(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 578 928**

(51) Int. Cl.:

C12P 7/06	(2006.01)
C12P 7/14	(2006.01)
C12P 7/16	(2006.01)
C12P 21/06	(2006.01)
C12N 1/20	(2006.01)
C12N 9/24	(2006.01)
C12N 15/00	(2006.01)
C07H 21/04	(2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **30.10.2002** **E 02792218 (6)**

(97) Fecha y número de publicación de la concesión europea: **27.04.2016** **EP 1472359**

(54) Título: **Endoglucanasa EGVII y ácidos nucleicos que codifican la misma**

(30) Prioridad:

18.12.2001 US 28244

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
02.08.2016

(73) Titular/es:

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

(72) Inventor/es:

DUNN-COLEMAN, NIGEL;
WARD, MICHAEL;
YAO, JIAN y
GOEDEGEBUUR, FRITS

(74) Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 578 928 T3

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

DESCRIPCIÓN

Endoglucanasa EGVII y ácidos nucleicos que codifican la misma

Campo

La presente descripción se refiere a secuencias de ácido nucleico *egl7* aisladas que codifican polipéptidos que poseen actividad endoglucanasa. La descripción se refiere a constructos de ácido nucleico, vectores, y células huésped que comprenden las secuencias de ácido nucleico así como métodos para producir polipéptidos EGVII recombinantes.

Referencias

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, April 13, 2001.

Aubert, *et. al.*, Ed., p11 et seq., Academic Press, 1988.

Ausubel, G. M., et. al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New 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.

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.

Cummings and Fowler, Curr. Genet. 29:227-233, 1996.

Dayhoff et al. in Atlas of Protein Sequence and Structure, Volume 5, Supplement 3, Chapter 22, pp. 345-352, 1978.

Deutscher, M.P., Methods Enzymol. 182:779-80, 1990.

Doolitle, R.F., OF URFS AND ORFs, University Science Books, CA, 1986.

Ellouz, S. et al., J. Chromatography 396:307, 1987.

Fields and Song, Nature 340:245-246, 1989.

Filho, et al. J. Microbiol. 42:1-5, 1996.

Fliess, A., et al., Eur. J. Appl. Microbiol. Biotechnol. 17:314, 1983.

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.

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, et al, Curr. Opin. Biotechnol. 6(5):561-6, 1995.

Jones et al., Nature 321:522-525, 1986.

Kawaguchi, T et al., Gene 173(2):287-8, 1996.

Knowles, J. et al., TIBTECH 5, 255-261, 1987.

Kohler and 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.

Lehtio, J. et al., FEMS Microbiology Letters 195:197-204, 2001.

Li and Ljungdahl Appl. Environ. Microbiol. 62:209-213, 1996.

Linder, M. and 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.

Ooi et al., Nucleic Acids Res. 18(19):5884, 1990.

Ortega et al., International Biodeterioration and Biodegradation 47:7-14, 2001.

Penttilia et al., Yeast 3:175-185, 1987.

Penttilia et al., Gene 63: 103-112, 1988.

Pere, J., et al., In Proc. Tappi. Pulping Conf., Nashville, TN, 27-31, pp. 693-696, 1996.

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.

Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989.

Schulein, Methods Enzymol., 160, 25, pages 234 et seq, 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.

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.

Tomaz, C. and 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 14:61-76, 1998.

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.

Wood, Biochem. Soc. Trans., 13, pp. 407-410, 1985.

Wood et al., METHODS IN ENZYMOLOGY, 160, 25, p. 87 et seq., Academic Press, New York, 1988.

Zoller et al., Nucl. Acids Res. 10:6487, 1987.

Antecedentes

La celulosa y la hemicelulosa son los materiales de plantas más abundantes producidos por fotosíntesis. Pueden ser degradados y usados como una fuente de energía por numerosos microorganismos, incluyendo bacterias, levaduras y hongos, que producen enzimas extracelulares capaces de la hidrólisis de los sustratos poliméricos a azúcares monoméricos (Aro *et al.,* 2001). Según se acercan los límites de los recursos no-renovables, el potencial de la celulosa para ser el principal recurso de energía renovable es enorme (Krishna *et al.,* 2001). La utilización eficaz de la celulosa a lo largo de los procesos biológicos es uno de los enfoques para superar la escasez de alimentos, piensos y combustibles (Ohmiya *et al.,* 1997).

Las celulasas son enzimas que hidrolizan celulosa (enlaces beta-1,4-glucan o beta D-glucosídicos) dando como resultado la formación de glucosa, celobiosa, celooligosacáridos, y similares. Las celulasas se han dividido tradicionalmente en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) ("CBH") y beta-glucosidasas (beta-D-glucosido 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 son también capaces de degradar la celulosa cristalina (Nevalainen and Penttila, 1995). Así, se requiere la presencia de una celobiohidrolasa en un sistema celulasa para la solubilización eficiente de la celulosa cristalina (Suurnakki, *et al.,* 2000). La beta-glucosidasa actúa para liberar unidades de D-glucosa a partir de la celobiosa, celooligosacáridos, y otros glucósidos (Freer, 1993).

Se sabe que las celulasas son producidas por un gran número de bacterias, levaduras y hongos. Ciertos hongos producen un sistema celulasa completo capaz de degradar formas cristalinas de celulosa, de tal manera que las celulasas son producidas fácilmente en grandes cantidades a través de la fermentación. Los hongos filamentosos juegan un papel esencial ya que muchas levaduras, como *Saccharomyces cerevisiae*, carecen de la capacidad de hidrolizar celulosa. Adviértase en, p.ej., Aro *et al.,* 2001; Aubert *et al.,* 1988; Wood *et al.,* 1988, y Coughlan, *et al..*

Las clasificaciones de la celulasa fúngica de CBH, EG y BG se pueden ampliar más a fondo para incluir múltiples componentes dentro de cada clasificación. Por ejemplo, se han aislado múltiples CBHs, EGs y BGs de una variedad de fuentes fúngicas incluyendo *Trichoderma reesei* que contiene genes conocidos para 2 CBHs, esto es, CBH I y CBH II, al menos 5 EGs, *esto es*, EG I, EG II, EG III, EG IV y EGV, y al menos 2 BGs, esto es, BG1 y BG2.

Para convertir eficazmente la celulosa cristalina en glucosa se necesita el sistema celulasa completo que comprende componentes de cada una de las clasificaciones de CBH, EG y BG, con componentes aislados menos eficaces en hidrolizar celulosa cristalina (Filho *et al.,* 1996). Se ha observado una relación sinérgica entre los componentes de la celulasa de diferentes clasificaciones. En particular, las celulasas tipo-EG y las celulasas tipo-CBH interactúan sinérgicamente para degradar de manera más eficiente la celulosa. Adviértase en, *p.ej.,* Wood, 1985.

Las celulasas son conocidas en la técnica por ser útiles en el tratamiento de textiles a efectos de mejorar la capacidad de limpieza de las composiciones de detergentes, para su uso como un agente suavizante, para mejorar el tacto y el aspecto de los tejidos de algodón, y similares (Kumar *et al.,* 1997).

Se han descrito las composiciones de detergente que contienen celulasa con un mejor rendimiento de limpieza (Pat. EE.UU. Nº. 4.435.307; GB App. Nºs. 2.095.275 y 2.094.826) y para su uso en el tratamiento de tejidos para mejorar el tacto y el aspecto del textil (Pat EE.UU. Nºs. 5.648.263, 5.691.178 y 5.776.757; GB App. Nº. 1.358.599; The Shizuoka Prefectural Hammamatsu Textile Industrial Research Institute Report, Vol. 24, pp.54-61, 1986).

Por consiguiente, las celulasas producidas por hongos y bacterias han recibido mucha atención. En particular, se ha mostrado que la fermentación de *Trichoderma spp. (p.ej., Trichoderma longibrachiatum o Trichoderma reesei)* produce un sistema de celulasa completo capaz de degradar las formas cristalinas de celulosa. En la Pat. EE.UU.

Nº. 5.475.101 se describe la purificación y la clonación molecular de una enzima útil en particular designada EGIII que se deriva de *Trichoderma tongibrachiatum*.

SALOHEIMO M. ET AL., EUROPEAN JOURNAL OF BIOCHEMISTRY/ FEBS. 15 OCT 1997, vol. 249, nº. 2, páginas 584-591, describe un cDNA de *Trichoderma reesei* que codifica una proteína con un C-terminal de unión a celulosa. La proteína presenta una similitud de aminoácidos en toda su longitud con la proteína inducida por celulosa CEL1 de Agaricus hisporus. Estas dos proteínas son miembros de la familia de las glicosil hidrolasas, número 61. La expresión del cDNA T. reesei en levadura mostró que codifica una proteína con actividad endoglucanasa y por lo tanto a la proteína se la llamó EGIV y al gen correspondiente egl4. Se prepararon anticuerpos policlonales frente a EGIV producida en *Escherichia coli* y detectaron una proteína de 56 kDa en el sobrenadante de cultivo de *T. reesei.* Mediante hibridación por Northern se reveló que eg14 de *T. reesei* está regulado de la misma manera que otros genes de celulasa de este hongo.

Aunque se han descrito previamente las composición de celulasa, sigue habiendo una necesidad de nuevas y mejoradas composiciones de celulasa para su uso en detergentes domésticos, composiciones para el lavado a la piedra o detergentes de lavandería, etc. Las celulasas que exhiben resistencia a los tensoactivos (p.ej., sulfonatos alquilo lineales, LAS), rendimiento mejorado bajo condiciones de estrés térmico, capacidad celulolítica aumentada o disminuida, y/o alto nivel de expresión *in vitro*, son de particular interés.

Compendio

La descripción proporciona una proteína celulasa aislada, identificada en el presente documento como EGVII, y ácidos nucleicos que codifican EGVII.

En un aspecto, EGVII, polipéptidos o proteínas comprenden una secuencia que tiene al menos un 80%, 85%, 90%, 95%, 98% o más de identidad de secuencia con la secuencia presentada como SEQ ID NO:2.

En un aspecto relacionado, la descripción incluye (i) fragmentos de EGVII, preferiblemente al menos aproximadamente 20-100 aminoácidos en longitud, más preferiblemente aproximadamente 100-200 aminoácidos en longitud, e (ii) una composición farmacéutica que comprende EGVII. En varias realizaciones, el fragmento corresponde al dominio N-terminal de EGVII o al dominio C-terminal de EGVII.

En otro aspecto la descripción incluye un polinucleótido aislado que tiene una secuencia que codifica EGVII, una secuencia complementaria a la secuencia codificadora de *egl7*, y una composición que comprende el polinucleótido. El polinucleótido puede ser mRNA, DNA, cDNA, DNA genómico, o un análogo antisentido del mismo.

Un polinucleótido *egl7* puede comprender una molécula de ácido nucleico aislada que hibrida con el ácido nucleico presentado como SEQ ID NO:1 bajo condiciones de astringencia de moderada a alta, donde la molécula de ácido nucleico codifica un polipéptido EGVII que exhibe actividad endoglucanasa.

El polinucleótido puede codificar una proteína EGVII que tiene al menos un 80%, 85%, 90%, 95%, 98% o más identidad de secuencia con la secuencia presentada como SEQ ID NO:1. En una realización específica, el polinucleótido comprende una secuencia sustancialmente idéntica a SEQ ID NO:1. La descripción también contempla fragmentos del polinucleótido, preferiblemente al menos aproximadamente de 15-30 nucleótidos en longitud.

La descripción proporciona adicionalmente vectores de expresión recombinantes que contienen una secuencia de ácido nucleico que codifica EGVII o un fragmento o variante de corte y empalme del mismo, operativamente unido a elementos reguladores eficaces para la expresión de la proteína en un huésped seleccionado. En un aspecto relacionado, la descripción incluye una célula huésped que contiene el vector.

La descripción incluye además un método para producir EGVII mediante técnicas recombinantes, mediante cultivo de células huésped procariotas o eucariotas recombinantes que comprenden secuencia de ácido nucleico que codifica EGVII bajo condiciones eficaces de promover la expresión de la proteína, y posterior recuperación de la proteína a partir de la célula huésped o del medio de cultivo celular.

En otro aspecto aún, la descripción incluye un anticuerpo específicamente inmunoreactivo frente a EGVII.

También forman parte de la descripción métodos analíticos para detectar los ácidos nucleicos *ecl7* y las proteínas EGVII.

En otro aspecto la descripción proporciona una composición enzimática útil en la conversión de celulosa en azúcares y/o etanol. En una realización preferida la composición enzimática comprende EGVII. La composición puede comprender además enzimas celulasa adicionales como otras endoglucanasas, beta-glucosidasas y/o celobiohidrolasas. La composición se puede enriquecer en EGVII.

Breve descripción de la figuras

La Figura 1 es una representación monocatenaria de una secuencia de ácido nucleico (SEQ ID NO:1), del cDNA de *egl7* de *T, reesei,* en donde la secuencia no-codificadora se indica en negrita.

La Figura 2 muestra la secuencia de aminoácidos predicha (SEQ ID NO:2) y la secuencia señal (SEQ ID NO:3) basada en la secuencia de nucleótidos proporcionada en la Figura 1, en donde la secuencia señal se indica en negrita.

Descripción detallada

I. Definiciones

A menos que se indique lo contrario, todos los términos técnicos y científicos empleados en esta memoria tienen el mismo significado como lo serían para un experto en la técnica. Los profesionales se dirigen particularmente a Sambrook *et al.,* 1989, y Ausubel FM *et al.,* 1993, para las definiciones y términos de la técnica. Se ha de entender que esta descripción no se limita a la metodología particular, métodos, y reactivos descritos, ya que éstos pueden variar.

El término “polipéptido” como se emplea en esta memoria se refiere a un compuesto formado por una sola cadena de residuos de aminoácidos unidos por enlaces peptídicos. El término “proteína” como se como se emplea en esta memoria puede ser sinónimo con el término “polipéptido” o se puede referir, además, a un complejo de dos o más polipéptidos.

El término “molécula de ácido nucleico” incluye moléculas de RNA, DNA y cDNA. Se ha de entender que, como resultado de la degeneración del código genético, se pueden producir una multitud de secuencias de nucleótidos que codifican una proteína dada como EGVII. La presente descripción contempla cada posible variante de secuencia de nucleótidos, que codifica EGVII, todas las cuales son posibles dada la degeneración del código genético.

Un constructo o secuencia de ácido nucleico “heterólogo” tiene una porción de secuencia, que no es nativa a la célula en que se expresa. Heterólogo, con respecto a una secuencia control se refiere a una secuencia control (esto es, promotor o potenciador) que no funciona en la naturaleza para regular al mismo gen cuya expresión está actualmente regulando. Generalmente, las secuencias de ácido nucleico heterólogo no son endógenas a la célula o a parte del genoma en el que están presentes, y se han añadido a la célula, por infección, transfección, transformación, microinyección, electroporación, o similares. Un constructo de ácido nucleico “heterólogo” puede contener una combinación de una secuencia control/secuencia de DNA codificadora que es la misma que, o diferente de una combinación de una secuencia control/secuencia de DNA codificadora encontrada en la célula nativa.

Como se emplea en esta memoria, el término "vector" se refiere a un constructo de ácido nucleico designado para la transferencia entre diferentes células huésped. Un “vector de expresión” se refiere a un vector que tiene la capacidad de incorporar y expresar fragmentos de DNA heterólogo en una célula extraña. Muchos vectores de expresión procariotas y eucariotas están disponibles comercialmente. La selección de los vectores apropiados está dentro del conocimiento de los expertos en la técnica.

Por consiguiente, un “casete de expresión” o “vector de expresión” es un constructo de ácido nucleico generado de manera recombinante o sintética, con una serie de elementos de ácido nucleico especificados que permiten la transcripción de un ácido nucleico particular en la célula diana. El casete de expresión recombinante se puede incorporar dentro de un plásmido, cromosoma, DNA mitocondrial, DNA plastídico, virus, o fragmento de ácido nucleico. Normalmente, la porción del casete de expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico que se transcribe y un promotor.

Como se emplea en esta memoria, el término “plásmido” se refiere a un constructo de DNA circular de doble cadena (ds) usado como un vector de clonación, y que forma un elemento genético extra-cromosómico auto-replicante en muchas bacterias y en algunas eucariotas.

Como se emplea en esta memoria, el término “secuencia de nucleótidos seleccionable con marcador-codificado” se refiere a una secuencia de nucleótidos que es capaz de expresarse en células y donde la expresión del marcador seleccionable confiere a las células que contienen el gen expresado la habilidad de crecer in presencia de un agente selectivo correspondiente, o bajo condiciones selectivas de crecimiento correspondientes.

Como se emplea en esta memoria, el término “promotor” se refiere a una secuencia de ácido nucleico que funciona para dirigir la transcripción de un gen corriente abajo. El promotor generalmente será apropiado a la célula huésped en la que el gen diana está siendo expresado. El promotor junto con otras secuencias de ácido nucleico reguladoras transcripcionales y traduccionales (también denominadas “secuencias control”) son necesarias para expresar un gen dado. En general, las secuencias reguladoras transcripcionales y traduccionales incluyen, pero no se limitan a, secuencias promotoras, sitios de unión al ribosoma, secuencias de inicio y de parada de la transcripción, secuencias de inicio y de parada de la traducción, y secuencias potenciadoras y activadoras.

“Gen quimérico” o “constructo de ácido nucleico heterólogo”, como se define en este documento se refiere a un gen no-nativo (esto es, uno que se ha introducido dentro de un huésped) que se puede componer de partes de distintos genes, incluyendo elementos reguladores. Un constructo de gen quimérico para la transformación de una célula huésped se compone normalmente de una región reguladora transcripcional (promotor) ligada operativamente a una secuencia codificadora de una proteína heteróloga, o, en un gen quimérico con un marcador seleccionable, a un gen con un marcador seleccionable que codifica una proteína que confiere resistencia a antibióticos a la célula transformada. Un gen quimérico típico para la transformación en una célula huésped, incluye una región reguladora transcripcional que es constitutiva o inducible, una secuencia que codifica una proteína, y una secuencia terminadora. Un constructo de gen quimérico puede incluir también una segunda secuencia de DNA que codifica un péptido señal si se desea la secreción de la proteína diana.

Un ácido nucleico está “operativamente ligado” cuando está localizado dentro de una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, DNA que codifica un líder secretorio está operativamente ligado a un DNA para un polipéptido si se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o potenciador está operativamente ligado a una secuencia codificadora si se afecta la transcripción de la secuencia; o un sitio de unión al ribosoma está operativamente ligado a una secuencia codificadora si está posicionado con el fin de facilitar la traducción. Generalmente, “operativamente ligado” significa que las secuencias de DNA que están ligadas son contiguas, y, en el caso del líder secretorio, contiguo y en marco de lectura. Sin embargo, los potenciadores no tienen que estar contiguos. La ligación se lleva a cabo por ligación en sitios de restricción convenientes. Si tales sitios no existen, se usan los adaptadores de oligonucleótidos sintéticos, enlazadores o cebadores para la PCR, de acuerdo con la práctica convencional.

Como se emplea en esta memoria, el término “gen” significa el segmento de DNA involucrado en la producción de una cadena polipeptídica, que puede o no incluir regiones que preceden o siguen a la región codificadora, *p.ej.* 5’ no traducida (5’ UTR) o secuencias “líder” y 3’ UTR o secuencias “remolque”, así como secuencias intermedias (intrones) entre segmentos codificadores individuales (exones).

En general, las moléculas de ácido nucleico que codifican EGVII, o un análogo u homólogo de la misma hibridará, bajo condiciones de astringencia de moderadas a altas con la secuencia que se proporciona en este documento como SEQ ID NO:1. Sin embargo, en algunos casos se emplea una secuencia que codifica EGVII que posee una utilización de codón sustancialmente diferente, mientras que la proteína codificada por la secuencia de nucleótido que codifica EGVII tiene la misma o sustancialmente la misma secuencia de aminoácidos que la proteína nativa. Por ejemplo, la secuencia codificadora se puede modificar para facilitar una expresión más rápida de EGVII en un sistema de expresión procariota o eucariota particular, de acuerdo con la frecuencia con la que un codón particular es utilizado por el huésped. Te’o, *et al.* (2000), por ejemplo, describe la optimización de genes para la expresión en hongos filamentosos.

Se considera que una secuencia de ácido nucleico es “selectivamente hibridable” con un ácido nucleico de referencia si las dos secuencias hibridan específicamente una con la otra bajo una astringencia de hibridación de moderada a alta y condiciones de lavado. Las condiciones de hibridación se basan en la temperatura de fusión (Tm) del complejo de unión de ácido nucleico o de la sonda. Por ejemplo, “astringencia máxima” normalmente ocurre a aproximadamente Tm-5ºC (5º por debajo de la Tm de la sonda); “astringencia alta” a aproximadamente 5-10º por debajo de la Tm; “astringencia intermedia” a aproximadamente 10-20º por debajo de la Tm de la sonda; y “astringencia baja” a aproximadamente 20-25º por debajo de la Tm. Funcionalmente, las condiciones de máxima astringencia se pueden usar para identificar secuencias que tienen identidad estricta o identidad casi estricta con la sonda de hibridación; mientras que las condiciones de astringencia alta se usan para identificar secuencias que tienen aproximadamente 80% o más de identidad de secuencia con la sonda.

Son bien conocidas en la técnica las condiciones de hibridación de astringencia moderada y alta (adviértase en, por ejemplo, Sambrook, *et al,* 1989, Capítulos 9 y 11, y en Ausubel, F.M., *et al.,* 1993). Un ejemplo de condiciones de astringencia alta incluye hibridación a aproximadamente 42ºC en formamida al 50%, SSC 5X, solución de Denhardt 5X, SDS al 0,5% y DNA portador desnaturalizado 100μg/ml seguido de lavado dos veces en SSC 2X y SDS al 0,5% a temperatura ambiente y dos veces adicionales en SSC 0,1X y SDS al 0,5% a 42ºC.

Como se emplea en esta memoria, “recombinante” incluye la referencia a una célula o vector, que ha sido modificada/o mediante la introducción de una secuencia de ácido nucleico heterólogo o que la célula se derive de una célula así modificada. Así, por ejemplo, las células recombinantes expresan genes que no se encuentran en forma idéntica dentro de la forma nativa (no-recombinante) de la célula o expresan genes nativos que se expresan anormalmente de otro modo, expresados a la baja o no expresados en absoluto como resultado de una intervención humana deliberada.

Como se emplea en esta memoria, los términos “transformado/a”, “establemente transformado/a” o “transgénico” con referencia a una célula significa que la célula tiene una secuencia de ácido nucleico no-nativa (heteróloga) integrada en su genoma o como un plásmido episomal que se mantiene a lo largo de múltiples generaciones.

Como se emplea en esta memoria, el término “expresión” se refiere al proceso mediante el cual un polipéptido se produce en función de la secuencia de ácido nucleico de un gen. El proceso incluye ambas, transcripción y traducción.

El término “introducido” en el contexto de insertar una secuencia de ácido nucleico dentro de una célula significa “transfección”, o “transformación” o “transducción” e incluye referencia a la incorporación de una secuencia de ácido nucleico dentro de una célula eucariota o procariota donde la secuencia de ácido nucleico puede ser incorporada dentro del genoma de la célula (por ejemplo, cromosoma, plásmido, plástido, o DNA mitocondrial), convertido en un replicón autónomo, o expresado transitoriamente (por ejemplo, mRNA transfectado).

Se deduce que el término “expresión de EGVII” se refiere a la transcripción y la traducción del gen *egl7*, los productos de las cuales incluyen RNA precursor, mRNA, polipéptido, polipéptidos procesados post-traduccionalmente y derivados de los mismos, incluyendo EGVII de especies relacionadas como *Trichoderma longibrachiatum* (*reesei*), *Trichoderma viride, Trichoderma koningii, Hypocrea jecorina y Hypocrea schweinitzii.* A modo de ejemplo, ensayos para la expresión de EGVII incluyen transferencia de Western para la proteína EGVII, análisis de transferencia de Northern y ensayos de reacción en cadena de la polimerasa con transcriptasa reversa (RT-PCR) para el mRNA de EGVII, ensayos de actividad de endoglucanasa como se describen en Shoemaker S.P. y Brown R.D. Jr. (Biochim. Biophys. Acta, 1978, 523:133-146) y Schulein (1988).

El término “corte y empalme alternativo” se refiere al proceso por el cual se generan múltiples isoformas polipeptídicas a partir de un único gen. Así un exón particular se podría conectar a cualquiera de los varios exones alternativos para formar RNAs mensajeros. Los mRNAs alternativamente cortados y empalmados producen polipéptidos (“variantes de corte y empalme”) en los que algunas partes son comunes mientras otras partes son diferentes.

El término “secuencia señal” se refiere a una secuencia de aminoácidos en la porción 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 secuencia señal la cual es escindida durante el proceso de secreción.

Mediante el término “célula huésped” se entiende una célula que contiene un vector y mantiene la replicación, y/o transcripción o transcripción y traducción (expresión) del constructo de expresión. Las células huésped 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.

El término “hongos filamentosos” significa cualquier y todos los hongos filamentosos reconocidos por los expertos en la técnica. Un hongo preferido se selecciona del grupo que consiste en Aspergillus, Trichoderma, Fusarium, Chrysosporium, Penicillium, Humicola, Neurospora, o formas sexuales alternativas de los mismos como Emericella, Hypocrea.

El término “celooligosacárido” se refiere a grupos de oligosacáridos que contienen de 2-8 unidades de glucosa y que tienen enlaces β-1,4, p.ej. celobiosa.

El término “celulasas” se refiere a una categoría de enzimas capaces de hidrolizar polímeros de celulosa a oligómeros más cortos de celo-oligosacáridos, celobiosa y/o glucosa. Numerosos ejemplos de celulasas, como exoglucanasas, exocelobiohidrolasas, endoglucanasas, y glucosidasas se han obtenido de organismos celulolíticos, particularmente incluyendo hongos, plantas y bacterias.

El término “dominio de unión a celulosa” como se emplea en esta memoria se refiere a una porción de la secuencia de aminoácidos de una celulasa o una región de la enzima que está involucrada en la actividad de unión a celulasa o un derivado de la misma. Los dominios de unión a celulosa generalmente funcionan mediante la unión no-covalente de la celulasa a la celulosa, un derivado de celulosa u otro polisacárido equivalente a la misma. Los dominios de unión a celulosa permiten o facilitan la hidrólisis de las fibras de celulosa mediante la región central estructuralmente distinta y funciona normalmente independiente del centro catalítico. Así, un dominio de unión a celulosa no tendrá la actividad hidrolítica significativa atribuible al centro catalítico. En otras palabras, un dominio de unión a celulosa es un elemento estructural de la estructura terciaria de la proteína enzima celulasa que es distinto del elemento estructural que tiene actividad catalítica.

Como se emplea en esta memoria, el término “tensoactivo” se refiere a cualquier compuesto generalmente reconocido en la técnica que tiene cualidades activas de superficie. Así, por ejemplo, los tensoactivos comprenden tensoactivos aniónicos, catiónicos y no iónicos como los que se encuentran comúnmente en detergentes. Los tensoactivos aniónicos incluyen alquilbencenosulfonatos lineales o ramificados; alquil o alquenil éter sulfatos que tienen grupos alquilo o grupos alquenilo lineales o ramificados.; alquil o alquenil sulfatos; o sulfatos de olefina; y alcanosulfonatos. Tensoactivos anfolíticos incluyen sales de sulfonatos de amonio cuaternario, y tensoactivos anfolíticos tipo betaína. Tales tensoactivos anfolíticos tienen tanto grupos de carga positiva como negativa en la misma molécula. Los tensoactivos no iónicos pueden comprender éteres de polioxiaquilenos, así como alcanolamidas de ácidos grasos superiores o un aducto de óxido de alquileno de los mismos, monoésteres de glicerina de ácidos grasos, y similares.

Como se emplea en esta memoria, el término “tejido que contiene celulosa” se refiere a cualquier tejido cosido o no cosido, hilados o fibras hechas de algodón o no algodón que contiene celulosa, o algodón o no-algodón que contiene mezclas de celulosa incluyendo celulósicos naturales y celulósicos artificiales (como yute, lino, ramio, rayón, y lyocell).

Como se emplea en esta memoria, el término “tejido que contiene algodón” se refiere a tejidos cosidos o no cosidos, hilos o fibras hechas de algodón puro o mezclas de algodón que incluyen tejidos de algodón, tejidos de punto de algodón, drils de algodón, hilos de algodón, algodón en bruto y similares.

Como se emplea en esta memoria, el término “composición de lavado a la piedra” se refiere a una formulación para su uso en el lavado a la piedra de tejidos que contienen celulosa. Las composiciones de lavado a la piedra se usan para modificar los tejidos que contienen celulosa antes de la venta, esto es, durante el proceso de fabricación. Por el contrario, las composiciones de detergente están destinadas para la limpieza de prendas sucias y no se usan durante el proceso de fabricación.

Como se emplea en esta memoria, el término “composición de detergente” se refiere a una mezcla que está destinada para su uso en un medio de lavado para el lavado de tejidos sucios que contienen celulosa.

En el contexto de la presente descripción dichas composiciones pueden incluir, además de celulasas y tensoactivos, enzimas hidrolíticas adicionales, detergentes de lavandería, agentes blanqueadores, activadores de blanqueo, agentes de azulado y colorantes fluorescentes, inhibidores de apelmazamiento, agentes enmascaradores, activadores de celulasa, antioxidantes, y solubilizantes.

Como se emplea en esta memoria, el término “disminución o eliminación de la expresión del gen *egl7*” significa que o bien el gen *egl7* ha sido eliminado del genoma y por lo tanto no puede ser expresado por el microorganismo huésped recombinante; o que el gen *egl7* ha sido modificado de tal manera que no se produce la enzima EGVII funcional por el microorganismo huésped recombinante.

El término “*egl7* alterado” o “gen *egl7* alterado” significa que la secuencia de ácido nucleico del gen se ha alterado mediante eliminación, adición, y/o manipulación de la secuencia codificadora o se ha modificado la secuencia de aminoácidos de la proteína expresada.

Como se emplea en esta memoria, el término “purificación” se refiere generalmente a someter el ácido nucleico transgénico o las células que contienen la proteína a purificación bioquímica y/o columna cromatográfica.

Como se emplea en esta memoria, los términos “activo” y “biológicamente activo” se refieren a una actividad biológica asociada con una proteína particular, como la actividad enzimática asociada con una proteasa. Se deduce que la actividad biológica de una proteína dada se refiere a cualquier actividad biológica normalmente atribuida a esa proteína por los expertos en la técnica.

Como se emplea en esta memoria, el término “enriquecido” significa que el EGVII se encuentra en una concentración que es mayor en relación a la concentración de EGVII encontrada en una composición de celulasa fúngica de tipo-salvaje, o que ocurre naturalmente. Los términos enriquecido, elevado y mejorado se puede usar de manera intercambiable en esta memoria.

Una composición de celulasa fúngica de tipo salvaje es una producida por una fuente fúngica de origen natural y que comprende uno o más componentes BGL, CBH y EG en donde cada uno de estos componentes se encuentra en la proporción producida por la fuente fúngica. Así, una composición EGVII enriquecida tendría EGVII en una proporción alterada en donde la proporción de EGVII con respecto a otros componentes de celulasa (esto es, CBHs, beta-glucosidadsas y otras endoglucanasas) está elevada. Esta proporción puede ser aumentada ya sea aumentando EGVII o disminuyendo (o eliminando) al menos uno de los componentes por cualquier medio conocido en la técnica.

Por lo tanto, para ilustrar, un sistema de celulasa de origen natural se puede purificar en componentes sustancialmente puros mediante técnicas de separación reconocidas bien publicadas en la literatura, incluyendo cromatografía de intercambio iónico a un pH adecuado, cromatografía de afinidad, exclusión por tamaño y similares. Por ejemplo, en la cromatografía de intercambio iónico (generalmente cromatografía de intercambio aniónico), es posible separar los componentes de celulasa median elución con un gradiente de pH, o un gradiente de sal, o ambos un gradiente de pH y un gradiente de sal. El EGVII purificado se puede entonces añadir a la solución enzimática dando como resultado una solución enriquecida de EGVII. También es posible elevar la cantidad de EGVII producido por un microbio usando métodos de genética molecular para sobreexpresar el gen que codifica EGVII, posiblemente en conjunción con la supresión de uno o más genes que codifican otras celulasas.

Las celulasas fúngicas pueden contener más de un componente EGVII. Los diferentes componentes generalmente tienen diferentes puntos isoeléctricos que permiten su separación por cromatografía de intercambio iónico y similares. Se pueden utilizar o bien un único componente EG o una combinación de componentes EG en una solución enzimática.

Cuando se emplean en soluciones enzimáticas, el componente EG se añade generalmente en una cantidad suficiente para permitir la tasa más alta de liberación de los azúcares solubles de la biomasa. La cantidad de componente EG añadido depende del tipo de biomasa que se sacarifica el cual puede ser determinado fácilmente por el experto en la materia. Sin embargo, cuando se emplea, el porcentaje en peso del componente EGVII en relación con cualquiera de los componentes de tipo CBH presentes en la composición de celulasa es de preferiblemente aproximadamente 1, preferiblemente aproximadamente 5, preferiblemente aproximadamente 10, preferiblemente aproximadamente 15, o preferiblemente aproximadamente 20% en peso hasta preferiblemente aproximadamente 25, preferiblemente aproximadamente 30, preferiblemente aproximadamente 35, preferiblemente aproximadamente 40, preferiblemente aproximadamente 45 o preferiblemente aproximadamente 50% en peso. Además, los intervalos preferidos pueden ser de aproximadamente 0,5 a aproximadamente 15% en peso, aproximadamente 0,5 a aproximadamente 20% en peso, de aproximadamente 1 a aproximadamente 10% en peso, de aproximadamente 1 a aproximadamente 15% en peso, de aproximadamente 1 a aproximadamente 20% en peso, de aproximadamente 1 a aproximadamente 25% en peso, de aproximadamente 5 a aproximadamente 20% en peso, de aproximadamente 5 a aproximadamente 25% en peso, de aproximadamente 5 a aproximadamente 30% en peso, de aproximadamente 5 a aproximadamente 35% en peso, de aproximadamente 5 a aproximadamente 40% en peso, de aproximadamente 5 a aproximadamente 45% en peso, de aproximadamente 5 a aproximadamente 50% en peso, de aproximadamente 10 a aproximadamente 20% en peso, de aproximadamente 10 a aproximadamente 25% en peso, de aproximadamente 10 a aproximadamente 30% en peso, de aproximadamente 10 a aproximadamente 35% en peso, de aproximadamente 10 a aproximadamente 40% en peso, de aproximadamente 10 a aproximadamente 45% en peso, de aproximadamente 10 a aproximadamente 50% en peso, de aproximadamente 15 a aproximadamente 20% en peso, de aproximadamente 15 a aproximadamente 25% en peso, de aproximadamente 15 a aproximadamente 30% en peso, de aproximadamente 15 a aproximadamente 35% en peso, de aproximadamente 15 a aproximadamente 30% en peso, de aproximadamente 15 a aproximadamente 45% en peso, de aproximadamente 15 a aproximadamente 50% en peso.

II. Organismos diana

A. Hongos filamentosos

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, quitosano, manano y otros polisacáridos complejos, con un crecimiento vegetativo por elongación de hifas y un catabolismo de carbono que es estrictamente aeróbico.

La célula parental fúngica filamentosa puede ser una célula de una especie de, pero no limitada a, *Trichoderma. e.g., Trichoderma longibrachiatum (reesei), Trichoderma viride, Trichoderma koningii, Trichoderma harzianum; Penicillium sp.; Numicola sp.,* incluyendo *Humicola insolens; Chrysosporium sp.,* incluyendo C. lucknowense; *Gliocladium sp,; Aspergillus sp.; Fusarium sp., Neurospora sp., Hypocrea sp., y Emericella sp.* Como se emplea en esta memoria, el término "*Trichoderma*" o "*Trichoderma sp.*" se refiere a cualquier cepa de hongos que se ha clasificado previamente como *Trichoderma* o está actualmente clasificada como *Trichoderma*.

En una realización preferida, la célula parental de hongo filamentoso es una célula *Aspergillus niger, Aspergillus awamori, Aspergillus aculeatus,* o *Aspergillus nidulans.*

En otra realización preferida, la célula parental de hongo filamentoso es una célula *Trichoderma reesei.*

III. Celulasas

Las celulasas son conocidas en la técnica como enzimas que hidrolizan celulosa (enlaces beta-1,4-glucan o beta D-glucosídicos) dando como resultado la formación de glucosa, celobiosa, celooligosacáridos, y similares. Como se expuso anteriormente, las celulasas se han dividido tradicionalmente en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) ("CBH") y beta-glucosidasas (EC 3.2.1.21) ("BG"). (Knowles, *et al.,* 1987; Schulein, 1988).

Ciertos hongos producen sistemas completos de celulasa que incluyen exo-celobiohidrolasas o celulasas tipo-CBD, endoglucanasas o celulasas tipo-EG y beta-glucosidasas o celulasas tipo-BG (Schulein, 1988). Sin embargo, algunas veces estos sistemas carecen de celulasas tipo-CBH y las celulasas bacterianas también incluyen normalmente pocas o ninguna celulasa. Además, se ha mostrado que los componentes EG y los componentes CBH interactúan sinérgicamente para degradar de manera más eficiente la celulosa. Adviértase p.ej., en Wood, 1985. Los diferentes componentes, esto es, las distintas endoglucanasas y exocelobiohidrolasas en un sistema de celulasa completo o multi-componente, tienen generalmente diferentes propiedades, como punto isoeléctrico, peso molecular, grado de glicosilación, especificidad de sustrato y patrones enzimáticos de acción.

Se cree que las celulasas tipo-endoglucanasa hidrolizan uniones beta-1,4-glucosídicas internas en regiones de baja cristalinidad de la celulosa y las celulasas tipo-exo-celobiohidrolasas hidrolizan celobiosa a partir de terminaciones reductoras o no-reductoras de celulosa. Se deduce que la acción de los componentes de endoglucanasa puede facilitar en gran medida la acción de las exo-celobiohidrolasas mediante la creación de nuevos extremos de cadena que son reconocidos por los componentes exo-celobiohidrolasa. Además, las celulasas tipo-exo-celobiohidrolasas

se ha demostrado que catalizan la hidrólisis alquilo y/o arilo β-D-glucósidos como metil β-D-glucósido y p-nitrofenil glucósido así como glicósidos que contienen solamente residuos de carbohidratos, como celobiosa. Esto produce glucosa como el único producto para el microorganismo y reduce o elimina celobiosa que inhibe las celobiohidrolasas y endoglucanasas.

Por consiguiente, las celulasas tipo-β-glucosidadsas se considera que son una parte integral del sistema de celulasa porque conducen a la reacción total hacia glucosa. Se ha mostrado que una expresión aumentada de BG en T. reesei mejora la degradación de celulosa a glucosa. Adviértase en EP0562003. Además, las β-glucosidasas pueden catalizar la hidrólisis de un número de diferentes sustratos, y por lo tanto encuentran utilidad en una variedad de aplicaciones diferentes. Algunas β-glucosidasas se pueden añadir a las uvas durante la elaboración del vino para potenciar el aroma potencial del producto vinícola terminado. Otra aplicación aún puede ser el uso de β-glucosidasa en la fruta para potenciar el aroma de la misma. Alternativamente, la β-glucosidasa se puede usar directamente en aditivos alimentarios o procesamiento del vino para potenciar el sabor y el aroma.

Las celulasas también encuentran un número de usos en composiciones detergentes incluyendo capacidad de limpieza, como agente suavizante y para mejorar el tacto de los tejidos de algodón (Hemmpel, 1991; Tyndall, 1992; Kumar *et al.,* 1997). Mientras el mecanismo no es parte de la descripción, las propiedades suavizantes y de restauración del color de la celulasa han sido atribuidas a los componentes de endoglucanasa alcalina en composiciones de celulasa, como se ejemplifica por la Patente EE.UU. Nºs 5.648.263, 5.691.178 y 5.776.757, que describen que esas composiciones de detergente que contienen una composición de celulasa enriquecida en un componente endoglucanasa alcalina específica imparten restauración del color y mejor suavizado a prendas de vestir tratadas en comparación con composiciones de celulasa no enriquecidas en dicho componente. Además, el uso de dichos componentes endoglucanasa alcalina en composiciones de detergente se ha demostrado que complementan los requerimientos de pH de la composición de detergente (p.ej., exhibiendo máxima actividad a un pH alcalino de 7,5 hasta 10, como se describe en las Patentes EE.UU. Nºs. 5.648.263, 5.691.178 y 5.776.757).

También se ha demostrado que las composiciones de celulasa degradan tejidos que contienen algodón, dando como resultado una reducción de la pérdida de fuerza en el tejido (Patente EE.UU. Nº 4.822.516), contribuyendo a la reticencia a usar composiciones de celulasa en aplicaciones de detergentes comerciales. Se ha sugerido que las composiciones de celulasa que comprenden componentes endoglucanasa exhiben una pérdida de fuerza reducida para los tejidos que contienen algodón en comparación con composiciones que comprenden un sistema de celulasa completo.

También se ha mostrado que las celulasas son útiles en la degradación de biomasa celulasa a etanol (en donde la celulasa degrada la celulosa a glucosa y la levadura y otros microbios fermentan además la glucosa a etanol), en el tratamiento de la pulpa de madera (Pere *et al.,* 1996), para uso como aditivo para piensos (WO 91/04673) y en la molienda húmeda de cereales.

La mayoría de las CBHs y EGs tienen una estructura multidominio que consiste en un dominio central separado de un dominio de unión a celulosa (CBD) mediante un péptido de unión (Suurnakki *et al.,* 2000). El dominio central contiene el sitio activo mientras que el CBD interactúa con la celulosa mediante la unión de la enzima a él (van Tilbeurgh *et al.,* 1986; Tomme *et al.,* 1988). Los CBDs son particularmente importantes en la hidrólisis de la celulosa cristalina. Se ha mostrado que la habilidad de las celobiohidrolasas para degradar la celulosa cristalina disminuye claramente cuando el CBD está ausente (Linder and Teeri, 1997). Sin embargo, el papel exacto y el mecanismo de acción de los CBDs sigue siendo un tema de especulación. Se ha sugerido que el CBD potencia la actividad enzimática aumentando simplemente la concentración efectiva de la enzima en la superficie de la celulosa (Stahlberg *et al.,* 1991), y/o aflojando las cadenas de celulosa individuales de la superficie de celulosa (Tormo *et al.,* 1996). La mayoría de los estudios sobre los efectos de los dominios de celulasa sobre diferentes sustratos se han llevado a cabo con proteínas centrales de celobiohidrolasas, ya que sus proteínas centrales se pueden producir fácilmente mediante proteólisis limitada con papaína (Tomme *et al*., 1988). Se han descrito numerosas celulasas en la literatura científica, ejemplos de las cuales incluyen: desde *Trichoderma reesei*: Shoemaker, S. *et al*., Bio/Technology, 1:691-696, 1983, que describe *CBHI*; Teeri, *et al*., Gene, 51:43-52, 1987, que describe *CBHII*; Penttila, M. *et al*., Gene, 45:253-263, 1986, que describe EG/; Saloheimo, M. et al., Gene, 63:11-22, 1988, que describe 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 Biology, 13:219-228, 1994, que describe *EGV*; Barnett, C. C., *et al*., Bio/Technology, 9:562-567, 1991, que describe *BGL1*, y Takashima, S. *et al*., J. Biochem., 125:728-736, 1999, que describe BGL2. Se han descrito también celulasas de especies distintas a Trichoderma p.ej., Ooi, *et al.,* 1990, que describe la secuencia de cDNA que codifica la endoglucanasa F1-CMC producida por *Aspergillus aculeatus*; Kawaguchi T *et al.,* 1996, que describe el clonaje y la secuenciación del cDNA que codifica la beta-glucosidasa 1 de *Aspergillus aculeatus*; Sakamoto *et al.,* 1995, que describe la secuencia de cDNA que codifica la endoglucanasa CMCasa-1 de *Aspergillus kawachii* IFO 4308; Saarilahti *et al.,* 1990 que describe una endoglucanasa de *Erwinia carotovara*; Spiliaert R, *et al.,* 1994, que describe el clonaje y la secuenciación de bgla, que codifica una beta-glucanasa termoestable de *Rhodothermus marinu*; y Halldorsdottir S *et al.,* 1998, que describe el clonaje, secuenciación y sobre-expresión de un gen de *Rhodothermus marinus* que codifica una celulasa termoestable de la familia de glicosil hidrolasa 12. Sin embargo, sigue habiendo una necesidad de identificación y caracterización de nuevas celulasas, con propiedades mejoradas, como actuación mejorada bajo

condiciones de estrés térmico o en presencia de tensoactivos, actividad específica aumentada, patrón de escisión de sustrato alterado, y/o alto nivel de expresión *in vitro*.

El desarrollo de nuevas y mejoradas composiciones de celulasa que comprenden cantidades variables de celulasas tipo-CBH, tipo-EG y tipo-BG es de interés para su uso: (1) en composiciones de detergente que exhiben capacidad de limpieza mejorada, funcionan como un agente suavizante y/o mejoran el tacto de tejidos de algodón (*p.ej.*, "lavado a la piedra" o "pulido biológico"); (2) en composiciones para degradar la pulpa de madera u otra biomasa en azúcares (*p.ej.*, para la producción de bio-etanol); y/o (3) en composiciones de pienso.

IV. Métodos de identificación de nuevas secuencias

Los marcos abiertos de lectura (ORFs) se analizan siguiendo secuenciación total o parcial del genoma de *T. reesei* o de clones de bibliotecas de cDNA derivadas de mRNA de T. reesei y se analizan adicionalmente usando software de análisis de secuencia, y determinando homología con secuencias conocidas en las bases de datos (públicas/privadas).

V. Ácidos nucleicos *egl7* y polipéptidos EGVII.

A. Ácidos nucleicos *egl7*

La invención de moléculas de ácido nucleico incluye la secuencia codificadora nativa de *egl7*. En una realización, la secuencia es la secuencia de cDNA de *egl7* presentada en esta memoria como SEQ. ID. NO:1 o SEQ. ID. NO:4; y homólogos de las mismas en otras especies, variantes alélicas y de corte y empalme naturales, fragmentos de ácidos nucleicos, y derivados biológicamente activos (funcionales) de las mismas, como, variantes de secuencia de aminoácidos de la molécula nativa y secuencias que codifican proteínas de fusión. Las secuencias están referidas de manera colectiva en esta memoria como "secuencias de ácido nucleico que codifican EGVII".

Se llevó a cabo una búsqueda básica BLASTN (http://www.ncbi.nlm.nih.gov/BLAST) de la base de datos no redundante de secuencia de ácidos nucleicos el 12 de Septiembre de 2001, con la secuencia del gen *egl7* presentada en la Figura 1 (SEQ ID NO:1), indicó que ninguna secuencia producía alineamientos significativos (esto es, con un valor de E de menos de 10^{-5}).

Una secuencia de ácido nucleico *egl7* puede ser una secuencia de DNA o RNA, derivada de DNA, cDNA, mRNA genómico o puede ser sintetizada entera o en parte. El DNA debe ser de doble cadena o de cadena sencilla y si es de cadena sencilla debe ser la cadena codificadora o la cadena no-codificadora (antisentido, complementaria). La secuencia de ácido nucleico puede ser clonada, por ejemplo, aislando DNA genómico de una fuente apropiada y amplificando y clonando la secuencia de interés usando una reacción en cadena de la polimerasa (PCR). Alternativamente, la secuencia de ácido nucleico se puede sintetizar, ya sea completamente o en parte, especialmente donde es deseable proporcionar secuencias preferidas del huésped para una óptima expresión. Así, se puede sintetizar toda o una porción del gen estructural deseado (esa porción del gen que codifica un polipéptido o una proteína) usando codones preferidos por un huésped seleccionado.

Debido a la degeneración del código genético se pueden usar secuencias de ácido nucleico que no sean la forma nativa que codifica sustancialmente la misma o una secuencia de aminoácidos funcionalmente equivalente para clonar y/o expresar las secuencias de ácido nucleico que codifica EGVII. Así, para una secuencia de ácido nucleico que codifica EGVII dada se valora que, como resultado de la degeneración del código genético, se pueden producir un número de secuencias codificadoras que codifican una proteína que tenga la misma secuencia de aminoácidos. Por ejemplo, el triplete CGT codifica al aminoácido arginina. La arginina es codificada alternativamente por CGA, CGC, CGG, AGA y AGG. Por lo tanto se valora que dichas sustituciones en la región codificadora caigan dentro las variantes de secuencia de ácido nucleico descritas en esta memoria. Todas y cada una de estas variantes de secuencia se pueden utilizar de la misma forma como se describe en la presente memoria para la forma nativa de una secuencia de ácido nucleico que codifica EGVII.

Una secuencia de ácido nucleico que codifica EGVII "variante" puede codificar una secuencia de aminoácidos de EGVII "variante" que está alterada en uno o más aminoácidos de la secuencia polipeptídica nativa o puede estar truncada por eliminación de uno o más aminoácidos de cualquier extremo de la secuencia polipeptídica, ambas incluidos dentro del alcance de la descripción. De manera similar, el término "forma modificada de", relativo a EGVII, significa un derivado o forma variante de la secuencia de ácido nucleico que codifica la proteína EGVII nativa o la secuencia de aminoácidos de la EGVII nativa.

De forma similar, los polinucleótidos incluyen secuencias que codifican proteínas EGVII nativas y variantes de corte y empalme de las mismas, secuencias complementarias a la secuencia codificadora de la proteína nativa, y fragmentos nuevos de polinucleótidos que codifican EGVII. Una secuencia de ácido nucleico que codifica EGVII puede contener uno o más secuencias intrónicas si es una secuencia de DNA genómico.

En una realización general, una secuencia de nucleótidos que codifica EGVII tiene al menos un 70%, preferiblemente un 80%, 85%, 90%, 95%, 98%, o más identidad de secuencia con la secuencia codificadora *egl7* presentada en esta memoria como SEQ ID NO:1.

En otra realización, una secuencia de nucleótidos que codifica EGVII hibridará bajo condiciones de astringencia de moderadas a altas con una secuencia de nucleótidos que codifica una proteína EGVII. En una realización relacionada, una secuencia de nucleótidos que codifica EGVII hibridará bajo condiciones de astringencia de moderadas a altas con la secuencia de nucleótidos presentada como SEQ ID NO:1.

Se aprecia que algunas variantes de secuencia de ácido nucleico que codifican EGVII puede o puede que no hibriden selectivamente con la secuencia parental. A modo de ejemplo, en situaciones donde la secuencia codificadora se ha optimizado en base a la degeneración del código genético, se puede producir una secuencia codificadora variante que codifica una proteína EGVII, pero que no hibrida con la secuencia de ácido nucleico que codifica EGVII nativa bajo condiciones de astringencia de moderadas a altas. Esto ocurriría, por ejemplo, cuando la variante de secuencia incluye un codón diferente para cada uno de los aminoácidos codificados por el nucleótido parental.

Como se entenderá adicionalmente por los expertos en la técnica, en algunos casos podría ser ventajoso producir secuencias de nucleótidos que posean codones no naturales, p.ej., inosina u otro nucleótido análogo no-natural. Los codones preferidos por un huésped eucariota particular se pueden seleccionar, por ejemplo, para aumentar la velocidad de expresión de la proteína EGVII o para producir transcritos de RNA recombinante que tengan las propiedades deseables, como vida media más larga que los transcritos producidos a partir de la secuencia natural. Por tanto, se puede diseñar una secuencia de nucleótidos que codifica EGVII nativa con el fin de alterar la secuencia codificadora por una variedad de razones, incluyendo pero no limitadas a, alteraciones que modifican el clonaje, procesamiento y/o expresión de la proteína EGVII por una célula.

Particularmente preferidas son las sustituciones, adiciones y supresiones en el ácido nucleico que son silenciosas de tal manera que no alteran las propiedades o actividades del polinucleótido nativo o polipéptido.

Las variaciones se pueden hacer usando 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 de casete (Wells *et al.,* 1985), mutagénesis de selección por restricción (Wells *et al.,* 1986) u otras técnicas conocidas se pueden llevar a cabo sobre el DNA clonado para producir una variante de DNA que codifica el polipéptido EGVII.

Sin embargo, en algunos casos puede ser ventajoso expresar variantes de *egl7* que carecen de las proteínas o las actividades del polinucleótido nativo de *egl7* o el polipéptido EGVII. En tales casos, se pueden generar formas mutantes o modificadas de la secuencia de ácido nucleico que codifica EGVII nativa.

B. Polipéptidos EGVII

En una realización preferida, se describe un polipéptido EGVII, que tiene una secuencia polipeptídica nativa de EGVII madura o entera que comprende la secuencia presentada en la Figura 2 (SEQ ID NO:2). Un polipéptido EGVII puede ser el polipéptido de EGVII madura, parte de una proteína de fusión o un fragmento o variante de la secuencia polipeptídica de EGVII presentada en la Figura 2 (SEQ ID NO:2).

Ordinariamente, un polipéptido de EGVII tiene al menos un 80% de identidad con la secuencia de aminoácidos de EGVII en toda su longitud. Más preferibles son las secuencias de polipéptido de EGVII que comprenden una región que tiene al menos un 80, 85, 90, 95, 98% o más identidad de secuencia con la secuencia del polipéptido de EGVII de la Figura 2 (SEQ ID NO:2), usando un programa de alineamiento, como se detalla en esta memoria.

Normalmente, una “forma modificada de” una proteína nativa EGVII o una “variante” de la proteína EGVII tiene una secuencia derivada que contiene al menos una sustitución, adición, supresión o inserción de un aminoácido, respectivamente.

Es bien conocido en la técnica que se pueden hacer ciertas sustituciones de aminoácidos en las secuencias de proteína sin afectar a la función de la proteína. Generalmente, las sustituciones conservativas de aminoácidos o sustituciones de aminoácidos similares son toleradas sin afectar a la función de la proteína. Aminoácidos similares pueden ser aquellos que son similares en tamaño y/o propiedades de carga, por ejemplo, aspartato y glutamato, e isoleucina y valina, son parejas ambas de aminoácidos similares. Se han evaluado en la técnica de muchas maneras la similitud entre parejas de aminoácidos. Por ejemplo, Dayhoff *et al.,* (1978), proporciona tablas de frecuencia para sustituciones de aminoácidos que pueden ser empleadas como una medida de similitud de aminoácidos. Las tablas de frecuencia de Dayhoff *et al.* se basan en comparaciones de secuencias de aminoácidos para proteínas que tienen la misma función de una variedad de fuentes evolutivamente diferentes.

Los fragmentos y variantes de la secuencia polipetídica de EGVII de la Figura 2 (SEQ ID NO:2) se considera que son una parte de ella. Un fragmento es un polipétido variante que tiene una secuencia de aminoácidos que es enteramente la misma como parte pero no toda la secuencia de aminoácidos de los polipéptidos previamente descritos. Los fragmentos pueden ser “independientes” o estar comprendidos dentro de un polipéptido grande de la que el fragmento forma una parte o una región, más preferiblemente como una única región continua. Los fragmentos preferidos son fragmentos biológicamente activos que son aquellos fragmentos que median actividades de los polipéptidos, incluyendo aquellos con actividad similar o actividad mejorada o con una actividad disminuida.

También están incluidos aquellos fragmentos que son antigénicos o inmunogénicos en un animal, particularmente un humano. En este aspecto, la descripción incluye (i) fragmentos de EGVII, preferiblemente al menos aproximadamente de 20-100 aminoácidos de longitud, más preferiblemente aproximadamente de 100-200 aminoácidos de longitud, y (ii) una composición farmacéutica que comprende EGVII. En varias realizaciones, el fragmento corresponde al dominio N-terminal de EGVII o al dominio C-terminal de EGVII.

Los polipéptidos EGVII también incluyen polipéptidos que varían del EGVII, secuencia polipeptídica de la Figura 2 (SEQ ID NO:2). Estas variantes exhiben normalmente la misma actividad biológica cualitativa que el análogo que ocurre naturalmente, aunque se pueden seleccionar variantes que tienen características modificadas como se describe más adelante.

Una “sustitución” resulta del reemplazo de uno o más nucleótidos o aminoácidos por diferentes nucleótidos o aminoácidos, respectivamente.

Una “inserción” o “adición” es aquel cambio en una secuencia de nucleótidos o un aminoácidos que ha dado lugar a la adición de uno o más residuos de nucleótidos o de aminoácidos, respectivamente, en comparación con la secuencia que ocurre naturalmente.

Una “supresión” se define como un cambio en cualquier secuencia de nucleótidos o de aminoácidos en la que están ausentes uno o más residuos de nucleótidos o aminoácidos, respectivamente.

Las sustituciones de aminoácidos son normalmente de residuos únicos; las inserciones por lo general estarán en el orden de aproximadamente 1 a 20 aminoácidos, aunque pueden tolerarse inserciones considerablemente más grandes. Las supresiones varían de aproximadamente 1 a aproximadamente 20 residuos, aunque en algunos casos las supresiones pueden ser mucho más grandes.

Las sustituciones, supresiones, inserciones o cualquier combinación de las mismas se pueden usar para llegar a un derivado final. Generalmente estos cambios se hacen en unos pocos aminoácidos para minimizar la alteración de la molécula. Sin embargo, se pueden tolerar cambios mayores en ciertas circunstancias.

Las sustituciones de aminoácidos pueden ser el resultado de sustituir un aminoácido con otro aminoácido que tenga propiedades estructurales y/o químicas similares, como la sustitución de una isoleucina con una valina, esto es, sustituciones conservativas de aminoácidos. Las inserciones o supresiones pueden estar opcionalmente en el intervalo de 1 a 5 aminoácidos.

Las sustituciones se hacen generalmente de acuerdo con “sustituciones conservativas” conocidas. Una “sustitución conservativa” se refiere a la sustitución de un aminoácido de una clase por un aminoácido de la misma clase, donde una clase se define por propiedades fisicoquímicas comunes de la cadena lateral del aminoácido y las altas frecuencias de sustitución en proteínas homólogas encontradas en la naturaleza (tal como se determina, p.ej., por una matriz estándar de intercambio de frecuencia Dayhoff o matriz BLOSUM). (Adviértase de forma general, Doolittle, R.F., 1986.)

Una “sustitución no conservativa” se refiere a la sustitución de un aminoácido de una clase con un aminoácido de otra clase.

Las variantes polipeptídicas de EGVII exhiben normalmente la misma actividad bilógica cualitativa que el análogo que ocurre naturalmente, aunque las variantes se seleccionan también para modificar las características del polipéptido EGVII, según sea necesario. Por ejemplo, se pueden alterar o eliminar los sitios de glicosilación, y más concretamente uno o más sitios de glicosilación ligados a N o ligados a O. Aquellos expertos en la técnica valoran que los cambios de aminoácidos puedan alterar los procesos post-traduccionales del polipéptido EGVII, como cambiar el número o la posición de los sitios de glicosilación o alterar las características de anclaje a la membrana o las características de secreción u otras características de localización celular.

También se incluyen dentro de la definición de polipéptidos EGVII otros polipéptidos EGVII relacionados. Así, se pueden usar sondas o secuencias cebadoras degeneradas de reacción en cadena de la polimerasa (PCR) para encontrar otros polipéptidos relacionados. Se pueden diseñar sondas o secuencias cebadoras útiles para: toda o parte de la secuencia polipeptídica de EGVII, o secuencias fuera de la región codificadora. Como se conoce generalmente en la técnica, los cebadores preferidos son desde aproximadamente 15 a aproximadamente 35 nucleótidos de longitud, prefiriéndose con desde aproximadamente 20 a aproximadamente 30, y pueden contener inosina según sea necesario. Las condiciones para la reacción de PCR se conocen generalmente en la técnica.

También se incluyen dentro del alcance de esta descripción modificaciones covalentes de polipéptidos EGVII. Por ejemplo, la descripción proporciona polipéptidos EGVII que son una proteína madura y pueden comprender aminoácidos adicionales amino- o carboxilo-terminales, o aminoácidos dentro del polipéptido maduro (por ejemplo, cuando la forma madura de la proteína tiene más de una cadena polpeptídica). Dichas secuencias pueden, por ejemplo, jugar un papel en el procesamiento de la proteína desde un precursor a una forma madura, permiten el transporte de proteína, acortan o alargan la vida media de la proteína, o facilitan la manipulación de la proteína en los ensayos de producción.

También se contemplan modificaciones dirigidas a la alteración de un sitio activo, alteración de un pH óptimo, temperatura óptima, y/o afinidad de sustrato de la enzima EGVII.

La Figura 2 muestra la secuencia de aminoácidos predicha (SEQ ID NO:2) de un polipéptido EGVII modelo basado en la secuencia de nucleótidos proporcionada en la Figura 1.

El peso molecular predicho del polipéptido EGVII codificado es de 26,8 kDa. Un péptido señal predicho de 19 aminoácidos precede al extremo amino terminal del EGVII maduro como se establece en la figura sugiriendo que el polipéptido EGVII se secreta (Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., Protein Engineering, 10: 1-, 1997).

Una búsqueda básica BLASTP (http://www.ncbi.nlm.nih.gov/BLAST) de la base de datos de proteínas no redundante, realizada el 12 de Septiembre de 2001 con la secuencia de aminoácidos de EGVII indicó un 46% de identidad con Número de Acceso a GenBank Y11113 (endoglucanasa IV de *Hypocrea jecorina* o *Trichoderma reesei*), un 41% de identidad de secuencia con Número de Acceso a GenBank M86356 (proteína específica de crecimiento de celulosa de *Agaricus bisporus*) y un 37% de identidad de secuencia con Número de Acceso a GenBank AB055432 (endoglucanasa B de *Aspergillus kawachii*). Estas similitudes de secuencia indican que EGVII es un miembro de la familia 61 de glicosil hidrolasas (Henrissat, B. and Bairoch, A. (1993) Biochem. J. 293:781-788).

C. Anticuerpos anti- EGVII

La descripción proporciona además anticuerpos anti-EGVII. Los anticuerpos pueden ser anticuerpos policlonales, monoclonales, humanizados, biespecíficos o heteroconjugados.

Los métodos de preparación de anticuerpos policlonales son conocidos por el técnico experto. El agente inmunizante puede ser un polipéptido EGVII o una proteína de fusión del mismo. Puede ser útil conjugar el antígeno con una proteína conocida para que sea inmunogénica en el mamífero que se inmuniza. El protocolo de inmunización puede ser determinado por un experto en la técnica basándose en protocolos estándar o experimentación de rutina.

Alternativamente, los anticuerpos anti-EGVII pueden ser monoclonales. Se pueden producir anticuerpos monoclonales por células inmunizadas en un animal o usando métodos de DNA recombinante (Adviértase, p.ej., Kohler *et al.*, 1975; Patente EE.UU. Nº 4.816.567).

Un anticuerpo anti-EGVII puede comprender además 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 inmunoglobulinas o fragmentos de los mismos (como Fv, Fab, Fab’, $F(ab')_2$ u otras secuencias parciales de unión a antígeno de anticuerpos) que contienen alguna porción de la secuencia derivada de un anticuerpo no-humano. Se conocen bien en la técnica los métodos para humanizar anticuerpos no-humanos, como se detalla en Jones *et al.*, 1986; Riechmann *et al.*, 1988; and Verhoeyen *et al.*, 1988. Son también conocidos en la técnica los métodos para producir anticuerpos humanos. Adviértase, p.ej., Jakobovits, A. *et al.*, 1995 y Jakobovits, A., 1995.

VI. Expresión de EGVII recombinante

Los métodos descritos en esta memoria dependen del uso de células para expresar EGVII, sin ningún método particular de expresión de EGVII requerido.

La descripción proporciona células huésped que han sido transducidas, transformadas o transfectadas con un vector de expresión que comprende una secuencia de ácido nucleico que codifica EGVII. Las condiciones de cultivo, como temperatura, pH y similares, son las usadas previamente por la célula parental huésped antes de la transducción, transformación o transfección y serán evidentes para los expertos en la técnica.

En una estrategia, una célula de hongo filamentoso 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 potenciadores (p.ej., una serie) que funciona en la célula huésped, operativamente ligado a un segmento de DNA que codifica EGVII de tal manera que EGVII se exprese en la línea celular.

A. Constructos de ácido nucleico/Vectores de expresión

Se pueden incorporar fragmentos de polinucleótidos naturales o sintéticos que codifican EGVII (“secuencias de ácido nucleico que codifican EGVII”) en constructos o vectores de ácido nucleico, capaces de introducirse dentro, y replicación en, una célula de hongo filamentoso o de levadura. Los vectores y métodos descritos en esta memoria son adecuados para usar en células huésped para la expresión de EGVII. Se puede usar cualquier vector mientras sea replicable y viable en las células en las que es introducido. Son conocidos por aquellos expertos en la técnica y están disponibles comercialmente un gran número de vectores y promotores adecuados. Se describen también vectores de clonaje y de expresión en Sambrook *et al.*, 1989, Ausubel FM *et al.*, 1989, y Strathern *et al.*, 1981. Se describen vectores de expresión apropiados para hongos 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, pp. 396-428. Se puede insertar la secuencia de DNA apropiada en un plásmido o vector (denominados colectivamente en esta memoria como "vectores") por una variedad de procedimientos. En general, la secuencia se inserta en un(os) sitio(s) de restricción de endonucleasa apropiado(s) por procedimientos estándar. Dichos procedimientos y procedimientos de sub clonaje relacionados se considera que están dentro del alcance del conocimiento de los expertos en la técnica.

Los hongos filamentosos recombinantes que comprenden la secuencia codificadora de EGVII se pueden generar mediante la introducción de un constructo de ácido nucleico heterólogo que comprende la secuencia codificadora de EGVII dentro de las células de una cepa seleccionada de los hongos filamentosos.

Una vez que se obtiene la forma deseada de una secuencia de ácido nucleico *egl7*, homólogo, variante o fragmento de la misma, se puede modificar de una variedad de formas. Donde la secuencia implica regiones flanqueantes no codificadoras, las regiones flanqueantes se pueden someter a resección, mutagénesis, etc. De esta forma, se pueden realizar transiciones, transversiones, supresiones e inserciones en la secuencia que ocurre naturalmente.

Se puede insertar una secuencia codificadora de *egl7* seleccionada en un vector adecuado de acuerdo a técnicas recombinantes y usar para transformar hongos filamentosos capaces de expresar EGVII. Debido a degeneración inherente del código genético, se pueden usar otras secuencias de ácido nucleico que codifican sustancialmente la misma secuencia de aminoácidos o una funcionalidad equivalente para clonar y expresar EGVII. Por lo tanto, se valora que dichas sustituciones en la región codificadora caigan dentro de las variantes de secuencia descritas en esta memoria. Todos y cada una de estas variantes de secuencia se pueden utilizar de la misma forma a como se describe en esta memoria para una secuencia de ácido nucleico que codifica EGVII parental.

La presente descripción también incluye constructos de ácido nucleico recombinante que comprenden una o más de las secuencias de ácido nucleico que codifican EGVII como se describió anteriormente. Los constructos comprenden un vector, como un plásmido o vector viral, en el que se ha insertado una secuencia, en una orientación hacia delante o hacia atrás.

Los constructos de ácido nucleico heterólogos pueden incluir la secuencia codificadora para *egl7*, o una variante, fragmento o variante de corte y empalme de la misma: (i) aisladamente; (ii) en combinación con secuencias codificadoras adicionales; como proteína de fusión o secuencias que codifican péptido señal, donde la secuencia codificadora *egl7* es la secuencia codificadora dominante; (iii) en combinación con secuencias no codificadoras, como intrones y elementos de control, como elementos promotores y terminadores o regiones no traducidas 5' y/o 3', eficaces para la expresión de la secuencia codificadora en un huésped adecuado; (iv) en un vector o ambiente hospedador en el que la secuencia codificadora de *egl7* es un gen heterólogo.

En un aspecto, se emplea un constructo de ácido nucleico heterólogo para transferir una secuencia de ácido nucleico codificadora de EGVII dentro de una célula *in vitro*, con líneas establecidas preferidas de hongos filamentosos y levaduras. Para producción de alto rendimiento a largo plazo se prefiere una expresión estable. Se deduce que se puede usar cualquier método eficaz de generar transformantes estables.

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 promotoras y de terminación. El vector puede comprender secuencias reguladoras, incluyendo, por ejemplo, secuencias no codificadoras, como intrones y elementos de control, esto es, elementos promotores y terminadores o regiones no traducidas 5' y/o 3', eficaces para la expresión de la secuencia codificadora en células huésped (y/o en un vector o ambiente de célula huésped en el que no se expresa normalmente una secuencia codificadora de un antígeno de proteína soluble modificada), operativamente ligadas a la secuencia codificadora. Se conocen un gran número de vectores y promotores adecuados por aquellos expertos en la técnica, muchos de los cuales están comercialmente disponibles y/o están descritos en Sambrook, *et al.*, (*supra*).

Promotores modelo incluyen tanto promotores constitutivos como promotores inducibles, ejemplos de los cuales incluyen un promotor CMV, un promotor temprano SV40, un promotor RSV, un promotor EF-1α, un promotor que contiene el elemento de respuesta tet (TRE) en el sistema tet-activado o tet-desactivado como se describe (ClonTech y BASF), el promotor de beta actina y el promotor de metalotionina que puede regularse al alza añadiendo sales de ciertos metales. Una secuencia promotora es una secuencia de DNA que es reconocida por el hongo filamentoso en particular con el propósito de expresión. Está ligado operativamente a la secuencia de DNA que codifica un polipéptido EGVII. Dicha unión comprende el posicionamiento del promotor con respecto al codón de iniciación de la secuencia de DNA que codifica el polipéptido en los vectores de expresión descritos. La secuencia promotora contiene secuencias control de la transcripción y de la traducción que median en la expresión del polipéptido EGVII. Ejemplos incluyen promotores de 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 proteinasa aspártica de *A. niger* o *Rhizomucor miehei*; los genes que codifican *cbh1*, *cbh2*, *egl1*, *egl2* u otras celulasas de *T. reesei*.

La elección del marcador seleccionable adecuado dependerá de la célula huésped, y se conocen bien en la técnica marcadores apropiados para los diferentes huéspedes. Los genes marcadores seleccionables normales 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*. Marcadores seleccionables modelo adicionales incluyen, pero no se limitan a, trpc, trp1, oliC31, niaD o leu2, que están incluidos en constructos de ácido nucleico heterólogo usados para transformar una cepa mutante como trp-, pyr-, leu- y similares.

Tales marcadores seleccionables confieren a los transformantes la habilidad de utilizar un metabolito que usualmente no se metaboliza por los hongos filamentosos. Por ejemplo, el gen amdS de *T. reesei* que codifica la enzima acetamidasa que permite crecer a las células transformantes en acetamida como fuente de nitrógeno. El marcador seleccionable (p.ej. pyrG) puede restaurar la capacidad de crecer de una cepa mutante auxótrofa 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 inhibidor o antibiótico.

La secuencia codificadora del marcador seleccionable se clona en cualquier plásmido adecuado usando métodos que se emplean generalmente en la técnica. Plásmidos modelo incluyen pUC18, pBR322, y pUG100.

La práctica de la presente descripción empleará, a no ser que se indique lo contrario, técnicas convencionales de biología molecular, microbiología, DNA recombinante e inmunología, que están dentro de la experiencia de la técnica. Dichas técnicas se explican completamente en la literatura. Adviértase, 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 EGVII

(i) Hongos filamentosos

De esta forma, la presente descripción proporciona hongos filamentosos que comprenden células que han sido modificadas, seleccionadas y cultivadas de una forma efectiva para dar como resultado una producción o expresión de EGVII potenciada en relación con la correspondiente de los hongos parentales no transformados.

Ejemplos de especies de hongos filamentosos parentales que se pueden tratar y/o modificar para una expresión de EGVII potenciada incluyen, pero no se limitan a *Trichoderma*, p.ej., *Trichoderma reesei*, *Trichoderma longibrachiatum, Trichoderma viride, Trichoderma koningii; Penicillium sp*., *Humicola sp*., incluyendo *Humicola insolens; Aspergillus sp., Chrystosporium sp., Fusarium sp., Hypocrea sp*., y *Emericella sp*.

Las células que expresan EGVII se cultivan bajo condiciones empleadas normalmente para cultivar la línea fúngica parental. Generalmente, las células se cultivan en un medio estándar que contiene sales y nutrientes fisiológicos, como se describe en Pourquie, J. *et al*., Biochemistry and Genetics of Cellulose Degradation, eds. Aubert, J. P. *et al*., Academic Press, pp. 71-86, 1988 y Limen, M. *et al*., Appl. Environ. Microbiol. 63:1298-1306, 1997. Las condiciones de cultivo son también estándar. p.ej., los cultivos se incuban a 28ºC en un agitador de cultivos o fermentadores hasta que se consiguen niveles deseados de expresión de EGVII.

Se pueden encontrar condiciones de cultivo preferidas para un hongo filamentoso dado en la literatura científica y/o del origen de los hongos como American Type Culture Collection (ATCC; http://www.atcc.org/). Después de establecerse el crecimiento fúngico, las células se exponen a condiciones efectivas para causar o permitir la sobre expresión de EGVII.

En casos donde una secuencia que codifica de EGVII está bajo el control de un promotor inducible, se añade el agente inductor, p.ej., un azúcar, sal metálica o antibióticos, al medio a una concentración efectiva para inducir altos niveles de expresión de EGVII.

(ii) Levaduras

La presente descripción también contempla el uso de levaduras como una célula huésped para la producción de EGVII. Se han expresado otros diversos genes que codifican enzimas hidrolíticas en diversas cepas del hongo *S. cerevisiae*. Estos incluyen secuencias que codifican dos endoglucanasas (Penttila *et al*., 1987), dos celobiohidrolasas (Penttila *et al*., 1988) y una beta-glucosidasa de *Trichoderma reesei* (Cummings y Fowler, 1996), una xilanasa de *Aureobasidlium pullulans* (Li y Ljungdahl, 1996), una alfa-amilasa de trigo (Rothstein *et al*., 1987), etc. Además, un casete de genes de celulasa que codifica la endo-[beta]-1,4-glucanasa (END1) de *Butyrivibrio fibrisolvens*, celobiohidrolasa (CBH1) de *Phanerochaete chrysosporium*, celodextrinasa (CEL1) de *Ruminococcus flavefaciens* y celobiasa (Bgl1) de *Endomyces fibrilizer* se expresó con éxito en una cepa de laboratorio de *S. cerevisiae* (Van Resburg *et al*., 1998).

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

La descripción proporciona además células y composiciones celulares que se han modificado genéticamente para comprender una secuencia de ácido nucleico que codifica EGVII proporcionada exógenamente. Una célula parental o línea celular se puede modificar genéticamente (esto es, transducida, transformada o transfectada) con un vector de clonaje o un vector de expresión. El vector puede ser, por ejemplo, en forma de un plásmido, una partícula viral, un fago, etc. como se describe anteriormente.

Se pueden emplear diversos métodos para distribuir un vector de expresión dentro de células *in vitro*. Después de construirse un vector adecuado, se usa para transformar cepas de hongos o levaduras. Se conocen métodos generales de introducción de ácidos nucleicos dentro de células para la expresión de una secuencia de ácido nucleico heteróloga por el técnico ordinario experto. Dichos métodos incluyen, pero no se limitan a, electroporación, microinyección nuclear o microinyección directa en células únicas; fusión de protoplastos bacterianos con células intactas; uso de policationes, p.ej., polibreno o poliomitina; fusión de membrana con liposomas, lipofectamina o transfección mediada por lipofección, bombardeo a alta velocidad con microproyectiles recubiertos con DNA; incubación con precipitado de DNA-fosfato cálcico; transfección mediada por DEAE-dextrano; infección con ácidos nucleicos virales modificados; y similares.

Métodos preferidos para introducir un constructo de ácido nucleico heterólogo (vector de expresión) dentro de hongos filamentosos (*p.ej.*, *T. reesei*) incluyen, pero no se limitan al uso de biobalística de partícula o de gen, permeabilización de las paredes de hongos filamentosos antes del proceso de transformación (p.ej., mediante el uso de altas concentraciones de álcali, *p.ej.*, de 0,05 M a 0,4 M de $CaCl_2$ o acetato de litio), fusión de protoplastos o transformación mediada por agrobacteria. Se describe un método modelo para la transformación de hongos filamentosos mediante el tratamiento de protoplastos o esferoplastos con polietilenglicol y $CaCl_2$ en Campbell, E.I. *et al.*, Curr. Genet. 16:53-56, 1989 y Penttila, M. *et al.*, Gene, 63:11-22, 1988.

Además, se pueden transcribir *in vitro* constructos de ácido nucleico heterólogo que comprenden una secuencia de ácido nucleico que codifica EGVII, y el RNA resultante introducido en una célula huésped mediante métodos conocidos, p.ej., mediante inyección.

Después de la introducción de un constructo de ácido nucleico heterólogo que comprende la secuencia codificadora de *egl7*, las células modificadas genéticamente se pueden cultivar en medios nutrientes convencionales modificados según sea apropiado para la activación de los promotores, selección de transformantes o amplificación de la expresión de una secuencia de ácido nucleico que codifica EGVII. Las condiciones de cultivo, como temperatura, pH y similares, son aquellas usadas previamente para la célula huésped seleccionada para la expresión, y serán evidentes para aquellos expertos en la técnica.

La progenie de células en las que han sido introducidos los constructos de ácido nucleico heterólogo se considera generalmente que comprende la secuencia de ácido nucleico que codifica EGVII encontrada en el constructo de ácido nucleico heterólogo.

La descripción también incluye transformantes nuevos y útiles de hongos filamentosos como *Trichoderma reesei* para su uso en la producción de composiciones de celulasa fúngica. La descripción incluye transformantes de hongos filamentosos especialmente hongos que comprenden la secuencia codificadora de *egl7*, comprenden una forma modificada de la secuencia codificadora de *egl7* o supresión de la secuencia codificadora de *egl7*.

Generalmente se pueden distinguir transformantes estables de hongos filamentosos de transformantes inestables mediante su ritmo más rápido de crecimiento y la formación de colonias circulares con una superficie lisa en lugar de un contorno irregular en medio de cultivo sólido. Adicionalmente, en algunos casos, se puede hacer además una prueba de estabilidad mediante el crecimiento de los transformantes en medio sólido no selectivo, recogiendo las esporas de este medio de cultivo y determinando el porcentaje de estas esporas que posteriormente germinarán y crecerán en medio selectivo.

VII. Análisis de las secuencias de ácido nucleico que codifican EGVII y/o de la expresión de proteína.

Con el fin de evaluar la expresión de EGVII por una línea celular que ha sido transformada con un constructo de ácido nucleico que codifica EGVII, se pueden llevar a cabo ensayos a nivel de la proteína, a nivel del RNA o mediante el uso de bioensayos funcionales en concreto de la actividad y/o producción de endoglucanasa.

En una aplicación modelo de las secuencias de ácido nucleico y de proteína de *egl7* descritas en esta memoria, se diseñó una cepa de hongos filamentosos genéticamente modificada, *p.ej., Trichoderma reesei*, para producir una cantidad aumentada de EGVII. Dichos hongos filamentosos modificados genéticamente serían útiles para producir un producto de celulasa con un mayor aumento de la capacidad celulolítica. En una estrategia, esto se logra mediante la introducción de la secuencia codificadora de *egl7* dentro de un huésped adecuado, *p.ej.*, un hongo filamentoso como *Trichoderma reesei.*

Por consiguiente, la descripción incluye métodos para expresar EGVII en un hongo filamentoso u otro huésped adecuado mediante la introducción de un vector de expresión que contiene la secuencia de DNA que codifica EGVII dentro de las células del hongo filamentoso u otra célula adecuada.

En otro aspecto, la descripción incluye métodos para modificar la expresión de EGVII en un hongo filamentoso u otro huésped adecuado. Dicha modificación incluye una disminución o eliminación de expresión, o de expresión de una forma alterada de EGVII. Una forma alterada de EGVII puede tener una secuencia de aminoácidos alterada o una secuencia de ácido nucleico alterada.

En general, los ensayos empleados para analizar la expresión de EGVII incluyen, transferencia de Northern, transferencia de punto (análisis de DNA o RNA), RT-PCR (reacción en cadena de la polimerasa con transcriptasa reversa), o hibridación *in situ*, usando una sonda marcada apropiada (basada en la secuencia codificadora de ácido nucleico) y transferencia convencional de Southern y autoradiografía.

Además, la producción y/o expresión de EGVII se puede medir directamente en una muestra, por ejemplo, mediante ensayos de actividad de endoglucanasa, expresión y/o producción. Dichos ejemplos se describen, por ejemplo, en Shoemaker, S.P. y Brown, R.D.Jr. (Biochim. Biophys. Acta. 1978, 523:133-146; Schulein (1988) y Patente EE.UU. Nºs 5.246.853 y 5.475.101. La capacidad de EGVII de hidrolizar sustratos aislados solubles e insolubles se puede medir usando ensayos descritos en Suumakki *et al.*, (2000) y Ortega *et al.*, (2001). Sustratos útiles para ensayar actividades celobiohidrolasa, endoglucanasa o β-glucosidasa incluyen celulosa cristalina, papel de filtro, celulosa empapada en ácido fosfórico, hidroxietilcelulosa, carboximetilcelulosa, celooligosacáridos, metilobeliferil lactósido, metilobeliferil celobiósido; ortonitrofenil lactósido, paranitrofenil lactósido, ortonitrofenil celobiósido, paranitrofenil celobiósido, ortonitrofenil glucósido, paranitrofenil glucósido, metilobeliferil glicósido.

Además, la expresión de proteína se puede evaluar mediante métodos inmunlógicos, como tinción inmunohistoquímica de las células, secciones de tejidos o inmunoensayos de medio de cultivo de tejido, p.ej., mediante transferencia de Western o ELISA. Dichos inmunoensayos se pueden usar para evaluar la expresión de EGVII cualitativa y cuantitativamente. Los detalles de dichos métodos son conocidos por aquellos expertos en la técnica y muchos reactivos para la práctica de tales métodos están comercialmente disponibles.

Se puede usar una forma purificada de EGVII para producir o bien anticuerpos monoclonales o policlonales específicos para la proteína expresada para usar en distintos inmunoensayos. (Adviértase, *p.ej.*, Hu *et al.*, 1991). Ejemplos modelo incluyen ELISA, inmunoensayos competitivos, radioinmunoensayos, transferencia de Western, ensayos de inmunofluorescencia indirecta y similares. En general, se pueden usar anticuerpos disponibles comercialmente y/o kits para los inmunoensayos cuantitativos del nivel de expresión de las proteínas endoglucanasa.

VIII. Aislamiento y purificación de la proteína EGVII recombinante

En general, una proteína EGVII producida en el cultivo celular se secreta al medio y se puede purificar o aislar, p.ej., eliminando componentes no deseados del medio de cultivo celular. Aunque, en algunos casos, se puede producir una proteína EGVII en una forma celular que necesita recuperación a partir del lisado celular. En dichos casos las proteínas EGVII se purifican de las células en las cuales se produce usando técnicas rutinariamente empleadas por aquellos expertos en la técnica. Ejemplos incluyen, pero no se limitan a, cromatografía de afinidad (Tilbeurg *et al.*, 1984), métodos cromatográficos de intercambio iónico (Goyal *et al.*, 1991; Fliess *et al.*, 1983; Bhikhabhai *et al.*, 1984; Ellouz *et al.*, 1987), incluyendo el uso de materiales de intercambio iónico con alto poder de resolución (Medve *et al.*, 1998), cromatografía de interacción hidrofóbica (Tomaz y Queiroz, 1999), y sistema de reparto en dos fases (Brumbauer, *et al.*, 1999).

Normalmente, las proteínas EGVII se fraccionan para separar proteínas que tienes propiedades seleccionadas, como afinidad de unión a agentes de unión particulares, p.ej., anticuerpos o receptores; que tienen un intervalo seleccionado de peso molecular, o intervalo de puntos isoeléctricos.

Una vez que se logra la expresión de una proteína EGVII dada, la proteína producida de esta manera se purifica de células o de cultivo celular. Procedimientos modelo adecuados para tal purificación incluyen los siguientes: columna de cromatografía de afinidad de anticuerpo, cromatografía de intercambio iónico; precipitación en etanol; HPLC de fase reversa; cromatografía en resina de sílice o de intercambio catiónico como DEAE; cromatoenfoque; SDS-PAGE; precipitación en sulfato amónico; y filtración en gel usando, p.ej., Sephadex G-75. Se pueden emplear diversos métodos de purificación de proteína y dichos métodos se conocen en la técnica y se describen p.ej. en Deutscher, 1990; Scopes, 1982. La(s) etapa(s) de purificación seleccionada(s) dependerá(n), p.ej., de la naturaleza del proceso de purificación usado y de la proteína particular producida.

IX. Utilidad de *egl7* y EGVII

Se puede apreciar que el nucleótido *egl7*, la proteína EGVII y las composiciones que comprenden actividad de proteína EGVII encuentran utilidad en una amplia variedad de aplicaciones, algunas de las cuales se describen a continuación.

Composiciones de celulasa nuevas y mejoradas que comprenden cantidades variables de celulasas tipo CBH, tipo EG y tipo BG encuentran utilidad en composiciones detergentes que presentan capacidad de limpieza mejorada, funcionan como un agente suavizante y/o mejoran el tacto de los tejidos de algodón (p.ej., “lavado a la piedra” o “pulido biológico”), en composiciones para degradar pulpa de madera en azúcares (p.ej., para la producción de bio-etanol), y/o en composiciones de pienso. El aislamiento y la caracterización de celulasa de cada tipo proporciona la capacidad de controlar los aspectos de dichas composiciones.

En una estrategia preferida, la celulasa encuentra utilidad en composiciones de detergente o en el tratamiento de tejidos para mejorar el tacto y la apariencia.

Ya que que se puede aumentar la tasa de hidrólisis de productos celulósicos usando un transformante que tiene al menos una copia adicional del gen *egl7* insertada en el genoma, se pueden degradar productos que contienen celulosa o heteroglicanos a una mayor velocidad y en mayor medida. Los productos hechos de celulosa como papel, algodón, pañales de celulosa y similares se pueden degradar de manera más eficiente en un vertedero. Así, el producto de fermentación asequible a partir de los transformantes o los transformantes solos se pueden usar en composiciones para ayudar a degradar por licuefacción una variedad de productos de celulosa que se suman en los vertederos repletos.

La sacarificación y fermentación por separado son unos procesos en donde la celulosa presente en la biomasa, p.ej., el rastrojo de maíz es convertido a glucosa y después las cepas de levaduras convierten la glucosa en etanol. La sacarificación y fermentación simultánea es un proceso en donde la celulosa presente en la biomasa, p.ej., el rastrojo de maíz es convertido a glucosa y, al mismo tiempo y en el mismo reactor, las cepas de levadura convierten la glucosa en etanol. Así, en otra estrategia preferida, la celulasa tipo glucosidasa encuentra utilidad en la degradación de biomasa a etanol. La producción de etanol a partir de fuentes fácilmente disponibles proporciona una fuente de combustible estable y renovable.

Las materias primas basadas en celulosa se componen de residuos agrícolas, pastos y bosques y otra biomasa de bajo valor como residuos urbanos (p.ej., papel reciclado, recortes de jardín, etc.). Se puede producir etanol por fermentación de cualquiera de estas materias primas celulósicas. Sin embargo, la celulosa debe ser primero convertida a azúcares antes de que pueda ser la conversión a etanol.

Se pueden usar con la endoglucanasa una gran variedad de materias primas y la seleccionada para su uso puede depender de la región donde se realiza la conversión. Por ejemplo, en el medio oeste de Estados Unidos pueden predominar los residuos agrícolas como paja de trigo, rastrojo de maíz y bagazo mientras en California puede predominar paja de arroz. Sin embargo, debe entenderse que se puede usar cualquier biomasa celulósica en cualquier región.

Una composición de celulasa que contiene una cantidad de endoglucanasa aumentada encuentra utilidad en la producción de etanol. El etanol de este proceso se puede usar además como un potenciador de octanos o directamente como un combustible en lugar de gasolina lo que es una ventaja porque el etanol como fuente de combustible es más respetuoso con el medio ambiente que los productos derivados del petróleo. Se sabe que el uso de etanol mejorará la calidad del aire y reducirá posiblemente los niveles de ozono y niebla tóxica locales. Además, la utilización de etanol en lugar de la gasolina puede ser de importancia estratégica en la amortiguación del impacto de los cambios repentinos en el suministro de energía no renovables y petroquímica.

El etanol se puede producir a través de procesos de sacarificación y fermentación a partir de biomasa celulósica como árboles, plantas herbáceas, residuos sólidos urbanos y residuos agrícolas y forestales. Sin embargo, la relación de ezimas de celulasa individuales dentro de una mezcla de celulasa que ocurre naturalmente por un microbio puede no ser la más eficaz para la conversión rápida de la celulosa de la biomasa en glucosa. Se sabe que las endoglucanasas actúan para producir nuevos extremos de cadena de celulosa que por si mismos son sustratos para la acción de celobiohidrolasas y con ello mejorar la eficiencia de la hidrólisis de todo el sistema de celulasa. Por lo tanto el uso de la actividad incrementada u optimizada de endoglucanasa mejoraría en gran medida la producción de etanol.

Así, la endoglucanasa encuentra uso en la hidrólisis de celulosa en sus azúcares componentes. En una realización, la endoglucanasa se añade a la biomasa antes de la adición de un organismo fermentativo. En una segunda realización, la endoglucanasa se añade a la biomasa al mismo tiempo que el organismo fermentativo. Opcionalmente, puede haber otros componentes celulasa presentes en cualquier realización.

En otra realización la materia prima celulósica se puede pretratar. El pretratamiento puede ser por elevación de la temperatura y la adición de cualquier ácido diluido, ácido concentrado o solución diluida de álcali. La solución de pretratamiento se añade durante un tiempo suficiente para al menos hidrolizar parcialmente los componentes hemicelulosa y luego se neutraliza.

En una estrategia alternativa, se prefiere una composición de celulasa que es deficiente o libre de endoglucanasa. La supresión del gen de endoglucanasa sería particularmente útil en la preparación de composiciones de celulasa para su uso en detergentes. Adicionalmente, dichas composiciones son útiles para la producción de celooligosacáridos. La supresión del gen *egl7* de cepas de T. reesei sería particularmente útil en la preparación de composiciones de celulasa para su uso en los detergentes y en el aislamiento de celooligosacáridos. Las enzimas celulasa se han usado en una variedad de composiciones de detergente para limpiar ropa enzimáticamente. Sin embargo, se sabe en la técnica que el uso de enzimas celulasa puede impartir la degradación de las fibras de celulosa en la ropa. Una posibilidad para disminuir el efecto de degradación es producir un detergente que no contenga endoglucanasa. Así, la supresión de esta proteína afectaría al sistema de celulasa para inhibir a los otros componentes mediante acumulación de celobiosa. Los microorganismos modificados son particularmente adecuados para preparar dichas composiciones porque el gen *egl7* se puede suprimir dejando los restantes componentes CBH y EG dando como resultado una mejor limpieza y beneficios suavizantes en la composición sin efectos degradativos.

Las composiciones de detergente pueden emplear además de la composición de celulasa (con independencia del contenido en endoglucanasa, esto es, libre de endoglucanasa, endoglucanasa sustancialmente libre, o endoglucanasa aumentada), un tensoactivo, incluyendo tensoactivos aniónicos, no iónicos y anfolíticos, una hidrolasa, agentes de levantamiento, agentes blanqueadores, agentes de azulado y colorantes fluorescentes, inhibidores de apelmazamiento, solubilizantes, tensoactivos catiónicos y similares. Todos estos componentes se conocen en la técnica de detergentes. La composición de celulasa como se describió anteriormente se puede añadir a la composición de detergente, ya sea en un diluyente líquido, en gránulos, en emulsiones, en geles, en pastas, y similares. Dichas formas son conocidas por el experto en la materia. Cuando se emplea una composición de detergente sólida, la composición de celulasa se formula preferiblemente en gránulos. Preferiblemente los gránulos se pueden formular de manera que contengan un agente protector de celulasa. Para una discusión más a fondo, adviértase la Patente EE.UU. Número 6.162.782 denominada “Composiciones de detergentes que contienen composiciones de celulasa deficientes en componentes tipo CBHI”.

En aún otra realización, las composiciones de detergente pueden contener también niveles de endoglucanasa aumentados o endoglucanasa alterada. En este sentido, realmente depende del tipo de producto que se desea usar en las composiciones de detergente para dar los efectos deseados.

Preferiblemente las composiciones de celulasa se emplean de aproximadamente 0,00005 de porcentaje en peso a aproximadamente 5 de porcentaje de peso con respecto a la composición de detergente total. Más preferiblemente, las composiciones de celulasa se emplean de aproximadamente 0,0002 de porcentaje de peso a aproximadamente 2 de porcentaje de peso con respecto a la composición de detergente total.

Las porciones de secuencia de ácido nucleico *egl7* que son capaces de unirse a celulosa se pueden usar para generar proteínas bacterianas quiméricas de superficie, permitiendo la inmovilización de toda la célula en filtros de celulosa u otros soportes sólidos fibrosos como se describe en Lehtio *et al.*, 2001.

Además la secuencia de ácido nucleico de *egl7* encuentra utilidad en la identificación y caracterización de secuencias de ácido nucleico relacionadas. Un número de técnicas útiles para la determinación (predicción o confirmación) de la función de genes relacionados o productos génicos incluyen, pero no se limitan a, (A) análisis de DNA/RNA, como (1) sobreexpresión, expresión ectópica y expresión en otras especies; (2) bloqueo de genes (genéticas en reverso, bloqueo dirigido, silenciamiento de genes inducido por virus (VIGS, adviértase Baulcombe, 1999); (3) análisis del estado de metilación del gen, especialmente regiones reguladoras flanqueantes; e (4) hibridación in situ; (B) análisis de producto génico como (1) expresión de proteína recombinante; (2) producción de antisueros, (3) inmunolocalización; (4) ensayos bioquímicos para actividad catalítica u otra; (5) estado de fosforilación; e (6) interacción con otras proteínas mediante análisis de doble híbrido en levadura; (C) análisis de ruta, como colocar un gen o producto génico dentro de una ruta bioquímica o de señalización particulares basado en su fenotipo de sobreexpresión o por homología de secuencia con genes relacionados; y (D) otros análisis que se pueden también llevar a cabo para determinar o confirmar la participación del gen aislado y su producto en una ruta metabólica o de señalización particulares, y ayudar a determinar la función génica.

Las endoglucanasas y beta-glucosidasas pueden ser responsables de la producción de disacáridos, como soforosa, a partir de celooligosacáridos y glucosa mediante reacciones de transglicosilación. La soforosa se sabe que es un potente inductor de la expresión del gen de celulasa (Ilmen, M., et al., 1997, Appl. Environ. Microbiol. 63:1298-1306 y referencias en ello). De esta forma los EGs y BGLs pueden jugar un papel importante en el proceso de inducción de la expresión del gen de celulasa. La sobreexpresión de ciertos EGs o BGLs en una cepa fúngica puede dar lugar a una mayor productividad general de celulasa por esa cepa.

A. Homología con secuencias conocidas

La función de una secuencia de ácido nucleico relacionada que codifica EGVII se puede determinar por homología con genes conocidos que tienen una función particular. Por ejemplo, se usa una comparación de la secuencia codificadora de una molécula de ácido nucleico identificada con bases de datos públicas de secuencias de ácido nucleico para confirmar la función por homología con genes conocidos o por extensión de la secuencia de ácido nucleico identificada.

El término “% de homología” se usa de forma intercambiable en esta memoria con el término “% de identidad” en esta memoria y se refiere al nivel de identidad de secuencia de ácido nucleico o de aminoácidos entre la secuencia de ácido nucleico que codifica EGVII o la secuencia de aminoácidos de EGVII, cuando se alinean usando un programa de alineamiento de secuencia.

Por ejemplo, como se usa en esta memoria, 80% de homología significa lo mismo que 80% de identidad de secuencia determinado mediante un algoritmo definido, y en consecuencia, un homólogo de una secuencia dada tiene más de un 80% de identidad de secuencia en una longitud de la secuencia dada. Niveles modelo de identidad de secuencia incluyen, pero no se limitan a, 80, 85, 90, 95, 98% o más de identidad de secuencia con una secuencia dada, p.ej., la secuencia codificadora de *egl7*, como se describe en esta memoria.

Programas de ordenador modelo que se pueden usar para determinar identidad entre dos secuencias incluyen, pero no se limitan a, el conjunto de programas BLAST, p.ej., BLASTN, BLASTX, y TBLASTX, BLASTP y TBLASTN,

disponibles públicamente en internet en http://www.ncbi.nlm.nih.gov/BLAST/. Adviértase también, Altschul, *et al.*, 1990 y Altschul, *et al.*, 1997.

Las búsquedas de secuencia normalmente se llevan a cabo usando el programa BLASTN cuando se evalúa una secuencia de ácido nucleico dada en relación con secuencias de ácido nucleico en las secuencias de DNA del GenBank y otras bases de datos públicas. El programa BLASTX se prefiere para la búsqueda de secuencias de ácido nucleico que se han traducido en todos los marcos de lectura frente a secuencias de aminoácidos en las secuencias de proteínas del GenBank u otras bases de datos públicas. Tanto BLASTN como BLASTX se ejecutan usando parámetros por defecto de una penalización de hueco abierto de 11,0, y una penalización de hueco extendido de 1,0, y utilizando la matriz de BLOSUM-62. (Adviértase, p.ej., Altschul, *et al.*, 1997).

Se realiza un alineamiento preferido de secuencias seleccionadas con el fin de determinar "% de identidad" entre dos o más secuencias usando por ejemplo, el programa CLUSTAL-W en la versión MacVector 6.5, operado con parámetros por defecto, incluyendo una penalización de hueco abierto de 10,0, una penalización de hueco extendido de 0,1, y una matriz de similitud BOLSUM 30.

En una estrategia modelo, la extensión de secuencia de un ácido nucleico que codifica *egl7* se puede llevar a cabo usando procedimientos convencionales de extensión del cebador como se describe en Sambrook *et al.*, *supra*, para detectar precursores de *egl7* e intermediarios de procesamiento de mRNA que pueden no haber sido transcritos en reverso en cDNA y/o para identificar ORFs que codifican una proteína de longitud completa.

En aún otro aspecto, la presente descripción incluye la secuencia de nucleótidos total o parcial de la secuencia de ácido nucleico de *egl7* para su uso como una sonda. Dicha sonda se puede usar para identificar y clonar secuencias de ácidos nucleicos homólogos de organismos relacionados.

El cribado de librería de cDNA o genómica con la sonda seleccionada se puede llevar a cabo usando procedimientos estándar, como se describe en Sambrook *et al.*, (1989). Las condiciones de hibridación, incluyendo astringencia moderada y astringencia alta, se proporcionan en Sambrook *et al., supra*.

Las sondas o porciones de las mismas también se pueden emplear en técnicas de PCR para generar un grupo de secuencias para la identificación de secuencias *egl7* estrechamente relacionadas. Cuando las secuencias *egl7* se destinan para uso como sondas, se puede usar una porción particular de una secuencia que codifica EGVII, por ejemplo, una porción altamente conservada de la secuencia codificadora.

Por ejemplo, se puede usar una secuencia de nucleótidos *egl7* como una sonda de hibridación para una librería de cDNA para aislar genes, por ejemplo, aquellas variantes codificadas que ocurren naturalmente de EGVII de otras especies fúngicas, bacterianas o de plantas, que tienen un nivel deseado de identidad de secuencia con la secuencia de nucleótidos de *egl7* descrita en la Figura 1 (SEQ ID NO: 1). Las sondas modelo tienen una longitud de aproximadamente 20 a aproximadamente 50 bases.

B. Análisis de doble híbrido

Las proteínas identificadas se pueden usar en el sistema de doble híbrido de levadura para "capturar" proteínas que unen proteína que son proteínas teóricas de vías de señalización. El sistema de doble híbrido de levadura se describe en Fields y Song, Nature 340:245-246 (1989). Brevemente, en un sistema de doble híbrido, se construye una fusión de un dominio de unión a DNA y un dominio *egl7* (*p.ej.*, fusión GAL4-*egl7*) y se transfecta en células de levadura. Se pueden usar todo el gen *egl7*, o subregiones del gen *egl7*. Se co-transfecta un segundo constructo que contiene la librería de parejas de unión potenciales fusionadas al dominio de activación de DNA. Las levaduras co-transformantes que albergan proteínas que se unen a la EGVII, la proteína se identifica mediante, por ejemplo, producción de beta-galactosidasa o luciferasa (un cribado), o supervivencia en placas que carecen de un nutriente esencial (una selección), según sea apropiado para los vectores usados.

C. Análisis por microarrays

Además, se puede usar el análisis por microarray, también conocido como perfil de expresión o perfil de transcripción para evaluar simultáneamente la presencia expresión de una secuencia de DNA dada, o cambios en la expresión de muchos genes diferentes. En una estrategia, un gran conjunto de secuencias de DNA (sondas), por lo general un conjunto amplio marcadores de secuencias expresadas, cDNA, fragmentos de cDNA, u oligonucleótidos específicos de secuencia, se exhiben en un soporte sólido como un portaobjetos de vidrio o una membrana de nylon. El marcaje de la diana para la hibridación con las sondas se genera mediante el aislamiento de mRNA de un tejido control e inducido, marcando después cada grupo de mRNA ya sea a través de un intermedio de cDNA o cRNA, con un marcador distinto, normalmente un colorante fluorescente. El microarray se hibrida con las sondas complejas, y la intensidad relativa de señal de hibridación asociada con cada localización en el array se puede cuantificar para cada colorante marcador. Se pueden medir las diferencias en expresión entre estados de control e inducidos como una relación de la señal de los dos colorantes marcadores. (Adviértase Baldwin, D *et al.*, 1999).

El análisis por microarray del organismo fuente del que se derivó *egl7* se puede llevar a cabo para facilitar la comprensión de la función del gen mediante identificación de otros genes que están coordinadamente regulados

como consecuencia de la sobre expresión de *egl7*. La identidad de genes coordinadamente regulados puede ayudar a colocar el gen *egl7* en una ruta particular. Alternativamente, dicho análisis se puede utilizar para identificar otros genes involucrados en la misma ruta usando análisis por microarray.

Aunque se ha descrito con referencia a métodos y realizaciones específicos, se valorará que se puedan hacer varias modificaciones y cambios sin apartarse de la descripción, como se define en las reivindicaciones adjuntas.

Ejemplo 1

En una realización modelo, se aísla un fragmento de cDNA para usar como sonda mediante extracción de RNA total a partir del micelio de una cepa de *T. reesei* creciendo bajo condiciones conocidas para inducir la producción de celulasa y obtener la fracción poliadenilada (polyA) de la misma. El RNA poliA se usa para producir un grupo de cDNA que es amplificado después usando cebadores basados en la secuencia de ácido nucleico de *egl7* proporcionada en esta memoria.

El RNA total se aísla del micelio usando métodos conocidos en la técnica, por ejemplo, como se describe en Timberlake *et al.*, 1981; Maniatis, *et al.*, 1989; Ausubel, *et al.*, 1993 y Sambrook *et al.*, 1989. Una vez aislado, se realizan transferencias de Northern para confirmar la expresión de celulasa y seleccionar un tiempo de inducción óptimo para la expresión de celulasa y correspondiente aislamiento de RNA.

El RNA mensajero (mRNA), que tiene una cola de poli (A) en el extremo 3' se puede purificar del mRNA total usando métodos conocidos en la técnica.

El RNA de T. reesei se usa como molde para métodos conocidos en la técnica que usan RT-PCR (Loftus, J. et al., Science, 249:915-918, 1990). Durante este procedimiento el mRNA se transcribe en reverso para producir cDNA de primera cadena. Después el cDNA sirve como molde para la amplificación de PCR de las secuencias de cDNA de *egl7* usando cebadores de oligonucleótidos específicos diseñados de acuerdo con SEQ ID Nº. 1 o SEQ ID Nº. 4.

Tabla 1. Secuencias proporcionadas

Descripción	SEQ. ID. NO.
Longitud total de la secuencia de ácido nucleico de cDNA de *egl7* de T. reesei	**1**
ATTTAAGGATCTAAGCCCCATCGATATGAAGTCCTGCGCCATTCTTGCAG CCCTTGGCTGTCTTGCCGGGAGCGTTCTCGGCCATGGACAAGTCCAAAACT TCACGATCAATGGACAATACAATCAGGGTTTCATTCTCGATTACTACTATC AGAAGCAGAATACTGGTCACTTCCCCAACGTTGCTGGCTGGTACGCCGAGG ACCTAGACCTGGGCTTCATCTCCCCTGACCAATACACCACGCCCGACATTG TCTGTCACAAGAACGCGGCCCCAGGTGCCATTTCTGCCACTGCAGCGGCCG GCAGCAACATCGTCTTCCAATGGGGCCCTGGCGTCTGGCCTCACCCCTACG GTCCCATCGTTACCTACGTGGTTGAGTGCAGCGGATCGTGCACGACCGTGA ACAAGAACAACCTGCGCTGGGTCAAGATTCAGGAGGCCGGCATCAACTAT AACACCCAAGTCTGGGCGCAGCAGGATCTGATCAACCAGGGCAACAAGTG GACTGTGAAGATCCCGTCGAGCCTCAGGCCCGGAAACTATGTCTTCCGCCA TGAACTTCTTGCTGCCCATGGTGCCTCTAGTGCGAACGGCATGCAGAACTA TCCTCAGTGCGTGAACATCGCCGTCACAGGCTCGGGCACGAAAGCGCTCCC TGCCGGAACTCCTGCAACTCAGCTTTACAAGCCCACTGACCCTGGCATCTT GTTCAACCCTTACACAACAATCACGAGCTACACCATCCCTGGCCCAGCCCT GTGGCAAGGCTAG**ATCCAGGGGTACGGTGTTGGCGTTCGTGAAGTCGG** **AGCTGTTGACAAGGATATCTGATGATGAACGGAGAGGACTGATGGGC** **GTGACTGAGTGTATATATTTTTGATGACCAAATTGTATACGAAATCCGA** **ACGCATGGTGATCATTGTTTATCCCTGTAGTATATTGTCTCCAGGCTGC** **TAAGAGCCCACGGGTGTATTACGGCAACAAAGTCAGGAATTTGGGTGG** **CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA**	
Secuencia de aminoácidos predicha de EGVII de T. reesei HGQVQNFTINGQYNQGFILDYYYQKQNTGHFPNVAGWYAEDLDLGFISPDQY TTPDIVCHKNAAPGAISATAAAGSNIVFQWGPGVWPHPYGPIVTYVVECSGSC	2

Descripción	SEQ. ID. NO.
TTVNKNNLRWVKIQEAGINYNTQVWAQQDLINQGNKWTVKIPSSLRPGNYVF RHELLAAHGASSANGMQNYPQCVNIAVTGSGTKALPAGTPATQLYKPTDPGI LFNPYTTITSYTIPGPALWQG	
Secuencia señal predicha de EGVII de T. reesei: MKSCAILAALGCLAGSVLG	3
Secuencia de ácido nucleico codificadora de *egl7* de T. reesei ATGAAGTCCTGCGCCATTCTTGCAGCCCTTGGCTGTCTTGCCGGGAGCGTT CTCGGCCATGGACAAGTCCAAAACTTCACGATCAATGGACAATACAATCA GGGTTTCATTCTCGATTACTACTATCAGAAGCAGAATACTGGTCACTTCCC CAACGTTGCTGGCTGGTACGCCGAGGACCTAGACCTGGGCTTCATCTCCCC TGACCAATACACCACGCCCGACATTGTCTGTCACAAGAACGCGGCCCCAG GTGCCATTTCTGCCACTGCAGCGGCCGGCAGCAACATCGTCTTCCAATGGG GCCCTGGCGTCTGGCCTCACCCCTACGGTCCCATCGTTACCTACGTGGTTG AGTGCAGCGGATCGTGCACGACCGTGAACAAGAACAACCTGCGCTGGGTC AAGATTCAGGAGGCCGGCATCAACTATAACACCCAAGTCTGGGCGCAGCA GGATCTGATCAACCAGGGCAACAAGTGGACTGTGAAGATCCCGTCGAGCC TCAGGCCCGGAAACTATGTCTTCCGCCATGAACTTCTTGCTGCCCATGGTG CCTCTAGTGCGAACGGCATGCAGAACTATCCTCAGTGCGTGAACATCGCCG TCACAGGCTCGGGCACGAAAGCGCTCCCTGCCGGAACTCCTGCAACTCAG CTTTACAAGCCCACTGACCCTGGCATCTTGTTCAACCCTTACACAACAATC ACGAGCTACACCATCCCTGGCCCAGCCCTGTGGCAAGGCTAG	4

REIVINDICACIONES

1. Un polinucleótido aislado seleccionado del grupo que consiste en:

(a) una secuencia de ácido nucleico que codifica o es complementaria a una secuencia que codifica un polipéptido que tiene al menos un 85% de identidad de secuencia con la secuencia de aminoácidos que se presenta como SEQ ID NO:2; y

(b) una secuencia de ácido nucleico presentada como SEQ ID NO:4, o la complementaria de la misma;

en donde dicho polinucleótido aislado codifica un polipéptido que tiene actividad endoglucanasa.

2. El polinucleótido aislado de acuerdo con la reivindicación 1 en donde la secuencia de ácido nucleico codifica o es complementaria a una secuencia que codifica un polipéptido que tiene al menos un 90% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2.

3. El polinucleótido aislado de acuerdo con la reivindicación 1 en donde la secuencia de ácido nucleico codifica o es complementaria a una secuencia que codifica un polipéptido que tiene al menos un 95% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2.

4. El polinucleótido aislado de acuerdo con la reivindicación 1 en donde la secuencia de ácido nucleico codifica o es complementaria a una secuencia que codifica un polipéptido que tiene la secuencia de aminoácidos presentada como SEQ ID NO: 2.

5. El polinucleótido aislado de cualquiera de las reivindicaciones de 1 a 4, en donde el % de identidad se calcula usando el programa CLUSTAL-W en la versión MacVector 6.5, operado con parámetros por defecto, incluyendo una penalización por hueco abierto de 10,0, una penalización de hueco extendido de 0,1, y una matriz de similitud BLOSUM 30.

6. El polinucleótido aislado de cualquiera de las reivindicaciones de 1 a 4 que codifica una enzima que tiene actividad endoglucanasa, en donde la enzima se deriva de una fuente Trichoderma.

7. El polinucleótido aislado de la reivindicación 6, en donde la enzima se deriva de *Trichoderma reesei*.

8. Un constructo de expresión incluyendo una secuencia de polinucleótidos (i) que codifica un polipéptido que tiene al menos un 85% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2, o (ii) que es capaz de hibridar con una sonda derivada de una secuencia polipeptídica descrita como SEQ ID NO: 2 bajo condiciones de astringencia de intermedia a alta, o (iii) que es complementaria a una secuencia de nucleótidos que codifica una polipéptido que tiene al menos un 85% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2;

en donde dicha secuencia de polinucleótidos codifica un polipéptido que tiene actividad endoglucanasa.

9. Un vector que incluye el constructo de expresión de la reivindicación 8.

10. Un vector que comprende un polinucleótido aislado de cualquiera de las reivindicaciones de 1 a 4, ligado operativamente a secuencias control reconocidas por una célula huésped transformada con el vector.

11. Una célula huésped transformada con el vector de la reivindicación 9 o de la reivindicación 10.

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 cualquiera de las reivindicaciones de 1 a 4.

15. La célula huésped de la Reivindicación 14, que es una célula procariota.

16. La célula huésped de la Reivindicación 14, que es una célula eucariota.

17. Un polipéptido sustancialmente purificado con la actividad biológica de una endoglucanasa, que comprende una secuencia seleccionada del grupo que consiste en:

(a) una secuencia de aminoácidos que tiene al menos un 85% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2; o (b) un fragmento biológicamente activo sustancialmente purificado de la secuencia de aminoácidos presentada como SEQ ID NO: 2.

18. El polipéptido sustancialmente purificado de acuerdo con la reivindicación 17, que comprende una secuencia de aminoácidos que tiene al menos un 90% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2.

19. El polipéptido sustancialmente purificado de acuerdo con la reivindicación 17, que comprende una secuencia de aminoácidos que tiene al menos un 95% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2.

20. El polipéptido sustancialmente purificado de acuerdo con la reivindicación 17, que comprende una secuencia de aminoácidos presentada como SEQ ID NO: 2.

21. Un método de producción de una enzima que tiene actividad endoglucanasa, que comprende:

(a) transformar de manera estable una célula huésped con un vector de expresión que comprende un polinucleótido como se define en cualquiera de las reivindicaciones de 1 a 4;

(b) cultivar dichas células huésped transformadas bajo condiciones adecuadas para que dicha célula produzca dicha endoglucanasa; y

(c) recuperar dicha endoglucanasa.

22. El método de la reivindicación 21 en donde la célula huésped es una célula de hongo filamentoso o de levadura.

23. Una composición de detergente, dicha composición comprende un polipéptido seleccionado del grupo que consiste en:

(a) una secuencia de aminoácidos que tiene al menos un 85% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2; o (b) un fragmento biológicamente activo sustancialmente purificado de la secuencia de aminoácidos presentada como SEQ ID NO: 2;

en donde dicho polipéptido tiene actividad endoglucanasa.

24. La composición de detergente de acuerdo con la reivindicación 23, que comprende una secuencia de aminoácidos que tiene al menos un 90% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2.

25. La composición de detergente de acuerdo con la reivindicación 23, que comprende una secuencia de aminoácidos que tiene al menos un 95% de identidad de secuencia con la secuencia de aminoácidos presentada como SEQ ID NO: 2.

26. La composición de detergente de acuerdo con la reivindicación 23, que comprende una secuencia de aminoácidos presentada como SEQ ID NO: 2.

27. Un método de expresar un polipéptido recombinante, el método comprende:

(i) proporcionar células huésped que han sido transducidas, transformadas o transfectadas con un vector de expresión que comprende una secuencia de ácido nucleico de acuerdo a cualquiera de las reivindicaciones de 1 a 4;

(ii) cultivar las células.

28. Un método como se reivindica en la reivindicación 27 en donde el vector de expresión incluye una secuencia que codifica un péptido señal adicional.

29. Un método como se reivindica en la reivindicación 27 en donde las células huésped son Aspergillus sp.

30. Un método de convertir glucosa a etanol, el método comprende:

(i) el uso de una composición enzimática que comprende el polipéptido de acuerdo con cualquiera de las reivindicaciones de 17 a 20 para hidrolizar la celulosa en sus componentes azúcares; y;

(ii) la fermentación de los componentes azúcares a etanol.

31. El método de la reivindicación 30 en donde la composición está enriquecida en el polipéptido de acuerdo con cualquiera de las reivindicaciones de 17 a 20.

32. El método de la reivindicación 30 o reivindicación 31 en donde la composición enzimática se añade a la biomasa celulósica antes de la adición de un organismo fermentativo.

33. El método de la reivindicación 30 o reivindicación 31 en donde la composición enzimática se añade a la biomasa celulósica al mismo tiempo que un organismo fermentativo.

34. El método de la reivindicación 33 que comprende un proceso simultaneo de sacarificación y fermentación por el que la celulosa presente en la biomasa es convertida a glucosa y al mismo tiempo y en el mismo reactor las cepas de levadura convierten la glucosa en etanol.

35. El método de cualquiera de las reivindicaciones de 30 a 34 en donde la composición comprende además enzimas celulasa adicionales seleccionadas de: otras endoglucanasas; beta-glucosidasas y/o celobiohidrolasas.

36. El método de cualquiera de las reivindicaciones de 30 a 35 en donde la celulosa se proporciona a través de: pulpa de madera, papel, algodón, pañales de celulosa, rastrojo de maíz, plantas herbáceas, residuos sólidos urbanos o residuos agrícolas y forestales.

37. Un método de cualquiera de las reivindicaciones de 30 a 36 en donde la celulosa se proporciona a través de materia prima celulósica que se pretrata con un pretratamiento seleccionado de:

pretratamiento con ácido diluido;

pretratamiento con temperatura elevada y la adición de cualquier ácido diluido, ácido concentrado o álcali diluido.

ATTTAAGGATCTAAGCCCCATCGATATGAAGTCCTGCGCCATTCTTGCAGCCCTTGGCTGTCTTGCCGGGAGCGTTCTCGG
CCATGGACAAGTCCAAAACTTCACGATCAATGGACAATACAATCAGGGTTTCATTCTCGATTACTACTATCAGAAGCAGAA
TACTGGTCACTTCCCCAACGTTGCTGGCTGGTACGCCGAGGACCTAGACCTGGGCTTCATCTCCCCTGACCAATACACCAC
GCCCGACATTGTCTGTCACAAGAACGCGGCCCCAGGTGCCATTTCTGCCACTGCAGCGGCCGGCAGCAACATCGTCTTCCA
ATGGGGCCCTGGCGTCTGGCCTCACCCCTACGGTCCCATCGTTACCTACGTGGTTGAGTGCAGCGGATCGTGCACGACCGT
GAACAAGAACAACCTGCGCTGGGTCAAGATTCAGGAGGCCGGCATCAACTATAACACCCAAGTCTGGGCGCAGCAGGATCT
GATCAACCAGGGCAACAAGTGGACTGTGAAGATCCCGTCGAGCCTCAGGCCCGGAAACTATGTCTTCCGCCATGAACTTCT
TGCTGCCCATGGTGCCTCTAGTGCGAACGGCATGCAGAACTATCCTCAGTGCGTGAACATCGCCGTCACAGGCTCGGGCAC
GAAAGCGCTCCCTGCCGGAACTCCTGCAACTCAGCTTTACAAGCCCACTGACCCTGGCATCTTGTTCAACCCTTACACAAC
AATCACGAGCTACACCATCCCTGGCCCAGCCCTGTGGCAAGGCTAG**ATCCAGGGGTACGGTGTTGGCGTTCGTGAAGTCGG**
AGCTGTTGACAAGGATATCTGATGATGAACGGAGAGGACTGATGGGCGTGACTGAGTGTATATATTTTTGATGACCAAATT
GTATACGAAATCCGAACGCATGGTGATCATTGTTTATCCCTGTAGTATATTGTCTCCAGGCTGCTAAGAGCCCACGGGTGT
ATTACGGCAACAAAGTCAGGAATTTGGGTGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Figura 1
Secuencia de nucleótidos de *egl7* de *T. reesei* (regiones codificantes y no-codificantes).

```
MKSCAILAAL GCLAGSVLGH GQVQNFTING QYNQGFILDY YYQKQNTGHF PNVAGWYAED LDLGFISPDQ
         10         20         30         40         50         60         70

YTTPDIVCHK NAAPGAISAT AAAGSNIVFQ WGPGVWPHPY GPIVTYVVEC SGSCTTVNKN NLRWVKIQEA
         80         90        100        110        120        130        140

GINYNTQVWA QQDLINQGNK WTVKIPSSLR PGNYVFRHEL LAAHGASSAN GMQNYPQCVN IAVTGSGTKA
        150        160        170        180        190        200        210

LPAGTPATQL YKPTDPGILF NPYTTITSYT IPGPALWQG
        220        230        240        249
```

Figura 2
Secuencia de aminoácidos predicha de *egl7* de *T. reesei.*