivindicaciones. Además, la composición puede constar de enzimas de celulasas adicionales como las endoglucanasas y/o celobiohidrolasas. La composición puede enriquecerse con BGL7. También se describe en la invención un anticuerpo específicamente inmunorreactivo con BGL7.

[0026] También se incluyen métodos analíticos para detectar ácidos nucleicos *bgl7* y proteínas BGL7.

Breve descripción de las figuras

[0027]

La Figura 1 es una descripción de una sola hebra de la secuencia de ácido nucleico (SEQ ID N°:1), del ADNc de *T. reesei bgl7*, en la que la secuencia no codificante aparece en negrita y subrayada.

La Figura 2 muestra la secuencia de aminoácidos prevista (SEQ ID N°:2) y la secuencia señal (SEQ ID N°:3) basada en la secuencia de nucleótidos proporcionada en la Figura 1, en la que la secuencia señal aparece en negrita y subrayada.

Descripción detallada de la invención

I. Definiciones

[0028] A no ser que se indique lo contrario, todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado que el que tendrían para cualquier experto del campo de la presente invención. En concreto los profesionales se centran en Sambrook *et al.*, 1989 y Ausubel FM *et al.*, 1993, para las definiciones y términos de la técnica. Debe entenderse que esta invención no se limita a la metodología, protocolos y reactivos concretos descritos, ya que pueden variar.

[0029] El término “polipéptido” como se utiliza en el documento se refiere a un compuesto formado por una cadena sencilla de residuos de aminoácidos unidos mediante enlaces peptídicos. El término “proteína” como se utiliza en el presente documento puede ser sinónimo del término “polipéptido” o puede referirse, además, a un complejo de dos o más polipéptidos”.

5 **[0030]** El término “molécula de ácido nucleico incluye moléculas de ARN, ADN y ADNc. Se entenderá que, como resultado de la degeneración del código genético, puede producirse una multitud de secuencias de nucleótidos que codifican una proteína determinada como la BGL7. La presente invención contempla todas las posibles variantes de secuencias de nucleótidos, que codifican BGL7, todas ellas son posibles dada la degeneración del código genético.

10 **[0031]** Una secuencia o construcción de ácido nucleico “heteróloga” tiene una parte de la secuencia que no es nativa con respecto a la célula en la que se expresa. Heteróloga, con respecto a una secuencia de control, se refiere a una secuencia de control (es decir, promotor o potenciador) que no funciona en naturaleza para regular el mismo gen cuya expresión se está regulando actualmente. En general, las secuencias de ácido nucleico heterólogas no son endógenas a la célula o parte del
15 genoma en el que están presentes, y se han añadido a la célula, mediante infección, transfección, transformación, microinyección, electroporación o similar. Una construcción de ácido nucleico “heteróloga” puede contener una combinación de secuencia de codificación de ADN/secuencia de control que es la misma o diferente de una combinación de secuencia de control/secuencia de codificación de ADN localizadas en la célula de origen.

20 **[0032]** Como se utiliza en el presente documento, el término “vector” se refiere a una construcción de ácido nucleico diseñada para la transferencia entre diferentes células huésped. Un “vector de expresión” hace referencia a un vector que tiene la capacidad de incorporar y expresar los fragmentos de ADN heterólogo en una célula extraña. Están disponibles en el mercado muchos vectores de expresión procariota o eucariota. La elección de los vectores de expresión apropiada forma parte del
25 conocimiento de los expertos en la técnica.

[0033] En consecuencia, un “casete de expresión” o “vector de expresión” es una construcción de ácido nucleico generada recombinante o sintéticamente, con una serie de elementos de ácido nucleico especificados que permiten la transcripción de un ácido nucleico concreto en una célula diana. El casete de expresión recombinante puede incorporarse en un plásmido, cromosoma, ADN
30 mitocondrial, ADN de plastidio, virus o un fragmento de ácido nucleico. Normalmente, la parte del casete de expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico para ser transcrita y un promotor.

[0034] Como se utiliza en el presente documento, el término “plásmido” hace referencia a una construcción de ADN bicatenario (ds, en inglés) circular que se usa como vector de clonación, y que
35 forma un elemento genético autorreplicante extracromosómico en muchas bacterias y en algunas eucariotas.

[0035] Según su uso en el presente documento, la expresión “secuencia de nucleótidos que codifica un marcador seleccionable” se refiere a una secuencia de nucleótidos que es capaz de expresión en las células y donde la expresión del marcador seleccionable confiere a las células que contienen el
40 gen expresado la capacidad de crecer en presencia de un agente selectivo correspondiente o en condiciones de crecimiento selectivo correspondientes.

[0036] Como se utiliza en este documento, el término “promotor” hace referencia a una secuencia de ácido nucleico que funciona para dirigir la transcripción de un gen en dirección 3'. En general, el promotor será adecuado para la célula huésped en la que se expresa el gen diana. El promotor, junto con otras secuencias de ácido nucleico reguladoras de la transcripción y de traducción (también llamadas “secuencias de control”) es necesario para expresar un gen determinado. En general, entre las secuencias reguladoras de la transcripción y de traducción se incluyen, sin carácter limitativo, secuencias promotoras, sitios de unión a ribosoma, secuencias de inicio y terminación de la transcripción y secuencias potenciadoras o activadoras.

[0037] Según el presente documento, “gen quimérico” o “construcción de ácido nucleico heteróloga” se refiere a un gen no nativo (es decir, uno que se ha introducido en un huésped) que puede estar compuesto de partes de genes diferentes, incluidos elementos reguladores. Una construcción de un gen quimérico para la transformación de una célula huésped está normalmente compuesta de una región reguladora de la transcripción (promotor) unida de modo funcional a una secuencia de codificación de proteínas heterólogas, o, en un gen quimérico con marcador seleccionable, a un gen marcador seleccionable que codifica una proteína que confiere resistencia antibiótica a las células transformadas. Un gen quimérico típico de la presente invención, para la transformación en una célula huésped, incluye una región reguladora de la transcripción que es constitutiva o inducible, una secuencia de codificación de proteína y una secuencia de terminación. Una construcción de gen quimérico también puede incluir una segunda secuencia de ADN que codifica un péptido señal si se desea la secreción de la proteína diana.

[0038] Un ácido nucleico está “unido de modo funcional” cuando está situado en una relación funcional con otra secuencia de ácidos nucleicos. Por ejemplo, el ADN que codifica una secuencia señal está unido de modo funcional al ADN para un polipéptido si éste se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o potenciador está unido de modo funcional a una secuencia de codificación si afecta a la transcripción de la secuencia; o un sitio de unión a ribosoma está unido de modo funcional a una secuencia de codificación si se coloca con el fin de facilitar la traducción. En general, “unido de modo funcional” significa que las secuencias de ADN que se unen son contiguas y, en el caso de una secuencia señal, contiguas y en el marco de lectura. Sin embargo, los potenciadores no tienen que ser contiguos. La unión se realiza mediante el enlace en los sitios de restricción convenientes. Si no existen dichos sitios, se utilizan adaptadores de oligonucleótidos sintéticos, enlaces o cebadores para PCR de acuerdo con la práctica convencional.

[0039] Como se utiliza en este documento, el término “gen” se refiere al segmento de ADN que participa en la producción de una cadena de polipéptidos, que puede o no incluir las regiones anterior y posterior a la región de codificación, p.ej., secuencias 5' no traducidas (5' UTR) o secuencias “líder” y secuencias 3' UTR o “tráiler”, así como secuencias intercalada (intrones) entre los segmentos de codificación individuales (exones).

[0040] En general, las moléculas de ácido nucleico que codifican BGL7 o un análogo u homólogo de la misma se hibridarán, en condiciones de moderada a alta astringencia, a la secuencia presentada en este documento como SEQ ID N°:1. No obstante, en algunos casos una secuencia de nucleótidos que codifica BGL7 que tiene un uso de codón sustancialmente diferentes, mientras que la proteína codificada por la secuencia de nucleótidos que codifica BGL7 tiene la misma o sustancialmente la

misma secuencia de aminoácidos que la proteína nativa. Por ejemplo, la secuencia de codificación puede ser modificada para facilitar la expresión más rápida de BGL7 en un sistema de expresión procariota o eucariota particular, de acuerdo con la frecuencia con la que el huésped utiliza un codón concreto. Te'o *et al.* (2000), por ejemplo, describe la optimización de genes para la expresión en
5 hongos filamentosos.

[0041] Una secuencia de ácido nucleico se considera que es “selectivamente hibridable” a una secuencia de ácido nucleico de referencia, si las dos secuencias se hibridan específicamente entre sí en condiciones de hibridación de moderada a alta astringencia y de lavado. Las condiciones de hibridación se basan en la temperatura de fusión (Tf) de la sonda o del complejo de unión a ácidos
10 nucleicos. Por ejemplo, la “máxima astringencia” normalmente se produce a aproximadamente Tf-5°C (5°C por debajo de la Tf de la sonda); la “alta astringencia” a aproximadamente 5-10°C por debajo de la Tf; “astringencia intermedia” aproximadamente a 10-20°C por debajo de la Tf de la sonda; y “baja astringencia” aproximadamente a 20-25°C por debajo de la Tf. Funcionalmente, pueden usarse las condiciones de máxima astringencia para identificar secuencias que tienen identidad estricta o casi
15 estricta con la sonda de hibridación; mientras que las condiciones de alta astringencia se utilizan para identificar secuencias que tienen aproximadamente un 80% o más de identidad con la sonda.

[0042] En la técnica, son bien conocidas las condiciones de hibridación de moderada a alta astringencia (véanse, por ejemplo, Sambrook, *et al.*, 1989, capítulos 9 y 11, y Ausubel, F.M., *et al.*, 1993). Un ejemplo de las condiciones de alta astringencia incluye la hibridación a aproximadamente
20 42°C en 50% de formamida, 5X de SSC, 5X de solución Denhardt, 0,5% de SDS y 100 mg/ml de ADN portador desnaturizado seguido de un lavado en dos veces en 2X de SSC y 0,5% de SDS a temperatura ambiente y dos veces más en 0,1X de SSC y 0,5% de SDS a 42°C.

[0043] Como se utiliza en el documento, “recombinante” se refiere a una célula o un vector, que ha sido modificado por la introducción de una secuencia de ácido nucleico heteróloga o que la célula se
25 deriva de una célula modificada de ese modo. En consecuencia, por ejemplo, las células recombinantes expresan genes que no se encuentran en forma idéntica en la forma nativa (no recombinante) de la célula o expresan genes nativos que de otra manera se expresan anormalmente, se expresan poco o no se expresan en absoluto como consecuencia de la intervención humana deliberada.

[0044] Como se usa en el presente documento, los términos “transformada”, “transformada de modo estable” o “transgénica”, con referencia a una célula, significan que la célula tiene una secuencia de
30 ácidos nucleicos no nativa (heteróloga) integrada en su genoma o como un plásmido episomal que se mantiene a través de múltiples generaciones.

[0045] Según su uso en este documento, el término “expresión” hace referencia al proceso mediante el cual se produce un polipéptido sobre la base de la secuencia de ácido nucleico de un gen. El
35 proceso incluye tanto la transcripción como la traducción.

[0046] El término “introducido” en el contexto de la inserción de una secuencia de ácido nucleico en una célula significa “transfección” o “transformación” o “transducción” e incluye la referencia a la
40 incorporación de una secuencia de ácido nucleico en una célula eucariota o procariota, donde la secuencia de ácido nucleico puede incorporarse en el genoma de la célula (por ejemplo, cromosoma,

plásmido, plástido o ADN mitocondrial), convertido en un replicón autónomo, o expresado de forma transitoria (por ejemplo, ARNm transfectado).

[0047] En consecuencia, el término “expresión de BGL7” hace referencia a la transcripción y traducción del gen *bgl7*, cuyos productos incluyen precursores de ARN, ARNm, polipéptido, polipéptidos procesados de manera postraducciona l y sus derivados, incluido BGL7 de especies relacionadas, como *Trichoderma longibrachiatum (reesei)*, *Trichoderma viride*, *Trichoderma koningii*, *Hypocrea jecorina* e *Hypocrea schweinitzii*. A modo de ejemplo, entre los ensayos para la expresión de BGL7 se incluyen: transferencia Western blot para proteínas BGL7, análisis de transferencia Northern blot y ensayos de reacción en cadena de polimerasa con transcriptasa inversa (RT-PCR) para ARNm de BGL7, y ensayos de actividad de glucosidasa como se describe en Chen *et al.* (1992) y Herr *et al.* (1978).

[0048] El término “corte y empalme alternativo” se refiere al proceso mediante el cual se generan múltiples isoformas del polipéptido a partir de un único gen, e implica el corte y empalme conjunto de exones no consecutivos durante el procesamiento de algunos, pero no todos, los transcritos del gen. De este modo, un exón determinado puede conectarse a uno cualquiera de los exones alternativos para formar ARN mensajero. Los ARNm cortados y empalmados de modo alternativo producen polipéptidos (“variantes de corte y empalme”) en los que algunas partes son comunes, mientras que otras partes son diferentes.

[0049] El término “secuencia señal” hace referencia a una secuencia de aminoácidos en la porción de extremo N-terminal de una proteína que facilita la secreción de la forma madura de la proteína fuera de la célula. La forma madura de la proteína extracelular carece de la secuencia señal que se separa durante el proceso de secreción.

[0050] Por el término “célula huésped” se entiende una célula que contiene un vector y soporta la replicación y/o la transcripción o la transcripción y la traducción (expresión) de la construcción de expresión. Las células huésped utilizadas en la presente invención pueden ser células procariotas, como *E.coli*, o células eucariotas, como células de levadura, planta, insecto, anfibio o mamífero. En general, las células huésped son hongos filamentosos.

[0051] El término “hongos filamentosos” hace referencia a todos y cada uno de los hongos filamentosos reconocidos por los expertos en la materia. Se selecciona un hongo preferido de un grupo formado por *Aspergillus*, *Trichoderma*, *Fusarium*, *Chrysosporium*, *Penicillium*, *Humicola*, *Neurospora* o formas sexuales alternativas de los mismos como *Emericella*, *Hypocrea*.

[0052] El término “celooligosacárido” se refiere a grupos de oligosacáridos que contienen 2-8 unidades de glucosa y tienen enlaces β -1,4, por ejemplo, celobiosa.

[0053] El término “celulasa” se refiere a una categoría de enzimas capaces de hidrolizar polímeros de celulosa a oligómeros celooligosacáridos más cortos, celobiosa y/o glucosa. Se han obtenido numerosos ejemplos de celulasas, como exoglucanasas, exocelobiohidrolasas, endoglucanasas y glucosidasas a partir de organismos celulolíticos, incluyendo en particular hongos, plantas y bacterias.

[0054] El término “dominio de unión a celulosa”, como se utiliza en el presente documento, se refiere a la parte de la secuencia de aminoácidos de una celulasa o a una región de la enzima que interviene en la actividad de unión a celulosa, de una celulasa o derivado de la misma. En general, los dominios de unión a celulosa funcionan mediante la unión no covalente de la celulasa a la celulosa, un

derivado de celulosa u otro polisacárido equivalente del mismo. Los dominios de unión a celulosa permiten o facilitan la hidrólisis de fibras de celulosa por la región del núcleo catalítico estructuralmente distinto, y habitualmente funcionan independientes del núcleo catalítico. De este modo, un dominio de unión a celulosa no tendrá la actividad hidrolítica significativa atribuible a un núcleo catalítico. En otras palabras, un dominio de unión a celulosa es un elemento estructural de la estructura terciaria de la proteína enzimática celulasa que es distinto del elemento estructural que posee la actividad catalítica.

[0055] Como se utiliza en el presente documento, el término “surfactante” se refiere a cualquier compuesto generalmente reconocido en la materia por tener cualidades tensoactivas. De este modo, por ejemplo, los surfactantes comprenden surfactantes aniónicos, catiónicos y no iónicos, como los que se encuentran habitualmente en detergentes. Los surfactantes aniónicos incluyen alquilbenzenosulfonatos lineales o ramificados; sulfatos de éter de alquilo o alquenilo que tienen grupos alquilo lineales o ramificados o grupos alquenilos; sulfatos de alquilo o alquenilo; olefinsulfonatos y alcanosulfonatos. Los surfactantes anfólicos incluyen sulfonatos de sal de amonio cuaternario y surfactantes anfólicos de tipo betaína. Dichos surfactantes anfólicos tienen tanto grupos de carga positiva como negativa en la misma molécula. Los surfactantes no iónicos pueden comprender éteres de polioxialquileo, así como alcanolamidas de ácidos grasos superiores o aductos de óxido de alquileo, monoésteres de glicerina de ácidos grasos y similares.

[0056] Como se usa aquí, el término “tejido que contiene celulosa” se refiere a cualquier tejido cosido o no cosido, hilos o fibras de algodón o no de algodón que contienen celulosa o mezclas de celulosa que contienen o no algodón, incluidos celulósicos naturales y celulósicos artificiales (como yute, lino, ramio, rayón y liocel).

[0057] Como se utiliza en este documento, el término “tejido que contiene algodón” se refiere a tejidos cosidos o no cosidos, hilos o fibras de algodón puro o mezclas de algodón, entre los que se incluye algodón en tejidos blanqueados de algodón, tricotados de algodón, vaqueros de algodón, hilos de algodón, algodón en rama y similares.

[0058] Tal y como se usa aquí, el término “composición de lavado a la piedra” se refiere a una formulación para uso en tejidos que contienen celulosa para el lavado a la piedra. Las composiciones de lavado a la piedra se utilizan para modificar los tejidos que contienen celulosa antes de su venta, es decir, durante el proceso de fabricación. Por el contrario, las composiciones detergentes están destinadas a la limpieza de prendas sucias y no se utilizan durante el proceso de fabricación.

[0059] Como se utiliza en el presente documento, el término “composición detergente” se refiere a una mezcla que está destinada para su uso en un medio de lavado de tejidos sucios que contienen celulosa. En el contexto de esta invención, dichas composiciones pueden incluir, además de celulasas y surfactantes, enzimas hidrolíticas adicionales, agentes de relleno, agentes de blanqueo, activadores de blanqueo, agentes de azulado y tintes fluorescentes, inhibidores de la aglomeración, agentes de enmascaramiento, activadores de celulasas, antioxidantes y solubilizantes.

[0060] Tal y como se utiliza en el presente documento, el término “reducción o eliminación de la expresión del gen *bgl7*” significa que el gen *bgl7* ha sido eliminado del genoma y, por tanto, no puede ser expresado por el microorganismo huésped recombinante, o que el gen *bgl7* ha sido modificado de manera que microorganismo huésped recombinante no produce una enzima BGL7 funcional.

[0061] El término “*bg17* alterado” o “gen *bg17* alterado” significa que la secuencia de ácido nucleico del gen se ha alterado mediante la eliminación, adición y/o manipulación de la secuencia de codificación o se ha modificado la secuencia de aminoácidos de la proteína expresada.

[0062] Como se usa en el documento, el término “purificar” generalmente se refiere a someter a los ácidos nucleicos transgénicos o las células que contienen proteínas a purificación bioquímica y/o cromatografía en columna.

[0063] Tal como se usa aquí, los términos “activo” y “biológicamente activos” se refieren a una actividad biológica asociada a una proteína determinada, como la actividad enzimática asociada con una proteasa. Se deduce que la actividad biológica de una proteína determinada se refiere a cualquier actividad biológica que suele atribuirse a esa proteína por parte de los expertos en la materia.

[0064] Como se utiliza en este documento, el término “enriquecida” significa que BGL7 se encuentra en una concentración que es mayor en relación a la concentración de BGL7 encontrada en una composición de celulasa fúngica de tipo salvaje, o de origen natural.

[0065] Una composición de celulasa fúngica de tipo salvaje es la producida por una fuente fúngica de origen natural y que comprende uno o más componentes BG, CBH y EG, donde cada uno de estos componentes se encuentra en la proporción producida por la fuente fúngica. De este modo, una composición de BGL7 enriquecida tendría BGL7 en una proporción alterada, donde la proporción de CBH con respecto a otros componentes de celulasa (es decir, CBHs y endoglucanasas) es elevada. Esta proporción puede incrementarse ya sea aumentando BGL7 o disminuyendo (o eliminando) por lo menos uno de los otros componentes mediante cualquier medio conocido en la técnica.

[0066] Así, a modo de ejemplo, un sistema de celulasas de origen natural puede ser purificado en componentes sustancialmente puros mediante técnicas de separación reconocidas descritas en publicaciones científicas, incluidas la cromatografía de intercambio iónico con un pH adecuado, cromatografía de afinidad, exclusión de tamaño y similares. Por ejemplo, en la cromatografía de intercambio iónico (normalmente cromatografía de intercambio aniónico) es posible separar los componentes de celulasa por elución con un gradiente de pH y un gradiente de sal, o ambos, un gradiente de pH y un gradiente de sal. La BGL7 purificada se puede añadir a continuación a la solución enzimática dando lugar a una solución de BGL7 enriquecida.

[0067] Las celulasas fúngicas pueden contener más de un componente BG. Los diferentes componentes generalmente tienen diferentes puntos isoeléctricos que permiten su separación mediante cromatografía de intercambio iónico y similares. Se puede utilizar un único componente BG, o una combinación de componentes BG en una solución enzimática.

[0068] Cuando se utiliza en soluciones enzimáticas, el componente BG se añade generalmente en una cantidad suficiente para evitar la inhibición por celobiosa de cualesquiera componentes de endoglucanasa y CBH encontrados en la composición de celulasa. La cantidad de componente BG añadido depende de de la cantidad de de celobiosa producida durante el proceso de sacarificación de biomasa que puede determinarse fácilmente por parte del experto en la materia. No obstante, cuando se utiliza, el porcentaje en peso del componente BGL7 con respecto a cualquier tipo de componente CBH o endoglucanasa presente en la composición de celulasa es preferentemente de aproximadamente 1, preferentemente aproximadamente 5, preferentemente aproximadamente 10,

preferentemente aproximadamente 15 o preferentemente aproximadamente 20 por ciento en peso a preferentemente 25 aproximadamente, preferentemente 30 aproximadamente, preferiblemente 35 aproximadamente, preferentemente 40 aproximadamente, preferentemente aproximadamente 15 o preferiblemente aproximadamente 50 por ciento en peso. Además, los intervalos preferidos pueden ser de aproximadamente 0,5 a aproximadamente 15 por ciento en peso, de aproximadamente 0,5 a aproximadamente a aproximadamente 20 por ciento en peso, de aproximadamente 1 a aproximadamente 10 por ciento en peso, de aproximadamente 1 a aproximadamente 15 por ciento en peso, de aproximadamente 1 a aproximadamente 20 por ciento en peso, de aproximadamente 1 a 25 por ciento en peso, de aproximadamente 5 a aproximadamente 20 por ciento en peso, de aproximadamente 5 a aproximadamente 25 por ciento en peso, de aproximadamente 5 a aproximadamente 30 por ciento en peso, de aproximadamente 5 a aproximadamente 35 por ciento en peso, de aproximadamente 5 a aproximadamente 40 por ciento en peso, de aproximadamente 5 a aproximadamente 45 por ciento en peso, de aproximadamente 5 a aproximadamente 50 por ciento en peso, de aproximadamente 10 a aproximadamente 20 por ciento en peso, de aproximadamente 10 a aproximadamente 25 por ciento en peso, de aproximadamente 10 a aproximadamente 30 por ciento en peso, de aproximadamente 10 a aproximadamente 35 por ciento en peso, de aproximadamente 10 a aproximadamente 40 por ciento en peso, de aproximadamente 10 a aproximadamente 45 por ciento en peso, de aproximadamente 10 a aproximadamente 50 por ciento en peso, de aproximadamente 15 a aproximadamente 20 por ciento en peso, de aproximadamente 15 a aproximadamente 25 por ciento en peso, de aproximadamente 15 a aproximadamente 30 por ciento en peso, de aproximadamente 15 a aproximadamente 35 por ciento en peso, de aproximadamente 15 a aproximadamente 40 por ciento en peso, de aproximadamente 15 a aproximadamente 45 por ciento en peso, de aproximadamente 15 a aproximadamente 50 por ciento en peso.

II. Organismos diana

25 A. Hongos filamentosos

[0069] Los hongos filamentosos incluyen todas las formas filamentosas de la subdivisión Eumycota y Oomycota. Los hongos filamentosos se caracterizan por un micelio vegetativo que tiene una pared celular compuesta de quitina, glucano, quiosano, manano y otros polisacáridos complejos, con crecimiento vegetativo por la elongación hifal y catabolismo de carbono que es obligadamente aeróbico.

[0070] En la presente invención, la célula madre de los hongos filamentosos puede ser una célula de una especie de, sin carácter limitativo, *Trichoderma*, por ejemplo, *Trichoderma longibrachiatum* (*reesei*), *Trichoderma viride*, *Trichoderma koningii*, *Trichoderma harzianum*; *Penicillium sp*; *Humicola sp.*, incluida *Humicola insolens*; *Chrysosporium sp.*; incluida *C. lucknowense*; *Gliocladium sp.*; *Aspergillus sp.*; *Fusarium sp*, *Neurospora sp.*, *Hypocrea sp.* y *Emericella sp.* Como se utiliza en el presente documento, el término "*Trichoderma*" o "*Trichoderma sp*" se refiere a cualquier cepa de hongos que ha sido previamente clasificada como *Trichoderma* o está actualmente clasificada como *Trichoderma*.

[0071] La célula madre de los hongos filamentosos puede ser una célula de *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus aculeatus* o *Aspergillus nidulans*.

[0072] En una realización preferida, la célula madre de los hongos filamentosos es una célula de *Trichoderma reesei*.

5 III. Celulasas

[0073] Las celulasas se conocen en la técnica como enzimas que hidrolizan celulosa (enlaces β -1,4-glucano o β D-glucosídicos), lo que resulta en la formación de glucosa, celobiosa, celooligosacáridos y similares. Como se ha señalado anteriormente, las celulasas se han dividido tradicionalmente en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 10 3.2.1.91) ("CBH") y β -glucosidasas (CE 3.2.1.21) ("BG"). (Knowles *et al.*, 1987; Schulein, 1988).

[0074] Algunos hongos producen sistemas de celulasas completos que incluyen exocelobiohidrolasas o celulasas de tipo CBH, endoglucanasas o celulasas de tipo EG y β -glucosidasas o celulasas de tipo BG (Schulein, 1988). No obstante, algunas veces estos sistemas carecen de celulasas de tipo CBH y las celulasas bacterianas también incluyen normalmente pocas o ninguna celulasa de tipo CBH. 15 Además, se ha demostrado que los componentes EG y los componentes CBH interactúan sinérgicamente para degradar la celulosa de modo más eficaz. Véase, por ejemplo, Wood, 1985. Los diferentes componentes, es decir, las diversas endoglucanasas y exocelobiohidrolasas en un sistema de celulasas multicomponentes o completo, generalmente tienen propiedades diferentes, como el punto isoeléctrico, peso molecular, grado de glucosilación, especificidad de sustrato y patrones de 20 acción enzimáticos.

[0075] Se cree que las celulasas de tipo endoglucanasa hidrolizan los enlaces β -1,4-glucosídicos internos en regiones de baja cristalinidad de la celulosa y que las celulasas de tipo exocelobiohidrolasa hidrolizan la celobiosa a partir del extremo reductor o no reductor de la celulosa. Se deduce que la acción de los componentes de las endoglucanasas puede facilitar mucho la acción 25 de las exocelobiohidrolasas mediante la creación de nuevos extremos de cadena que son reconocidos por componentes de las exocelobiohidrolasas. Además, se ha demostrado que las celulasas de tipo β -glucosidasa catalizan la hidrólisis de β -D-glucósidos de alquilo y/o arilo, como β -D-glucósidos de metilo y glucósido de p-nitrofenilo, así como glucósidos que contienen únicamente residuos de carbohidratos, como celobiosa. Lo que produce glucosa como el único producto para los 30 microorganismos y reduce o elimina la celobiosa que inhibe las celobiohidrolasas y las endoglucanasas.

[0076] En consecuencia, se considera que las celulasas de tipo β -glucosidasa son una parte integrante del sistema de celulasas porque dirigen la reacción general a la glucosa. Se ha demostrado que la expresión aumentada de BG en *T. reesei* mejora la degradación de la celulosa en 35 glucosa. Véase EP0562003. Además, las β -glucosidasas pueden catalizar la hidrólisis de una serie de diferentes sustratos, y por lo tanto, son útiles en una variedad de aplicaciones distintas. Pueden añadirse algunas β -glucosidasas a las uvas durante la elaboración del vino para potenciar el aroma del producto vinícola acabado. Otra aplicación más puede consistir en utilizar β -glucosidasa en la

fruta para potenciar el aroma de la misma. Por otro lado, la β -glucosidasa puede usarse directamente en los aditivos alimentarios o en la elaboración de vino para realzar el sabor y el aroma.

[0077] Las celulasas también pueden usarse en composiciones detergentes para mejorar la capacidad de limpieza, como agente suavizante y mejorar el tacto de los tejidos de algodón (Hemmpel, 1991; Tyndall, 1992; Kumar *et al.*, 1997). Aunque el mecanismo no es parte de la invención, se han atribuido las propiedades de restauración de la suavidad y el color de la celulasa a los componentes de la endoglucanasa alcalina en las composiciones de celulasas, como lo demuestran las patentes de EE. UU. núms. 5.648.263, 5.699.178 y 5.776.757, que divulgan que las composiciones detergentes que contienen una composición de celulasa enriquecida en un componente de endoglucanasa alcalina específico proporcionan una restauración del color y mejora de la suavidad de las prendas de vestir tratadas en comparación con las composiciones de celulasa no enriquecidas con dicho componente. Además, se ha demostrado que el uso de estos componentes de endoglucanasa alcalina en las composiciones detergentes cumple con los requisitos de pH de la composición detergente (por ejemplo, exhibiendo su actividad máxima en un pH alcalino de 7,5 a 10, tal como se describe en las Patentes de EE. UU. núms. 5.648.263, 5.691.178 y 5.776.757).

[0078] También se ha demostrado que las composiciones de celulasa degradan tejidos que contienen algodón, lo que resulta en una pérdida de resistencia reducida en el tejido (Patente de EE. UU. Nº: 4.822.516), lo que contribuye a la resistencia al uso de composiciones de celulasas en aplicaciones comerciales de detergente. Se ha sugerido que las composiciones de celulasas formadas por componentes de endoglucanasa presentan una pérdida de resistencia reducida para los tejidos que contienen algodón en comparación con las composiciones que comprenden un sistema de completo de celulasas.

[0079] También se ha demostrado que las composiciones de celulasas son útiles en la degradación de biomasa de celulasas en etanol (en la que la celulasa degrada celulosa a glucosa y levadura u otros microbios fermentan además la glucosa en etanol), en el tratamiento de pulpa mecánica (Pere *et al.*, 1996), para su uso como aditivo de piensos (WO 91/04673) y en la molienda húmeda del grano.

[0080] En las publicaciones científicas, se han descrito numerosas celulasas, cuyos ejemplos incluyen: de *Trichoderma reesei*: Shoemaker, S. *et al.*, *Bio/Technology*, 1:691-696, 1983, que presenta *CBHI*; Teeri, T. *et al.*, *Gene*, 51: 43-52, 1987, que describe *CBHII*; Penttila, M. *et al.*, *Gene*, 45:253-263, 1986, que presenta *EGI*; Saloheimo, M. *et al.*, *Gene*, 63:11-22, 1988, que divulga *EGII*; Okada, M. *et al.*, *Appl. Environ. Microbiol.*, 64:555-563, 1988, que describe *EGIII*; Saloheimo, M. *et al.*, *Eur. J. Biochem.*, 249:584-591, 1997, que describe *EGIV*, Saloheimo, A. *et al.*, *Molecular Microbiology*, 13:219-228, 1994, que presenta *EGV*; Barnett, C.C., *et al.*, *Bio/Technology*, 9:562-567, 1991, que divulga *BGL1* y Takashima, S. *et al.*, *J. Biochem.*, 125:728-736, 1999, que describe *BGL2*. Las celulasas de especies diferentes de *Trichoderma* también se han descrito, por ejemplo, Ooi *et al.*, 1990, que presenta la secuencia de ADNc que codifica la endoglucanasa F1-CMC producida por *Aspergillus aculeatus*, Kawaguchi T *et al.*, 1996, que describe la clonación y la secuenciación de ADNc que codifica la β -glucosidasa 1 de *Aspergillus aculeatus*; Sakamoto *et al.*, 1995, que presenta la secuencia de ADNc que codifica la endoglucanasa CMCCase-1 de *Aspergillus kawachii* IFO 4308;

Saarilahti *et al.*, 1990, que divulga una endoglucanasa de *Erwinia carotovora*; Spilliaert R, *et al.*, 1994, que describe la clonación y la secuenciación de *bglA*, que codifica una β -glucanasa termoestable de *Rhodothermus marinus* y Halldorsdottir S *et al.*, 1998, que presenta la clonación, la secuenciación y la sobreexpresión de un gen de *Rhodothermus marinus* que codifica una celulasa termoestable de la familia 12 de glicosil hidrolasa. Sin embargo, siguen siendo necesarias la identificación y caracterización de nuevas celulasas, con propiedades mejoradas, como mejor rendimiento en condiciones de tensión térmica o en la presencia de surfactantes, mayor actividad específica, patrón de escisión del sustrato alterado y/o nivel alto de expresión *in vitro*.

[0081] El desarrollo de nuevas y mejoradas composiciones de celulasa que constan de cantidades variables de celulasas de tipo CBH, de tipo EG y de tipo BG es de interés para su uso: (1) en composiciones detergentes que muestran una mayor capacidad de limpieza, que funcionan como un agente suavizante y/o de mejor del tacto de los tejidos de algodón (por ejemplo, “lavado a la piedra” o “biopulido”); (2) en composiciones para degradar la pulpa de madera u otras biomásas en azúcares (por ejemplo, para la producción de bioetanol) y/o (3) en composiciones de piensos.

IV. Métodos para la identificación de nuevas secuencias

[0082] Se analizan los marcos de lectura abierta (ORF, en inglés) después de la secuenciación total o parcial del genoma de *T. reesei* o de clones de las bibliotecas de ADNc de ARNm de *T. reesei* y se analizan posteriormente usando un programa de análisis de secuencia, y determinando la homología con secuencias conocidas en bases de datos (públicas/privadas).

V. Ácidos nucleicos *bgl7* y polipéptidos BGL7

A. Ácidos nucleicos *bgl7*

[0083] Las moléculas de ácido nucleico descritas en el presente documento incluyen las secuencias de codificación nativas, la secuencia de ADNc para *bgl7* presentada aquí como SEQ ID N°:1 y homólogos de las mismas en otras especies, variantes de corte y empalme y alélicas de origen natural, fragmentos de ácidos nucleicos y derivados de los mismos biológicamente activos (funcionales), como variantes de secuencias de aminoácidos de la molécula nativa y secuencias que codifican proteínas de fusión. Las secuencias se denominan colectivamente en este documento “secuencias de ácido nucleico que codifican BGL7”.

[0084] Se realizó una búsqueda con BLASTN Basic (<http://www.ncbi.nlm.nih.gov/BLAST>) de la base de datos de secuencias de ácidos nucleicos no redundantes el 18 de octubre de 2002, con la secuencia del gen *bgl7* mostrada en la Figura 1, que indicó que no hay secuencias que produzcan alineamientos significativos (es decir, con un valor E de 10^{-5} o menor).

[0085] Una secuencia de ácido nucleico de *bgl7* puede ser una secuencia de ADN o ARN, derivada del genómico ADN, ADNc, ARNm o puede sintetizarse en su totalidad o en parte. El ADN puede ser bicatenario o monocatenario, y si es monocatenario puede ser la cadena de codificación o la cadena de no codificación (antisentido, complementaria). Puede clonarse la secuencia de ácido nucleico, por ejemplo, aislando el ADN genómico de una fuente adecuada, y amplificando y clonando la secuencia de interés utilizando una reacción en cadena de la polimerasa (PCR). Como alternativa, puede

sintetizarse la secuencia de ácido nucleico, ya sea en su totalidad o en parte, especialmente cuando se desea proporcionar secuencias preferidas como huésped para la expresión óptima. Por tanto, puede sintetizarse todo o parte del gen estructural deseado (la parte del gen que codifica un polipéptido o proteína) utilizando codones preferidos por un huésped seleccionado.

5 **[0086]** Debido a la degeneración inherente del código genético, se pueden utilizar secuencias de ácido nucleico distintas de la forma nativa que codifican sustancialmente la misma secuencia o una secuencia de aminoácidos funcionalmente equivalente para clonar y/o expresar las secuencias de ácido nucleico que codifican BGL7. Por tanto, para una secuencia determinada de ácido nucleico que codifica BGL7, se apreciará que como resultado de la degeneración del código genético, puede
10 generarse una serie de secuencias de codificación que codifican una proteína que tiene la misma secuencia de aminoácidos. Por ejemplo, el triplete CGT codifica el aminoácido arginina. La arginina es codificada alternativamente por CGA, CGC, CGG, AGA y AGG. Por lo tanto, se aprecia que dichas sustituciones en la región de codificación se incluyen en las variantes de la secuencia de ácidos nucleicos presentada en la presente invención. Se pueden usar todas y cada una de estas variantes
15 de la secuencia puede utilizarse del mismo modo descrito en el presente documento para la forma nativa de una secuencia de ácido nucleico que codifica BGL7.

[0087] Una “variante” de la secuencia de ácido nucleico que codifica BGL7 puede codificar una “variante” de la secuencia de aminoácidos de BGL7 que está alterada por uno o más aminoácidos respecto a la secuencia de polipéptido nativo o puede alterarse por la eliminación de uno o más
20 aminoácidos de cualquiera de los dos extremos de la secuencia del polipéptido; ambas están incluidas en el objetivo de la invención. Del mismo modo, la expresión “forma modificada de”, en relación con BGL7, se refiere a una forma o derivado variante de la secuencia nativa de ácido nucleico que codifica la proteína BGL7 o la secuencia nativa de aminoácidos de BGL7.

[0088] Del mismo modo, los polinucleótidos descritos en este documento incluyen secuencias que
25 codifican proteínas BGL7 nativas y variantes de corte y empalme de las mismas, secuencias complementarias de la secuencia de codificación de la proteína nativa y nuevos fragmentos de polinucleótidos que codifican BGL7. Una secuencia de ácido nucleico que codifica BGL7 puede contener una o más secuencias de intrones si se trata de una secuencia de ADN genómico.

[0089] Una secuencia de nucleótidos que codifica BGL7 puede tener al menos el 70%,
30 preferentemente el 80%, 85%, 90%, 95%, 98% o mayor porcentaje de identidad de secuencia con la secuencia de codificación de *bgl7* presentada en este documento como la SEQ ID N°:1.

[0090] En otro ejemplo, una secuencia de nucleótidos que codifica BGL7 hibridará en condiciones de moderada a alta astringencia con una secuencia de nucleótidos que codifica una proteína BGL7. En un ejemplo relacionado, una secuencia de nucleótidos que codifica BGL7 hibridará en condiciones de
35 moderada a alta astringencia con la secuencia de nucleótidos presentada como la SEQ ID N°:1.

[0091] Se aprecia que algunas variantes de la secuencia de ácidos nucleicos que codifican BGL7 pueden hibridarse o no selectivamente con la secuencia original. A modo de ejemplo, en situaciones en las que la secuencia de codificación ha sido optimizada basándose en la degeneración del código genético, se puede producir una variante de la secuencia de codificación que codifica una proteína
40 BGL7, pero que no hibrida con una secuencia nativa de ácido nucleico que codifica BGL7 en condiciones de moderada a alta astringencia. Lo que ocurriría, por ejemplo, cuando la variante de la

secuencia incluye un codón diferente para cada uno de los aminoácidos codificados por el nucleótido original.

[0092] Como entenderán mejor los expertos en la materia, en algunos casos, puede ser una ventaja producir secuencias de nucleótidos que poseen codones de origen no natural, por ejemplo, inosina u otros análogos de nucleótidos de origen no natural. Pueden seleccionarse los codones preferidos por un huésped eucariota concreto, por ejemplo, para aumentar la tasa de expresión de la proteína BGL7 o para producir transcritos de ARN recombinante que tengan las propiedades deseadas, como una vida media más larga que la de los transcritos producidos a partir de una secuencia de origen natural. Por tanto, puede diseñarse una secuencia de nucleótidos que codifican BGL7 nativa con el fin de alterar la secuencia de codificación por diversas razones, incluidas, sin carácter limitativo, alteraciones que modifican la clonación, procesamiento y/o expresión de la proteína BGL7 por una célula.

[0093] En particular, se prefieren las sustituciones, adiciones y eliminaciones del ácido nucleico que son silenciosas, de manera que no alteran las propiedades o actividades del polipéptido o polinucleótido nativo.

[0094] Las variaciones pueden realizarse utilizando métodos conocidos en la técnica como mutagénesis mediada por oligonucleótido (dirigida al sitio) y mutagénesis por PCR. La mutagénesis dirigida al sitio (Carter *et al.*, 1986; Zoller *et al.*, 1987), mutagénesis por inserción de un casete (Wells *et al.*, 1985), mutagénesis por selección de restricción (Wells *et al.*, 1986) u otras técnicas conocidas pueden llevarse a cabo en el ADN clonado para producir la variante de ADN que codifica el polipéptido BGL7.

[0095] No obstante, en algunos casos, puede ser una ventaja expresar las variantes de *bg/7* que no tienen las propiedades o actividades del polinucleótido *bg/7* nativo o del polipéptido BGL7. En estos casos, se pueden generar formas modificadas o mutantes de la secuencia nativa de ácido nucleico que codifica BGL7, usando técnicas empleadas habitualmente por los expertos en la técnica.

B. Polipéptidos BGL7

[0096] En el presente documento, se describe un polipéptido BGL7, que tiene una secuencia del polipéptido BGL7 de longitud completa o madura nativa, que consta de una secuencia presentada como SEQ ID N°:2. Un polipéptido de la invención puede ser el polipéptido BGL7 maduro, parte de una proteína de fusión o un fragmento o variante de la secuencia del polipéptido BGL7 presentada como SEQ ID N°:2 como se define en las reivindicaciones.

[0097] Por lo general, un polipéptido BGL7 tiene una identidad de al menos el 85% con una secuencia de aminoácidos de BGL7 en toda su longitud. Son más preferidas las secuencias de polipéptidos BGL7 compuestas de una región que presenta una identidad de secuencia de al menos el 85, 90, 95, 98% o un porcentaje mayor con la secuencia del polipéptido BGL7 presentada como la SEQ ID N°:2, usando un programa de alineamiento de secuencias, como se detalla en el presente documento.

[0098] Normalmente, una “forma modificada de una proteína BGL7 nativa o una “variante” de proteína BGL7 tiene una secuencia derivada compuesta de al menos una sustitución, adición, eliminación o inserción de aminoácidos, respectivamente.

[0099] Se conoce bien en la técnica que pueden realizarse determinadas sustituciones de aminoácidos en secuencias de proteínas sin afectar a la función de la proteína. En general, se toleran las sustituciones de aminoácidos conservadoras o sustituciones de aminoácidos similares sin afectar a la función de la proteína. Los aminoácidos similares pueden ser los que son similares en propiedades de tamaño y/o carga, por ejemplo, el aspartato y glutamato, y la isoleucina y valina, son dos pares de aminoácidos similares. La similitud entre pares de aminoácidos se ha estudiado en la técnica de varios modos. Por ejemplo, Dayhoff *et al* (1978), que se incluye como referencia en este documento, proporcionan tablas de frecuencias para las sustituciones de aminoácidos que pueden emplearse como medida de similitud de aminoácidos. Las tablas de frecuencia de Dauhoff *et al.* se basan en comparaciones de secuencias de aminoácidos para proteínas que tienen la misma función a partir de una diversidad de fuentes evolutivamente distintas.

[0100] También se describen en el presente documento los fragmentos y variantes de la secuencia de polipéptido BGL7 presentada como la SEQ ID N°:2. Un fragmento es una variante de polipéptido que tiene una secuencia de aminoácidos que es completamente la misma como parte, pero no es toda la secuencia de aminoácidos de los polipéptidos descritos anteriormente. Los fragmentos pueden ser “independientes” o estar comprendidos en un polipéptido más largo cuyo fragmento forma una parte o una región, más preferiblemente como una única región continua. Los fragmentos preferidos son fragmentos biológicamente activos que son los fragmentos que median actividades de los polipéptidos, entre los que se incluyen aquellos con actividad similar o mejorada o con una menor actividad. También se incluyen aquellos fragmentos que son antigénicos o inmunogénicos en un animal, particularmente en un ser humano. En este aspecto, se describe en la presente invención: (i) fragmentos de BGL7, preferentemente de al menos aproximadamente 20-100 aminoácidos de longitud, más preferentemente aproximadamente 100-200 aminoácidos de longitud y (ii) una composición farmacéutica que contiene BGL7. En varios modos de realización, el fragmento se corresponde con el dominio del extremo N-terminal de BGL7 o con el dominio de extremo C-terminal de BGL7.

[0101] Los polipéptidos BGL7 también incluyen polipéptidos que varían de la secuencia del polipéptido BGL7 de la Figura 2. Estas variantes pueden ser variantes de sustitución, inserción o eliminación. Las variantes normalmente presentan la misma actividad biológica cualitativa que el análogo de origen natural, aunque también se pueden seleccionar las variantes que tienen características modificadas como se describe con más adelante a continuación.

[0102] Una “sustitución” es el resultado del reemplazo de uno o más nucleótidos o aminoácidos por diferentes nucleótidos o aminoácidos, respectivamente.

[0103] Una “inserción” o “adición” es el cambio en una secuencia de aminoácidos o nucleótidos que ha dado como resultado la adición de uno o más residuos de aminoácidos o nucleótidos, respectivamente, en comparación con la secuencia de origen natural.

[0104] Una “eliminación” se define como un cambio en la secuencia de nucleótidos o aminoácidos en el que uno o más nucleótidos o restos de aminoácidos, respectivamente, están ausentes.

[0105] Las sustituciones de aminoácidos son normalmente de restos individuales; las inserciones generalmente serán de aproximadamente 1 a 20 aminoácidos, aunque se tolerarán inserciones considerablemente mayores. Las eliminaciones varían de aproximadamente 1 a aproximadamente 20 restos, aunque en algunos casos las eliminaciones pueden ser mucho mayores.

5 **[0106]** Pueden utilizarse sustituciones, eliminaciones, inserciones o cualquier combinación de las mismas para obtener un derivado final. En general, estos cambios se llevan a cabo en unos pocos aminoácidos para minimizar la alteración de la molécula. Sin embargo, pueden tolerarse cambios mayores en algunas circunstancias.

10 **[0107]** Las sustituciones de aminoácidos pueden ser el resultado de sustituir un aminoácido con otro aminoácido que tiene propiedades químicas y/o estructurales similares, como la sustitución de una isoleucina por una valina, es decir, sustituciones de aminoácidos conservativas. Las inserciones o eliminaciones pueden estar opcionalmente en el intervalo de 1 a 5 aminoácidos.

15 **[0108]** Las sustituciones se realizan generalmente según las “sustituciones conservativas” conocidas. Una “sustitución conservativa” hace referencia a la sustitución de un aminoácido de una clase por un aminoácido de la misma clase, donde una clase se define como las propiedades fisicoquímicas comunes de la cadena lateral de aminoácidos y frecuencias de sustitución altas en proteínas homólogas encontradas en la naturaleza (como se determina, por ejemplo, con una matriz de intercambio de frecuencia de Dayhoff estándar o matriz BLOSUM). (En general, véase, Doolittle, R.F., 1986).

20 **[0109]** Una “sustitución no conservativa” se refiere a la sustitución de un aminoácido de una clase por un aminoácido de otra clase.

25 **[0110]** Normalmente, las variantes de polipéptidos BGL7 presentan la misma actividad biológica cualitativa que el análogo de origen natural, aunque también se seleccionan las variantes para modificar las características del polipéptido BGL7, según sea necesario. Por ejemplo, se pueden alterar o eliminar los sitios de glucosilación unidos a O o unidos a N. Los expertos en la materia apreciarán que los cambios de aminoácidos pueden alterar los procesos postraduccionales del polipéptido BGL7, como cambiar el número o posición de los sitios de glucosilación o alterar las características de anclaje de la membrana o las características de secreción u otras características de localización celular.

30 **[0111]** En la definición de polipéptidos BGL7 también se incluyen otros polipéptidos BGL7 relacionados. Por tanto, las secuencias de sondas o cebadores de la reacción en cadena de la polimerasa (PCR) degeneradas se pueden utilizar para encontrar otros polipéptidos relacionados. Se pueden diseñar secuencias de cebadores o sondas útiles para: la totalidad o parte de la secuencia del polipéptido BGL7, o secuencias de fuera de la región de codificación. Como se conoce
35 generalmente en la técnica, los cebadores de la PCR preferidos son de aproximadamente 15 a aproximadamente 35 nucleótidos de longitud, más preferentemente de aproximadamente 20 a aproximadamente 30 y puede contener inosina si es necesario. Las condiciones para la reacción por PCR se conocen generalmente en la técnica.

40 **[0112]** También se describen modificaciones covalentes de polipéptidos BGL7. Por ejemplo, en el presente documento se describen polipéptidos BGL7 que son una proteína madura y pueden constar de aminoácidos de extremo amino o carboxilo terminal, o aminoácidos en el polipéptido maduro (por

ejemplo, cuando la forma madura de la proteína tiene más de una cadena de polipéptidos). Dichas secuencias pueden, por ejemplo, desempeñar una función en el procesamiento de la proteína desde un precursor a una forma madura, permitir el transporte de proteína, acortar o prolongar la vida media de la proteína o facilitar la manipulación de la proteína en ensayos o en la producción.

5 **[0113]** También se contemplan las modificaciones dirigidas a alterar un sitio activo, alterar el pH óptimo, la temperatura óptima y/o la afinidad del sustrato de la enzima BGL7.

[0114] La Figura 2 muestra la secuencia de aminoácidos predicha de un ejemplo de polipéptido BGL7 basado en la secuencia de nucleótidos proporcionada en la Figura 1. El peso molecular predicho del polipéptido maduro BGL7 codificado es 83 kDa. Una secuencia que se parece a un péptido señal
10 (Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., *Protein Engineering*, 10:1-6, 1997) está presente en el extremo amino de BGL7, lo que sugiere que se secreta el polipéptido BGL7.

[0115] Se llevó a cabo una búsqueda con BLASTP Basic (<http://www.ncbi.nlm.nih.gov/BLAST>) de la base de datos de proteínas no redundantes el 18 de octubre de 2002 con la secuencia de aminoácidos de BGL7 que indicó una identidad de secuencia del 51% con el número de acceso del
15 GenBank U16259 (precursor de β -glucosidasa de *Candida molischiana*), identidad de secuencia del 41% con el número de acceso del GenBank U24701 (β -1,2-glucosidasa de *Septoria lycopersici*), identidad de secuencia del 41% con el número de acceso del GenBank U35463 (avenacinasa de *Gaeumannomyces graminis*), identidad de secuencia del 42% con el número de acceso del GenBank U09580 (glucohidrolasa β -D-glucosidasa de *Hypocrea jecorina*), y una identidad de secuencia del
20 44% con el número de acceso D64088 (precursor de β -glucosidasa 1 *Aspergillus aculeatus*). Las diez secuencias que tienen la mayor identidad pero menor de 51% de identidad con BGL7 se anotaron todas como β -glucosidasas o sinónimos de β -glucosidasas. Estas similitudes de secuencia indican que BGL7 forma parte de la familia 3 de las glucosil hidrolasas (Henrissat, B. y Balroch, A. (1993) *Biochem. J.* 293:781-788).

25 C. Anticuerpos anti-BGL7

[0116] También se describen anticuerpo anti-BGL7. Los anticuerpos pueden ser anticuerpos policlonales, monoclonales, humanizados, biespecíficos o heteroconjugados.

[0117] Los profesionales en la materia conocen los métodos de preparación de anticuerpos policlonales. El agente inmunizante puede ser un polipéptido BGL7 o una proteína de fusión del
30 mismo. Puede ser útil conjugar el antígeno con una proteína que se sabe que es inmunogénica en el mamífero que se está inmunizando. El protocolo de inmunización lo puede determinar un experto en la técnica basándose en protocolos estándares o experimentación rutinaria.

[0118] Como alternativa, los anticuerpos anti-BGL7 pueden ser anticuerpos monoclonales. Los anticuerpos monoclonales se pueden producir mediante células inmunizadas en un animal o usando
35 métodos de ADN recombinante. (Véanse, p.ej., Kohler *et al.*, 1975; Patente de EE. UU. N° 4.816.567).

[0119] Un anticuerpo anti-BGL7 también puede comprender un anticuerpo humanizado o humano. El término "anticuerpo humanizado" se refiere a formas humanizadas de anticuerpos no humanos (p.ej., murinos) que son anticuerpos quiméricos, cadenas de inmunoglobulina o fragmentos de las mismas (como Fv, Fab, Fab', F(ab')₂ u otras secuencias parciales de anticuerpos de unión al antígeno) que

contienen alguna parte de la secuencia derivada de un anticuerpo no humano. Los métodos para la humanización de anticuerpos no humanos son bien conocidos en la técnica, como se especifica en Jones *et al.*, 1986; Riechmann *et al.*, 1988; Verhoeyen *et al.*, 1988. Los métodos para la producción de anticuerpos humanos también son conocidos en la técnica. Véanse, p.ej., Jakobovits, A., *et al.*, 1995 y Jakobovits, A, 1995.

VI. Expresión de BGL7 recombinante

[0120] Los métodos descritos en el presente documento se basan en el uso de células para expresar BGL7, sin que se requiera un método particular de expresión de BGL7.

[0121] En este documento se describen células huésped que se han transducido, transformado o transfectado con un vector de expresión que comprende una secuencia de ácido nucleico que codifica BGL7. Las condiciones de cultivo, como la temperatura, pH y similares, son las utilizadas anteriormente para la célula huésped original antes de la transducción, transformación o transfección y serán evidentes para los expertos en la materia.

[0122] En una estrategia, una célula de hongos filamentosos o célula de levadura se transfecta con un vector de expresión que tiene un promotor o fragmento de promotor biológicamente activo o uno o más (p.ej., una serie de) potenciadores que funcionan en la línea celular huésped, unidos de modo funcional a un segmento de ADN que codifica BGL7, de manera que BGL7 se expresa en la línea celular.

A. Vectores de expresión/Construcciones de ácido nucleico

[0123] Se pueden incorporar fragmentos de polinucleótidos sintéticos o naturales que codifican BGL7 ("secuencias de ácido nucleico que codifican BGL7") en construcciones o vectores de ácido nucleico heterólogos, capaces de la introducción y la replicación en una célula de hongo filamentoso o una célula de levadura. Los vectores y métodos descritos aquí son útiles para usar en células huésped para la expresión de BGL7. Se puede usar cualquier vector siempre y cuando sea replicable y viable en las células en las que se introduce. Los expertos en la materia conocen un gran número de vectores y promotores adecuados, y estos están disponibles en el mercado. Los vectores de clonación y expresión también se describen en Sambrook *et al.*, 1989, Ausubel FM *et al.*, 1989 y Strathern *et al.*, 1981. Los vectores de expresión apropiados para los hongos se describen en van den Hondel, C.A.M.J.J. *et al.* (1991) en: Bennett, J.W. y Lasure, L.L. (eds.) *More Gene Manipulations in Fungi*. Academic Press, págs. 396-428. La secuencia de ADN apropiada puede insertarse en un plásmido o vector (en el presente documento denominados "vectores") mediante una variedad de procedimientos. En general, la secuencia de ADN se inserta en un sitio o sitios de endonucleasas de restricción apropiados por procedimientos estándares. Estos procedimientos y los de subclonación relacionados se consideran parte de los conocimientos de los expertos en la materia.

[0124] Los hongos filamentosos recombinantes compuestos por una secuencia de codificación para BGL7 pueden producirse por la introducción de una construcción de ácido nucleico heteróloga que consta de la secuencia que codifica BGL7 en las células de una cepa seleccionada de los hongos filamentosos.

[0125] Una vez que se obtiene la forma deseada de una secuencia de ácido nucleico de *bgl7*, homólogo, variante o fragmento de la misma, se puede modificar en una variedad de maneras. Cuando la secuencia implica regiones flanqueadoras no codificantes, las regiones flanqueadoras pueden someterse a resección, mutagénesis, etc. De este modo, pueden realizarse transiciones, transversiones, eliminaciones e inserciones en la secuencia de origen natural.

[0126] Puede insertarse una secuencia que codifica *bgl7* seleccionada en un vector adecuado de acuerdo con técnicas recombinantes conocidas y utilizarse para transformar los hongos filamentosos capaces de expresar BGL7. Debido a la degeneración inherente del código genético, pueden utilizarse otras secuencias de ácidos nucleicos que codifican sustancialmente la misma secuencia de aminoácidos o una secuencia funcionalmente equivalente para clonar y expresar BGL7. Por tanto, se entiende que dichas sustituciones en la región de codificación están dentro de las variantes de secuencia descritas en la presente invención. Todas y cada una de las variantes de secuencia pueden utilizarse de la misma forma descrita en el presente documento para una secuencia original de ácido nucleico que codifica BGL7.

[0127] En este documento, también se describen construcciones de ácido nucleico recombinante que constan de una o más secuencias de ácido nucleico que codifican BGL7, tal como se describe anteriormente. Las construcciones están compuestas de un vector, como un plásmido o vector viral, en el que se ha insertado una secuencia de la invención, en una orientación directa o inversa.

[0128] Las construcciones de ácido nucleico heterólogas pueden incluir la secuencia que codifica *bgl7*, o una variante, un fragmento o variante de empalme de la misma: (i) en aislamiento, (ii) en combinación con secuencias de codificación adicionales; como secuencias de proteína de fusión o secuencias de codificación de péptido señal, en las que la secuencia que codifica *bgl7* es la secuencia de codificación dominante; (iii) en combinación con secuencias no codificantes, como intrones y elementos de control, como elementos de promotor y de terminación o regiones 5' o 3' no traducidas, eficaces para la expresión de la secuencia de codificación en un huésped apropiado; y/o (iv) en un vector o medio de huésped en el que la secuencia que codifica *bgl7* es un gen heterólogo.

[0129] En la presente invención se describe una construcción de ácido nucleico heteróloga que se utiliza para transferir una secuencia de ácido nucleico que codifica BGL7 en una célula *in vitro*, pero se prefieren las líneas de levaduras y hongos filamentosos establecidas. Para la producción de alto rendimiento de BGL7 a largo plazo, se prefiere la expresión estable. De lo que se deduce que se puede utilizar cualquier método eficaz para generar transformantes estables.

[0130] Los vectores apropiados están normalmente equipados con una secuencia de ácido nucleico que codifica un marcador seleccionable, sitios de inserción y elementos de control adecuados, como secuencias de promotores y de terminación. El vector puede comprender secuencias reguladoras, incluyendo, por ejemplo, secuencias no codificantes, como intrones y elementos de control, es decir, elementos promotores y de terminación o regiones 5' y/o 3' no traducidas, eficaces para la expresión de la secuencia de codificación en las células huésped (y/o en un vector o medio de células huésped en el que una secuencia de codificación de antígeno de proteína soluble modificada no se expresa normalmente), unida de modo funcional a la secuencia de codificación. Los expertos en la técnica conocen un gran número de vectores y promotores adecuados, muchos de los cuales están disponibles en el mercado y/o se describen en Sambrook, *et al.*, (*supra*).

[0131] Entre los promotores de ejemplo se incluyen promotores constitutivos y promotores inducibles, cuyos ejemplos incluyen un promotor de CMV, un promotor temprano de SV40, un promotor de RSV, un promotor de EF-1 α , un promotor que contiene un elemento de respuesta a tet (TRE) en el sistema tet-on o tet-off, como se ha descrito (ClonTech y BASF), el promotor de β -actina y el promotor de metalotionina que puede estimularse mediante la adición de determinadas sales de metales. Una secuencia promotora es una secuencia de ADN que es reconocida por el hongo filamentoso particular para los propósitos de expresión. Está unida de modo funcional a una secuencia de ADN que codifica un polipéptido BGL7. Dicha unión comprende la colocación del promotor en relación con el codón de iniciación de la secuencia de ADN que codifica el polipéptido BGL7 en los vectores de expresión descritos. La secuencia promotora contiene secuencias de control de la transcripción y traducción que median la expresión del polipéptido BGL7. Entre los ejemplos se incluyen los promotores de los genes que codifican glucoamilasa, alfa-amilasa o alfa-glucosidasa de *Aspergillus niger*, *A. awamori* o *A. oryzae*; los genes *gpdA* o *trpC* de *A. nidulans*; los genes *cbh1* o *trp1* de *Neurospora crassa*; los genes que codifican la aspártico proteinasa de *A. niger* o *Rhizomucor miehei*, los genes *cbh1*, *cbh2*, *egl1*, *egl2* de *T. reesei*, u otros genes que codifican celulasas.

[0132] La elección del marcador seleccionable adecuado dependerá de la célula huésped, y los marcadores adecuados para los diferentes huéspedes son bien conocidos en la técnica. Los genes de marcadores seleccionables típicos incluyen *argB* de *A. nidulans* o *T. reesei*, *amdS* de *A. nidulans*, *pyr4* de *Neurospora crassa* o *T. reesei*, *pyrG* de *Aspergillus niger* o *A. nidulans*. Entre los marcadores seleccionables de ejemplo se incluyen, sin carácter limitativo, *trpc*, *trp1*, *oliC31*, *niaD* o *leu2*, que se incluyen en construcciones de ácido nucleico heterólogas utilizadas para transformar una cepa mutante como *trp-*, *pyr-*, *leu-* y similares.

[0133] Dichos marcadores seleccionables confieren a los transformantes la capacidad de utilizar un metabolito que normalmente no es metabolizado por los hongos filamentosos. Por ejemplo, el gen *amdS* de *T. reesei* que codifica la enzima acetamidasa que permite que las células transformantes se desarrollen en acetamida como fuente de nitrógeno. El marcador seleccionable (p.ej., *pyrG*) puede restablecer la capacidad de una cepa mutante auxotrófica para crecer en un medio mínimo selectivo o el marcador seleccionable (p.ej. *olic31*) puede conferir a los transformantes la capacidad de crecer en presencia de un fármaco o antibiótico inhibidor.

[0134] La secuencia que codifica el marcador seleccionable se clona en cualquier plásmido adecuado utilizando métodos empleados generalmente en la técnica. Entre los plásmidos de ejemplo se incluyen: pUC18, pBR322 y pUC100.

[0135] La práctica de la presente invención empleará, a no ser que se indique lo contrario, técnicas convencionales de biología molecular, microbiología, ADN recombinante e inmunología, que son parte de la técnica. Estas técnicas se explican con detalle en la bibliografía. Véanse, por ejemplo, Sambrook *et al.*, 1989; Freshney, 1987; Ausubel, *et al.*, 1993; y Coligan *et al.*, 1991.

B. Células huésped y condiciones de cultivo para la producción mejorada de BGL7

(i) Hongos filamentosos

[0136] De este modo, la presente invención proporciona hongos filamentosos compuestos de células que se han modificado, seleccionado y cultivado de un modo efectivo para dar lugar a una producción o expresión mejorada de BGL7 con respecto a los correspondientes hongos originales no transformados.

5 **[0137]** Entre los ejemplos de especies de hongos filamentosos originales que se pueden tratar y o modificar para mejorar la expresión de BGL7 se incluyen, sin carácter limitativo, *Trichoderma*, p.ej, *Trichoderma reesei*, *Trichoderma longibrachiatum*, *Trichoderma viride*, *Trichoderma koningii*; *Penicillium sp.*, *Humicola sp.*, incluidas *Humicola insolens*; *Aspergillus sp.*, *Crysosporium sp.*, *Fusarium sp.*, *Hypocrea sp.* y *Emericella sp.*

10 **[0138]** Las células que expresan BGL7 se cultivan en condiciones que se emplean habitualmente para el cultivo de la línea fúngica original. En general, las células se cultivan en un medio estándar que contiene sales fisiológicas y nutrientes, como se describe en Pourquie, J. *et al.*, *Biochemistry and Genetics of Cellulose Degradation*, eds. Aubert, J.P. *et al.*, Academic press, págs. 71-86, 1988 y Ilmen, M. *et al.*, *Appl. Environ. Microbiol.* 63:1298-1306, 1997. Las condiciones de cultivo son también
15 estándares, por ejemplo, los cultivos se incuban a 28°C en cultivos con agitación o fermentadores hasta lograr los niveles deseados de expresión de BGL7.

[0139] Las condiciones de cultivo preferidas para un hongo filamentoso determinado se pueden encontrar en la bibliografía científica y/o en la fuente de hongos como la American Type Culture Collection (ATCC; "<http://www.atcc.org/>"). Después de establecer el crecimiento fúngico, se exponen
20 las células a condiciones eficaces para producir o permitir la sobreexpresión de BGL7.

[0140] En los casos en los que una secuencia que codifica BGL7 está controlada por un promotor inducible, el agente de inducción, p.ej., un azúcar, sal de metal o antibióticos, se añade al medio en una concentración eficaz para inducir un nivel alto de expresión de BGL7.

(ii) Levadura

25 **[0141]** También se describe el uso de la levadura como célula huésped para la producción de BGL7. Se han expresado muchos otros genes que codifican enzimas hidrolíticas en diversas cepas de la levadura *S. cerevisiae*. Lo que incluye secuencias que codifican dos endoglucanasas (Penttila *et al.*, 1987), dos celobiohidrolasas (Penttila *et al.*, 1988) y una β -glucosidasa de *Trichoderma reesei* (Cummings y Fowler, 1996), una xilanasa de *Aureobasidium pullulans* (Li y Ljungdahl, 1996), una
30 alfa-amilasa de trigo (Rothstein *et al.*, 1987), etc. Además, se expresó con éxito un casete de genes de celulasa que codifica la endo- $[\beta]$ -1,4-glucanasa (END1) de *Butyrivibrio fibrisolvens*, celobiohidrolasa (CBH1) de *Phanerochaete chrysosporium*, la celodextrinasa (CEL1) de *Ruminococcus flavefaciens* y la celobiasa (Bgl1) de *Endomyces fibrilizer*, en una cepa de laboratorio de *S. cerevisiae* (Van Rensburg *et al.*, 1998).

35 C. Introducción de una secuencia de ácido nucleico que codifica BGL7 en células huésped

[0142] La invención también describe células y composiciones celulares que han sido genéticamente modificadas para que comprendan una secuencia de ácido nucleico que codifica BGL7 proporcionada de modo exógeno. Se puede modificar una célula o línea celular original genéticamente (es decir,

transducir, transformar o transfectar) con un vector de clonación o un vector de expresión. El vector puede tener forma, por ejemplo, de un plásmido, una partícula viral, un fago, etc., como se ha descrito con más detalle anteriormente.

[0143] Se pueden utilizar varios métodos para suministrar un vector de expresión en células *in vitro*.

5 Después de construir un vector de expresión adecuado, se utiliza para transformar cepas de hongos o levaduras. Los profesionales conocen los métodos generales para introducir ácidos nucleicos en células para expresar secuencias de ácido nucleico heterólogas. Entre estos métodos se incluyen, sin carácter limitativo: electroporación; microinyección nuclear o microinyección directa en células simples; fusión de protoplastos bacterianos con células intactas; uso de policonaciones, por ejemplo, 10 polibreno o poliornitina; fusión de membrana con liposomas, transfección mediada por lipofectamina, o lipofección; bombardeo de alta velocidad con microproyectiles recubiertos de ADN; incubación con precipitación del ADN-fosfato de calcio; transfección mediada con DEAE-dextrano; infección con ácidos nucleicos virales modificados y similares.

[0144] Los métodos preferidos para introducir una construcción de ácido nucleico heteróloga (vector 15 de expresión) en hongos filamentosos (p.ej., *T. reesei*) incluyen, sin carácter limitativo, el uso de una pistola de partículas o genes, permeabilización de paredes celulares de hongos filamentosos antes del proceso de transformación (por ejemplo, mediante el uso de concentraciones altas de álcali, p.ej., de 0,05 M a 0,4 M de CaCl₂ o acetato de litio), fusión de protoplastos o transformación mediada por agrobacterias. Un método de ejemplo para transformar hongos filamentosos mediante el tratamiento 20 de protoplastos o esferoplastos con polietilenglicol y CaCl₂ se describe en Campbell, E.I. *et al.*, *Curr. Genet.* 16:53-56, 1989 y Penttila, M. *et al.*, *Gene*, 63:11-22, 1988.

[0145] Además, las construcciones de ácido nucleico heterólogas que constan de una secuencia de ácido nucleico que codifica BGL7 pueden transcribir *in vitro*, e introducir el ARN resultante en una célula huésped por métodos conocidos, por ejemplo, mediante inyección.

25 **[0146]** Después de introducir una construcción de ácido nucleico heteróloga compuesto por la secuencia que codifica *bgl7*, las células genéticamente modificadas se pueden cultivar en un medio de nutrientes convencionales modificados según sea apropiado para activar los promotores, seleccionar transformantes o amplificar la expresión de una secuencia de ácido nucleico que codifica BGL7. Las condiciones de cultivo, como la temperatura, pH y similares, son las que se han utilizado 30 previamente para la célula huésped seleccionada para la expresión, y serán evidente para aquellos expertos en la técnica.

[0147] Se considera que la progenie de las células en las que estas construcciones de ácidos nucleicos heterólogas se han introducido está compuesto por la secuencia de ácido nucleico que codifica BGL7 en la construcción de ácido nucleico heterólogo.

35 **[0148]** También se describen transformantes novedosos y útiles de hongos filamentosos como *Trichoderma reesei* para su uso en la producción de composiciones de celulasa fúngica. Por ejemplo, se describen transformantes de hongos filamentosos, especialmente hongos que comprenden la secuencia que codifica *bgl7*, que comprenden una forma modificada de la secuencia que codifica *bgl7* o la eliminación de la secuencia que codifica *bgl7*.

40 **[0149]** Los transformantes estables de los hongos filamentosos generalmente pueden distinguirse de los transformantes inestables por su velocidad de crecimiento más rápida y la formación de colonias

circulares con un perfil liso más que rugoso en el medio de cultivo sólido. Además, en algunos casos, puede llevarse a cabo un ensayo adicional de estabilidad mediante el cultivo de los transformantes en medio sólido no selectivo, la recolección de las esporas de este medio de cultivo y la determinación del porcentaje de estas esporas que germinarán posteriormente y crecerán en un medio selectivo.

5 VII. Análisis de la expresión de proteína y/o secuencias que codifican el ácido nucleico de BGL7

[0150] Con el fin de evaluar la expresión de BGL7 por una línea celular que se ha transformado con una construcción de ácido nucleico que codifica BGL7, pueden realizarse ensayos a nivel de proteínas, a nivel del ARN o mediante el uso de bioensayos funcionales en particular de la actividad y/o producción de glucosidasa.

10 **[0151]** En una aplicación de ejemplo de las secuencias de ácido nucleico de *bgl7* descritas en la presente invención, se diseña una cepa genéticamente modificada de hongos filamentosos, p.ej., *Trichoderma reesei*, para producir una mayor cantidad de BGL7. Estos hongos filamentosos genéticamente modificados serían útiles para producir un producto de celulasa con capacidad
15 celulolítica aumentada. En un enfoque, esto se logra mediante la introducción de la secuencia de codificación para *bgl7* en un huésped adecuado, por ejemplo, un hongo filamentosos como *Trichoderma reesei*.

[0152] Por consiguiente, la invención incluye procedimientos para expresar BGL7 en un hongo filamentosos u otro huésped adecuado, mediante la introducción de un vector de expresión que contiene la secuencia de ADN que codifica BGL7 en células del hongo filamentosos u otro huésped
20 adecuado. También se describen métodos para modificar la expresión de BGL7 en un hongo filamentosos u otro huésped adecuado. Esta modificación incluye una reducción o eliminación de la expresión, o de la expresión de una forma alterada de BGL7. Una forma alterada de BGL7 puede tener una secuencia de aminoácidos alterada o una secuencia de ácido nucleico alterada.

[0153] En general, entre los ensayos empleados para analizar la expresión de BGL7 se incluyen:
25 transferencia Northern blot, dot blot (análisis de ADN o ARN), RT-PCR (reacción en cadena de la polimerasa con transcriptasa inversa) o hibridación *in situ*, utilizando una sonda marcada de manera apropiada (basada en la secuencia que codifica el ácido nucleico) y transferencia Southern blot y autoradiografía convencionales.

[0154] Además, la producción y/o la expresión de BGL7 se pueden medir en una muestra
30 directamente, por ejemplo, mediante ensayos de la actividad, expresión y/o producción de la glucosidasa. Dichos ensayos se describen, por ejemplo, en Chen *et al.*, (1992), Herr *et al.*, (1978) y la Patente de EE. UU. Nº: 6.184.018 (Li *et al.*; 2001). La capacidad de BGL7 para hidrolizar sustratos aislados solubles e insolubles puede medirse con el uso de los ensayos descritos en Suumakki *et al.* (2000) y Ortega *et al.* (2001). Los sustratos útiles para ensayar las actividades de celobiohidrolasa, endoglucanasa o β -glucosidasa incluyen celulosa cristalina, papel de filtro, celulosa hinchada por ácido fosfórico, hidroxietilcelulosa, carboximetilcelulosa, celooligosacáridos, metilumbeliferil lactósido, metilumbeliferil celobiósido, ortonitrofenil lactósido, ortonitrofenil celobiósido,
35 paranitrofenil celobiósido, ortonitrofenil glucósido, paranitrofenil glucósido y metilumbeliferil glucósido.

Los tres últimos son particularmente útiles para ensayar β -glucosidasas, Los ensayos de β -glucosidasas son muy conocidos en la técnica. Véase Cummings y Fowler (1996).

[0155] Además, la expresión de proteínas se puede evaluar con métodos inmunológicos, como la tinción inmunohistoquímica de células, secciones de tejido o inmunoensayo de medio de cultivo de tejidos, por ejemplo, por análisis Western blot o ELISA. Pueden utilizarse estos ensayos para evaluar cualitativa y cuantitativamente la expresión de BGL7. Los detalles de dichos métodos son conocidos por aquellos expertos en la materia y muchos reactivos para la práctica de dichos métodos están disponibles en el mercado

[0156] Se puede usar una forma purificada de BGL7 para producir anticuerpos monoclonales o policlonales específicos para la proteína expresada para su uso en varios inmunoensayos. (Véase, p.ej., Hu *et al.*, 1991). Los ensayos de ejemplo incluyen ELISA, inmunoensayos competitivos, radioinmunoensayos, Western blot, ensayos de inmunofluorescencia indirecta y similares. En general, se pueden utilizar anticuerpos y/o kits disponibles comercialmente para el inmunoensayo cuantitativo del nivel de expresión de las proteínas glucosidasa.

VIII. Aislamiento y purificación de proteína BGL7 recombinante

[0157] En general, una proteína BGL7 producida en cultivo celular es secretada en el medio y se puede purificar o aislar, por ejemplo, mediante la eliminación de los componentes no deseados del medio de cultivo celular. No obstante, en algunos casos, se puede producir una proteína BGL7 en una forma celular que necesita la recuperación de un lisado celular. En dichos casos, la proteína BGL7 se purifica a partir de las células en las que se produjo utilizando técnicas habitualmente empleadas por los expertos en la técnica. Entre los ejemplos se incluyen, sin carácter limitativo, cromatografía de afinidad (Tilbeurgh *et al.* 1984), métodos de cromatografía de intercambio iónico (Goyal *et al.*, 1991; Fliess *et al.*, 1983; Bhikhabhai *et al.*, 1984; Ellouz *et al.*, 1987), incluyendo materiales que usan el intercambio iónico con alto poder de resolución (Medve *et al.*, 1998), cromatografía de interacción hidrófoba (Tomaz and Queiroz, 1999) y separación en dos fases (Brumbauer, *et al.*, 1999).

[0158] Normalmente, la proteína BGL7 se fracciona para segregar proteínas que tienen las propiedades seleccionadas, tales como la afinidad de unión a agentes de unión particulares, p.ej., anticuerpos o receptores; o que tienen un intervalo de peso molecular seleccionado o un intervalo de puntos isoeléctricos.

[0159] Una vez lograda la expresión de una proteína BGL7 determinada, la proteína BGL7 producida de este modo es purificada a partir de las células o cultivo celular. Los procedimientos de ejemplo adecuados para esta purificación incluyen los siguientes: cromatografía de columna de afinidad con anticuerpos, cromatografía de intercambio iónico; precipitación con etanol; HPLC en fase inversa; cromatografía en sílice o en una resina de intercambio iónico, como DEAE; cromatografía en SDS-PAGE; precipitación con sulfato de amonio y filtración en gel empleando, por ejemplo, Sephadex G-75. Se pueden utilizar diversos métodos de purificación de proteínas y estos métodos son conocidos en la técnica y descritos, p.ej., en Deutscher, 1990; Scopes 1982. La etapa o etapas de purificación

seleccionadas dependerán, por ejemplo, de la naturaleza del proceso de producción utilizado y de la proteína concreta producida.

IX. Utilidad de *bgl7* y BGL7

[0160] Se puede observar que el nucleótido *bgl7*, la proteína BGL7 y las composiciones que contienen actividad de la proteína BGL7 son útiles en una amplia variedad de aplicaciones, algunas de las cuales se describen a continuación.

[0161] Las composiciones nuevas y mejoradas de celulasas que comprenden cantidades variables de celulasas de tipo CBH, de tipo EG y de tipo BG son útiles en composiciones detergentes que presentan una capacidad de limpieza mejorada, funcionan como agente suavizante y/o mejoran el tacto de los tejidos de algodón (por ejemplo, “lavado a la piedra” o “biopulido”), en las composiciones para degradar la pulpa de madera en azúcares (por ejemplo, para la producción de bioetanol) y/o en composiciones de piensos. El aislamiento y caracterización de las celulasas de cada tipo proporcionan la capacidad de controlar los aspectos de dichas composiciones.

[0162] En un enfoque preferido, la celulasa es útil para las composiciones detergentes o para tratamientos de tejidos para mejorar el tacto y el aspecto.

[0163] Las β -glucosidasas pueden utilizarse en diferentes aplicaciones. Por ejemplo, puede añadirse β -glucosidasa a las uvas durante la elaboración del vino para aumentar el aroma potencial del producto vinícola acabado. Otra aplicación más puede consistir en utilizar β -glucosidasa en la fruta para potenciar el aroma de la misma. Como alternativa, el producto de la fermentación recombinante aislado que contiene la β -glucosidasa mejorada, se puede utilizar directamente en aditivos alimentarios o en la elaboración de vino para mejorar el sabor o el aroma.

[0164] Puesto que la velocidad de hidrólisis de los productos celulósicos puede incrementarse mediante el uso de un transformante que tenga al menos una copia adicional del gen *bgl7* insertado en el genoma, los productos que contienen celulosa o heteroglucanos pueden degradarse a un ritmo más rápido y en mayor medida. Los productos producidos a base de celulosa como el papel, algodón, pañales de celulosa y similares se pueden degradar de manera más eficaz en un vertedero. De este modo, el producto de fermentación que se puede obtener de los transformantes o los transformantes solos se pueden utilizar en composiciones para ayudar a degradar mediante licuefacción de una gran variedad de productos de celulosa que se añaden a los vertederos abarrotados.

[0165] La sacarificación y la fermentación separadas es un proceso por el que la celulosa presente en la biomasa, por ejemplo, forraje de maíz, se convierte en glucosa y, posteriormente, las cepas de levadura convierten la glucosa en etanol. La sacarificación y fermentación simultáneas es un proceso por el que la celulosa presente en la biomasa, por ejemplo, forraje de maíz, se convierte en glucosa y, al mismo tiempo y en el mismo reactor, las cepas de levadura convierten la glucosa en etanol. Por lo tanto, en otra estrategia preferida, la celulasa de tipo glucosidasa es útil en la degradación de la biomasa en etanol. La producción de etanol a partir de fuentes fácilmente disponibles de celulosa proporciona una fuente de combustible renovable estable.

[0166] Las materias primas a base de celulosa están compuestas de residuos agrícolas, pastos y bosques y otra biomasa de bajo valor, como los residuos municipales (p.ej., papel reciclado, hierba

cortada de los jardines, etc.) El etanol puede producirse a partir de la fermentación de cualquiera de estas materia primas de celulosa. No obstante, primero debe convertirse la celulosa en azúcares antes de que pueda producirse la conversión en etanol.

[0167] Se puede utilizar una gran variedad de materias primas con la β -glucosidasa y la seleccionada para su uso puede depender de la región en la que se realiza la conversión. Por ejemplo, en el Medio Oeste de Estados Unidos los residuos agrícolas como paja del trigo, forraje de maíz y bagazo pueden predominar mientras en California puede predominar la paja de arroz. No obstante, debe entenderse que puede utilizarse cualquier biomasa disponible de celulosa en cualquier región.

[0168] Una composición de celulosa que contiene una cantidad mayor de β -glucosidasa es útil en la producción de etanol. El etanol obtenido en este proceso también se puede utilizar como un potenciador de octanaje o directamente como combustible en lugar de gasolina, lo que es una ventaja porque el etanol como fuente de combustible es más ecológico que los productos derivados del petróleo. Se sabe que el uso de etanol mejorará la calidad del aire y posiblemente reducirá los niveles locales de ozono y la contaminación. Además, la utilización de etanol en lugar de gasolina puede ser una estrategia importante en la amortiguación del impacto de los cambios repentinos en los suministros de energía no renovable y petroquímica.

[0169] El etanol se puede producir a través de procesos de sacarificación y fermentación a partir de biomasa de celulosa, como árboles, plantas herbáceas, residuos sólidos municipales y residuos agrícolas y forestales. No obstante, un problema importante en este proceso es la falta de β -glucosidasa en el sistema para convertir la celobiosa en glucosa. Se sabe que la celobiosa actúa como inhibidor de las celobiohidrolasas y endoglucanasas y, por lo tanto, reduce la tasa de hidrólisis del sistema de celulosa completo. Por eso, la utilización de una actividad mayor de β -glucosidasa para convertir la celobiosa en glucosa incrementaría considerablemente la producción de etanol.

[0170] De este modo, la β -glucosidasa es útil en la hidrólisis de celulosa en sus componentes de azúcar. Puede añadirse β -glucosidasa a la biomasa antes de la adición de un organismo de fermentación. La β -glucosidasa puede añadirse a la biomasa al mismo tiempo que un organismo de fermentación. De modo opcional, puede haber otros componentes de celulosa presentes.

[0171] En otro ejemplo, la materia prima de celulosa se puede pretratar. El tratamiento previo puede ser mediante temperatura elevada y la adición de una solución de ácido diluido, ácido concentrado o álcali diluida. La solución de tratamiento previo se añade durante un tiempo suficiente para, por lo menos parcialmente, hidrolizar los componentes de hemicelulosa y después neutralizarlos.

[0172] En un enfoque alternativo, se prefiere una composición de celulasas que es deficiente en β -glucosidasa o que carece de ésta. La eliminación del gen de la β -glucosidasa descrito en el presente documento sería. Además, dichas composiciones son útiles para la producción de celobiosa y otros celooligosacáridos. La eliminación del gen *bgl7* de las cepas de *T.reesei* sería particularmente útil para preparar composiciones de celulosa para usar en detergentes y para aislar celobiosa. Las enzimas celulasas se han utilizado en una variedad de composiciones detergentes para limpiar la ropa de modo enzimático. No obstante, se sabe en la técnica que el uso de enzimas celulosa puede producir degradación de las fibras de celulosa en la ropa. Una posibilidad para reducir el efecto de degradación es producir un detergente que no contenga β -glucosidasa. Por consiguiente, la eliminación de esta proteína haría que el sistema de celulosa inhibiera los otros componentes por

acumulación de celobiosa. Los microorganismos modificados descritos en el presente documento son particularmente adecuados para la preparación de estas composiciones porque el gen *bgl7* se puede eliminar dejando los componentes CBH y EG restantes, lo que conllevaría la mejora de los beneficios de limpieza y de suavizado en la composición, sin los efectos de la degradación.

5 **[0173]** Las composiciones detergentes de la invención pueden emplear, además de la composición de celulasa (con independencia del contenido de β -glucosidasa, es decir, β -glucosidasa libre, β -glucosidasa sustancialmente libre o β -glucosidasa mejorada), un surfactante, incluyendo surfactantes aniónicos, no iónicos y anfólicos, una hidrolasa, agentes de relleno, agentes blanqueantes, agentes azulantes y tintes fluorescentes, inhibidores de la aglomeración, solubilizantes, surfactantes catiónicos
10 y similares. Todos estos componentes son conocidos en el sector de los detergentes. La composición de celulasa descrita anteriormente puede añadirse a la composición detergente, ya sea en un diluyente líquido, en gránulos, en emulsiones, en geles, en pastas y similares. Estas formas son bien conocidas por parte de los expertos en la materia. Cuando se utiliza una composición detergente sólida, la composición de celulasa se formula preferentemente en gránulos. Preferentemente, los
15 gránulos pueden formularse de manera que contengan un agente protector de celulasa. Para una discusión más completa, véase la Patente de EE. UU. N^o: 6.162.782, titulada "Detergent compositions containing cellulase compositions deficient in CBH I type components".

[0174] Las composiciones detergentes también pueden contener niveles mejorados de β -glucosidasa o β -glucosidasa alterada. En relación con este aspecto, depende realmente del tipo de producto que
20 se desea utilizar en las composiciones detergentes para ofrecer los efectos apropiados.

[0175] Preferentemente, se emplean composiciones de celulasa a partir de aproximadamente un 0,00005 por ciento en peso a aproximadamente un 5 por ciento en relación con la composición detergente total. Más preferiblemente, se utilizan composiciones de celulasa de aproximadamente un
25 0,0002 por ciento en peso a aproximadamente un 2 por ciento en peso en relación con la composición detergente total.

[0176] La eliminación del gen *bgl7* también proporcionaría la acumulación de celobiosa en el sistema de celulasa, que puede purificarse a partir de la misma. En este sentido, en la presente invención se presenta la posibilidad de aislar celobiosa de microorganismos de un modo fácil y eficaz.

[0177] Las partes de la secuencia de ácido nucleico de *bgl7* que son capaces de unirse a la celulosa
30 se pueden usar para generar proteínas de superficie quiméricas bacterianas, lo que permite la inmovilización de la célula completa en filtros de celulosa u otros soportes sólidos fibrosos como se describe en Lehtio *et al.*, 2001.

[0178] Además, la secuencia de ácidos nucleicos de *bgl7* es útil en la identificación y caracterización de secuencias de ácido nucleico relacionadas. Una serie de técnicas útiles para determinar (predecir
35 o confirmar) la función de genes o productos génicos relacionados incluyen sin carácter limitativo: (a) análisis de ADN/ARN, como (1) la sobreexpresión, expresión ectópica y expresión en otras especies; (2) *knock out* génico (genética inversa, *knock-out* dirigido, silenciamiento génico inducido viral (VIGS, véase Baulcombe, 1999); (3) análisis del estado de metilación del gen, especialmente en regiones reguladoras flanqueadoras e (4) hibridación *in situ*; (B) análisis de productos génicos como (1)
40 expresión de proteína recombinante; (2) producción de antisueros; (3) inmunolocalización; (4) ensayos bioquímicos de la actividad catalítica u otra actividad; (5) estado de fosforilación; e (6)

interacción con otras proteínas a través de análisis de dos híbridos de levaduras; (C) análisis de la vía, como la colocación de un gen o producto génico en una vía bioquímica particular o de señalización, basado en su fenotipo de sobreexpresión o mediante homología de la secuencia con genes relacionados; y (D) otros análisis que también se pueden llevar a cabo para determinar o
5 confirmar la participación del gen aislado y su producto en una vía metabólica o de señalización concreta, y ayudar a determinar la función del gen.

[0179] Las endoglucanasas y las β -glucosidasas pueden ser responsables de la producción de disacáridos, como soforosa, a partir de celooligosacáridos y glucosa por reacciones de transglucosilación. Se sabe que la soforosa es un inductor muy potente de la expresión del gen de
10 celulasa (limen, M. *et al.*, *Appl. Environ. Microbiol.* 63:1298-1306 y referencias en la misma). De este modo, las EG y BGL desempeñan una función importante en el proceso de inducción de la expresión del gen de celulasa. La sobreexpresión de determinadas EG y BGL en una cepa fúngica puede conducir a una mayor productividad general de celulasa mediante esa cepa.

A. Homología con secuencias conocidas

[0180] La función de una secuencia relacionada de ácido nucleico que codifica BGL7 puede determinarse por homología con genes conocidos que tienen una función concreta. Por ejemplo, se utiliza una comparación de la secuencia de codificación de una molécula de ácido nucleico identificada con bases de datos públicas de secuencias de ácidos nucleicos, con el fin de confirmar la
15 función por homología con genes conocidos o por extensión de la secuencia de ácido nucleico identificada.

[0181] La expresión “% de homología” se utiliza en el presente documento de forma intercambiable con la expresión “% de identidad” en este escrito y hace referencia al nivel de identidad de secuencia de los aminoácidos y del ácido nucleico entre la secuencia de ácido nucleico que codifica BGL7 o la
20 secuencia de aminoácidos de BGL7, cuando se alinean usando un programa de alineamiento de secuencias.

[0182] Por ejemplo, como se utiliza en este documento, 80% de homología es sinónimo de una identidad de secuencia de 80% determinada por un algoritmo definido y, en consecuencia, un homólogo de una secuencia determinada presenta más de un 80% de identidad de secuencia a lo largo de una secuencia dada. Los niveles de ejemplo de identidad de secuencia incluyen, sin carácter
25 limitativo, el 80, 85, 90, 95, 98% o más de identidad de secuencia con una secuencia determinada, p. ej., la secuencia de codificación para *bgl7*, como se describe en este documento.

[0183] Entre los programas de ordenador que pueden utilizarse para determinar la identidad entre dos secuencias se incluyen, por ejemplo, sin carácter limitativo, el paquete de programas BLAST, p.ej., BLASTN, BLASTX y TBLASTX, BLASTP y TBLASTN, disponibles al público en Internet en
30 <http://www.ncbi.nlm.nih.gov/BLAST/>. Véanse también, Altschul, *et al.*, 1990 y Altschul, *et al.*, 1997.

[0184] Las búsquedas de secuencias se llevan a cabo normalmente empleando el programa BLASTN cuando se evalúa una secuencia de ácido nucleico en relación con secuencias de ácidos nucleico en las secuencias de ADN del GenBank y otras bases de datos públicas. Se prefiere el programa BLASTX para la búsqueda de secuencias de ácidos nucleicos que se han transducido en todos los

marcos de lectura frente a secuencias de aminoácidos en las secuencias de proteínas del GenBank y otras bases de datos públicas. Tanto BLASTN como BLASTX se ejecutan utilizando parámetros por defecto de una penalización por hueco (*gap*) abierto de 11,0 y una penalización por hueco extendido de 1,0 y utilizan la matriz BLOSUM-62.

5 (Véase, p.ej., Altschul, *et al.*, 1997.)

[0185] Un alineamiento de secuencias seleccionadas con el fin de determinar el “% de identidad” entre dos o más secuencias se realiza empleando, por ejemplo, el programa CLUSTAL-W en MacVector versión 6.5, ejecutado con los parámetros por defecto, incluyendo la penalización por hueco abierto de 10,0, una penalización por hueco extendido de 0,1 y una matriz de similitud BLOSUM 30.

10 **[0186]** En un método de ejemplo, la extensión de secuencia de un ácido nucleico que codifica *bgl7* puede realizarse utilizando procedimientos convencionales de extensión de cebador como se describe en Sambrook *et al.*, *supra*, para detectar precursores de *bgl7* y los procesos intermedios de ARNm, que pueden no haber experimentado transcripción inversa en ADNc y/o para identificar ORF que codifican una proteína de longitud completa.

15 **[0187]** En el presente documento se describe la secuencia de nucleótidos completa o parcial de la secuencia del ácido nucleico de *bgl7* para utilizarla como sonda. Esta sonda se utilizaría para identificar y clonar secuencias de ácidos nucleicos homólogos de organismos relacionados.

[0188] El cribado de una biblioteca genómica o de ADNc con la sonda seleccionada se llevaría a cabo empleando procedimientos estándares, como los descritos en Sambrook *et al.*, (1989). Las condiciones de hibridación, incluidas la astringencia moderada y la astringencia alta, se proporcionan en Sambrook *et al.*, *supra*.

20 **[0189]** Las sondas o partes de las mismas también pueden emplearse en técnicas de PCR para generar un grupo de secuencias para la identificación de secuencias estrechamente relacionadas con *bgl7*. Cuando las secuencias de *bgl7* son para utilizar como sondas, se puede utilizar una parte concreta de una secuencia que codifica BGL7, por ejemplo, una parte muy conservada de la secuencia de codificación.

25 **[0190]** Por ejemplo, puede utilizarse una secuencia de nucleótidos de *bgl7* como una sonda de hibridación para una biblioteca de ADNc para aislar genes, por ejemplo, los que codifican variantes de origen natural de BGL7 de otras especies fúngicas, bacterianas o vegetales, que tienen el nivel deseado de identidad de secuencia con la secuencia de nucleótidos de *bgl7* descritas en la Figura 1 (SEQ ID N°:1). Las sondas de ejemplo tienen una longitud de aproximadamente 20 a aproximadamente 50 bases.

B: Análisis de dos híbridos

35 **[0191]** Las proteínas identificadas como se describe en este documento pueden utilizarse en el sistema de dos híbridos para “capturar” las proteínas de unión a la proteína que son proteínas de vía de señal aparentes. El sistema de dos híbridos de levadura se describe en Fields y Song, *Nature* 340:245-246 (1989). En resumen, en un sistema de dos híbridos, se construye una fusión de un dominio-*bgl7* de unión al ADN (p.ej., fusión GAL4-*bgl7*) y se transfecta en células de levadura. Puede

utilizarse el gen *bgl7* completo, o subregiones del gen *bgl7*. Una segunda construcción que contiene la librería de ligandos potenciales fusionados al dominio de activación de ADN se cotransfecta. Los cotransformantes de levadura que albergan proteínas que se unen a la proteína BGL7 se identifican mediante, por ejemplo, la producción (un cribado) de β -galactosidasa o luciferasa, o la supervivencia en placas que carecen de un nutriente esencial (una selección), adecuado para los vectores utilizados.

C. Análisis de micromatriz

[0192] Además, se puede utilizar el análisis de micromatrices, también conocido como descripción del perfil de expresión o perfiles de transcripción, para evaluar simultáneamente la presencia o expresión de secuencias dadas de ADN, o cambios de la expresión de muchos genes diferentes. En un procedimiento, un gran conjunto de secuencias de ADN (sondas), normalmente una amplia gama de marcas de secuencias expresadas, ADNc, fragmentos de ADNc u oligonucleótidos específicos de secuencia, se coloca sobre un soporte sólido como una placa de vidrio o una membrana de nailon. La diana marcada para la hibridación con las sondas se genera aislando ARNm del tejido de control e inducido, después marcando cada grupo de ARNm ya sea directamente o a través de ADNc o ARNc, con un marcador distinto, normalmente un tinte fluorescente. La micromatriz se hibrida con las sondas complejas y se puede cuantificar la intensidad relativa de la señal de hibridación asociada con cada localización de la matriz para cada tinte marcador. Las diferencias en la expresión entre los estados de control e inducidos se pueden medir como una relación de las señales de los dos tintes marcadores.

(Véase Baldwin, D *et al.*, 1999)

[0193] Puede llevarse a cabo el análisis de la micromatriz del organismo fuente del que se ha obtenido *bgl7*, con el fin de facilitar la comprensión de la función génica mediante la identificación de otros genes que son regulados de forma coordinada como consecuencia de la sobreexpresión de *bgl7*. Para identificar los genes regulados de forma coordinada puede ayudar el colocar el gen *bgl7* en una vía concreta. Como alternativa, puede utilizarse dicho análisis para identificar otros genes involucrados en la misma vía empleando un análisis de micromatrices.

[0194] Aunque en la invención se han indicado referencias a métodos y modos de realización específicos, se apreciará que se puedan realizar diversas modificaciones y cambios del modo en el que se establece en las reivindicaciones.

EJEMPLO 1

[0195] En un procedimiento de ejemplo, se aísla un fragmento de ADNc para usar como sonda mediante la extracción total de ARN de micelios de una cepa de *T. reesei* cultivada en condiciones que se sabe que inducen la producción de celulasa y la obtención de la fracción poliadenilada (poliA) de la misma. El poliA ARN se utiliza para producir una mezcla de ADNc que a continuación se amplifica utilizando cebadores específicos basados en la secuencia de ácido nucleico de *bgl7* proporcionada aquí.

[0196] El ARN total se aísla de micelios utilizando métodos conocidos en la técnica, por ejemplo como describen Timberlake *et al.*, 1981; Maniatis, *et al.*, 1989; Ausubel, *et al.*, 1993 y Sambrook *et al.*, 1989, cada uno de los cuales se incluye expresamente como referencia en este documento. Una vez aislado, se realizans transferencias Northern blot para confirmar la expresión de la celulasa y seleccionar un tiempo de inducción óptimo para la expresión de la celulasa y el aislamiento de ARN correspondiente.

[0197] El ARN mensajero (ARNm), que tiene una cola de poli(A) en el extremo 3', se puede purificar del ARN total utilizando métodos conocidos en la técnica.

[0198] El ARN de *T. reesei* se utiliza como modelo para la RT-PCR utilizando métodos conocidos en la técnica (Loftus, J. *et al.*, *Science*, 249:915-918, 1990). Durante este proceso, se lleva a cabo la transcripción inversa del ARNm para producir la primera cepa de ADNc. Posteriormente, el ADNc se utiliza como modelo para la amplificación por PCR de las secuencias de ADNc de *bgI7* empleando cebadores de oligonucleótidos específicos diseñados según la SEQ ID N°1 o SEQ ID N°4.

Tabla 1. Secuencias proporcionadas para apoyar la invención

Descripción	SEQ. ID N°.
Secuencia de ácido nucleico de ADNc de <i>bgI7</i> de longitud completa	1

<p><u>GTAGTAGCAAACACGTCGCCCATCAACATGCGGCTGTGTGACTTATCCAG CCTTGCGAGCTGGGTCTCGTGACAGTAGCTCTGCCCTCGAGCGGCGCT GCTGCCAAAGGCGTCTCGCAAATACCTTCAACACATTC AAGTCAAAGCAA AGGAAATGGACCGTGGGCTCACGCGTATCGTCGCGCCGAGAAGTTAGTG CGACAAATGACACTCGAAGAAAAGGCCAACATCACGCGCGGATTCACCG GCGACAATGTCTGTGCCGGCAACACTGGCTCTGTTCCCTCGCCTGGGATG GCCCGGCATGTGTGTCCACGATGCCGGCAACGGAGTTCGCGCAACCGA CTTGGTCAATTCCTATCCCTCTGGCATCCACGTCGGGGCGAGCTGGGAT CGAAACCTGACGTACGAGAGGGGGCTTCATATGGGCGGGGAGTTCAAAG CAAAAGGAGTCAACGTCCCCTCGGTCCCAATGCTGGCCCGCTAGGGCG AACACCTCTGGGTGGTGC AAACTGGGAGGGTTTCTCCATCGATCCGTATC TCTCTGGCCAATTGAACGCAGAGACAATCACTGGAATGCAAGATGCCGG AGTGATTGCGAACATCAAGCATTTCATCGCCAACGAACAAGAGACGCTTC GGCGTCCCTACTTTGGTGTGGAAGCTGTTTCTGCAAATATCGATGACAGA ACCTACACGAATACTATCTCTGGCCCTTTATGGATAGTGTGCATGCTGG CGTGGGATCCGT CATGTGCTCCTATAACAGGATCAACAACACGTACGGAT GCATGAACGACAAGCTTATGAACGGAATTCTCAAGGCTGAATTGGGCTTT CAAGGTTTCGT CATGCTTGACTGGAATGCTCAGCACGATCTGCAAAGCGC CAATGCCGGACTCGACATGGTGATGCCCTCGGTGGTCTTTGGGGCAAG AATCTGACAGATGCTGTTGCAAACGGGACGGTCAGCGAGTCTCGGATTA CGGACATGGCCACGAGGATTGCTGCATGGTACTTAGTCGGTCAAGA TGCCAACAACCTTCCAGTACCGGGCATCGGCTTCAAACAGCTCACGAAAC CGCACGAGCAAGTCGACGCACGCGATCCCGCATCGAAGCCCGTGCTTCT GGAGGGCGCCATTGCAGGACACGTTCTAGTCAAGAACGAAAACAATGCG CTACCGTTCAACAAGAAGCTAACCATGATCTCCGTCTTTGGCTACGATGC TACGATCCCACGCACAAAGAATACCGACATTCTTTCCAGCTCGGATATA CCTCTTCGCCGGAGATGGCTCAGGCCGTA CTGGCAATGAGGCGCATTTC CGACCATGGCCAGCAAAGGGAGGGACAATTATGACTGGCGGGCGAGCTGG CGCAAACGCTCCATCATAATCGACGATCCGCTTGCTGCTATCCAACGTC GAGCCCGCAAAGATGATACTTGGGTA AATTGGGACCTGGACTCCTTCAAT CCGGAAGTCAATGCTGCTTCAGATGCTTGCTTGGTCTTCATCAATGCCAT CGCAACAGAGGGCTGGGACCGTGACGGCCTCCATGACGATTTTAGTGAC GGCCTTGCTTGAATGTAGCCGCCAACTGCTCCAACACGATTGTCGTGCT TCAGGCCGCGGGCACTCGCCTGTTGACCAATGGATTGAGCATCCCAAT TTGGCAAGCCGCTGCACAAGTCACCGCAAGCATACCAACAGTGGCGGC ATGTCTGGAAGTGAGGTTGCGCAGCTGTACTTGGCCATTCAAATAGCCC GCCAAAGCAATTGCGCGGATTCAACA AACTGTTGCTGCGTCCACATGAGT CTGGAAGTGTCACTTTGGACTCACGAAGCGAGACTTAAGTGTGTTGGGAT GTTGTTTCTCAGTCGTGGGTTATT CAGGAGGGTGAGTACAAGGTATTTGT TGGGGCGAGCAGCCGCGATATTGACTCAGTGGAAAACGATATTTAG GGAGCATAGCTTATTGAGCGAATTCTTTCTCTTTCAAAAAAAAAAAAAAA A</u></p>	
<p>secuencia de aminoácidos predicha de <i>bgl7</i></p>	<p>2</p>

<p>AKGVSQIPSTHSSQSKGNGPWAHAYRRAEKLVRQMTLEEKANITRGFTGDNV CAGNTGSVPRLGWPGMVCVHDAGNGVRATDLVNSYPSGIHVGASWDRNLTYE RGLHMGGEFKAKGVNVPLGPNAGPLGRTPLGGRNWEFGSIDPYLSGQLNAET ITGMQDAGVIANIKHFIANEQETLRRPYFGVEAVSANIDDRTLHEYYLWPFMD SVHAGVGSVMCSYNRINNTYGCMNDKLMNGILKAELGFQGFVMLDWNAQH DLQSANAGLDMVMPLGGSWGNLTDVANGTVSESRTDMATRIIAAWYLV GQDGNNFPVPGIGLKQLTKPHEQVDARDPASKPVLLEGAIAGHVLVKNENNA LPFNKCLTMISVFGYDATIPRTKNTDILFQLGYTSSPEMAQAVLGNEAHFDQA AKGGTIMTGGRAGANAPSYIDDPLAAIQRRARKDDTWVNWDLDSFNPEVNA ASDACLVFINAJATEGWDRDGLHDDFSDGLVLNVAANCSNTIVVVHAAGTRL VDQWIEHPNVTAAVIAHLPQDSGRALVKLLYGEANFSGKLPYTIKKNESDYS VYTPCQRRSPEDTDPQCDFTEGVYLDYRAFDANNMTPRFEFGYGLSYTSFNYS ALSIKKAKGLRQSRCTDDLWQAAAQVTASITNSGGMSGSEVAQLYLAIPNSPP KQLRGFNKLLLRPHESGTVHFGLTKRDLVWDVVSQSWVIQEGEYKVFVGAS SRDIRLSGKLI</p>	
<p>Secuencia señal predicha de la proteína BGL7:</p>	<p>3</p>
<p>MRLCDLSSLASWVLVTVALPSSGAA</p>	
<p>Secuencia que codifica el ácido nucleico de <i>bgl7</i></p>	<p>4</p>

ATGCGGCTGTGTGACTTATCCAGCCTTGCGAGCTGGGTCTCGTGACAG
 TAGCTCTGCCTTCGAGCGGCGCTGCTGCCAAAGGCGTCTCGAAATACC
 TTCAACACATTCAAGTCAAAGCAAAGGAAATGGACCGTGGGCTCACGCGT
 ATCGTCGCGCCGAGAAGTTAGTGCGACAAATGACACTCGAAGAAAAGGC
 CAACATCACGCGCGGATTCACCGGCGACAATGTCTGTGCCGGCAACT
 GGCTCTGTCTCGCCTGGGATGGCCCGGCATGTGTGCCACGATGCCG
 GCAACGGAGTTCGCGCAACCGACTTGGTCAATTCTTATCCCTCTGGCATC
 CACGTCGGGGCGAGCTGGGATCGAAACCTGACGTACGAGAGGGGGCTT
 CATATGGGCGGGGAGTTCAAAGCAAAGGAGTCAACGTCCCCTCGGTC
 CCAATGCTGGCCCGCTAGGGCGAACACCTCTGGTGGTTCGAAACTGGGA
 GGGTTTCTCCATCGATCCGTATCTCTCTGGCCAATTGAACGCAGAGACAA
 TCACTGGAATGCAAGATGCCGGAGTGATTGCGAACATCAAGCATTTTCATC
 GCCAACGAACAAGAGACGCTTCGGCGTCCCTACTTTGGTGTGGAAGCTG
 TTTCTGCAAATATCGATGACAGAACCCTACACGAATACTATCTCTGGCCCT
 TTATGGATAGTGTGCATGCTGGCGTGGGATCCGTGATGTGCTCCTATAAC
 AGGATCAACAACACGTACGGATGCATGAACGACAAGCTTATGAACGGAAT
 TCTCAAGGCTGAATTGGGCTTTCAAGGTTTCGTGATGCTTGACTGGAATG
 CTCAGCAGATCTGCAAAGCGCCATGCCGGACTCGACATGGTGTGATGCC
 CCTCGTGGTCTTGGGGCAAGAATCTGACAGATGCTGTTGCAAACGGG
 ACGGTCAGCGAGTCTCGGATTACGGACATGGCCACGAGGATCATTGCTG
 CATGGTACTTAGTCCGGTCAAGATGGCAACAACCTTCCAGTACCGGGCATC
 GGCTTGAAACAGCTCACGAAACCGCACGAGCAAGTCGACGCACGCGATC
 CCGCATCGAAGCCCGTGTCTGGAGGGCGCCATTGCAGGACACGTTCT
 AGTCAAGAACGAAAACAATGCGCTACCGTTCAACAAGAAGCTAACCATGA
 TCTCCGTCTTGGCTACGATGCTACGATCCCACGCACAAAGAATACCGAC
 ATTCTTTTCCAGCTCGGATATACCTTTCGCCGGAGATGGCTCAGGCCGT
 ACTTGGCAATGAGGCGCATTTGACCAGGCAGCAAAGGGAGGGACAATT
 ATGACTGGCGGGCGAGCTGGCGCAAACGCTCCATCATACATCGACGATC
 CGCTTGCTGCTATCCAACGTGAGCCCGCAAAGATGATACTTGGGTAAT
 TGGGACCTGGACTCCTTCAATCCGGAAGTCAATGCTGCTTCAGATGCTTG
 CTTGGTCTTCATCAATGCCATCGCAACAGAGGGCTGGGACCGTGACGGC
 CTCCATGACGATTTTAGTGACGGCCTTGTCTTGAATGTAGCCGCCAACTG
 CTCCAACACGATTTGTCGTGTTACGCCCGGGGCACTCGCCTGGTTGAC
 CAATGGATTGAGCATCCCAATGTTACTGCCGCCGTATCGCGCATCTTCC
 AGGCCAGGACAGCGGTAGAGCCCTCGTGAAGCTTCTTATGGCGAAGCC
 AACTTCTCTGGCAAACCTCCCTATACAATTGCCAAGAACGAGAGCGATTA

 CTCAGTTTACACCCCATGCCAGCGACGCTCTCCCGAAGACACCGATCCC
 CAGTGGGATTTACCGAAGGCGTCTATCTCGATTATCGCGCTTTTGATGC
 GAACAACATGACTCCCCGCTTCGAGTTCGGATACGGGCTCAGCTACACG
 TCGTTCAATTAATCAGCTCTCTCCATCAAAGGCAAAGGGCCTTCGGCA
 GTCAAGGTGTACCGACGATCTTTGGCAAGCCGCTGCACAAGTCAACGCA
 AGCATCACCACAGTGGCGGCATGTCTGGAAGTGAGGTTGCGCAGCTGT
 ACTTGGCCATTCAAATAGCCCGCAAAGCAATTGCGCGGATTCAACAAA
 CTGTTGCTGCGTCCACATGAGTCTGGAAGTGTCACTTTGGACTCACGAA
 GCGAGACTTAAGTGTGGGATGTTGTTTCTCAGTCGTGGGTTATTCAGG
 AGGTTGAGTACAAGGTATTTGTTGGGGCGAGCAGCCGCGATATTCGACT
 CAGTGAAAACTGCATATTTAG

Reivindicaciones

1. Un polinucleótido aislado de origen fúngico, donde el polinucleótido consta de una secuencia de nucleótidos que codifica una enzima que tiene actividad de β -glucosidasa, donde se selecciona dicho polinucleótido aislado del grupo compuesto de:
 - 5 **(a)** una secuencia de ácido nucleico que codifica o que es complementaria de una secuencia que codifica un polipéptido BGL7 que tiene una identidad de secuencia de al menos el 85% con la secuencia de aminoácidos presentada como la SEQ ID N°:2;
 - (b)** una secuencia de ácido nucleico que codifica o que es complementaria de una secuencia que codifica un polipéptido BGL7 que tiene una identidad de secuencia de
10 al menos el 90% con la secuencia de aminoácidos presentada como la SEQ ID N°:2;
 - (c)** una secuencia de ácido nucleico que codifica o que es complementaria de una secuencia que codifica un polipéptido BGL7 que tiene una identidad de secuencia de al menos el 95% con la secuencia de aminoácidos presentada en la Figura 2;
 - (d)** una secuencia de ácido nucleico que codifica o que es complementaria de una
15 secuencia que codifica un polipéptido BGL7 que tiene la secuencia de aminoácidos presentada en la Figura 2;
 - (e)** una secuencia de ácido nucleico que codifica o que es complementaria de una secuencia que codifica un polipéptido BGL7 que tiene una identidad de secuencia de al menos el 95% con la secuencia de aminoácidos presentada en la SEQ ID N°:2;
 - 20 **(f)** una secuencia de ácido nucleico que codifica o que es complementaria de una secuencia que codifica un polipéptido BGL7 que tiene una secuencia de aminoácidos presentada como la SEQ ID N°:2;
 - (g)** una secuencia de ácido nucleico presentada como la SEQ ID N°:4, o el complemento de la misma; y
 - 25 **(h)** una secuencia de ácido nucleico que se hibrida, en condiciones de alta astringencia, a la secuencia presentada como la SEQ ID N°:4; el complemento de la misma.
2. El polinucleótido aislado de la reivindicación 1, en el que se calcula el % de identidad utilizando el programa CLUSTAL-W en MacVector versión 6.5, ejecutado con parámetros por defecto, que incluye una penalización por huecos abiertos de 10,0, una penalización de hueco extendido de
30 0,1 y una matriz de similitud BLOSUM 30.
3. El polinucleótido aislado de la reivindicación 1, en el que la hibridación se lleva a cabo a 42°C en 50% de formamida, 6X de SSC, 5X de solución de Denhardt, 0,5% de SDS y 100 μ g/ml de ADN portador desnaturalizado seguido de lavado dos veces en 2X SSPE y 0,5% de SDS a temperatura ambiente y dos veces más adicionales en 0,1 de SSPE y 0,5% de SDS a 42°C.
- 35 4. El polinucleótido aislado de la reivindicación 1, donde dicho polinucleótido es una molécula de ARN.
5. El polinucleótido aislado de la reivindicación 2, que codifica una enzima que tiene una actividad de β -glucosidasa, en el que la enzima deriva de una fuente de *Trichoderma*.
6. Un polinucleótido aislado de la reivindicación 4, en el que la enzima deriva de *Trichoderma*

reesei.

7. Una construcción de expresión que incluye una secuencia de polinucleótido que (i) codifica un polipéptido que presenta una identidad de secuencia de al menos el 85% con la secuencia de aminoácidos presentada como la SEQ ID N°:2, o (ii) es capaz de hibridarse a una secuencia de nucleótidos que codifica el polipéptido presentado como la SEQ ID N°:2 en condiciones de intermedia a alta astringencia, o (iii) es complementaria de una secuencia de nucleótidos que codifica un polipéptido que tiene una identidad de secuencia de al menos el 85% con una secuencia de aminoácidos presentada como la SEQ ID N°: 2, en la que la secuencia de polinucleótido o su complemento codifica un polipéptido que tiene actividad de β -glucosidasa.
8. Un vector que incluye la construcción de expresión de la reivindicación 7.
9. Un vector compuesto de un polinucleótido aislado de la reivindicación 1, unido de modo funcional a secuencias de control reconocidas por una célula huésped transformada con el vector.
10. Una célula huésped que comprende el vector de la reivindicación 8.
11. Una célula huésped que comprende el vector de la reivindicación 9.
12. La célula huésped de la reivindicación 11, que es una célula procariota.
13. La célula huésped de la reivindicación 11, que es una célula eucariota.
14. Una célula huésped recombinante que comprende un polinucleótido de la reivindicación 1.
15. La célula huésped recombinante de la reivindicación 14, que es una célula procariota.
16. La célula huésped recombinante de la reivindicación 12, que es una célula eucariota.
17. Un polipéptido BGL7 purificado con la actividad biológica de una β -glucosidasa, que consta de una secuencia seleccionada del grupo que consiste en:
 - (a) una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 85% con la secuencia de aminoácidos presentada en la Figura 2;
 - (b) una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 90% con la secuencia de aminoácidos presentada en la Figura 2;
 - (c) una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 95% con la secuencia de aminoácidos presentada en la Figura 2;
 - (d) una secuencia de aminoácidos presentada en la Figura 2;
 - (e) una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 95% con la secuencia de aminoácidos presentada como la SEQ ID N°:2;
 - (f) una secuencia de aminoácidos presentada como la SEQ ID N°:2; y
 - (g) un fragmento biológicamente activo purificado de la secuencia de aminoácidos presentada como la SEQ ID N°:2 de al menos 20-100 aminoácidos de longitud, donde el fragmento tiene actividad de β -glucosidasa.
18. Un método para producir una enzima que tiene actividad de β -glucosidasa que comprende los pasos de:
 - (a) transformar de modo estable una célula huésped con un vector de expresión que

consta de un polinucleótido tal y como se define en la reivindicación 1;

(b) cultivar dicha célula huésped transformada en condiciones apropiadas para que dicha célula huésped produzca dicha β -glucosidasa; y

(c) recuperar la mencionada β -glucosidasa.

5 19. El método de la reivindicación 18, en el que la célula huésped es un hongo filamentoso o una célula de levadura.

20. Una enzima purificada que tiene actividad de β -glucosidasa preparada por el procedimiento de la reivindicación 18.

10 21. Una célula huésped recombinante que consta de una eliminación o inserción u otra modificación en el gen *BGL7* que inactiva el gen y previene la producción de polipéptido BGL7, en la que el gen *BGL7* presenta la secuencia mostrada en la SEQ ID N°:1.

15 22. Un oligonucleótido antisentido complementario de un ARN mensajero que codifica un polipéptido BGL7 que tiene la secuencia presentada como SEQ ID N°:2, donde tras la exposición a una célula huésped que produce β -glucosidasa, dicho oligonucleótido disminuye o inhibe la producción de β -glucosidasa por dicha célula huésped.

23. El oligonucleótido antisentido de la reivindicación 22, donde la célula huésped es un hongo filamentoso.

24. Una composición detergente, dicha composición consta de un polipéptido seleccionado del grupo formado por: los polipéptidos según la reivindicación 17 del apartado (a) al (f).

20 25. Un método para expresar un polipéptido heterólogo que tiene actividad de β -glucosidasa en una especie *Aspergillus*, que comprende:

(a) proporcionar un *Aspergillus* huésped con un vector de expresión que comprende un polinucleótido que codifica una secuencia señal de β -glucosidasa de *Aspergillus* unido a un polinucleótido que codifica una β -glucosidasa heteróloga, que codifica de este modo un polipéptido quimérico, en la que la β -glucosidasa heteróloga presenta una identidad de secuencia de al menos el 85% con la SEQ ID N°:2;

25

(b) cultivar dicho *Aspergillus* huésped en condiciones adecuadas para que dicho *Aspergillus* produzca dicho polipéptido quimérico, donde se produce dicho polipéptido quimérico.

30 26. Un método para la producción de etanol, dicho método sigue las etapas de:

(a) poner en contacto una composición de biomasa con una composición enzimática que consta de β -glucosidasa 7 para lograr una solución de azúcar, en la que la β -glucosidasa 7 presenta una identidad de secuencia de al menos el 85% con la SEQ ID N°:2;

35

(b) añadir a la solución de azúcar un microorganismo de fermentación; y

(c) cultivar el microorganismo de fermentación en condiciones suficientes para producir etanol,

en el que la composición de biomasa puede ser pretratada opcionalmente.

27. El método de la reivindicación 26, en el que la etapa (a) también consta de la adición de al

menos una endoglucanasa.

28. El método de la reivindicación 26, en el que la etapa (a) también consta de la adición de al menos una celobiohidrolasa.

5 **29.** El método de la reivindicación 27, en el que la etapa (a) también consta de la adición de al menos una celobiohidrolasa.

30. El método de la reivindicación 26, en el que el tratamiento previo es con un ácido diluido.

31. Un método para producir etanol, dicho método comprende las etapas de:

10 **(a)** poner en contacto una composición de biomasa con una composición enzimática que comprende β -glucosidasa 7, en la que la β -glucosidasa 7 presenta una identidad de secuencia de al menos el 85% con la SEQ ID N°:2; y

(b) cultivar el microorganismo de fermentación en condiciones suficientes para producir etanol.

 en el que la composición de biomasa puede ser pretatada opcionalmente.

15 **32.** El método de la reivindicación 31, en el que la etapa (a) también consta de la adición de al menos una endoglucanasa.

33. El método de la reivindicación 31, en el que la etapa (a) también consta de la adición de al menos una celobiohidrolasa.

34. El método de la reivindicación 32, en el que la etapa (b) también consta de la adición de al menos una celobiohidrolasa.

20 **35.** El método de la reivindicación 31, en el que el tratamiento previo es con un ácido diluido.

3TAGTAGCAA	ACACGTCGCC	CATCAACATG	CGGCTGTGTG	ACTTATCCAG	50
CCTTGCGAGC	TGGGTCTCTG	TGACAGTAGC	TCTGCCTTCG	AGCGGCGCTG	100
CTGCCAAAGG	CGTCTCGCAA	ATACCTTCAA	CACATTCAAG	TCAAAGCAAA	150
GGAAATGGAC	CGTGGGCTCA	CGCGTATCGT	CGCGCCGAGA	AGTTAGTGCG	200
ACAAATGACA	CTCGAAGAAA	AGGCCAACAT	CACGCGCGGA	TTCACCGGCG	250
ACAATGTCTG	TGCCGGCAAC	ACTGGCTCTG	TTCCTCGCCT	GGGATGGCCC	300
GGCATGTGTG	TCCACGATGC	CGGCAACGGA	GTTTCGCGAA	CCGACTTGGT	350
CAATTCTTAT	CCCTCTGGCA	TCCACGTCGG	GGCGAGCTGG	GATCGAAACC	400
TGACGTACGA	GAGGGGGCTT	CATATGGGCG	GGGAGTTCAA	AGCAAAAGGA	450
GTCAACGTCC	CACTCGGTCC	CAATGCTGGC	CCGCTAGGGC	GAACACCTCT	500
GGGTGGTCTG	AACTGGGAGG	GTTTCTCCAT	CGATCCGTAT	CTCTCTGGCC	550
AATTGAACGC	AGAGACAATC	ACTGGAATGC	AAGATGCCGG	AGTGATTGCG	600
AACATCAAGC	ATTTTCATCGC	CAACGAACAA	GAGACGCTTC	GGCGTCCCTA	650
CTTTGGTGTG	GAAGCTGTTT	CTGCAAATAT	CGATGACAGA	ACCTTACACG	700
AATACTATCT	CTGGCCCTTT	ATGGATAGTG	TGCATGCTGG	CGTGGGATCC	750
GTCATGTGCT	CCTATAACAG	GATCAACAAC	ACGTACGGAT	GCATGAACGA	800
CAAGCTTATG	AACGGAATTC	TCAAGGCTGA	ATTGGGCTTT	CAAGGTTTCG	850
TCATGCTTGA	CTGGAATGCT	CAGCACGATC	TGCAAAGCGC	CAATGCCGGA	900
CTCGACATGG	TGATGCCCTT	CGGTGGTCTT	TGGGGCAAGA	ATCTGACAGA	950
TGCTGTTGCA	AACGGGACGG	TCAGCGAGTC	TCGGATTACG	GACATGGCCA	1000
CGAGGATCAT	TGCTGCATGG	TACTTAGTCG	GTCAAGATGG	CAACAACCTT	1050
CCAGTACCGG	GCATCGGCTT	GAAACAGCTC	ACGAAACCGC	ACGAGCAAGT	1100
CGACGCACGC	GATCCCAGAT	CGAAGCCCGT	GCTTCTGGAG	GGCGCCATTG	1150
CAGGACACGT	TCTAGTCAAG	AACGAAAACA	ATGCGCTACC	GTTCAACAAG	1200
AAGTAACCA	TGATCTCCGT	CTTTGGCTAC	GATGCTACGA	TCCCACGCAC	1250
AAAGAATACC	GACATTCTTT	TCCAGCTCGG	ATATACCTCT	TCGCCGGAGA	1300
TGGCTCAGGC	CGTACTTGGC	AATGAGGCGC	ATTTTCGACCA	GGCAGCAAAG	1350
GGAGGGACAA	TTATGACTGG	CGGGCGAGCT	GGCGCAAACG	CTCCATCATA	1400
CATCGACGAT	CCGCTTGCTG	CTATCCAACG	TCGAGCCCGC	AAAGATGATA	1450
CTTGGGTAAG	TTGGGACCTG	GACTCCTTCA	ATCCGGAAGT	CAATGCTGCT	1500
TCAGATGCTT	GCTTGGTCTT	CATCAATGCC	ATCGCAACAG	AGGGCTGGGA	1550
CCGTGACGGC	CTCCATGACG	ATTTTAGTGA	CGGCCTTGTC	TTGAATGTAG	1600
CCGCCAACCG	CTCCAACACG	ATTGTCTGTC	TTCACGCCGC	GGGCACTCGC	1650
CTGTTTGACC	AATGGATTGA	GCATCCCAAT	GTTACTGCCG	CCGTCATCGC	1700
GCATCTTCCA	GGCCAGGACA	GCGGTAGAGC	CCTCGTGAAG	CTTCTTTATG	1750
GCGAAGCCAA	CTTCTCTGGC	AACTTCCCTT	ATACAATTGC	CAAGAACGAG	1800
AGCGATTACT	CAGTTTACAC	CCCATGCCAG	CGACGCTCTC	CCGAAGACAC	1850
CGATCCCCAG	TGCGATTTCA	CCGAAGGCGT	CTATCTCGAT	TATCGCGCTT	1900
TTGATGCGAA	CAACATGACT	CCCCGCTTCG	AGTTCGGATA	CGGGCTCAGC	1950
TACACGTCGT	TCAATTACTC	AGCTCTCTCC	ATCAAAAAGG	CAAAGGCCTT	2000
TCGGCAGTCA	AGGTGTACCG	ACGATCTTTG	GCAAGCCGCT	CCACAAGTCA	2050
CCGCAAGCAT	CACCAACAGT	GGCGGCATGT	CTGGAAGTGA	GGTTGCGCAG	2100
CTGTACTTGG	CCATTCCAAA	TAGCCCGCCA	AAGCAATTGC	GCGGATTCAA	2150
CAAACGTTTG	CTGCGTCCAC	ATGAGTCTGG	AACTGTTTAC	TTTGACTTCA	2200
CGAAGCGAGA	CTTAAGTGTT	TGGGATGTTG	TTTCTCAGTC	GTGGGTTATT	2250
CAGGAGGGTG	AGTACAAGGT	ATTTGTTGGG	GCGAGCAGCC	GCGATATTTCG	2300
ACTCAGTGGG	AAACTGCATA	TTTAGGGAGC	ATAGCTTATT	GAGCGAATTC	2350
GTTTCTCTTT	TCAAAAAAAA	AAAAAAA			2377

Fig. _1

MRLCDLSSLA	SWVLVTVALP	SSGAAAKGVS	QIPSTHSSQS	KGNGPWAHAY	50
RRAEKLVRQM	TLEEKANITR	GFTGDNVCAG	NTGSVPRLGW	PGMCVHDAGN	100
GVRATDLVNS	YPSGIHVGAS	WDRNLTYERG	LHMGGEFKAK	GVMVPLGPNA	150
GPLGRTPGG	RNWEFSDIP	YLSGQLNAET	ITGMQDAGVI	ANIKHFIANE	200
QETLRRPYFG	VEAVSANIDD	RTLHEYLLWP	FMDSVHAGVG	SVMCSYNRIN	250
NTYGCMNDKL	MNGILKAELG	FQGFVMLDWN	AQHDLQSANA	GLDMVMPLGG	300
SWGKNLDAV	ANGTVSESRI	TDMATRIIAA	WYLVGQDGNN	FPVPGIGLKQ	350
LTKPHEQVDA	RDFASKPVLL	EGAIAGHVLV	KNENNALPFN	KKLTMISVFG	400
YDATIPRTKN	TDILFQLGYT	SSPEMAQAVL	GNEAHFDQAA	KGGTIMTGGR	450
AGANAPSYID	DPLAAIQRRR	RKDDTWVNW	LDSFNPEVNA	ASDACLVFIN	500
AIATEGWDRD	GLHDDFSDGL	VLNVAANCSN	TIVVVHAAGT	RLVDQWIEHP	550
NVTAAVIAHL	PGQDSGRALV	KLLYGEANFS	GKLPYTIAKN	ESDYSVYTPC	600
QRRSPEDTDP	QCDFTEGVYL	DYRAFDANNM	TPRFEFGYGL	SYTSEFNYSAL	650
SIKKAKGLRQ	SRCTDDLWQA	AAQVTASITN	SGMSGSEVA	QLYLAI PNSP	700
PKQLRGFNKL	LLRPHESGTV	HFGLTKRDLS	VWDVVSQSWV	IQEGEYKVFV	750
GASSRDIRLS	GKLHI				765

Fig. 2