

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 371 362**

51 Int. Cl.:
C07K 14/37 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07014764 .0**
96 Fecha de presentación: **28.05.2004**
97 Número de publicación de la solicitud: **1862626**
97 Fecha de publicación de la solicitud: **05.12.2007**

54 Título: **NUEVOS GENES DE TRICHODERMA.**

30 Prioridad:
29.05.2003 US 474411 P
03.06.2003 US 475826 P

45 Fecha de publicación de la mención BOPI:
30.12.2011

45 Fecha de la publicación del folleto de la patente:
30.12.2011

73 Titular/es:
GENENCOR INTERNATIONAL, INC.
925 Page Mill Road
Palo Alto, California 94304, US

72 Inventor/es:
Foreman, Pamela;
Goedegebuur, Frits;
Van Solingen, Pieter y
Ward, Michael

74 Agente: **Ponti Sales, Adelaida**

ES 2 371 362 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

Nuevos genes de *Trichoderma*

5 CAMPO DE LA INVENCION

[0001] Se describen en la presente invención cuatro genes – dos genes que codifican proteínas que comprenden un dominio de unión a celulosa, una arabinofuranosidasa, y una acetilxilano esterasa. También se describen aquí las proteínas deducidas, y las composiciones que tienen las nuevas proteínas. Estas composiciones son especialmente
10 útiles en aplicaciones textiles, detergentes, conversión de biomasa, pienso y alimentos, y las industrias de la pulpa y el papel. Los genes se aislaron a partir de un hongo filamentoso, *Trichoderma reesei* (también denominado indistintamente aquí *Hypocrea jecorina*).

15 ANTECEDENTES DE LA INVENCION

[0002] La celulosa y la hemicelulosa son los materiales más abundantes de las plantas producidos por fotosíntesis. Se pueden degradar y utilizar como 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., J. Biol. Chem., 10.1074/ M003624200, 13 de abril de 2001). Dado que se acercan
20 los límites de los recursos no renovables, el potencial de la celulosa para convertirse en una energía renovable principal es enorme (Krishna et al., Bioresource Tech. 77:193-196, 2001). La utilización eficaz de la celulosa a través de procesos biológicos es una estrategia para superar la escasez de alimentos, pienso y combustibles (Ohmiya et al., Biotechnol. Gen. Engineer. Rev. 14:365-414, 1997).

[0003] La celulosa es un polisacárido lineal de residuos de glucosa conectados mediante uniones β -1,4. En la naturaleza, la celulosa está normalmente asociada con lignina junto con hemicelulosas, tales como xilanos y glucomananos. El uso práctico de las celulasas está dificultado por la naturaleza de las celulasas conocidas, que a menudo son mezclas de celulasas que tienen una variedad actividades y especificidades de sustrato. Por esta razón, es deseable identificar celulasas que tienen sólo las actividades deseadas o proteínas que pueden facilitar la acción de la celulasa.
30

[0004] La hemicelulosa es cualquiera de los diversos heteropolímeros (polisacáridos en matriz) presentes en casi todas las paredes celulares junto con celulosa. Sus pesos moleculares son habitualmente inferiores a los de la celulosa y presentan una estructura no diferenciada débil en comparación con la celulosa cristalina. Las cadenas forman una “base” - se unen con pectina a la celulosa para formar un red de fibras reticuladas. De este modo, sería beneficioso aumentar la degradación de la hemicelulosa.
35

[0005] Las O-Glicosil hidrolasas (EC 3.2.1.-) con un grupo amplio de enzimas que hidrolizan el enlace glicosídico entre dos o más carbohidratos, o entre un carbohidrato y un grupo no carbohidrato. El sistema de clasificación para las glicosil hidrolasas, basada en la similitud de secuencias, ha conducido a la definición de hasta 60 familias diferentes [HENRISSAT, B. AND BAIROCH, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. BIOCHEM.J. 293 781-788 (1993); HENRISSAT, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. BIOCHEM.J. 280 309-316 (1991); DAVIES, G. AND HENRISSAT, B. Structures and mechanisms of glycosyl hydrolases. STRUCTURE 3 853-859 (1995); and HENRISSAT, B. AND BAIROCH, A. Updating the sequence-based classification of glycosyl hydrolases. BIOCHEM.J. 316 695-696 (1996)]. Las acetil xilano esterasas (EC 3.1.1.72) son un grupo de enzimas que eliminan grupos laterales acetilo a partir de xilano. El sistema de clasificación para carbohidrato esterasas, basado en la similitud de secuencias, ha conducido a la definición de 13 familias, siete de las cuales contienen acetil xilano esterasas (COUTINHO, P.M. AND HENRISSAT, B., 1999 Carbohydrate-active enzymes server at URL: <<http://afmb.cnrs-mrs.fr/CAZY/index.html>>).
40
45
50

[0006] Con el fin de ser eficientes, la digestión de celulosa requiere de varios tipos de enzimas actuando de manera cooperativa. Son necesarias por lo menos tres categorías de enzimas para convertir la celulosa en glucosa: endo (1,4)-beta-D-glucanasas (EC 3.2.1.4) que cortan las cadenas de celulosa al azar; celobiohidrolasas (EC 3.2.1.91) que dividen las unidades de celobiosilo de los extremos de las cadenas de celulosa y beta-glucosidasas (EC 3.2.1.21) que convierten la celobiosa y las celodextrinas solubles en glucosa.
55

[0007] WO 95/16782 describe enzimas celulasa y sistemas para su expresión.

60 [0008] WO 01/46357 describe proteínas de fusión que comprenden un dominio de unión a celulosa y un dominio que

tiene una afinidad elevada por otro ligando.

[0009] WO03/031477 describe proteínas de fusión que comprenden un dominio de unión a carbohidrato y por lo menos un dominio de unión adicional capaz de unirse a (a) un ligando o sitio específico que forma parte de un organismo vivo, o (b) una microcápsula o micropartícula.

[0010] Un objetivo de la presente invención es proporcionar proteínas mejoradas que tienen una actividad degradante de celulosa o hemicelulosa y polinucleótidos que codifican las proteínas. Un objetivo de la presente invención es proporcionar proteínas mejoradas que tienen actividad de unión a celulosa o hemicelulosa y polinucleótidos que codifican las proteínas. Las proteínas mejoradas pueden mejorar la degradación del material de la pared celular, por ejemplo, celulosa y/o hemicelulosa. Las proteínas también pueden mejorar la estabilidad o la actividad de otras enzimas implicadas en la degradación del material de la pared celular de la planta, por ejemplo, biomasa.

DESCRIPCIÓN RESUMIDA DE LA INVENCION

[0011] Se describen en la presente invención genes nuevos, denominados aquí *cip1*, *cip2*, *axe2* y *abf2*. También se describen en la presente invención los productos génicos codificados por los genes nuevos. Por lo menos dos genes se coexpresan con los genes en la familia de celulasa.

CIP2

[0012] En un primer aspecto, la presente invención se refiere a un polinucleótido tal como se define en las reivindicaciones. El polinucleótido puede ser ARNm, ADN, ADNc, ADN genómico o un análogo no codificante de los mismos. En una realización específica, el polinucleótido comprende una secuencia sustancialmente idéntica a la SEC ID No. 6.

[0013] En un segundo aspecto, se proporcionan polipéptidos o proteínas CIP2 tal como se definen en las reivindicaciones.

[0014] En una realización, la presente invención incluye (i) fragmentos de CIP2 de por lo menos 200 aminoácidos de longitud, y (ii) una composición que comprende CIP2.

[0015] En un tercer aspecto, la presente invención se refiere a una construcción de ácido nucleico que comprende la secuencia de nucleótidos, que codifica el polipéptido de la presente invención, unida operativamente a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.

[0016] En un cuarto aspecto, la presente invención se refiere a un vector de expresión recombinante que comprende la construcción de ácido nucleico de la invención.

[0017] La presente invención proporciona además vectores de expresión recombinantes que contienen una secuencia de ácido nucleico que codifica CIP2 o un fragmento de la misma, unida operativamente a elementos reguladores eficaces para la expresión de la proteína en un huésped seleccionado. En un aspecto relacionado, la presente invención incluye una célula huésped que contiene el vector.

[0018] En un quinto aspecto, la presente invención se refiere a una célula huésped recombinante que comprende la construcción de ácido nucleico de la invención.

[0019] La presente invención incluye además un método para producir CIP2 mediante técnicas recombinantes, mediante el cultivo de células huésped procariotas o eucariotas recombinantes que comprenden una secuencia de ácido nucleico que codifica CIP2 bajo condiciones efectivas para promover la expresión de la proteína, y la posterior recuperación de la proteína de la célula huésped o del medio de cultivo celular.

[0020] En un sexto aspecto, la presente invención se refiere a un método para producir un polipéptido de la invención, comprendiendo el método: (a) cultivar un microorganismo tal como se define en las reivindicaciones; y (b) recuperar el polipéptido.

BREVE DESCRIPCIÓN DE LAS FIGURAS

[0021] La figura 1 es una representación en cadena sencilla de la secuencia de ácidos nucleicos (SEC ID NO:1), del

ADNc de *cip1* de *T. reesei*, donde la secuencia no codificante está subrayada.

[0022] La figura 2 es la secuencia codificante para *cip1* de *T. reesei* (SEC ID No. 2), donde la secuencia señal codificada se indica con los nucleótidos en negrita.

[0023] La figura 3 muestra la secuencia de aminoácidos prevista de CIP1 (SEC ID no. 3), la secuencia señal (SEC ID No. 4) y la secuencia de la proteína madura (SEC ID No. 5) en base a la secuencia de nucleótidos proporcionada en la figura 1.

[0024] La figura 4 es la secuencia codificante para *cip2* de *T. reesei* (SEC ID NO:6).

[0025] La figura 5 muestra la secuencia de aminoácidos prevista de CIP2 (SEC ID No. 7), la secuencia señal (SEC ID No. 8) y la secuencia de la proteína madura (SEC ID No. 9) en base a la secuencia de nucleótidos proporcionada en la figura 4.

[0026] La figura 6 es una alineación de CIP2 con *R.flavefaciens cesA CAB55348*. CIP2 tiene la secuencia señal N-terminal prevista de 17 aminoácidos seguido de 36 aminoácidos que comprende una molécula de unión a carbohidrato de la familia CBM1 y región de unión que acaba aproximadamente en el aminoácido 95.

[0027] La figura 7 es una representación en cadena sencilla de la secuencia de ácido nucleico (SEC ID No. 10), del gen *abf2* de *T. reesei* que codifica una arabinofuranosidasa.

[0028] La figura 8 muestra la secuencia de aminoácidos prevista de ABF2 (SEC ID No. 11), la secuencia señal (SEC ID No. 12) y la secuencia de la proteína madura (SEC ID No. 13) basadas en la secuencia de nucleótidos proporcionada en la figura 7.

[0029] La figura 9 es una alineación de ABF2 con ARF1 de *C. carbonum* y *stxIV* de *S. thermoviolaceus*.

[0030] La figura 10 es la secuencia de ADNc para el gen *axe2* (SEC ID No. 14) que codifica una acetilxilano esterasa.

[0031] La figura 11 muestra la secuencia de aminoácidos prevista de AXE2 (SEC ID no. 15), la secuencia señal (SEC ID No. 16) y la secuencia de la proteína madura (SEC ID No. 17) en base a la secuencia de nucleótidos proporcionada en la figura 9.

[0032] La figura 12 es una alineación de secuencia de AXE2 con AXE1 de *T. reesei*.

[0033] La figura 13 muestra una transferencia Northern para dos cepas fúngicas bajo condiciones VARIANTEes. Os cultivos de QM6a y RLP-37 se desarrollaron en glucosa (carriles A), celulosa (Carriles B), glicerol (Carriles C) o glicerol complementado con soforosa (carriles D), El ARNm de cada uno de los cultivos se analizó mediante transferencia Northern. El grupo de bandas superior para cada gen se sondó con ADNcs marcados tal como se indica. El grupo de bandas inferior para cada gen se sondó con una sonda de actina para corregir las diferencias en la carga y las diferencias en los tiempos de expansión requeridos para visualizar las bandas.

[0034] La figura 15 es una representación de los resultados a partir de un análisis microarray realizado para valorar los niveles de expresión para cada uno de los genes indicados. A) Análisis de matriz de agitación de la inducción soforosa en dos cepas diferentes. El ARNm de los cultivos de QM6a y RL-P37 desarrollados en glicerol o glicerol complementado soforosa 1 mM se marcaron individualmente con los colorantes fluorescentes Cy5 y con Cy3, El ARNm marcado de los cultivos desarrollados con soforosa se combinó con ARNm marcado recíprocamente de los cultivos desarrollados con glicerol y se hibridaron a microarrays. La proporción logarítmica de las dos muestras de ARNm marcadas diferentes que se unieron a sondas para cada uno de los genes se indica según la barra de colores siguiente. L.R.: proporción logarítmica. El color refleja la magnitud de la inducción mediada con soforosa de cada uno de los genes indicados. Columna 1: inducción por soforosa en cultivos RL-P37. Columna 2: inducción por soforosa en cultivos QM6a. Columna 3: ARNm marcado fluorescentemente de cultivos RL-P37 inducidos por soforosa se cohibridizó con ARNm marcado recíprocamente de cultivos QM6a inducidos por soforosa desarrollados en condiciones similares. El color refleja la abundancia de ARNm correspondiente a cada uno de los genes en RL-P37 en relación a QM6a. B) Análisis de los niveles de expresión durante el cultivo en diferentes fuentes de carbono en fermentadores. Se desarrollaron inicialmente los micelios de RLP-37 y QM6a en medio que contiene glucosa. Una hora más tarde se haber utilizado la glucosa completamente; los cultivos fueron alimentados con lactosa a una velocidad que evitaba la acumulación en el medio. Las muestras se obtuvieron durante la alimentación de glucosa

durante la privación de carbono y después de 24 y 48 horas después del comienzo de la alimentación de lactosa. Se utilizaron microarrays para determinar los niveles de expresión en cada uno de los tiempos relacionados con la expresión de la privación de carbono. Columna 4: alimentación de RL-P37 con glucosa, columna 5: alimentación de QM6a con glucosa, columna 6: alimentación de 24 horas de RL-P37 con lactosa, columna 7: alimentación de 48 horas de RL-P37 con lactosa, columna 8: alimentación de 24 horas de QM6a con lactosa, columna 9: alimentación de 48 horas de QM6a con lactosa.

[0035] La figura 15 es un mapa esquemático del vector pREP3Y.

[0036] La figura 16 es una alineación de secuencias de CIP1 con potencial hidrolasa secretada de *Streptomyces coelicolor* A3 (número de acceso CAA18323).

[0037] La figura 17 es un esquema del vector pTrex3g.

[0038] La figura 18 es un esquema del vector pENTR/D-TOPO (Invitrogen).

[0039] La figura 19 es un esquema de la construcción de pExpression que comprenderá un gen de interés. El gen de interés se selecciona entre *cip1* o *cip2* o *axe2* o *abf2*.

[0040] La figura 20 es una fotografía de un gel de SDS-PAGE de sobrenadante en el matraz de agitación de células huésped transformadas con un vector de expresión que comprende el gen *cip1*. El carril 1 contiene marcadores de peso molecular, Mark12, de Invitrogen. Los carriles 2-12 el sobrenadante de cepas transformantes individuales. La flecha en la izquierda del gel designa allí donde se localizaría en el gel la proteína CIP1 si se expresara y secretara en cantidades detectables.

[0041] La figura 21 es una fotografía de un gel de SDS-PAGE de sobrenadante en el matraz de agitación de células huésped transformadas con un vector de expresión que comprende el gen *cip2*. El carril 1 contiene marcadores de peso molecular, Mark12, de Invitrogen. Los carriles 2-12 el sobrenadante de cepas transformantes individuales. La flecha en la izquierda del gel designa allí donde se localizaría en el gel la proteína CIP2 si se expresara y secretara en cantidades detectables.

[0042] La figura 22 es una fotografía de un gel de SDS-PAGE de sobrenadante en el matraz de agitación de células huésped transformadas con un vector de expresión que comprende el gen *abf2*. El carril contiene el sobrenadante de de la cepa de *Trichoderma* con eliminación de quad descrita aquí. El carril 2 contiene marcadores de peso molecular, Mark12, de Invitrogen. El carril 3 contiene el sobrenadante de un transformante individual de *abf2*. La flecha en la derecha del gen designa la banda correspondiente donde se localizaría la proteína ABF2 si se expresara.

DESCRIPCIÓN DETALLADA DE LA INVENCION

[0043] La presente invención se describirá en detalle a modo de referencia utilizando sólo las siguientes definiciones y ejemplos.

[0044] A menos que se defina aquí lo contrario, todos los términos técnicos y científicos utilizados aquí tienen el mismo significado que entiende normalmente un experto en la materia al que pertenece la presente invención. Singleton, et al., *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY*, 2D ED., John Wiley and Sons, New York (1994), y Hale & Marham, *THE HARPER COLLINS DICTIONARY OF BIOLOGY*, Harper Perennial, NY (1991) proporcionan a un experto un diccionario general de muchos de los términos utilizados en la presente invención. Aunque se pueden utilizar en la práctica o el análisis de la presente invención cualquier método y material similar o equivalente a los descritos aquí, se describen los métodos y materiales preferidos. Los intervalos numéricos incluyen los números que definen el intervalo. A menos que se indique lo contrario, los ácidos nucleicos se escriben de izquierda a derecha en la orientación 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en la orientación de amino a carboxi, respectivamente. Los profesionales deben dirigirse a Sambrook et al, *MOLECULAR CLONING: A LABORATORY MANUAL* (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989, y Ausubel FM et al., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, John Wiley & Sons, New York, N.Y., 1993, par alas definiciones y términos de la técnica. Debe entenderse que la presente invención no se limita a una metodología, protocolos y reactivos concretos descritos, ya que éstos pueden variar.

[0045] La presente invención proporciona las secuencias de nucleótidos de los genes de *Trichoderma reesei* implicados en la degradación de la celulosa y la biomasa. Los genes codifican proteínas con una actividad

enzimática que se utiliza en la industria o es de interés para la industria. Las secuencias genómicas de la presente invención que codifican las enzimas se identifican principalmente por comparación de las secuencias de nucleótidos de ADN genómico de *T. reesei* y las secuencias de nucleótidos de genes de enzimas conocidas de otros microorganismos. Antes de la presente invención, no se conocían las secuencias de nucleótidos de estos genes de *T. reesei*, los marcos de lectura, las posiciones de exones e intrones, la estructura de las enzimas, y su utilidad potencial en diversas industrias, tales como las implicadas en la fabricación de comida y pienso, bebidas, tejidos, detergentes. Sin limitarse, los polinucleótidos de los genes de enzimas se pueden utilizar para expresar las enzimas recombinantes para la caracterización, modificaciones o usos industriales; para comparar con la secuencia de ácidos nucleicos de *Trichoderma reesei* para identificar genes duplicados o parálogos que tiene una actividad bioquímica y/o función igual o similar; para comparar con las secuencias de ácidos nucleicos de otros organismos fúngicos relacionados o distantes para identificar potenciales genes de enzimas ortólogas; para seleccionar y fabricar oligómeros para la unión a un conjunto de ácidos nucleicos para el examen de los patrones de expresión; y para desarrollar anticuerpos anti-proteínas utilizando técnicas de inmunización de ácidos nucleicos. La información de secuencia proporcionada aquí también puede formar la base para el diseño y el análisis de enzimas modificadas genéticamente que poseen características químicas y físicas deseables.

[0046] El término "polipéptido" tal como se utiliza aquí, se refiere a un compuesto formado de una cadena única de residuos de aminoácidos unidos por enlaces peptídicos. El término "proteína" tal como se utiliza aquí, se utiliza indistintamente con el término "polipéptido".

[0047] El término "molécula de ácido nucleico" incluye moléculas de ARN, ADN y ADNc. Se entenderá que, como resultado de la degeneración del código genético, se pueden producir una multitud de secuencias de nucleótidos que codifican una proteína determinada.

[0048] Una construcción o secuencia de ácido nucleico "heteróloga" tiene una parte de la secuencia que no es nativa para la célula en la que se expresa. Heteróloga, con respecto a una secuencia de control se refiere a una secuencia de control (es decir, promotor o potenciador) que no actúa en la naturaleza para regular el mismo gen, la expresión del cual se está regulando. En general, las secuencias de ácido nucleico heterólogas no son endógenas a la célula o parte del genoma en la que están presentes, y se han añadido a la célula, mediante infección, transfección, transformación, microinyección, electroporación o similar. Una construcción de ácido nucleico "heteróloga" puede contener una combinación de la secuencia de control/secuencia codificante de ADN que es la misma que, o es diferente de una secuencia de control/secuencia codificante de ADN hallada en la célula nativa.

[0049] Tal como se utiliza aquí, el término "vector" se refiere a una construcción de ácido nucleico diseñado para transferirse entre diferentes células huésped. Un "vector de expresión" se refiere a un vector que tiene la capacidad de incorporar y expresar fragmentos de ADN heterólogos en una célula foránea. Existen muchos vectores de expresión procariotas y eucariotas. La selección de vectores de expresión adecuados se encuentra dentro del conocimiento de los expertos en la materia.

[0050] Por consiguiente, un "cassette de expresión" o "vector de expresión" es una construcción de ácido nucleico generado de manera recombinante o sintéticamente, con una serie de elementos de ácido nucleico específicos que permiten la transcripción de un ácido nucleico concreto en una célula diana. El cassette de expresión recombinante se puede incorporar en un plásmido, cromosoma, ADN mitocondrial, ADN de plástido, virus o fragmento de ácido nucleico. Habitualmente, la parte cassette de la expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico a transcribir y un promotor.

[0051] Tal como se utiliza aquí, el término "plásmido" se refiere a una construcción de ADN de doble cadena (ds) circular como vector de clonación, y que forma un elemento genético auto-replicante extracromosómico en muchas bacterias y algunas eucariotas.

[0052] Tal como se utiliza aquí, el término "secuencia de nucleótidos que codifica el marcador seleccionable" 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 capacidad de crecer en presencia de un agente selectivo correspondiente, o bajo condiciones de crecimiento selectivas correspondientes.

[0053] Tal como se utiliza aquí, el término "promotor" se refiere a una secuencia de ácido nucleico que funciona para dirigir la transcripción de un gen en dirección 3'. El promotor será generalmente apropiado para la célula huésped en el que se expresa el gen diana. El promotor junto con otras secuencias de ácido nucleico reguladoras de transcripción y la traducción (también denominadas "secuencias de control") son necesarios para expresar un gen determinado. En general, las secuencias reguladoras de la transcripción y la traducción incluyen, pero sin limitación,

secuencias de promotor, sitios de unión a ribosoma, secuencias de inicio y parada de la transcripción, secuencias de inicio y para de la traducción, y secuencias potenciadoras o activadoras.

5 [0054] "Gen quimérico" o "construcción de ácido nucleico heteróloga", tal como se define aquí se refiere a un gen no nativo (es decir, uno que se ha introducido en un huésped) que puede estar compuesto de partes de genes diferentes, incluyendo elementos reguladores. Una construcción de gen quimérico para la transformación de una célula huésped está compuesta habitualmente de una región reguladora de la transcripción (promotor) unida operativamente a una secuencia codificante de proteína heteróloga o, en un gen quimérico marcador seleccionable, a un gen marcador seleccionable que codifica una proteína que confiere resistencia a antibiótico a las células transformadas. Un gen quimérico habitual de la presente invención, para la transformación en una célula huésped, incluye una región reguladora de la transcripción que es constitutiva o inducible, una secuencia codificante de proteína y una secuencia terminadora. Una construcción gen quimérico puede incluir también una segunda secuencia de ADN que codifica un péptido señal si se desea la secreción de la proteína diana.

15 [0055] Un ácido nucleico está "unido operativamente" cuando se le coloca en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, el ADN que codifica una secuencia líder secretora está unido operativamente al ADN de un polipéptido si se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o un potenciador está unido operativamente a una secuencia codificante si afecta a la transcripción de la secuencia; o un sitio de unión a ribosomas está unido operativamente a una secuencia codificante si está localizado para facilitar la traducción. Por lo general, "unido operativamente" significa que las secuencias de ADN que se unen son contiguas y, en el caso de una secuencia líder secretora, contiguas y en el mismo marco de lectura. Sin embargo, los potenciadores no tienen que ser contiguos. La unión se logra mediante ligación en los sitios de restricción convenientes. Si dichos sitios no existieran, se utilizarían adaptadores oligonucleotídicos sintéticos, enlazadores o cebadores para PCR y se utilizarían de acuerdo con la práctica convencional.

25 [0056] Tal como se utiliza aquí, el término "gen" significa el segmento de ADN implicado en la producción de una cadena polipeptídica, que puede o no incluir regiones que preceden o siguen a la región codificante, por ejemplo, secuencias 5' no traducidas (5'UTR) o secuencias "líder" y secuencias 3' UTR o "trailer", así como secuencias de intervención (intrones) entre segmentos de codificación individual (exones).

35 [0057] En general, las moléculas de ácido nucleico que codifican una proteína nueva tal como se describe aquí o un análogo u homólogo de las mismas se hibridará, bajo condiciones de moderada a alta astringencia a la secuencia de ácido nucleico correspondiente de la proteína proporcionada aquí. Sin embargo, en algunos casos, se utiliza una secuencia de nucleótidos que codifica la proteína nueva que posee un uso de codón sustancialmente diferente, mientras que la proteína nueva codificada por la secuencia de nucleótidos que codifica la proteína tiene la misma o sustancialmente la misma secuencia de aminoácidos que la proteína nativa. Por ejemplo, la secuencia codificante se puede modificar para facilitar una expresión más rápida de la proteína nueva en un sistema de expresión procarionota o eucariota concreto, según la frecuencia con que un codón concreto es utilizado por el huésped. Te'o, et al. FEMS Microbiology Letters 190:13-19, (2000), por ejemplo, describe la optimización de genes para la expresión en hongos filamentosos.

45 [0058] Una secuencia de ácido nucleico se considera que es "selectivamente hibridable" a una secuencia de ácido nucleico de referencia si las dos secuencias se hibridan específicamente entre sí bajo condiciones de hibridación y lavado de moderada a alta astringencia. Las condiciones de hibridación se basan en la temperatura de fusión (T_f) del complejo o sonda de unión a ácido nucleico. Por ejemplo, la "máxima astringencia" tiene lugar habitualmente a aproximadamente T_f-5°C (5° por debajo de la T_f de la sonda); "astringencia alta" a aproximadamente 5-10° por debajo de T_f; "astringencia intermedia" a aproximadamente 10-20° por debajo de la T_f de la sonda; y "baja astringencia" a aproximadamente 20-25° por debajo de la T_f. Funcionalmente, las condiciones de astringencia máxima se pueden utilizar para identificar secuencias que tiene una identidad estricta o una identidad casi estricta con la sonda de hibridación; aunque las condiciones de alta astringencia se utilizan para identificar secuencias que tienen aproximadamente un 80% o más de identidad en la secuencia con la sonda.

55 [0059] Las condiciones de hibridación moderadas y altas son conocidas en la técnica (véase, por ejemplo, Sambrook, et al, 1989, Capítulos 9 y 11, y en Ausubel, F.M., et al., 1993, incorporados expresamente por referencia aquí). Un ejemplo de condiciones de alta astringencia incluye la hibridación a aproximadamente 42°C en formamida al 50%, 5X SSC, 5X solución de Denhardt, SDS al 0,5% y 100 mg/ml de ADN portador desnaturalizado seguido de dos veces un en 2X SSC y SDS al 0,5% a temperatura ambiente y dos veces adicionales en 0,1 X SSC y SDS al 0,5% a 42°C.

60

- 5 [0060] Tal como se utiliza aquí, "recombinante" incluye la referencia a una célula o vector, que se ha modificado mediante la introducción de una secuencia de ácido nucleico heteróloga o que la célula se deriva de una célula modificada de esta manera. De este modo, por ejemplo, las células recombinantes que expresan genes que no se hallan en la forma idéntica a la forma nativa (no recombinante) de la célula o expresan genes nativos que en cambio se expresan anormalmente, subexpresan o no se expresan como resultado de intervención humana deliberada.
- 10 [0061] Tal como se utiliza aquí, los términos "transformada", "transformada de manera estable" o "transgénica" 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 través de múltiples generaciones.
- 15 [0062] Tal como se utiliza aquí, el término "expresión" se refiere al proceso por el cual se produce un polipéptido basado en la secuencia de ácido nucleico de un gen. El proceso incluye tanto la transcripción como la traducción.
- 20 [0063] El término "introducido" en el contexto de insertar una secuencia de ácido nucleico en una célula, significa "transfección" o "transformación" o "transducción" e incluye la referencia a la incorporación de una secuencia de ácido nucleico en una célula eucariota o procariota, donde la secuencia de ácido nucleico se puede incorporar en el genoma de la célula (por ejemplo, cromosoma, plásmido, plástico o ADN mitocondrial), convertirse en una replicación autónoma o se expresa manera transitoria (por ejemplo, ARNm transfectado).
- 25 [0064] Tal como se utiliza aquí, la frase "proteína nueva" se refiere a por lo menos uno de las cuatro proteínas nuevas descritas aquí, ABF2, AXE2, CIP1 y/o CIP2.
- 30 [0065] Se deduce que el término "expresión de proteína nueva" se refiere a la transcripción y traducción del gen que codifica la proteína nueva, los productos de la cual incluyen ARN precursor, ARNm, polipéptido, polipéptidos procesados después de traducción, y derivados de los mismos, incluyendo las correspondientes proteínas nuevas de especies relacionadas, tales como *Trichoderma longibrachiatum (reesei)*, *Trichoderma viride*, *Trichoderma koningii*, *Hypocrea jecorina* y *Hypocrea schweinitzii*. A modo de ejemplo, los ensayos para la expresión de proteínas nuevas incluyen la transferencia Western para la proteína nueva, el análisis por transferencia Northern y ensayos con la reacción en cadena de la polimerasa de la transcriptasa inversa (RT-PCR) para el ARNm de la proteína nueva.
- 35 [0066] El término "empalme (splicing) alternativo" se refiere al proceso por el cual se generan múltiples isoformas de polipéptidos a partir de un único gen, e implica el empalme conjunto de exones no consecutivos durante el procesamiento de algunos, pero no todos, los transcritos del gen. De este modo, un exón particular se puede conectar a cualquiera de varios exones alternativos para formar ARNs mensajeros. Los ARNm empalmados de forma alternativa producen polipéptidos ("VARIANTES por empalme") en los que algunas partes son comunes, mientras que otras partes son diferentes.
- 40 [0067] El término "secuencia señal" se refiere a una secuencia de aminoácidos en la parte N-terminal de una proteína que facilita la secreción de la forma madura de la proteína fuera de la célula. La forma madura de la proteína extracelular carece de la secuencia señal que se escinde durante el proceso de secreción.
- 45 [0068] Por el término "célula huésped" se entiende que una célula que contiene un vector y admite la replicación y/o la transcripción o la transcripción y traducción (expresión) de la construcción de expresión. Las células huésped para su uso en la presente invención pueden ser células procariotas, tales como *E. coli*, o células eucariotas, tales como células de levadura, plantas, insectos, anfibios o mamíferos. En general, las células huésped son hongos filamentosos.
- 50 [0069] El término "hongo filamentosos" significa cualquier y todos los hongos filamentosos reconocidos por los expertos en la materia. Un hongo preferido se selecciona del grupo que consiste en *Aspergillus*, *Trichoderma*, *Fusarium*, *Chrysosporium*, *Penicillium*, *Humicola*, *Neurospora*, o formas sexuales alternativas de los mismos, tales como *Emericella*, *Hypocrea*.
- 55 [0070] El término "celooligosacárido" se refiere a grupos de oligosacáridos que contienen de 2 a 8 unidades de glucosa y que tienen uniones β -1,4, por ejemplo, celobiosa.
- 60 [0071] El término "celulasa" se refiere a una categoría de enzimas capaces de hidrolizar polímeros de celulosa a oligómeros de celooligosacáridos más cortos, celobiosa y/o glucosa. Se han obtenido numerosos ejemplos de celulasas, tales como exoglucanasas, exocelobiohidrolasas, endoglucanasas y glucosidasas a partir de organismos celololíficos, particularmente incluyendo hongos, plantas y bacterias.

[0072] Los términos "dominio de unión a celulosa" o "CBD" o "módulo de unión a celulosa" o "CBM" tal como se utiliza aquí, se refieren a una parte de la secuencia de aminoácidos de una proteína o una región de la enzima que está implicada en la actividad de unión a celulosa de una enzima celulolítica o derivado de la misma. Un dominio es una parte estable de una proteína con diferentes dominios de proteína que llevan a cabo diferentes funciones. De este modo, un dominio central catalítico (o simplemente el núcleo) contiene el sitio activo y lleva a cabo la reacción enzimática. De manera similar, los dominios de unión a celulosa actúan generalmente mediante unión no covalente de la celulosa a la celulosa, un derivado de la celulosa u otro polisacárido equivalente del mismo. Los dominios de unión a celulosa permiten o facilitan la hidrólisis de fibras de celulosa por la región central catalítica estructuralmente distinta y habitualmente actúan de manera independiente al centro catalítico. De este modo, un dominio de unión a celulosa no poseerá la actividad hidrolítica significativa atribuible a un centro catalítico. En otras palabras, un dominio de unión a celulosa es un elemento estructural de la estructura terciaria de la proteína enzima celulolítica que es diferente del elemento estructural que posee actividad catalítica. Si una proteína posee más de un dominio, los dominios se conectan habitualmente mediante un enlazador.

[0073] Tal como se utiliza aquí, el término "disminución o eliminación de la expresión del gen que codifica una proteína nueva" significa que o el gen que codifica la proteína nueva se ha eliminado del genoma y por tanto no puede ser expresado por el microorganismo huésped recombinante; o que el gen que codifica la proteína nueva se ha modificado, de manera que el microorganismo huésped recombinante no produce una proteína nueva funcional, o el uso de ARNi para dirigir específicamente un producto nuevo del gen, dando lugar a fenotipos nulos o hipomórficos.

[0074] El término "% homología" se utiliza indistintamente aquí con el término "% identidad" y se refiere al nivel de identidad en la secuencia de ácidos nucleicos o aminoácidos entre la secuencia de ácidos nucleicos que codifica cualquiera de los polipéptidos de la invención o la secuencia de aminoácidos del polipéptido de la invención, cuando se alinean utilizando un programan de alineación de secuencias.

[0075] Por ejemplo, tal como se utiliza aquí, el 80% de homología significa lo mismo que un 80% de identidad en la secuencia determinada mediante un algoritmo definido y, por consiguiente, un homólogo de una secuencia determinada tiene más de un 80% de identidad en la secuencia sobre la longitud de la secuencia determinada. Entre los niveles de identidad en la secuencia se incluyen, pero sin limitación, un 80, 85, 90, 95, 98% o más de identidad en la secuencia con una secuencia determinada, por ejemplo, la secuencia codificante para cualquiera de los polipéptidos de la invención, tal como se describe aquí.

[0076] Entre los programas informáticos de ejemplo que se pueden utilizar para determinar la identidad entre dos secuencias se incluyen, pero sin limitación, el conjunto de programas BLAST, por ejemplo, BLASTN, BLASTX, y TBLASTX, BLASTP y TBLASTN, disponibles públicamente en Internet en www.ncbi.nlm.nih.gov/BLAST. Véase también, Altschul, et al., 1990 y Altschul, et al., 1997.

[0077] Las búsquedas de secuencias se llevan a cabo habitualmente utilizando el programa BLASTN cuando se evalúa una secuencia de ácido nucleico determinado en relación a secuencias de ácido nucleico en las secuencias de ADN del Banco de Genes y otras bases de datos públicas. Se prefiere el programa BLASTX para la búsqueda de secuencias de ácido nucleico que se han traducido en todos los marcos de lectura contra las secuencias de aminoácidos en las Secuencias de Proteínas del Banco de Genes y otras bases de datos públicas. Tanto BLASTN y BLASTX se desarrollan utilizando parámetros por defecto de una penalización por espacio abierto de 11,0 y una penalización por espacio extendido de 1,0 y utilizan la matriz BLOSUM-62 (véase por ejemplo, Altschul, S. F., et al., Nucleic Acids Res. 25:3389-3402, 1997.)

[0078] Una alineación preferida de las secuencias seleccionadas con el fin de determinar el "% de identidad" entre dos o más secuencias, se realiza utilizando, por ejemplo, el programa CLUSTAL-W en MacVector version 6.5, operado con los parámetros por defecto, incluyendo una penalización por la apertura del espacio ("open gap penalty") de 10,0 y una penalización por la extensión del espacio ("extended gap penalti") de 0,1, y una matriz de similitud de BLOSUM 30.

[0079] El término "gen alterado" o "gen alterado que codifica la proteína nueva" significa que la secuencia de ácido nucleico del gen ha sido alterada mediante la eliminación, adición y/o manipulación de la secuencia codificante o se ha modificado la secuencia de aminoácidos de la proteína expresada.

[0080] Tal como se utiliza aquí, el término "purificar" se refiere en general a someter células que contienen el ácido nucleico o proteína a purificación bioquímica y/o cromatografía en columna.

[0081] Los términos "aislado" o "purificado" tal como se utiliza aquí, se refiere a un ácido nucleico o proteína que se extrae de por lo menos un componente con el que está asociado de forma natural.

[0082] En el presente contexto, el término "polipéptido sustancialmente puro" significa una preparación de polipéptidos que contiene como máximo un 10% en peso de otro material polipeptídico con el que está asociado de forma nativa (se prefiere porcentajes inferiores del otro material polipeptídico, por ejemplo, como máximo un 8% en peso, como máximo un 6% en peso, como máximo un 5% en peso, como máximo un 4% en peso, como máximo un 3% en peso, como máximo un 2% en peso, como máximo un 1% en peso, y como máximo un 1/2% en peso). De este modo, se prefiere que el polipéptido sustancialmente puro sea por lo menos un 92% puro, es decir, que el polipéptido constituya por lo menos un 92% en peso del material polipeptídico total presente en la preparación, y se prefieren porcentajes superiores, tales como, por lo menos un 94% puro, por lo menos un 95% puro, por lo menos un 96% puro, por lo menos un 96% puro, por lo menos un 97% puro, por lo menos un 98% puro, por lo menos un 99% puro, y como máximo un 99,5% puro. Los polipéptidos descritos aquí están preferiblemente en una forma sustancialmente pura. En particular, se prefiere que los polipéptidos descritos estén en forma sustancialmente pura. En particular, se prefiere que los polipéptidos descritos aquí se encuentren en "forma esencialmente pura", es decir, que la preparación de polipéptidos esté esencialmente libre de otro material polipeptídico con el que está asociado de forma nativa. Esto se puede realizar, por ejemplo, mediante la preparación del polipéptido por medio de métodos recombinantes conocidos. Aquí, el término "polipéptido sustancialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada".

[0083] Tal como se utiliza aquí, los términos "activo" y "biológicamente activo" se refieren a una actividad biológica asociada con una proteína particular, tal como la actividad enzimática asociada con una proteasa. Se deduce que la actividad biológica de una proteína determinada se refiere a cualquier actividad biológica atribuida habitualmente a esa proteína por los expertos en la materia.

[0084] Tal como se utiliza aquí, el término "enriquecida" significa que la proteína nueva se encuentra en una concentración que es superior en relación con la concentración de proteína nueva hallada en una composición de celulosa fúngica de tipo salvaje o natural.

[0085] Cuando se utiliza en soluciones enzimáticas, el componente proteico nuevo se añade en general en una cantidad suficiente a: para las proteínas CIP, aumentan la acción de los componentes CBH y endoglucanasa hallada en la composición de celulosa; para la arabinofuranosidasa y acetilxilanoesterasa, aumentan la acción de una xilanasasa. La cantidad de componente proteico nuevo añadido depende del nivel de acción aumentada deseada proporcionada por la proteína nueva, que se puede determinar fácilmente por el técnico. Sin embargo, cuando se utiliza, el porcentaje en peso del componente proteico nuevo es preferiblemente de aproximadamente 1, preferiblemente de aproximadamente 5, preferiblemente de aproximadamente 10, preferiblemente de aproximadamente 15, o preferiblemente de aproximadamente 20 por ciento en peso a preferiblemente aproximadamente 25, preferiblemente aproximadamente 30, preferiblemente aproximadamente 35, preferiblemente aproximadamente 40, preferiblemente aproximadamente 45 ó preferiblemente aproximadamente 50 por ciento en peso. Además, los intervalos preferidos pueden ser de aproximadamente 0,5 a aproximadamente 15 por ciento en peso, de aproximadamente 0,5 a aproximadamente 20 por ciento en peso, de aproximadamente 1 a aproximadamente 10 por ciento en peso, de aproximadamente 1 a aproximadamente 15 por ciento en peso, de aproximadamente 1 a aproximadamente 20 por ciento en peso, de aproximadamente 1 a aproximadamente 25 por ciento en peso, de aproximadamente 5 a aproximadamente 20 por ciento en peso, de aproximadamente 5 a aproximadamente 25 por ciento en peso, de aproximadamente 5 a aproximadamente 30 por ciento en peso, de aproximadamente 5 a aproximadamente 35 por ciento en peso, de aproximadamente 5 a aproximadamente 40 por ciento en peso, de aproximadamente 5 a aproximadamente 45 por ciento en peso, de aproximadamente 5 a aproximadamente 50 por ciento en peso, de aproximadamente 10 a aproximadamente 20 por ciento en peso, de aproximadamente 10 a aproximadamente 25 por ciento en peso, de aproximadamente 10 a aproximadamente 30 por ciento en peso, de aproximadamente 10 a aproximadamente 35 por ciento en peso, de aproximadamente 10 a aproximadamente 40 por ciento en peso, de aproximadamente 10 a aproximadamente 45 por ciento en peso, de aproximadamente 10 a aproximadamente 50 por ciento en peso, de aproximadamente 15 a aproximadamente 20 por ciento en peso, de aproximadamente 15 a aproximadamente 25 por ciento en peso, de aproximadamente 15 a aproximadamente 30 por ciento en peso, de aproximadamente 15 a aproximadamente 35 por ciento en peso, de aproximadamente 15 a aproximadamente 40 por ciento en peso, de aproximadamente 15 a aproximadamente 45 por ciento en peso, de aproximadamente 15 a aproximadamente 50 por ciento en peso.

Alcance de la invención

[0086] Las cepas de *Trichoderma reesei* utilizadas en este estudio se obtuvieron de la American Type Culture

collection. Sin embargo, debe entenderse que se pueden utilizar otras fuentes microbianas para identificar los correspondientes homólogos de polipéptido. Debe indicarse que el nombre *Hypocrea jecorina* se puede utilizar indistintamente en la presente invención con *Trichoderma reesei*.

5 [0087] Los títulos proporcionados aquí no son limitaciones de los diversos aspectos o realizaciones de la invención que se pueden disponer por referencia a la memoria de manera global. Por consiguiente, los términos definidos inmediatamente a continuación se definen de manera más completa por referencia a la memoria de manera global.

10 I. ORGANISMOS HUÉSPED

[0088] 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 crecimiento vegetativo mediante elongación hifal y catabolismo del carbono que es obligatoriamente aeróbico.

15 [0089] En la presente invención, la célula parental de hongo filamentoso puede ser una célula de una especie de, pero sin limitación, *Trichoderma*, por ejemplo, *Trichoderma longibrachiatum*, *Trichoderma viride*, *Trichoderma koningii*, *Trichoderma harzianum*; *Penicillium sp.*; *Humicola sp.*, incluyendo *Humicola insolens* y *Humicola grisea*; *Chrysosporium sp.*, incluyendo *C. lucknowense*; *Gliocladium sp.*; *Aspergillus sp.*; *Fusarium sp.*, *Neurospora sp.*, *Hypocrea sp.*, y *Emericella sp.* Tal como se utiliza aquí, el término "*Trichoderma*" o "*Trichoderma sp.*" se refiere a cualquier cepa fúngica que se ha clasificado previamente como *Trichoderma* o se clasifica actualmente como *Trichoderma*.

20 [0090] En una realización preferida, la célula parental de hongo filamentoso es una célula *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus tubingensis*, *Aspergillus foetidus*, *Aspergillus oryzae*, *Aspergillus sojae*, *Aspergillus aculeatus*, o *Aspergillus nidulans*.

[0091] En otra realización preferida, la célula parental de hongo filamentoso es una célula *Trichoderma reesei*.

30 II. BIOLOGÍA MOLECULAR

[0092] En una realización, la presente invención proporciona la expresión del gen *cip2* tal como se define en las reivindicaciones bajo el control de un promotor funcional en un hongo filamentoso. Por lo tanto, la presente invención se refiere a técnicas de rutina en el campo de la genética recombinante. Los textos básicos que describen los métodos generales de uso en esta invención incluyen Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); y Ausubel et al., eds., *Current Protocols in Molecular Biology* (1994).

40 A. Métodos de identificación de secuencias nuevas

[0093] Las técnicas que se pueden utilizar para aislar las nuevas secuencias de ADN que codifican proteínas son conocidas en la técnicas e incluyen, pero sin limitación, el cribado de bibliotecas genómicas y/o ADNc con una sonda de ADN homóloga y el cribado de la expresión con ensayos de actividad o anticuerpos contra las proteínas nuevas. Cualquiera de estos métodos se puede hallar en Sambrook, et al. o en *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, F. Ausubel, et al., ed. Greene Publishing and Wiley-Interscience, New York (1987) ("Ausubel").

[0094] Se secuenciaron parcial o totalmente por encima de 5000 ADNc de *T. reesei*. Se descubrieron cuatro ADNc que codifican nuevas enzimas con supuestos papeles en la degradación de la biomasa.

50 [0095] Se analizan los marcos de lectura abiertos (ORFs) siguiendo la secuenciación total o parcial de los clones de las bibliotecas de ADNc derivadas de ARNm de *T. reesei* y se analizan posteriormente utilizando un software de análisis de secuencias y mediante la determinación de la homología a secuencias conocidas en bases de datos (públicas/privadas).

55 [0096] Las secuencias de nucleótidos se anotaron inicialmente mediante programas software, tales como Genescan y Glimmer M (The Institute of Genome Research, Rockville, NM), que pueden identificar posibles regiones codificantes, intrones y uniones por empalme. El tratamiento posterior automatizado y manual de las secuencias de nucleótidos se realizó para refinar y establecer una caracterización exacta de las regiones codificantes y otras características génicas.

B. Construcciones/vectores de expresión de ácidos nucleicos.

- 5 [0097] Se pueden incorporar fragmentos e polinucleótidos naturales o sintéticos que codifican una nueva proteína en construcciones o vectores de ácidos nucleicos, capaces de la introducción y la replicación en una célula de hongo filamentoso o levadura. Los vectores y métodos descritos aquí son adecuados para su uso en células huésped para la expresión de una proteína nueva. Se puede utilizar cualquier vector, siempre y cuando sea replicable y viable en las células en las que se introduce. Se conocen una gran cantidad de vectores y promotores adecuados para los expertos en la materia, y se encuentran disponibles comercialmente. Los vectores de clonación y expresión también se describen en Sambrook et al., 1989, Ausubel FM et al., 1989, y Strathern et al., The Molecular Biology of the Yeast *Saccharomyces* 1981, cada uno se incorpora expresamente por referencia en la presente invención. Los vectores de expresión apropiados para los hongos se describen en van den Hondel, C.A.M.J.J. et al. (1991) en: Bennett, J.W. y Lasure, L.L. (eds.) *More Gene Manipulations in Fungi*. Academic Press, pp. 396-428. La secuencia de ADN apropiada se puede insertar en un plásmido o un vector (referido en general en la presente invención como "vectores") mediante una serie de procedimientos. En general, la secuencia de ADN se inserta en un sitio o sitios para endonucleasas de restricción apropiados mediante procedimientos estándar. Dichos procedimientos y procedimientos de subclonación relacionados se entiende que se encuentran en el alcance del conocimiento los expertos en la materia.
- 10
- 15
- 20 [0098] Los hongos filamentosos recombinantes que comprende la secuencia codificante para una nueva proteína se pueden producir mediante la introducción de una construcción de ácido nucleico heterólogo que comprende la proteína nueva que codifica la secuencia en las células de una cepa seleccionada de los hongos filamentosos.
- 25 [0099] Una vez se obtiene la forma deseada de una secuencia de ácido nucleico para la nueva proteína, se puede modificar de diversas maneras. Cuando la secuencia implica regiones flanqueantes no codificantes, las regiones flanqueantes se pueden someter a resección, mutagénesis, etc. De este modo, se pueden realizar transiciones, transversiones, deleciones e inserciones en la secuencia natural.
- 30 [0100] Se puede insertar una secuencia codificante de una proteína nueva seleccionada en un vector adecuado según técnicas recombinantes conocidas y se pueden utilizar para transformar hongos filamentosos capaces de la expresión heteróloga de la proteína. Debido a la degeneración inherente del código genético, se pueden utilizar otras secuencias de ácidos nucleicos que codifican sustancialmente la misma secuencias de aminoácidos o una funcionalmente equivalente para clonar y expresar una proteína nueva. Por lo tanto, se aprecia que dichas sustituciones en la región codificante se encuentran en las VARIANTES de secuencias cubiertas por la presente invención, siempre que dichas VARIANTES sean tal como se define en las reivindicaciones.
- 35
- 40 [0101] La presente invención también incluye construcciones de ácido nucleico recombinante que comprenden una o más de las secuencias de ácido nucleico que codifican CIP2 tal como se define en las reivindicaciones. Las construcciones comprenden un vector, tal como un plásmido o vector viral, en el que se ha insertado una secuencia de la invención, en una orientación directa o inversa.
- 45 [0102] Las construcciones de ácido nucleico heterólogas pueden incluir la secuencia codificante para una proteína nueva: (i) aislada; (ii) en combinación con secuencias codificantes adicionales, tales como secuencias codificantes de proteínas de fusión o péptido señal, donde la secuencia codificante de la proteína nueva es la secuencia codificante dominante; (iii) en combinación con secuencias no codificantes, tales como intrones y elementos de control, tales como los elementos promotor y terminador o regiones no traducidas 5' y/o 3', eficaces para la expresión de la secuencia codificante en un huésped adecuado, y/o (iv) en un medio vector o huésped en el que la secuencia codificante de la proteína nueva es un gen heterólogo.
- 50 [0103] En un aspecto de la presente invención, se utiliza una construcción de ácido nucleico heterólogo para transferir una secuencia de ácido nucleico que codifica CIP 2 tal como se define en las reivindicaciones en una célula in vitro, siendo preferidas las líneas fúngicas filamentosas y de levadura establecidas. Para la producción a largo plazo de una proteína nueva, se prefiere la expresión estable. Se deduce que se puede utilizar cualquier método eficaz para generar transformantes estables en la práctica de la invención.
- 55
- 60 [0104] Los vectores apropiados están equipados habitualmente con una secuencia de ácido nucleico que codifica un marcador seleccionable, sitios de inserción y elementos de control adecuados, tales como las secuencias de promotor y terminación. El vector puede comprender secuencias reguladoras incluyendo, por ejemplo, secuencias no codificantes, tales como intrones y elementos de control, es decir, elementos promotor y terminador o regiones no traducidas 5' y/o 3', eficaces para la expresión de la secuencia codificante en células huésped (y/o en un medio

vector o célula huésped en el que no se expresa normalmente una secuencia codificante de una proteína antígeno soluble modificada), unida operativamente a la secuencia codificante. Los expertos en la materia conocen una gran cantidad de vectores y promotores adecuados, muchos de los cuales están disponibles comercialmente y/o se describen en Sambrook, et al., (supra).

5 [0105] Entre los promotores de ejemplo se incluyen promotores constitutivos y promotores inducibles, ejemplos de los cuales incluyen un promotor CMV, un promotor temprano para SV40, un promotor RSV, un promotor EF-1 α , un promotor que contiene un elemento de respuesta de tet (TRE) en el sistema de tet-on o tet-off descrito (ClonTech y BASF), y el promotor de beta actina y el promotor de metalotionina que pueden favorecer la expresión mediante la adición de ciertas sales metálicas. Una secuencia de promotor es una secuencia de ADN que es reconocida por el hongo filamentoso particular para fines de expresión. Está unida operativamente a la secuencia de ADN que codifica una proteína nueva. Dicha unión comprende el posicionamiento del promotor con respecto al codón de iniciación de la secuencia de ADN que codifica la proteína nueva en los vectores de expresión descritos. La secuencia de promotor contiene la secuencia de control de la transcripción y la traducción que media en la expresión de la proteína nueva. Entre los ejemplos se incluyen promotores de los genes que codifican glucoamilasa, alfa-amilasa, o alfa-glucosidasa de *Aspergillus niger*, *A. awamori* o *A. oryzae*; los genes *gpdA*, *oliC* o *trpC* de *A. nidulans*; genes *cbh1* o *trp1* de *Neurospora crassa*; los genes que codifican aspártico proteinasa de *A. niger* o *Rhizomucor miehei*; los genes *cbh1*, *cbh2*, *egl1*, *egl2* o que codifican otra celulasa de *T. reesei* (*Hypocrea jecorina*).

20 [0106] La elección del marcador seleccionable correcto dependerá de la célula huésped, y los marcadores apropiados para los diferentes huéspedes son conocidos en la técnica. Los genes marcadores seleccionables habituales incluyen *argB* de *A. nidulans* o *T. reesei* (*H. jecorina*), *amdS* de *A. nidulans*, *pyr4* de *Neurospora crassa* o *H. jecorina*, *pyrG* de *Aspergillus niger* o *A. nidulans*. Marcadores seleccionables de ejemplo adicionales incluyen, pero sin limitación, *trpc*, *trp1*, *oliC31*, *niaD* o *leu2*, que se incluyen en construcciones de ácidos nucleicos heterólogos utilizadas para transformar una cepa mutante, tal como *trp-*, *pyr-*, *leu-* y similares.

30 [0107] Dichos marcadores seleccionables confieren a los transformantes la capacidad de utilizar un metabolito que normalmente no es metabolizado por los hongos filamentosos. Por ejemplo, el gen *amdS* de *H. jecorina* que codifica la enzima acetamidasa que permite que las células transformantes crezcan en acetamida como una fuente de nitrógeno. El marcador seleccionable (por ejemplo, *pyrG*) puede restaurar la capacidad de una cepa mutante autotrófica para crecer en un medio mínimo selectivo o el marcador seleccionable (por ejemplo, *olic31*) puede conferir a los transformantes la capacidad de crecer en presencia de un fármaco inhibidor o antibiótico.

35 [0108] La secuencia codificante del marcador seleccionable se clona en cualquier plásmido adecuado utilizando los métodos utilizados generalmente en la técnica. Entre los plásmidos de ejemplo se incluyen pUC18, pBR322, pRAX y pUC100. El plásmido pRAX contiene secuencias AMA1 de *A. nidulans*, que posibilita la replicación en *A. niger*.

40 [0109] La práctica de la presente invención utilizará, a menos que se indique lo contrario, técnicas convencionales de biología molecular, microbiología, ADN recombinante, e inmunología, que se encuentran en el campo de la invención. Dichas técnicas se explican de manera completa en la literatura. Véase, por ejemplo, Sambrook et al., 1989; Freshney, ANIMAL CELL CULTURE, 1987; Ausubel, et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y., 1993; y Coligan et al., CURRENT PROTOCOLS IN IMMUNOLOGY 1991.

45 C. Métodos para transformar una célula huésped

[0110] En la presente invención, la célula parental de hongo filamentoso puede ser una célula de una especie de, pero sin limitación, *Trichoderma*, por ejemplo, *Trichoderma longibrachiatum* (*reesei*), *Trichoderma viride*, *Trichoderma koningii*, *Trichoderma harzianum*; *Penicillium* sp.; *Humicola* sp., including *Humicola insolens*, *Chrysosporium* sp., incluyendo *C. lucknowense*; *Gliocladium* sp.; *Aspergillus* sp.; *Fusarium* sp., *Neurospora* sp., *Hypocrea* sp., y *Emericella* sp. Tal como se utiliza aquí, el término "*Trichoderma*" o "*Trichoderma* sp." Se refiere a cualquier cepa fúngica que se ha clasificado previamente como *Trichoderma* o se clasifica actualmente como *Trichoderma*.

55 [0111] Entre los ejemplos de líneas de células parentales que se pueden tratar y/o modificar para la expresión de proteínas nuevas se incluyen, pero sin limitación, células de hongos filamentosos. Entre los ejemplos de tipos de células primarias apropiadas para su uso en la práctica de la invención se incluyen, pero sin limitación, *Aspergillus* y *Trichoderma*.

60 [0112] en una realización, la célula parental de hongo filamentoso es una célula de *Aspergillus niger*, *Aspergillus*

awamori, Aspergillus tubingensis, Aspergillus foetidus, Aspergillus oryzae, Aspergillus sojae, Aspergillus aculeatus, o Aspergillus nidulans.

5 [0113] en una segunda realización, la célula parental de hongo filamentoso es una célula de *Hypocrea jecorina*. A esta célula se hizo referencia previamente como *T. reesei*.

10 [0114] Después de clonar las secuencias de ADN que codifican la proteína nueva en construcciones de ADN, se utiliza ADN para transformar microorganismos. El microorganismo a transformar para el objetivo de expresar una proteína nueva según la presente invención puede comprender de manera ventajosa una cepa derivada de *Trichoderma sp.* Por tanto, un modo preferido para preparar una proteína nueva según la presente invención comprende transformar una célula huésped de *Trichoderma sp.* con una construcción de ADN tal como se define en las reivindicaciones. La construcción de ADN se unirá generalmente funcionalmente, es decir, estará unido operativamente a un promotor. A continuación, la célula huésped transformada crece bajo condiciones para expresar la proteína nueva. Posteriormente, la proteína nueva se puede aislar. Puede ser deseable tener la proteína nueva en una forma sustancialmente pura. De forma similar, puede ser deseable tener la proteína nueva en una forma esencialmente pura.

20 [0115] Sin embargo, de hecho puede ser que el mejor vehículo de expresión para un ADN determinado que codifica una proteína nueva pueda diferir de *H. jecorina* (es decir, *T. reesei*). Por tanto, lo más ventajoso puede ser expresar una proteína en un huésped de transformación que porta una similitud filogenética con el organismo fuente para la proteína nueva. En una realización alternativa, se puede utilizar *Aspergillus niger* como vehículo de expresión. Para la descripción de las técnicas de transformación con *A. niger*, véase WO 98/31821.

25 [0116] Por consiguiente, la presente descripción del sistema de expresión de *Trichoderma spp.* se proporciona sólo con fines ilustrativos y como una opción para expresar la proteína nueva de la invención. Un experto en la materia, sin embargo, puede inclinarse a expresar el ADN que codifica la proteína nueva en una célula huésped diferente si es apropiada y debería entenderse que la fuente de la proteína nueva debe considerarse en la determinación del huésped de expresión óptimo. Adicionalmente, el experto en la materia será capaz de seleccionar el mejor sistema de expresión para un gen concreto mediante técnicas de rutina que utilizan herramientas disponibles en la técnica.

30 D. Métodos para expresar una proteína nueva

35 [0117] Los métodos de la presente invención se refieren al uso de células para expresar una proteína nueva sin que se requiera ningún método de expresión en concreto.

40 [0118] La presente invención proporciona células huésped que se han transducido, transformado o transfectado con un vector de expresión que comprende una secuencia de ácido nucleico nueva que codifica CIP 2 tal como se define en las reivindicaciones. Las condiciones de cultivo, tales como la temperatura, pH y similares, son las utilizadas previamente para la célula huésped parental antes de la transducción, transformación o transfección y serán evidentes para el experto en la materia.

45 [0119] En una estrategia, se transfecta una célula de hongo filamentoso o célula de levadura con un vector de expresión que tenga un promotor o fragmento de promotor biológicamente activo o uno o más (por ejemplo, un conjunto) potenciadores que funcionan en la línea de célula huésped, unidos operativamente a un segmento de ADN que codifica una proteína nueva, de manera que la proteína nueva se expresa en la línea celular.

50 [0120] Por tanto, la presente invención proporciona hongos filamentosos que comprenden células que han sido modificadas, seleccionadas y cultivadas de una manera eficaz para dar lugar a la producción o expresión de CIP2 en relación con los correspondientes hongos parentales no transformados.

55 [0121] Entre los ejemplos de especies de hongos filamentosos parentales que se pueden tratar y/o modificar para la expresión de proteína nueva incluyen, pero sin limitación, *Trichoderma, Penicillium sp., Humicola sp., including Humicola insolens; Aspergillus sp., incluyendo Aspergillus niger, Chrysosporium sp., Fusarium sp., Hypocrea sp., y Emericella sp.*

60 [0122] Las células que expresan una proteína nueva se cultivan bajo condiciones utilizadas habitualmente para cultivar la línea fúngica parental. En general, las células se cultivan en un medio estándar que contiene sales fisiológicas y nutrientes, tales como se los describen en Pourquie, J. et al., *Biochemistry and Genetics of Cellulose Degradation*, eds. Aubert, J. P. et al., Academic Press, pp. 71-86, 1988 e Ilmen, M. et al., *Appl. Environ. Microbiol.* 63:1298-1306, 1997. Las condiciones de cultivo también son estándar, por ejemplo, los cultivos se incuban a 28°C

en cultivadores o fermentadores con agitación hasta conseguir los niveles deseados de la expresión de la proteína.

[0123] Las condiciones de cultivo preferidas para un hongo filamentoso determinado se puede encontrar en la literatura científica y/o a partir de una fuente de hongos, tales como la American Type Culture Collection (ATCC; <www.atcc.org>). Después de establecer el crecimiento fúngico, se exponen las células a las condiciones eficaces para causar o permitir la expresión de una proteína nueva.

[0124] En los casos en los que la secuencia que codifica una proteína nueva se encuentra bajo el control de un promotor inducible, el agente inductor, por ejemplo, un azúcar, sal metálica o antibióticos, se añade al medio en una concentración eficaz para inducir la expresión de la proteína nueva.

[0125] En una realización, la cepa comprende *Aspergillus niger*, que es una cepa útil para obtener proteína sobreexpresada. Por ejemplo, se sabe que dgr246 de *A. niger var awamori* secreta cantidades elevadas de celulasas secretadas (Goedegebuur et al, Curr. Genet (2002) 41: 89-98). Se conocen otras cepas de *Aspergillus niger var awamori*, tales como GCDAP3, GCDAP4 y GAP3-4 (Ward, M, Wilson, L.J. and Kodama, K.H., 1993, Appl. Microbiol. Biotechnol. 39:738-743).

[0126] En otra realización, la cepa comprende *Trichoderma reesei*, que es una cepa útil para obtener proteína sobreexpresada. Por ejemplo, se sabe que RL-P37, descrita por Sheir-Neiss, et al., Appl. Microbiol. Biotechnol. 20:46-53 (1984), secreta cantidades elevadas de enzimas celulasa. Los equivalentes funcionales de RL-P37 incluyen la cepa RUT-C30 de *Trichoderma reesei* (ATCC No. 56765) y la cepa QM9414 (ATCC No. 26921). Se contempla que estas cepas también serían útiles en la sobreexpresión de una proteína nueva.

[0127] Cuando se desea obtener la proteína nueva deseada en ausencia de actividad celulolítica nativa potencialmente perjudicial, es útil para obtener una cepa de célula huésped a la que se han eliminado uno o más genes de celulasa antes de la introducción de una construcción de ADN o ADN plasmídico que contiene el fragmento de ADN que codifica la proteína nueva. Dichas cepas se pueden preparar mediante el método descrito en la Patente de Estados Unidos No. 5,246,853 y WO 92/06209. al expresar una proteína nueva en un microorganismo huésped que carece de uno o más genes de celulasa, se simplifican los procedimientos de identificación y purificación posterior. Se puede eliminar cualquier gen de *Trichoderma sp.* que haya sido clonado, por ejemplo, los genes *cbh1*, *cbh2*, *egl1*, y *egl2*, así como aquellos que codifican la proteína EGV (véase, por ejemplo, la Patente de Estados Unidos No. 5,475,101 y WO 94/28117, respectivamente).

[0128] La eliminación de genes se puede llevar a cabo mediante la inserción de una forma del gen deseado a eliminar o alterar en un plásmido mediante métodos conocidos en la técnica. El plásmido de delección se corta a continuación en un sitio o sitios para enzimas de restricción adecuados, interno a la región codificante del gen deseado, y la secuencia codificante del gen o parte de la misma se sustituye por un marcador seleccionable. Las secuencias de ADN flanqueantes del locus del gen a eliminar o alterar, preferiblemente entre aproximadamente 0,5 a 2,0 kb, permanecen en cualquier lado del gen marcador seleccionable. Un plásmido de delección apropiado tendrá en general sitios para enzimas de restricción únicos presentes en el mismo para permitir que el fragmento que contiene el gen eliminado, incluyendo las secuencias de ADN flanqueantes, y el gen marcador seleccionable, se eliminen como una pieza lineal única.

[0129] Debe elegirse un marcador seleccionable para permitir la detección del microorganismo transformado. Cualquier gen marcador seleccionable que se exprese en el microorganismo seleccionado será adecuado. Por ejemplo, con *Aspergillus sp.*, el marcador seleccionable se elige de manera que la presencia del marcador seleccionable en los transformantes no afectará significativamente en las propiedades del mismo. Dicho marcador seleccionable puede ser un gen que codifica un producto analizable. Por ejemplo, se puede utilizar una copia funcional de un gen de *Aspergillus sp.* que si está ausente en la cepa huésped da lugar a la cepa huésped que expresa un fenotipo auxotrófico.

[0130] En una realización, una cepa derivada de *pyrG* de *Aspergillus sp.* se transforma con un gen funcional *pyrG*, el cual, de este modo, proporciona un marcador seleccionable para la transformación. Se puede obtener una cepa derivada de *pyrG* mediante la selección de cepas de *Aspergillus sp.* que son resistentes de ácido fluoroorótico (FOA). El gen *pyrG* codifica orotidin-5'-monofosfato descarboxilasa, una enzima necesaria para la biosíntesis de uridina. Las cepas con un gen *pyrG* intacto crecen en un medio que carece de uridina, pero son sensibles a ácido fluoroorótico. Es posible seleccionar cepas derivadas de *pyrG* que carecen de una enzima funcional orotidin monofosfato descarboxilasa y requieren uridina para el crecimiento mediante selección por resistencia a FOA. Utilizando la técnica de selección con FOA, también es posible obtener cepas que requieren uridina que carecen de una orotato pirofosforibosil transferasa funcional. Es posible transformar estas células con una copia funcional del

gen que codifica esta enzima (Berges & Barreau, Curr. Genet. 19:359-365 (1991), y van Hartingsveldt et al., (1986) Development of a homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. Mol. Gen. Genet. 206:71-75). La selección de cepas derivadas se realiza fácilmente utilizando la técnica de resistencia a FOA referida anteriormente, y de este modo, el gen *pyrG* se utiliza preferiblemente como marcador seleccionable. En otra realización, se transforma una cepa derivada de *pyr4* of *Trichoderma sp.* con un gen funcional *pyr4*, el cual, de este modo, proporciona un marcador seleccionable para la transformación. A continuación se describe para el sistema de *Aspergillus*, se pueden utilizar procedimientos similares para *Trichoderma* y otros sistemas fúngicos tal como entenderá un experto en la materia.

[0131] Para transformar *pyrG* de *Aspergillus sp.* para que carezca de la capacidad de expresar uno o más genes de celulasa, a continuación se aísla un fragmento de ADN sencillo que comprende un gen de celulasa alterado o eliminado del plásmido de delección y se utiliza para transformar un huésped *pyf Aspegillus* apropiado. A continuación, se identifican los transformantes y se seleccionan en base a su capacidad de expresar el producto génico de *pyrG* y, de este modo, complementa la auxotrofia a uridina de la cepa huésped. A continuación, se lleva a cabo un análisis por transferencia Southern en los transformantes resultantes para identificar y confirmar una integración por doble entrecruzamiento que sustituye parte o toda la región codificante de la copia genómica del gen a eliminar por los marcadores seleccionables de *pyr4*.

[0132] Aunque los vectores plasmídicos específicos descritos anteriormente se relacionan con la preparación de transformantes de *pyr*, la presente invención no se limita a estos vectores. Se pueden eliminar y sustituir varios genes en la cepa de *Aspergillus sp.* utilizando las técnicas anteriores. Además, se puede utilizar cualquier marcador seleccionable adecuado, tal como se ha descrito anteriormente. De hecho, cualquier gen de *Aspergillus sp.* que se haya clonado y, por tanto, identificado, se puede eliminar del genoma utilizando la estrategia descrita anteriormente.

[0133] Tal como se ha afirmado anteriormente, las cepas huésped utilizadas son derivados de *Aspergillus sp.* que carecen o tiene un gen no funcional o genes que corresponden con el marcador seleccionable elegido. Por ejemplo, si se elige el marcador seleccionable de *pyrG*, entonces se utiliza una cepa derivada de *pyrG* específica como receptora en el procedimiento de transformación. De manera similar, se pueden utilizar marcadores seleccionables que comprenden genes de *Aspergillus sp.* equivalentes a los genes de *Aspergillus nidulans* *amdS*, *argB*, *trpC*, *niaD*. La cepa receptora correspondiente debe ser por tanto una cepa derivada, tal como *argB-*, *trpC-*, *niaD-*, respectivamente.

[0134] El ADN que codifica la proteína nueva se prepara a continuación para la inserción en un microorganismo apropiado. Tal como se ha descrito aquí, el ADN que codifica una proteína nueva comprende el ADN necesario para codificar una proteína que tiene actividad funcional, por ejemplo, actividad enzimática y/o unión a sustrato. El fragmento de ADN que codifica la proteína nueva se puede unir funcionalmente a una secuencia de promotor fúngico, por ejemplo del promotor del gen *glaA*.

[0135] También se contempla que más de una copia de ADN que codifica una proteína se pueda combinar en la cepa para facilitar la sobreexpresión. El ADN que codifica la proteína nueva se puede preparar mediante la construcción de un vector de expresión que porta el ADN que codifica la proteína nueva. El vector de expresión que porta el fragmento de ADN insertado que codifica la proteína nueva puede ser cualquier vector que es capaz de replicarse autónomamente en un organismo huésped determinado o de integrarse en el ADN del huésped, habitualmente un plásmido. En realizaciones preferidas, se contemplan dos tipos de vectores de expresión para obtener la expresión de genes. El primero contiene secuencias de ADN en las que la secuencia del promotor, la región codificante del gen y el terminador se originan todas del gen a expresar. El truncamiento de genes se puede obtener cuando se desee mediante la eliminación de secuencias de ADN no deseadas (por ejemplo, que codifican dominios no deseados) para dejar que el dominio se exprese bajo el control de sus propias secuencias reguladoras de la transcripción y la traducción. Un marcador seleccionable también está contenido en el vector que permite la selección de la integración en el huésped de múltiples copias de las secuencias génicas nuevas.

[0136] El Segundo tipo de vector de expresión se preensambla y contiene secuencias necesarias para una transcripción a alto nivel y un marcador seleccionable. Se contempla que la región codificante para un gen o parte del mismo se pueda insertar en este vector de expresión de objetivo general, de manera que se encuentre bajo el control transcripcional de las secuencias de promotor y terminador de cassettes de expresión. Por ejemplo, PRAY es dicho vector de expresión de objetivo general. Los genes o partes del mismo se pueden insertar en dirección 3' con respecto al promotor fuerte de *glaA*. Un ejemplo, de un vector de expresión integrador es el vector pTrex. Los genes o partes de los mismos se pueden insertar en dirección 3' con respecto al promotor fuerte de *cbh1*.

[0137] En el vector, la secuencia de ADN que codifica la proteína CIP 2 nueva de la presente invención debería unirse operativamente a las secuencias transcripcionales y traduccionales, es decir, una secuencia de promotor adecuado y secuencia señal, en el marco de lectura al gen estructural. El promotor puede ser cualquier secuencia de ADN que muestra actividad transcripcional en la célula huésped y puede derivar de genes que codifican proteínas homólogas o heterólogas a la célula huésped. Un péptido señal opcional para la producción extracelular de la proteína nueva. El ADN que codifica la secuencia señal es preferiblemente aquel que está asociado de forma natural con el gen a expresar, aunque la presente invención contempla la secuencia señal de cualquier origen.

[0138] Los procedimientos utilizados para fusionar las secuencias de ADN que codifican para la proteína nueva de la presente invención con el promotor en vectores adecuados son conocidos en la técnica.

[0139] Se pueden utilizar varios métodos para liberar un vector de expresión, un vector o construcción de ADN descritos anteriormente en células in vitro. Los métodos de introducción de ácidos nucleicos en células para la expresión de secuencias heterólogas de ácidos nucleicos también son conocidos para el técnico en la materia incluyendo, pero sin limitación, electroporación; microinyección nuclear o microinyección directa en células individuales; fusión de protoplastos bacterianos con células intactas; uso de policones, por ejemplo, polibreno o poliornitina; fusión de membranas con liposomas, lipofectamina o transfección mediada por lipofección; bombardeo de velocidad elevada con microproyectiles recubiertos de ADN; incubación con precipitado de fosfato de calcio-ADN; transfección mediada por DEAE dextrano; infección con ácidos nucleicos virales modificados; transferencia mediada por *Agrobacterium* de ADN; y similares. Además, las construcciones de ácidos nucleicos heterólogas que comprenden una secuencia de ácido nucleico que codifica una proteína nueva se puede transcribir in vitro y el ARN resultante se puede introducir en la célula huésped mediante métodos conocidos, por ejemplo, mediante inyección.

[0140] El método preferido en la presente invención para preparar *Aspergillus sp.* para la transformación implica la preparación de protoplastos de micelio fúngico. Véase Campbell et al. Improved transformation efficiency of *A. niger* using homologous niaD gene for nitrate reductase. Curr. Genet. 16:53-56; 1989. El micelio se puede obtener a partir de esporas vegetativas germinadas. El micelio se trata con una enzima que digiere la pared celular dando lugar a protoplastos. Los protoplastos a continuación se protegen mediante la presencia de un estabilizante osmótico en el medio en suspensión. Estos estabilizantes incluyen sorbitol, manitol, cloruro de potasio, sulfato de magnesio y similares. Normalmente la concentración de estos estabilizantes varía entre 0,8 M y 1,2 M. Es preferible utilizar aproximadamente una solución de sorbitol 1,2 M en el medio de suspensión.

[0141] La captación del ADN en la cepa de *Aspergillus sp.* huésped depende de la concentración de ion calcio. En general se utiliza ente aproximadamente 10 mM de CaCl_2 y 50 mM de CaCl_2 en una solución de captación. Antes de la necesidad de ion calcio en la solución de captación, otros elementos generalmente incluidos son un sistema tamponador, tal como un tampón TE (10 mM Tris, pH 7,4; EDTA 1 mM) o 10 mM de MOPS, tampón a pH 6,0 (ácido morfolinopropanosulfónico) y polietilenglicol (PEG). Se cree que el polietilenglicol actúa fusionando las membranas celulares, permitiendo así que el contenido del medio se libere en el citoplasma de la cepa de *Aspergillus sp.* y el ADN plasmídico se transfiere al núcleo. Esta fusión deja frecuentemente copias múltiples del ADN plasmídico integrado de forma sensible en el cromosoma huésped.

[0142] Normalmente se utiliza una suspensión que contiene los protoplastos o células de *Aspergillus sp.* que se han sometido a un tratamiento de permeabilidad a una densidad de 10^5 a 10^8 /mL, preferiblemente 2×10^5 /mL en la transformación. Se mezcla un volumen de 100 mL de estos protoplastos o células en una solución apropiada (por ejemplo, sorbitol 1,2 M; CaCl_2 50 mM) con el ADN deseado. En general, se añade una concentración elevada de PEG a la solución de captación. Se pueden añadir de 0,1 a 1 volumen de PEG4000 al 25% a la suspensión de protoplastos. Sin embargo, es preferible añadir aproximadamente 0,25 volúmenes a la suspensión de protoplastos. También se pueden añadir aditivos, tales como dimetil sulfóxido, heparina, espermidina, cloruro potásico y similares a la solución de captación y ayudar en la transformación. Existen procedimientos similares para otras células huésped fúngicas. Véase, por ejemplo, la Patente de Estados Unidos No. 6,268,328.

[0143] En general, la mezcla se incuba a continuación a aproximadamente 0°C durante un periodo entre 10 y 30 minutos. A continuación se añade PEG adicional a la mezcla para aumentar adicionalmente la captación del gen o secuencia de ADN deseados. SE añade en general PEG 4000 al 25% en volúmenes de 5 a 15 veces el volumen de la mezcla de transformación; sin embargo, pueden ser adecuados volúmenes mayores y menores. El PEG 4000 al 25% es preferiblemente aproximadamente 10 veces el volumen de la mezcla de transformación. Después de añadir el PEG, la mezcla de transformación se incuba a continuación a temperatura ambiente o sobre hielo antes de la adición de una solución de sorbitol y CaCl_2 . La suspensión de protoplastos se añade a continuación a alícuotas fundidas de un medio de crecimiento. Este medio de crecimiento permite el crecimiento de únicamente transformantes. Se puede utilizar cualquier medio de crecimiento en la presente invención que sea adecuado para el

desarrollo de los transformantes deseados. Sin embargo, si se seleccionan transformantes Pyr⁺ es preferible utilizar un medio de crecimiento que no contenga uridina. Las posteriores colonias se transfieren y purifican en un medio de crecimiento agotado en uridina.

5 [0144] En esta etapa, se pueden diferenciar transformantes estables de transformantes inestables por su velocidad de crecimiento más rápida y la formación de colonias circulares con un perfil más uniforme que desigual en el medio de cultivo sólido que carece de uridina. Adicionalmente, en algunos casos, se puede realizar un test adicional de estabilidad mediante el crecimiento de los transformantes en un medio sólido no selectivo (es decir, que contiene uridina), la recogida de esporas a partir de este medio de cultivo y la determinación del porcentaje de estas esporas
10 que posteriormente germinarán y crecerán en un medio selectivo que carece de uridina. Alternativamente, se pueden utilizar otros métodos conocidos en la técnica para seleccionar transformantes.

[0145] En una realización particular del método anterior, la proteína nueva se recupera en la forma activa de la célula huésped después del crecimiento en medio líquido como resultado del procesamiento apropiado después de la traducción de la proteína nueva.
15

E. Métodos de análisis de las secuencias codificantes de ácidos nucleicos para la proteína nueva y/o expresión de la proteína.

20 [014] Con el fin de evaluar la expresión de una proteína nueva mediante una línea celular que se ha transformado con una construcción de ácidos nucleicos que codifica una proteína nueva, se pueden llevar a cabo análisis a nivel de proteína, el nivel de ARN o mediante el uso de bioensayos funcionales particulares a la actividad y/o producción de la proteína nueva.

25 [0147] En una aplicación de ejemplo de las secuencias de ácido nucleico de la proteína nueva y las secuencias de proteína nueva descritas aquí, se diseña una cepa modificada genéticamente de hongos filamentosos, por ejemplo, *Trichoderma reesei*, para producir una mayor cantidad de proteína nueva. Dichos hongos filamentosos modificados genéticamente serían útiles para producir un producto de celulasa o hemicelulasa con una mayor capacidad
30 celulolítica o hemocelulolítica. En una estrategia, esto se consigue mediante la introducción de la secuencia codificante para una proteína nueva en un huésped adecuado, por ejemplo, un hongo filamentoso, tal como *Aspergillus niger*.

[0148] Por consiguiente, la presente invención incluye métodos para expresar una proteína CIP2 nueva tal como se define en las reivindicaciones en un hongo filamentosos u otro huésped adecuado mediante la introducción de un vector de expresión que contiene la secuencia de ADN que codifica una proteína nueva en células del hongo filamentosos u otro huésped adecuado.
35

[0149] En otro aspecto, la presente invención incluye métodos para modificar la expresión de una proteína CIP2 nueva tal como se define en las reivindicaciones en un hongo filamentosos u otro huésped adecuado. Dichas modificación incluye un descenso o eliminación en la expresión de la proteína CIP2 endógena.
40

[0150] En general, los análisis utilizados para analizar la expresión de una proteína nueva incluyen, transferencia Northern, transferencia de puntos (análisis de ADN o ARN), RT-PCR (reacción en cadena de la polimerasa con transcriptasa inversa), o hibridación in situ, utilizando una sonda marcada de manera apropiada (en base a la secuencia codificante de ácido nucleico) y transferencia Southern convencional y autorradiografía.
45

[0151] La producción y/o expresión de una proteína nueva se puede medir directamente en una muestra de lisado celular o sobrenadante de cultivo mediante electroforesis en gel de poliacrilamida con dodecil sulfato sódico utilizando métodos conocidos en la técnica. Después de la separación de las proteínas en muestra mediante electroforesis y tinción con un colorante adecuado (por ejemplo azul brillante de Coomassie), se demostraría la producción de una proteína nueva mediante la aparición de una banda de proteína nueva. Además, la producción y/o expresión de una proteína nueva se pueden medir directamente en una muestra, por ejemplo, mediante ensayos para la actividad enzimática, expresión y/o producción.
50

[0152] Además, la expresión de proteína, se puede evaluar mediante métodos inmunológicos, tales como tinción inmunohistoquímica de células, secciones de tejido o inmunoensayo de medio de cultivo celular, por ejemplo, mediante transferencia Western o ELISA. Dichos inmunoensayos se pueden utilizar para evaluar cualitativa y cuantitativamente la expresión de una proteína nueva. Los detalles de dichos métodos son conocidos por el experto en la materia y muchos reactivos para realizar dichos métodos están disponibles comercialmente.
55
60

[0153] Se puede utilizar una forma purificada de una proteína nueva para producir anticuerpos monoclonales o policlonales específicos para la proteína expresada para su uso en diversos inmunoensayos. (Véase, por ejemplo, Huet al., Mol Cell Biol. 11:5792-9, 1991). Entre los ensayos de ejemplo se incluyen ELISA, inmunoensayos competitivos, radioinmunoensayos, transferencia Western, ensayos de inmunofluorescencia indirectos, y similares.

5

F. Métodos para purificar una proteína nueva

[0154] En general, una proteína nueva producida en un cultivo celular se secreta en el medio, y se puede purificar o aislar, por ejemplo, mediante la eliminación de componentes no deseados del medio de cultivo celular. Sin embargo, en algunos casos, se pueden producir una proteína nueva en una forma celular que necesita la recuperación a partir de un lisado celular. En dichos casos, la proteína nueva se purifica a partir de las células en las que se produjo utilizando técnicas utilizadas habitualmente por los expertos en la materia. Entre los ejemplos se incluyen, pero sin limitación, cromatografía por afinidad (Tilbeurgh et al., FEBS Lett. 16:215, 1984), métodos cromatográficos de intercambio iónico (Goyal et al., Bioresource Technol. 36:37, 1991; Fliess et al., Eur. J. Appl. Microbiol. Biotechnol. 17:314, 1983; Bhikhabhai et al., J. Appl. Biochem. 6:336, 1984; Ellouz et al., J. Chromatography 396:307, 1987), incluyendo intercambio iónico utilizando materiales con un alto poder de resolución (Medve et al., J. Chromatography A 808:153, 1998), cromatografía por interacción hidrofóbica (Tomaz and Queiroz, J. Chromatography A 865:123-128, 1999), y separación de dos fases (Brumbauer, et al., Bioseparation 7:287-295, 1999).

10

15

20

[0155] Habitualmente, la proteína nueva se fracciona para segregar proteínas que tienen propiedades seleccionadas, tales como una afinidad de unión con agentes de unión concretos, por ejemplo, anticuerpos o receptores; o que tiene un rango de peso molecular o de puntos isoeléctricos seleccionados.

25

30

[0156] Una vez se consigue la expresión de una proteína nueva determinada, la proteína nueva producida de este modo se purifica a partir de las células o el cultivo celular. Entre los procedimientos de ejemplo adecuados para dicha purificación se incluyen los siguientes: cromatografía en columna de afinidad con anticuerpos, cromatografía de intercambio iónico; precipitación con etanol; HPLC de fase inversa; cromatografía sobre sílice o en una resina de intercambio catiónico, tal como DEAE; "chromatofocusing"; SDS-PAGE; precipitación con sulfato de amonio; y filtración en gel utilizando, por ejemplo, Sephadex G-75. Se pueden utilizar diversos métodos de purificación de proteínas y dichos métodos son conocidos en la técnica y se describen, por ejemplo, en Deutscher, Methods Enzymol. 182:779-80, 1990; Scopes, Methods Enzymol. 90 Pt E:479-90, 1982. La etapa o etapas de purificación seleccionadas de penderán, por ejemplo, de la naturaleza del proceso de producción utilizado y la proteína particular producida.

35

III. Características bioquímicas de la proteína nueva

CIP2

[0157] La proteína CIP2 se prevé que tiene 460 aminoácidos y un peso molecular de aproximadamente 48 kDaltons. La proteína prevista está compuesta de 24 aminoácidos fuertemente básicos (+) (K,R), 24 aminoácidos fuertemente ácidos (-) (D,E), 160 aminoácidos hidrofóbicos (A,I,L,F,W,V), y 165 aminoácidos polares (N,C,Q,S,T,Y). Se prevé que CIP2 tiene un punto isoeléctrico de 7,1 y una carga de 0,27 a pH 7,0.

40

45

[0158] CIP2 tiene una secuencia señal N-terminal prevista de 17 aminoácidos seguido de 36 aminoácidos que comprenden un módulo de unión a carbohidrato de la familia CBM1 y una región adaptadora que termina en aproximadamente el aminoácido 95. Véanse las figuras 5 y 6.

IV. Utilidad de los genes identificados de interés

50

[0159] Los genes de la invención se pueden utilizar en un conjunto de aplicaciones diferentes.

CIP1 y CIP 2 (proteína 1 y 2 inducidas por celulosa)

[0160] El CBD que contiene CIP1 codificado por las SEC ID NOS: 1 y/o 2 (véase las figuras 1 y 2) y CIP2 codificado por la SEC ID NO:6 (véase la Figura 4) puede tener utilidad (sinérgica) en la conversión de biomasa o en otras aplicaciones para las que los genes que contienen CBD son apropiados. De este modo, el producto génico puede ser útil en aplicaciones, tales como detergentes, tejidos, conversión de biomasa, aplicaciones en alimentos y piensos, y las industrias de la pulpa y el papel.

55

60

[0161] Las siguientes preparaciones y ejemplos se ofrecen para permitir a los expertos en la materia entender más

claramente y realizar la presente invención. No deberían considerarse como limitantes del alcance de la invención, sino que simplemente son ilustrativos y representativos de la misma.

[0162] En el desarrollo experimental que se acompaña, se aplican las siguientes abreviaturas: eq (equivalentes); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (milimoles); μ mol (micromoles); nmol (nanomoles); g (gramos); mg (miligramos); kg (kilogramos); μ g (microgramos); L (litros); ml (mililitros); μ l (microlitros); cm (centímetros); mm(milímetros); μ m (micrometros); nm (nanómetros); °C. (grados centígrados); h (horas); min (minutos); s (segundos); ms (milisegundos); Ci (Curies) mCi (miliCuries); μ Ci (microCuries); TLC (cromatografía en capa fina); Ts (tosilo); Bn (bencilo); Ph (fenilo); Ms (mesilo); Et (etilo), Me (metilo).

EJEMPLOS

[0163] Los siguientes ejemplos se ofrecen para ilustrar, pero sin limitar, la presente invención.

Ejemplo 1

Construcción de bibliotecas de ADNc de *T. Reesei*

[0164] Se desarrolló *T. reesei* (ATCC 13631) bajo condiciones diferentes para generar micelio que expresa perfiles de ARNm específicos de las condiciones de crecimiento. A continuación, se aisló el ARN, se agrupó y se generaron bibliotecas de ADNc.

1A. Crecimiento de micelios de *T. Reesei*

[0165] Todos los cultivos se desarrollaron en un medio líquido de extracto de levadura/glucosa (YEG) durante la noche a 28°C. Se transfirieron a continuación en las siguientes condiciones y se cultivaron durante el tiempo indicado a 28°C, a menos que se indique lo contrario:

Experimento 1:

[0166]

- A. Solución de Vogel + avicel al 2%, 3 días y 6 días
- B. Solución de Vogel + solkafloc al 2%, 3 días y 6 días
- C. Solución de Vogel + salvado de trigo al 2%, 6 días
- D. Solución de Vogel + pulpa de remolacha al 2%, 6 días
- E. Cultivo en estado sólido en salvado de trigo (15 g de salvado de trigo, 1 g Proflo, 1 g solkafloc, 30 ml agua), 7 días
- F. cultivo en estado sólido en pulpa de remolacha (15 g de pulpa de remolacha, 1 g Proflo, 1 g solkafloc, 30 ml agua), 9 días

Experimento 2:

[0167]

- A. Solución de Vogel + glucosa al 2%, 24h
- B. Solución de Vogel + lactosa al 2%, 24h
- C. Solución de Vogel + xilosa al 2%, 24h
- D. Solución de Vogel + fructosa al 2%, 24h
- E. Solución de Vogel + maltosa al 2%, 24h
- F. Solución de Vogel sin ningún carbono añadido, 24h
- G. Solución de Vogel sin ningún nitrógeno añadido, 24h
- H. Solución de Vogel + salvado de trigo al 2%, 3 días
- I. Solución de Vogel + salvado de trigo al 2%, 6 días
- J. Solución de Vogel + 2% solkafloc, 3 días
- K. Solución de Vogel + 2% solkafloc, 6 días
- L. Solución de Vogel + 2% avicel, 3 días
- M. Solución de Vogel + 2% avicel, 6 días
- N. Vogels + celulosa hinchada con fosfórico al 2%, 3 días
- O. Estado sólido (15 g salvado de trigo, 1 g Proflo, 1 g solkafloc, 30 mL agua), 6 días
- P. YEG, 42°C durante 1,5h (choque térmico)

Q. YEG, 20 mM DTT durante 1,5h (estrés redox)

R. YEG, no agitado en recipiente cerrado durante 1,5h a temperatura ambiente (anoxia)

Preparación de medios

5	[0168]	Medio extracto de levadura/glucosa	- 1 litro	
		1. dH ₂ O		1000 ml
10		2. Extracto de levadura	5 g	
		3. Glucosa		20 g
	[0169]	Solución de Vogel		- 1 litro
15		1. 50X Solución madre de Vogel	25 ml	
		2. dH ₂ O		975 ml
		3. Autoclave		
	[0170]	50X solución madre de Vogel		- 1 litro
20		1. citrato sódico		150 g
		2. KH ₂ PO ₄		250 g
		3. NH ₄ NO ₃	100 g	
		4. MgSO ₄ *7H ₂ O		10 g
25		5. CaCl ₂ *2H ₂ O		5 g
		6. Solución de elementos traza		5 ml
		7. Solución de Biotina		2,5 ml
		8. en dH ₂ O, se lleva a un volumen final de 1 litro		
30	[0171]	Solución de elementos traza		- 1 litro
		1. ácido cítrico		50 g
		2. ZnSO ₄ *7H ₂ O		50 g
		3. Fe(NH ₄) ₂ SO ₄ *6H ₂ O		10 g
35		4. CuSO ₄ *5H ₂ O		2,5 g
		5. MnSO ₄ *4H ₂ O		0,5 g
		6. H ₃ BO ₃		0,5 g
		7. NaMoO ₄ *2H ₂ O		0,5 g
		8. en dH ₂ O, se lleva a un volumen final de 1 litro		
40	[0172]	Solución de Biotina		- 1 litro
		1. d-Biotina	0,1 g	
		2. en dH ₂ O, se lleva a un volumen final de 1 litro		

45 1B. Aislamiento de ARN

[0173] Se aisló el ARN total utilizando el Reactivo TRIZOL® de Life Technologies™ (No. Catálogo 15596-026) y una ligera modificación de su protocolo de aislamiento de ARN que se acompaña (incorporado aquí en su totalidad). A menos que se afirme lo contrario, se llevó a cabo el procedimiento entre 15 y 30°C.

[0174] Se filtró el micelio de *T. reesei* de los diferentes cultivos descritos en 1A para extraer el líquido en exceso y se congeló en nitrógeno líquido. El micelio congelado se molió en un mortero y una mano de mortero y se añadió a reactivo TRIZOL (aproximadamente 9 ml por 1 ml de micelio molido). A continuación, el homogenato se centrifugó a 12,000 x g durante 10 minutos entre 2 y 8°C. La solución de homogenato purificado (sobrenadante) se transfirió a un tubo nuevo.

[0175] Las muestras homogeneizadas se incubaron durante 5 minutos entre 15 y 30°C para permitir la disociación completa de los complejos de nucleoproteínas. A continuación, se añadieron 0,2 mL de cloroformo por 1 mL de Reactivo TRIZOL y os tubos de muestra se taparon por seguridad. Los tubos se agitaron vigorosamente a mano durante 15 segundos, a continuación se incubaron entre 15 y 30°C durante 2 a 3 minutos. A continuación, las

muestras se centrifugaron a no más de 12,000 x g durante 15 minutos entre 2 y 8°C. Después de la centrifugación, la mezcla se separa en una fase inferior roja de fenol-cloroformo, una interfase, una fase acuosa superior incolora. La fase acuosa (aproximadamente un 60% del volumen de reactivo) se transfirió a continuación a un tubo nuevo.

5 [0176] El ARN de la fase acuosa se precipitó mediante la adición de 0,25 mL de isopropanol seguido de 0,25 mL de una solución de precipitación elevada en sal (citrato sódico 0,8 M y NaCl 1,2 M) por 1 ml de Reactivo TRIZOL utilizado para la homogenización. La solución resultante se mezcló y las muestras se incubaron entre 15 y 30°C durante 10 minutos, a continuación se centrifugaron a no más de 12,000 x g durante 10 minutos entre 2 y 8°C.

10 [0177] Se extrajo el sobrenadante y se lavó el residuo de ARN de tipo gel una vez con etanol al 75% (fabricado con agua libre de ARNasa), utilizando por lo menos 1 mL de etanol al 75% por 1 mL de Reactivo TRIZOL utilizado para la homogenización inicial. La muestra se mezcló a continuación mediante centrifugación a no más de 7.500 xg durante 5 minutos entre 2 y 8°C.

15 [0178] El sobrenadante se extrajo de nuevo y se secó brevemente el residuo de ARN (secado al aire o al vacío durante 5-10 minutos). El ARN se disolvió en agua libre de ARNasa mediante el paso de la solución varias veces a través de una punta de pipeta y a continuación la incubación durante 10 minutos entre 55 y 60°C.

20 [0179] La pureza del ARN aislado se comprobó mediante electroforesis en gel.

20 1C. Construcción de bibliotecas de ADNc

25 [0180] Se agruparon volúmenes iguales de ARN obtenido de cada una de las condiciones de crecimiento descritas para el Experimento 1 en 1A y se remitieron un total de 2 mg a Life Technologies (Rockville, MD; ahora Invitrogen) para la construcción de una biblioteca de ADNc. De forma similar, el ARN del Experimento 2 en 1A se agrupó y se remitió a Life Technologies para la construcción de ADNc. Las bibliotecas de ADNc se produjeron utilizando procedimientos estándar en la técnica. A continuación, se indica un resumen de las etapas realizadas.

30 [0181] Se aisló Poli-A ARN del ARN total mediante cromatografía. El ARN total se desarrolló en una columna de celulosa con oligo(dT), y posteriormente se eluyó el poli-A ARN (ARNm).

35 [0182] A partir del ARNm, se generaron ADNc por Life Technologies (Rockville, MD) utilizando el Sistema de Síntesis de ADNc de Life Technologies™ (el manual de instrucciones para el cual se incorpora en la presente invención en su totalidad). A continuación, se resumen los procedimientos a utilizar.

35 Síntesis de la primera cadena

40 [0183] Los componentes de reacción para la producción de una primera cadena de ADNc del ARNm de *T. reesei* aislado se combinan en un tubo de microcentrifuga de 1,5 ml en hielo. La mezcla de reacción, en un volumen de 50 µl, contiene los siguientes componentes:

- 50 mM Tris-HCl (pH 8,3)

- 75 mM KCl

- 3mM MgCl₂

- 10m MDTT

45 - 500 µM de cada dATP, dCTP, dGTP y dTTP

- 50 µg/ml oligo(dT)₁₂₋₁₈

- 100 µg/ml poli (A) ARN (de *T. reesei*)

- 10,000 unidades/ml de la transcriptasa inversa del Virus de la Leucemia Murina Moloney (M-MLV).

50 [01824] Se añade al final transcriptasa inversa, con mezclado, para iniciar la reacción. Opcionalmente, se extrae inmediatamente una alícuota de 10 µl y se transfiere a un tubo separado que contiene 1 µCi de trazador [α -³²P]dCTP. A continuación, ambos tubos se incuban a 37°C durante 1 hora. Los tubos se vuelven a colocar en hielo después de la incubación y la reacción se termina mediante la adición de 1 µl de Na₂EDTA 0,25 M (pH 7,5). La mezcla de reacción de 40 µl se utiliza para la síntesis del ADNc de la segunda cadena.

55 [0185] Si se produce, la mezcla de trazador se diluye con 89 µl de agua y alícuotas de 5 µl por duplicado se depositan sobre filtros (por ejemplo, filtros de fibra de vidrio). El segundo filtro se lava tres veces (secuencialmente), 5 minutos cada uno, con aproximadamente 50 µl por lavado de TCA enfriado en hielo. A continuación, se lava el segundo filtro con 50 ml de etanol al 95% durante aproximadamente 5 minutos a temperatura ambiente, a continuación se seca. Los dos filtros se cuentan por centelleo estándar para determinar la cantidad de ³²P en la

60

mezcla (del primer filtro) y la cantidad de ^{32}P incorporado en el ADNc de la primera cadena (del segundo filtro) para determinar el rendimiento de ADNc de la primera cadena.

5 [0186] El resto de la mezcla de trazador se extrae con fenol y se precipita con etanol. El residuo se aísla y se realiza una electroforesis en gel utilizando gel de agarosa alcalino para determinar el tamaño de productos de cadena única.

Síntesis de la segunda cadena

10 [0187] El ADNc de doble cadena se puede producir utilizando un procedimiento ajustado a la producción de ADNc al que se añadirán enlazadores.

15 [0188] En un tubo de microcentrífuga en hielo, se añaden los componentes a 40 μl del producto de reacción de la primera cadena para producir 300 μl de una mezcla de reacción de la segunda cadena. Los componentes se añaden en el siguiente orden: agua tratada con DEPC, mezcla de dNTPs, solución concentrada de tampón/sal, ADN polimerasa I de *E. coli*, ARNasa H de *E. coli* y ADN ligasa de *E. coli*. La mezcla de reacción final tiene la siguiente composición, además de los componentes originales en el producto de reacción de la primera cadena:

- Tris-HCl 25 mM (pH 8,3)
- KCl 100 mM
- $(\text{NH}_4)_2\text{SO}_4$ 10 mM
- 20 - MgCl_2 5 mM
- dATP, dCTP (incluyendo 10 mCi de [^{32}P] dCTP), dGTP, aproximadamente 250 mM
- NAD 0,15 mM
- DTT 5 mM
- ADN polimerasa I 250 U/ml
- 25 - ARNasa H 8,5 U/ml
- ADN ligasa 30 U/ml

30 [0189] El tubo se centrifuga suavemente para mezclar y se incuba a 16° C durante 2 horas. A continuación, se coloca el tubo en hielo y se añaden 25 ml de Na_2EDTA (pH 7,5).

35 [0190] Se añaden 10 ml de la mezcla a 90 ml de agua. Se dispone una alícuota de 5 ml sobre un primer filtro de fibra de vidrio y se seca. Se disponen otros 10 ml de la mezcla no diluida sobre un segundo filtro de fibra de vidrio, que se lava tres veces, 5 minutos cada vez, con 50 ml de TCA enfriado en hielo para cada lavado. A continuación, se lava el segundo filtro una vez a temperatura ambiente con etanol al 95% durante 5 minutos. Se recuentan los filtros en un centelleador estándar, el primero se utiliza para determinar la cantidad de ^{32}P en la mezcla (actividad específica) y la segunda se utiliza para determinar la cantidad de ^{32}P incorporado en el ADNc.

40 [0191] El resto de la mezcla de reacción se extrae en fenol y se precipita en etanol. El residuo se disuelve a continuación en 200 μl de tampón TE estéril (Tris-HCl 10 mM (pH 7,5), Na_2EDTA 1 mM), al que se añaden 100 μl de acetato de amonio 7,5 M, seguido de 500 μl de etanol para la precipitación. El residuo se seca, se disuelve a continuación en 20 μl de tampón TE estéril, se extraen 2 μl y se analizan mediante electroforesis en gel de agarosa alcalino. Se añaden enlazadores o adaptadores al resto para su incorporación en un vector.

45 [0192] Para la adición de enlazadores, el ADNc se metila primero con una metilasa específica para el enlazador a utilizar para proteger los sitios de restricción internos. Los extremos de ADNc se reparan con ADN polimerasa de T4 y, a continuación, se añaden enlazadores mediante unión por los extremos romos. Los enlazadores deberían disponerse a una concentración elevada para una adición eficaz. El ADNc se digiere con la endonucleasa o endonucleasas de restricción seleccionadas, a continuación se purifica a partir de los productos de digestión pequeños (por ejemplo, mediante cromatografía en columna). El vector se digiere con la misma endonucleasa o endonucleasas de restricción y se combina con el ADNc que, a continuación, se une en el vector como un inserto.

50 [0193] Los enlazadores o adaptadores añadidos a los sitios de endonucleasa de restricción contenidos en ADNc, de manera que un sitio *Sall* se halla en 5' con respecto a la secuencia de ADNc correspondiente al ARNm original y un sitio *NotI* 3' con respecto a la secuencia de ADNc correspondiente al ARNm original. A continuación, se insertan los ADNc en un vector transportador ("shuttle") pREP3Y. El vector pREP3Y es un vector pREP3X (ATTC número 87603) modificado, donde el vector se digirió con endonucleasa *Bam*HI y se insertó un oligonucleótido sintético para añadir sitios de restricción. El vector resultante tiene los siguientes sitios de restricción en la región de polienlazadores: *Xho*I, *Sall*, *Xba*I, *Bam*HI, *Sma*I, *Not*I y *Sma*I. El vector y el ADNc se digirieron en *Sall*/*Not*I y el ADNc se insertó en el vector. Véase la figura 15.

60

[0194] Se construyeron bibliotecas de ADNc por Invitrogen Life Technologies en el vector pREP3Y. La cepa DH12S de *E. coli* se transformó con el vector para crear la biblioteca de ADNc. Se devolvió una biblioteca no amplificada a los inventores. A continuación, se emplacaron los clones individuales y se desarrollaron en LA más carbenicilina 50 mg/ml (bacto triptona 20 g/l; extracto de levadura de Bacto 10 g/l; NaCl 1 g/l; agar de Bacto 17,5 g/l; 1 ml/l de 50 mg/ml de carbenicilina añadida después de la esterilización en autoclave y el enfriamiento, pero antes de la solidificación).

Ejemplo Comparativo 2.

10 Identificación de LT1-24 y el gen *cip1*

[0195] Se utilizó el siguiente razonamiento para encontrar el gen *cip1*:

1) Desarrollar las bibliotecas en membranas Hybond+ de Amersham; 2) Lisar las células y fijar el ADN de las bibliotecas en la membrana; 3) Hibridar la transferencia ("blot") con las sondas específicas de genes; 4) Hibridar la transferencia ("blot") una segunda vez, pero ahora con una sonda de CBM mixta; 5) Extraer el gen específico de las marcas de CBM; y Seleccionar y analizar las nuevas marcas.

Aislamiento de colonias

20 [0196] La biblioteca de ADNc inducida por celulosa de *T. reesei* se utilizó para estos experimentos de hibridación. La biblioteca de ADNc de *E. coli* se emplacó en placas de agar (20 x 20 cm) para obtener una cantidad suficiente de clones.

25 [0197] Las bibliotecas de ADNc se emplacaron en 200 ml 2xTY (bactotriptona 16 g/l, extracto de levadura 10 g/l, NaCl 5 g/l) solidificado con agar al 1,5% en presencia de ampicilina (AMP) 100 µg/ml. Se puede obtener una recogida eficaz cuando se emplacan 1500 cfu en placas de agar de 20x20cm (Genetix, Q-Tray). Se emplacó 1 ml de la dilución apropiada utilizando partículas de vidrio. Las placas se dejaron crecer toda la noche a 37°C.

30 [0198] Las colonias se recogieron y transfirieron a placas de microtitulación utilizando Q-Pix (Genetix Ltd.).

[0199] Esto dio lugar al crecimiento y almacenamiento de 45312 clones. Las placas de microtitulación se pueden guardar a -80C con glicerol al 10% hasta estar listas para su uso. A partir de estos clones que contienen ADN de *T. reesei*, se dispusieron 34500 clones en membranas de nylon y se utilizaron para experimentos de hibridación.

35 [0200] Se utiliza Q-Pix (Genetix Ltd.) para pasar los clones a MTPs de 384 pocillos. Después del crecimiento, se utiliza Q-Pix para situar en rejilla las MTP de 384 pocillos en filtros de membrana. Estos filtros de membrana se utilizaron para experimentos de hibridación con sondas que contienen CBM para buscar celulasas que contienen CBM nuevos.

40 Preparación de sondas

[0201] Las sondas se generaron utilizando cebadores especificados en la Tabla 1. Las sondas de CBM se diseñaron utilizando secuencias conocidas de módulos de unión a carbohidratos de *Trichoderma reesei*. Véase, Paul Birch, Curr. Genet (1998) 33; 70-76. Brevemente, para las sondas de CBM, se mezcló ADN genómico total de QM6A de *T. reesei* (100 ng/50 µl) con un volumen 1 µl/50 µl 10 mM de FRG164 y un volumen 1 µl/50 µl 100 mM de FRG165, FRG166 o FRG167. FRG166 no dio lugar a la amplificación (el codón de Ser era AGY), mientras que FRG167 daba lugar a una amplificación (el codón de Ser era TCN). De este modo, se utilizó el cebador FRG167 en la amplificación. Este fragmento se mezcló con el fragmento producido con FRG165 como cebador. Los dos fragmentos separados se mezclaron y contenían una mezcla de secuencias de CBM presentes en *T. reesei* y se utilizaron como la sonda de CBM. En resumen: la sonda de CBM se preparó mediante la mezcla de los fragmentos obtenidos mediante PCR utilizando las combinaciones: FRG164+FRG165 y FRG164+FRG167, 2,5 unidades de platinum TAQ polimerasa, 5 µl 10x tampón TAQ, 1,5 µl MgCl₂ y 1 µl de dNTPs 10mM. La PCR se realizó de la siguiente manera:

1 ciclo:

55 1 minuto a 98°C

10 ciclos:

1 minuto a 94°C

1,5 minutos a 65-50°C

1 minuto a 72°C

60 25 ciclos:

1 minuto a 94°C
 1,5 minutos a 50°C
 1 minuto a 72°C
 Parar la reacción y guardar a 15°C

5 [0202] Para sondas del núcleo catalítico (es decir, específicas de genes), se mezcló el ADN genómica total de QM6A de *T. reesei* (100 ng/50 µl) con 1 µl de concentración de cebador 10 mM en un volumen total de 50 µl, 2,5 unidades de Platinum Taq polimerasa, 5 µl 10 X tampón TAQ, 1,5 µl de MgCl₂ y 1 µl de dNTPs 10 mM utilizando el protocolo anterior, pero en lugar de 50°C, se utilizaron 55°C.

10 [0203] Las sondas se purificaron utilizando métodos estándar. En esta serie de experimentos, las sondas se purificaron mediante purificación en gel utilizando el Kit de purificación en gel Qiagen.

15 Detección

[0204] Las muestras de colonias recogidas de las placas de microtitulación se dispusieron en filtros de membrana de nylon de 20 x 20 cm (Hybond+ (RPN.82B), Amersham) y se desarrollaron durante toda la noche a 37°C después de colocar los filtros en una agar amplio 2xTY (100 µg/ml ampicilina). Cada membrana de 20 x 20 cm contenía 4600 clones por duplicado. A continuación, las placas se procesaron mediante ECL según las instrucciones del fabricante por la presencia de cualquiera de las secuencias específicas de genes o CBM.

20 [0205] Se realizó una prehibridación en tampón ECL provisto con el kit ECL Direct durante 20 minutos, durante los cuales la sonda se marcó (exactamente según el protocolo). La sonda se añade directamente a la solución de prehibridación hasta una concentración de 10 ng/ml y se hibridó durante aproximadamente 60 minutos a 42°C. A continuación, se lavaron los filtros dos veces en tampón primario (urea 6M, 0,5x SSC, SDS al 0,4%) a 42°C durante 20 minutos/lavado y dos veces en tampón secundario (2xSSC) a temperatura ambiente durante 5 minutos/lavado. Después de drenar el tampón de lavado en exceso, se llevó a cabo la detección mediante la adición de una mezcla de volumen igual de los reactivos de detección ECL directamente a los filtros hasta 0,125 ml/cm². Después de un minuto, se drenó el exceso de reactivo, las membranas se envolvieron en SaranWrap™ y se expusieron a Hyperfilm™ ECL (RPN.2 103) durante menos de dos horas, habitualmente durante 10 minutos. Las colonias que muestran una señal positiva se seleccionaron a continuación para un análisis posterior mediante otros métodos, tales como subclonación, secuenciación de ADN, localización por restricción y PCR.

25 [0206] Debido a que el sistema ECL utiliza un marcador enzimático y éste se desactiva tras la reacción quimioluminiscente, no es necesario extraer la transferencia ("blot") de la vieja sonda antes de iniciar la segunda hibridación y posteriores. La transferencia ("blot") debe mantenerse en el reactivo de detección durante la noche antes de volver a sondar la transferencia según el protocolo descrito.

30 [0207] Las 9 sondas específicas de genes, es decir, sondas para módulos catalíticos, se mezclaron y utilizaron como una "megasonda". Después de esta hibridación, se reutilizaron las mismas transferencias ("blots") y se hibridaron con la sonda de los CBM. Mediante la sustracción de los puntos de la "megasonda" de los CBM, se detectaron los desconocidos. Se cribaron un total de 34500 clones; se escogieron de forma óptima 264 clones y se sondaron con sondas catalíticas específicas; se hibridaron 20 clones con una sonda específica de gen de LT1-24. Los cebadores utilizados fueron los siguientes: Cebador directo: P002248: GAC AAT CCA AAC GAC GCT; y cebador inverso: PVS173: CAA TCG AGA TGT CGT CGA AC.

35 [0208] Se identificó un clon, LT1-24, que comprende cip1, que produjo una señal cuando se sondó con un grupo mixto de sondas de CBM, pero no consiguió generar una señal cuando se sondó con las sondas agrupadas de dominios catalíticos. De este modo, la hibridación por sustracción dio lugar a la identificación de un nuevo gen de *T. reesei*, cip1, que contiene CBM. La secuencia completa del ADNc de este gen se determinó utilizando técnicas conocidas en el sector. Tiene una señal de secreción prevista, una región "catalítica" de función desconocida, una región de enlace y un módulo de unión a celulosa C-terminal (CBM).

Tabla 1

Cebadores específicos de genes (dominio catalítico) y degenerados (CBM) de los genes conocidos que contienen CBM en <i>T. reesei</i> (ADN cromosómico:QM6A). N(=A o C o G o T), R(=A o G), Y(=C o T), D(G o A o T)			
cebadores específicos de genes (dominio catalítico)			
Gen	Orientación	Cebador	Secuencia
cbh1	directo	FRG168	CTC CTC CAC ACC CGG TGC CG

	inverso	FRG169	TGC TGC CAA TGG GTC CG	
cbh2	directo	FRG170	ACG TAT TCA GGC AAC CC	
	inverso	FRG171	GCA GTG GCC ATG GCT CC	
egl1	directo	FRG172	CCA GTA CAT GAA CTG GC	
	inverso	FRG173	AGA CCC AAT GTC TCC CC	
egl2	directo	FRG184	CGA ATT GTG CTC CTG GC	
	inverso	FRG185	GTG GTT GGA CCG GAT GG	
egl4	directo	FRG176	CCT ACC GTG GTA TCA GG	
	inverso	FRG177	TGG TTC TGC TGG TCG GG	
egl5	directo	FRG178	CAT TTC GAC ATC ATG GC	
	inverso	FRG179	CTG TCC CAC GCA GAG GC	
axe1	directo	FRG180	CCG GCT GGC TTC GTC TG	
	inverso	FRG181	TGG CCG TAA CCT TGG TG	
man1	directo	FRG182	CCT CTC TCA CGA CTC GC	
	inverso	FRG183	GTT CGA TGA GTT GTA CC	
swo1	directo	PVS159	CCC CCA AAC GGA ACA ACT TCC	
	inverso	PVS160	CTG TAT CTG TGG TTG TGT AGG	
Cebadores degenerados de CBM				
Caja	orientación	cebador	secuencia	deg.
GQCGG	directa	FRG164	GGNCARTGYGGN GG	64X
YSQC(L/I)	Inversa	FRG165	AD RCA YTG NGA RTA	96X
YSQC(L/I)	Inversa	FRG166	AD RCA YTG RCT RTA	32X
YAQC(L/I)	Inversa	FRG167	AD RCA YTG NGC RTA	96X

Ejemplo 3

Identificación y análisis de secuencia de los genes de *T. reesei* de interés

- 5 [0209] La secuenciación parcial de clones de ADNc anónimos es una técnica ampliamente utilizada para la identificación de genes. Estas secuencias de ADNc parciales o marcadores de secuencia expresada (ESTs) tienen una aplicación potencial para la identificación de genes importantes implicados en la degradación de celulosa.
- 10 [0210] Se aisló el plásmido que contiene un inserto de ADNc de clones de las bibliotecas descritas en el ejemplo 1 y se obtuvo una secuencia 5' de paso único del inserto de ADNc de aproximadamente 18.000 clones en la North Carolina State University (Fungal Genomics Laboratory, College of Agriculture and Life Sciences, Raleigh, NC). Las secuencias de los ADNc se obtuvieron utilizando un cebador correspondiente a la secuencia del vector adyacente al extremo 5' del inserto de ADNc. Las secuencias de lecturas de secuencia individuales se compararon y se ensamblaron segmentos solapantes para formar 2102 contigs (secuencias ensambladas) que consisten en dos o más lecturas. 3030 lecturas individuales no presentaron un solapamiento de secuencia significativo con ninguna otra lectura en el conjunto de datos. Las regiones codificantes previstas del grupo de EST se compararon mediante BLAST (Véase Altschul et al. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410) para todas las bases de datos de secuencias disponibles públicamente.
- 15 [0211] Los clones que contienen secuencias de ADNc similares a glicosil hidrolasas conocidas, carbohidrato estererasas o moléculas de unión a carbohidrato se identificaron para una investigación posterior utilizando el programa BLAST (BLASTX y BLASTN) utilizando los parámetros por defecto. Véase, Altschul et al., 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410. Los ADNc de longitud completa correspondientes a estos productos génicos se secuenciaron en su totalidad utilizando técnicas conocidas en el sector.
- 20 [0212] Las secuencias se analizaron utilizando el paquete informático DNASTar o Vector NTI utilizando los parámetros por defecto.
- 25 [0213] Los genes de interés identificados mediante este método se muestran en la Tabla 2.

TABLA 2. Actividades de degradación de biomasa y sus genes en *Trichoderma reesei*

Gen	Familia ^a	Función ^b	Características ^c	Acceso#
axe2	CE5	Acetil xilano esterasa	SS, GPI	AY281376
cip2	No asignado	Desconocida	SS,CBM	AY281368
abf2	GH62	Arabinofuranosidasa	SS	AY281369
cip1	No asignado	Desconocida	SS,CBM	AY281370

^aFamilia de glicosil hidrolasa, GH; CE, familia de carbohidrato esterasa,
^bLa función de proteínas identificadas en este estudio se prevé a partir de la secuencia de aminoácidos codificada.
^cSS, secuencia señal N-terminal; CBM, módulo de unión a carbohidrato; GPI, ancla de glicosil fosfatidilinositol. Las características de las proteínas identificadas en este estudio se prevén a partir de la secuencia de aminoácidos codificada.

Ejemplo 4

Coregulación de los genes cip1 y cip2 con otros genes de celulasa

5 [0214] Se inducen endoglucanasas previamente identificadas durante el crecimiento en medio que contiene celulosa, soforosa o lactosa. Para determinar si los polipéptidos recién descubiertos, CIP1 y CIP2, con potenciales papeles en la degradación de la biomasa se regulan de manera similar, se examinaron los niveles de ARNm para cada uno de los productos génicos mediante transferencia Northern. Se utilizaron dos cepas diferentes: QM6a, un aislado de tipo salvaje de *T. reesei* y RL-P37, una cepa que se ha seleccionado por la producción mejorada de enzimas celulolíticas. Los micelios de cada una de estas cepas se desarrollaron en matraces en medio mínimo que contenía glucosa, celulosa cristalina (avicel) o glicerol como única fuente de carbono, o glicerol complementado con soforosa 1 mM.

15 [0215] Se utilizaron microarrays para examinar la regulación de los genes *cip1* y *cip2*.

Generación de ARNm

20 [0216] Se obtuvieron cepas de *Trichoderma reesei* de la American Type Culture collection.

[0217] Para el análisis de transferencia Northern, se inocularon $\sim 1 \times 10^7$ esporas en 50 ml de medio mínimo complementado con glucosa al 5% y se desarrollaron durante 24 horas. Se recogieron los micelios mediante centrifugación, se lavaron en medio sin carbono y se resuspendieron hasta una densidad óptica de -0,3 en 50 ml de medio mínimo complementado con glucosa al 5%, avicel al 2%, glicerol al 2% o glicerol al 2% que contiene soforosa 1 mM (Sigma). Los cultivos se desarrollaron a 30°C en matraces con aireación vigorosa durante 20 horas.

30 [0218] Los micelios se recogieron mediante filtración a través de miracloth y se congelaron rápidamente en nitrógeno líquido. Se preparó ARN de los micelios mediante molienda bajo nitrógeno líquido con un mortero y mano de mortero y la extracción utilizando reactivo Trizol (Invitrogen Life Technologies) según las instrucciones del fabricante. Se seleccionó ARN poliadenilado 2 veces utilizando Oligotex (Qiagen). La transferencia se realizó utilizando un kit NorthernMax-Gly (Ambion). Se generaron sondas marcadas con ³²P utilizando un kit aDECAprime (Ambion). La hibridación se realizó utilizando el tampón de hibridación ultrasensible ULTRAhyb (Ambion).

35 [0219] Las endoglucanasas de *T. reesei* conocidas son inducidas durante el crecimiento en medio que contiene celulosa, soforosa o lactosa. Para determinar si *cip1* se regula de manera similar, se examinaron los niveles de ARNm para las endoglucanasas *egII*, *egIII* y *cip1* mediante transferencia Northern. Se utilizaron dos cepas diferentes: QM6a, un aislado de tipo salvaje de *T. reesei* y RL-P37, una cepa que se ha seleccionado por la producción mejorada de enzimas celulolíticas. Los micelios de cada una de estas cepas se desarrollaron en matraces en medio mínimo que contenía glucosa, celulosa cristalina (avicel) o glicerol como única fuente de carbono, o glicerol complementado con soforosa 1 mM. Tal como se muestra en la figura 13, las endoglucanasas se regularon de forma muy similar entre sí y a *cip1*. La inducción mediante soforosa dio lugar a niveles más elevados de expresión que con el crecimiento en celulosa durante el periodo de tiempo examinado. Además, la expresión de estos genes era sustancialmente superior en la cepa RL-P37 que en QM6a.

45 Microarrays

[0220] Para medir los niveles de expresión de *cip1* y *cip2* se construyeron microarrays, donde sondas de oligonucleótidos de sesenta pb que contenían secuencias únicas, de las que cada uno de los EST se diseñó para averiguar la abundancia de sus correspondientes ARNm. Se sintetizaron sondas de oligonucleótidos y se

seleccionaron tal como se describe en Hughes et al. (2001) Nature Biotechnol 19:342-347 by Agilent Technologies, Palo Alto, CA. En todos los experimentos realizados, se utilizaron microarrays para determinar los niveles de expresión entre dos muestras diferentes.

5 [0221] Los ARNs que comprenden las muestras de interés se marcaron con los colorantes fluorescentes Cy5 y Cy3 Perkin Elmer/NEN. Se combinaron las parejas de muestras marcadas recíprocamente y se cohibridaron a los "arrays". El logaritmo de la proporción (proporción en logaritmo) de dos especies fluorescentes unidas a cada una de las sondas refleja los niveles de expresión relativa de los genes afines en las dos muestras. Véase Hughes et al. (2001), supra, and DeRisi et al. (1996) Nat Genet 14:457-460.

10 [0222] Aunque las dos enzimas glicolíticas potenciales, *cip1* y *cip2*, no se ajustan a ninguna clase actualmente definida de GHs, su regulación es análoga a GHs conocidas. La regulación de *cip1* entre cepas con capacidades de producción de celulosa variables y a lo largo de una serie de condiciones es indistinguible de las endoglucanasas y particularmente la celobiohidrolasa *cbh1/cel7a* (Figuras 13 y 14). De manera similar, *cip1* tiene un patrón de expresión en común con estos genes, particularmente en RL-P37. La coregulación de estos genes con componentes de celulosa regulados canónicamente y el hecho de que contienen diferentes módulos de unión a celulosa favorece la idea de que *cip1* y *cip2* codifican actividades no reconocidas previamente con potenciales papeles en la degradación de la biomasa.

15 [0223] Los genes que codifican la hemicelulasa *axe2* y *abf2* parecen estar inducidos diferencialmente por lactosa y por soforosa en QM6a o en RL-P37 o ambas. De manera más destacable, *abf2* se indujo más sustancialmente en RL-P37 durante el crecimiento en lactosa que en soforosa (datos no mostrados).

20 [0224] Esto muestra que los genes nuevos, *cip1* y *cip2*, se regulan de manera coordinada con otras enzimas degradantes de celulosa.

25 Ejemplo 5

Construcción de una cepa de *Trichoderma reesei* con cuatro genes de celulosa eliminados (eliminación de quad)

30 [0225] Este ejemplo describe la construcción de un huésped de expresión adecuado. Más específicamente, en este ejemplo se describe la construcción de un huésped de expresión de *Trichoderma* al que se han eliminados los principales genes de celulosa. Los métodos utilizados aquí se han descrito previamente en, por ejemplo, las Patentes de Estados Unidos 5,650,322, 5,874,276 y 6,286,196.

35 [0226] Se construyó una cepa de *T. reesei* en la que los genes que codifican celobiohidrolasa I (CBHI, Cel7a), celobiohidrolasa II (CBHII, Cel6a), endoglucanasa I (EGI, Cel7b), y endoglucanasa II (EGII, Cel5a) se desactivaron por eliminación o alteración utilizando técnicas de genética molecular. Esta cepa (una cepa con quad eliminado) es útil como huésped para la sobreexpresión de genes que codifican otras proteínas secretadas por *T. reesei*.

40 [0227] La cepa huésped de *T. reesei* utilizada fue la cepa RL-P37 que se había utilizado previamente para producir preparaciones de celulosa comercial por Genencor International, Inc. La derivación y caracterización de esta cepa se ha publicado previamente (Sheir-Neiss, G. and Montenecourt, B.S. (1984) Appl. Microbiol. Biotechnol. 20:46-53; Patente de Estados Unidos 4,797,361). Es una cepa mutante sobreproductora de celulosa que se ha obtenido como resultado de varias etapas de mutagénesis de la cepa de tipo salvaje (QM6a).

1) Aislamiento de la cepa mutante con *pyr4*.

45 [0228] Con el fin de preparar la cepa RL-P37 para la transformación con el ADN plasmídico fue necesario aislar un derivado que tenía una mutación nula en el gen *pyr4*.

50 [0229] El gen *pyr4* codifica la orotidin-5'-monofosfato descarboxilasa, una enzima necesaria para la biosíntesis de uridina. El inhibidor tóxico ácido 5-fluoroorótico (FOA) es incorporado en la uridina por las células de tipo salvaje y, de este modo, envenena las células. Sin embargo, las células defectuosas en el gen *pyr4* son resistentes a este inhibidor, pero necesitan uridina para el crecimiento. Por tanto, es posible seleccionar cepas mutantes con *pyr4* utilizando FOA. En la práctica, las esporas de la cepa RL-P37 de *T. reesei* se extendieron por la superficie de un medio solidificado que contenía 2mg/ml de uridina y 1,2 mg/ml de FOA. Aparecieron colonias espontáneas resistentes a FOA en tres o cuatro días. Posteriormente se identificaron los mutantes resistentes a FOA que requerían uridina para el crecimiento. Con el fin de identificar aquellos mutantes que presentaban específicamente un gen *pyr4* defectuoso se generaron protoplastos y se transformaron con un plásmido que contenía un gen *pyr4* de

tipo salvaje. (Smith, J.L., Bayliss, F.T. y Ward, M. (1991) Curr. Genet. 19:27-33). Tras la transformación, se emplacaron protoplastos en un medio que carecía de uridina. El posterior crecimiento de colonias transformadas demostró la complementación de un gen *pyr4* defectuoso por el gen *pyr4* que porta el plásmido. De esta manera, se identificó la cepa GC69 como un mutante en *pyr4* de la cepa RL-P37.

5

2) Construcción de un plásmido diseñado para eliminar el gen que codifica CBHI

[0230] El gen *cbh1*, que codifica la proteína CBHI, se clonó a partir del ADN genómico de la cepa RL-P37 mediante hibridación con una sonda de oligonucleótidos diseñada en base a la secuencia publicada para este gen (Shoemaker, S., Schweickart, V., Ladner, M., Gelfand, D., Kwok, S., Myambo, K. y Innis, M. (1983) Biotechnology 1:691-696). El gen *cbh1* reside en un fragmento PstI de 6,5 kb y se insertó en el sitio PstI de pUC4K (Farmacia Inc., Piscataway, NJ, USA) sustituyendo el gen de resistencia a kanamicina de este vector. El plásmido resultante pUC4K::cbh1 se cortó a continuación con HindIII y se aisló el fragmento más grande y se volvió a unir para producir pUC4K::cbh1ΔH/H. Este procedimiento extrajo la secuencia codificante de *cbh1* completa y aproximadamente 1,2 kb de las secuencias flanqueadas 5' y 1,5 kb de la secuencia flanqueada 3'. Permaneció aproximadamente 1 kb del ADN flanqueado de cualquiera de los extremos del fragmento PstI original.

10

15

[0231] El gen *pyr4* de *T. reesei* se clonó como un fragmento HindIII de 6,5 kb de ADN genómico en pUC18 para formar pTpyr2 (Smith, J.L., Bayliss, F.T. y Ward, M. (1991) Curr. Genet. 19:27-33). El plásmido pUC4K::cbh1ΔH/H se cortó con HindIII y los extremos se desfosforilaron con fosfatasa alcalina intestinal de ternera. Este ADN se unió con el fragmento HindIII de 6,5 kb que contenía el gen *pyr4* para producir pΔCBHIpyr4.

20

[0232] La digestión de pΔCBHIpyr4 con EcoRI liberó un fragmento más grande que consistía en regiones flanqueadas del locus de *cbh1* en cualquier extremo con el gen *pyr4* sustituyendo la secuencia codificante de *cbh1* en el centro. El único ADN en este fragmento que no derivaba de *T. reesei* era un fragmento de 21 pb derivado del sitio de clonación múltiple de pUC4K.

25

3) Eliminación del gen *cbh1* de *T. reesei*.

[0233] Los protoplastos de micelio de la cepa GC69 se transformaron con el plásmido pΔCBHIpyr4 digerido con EcoRI utilizando los métodos indicados Smith et al., 1991. Se obtuvieron transformantes estables y se identificaron aquellos a los que se había eliminado el gen *cbh1* tal como se describe a continuación.

30

[0234] El ADN total se aisló de los transformantes, se digirió con PstI, se sometió a electroforesis en gel de agarosa y se transfirió a un filtro de membrana. A continuación, el filtro se hibridó con pΔCBHIpyr4 marcado con ³²P y se observó el patrón de hibridación mediante autoradiografía. Esta sonda se hibridó con los genes *cbh1* y *pyr4* nativos en una cepa no transformada. En un transformante (cepa P37PΔCBHI), se observó un patrón de hibridación que se hubiera previsto si hubiera tenido lugar una integración por doble entrecruzamiento. Es decir, el gen *cbh1* se había eliminado mediante la integración de una única copia del fragmento EcoRI más grande obtenido de pΔCBHIpyr4 en el locus *cbh1* de la cepa RL-P37.

35

40

[0235] También se realizó un análisis Southern como antes a excepción que la sonda utilizada fue pIntCBHI radiomarcada. Este plásmido consiste en un vector pUC que contiene un fragmento BglII de 2 kb del locus *cbh1* en la región que se eliminó en pUC4K::cbh1ΔH/H. Este plásmido se hibridó al locus *cbh1* de la cepa GC69, pero no se hibridó al ADN de la cepa P37PΔCBHI. Esto confirma que el gen *cbh1* se había eliminado y que el fragmento de ADN pUC de pΔCBHIpyr4 no había sido incorporado por la cepa eliminada.

45

[0236] El análisis de las proteínas secretadas mediante separación en geles por enfoque isoelectrico mostró que la proteína CBHI no era producida por la cepa P37PΔCBHI.

50

4) Generación de un mutante nulo en *pyr4* de P37PΔCBHI.

[0237] Se extendieron esporas del transformante (P37PΔCBHI) el cual fue eliminado para el gen *cbh1* sobre un medio que contenía FOA. Posteriormente se obtuvo un derivado de este transformante deficiente en *pyr4* utilizando los métodos descritos en la sección 1 anterior. La cepa deficiente en *pyr4* se designó como P37PΔCBHIPyr-26. El análisis southern mostró que había tenido lugar una delección espontánea cuando se seleccionó la cepa P37PΔCBHIPyr-26. Esta delección eliminaba completamente el gen *pyr4* que se había integrado en el locus *cbh1* en la cepa P37PΔCBHI, así como el ADN flanqueado del locus *cbh1* más allá de la extensión del fragmento PstI de 6,5 kb del ADN genómico que se clonó originalmente.

55

60

5) Construcción de un vector diseñado para eliminar el gen *cbh2*

[0238] el gen *cbh2* de *T. reesei*, que codifica la proteína CBHII, se clonó como un fragmento *EcoRI* de 4,1 kb de ADN genómico (Chen et al., 1987, *Biotechnology* 5:274-278). Este fragmento de 4,1 kb se insertó entre los sitios *EcoRI* de pUC4XL. Este último plásmido es un derivado de pUC (construido por R. M. Berka, Genencor International Inc.) que contiene un sitio de clonación múltiple con un patrón simétrico de sitios para endonucleasas de restricción dispuestos en el orden que se muestra aquí. *EcoRI*, *BamHI*, *SacI*, *SmaI*, *HindIII*, *XhoI*, *BglII*, *Clal*, *BglI*, *XhoI*, *HindIII*, *SmaI*, *SacI*, *BamHI*, *EcoRI*. Se construyó el plásmido p Δ CBHII, en donde se ha eliminado una región central de 1,7 kb de este clon *cbh2*, entre un sitio *HindIII* (a 74 pb en 3' del sitio de iniciación de la traducción de CBHII) y un sitio *Clal* (a 265 pb en 3' del último codón de CBHII), y se ha sustituido por un fragmento de ADN *HindIII-Clal* de 1,6 kb que contiene el gen *pyr4* de *T. reesei* obtenido de la siguiente manera. El gen *pyr4* de *T. reesei* se escindió de pTpyr2 en un fragmento *NheI-SphI* de 1,6 kb y se insertó entre los sitios *SphI* y *XbaI* de pUC219 (derivado de pUC119 mediante la expansión del sitio de clonación múltiple para incluir los sitios de restricción para *BglII*, *Clal* y *XhoI*; Wilson et al., 1989, *Gene* 77:69-78) para crear p219M (Smith et al., 1991, *Curr. Genet.* 19:27-33). A continuación, el gen *pyr4* se pudo extraer como un fragmento *HindIII-Clal* que tiene siete pb de ADN en un extremo y seis pb de ADN en el otro extremo derivado del sitio de clonación múltiple de pUC219 y se insertó en los sitios *HindIII* y *Clal* del gen *cbh2* para formar el plásmido p Δ CBHII.

[0239] La digestión de este plásmido con *EcoRI* liberó un fragmento que tenía 0,7 kb de ADN flanqueante del locus *cbh2* en un extremo, 1,7 kb del ADN flanqueante del locus *cbh2* en el otro extremo y el gen *pyr4* de *T. reesei* en el centro. El único ADN en este fragmento que no derivaba de *T. reesei* eran los fragmentos de 6 pb y 7 pb del sitio de clonación múltiple de pUC219 en cualquiera de los extremos del gen *pyr4*.

6) Deleción del gen *cbh2* de la cepa P37P Δ CBHIPyr-26.

[0240] Los protoplastos de la cepa P37P Δ CBHIPyr-26 se generaron y se transformaron con p Δ CBHII digeridos con *EcoRI* según los métodos descritos en 3 anterior. Los transformantes estables se cultivaron en matraces de agitación y se examinó la proteína en los sobrenadantes de cultivo mediante enfoque isoeléctrico. Se identificó un transformante (designado como P37P Δ CBH67) que no producía ninguna proteína CBHII (ni CBHI).

[0241] Se extrajo ADN de la cepa P37P Δ CBH67, se digirió con *EcoRI* y *Asp718* y se sometió a electroforesis en gel de agarosa. El ADN de este gel se transfirió a un filtro de membrana y se hibridó con p Δ CBHI marcado con ³²P. El fragmento *EcoRI* de 4,1 kb que contenía el gen *cbh2* de tipo salvaje se observó en el ADN de una cepa de control no transformada. En cambio, en la cepa P37P Δ CBH67 se eliminó la única banda de 4,1 kb y se sustituyó por dos bandas de aproximadamente 0,9 y 3,1 kb. Éste es el patrón esperado si una copia única del fragmento *EcoRI* más amplio de p Δ CBHI se hubiera integrado de manera precisa en el locus *cbh2* y se hubiera eliminado el gen *cbh2*.

[0242] Las mismas muestras de ADN se digirieron también con *EcoRI* y se realizó un análisis Southern al igual que antes. En este ejemplo, la sonda era plntCBHII marcada con ³²P. Este plásmido contiene una parte de la secuencia codificante del gen *cbh2* de la que el segmento del ADN de *cbh2* se había eliminado en el plásmido p Δ CBHII. No se observó hibridación con el ADN de la cepa P37P Δ CBH67 confirmando que se eliminó el gen *cbh2* y que el fragmento de plásmido pUC de p Δ CBHII no había sido incorporado por esta cepa.

7) Selección de un mutante nulo en *pyr4* de la cepa P37P Δ CBH67.

[0243] Se extendieron esporas del transformante (P37P Δ CBH67) al que se habían eliminado los genes *cbh1* y *cbh2* en un medio que contenía FOA. Posteriormente se obtuvo un derivado de este transformante deficiente en *pyr4* utilizando los métodos descritos en la sección 1 anterior. Esta cepa deficiente en *pyr4* se designó como P37P Δ CBH67Pr-1. El análisis southern mostró que había tenido lugar una deleción espontánea cuando se seleccionó la cepa P37P Δ CBH67Pr-1. Esta deleción eliminaba completamente el gen *pyr4* que se había integrado en el locus *cbh2* en la cepa P37P Δ CBH67, así como el ADN flanqueante del locus *cbh2* más allá de la extensión del fragmento *EcoRI* de 4,1 kb del ADN genómico que se clonó originalmente. Los fragmentos cortos (6 pb y 7 pb) de ADN derivado del sitio de clonación múltiple pUC219 que estaban presentes en cualquier extremo del gen *pyr4* también se habrían eliminado del genoma mediante esta deleción.

8) Construcción de un plásmido diseñado para alterar el gen *egl2*.

[0244] Se clonó el gen *egl2*, que codificaba EGII (previamente referido como EGIII por algunos autores), a partir de *T. reesei* y la secuencia de ADN publicada (Saloheimo et al., 1988, *Gene* 63:11-21). Se obtuvo el gen de la cepa RL-P37 como un fragmento *PstI-XhoI* de aproximadamente 4 kb del ADN genómico insertado entre los sitios *PstI* y *XhoI*

de pUC219. El gen *pyr4* de *T. reesei*, presente en un fragmento *SaII* de 2,7 kb de ADN genómico obtenido de pTpyr2, se insertó en un sitio *SaII* en la secuencia codificante de EGII para crear el plásmido pEGII::P-1. Esto dio lugar a alteración de la secuencia codificante de EGII, pero sin eliminar ninguna de las secuencias. El plásmido pEGII::P-1 se puede digerir con *HindIII* y *BamHI* para producir un fragmento lineal de ADN derivado exclusivamente de *T. reesei* a excepción de 5 pb en un extremo y 16 pb en el otro extremo, ambos derivados del sitio de clonación múltiple de pUC219.

9) Alteración del gen *egl2* de la cepa P37PΔCBH67Pyr-1.

[0245] La cepa P37PΔCBH67Pyr-1 se transformó con pEGII::P-1 que se había digerido previamente con *HindIII* y *BamHI* y se seleccionaron transformantes estables. Se aisló el ADN total de los transformantes y se utilizó un análisis Southern para identificar cepas en las que se había integrado el fragmento del ADN plasmídico que contenía los genes *pyr4* y *egl2* en el locus *egl2* y consecuentemente se había alterado la secuencia codificante de EGII. Se realizó el análisis southern utilizando como sonda un fragmento *PstI* de aproximadamente 4 kb de ADN de *T. reesei* que contenía el gen *egl2*. Cuando se digirió el ADN aislado de la cepa P37PΔΔ67P-1 con *PstI* para el análisis Southern, el locus *egl2* se visualizó posteriormente como una única banda de 4 kb en la autorradiografía. Sin embargo, para un transformante alterado para el gen *egl2*, esta banda se perdió y se sustituyó por dos nuevas bandas tal como se esperaba. Cuando se digirió el ADN con *BglII* o *EcoRV*, el tamaño de la banda correspondiente al gen *egl2* aumentó en aproximadamente 2,7 kb (el tamaño del fragmento *pyr4* insertado) entre la cepa P37PΔΔ67P-1 no transformada y el transformante alterado para *egl2*. Éste último transformante, ahora con los genes *cbh1*, *cbh2*, y *egl2* eliminados, se designó como cepa B31. Un análisis southern adicional confirmó que el fragmento de ADN pUC de pEGII::P-1 no se incorporó en esta cepa.

10) Selección de un mutante nulo en *pyr4* de la cepa B31

[0246] Se extendieron esporas del transformante (B31) al que se habían eliminado los genes *cbh1*, *cbh2* y *egl2* en un medio que contenía FOA. Posteriormente se obtuvo un derivado de este transformante deficiente en *pyr4* utilizando los métodos descritos en la sección 1 anterior. Esta cepa deficiente en *pyr4* se designó como B31 P6. El análisis southern mostró que había tenido lugar una delección espontánea cuando se seleccionó la cepa B31 P6. Esta delección eliminaba la mayoría del gen *pyr4* que se había integrado en el locus *egl2* en la cepa B31, pero no se extendió en el ADN flanqueante del locus *egl2*.

11) Construcción de un plásmido diseñado para eliminar el gen *eal1*

[0247] El gen *egl1* de *T. reesei* se ha clonado y la secuencia de ADN del gen se ha publicado (Penttila et al., 1986, Gene 45:253-263; van Arsdell et al., 1987, Bio/technology 5:60-64). Se obtuvo este gen de la cepa RL-P37 de *T. reesei* como un fragmento *HindIII* de 4,2 kb de ADN genómico insertado en el sitio *HindIII* de pUC100 (un derivado de pUC18 con un oligonucleótido insertado en el sitio de clonación múltiple que añade sitios de restricción para *BglII*, *ClaI* y *XhoI*) para producir pUCEGI. Se extrajo un fragmento *EcoRV* de aproximadamente 1 kb que se extiende desde una posición próxima al dentro de la secuencia codificante de EGI hasta una posición más allá del extremo 3' de la secuencia codificante y se sustituyó por un fragmento *Scal* de 3,5 kb de ADN de *T. reesei* que contiene el gen *pyr4* obtenido de pTpyr2. El plásmido resultante se denominó pΔEGI.

[0248] El plásmido, pΔEGI, se pudo digerir con *HindIII* para liberar un fragmento de ADN que comprendía sólo ADN genómico de *T. reesei* que tenía un segmento del gen *egl1* en cualquier extremo y el gen *pyr4*, sustituyendo parte de la secuencia codificante de EGI, en el centro.

12) Delección del gen *egl1* en la cepa B31P6

[0249] Se construyeron dos formas de pΔEG1 que diferían sólo en la orientación del gen *pyr4* con respecto a las regiones flanqueantes de *egl1*. La cepa B31 P6 se transformó con una mezcla de ambas formas del plásmido después de haberse digerido con *HindIII*. Se extrajo el ADN total de los transformantes estables, se digirieron con *HindIII* y se sometieron a análisis Southern. La sonda utilizada fue pUCEGI radiomarcada. Se observó hibridación a un fragmento de ADN de 4,2 kb de la cepa B31 P6 que representa el gen *egl1* no eliminado. Se identificó un transformante (cepa 1A52) en el que este fragmento de 4,2 kb ya no está presente porque se había sustituido por un fragmento de aproximadamente 6,8 kb. Éste es el patrón esperado si el fragmento *HindIII* más grande de pΔEG1 se hubiera integrado de manera precisa tal como se predijo en el locus *egl1* conduciendo a la delección de parte de la secuencia codificante de EGI y la inserción de *pyr4* en esta posición. Utilizando un plásmido pUC como sonda para el análisis Southern se confirmó que el fragmento de ADN pUC de pΔEG1 no se había incorporado en la cepa 1A52.

Ejemplo 6

Construcción del vector de expresión pTrex3g

[0250] Este ejemplo describe la construcción del vector básico para expresar los genes de interés.

[0251] Este vector se basa en el vector de *E. coli* pSL1180 (Pharmacia Inc., Piscataway, NJ, USA) que es un vector basado en el fagémido pUC118 (Brosius, J. (1989) DNA 8:759) con un sitio de clonación múltiple extendido que contiene secuencias de reconocimiento de enzimas de restricción de 64 hexámeros. Se diseñó como vector de destino Gateway (Hartley, J.L., Temple, G.F. y Brasch, M.A. (2000) Genome Research 10:1788-1795) para permitir la inserción utilizando tecnología Gateway (Invitrogen) de cualquier marco de lectura abierto deseado entre las regiones del promotor y el terminador del gen *cbh1* *T. reesei*. También contiene el gen de *Aspergillus nidulans amdS* para su uso como marcador seleccionable en la transformación de *T. reesei*.

[0252] Los detalles de pTrex3g son los siguientes (véase la figura 17). El vector tiene 10,3 kb de tamaño. En la región polienlazadora de pSL1180 se encuentran insertados los siguientes fragmentos de ADN:

1. Un segmento de ADN de 2,2 pb de la región de promotor del gen *cbh1* de *T. reesei*.
2. El cassette del marco de lectura A de Gateway de 1,7 kb adquirido de Invitrogen que incluye la recombinación de los sitios attR1 y attR2 en cualquier extremo que flanquea el gen de resistencia a cloramfenicol (CmR) y el gen *ccdB*.
3. Un segmento de ADN de 336 pb de la región de terminador del gen *cbh1* de *T. reesei*.
4. Un fragmento de ADN de 2,7 kb que contiene el gen de *Aspergillus nidulans amdS* con sus regiones de promotor y terminador nativas.

Ejemplo comparativo 7

Inserción de la región codificante de *cip1* en pTrex3g

[0253] Este ejemplo describe la construcción del vector de expresión para *cip1*.

[0254] El marco de lectura abierto de *cip1* se amplificó mediante la reacción en cadena de la polimerasa (PCR) utilizando ADN genómico purificado de la cepa QM6A de *Trichoderma reesei* (ATCC 13631) como molde. La máquina de PCR utilizada fue un Peltier Thermal Cycler PTC-200 (MJ Research). La ADN polimerasa utilizada en la PCR fue la Herculasa (Stratagene). Los cebadores utilizados para amplificar el gen *cip1* fueron el cebador 170 (directo) 5'-CACCATGGTTCGCCGGACTGCTCTG-3', y el cebador 171 (inverso) 5'-TTATAAGCACTGGGAGTAGTATGG-3'. El cebador directo contenía cuatro nucleótidos adicionales (secuencia -CACC) en el extremo 5' que no correspondían con el gen *cip1*, pero eran necesarios para la clonación en el vector pENTR/DTOPO. Las condiciones PCR para amplificar el marco de lectura abierto de *cip1* fueron las siguientes: Etapa 1: 94C durante 2 min. Etapa 2: 94 C durante 30 s. Etapa 3: 58C durante 30 s. Etapa 4: 72C durante 35 s. Las etapas 2, 3 y 4 se repitieron durante 21 ciclos adicionales. Etapa 5: 72C durante 5 min.

[0255] El producto PCR se purificó utilizando un Kit de Purificación de PCR Qiaquick (Qiagen). El producto PCR purificado se clonó inicialmente en el vector pENTR/D-TOPO (Invitrogen, Fig. 18), se transformó en las células de *E. coli* químicamente competentes TOP10 (Invitrogen) y se emplacaron en placas LA con 50 ppm de kanamicina. El ADN plasmídico se obtuvo de los transformantes de *E. coli* utilizando un kit de preparación de plásmidos QIAspin (Qiagen). Se obtuvieron datos de la secuencia para el ADN insertado en el vector pENTR/D-TOPO utilizando los cebadores directo e inverso M13. Se recombinó un vector pENTR/D-TOPO con la secuencia de ADN correcta insertada con el vector pTrex3g utilizando LRclonasa (Invitrogen) según las instrucciones del fabricante. El producto de la reacción con LR clonasa se transformó posteriormente en células de *E. coli* químicamente competentes TOP10 que a continuación se emplacaron en LA que contenía 50 ppm de carbenicilina. La construcción pExpression resultante (Figura 19) fue pTrex3g que contenía el gen *cip1* que resultaba de la recombinación entre los sitios attR1 y attR2 de pTrex3g y los sitios attL1 y attL2 de pENTR/D-TOPO. El ADN de la construcción pExpression que contenía el marco de lectura abierto de *cip1* se aisló utilizando un kit de minipreparación Qiagen para la transformación biolística de esporas de *Trichoderma reesei*.

Ejemplo 8

Inserción de la región codificante *cip2* de pTrex3g

[0256] Este ejemplo describe la construcción del vector de expresión para *cip2*.

[0257] El marco de lectura abierto de *cip2* se amplificó mediante PCR utilizando ADN genómico purificado de la cepa QM6A de *Trichoderma reesei* como molde. La máquina de PCR utilizada fue un Peltier Thermal Cycler PTC-200 (MJ Research). La ADN polimerasa utilizada en la PCR fue la Herculasa (Stratagene). Los cebadores utilizados para amplificar el gen *cip2* fueron el cebador 230 (directo) 5'-CACCATGGCTTCCCGCTTCTTG-3', y el cebador 231 (inverso) 5'-TCAACTCAGCGTTGGGGTTG-3'. El cebador directo contenía cuatro nucleótidos adicionales (secuencia - CACC) en el extremo 5' que no correspondían con el gen *cip2*, pero eran necesarios para la clonación en el vector pENTR/DTOPO. Las condiciones PCR para amplificar el marco de lectura abierto de *cip2* fueron las siguientes: Etapa 1: 94C durante 2 min. Etapa 2: 94 C durante 30 s. Etapa 3: 56C durante 30 s. Etapa 4: 72C durante 1 min 15 s. Las etapas 2, 3 y 4 se repitieron durante 21 ciclos adicionales. Etapa 5: 72C durante 5 min.

[0258] El producto PCR se purificó utilizando un Kit de Purificación de PCR Qiaquick (Qiagen). El producto PCR purificado se clonó inicialmente en el vector pENTR/D-TOPO (Invitrogen, Fig. 18), se transformó en las células de *E. coli* químicamente competentes TOP10 (Invitrogen) y se emplacaron en placas LA con 50 ppm de kanamicina. El ADN plasmídico se obtuvo de los transformantes de *E. coli* utilizando un kit de preparación de plásmidos QIASpin (Qiagen). Se obtuvieron datos de la secuencia para el ADN insertado en el vector pENTR/D-TOPO utilizando los cebadores directo e inverso M13. Se recombinó un vector pENTR/D-TOPO con la secuencia de ADN correcta insertada con el vector pTrex3g utilizando LRclonasa (Invitrogen) según las instrucciones del fabricante. El producto de la reacción con LR clonasa se transformó posteriormente en células de *E. coli* químicamente competentes TOP10 que a continuación se emplacaron en LA que contenía 50 ppm de carbenicilina. La construcción pExpression resultante (Figura 19) fue pTrex3g que contenía el gen *cip2* que resultaba de la recombinación entre los sitios attR1 y attR2 de pTrex3g y los sitios attL1 y attL2 de pENTR/D-TOPO. El ADN de la construcción pExpression que contenía el marco de lectura abierto de *cip2* se aisló utilizando un kit de minipreparación Qiagen para la transformación biolística de esporas de *Trichoderma reesei*.

Ejemplo Comparativo 9

Inserción de la región codificante *cip2* de pTrex3g

[0259] Este ejemplo describe la construcción del vector de expresión para *abf2*.

[0260] El marco de lectura abierto de *abf2* se amplificó mediante PCR utilizando ADN genómico purificado de la cepa QM6A de *Trichoderma reesei* como molde. La máquina de PCR utilizada fue un Peltier Thermal Cycler PTC-200 (MJ Research). La ADN polimerasa utilizada fue Pfu Turbo cx Hotstart (Stratagene). Los cebadores utilizados para amplificar el gen *abf2* fueron NSP071 (directo) 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGAGCTTAAAGCACTCAGTGCCG -3', y NSP072 (inverso) 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTTCAGCGCTGGAGAGTTAGCAGC -3'. Los cebadores directo e inverso contenían 29 nucleótidos en el extremo 5' que no correspondían con el gen *abf2*, pero representaban el sitio attB1 necesario para la clonación en el vector pDONR201 (Invitrogen). Las condiciones PCR para amplificar el marco de lectura abierto de *abf2* fueron las siguientes: Etapa 1: 95C durante 2 min. Etapa 2: 95 C durante 30 s. Etapa 3: 68C durante 30 s. Etapa 4: 72C durante 3 min. Las etapas 2, 3 y 4 se repitieron durante 29 ciclos adicionales. Etapa 5: 72C durante 1 min.

[0261] El producto PCR se clonó en el vector pDONR201 a través de la reacción con BP clonasa utilizando el kit de clonación de PCR con tecnología Gateway® (Invitrogen) según las instrucciones del fabricante. Se obtuvieron datos de la secuencia para el ADN insertado en el vector pDONR201 utilizando los cebadores directo e inverso M13. Se recombinó un vector pDONR201 con la secuencia de ADN correcta insertada con el vector pTrex3g utilizando LRclonasa (Invitrogen) según las instrucciones del fabricante. El producto de la reacción con LR clonasa se transformó posteriormente en células de *E. coli* químicamente competentes TOP10 que a continuación se emplacaron en LA que contenía 50 ppm de carbenicilina. La construcción pExpression resultante (Figura 8) fue pTrex3g que contenía el gen *abf2* que resultaba de la recombinación entre los sitios attR1 y attR2 de pTrex3g y los sitios attL1 y attL2 de pDONR201. El ADN de la construcción pExpression que contenía el marco de lectura abierto de *abf2* se aisló utilizando un kit de minipreparación Qiagen para la transformación biolística de esporas de *Trichoderma reesei*.

Ejemplo comparativo 10

Inserción de la región codificante *axe2* de pTrex3g

[0262] Este ejemplo describe la construcción del vector de expresión para *axe2*.

[0263] El marco de lectura abierto de *axe2* se amplificó mediante PCR utilizando ADN genómico purificado de la cepa QM6A de *Trichoderma reesei* como molde. La máquina de PCR utilizada fue un Peltier Thermal Cycler PTC-200 (MJ Research). La ADN polimerasa utilizada fue Pfu Turbo cx Hotstart (Stratagene). Los cebadores utilizados para amplificar el gen *axe2* fueron NSP111 (directo) 5'-GGGGACAAGTTTGTACAAAAAGCAGGCTATGCGCGCCCTCTCACTCTCC -3', y NSP112 (inverso) 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTACAGCATCTGAGACACCGCC -3'. Los cebadores directo e inverso contenían 29 nucleótidos en el extremo 5' que no correspondían con el gen *axe2*, pero representaban el sitio *attB1* necesario para la clonación en el vector pDONR201 (Invitrogen). Las condiciones PCR para amplificar el marco de lectura abierto de *abf2* fueron las siguientes: Etapa 1: 95C durante 2 min. Etapa 2: 95 C durante 30 s. Etapa 3: 68C durante 30 s. Etapa 4: 72C durante 3 min. Las etapas 2, 3 y 4 se repitieron durante 29 ciclos adicionales. Etapa 5: 72C durante 1 min.

[0264] El producto PCR se clonó en el vector pDONR201 a través de la reacción con BP clonasa utilizando el kit de clonación de PCR con tecnología Gateway® (Invitrogen) según las instrucciones del fabricante. Se obtuvieron datos de la secuencia para el ADN insertado en el vector pDONR201 utilizando los cebadores directo e inverso M13. Se recombinó un vector pDONR201 con la secuencia de ADN correcta insertada con el vector pTrex3g utilizando LRclonasa (Invitrogen) según las instrucciones del fabricante. El producto de la reacción con LR clonasa se transformó posteriormente en células de *E. coli* químicamente competentes TOP10 que a continuación se emplacaron en LA que contenía 50 ppm de carbenicilina. La construcción pExpression resultante (Figura 19) fue pTrex3g que contenía el gen *axe2* que resultaba de la recombinación entre los sitios *attR1* y *attR2* de pTrex3g y los sitios *attL1* y *attL2* de pDONR201.

Ejemplo 11

25

Transformación de una cepa de *T. reesei* con quad eliminado

[0265] Este ejemplo describe la transformación de una cepa de *Trichoderma* con una construcción de expresión. Se realizó una transformación biolística de *T. reesei* con los vectores de expresión pTrex3g con marcos de lectura abiertos de *cip1*, *cip2* o *abf2* (construcciones de pExpression) utilizando el protocolo resumido más adelante.

[0266] Se preparó una suspensión de esporas (aproximadamente 5×10^8 esporas/ml) a partir de una cepa con quad eliminado de *T. reesei*. Se extendieron 100 μ l – 200 μ l de suspensión de esporas en el centro de las placas de medio acetamida MM. El medio de acetamida MM presentaba la siguiente composición: acetamida 0,6 g/L; CsCl 1,68 g/L; glucosa 20 g/L; KH_2PO_4 20 g/L; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 0,6 g/L; solución de 1000X elementos traza 1 ml/L; agar Noble 20 g/L; pH 5,5. La solución de 1000X elementos traza contenía $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 5,0 g/l, $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ 1,6 g/l, $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 1,4 g/l y 1,0 g/l $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$. Se dejó secar la suspensión de esporas en la superficie del medio de acetamida MM.

[0267] Se realizó la transformación de *T. reesei* mediante biolística utilizando un Sistema de Liberación de Partículas PDS-1000/He Biolistic® de Bio-Rad (Hercules, CA) siguiendo las instrucciones de los fabricantes. Brevemente, se colocaron 60 mg de partículas de tungsteno M10 en un tubo de microcentrifuga. Se añadió 1 mL de etanol y se dejó reposar durante 15 minutos. Se centrifugaron las partículas a 15.000 rpm durante 15 segundos. Se extrajo el etanol y se lavaron las partículas tres veces con dH_2O estéril antes de añadir 1 mL de glicerol estéril al 50% (v/v). Se colocaron 25 μ l de suspensión de partículas de tungsteno en un tubo de microcentrifuga. Mientras se agitaba vigorosamente de forma continua, se añadió lo siguiente; 0,5-5 μ l (100-200 ng/ μ l) de ADN plasmídico, 25 μ l de CaCl_2 2,5M y 10 μ l de espermidina 0.1M. Se centrifugaron las partículas durante 3 segundos. Se extrajo el sobrenadante, se lavaron las partículas con 200 μ l de etanol al 70% (v/v) y se centrifugaron durante 3 segundos. Se extrajo el sobrenadante, se lavaron las partículas con 200 μ l de etanol al 100% y se centrifugaron durante 3 segundos. Se extrajo el sobrenadante y se añadieron 24 μ l de etanol al 100%, se mezcló mediante pipeteo, y se colocó el tubo en un baño de lavado ultrasónico durante aproximadamente 15 segundos. Mientras el tubo estaba en el baño ultrasónico, se extrajeron alícuotas de 8 μ l de partículas y se colocaron en el centro del disco microportador sostenido en un desecador. Una vez se hubo secado la solución de tungsteno/ADN, se colocó el disco microportador en la cámara de bombardeo junto con la placa de acetamida MM con esporas y se realizó el proceso de bombardeo de acuerdo con las instrucciones de los fabricantes. Después del bombardeo de las esporas emplacadas con las partículas de tungsteno/ADN, las placas se colocaron y se incubaron a 28°C. Se pasaron las colonias transformadas a placas nuevas de acetamida MM después de 4 días.

[0268] Después de 5 días de crecimiento en placas de acetamida MM se inocularon transformantes que mostraban una morfología estable en matraces de 250 ml con agitación que contenían 30 ml de medio Proflo. El medio Proflo contiene: α -lactosa 30 g/L; $(\text{NH}_4)_2\text{SO}_4$ 6,5 g/L; KH_2PO_4 2g/L; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 0,3 g/L; CaCl_2 0,2g/L; solución de sales con 1000X elementos traza 1 ml/L; Tween 80 al 10% 2 ml/L; harina de semilla de algodón Proflo 22,5g/L (Traders

60

Protein, Memphis, TN); CaCO₃ 0,72g/L. Después de dos días de crecimiento a 28°C y 140 rpm, se transfirió un 10% del cultivo Proflo a un matraz de 250 ml con agitación que contenía 30 ml de Medios Definidos de Lactosa. La composición del Medio Definido de Lactosa fue la siguiente: (NH₄)₂SO₄ 5g/L; tampón PIPPS 33g/L; casaminoácidos 9g/L; KH₂PO₄ 4,5 g/L; MgSO₄·7H₂O 1g/L; antiespuma Mazu DF60-P 5 ml/L (Mazur Chemicals, Gurnee, IL); solución de 1000X elementos traza 1 ml/L; pH 5,5. Se añadió 40 ml/L de una solución de lactosa al 40% (p/v) al medio tras la esterilización. Se incubaron matraces con agitación de medio Definido de Lactosa a 28°C, 140 rpm durante 4-5 días. Se mezclaron muestras de sobrenadante de cultivo con un volumen apropiado de 2X tampón de carga de muestra con agente reductor y se sometieron a electroforesis en gel de dodecil sulfato sódico - poliacrilamida (SDS-PAGE) utilizando geles premoldeados según las instrucciones de los fabricantes (el Sistema de Electroforesis Bis-Tris de NuPAGE de Invitrogen Corporation, Carlsbad, CA. Se utilizaron geles Bis-Tris al 10% de NuPAGE o Bis-Tris al 4-12% de NuPAGE con tampón MOPS. Se utilizaron tampón de muestra LDS de NuPAGE y agente reductor NuPAGE.). Se tiñeron los geles para las proteínas con colorante Azul de Coomassie Brillante.

[0269] En un análisis SDS-PAGE, una banda de proteína que no se observó en el sobrenadante de una cepa con quad eliminado, sí se observó en el sobrenadante de algunos transformantes con el vector pTrex3g que contenían el marco de lectura abierto de *cip1* (Fig. 20). Esta nueva banda de proteína tenía una masa molecular aparente de aproximadamente 50 kDa. Ésta es algo más elevada que el tamaño de 33 kDa previsto a partir de la secuencia genética. La discrepancia se podía justificar por la adición después de la traducción de la glicosilación. Además, se sabe que algunas proteínas no migran según su tamaño en el SDS-PAGE (Saloheimo et al., 1997). Este resultado confirma que Cip1 es una proteína secretada.

[0270] En un análisis SDS-PAGE, una banda de proteína que no se observó en el sobrenadante de una cepa con quad eliminado, sí se observó en el sobrenadante de algunos transformantes con el vector pTrex3g que contenían el marco de lectura abierto de *cip2* (Fig. 21). Esta nueva banda de proteína tenía una masa molecular aparente de aproximadamente 56 kDa. Ésta es algo más elevada que el tamaño de 48 kDa previsto a partir de la secuencia genética. La discrepancia se podía justificar por la adición después de la traducción de la glicosilación. Además, se sabe que algunas proteínas no migran según su tamaño en el SDS-PAGE (Saloheimo et al., 1997). Este resultado confirma que Cip2 es una proteína secretada.

[0271] En un análisis SDS-PAGE, una banda de proteína que no se observó en el sobrenadante de una cepa con quad eliminado, sí se observó en el sobrenadante de algunos transformantes con el vector pTrex3g que contenían el marco de lectura abierto de *abf2* (Fig. 22). Esta nueva banda de proteína tenía una masa molecular aparente muy próxima a los 35 kDa previstos a partir de la secuencia genética. Este resultado confirma que Abf2 es una proteína secretada.

Ejemplo comparativo 12

Purificación de la proteína Cip1 y ensayos de actividad

[0272] Se purificó la proteína Cip1 del sobrenadante de cultivo utilizando un equipo de trabajo de cromatografía BioCAD Sprint (Perseptive Biosystems, Cambridge, MA) mediante el protocolo siguiente. Se equilibró una columna Poros 20 HP2 10 que era de una columna de cromatografía de interacción hidrofóbica de Perseptive Biosystems (Cambridge, MA) con 5 volúmenes de columna de (NH₄)₂SO₄ 0,5 M/NaH₂PO₄ 0,02M, pH 6,80. Se determinó la concentración de proteína total en la muestra de sobrenadante utilizando un kit de ensayo de proteínas Bio-Rad (Hercules, CA) según las instrucciones de los fabricantes y se aplicó a la columna un 20% de la capacidad de la columna (20 mg/ml). Se lavó la columna con 10 volúmenes de columna de (NH₄)₂SO₄ 0,5 M/NaH₂PO₄ 0,02M, pH 6,80. Se eluyó la proteína Cip1 con 5 volúmenes de columna de NaH₂PO₄ 0,02M, pH 6,80. En este punto, Cip1 era aproximadamente un 70% pura. Se concentró el eluato hasta 13 ml mediante ultrafiltración utilizando unidades de filtración centrífuga con un límite de peso molecular nominal de 5.000 (Biomax 5K; Millipore, Bedford MA). Se equilibró una columna de filtración en gel (Superdex 75, Amersham Biosciences) con dos volúmenes de columna de NaH₂PO₄ 0,02M, pH 6,80 y se aplicó el eluato concentrado de la columna anterior. Se recogieron fracciones y se analizaron por el peso molecular de las proteínas mediante SDS-PAGE y por la actividad contra p-nitrofenil-β-D-celobiosido (p-NPC). En este punto la proteína Cip1 era pura en más de un 95% .

[0273] Para ensayos de p-NPC se mezclaron 20 μl de p-NPC (7,5 mg/ml) con 10 μl de muestra y 100 μl de acetato sódico 20 mM, pH 5,0. Después de la incubación a 50°C durante 30 minutos se paró la reacción mediante la adición de 100 μl de glicina 100 mM, pH 10. Se midió la densidad óptica a una longitud de onda de 405 nm. Aunque no se determinó una actividad específica era obvio que Cip1 tenía actividad contra p-NPC. En un experimento, la reacción de base sin enzima añadida produjo una DO405 de 0,071, mientras que con Cip1 la DO405 fue de 0,121. Esto demuestra que la proteína Cip1 tiene cierta actividad en un sustrato utilizado habitualmente para medir la actividad

de celulasas (tanto endoglucanasas como celobiohidrolasas).

[0274] Aunque se ha descrito la invención anterior con cierto detalle a modo de ilustración y ejemplo con propósitos de claridad y comprensión, será obvio que pueden realizarse ciertos cambios y modificaciones en el alcance de las reivindicaciones adjuntas.

5

ES 2 371 362 T3

LISTADO DE SECUENCIAS

<110> GENENCOR INTERNATIONAL, INC.

5 <120> Nuevos genes de Trichoderma

<130> GRF/FP6343529

<140> 04753666.9

10 <141> 2004-05-28

<150> US 60/475,826

<151> 2003-06-03

15 <150> US 60/474,411

<151> 2003-05-29

<160> 59

20 <170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1221

<212> ADN

25 <213> Trichoderma reesei

<400> 1

30	gactagttca taatacagta gttgagttca tagcaacttc actctctagc tgaacaaatt	60
	atctgcgcaa acatggttcg ccggactgct ctgctggccc ttggggctct ctcaacgctc	120
	tctatggccc aaatctcaga cgacttcgag tccggctggg atcagactaa atggccatt	180
	tcggcaccag actgtaacca gggcggcacc gtcagcctcg acaccacagt agcccacagc	240
	ggcagcaact ccatgaaggt cgttgggtggc cccaatggct actgtggaca catcttcttc	300
	ggcactacc aggtgccaac tggggatgta tatgtcagag cttggattcg gcttcagact	360
	gctctcggca gcaaccagct cacattcatc atcatgccag acaccgctca gggagggag	420
35	cacctccgaa ttggtggcca aagccaagtt ctcgactaca accgcgagtc cgacgatgcc	480
	actcttccgg acctgtctcc caacggcatt gcctccaccg tcaactctgcc taccggcgcg	540
	ttccagtgtc tcgagtacca cctgggcact gacggaacca tcgagacgtg gctcaacggc	600
	agcctcatcc cgggcatgac cgtgggcccct ggcgtcgaca atccaaacga cgctggctgg	660
	acgagggcca gctatattcc ggagatcacc ggtgtcaact ttggctggga ggcctacagc	720
40	ggagacgtca acaccgtctg gttcgacgac atctcgattg cgtcgaccg cgtgggatgc	780
	ggccccgca gccccggcg tcttgggaag tcgacgactg ggcgtagcag cacctcgggc	840
	ccgacgagca cttcgaggcc aagcaccacc attccgccac cgacttccag gacaacgacc	900
	gccacgggtc cgactcagac aactatggc cagtgcggag ggattgggta cagcggcct	960
	acggtctgcg cgagcggcac gacctgccag gtcctgaacc catactactc ccagtgttta	1020
45	taaggggatg agcatggagt gaagtgaagt gaagtggaga gagttgaagt ggcattgcgc	1080
	tgggctgggt agataaaaagt cagcagctat gaatactcta tgtgatgctc attggcgtgt	1140
	acgttttaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	1200
	aaaaaaaaaag ggggcggccc c	1221

50 <210> 2

<211> 951

<212> ADN

<213> Trichoderma reesei

55 <400> 2

60	atggttcgcc ggactgctct gctggccctt ggggctctct caacgctctc tatggcccaa	60
	atctcagacg acttcgagtc gggctgggat cagactaaat ggccatttc ggcaccagac	120
	tgtaaccagg gcggcaccgt cagcctcgac accacagtag cccacagcgg cagcaactcc	180
	atgaaggtcg ttggtggccc caatggctac tgtggacaca tcttcttcgg cactaccag	240
	gtgccaaactg gggatgtata tgtcagagct tggattcggc ttcagactgc tctcggcagc	300
	aaccacgtca cattcatcat catgccagac accgctcagg gagggaagca cctccgaatt	360

ES 2 371 362 T3

5 ggtggccaaa gccaaagtct cgactacaac cgcgagtccg acgatgccac tcttccggac 420
 ctgtctccca acggcattgc ctccaccgtc actctgccta ccggcgcggt ccagtgcctc 480
 gagtaccacc tgggcactga cggaaaccatc gagacgtggc tcaacggcag cctcatcccg 540
 ggcatgaccg tgggcccctgg cgtcgacaat ccaaacgacg ctggctggac gagggccagc 600
 5 tatattccgg agatcaccgg tgtcaacttt ggctgggagg cctacagcgg agacgtcaac 660
 accgtctggt tcgacgacat ctcgattgcg tcgacccgcg tgggatgcgg ccccggcagc 720
 cccggcggtc ctggaagctc gacgactggg cgtagcagca cctcgggccc gacgagcact 780
 tcgaggccaa gcaccacat tccgccaccg acttccagga caacgaccgc cacgggtccg 840
 10 actcagacac actatggcca gtgctggagg attggttaca gcgggcctac ggtctgcgcg 900
 agcggcacga cctgcccagg cctgaacca tactactccc agtgcttata a 951

<210> 3

<211> 316

<212> PRT

15 <213> Trichoderma reesei

<400> 3

20 Met Val Arg Arg Thr Ala Leu Leu Ala Leu Gly Ala Leu Ser Thr Leu
 1 Ser Met Ala Gln Ile Ser Asp Asp Phe Glu Ser Gly Trp Asp Gln Thr
 Lys Trp Pro Ile Ser Ala Pro Asp Cys Asn Gln Gly Gly Thr Val Ser
 25 Leu Asp Thr Thr Val Ala His Ser Gly Ser Asn Ser Met Lys Val Val
 50 Gly Gly Pro Asn Gly Tyr Cys Gly His Ile Phe Phe Gly Thr Thr Gln
 65 Val Pro Thr Gly Asp Val Tyr Val Arg Ala Trp Ile Arg Leu Gln Thr
 80 Ala Leu Gly Ser Asn His Val Thr Phe Ile Ile Met Pro Asp Thr Ala
 100 Gln Gly Gly Lys His Leu Arg Ile Gly Gly Gln Ser Gln Val Leu Asp
 115 Tyr Asn Arg Glu Ser Asp Asp Ala Thr Leu Pro Asp Leu Ser Pro Asn
 130 Gly Ile Ala Ser Thr Val Thr Leu Pro Thr Gly Ala Phe Gln Cys Phe
 145 Glu Tyr His Leu Gly Thr Asp Gly Thr Ile Glu Thr Trp Leu Asn Gly
 165 Ser Leu Ile Pro Gly Met Thr Val Gly Pro Gly Val Asp Asn Pro Asn
 180 Asp Ala Gly Trp Thr Arg Ala Ser Tyr Ile Pro Glu Ile Thr Gly Val
 195 Asn Phe Gly Trp Glu Ala Tyr Ser Gly Asp Val Asn Thr Val Trp Phe
 210 Asp Asp Ile Ser Ile Ala Ser Thr Arg Val Gly Cys Gly Pro Gly Ser
 225 Pro Gly Gly Pro Gly Ser Ser Thr Thr Gly Arg Ser Ser Thr Ser Gly
 245 Pro Thr Ser Thr Ser Arg Pro Ser Thr Thr Ile Pro Pro Pro Thr Ser
 260 Arg Thr Thr Thr Ala Thr Gly Pro Thr Gln Thr His Tyr Gly Gln Cys
 275 Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys Ala Ser Gly Thr Thr
 290 Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys Leu
 305 310 315

<210> 4

60 <211> 19

<212> PRT

<213> Trichoderma reesei

ES 2 371 362 T3

<400> 4
 Met Val Arg Arg Thr Ala Leu Leu Ala Leu Gly Ala Leu Ser Thr Leu
 1 5 10 15
 5 Ser Met Ala

<210> 5
 <211> 297
 <212> PRT
 10 <213> Trichoderma reesei

<400> 5
 Gln Ile Ser Asp Asp Phe Glu Ser Gly Trp Asp Gln Thr Lys Trp Pro
 1 5 10 15
 15 Ile Ser Ala Pro Asp Cys Asn Gln Gly Gly Thr Val Ser Leu Asp Thr
 20 Thr Val Ala His Ser Gly Ser Asn Ser Met Lys Val Val Gly Gly Pro
 25 Asn Gly Tyr Cys Gly His Ile Phe Phe Gly Thr Thr Gln Val Pro Thr
 30 Gly Asp Val Tyr Val Arg Ala Trp Ile Arg Leu Gln Thr Ala Leu Gly
 35 Ser Asn His Val Thr Phe Ile Ile Met Pro Asp Thr Ala Gln Gly Gly
 40 Lys His Leu Arg Ile Gly Gly Gln Ser Gln Val Leu Asp Tyr Asn Arg
 45 Glu Ser Asp Asp Ala Thr Leu Pro Asp Leu Ser Pro Asn Gly Ile Ala
 50 Ser Thr Val Thr Leu Pro Thr Gly Ala Phe Gln Cys Phe Glu Tyr His
 55 Leu Gly Thr Asp Gly Thr Ile Glu Thr Trp Leu Asn Gly Ser Leu Ile
 60 Pro Gly Met Thr Val Gly Pro Gly Val Asp Asn Pro Asn Asp Ala Gly
 65 Trp Thr Arg Ala Ser Tyr Ile Pro Glu Ile Thr Gly Val Asn Phe Gly
 70 Trp Glu Ala Tyr Ser Gly Asp Val Asn Thr Val Trp Phe Asp Asp Ile
 75 Ser Ile Ala Ser Thr Arg Val Gly Cys Gly Pro Gly Ser Pro Gly Gly
 80 Pro Gly Ser Ser Thr Thr Gly Arg Ser Ser Thr Ser Gly Pro Thr Ser
 85 Thr Ser Arg Pro Ser Thr Thr Ile Pro Pro Thr Ser Arg Thr Thr
 90 Thr Ala Thr Gly Pro Thr Gln Thr His Tyr Gly Gln Cys Gly Gly Ile
 95 Gly Tyr Ser Gly Pro Thr Val Cys Ala Ser Gly Thr Thr Cys Gln Val
 100 Leu Asn Pro Tyr Tyr Ser Gln Cys Leu
 105 290 295

<210> 6
 <211> 1383
 <212> ADN
 55 <213> Trichoderma reesei

<400> 6
 atggccttccc gcttctttg ctttctcctt tttagcgatcc caatccaggc ccaatctcca 60
 gtctgggggac aatgtggtgg aattggttgg tctggcccaa caacttggtg tggagggtgcg 120
 60 acttggtgat catataacc ttattactcg caatgtattc ccagtacaca ggcttcatcg 180
 agcatagcct ctacaacgct ggtcacatca tttacgacca ccactgctac gaggacttcg 240
 gcatcaacgc ctccagcgag cagtacaggt gcaggcgcg caacatgctc agcactgccc 300

ES 2 371 362 T3

```

ggctccatta ccctgagatc caacgcaaag ctcaacgatc tgtttacaat gttcaatgga 360
gataaggatc ccacgaaaga caaattctcg tgccgccagg cagagatgtc ggagctaata 420
caacgatatg agctcggcac cctgcccgga cgaccaagca ctctcacagc ctctattctcg 480
ggcaatacgt tgaccatcaa ttgctggagag gccggaaaagt caatttcatt cacagtcacg 540
5 atcactttatc catctttccgg aacagcacca taccctgcga ttatcggcta tggaggcggc 600
agtctttccag ctcccgccgg ggttgccatg atcaacttta acaatgacaa catagcagcc 660
caagttaata caggcagccg cggacagggc aagttctacg atctctacgg gagctcgcac 720
tccgcggggcg ccatgaccgc atgggcctgg ggagtaagcc gagtcattga tgctcttgag 780
cttgaccag gcgcaagaat agacaccacc aagattggcg tgacggggtg ttcacgaaat 840
10 ggcaaaggcg caatggctcg agggtgctttc gagaaacgaa tcgttctgac acttccccag 900
gagtcggggcg ccggtggctc tgcgtgctgg aggatttcag actacttaaa gtcccaagga 960
gccaatatcc agaccgcgct tgagatcatt ggcgaagacc cctggttctc gactactttc 1020
aacagctacg tcaaccaagt gccgggtgtg ccgtttgacc accattcgct tgctgccttg 1080
atagccccga gaggattatt cgatcatcgac aacaatattg actggctcgg cccacaaagc 1140
15 tgctttggct gtatgacagc tgctcacatg gcatggcaag ctttgggtgt ctcggaccac 1200
atgggctatt cgcagattgg agctcacgca cactgcgcgt tcccatcaaa ccagcaatcg 1260
caacttactg cttttgttca gaaattcttg ctgggccagt ccacaaatac ggcgattttc 1320
caaagcgact tttcggccaa tcaaagccaa tggatcgact ggacaacccc aacgctgagt 1380
tga 1383

<210> 7
<211> 460
<212> PRT
<213> Trichoderma reesei

25
<400> 7
Met Ala Ser Arg Phe Phe Ala Leu Leu Leu Leu Ala Ile Pro Ile Gln
1 5 10 15
Ala Gln Ser Pro Val Trp Gly Gln Cys Gly Gly Ile Gly Trp Ser Gly
30 20 25 30
Pro Thr Thr Cys Val Gly Gly Ala Thr Cys Val Ser Tyr Asn Pro Tyr
35 35 40 45
Tyr Ser Gln Cys Ile Pro Ser Thr Gln Ala Ser Ser Ser Ile Ala Ser
50 55 60
35 Thr Thr Leu Val Thr Ser Phe Thr Thr Thr Ala Thr Arg Thr Ser
65 70 75 80
Ala Ser Thr Pro Pro Ala Ser Ser Thr Gly Ala Gly Gly Ala Thr Cys
85 90 95
40 Ser Ala Leu Pro Gly Ser Ile Thr Leu Arg Ser Asn Ala Lys Leu Asn
100 105 110
Asp Leu Phe Thr Met Phe Asn Gly Asp Lys Val Thr Thr Lys Asp Lys
115 120 125
Phe Ser Cys Arg Gln Ala Glu Met Ser Glu Leu Ile Gln Arg Tyr Glu
130 135 140
45 Leu Gly Thr Leu Pro Gly Arg Pro Ser Thr Leu Thr Ala Ser Phe Ser
145 150 155 160
Gly Asn Thr Leu Thr Ile Asn Cys Gly Glu Ala Gly Lys Ser Ile Ser
165 170 175
50 Phe Thr Val Thr Ile Thr Tyr Pro Ser Gly Thr Ala Pro Tyr Pro
180 185 190
Ala Ile Ile Gly Tyr Gly Gly Gly Ser Leu Pro Ala Pro Ala Gly Val
195 200 205
Ala Met Ile Asn Phe Asn Asn Asp Asn Ile Ala Ala Gln Val Asn Thr
210 215 220
55 Gly Ser Arg Gly Gln Gly Lys Phe Tyr Asp Leu Tyr Gly Ser Ser His
225 230 235 240
Ser Ala Gly Ala Met Thr Ala Trp Ala Trp Gly Val Ser Arg Val Ile
245 250 255
60 Asp Ala Leu Glu Leu Val Pro Gly Ala Arg Ile Asp Thr Thr Lys Ile
260 265 270
Gly Val Thr Gly Cys Ser Arg Asn Gly Lys Gly Ala Met Val Ala Gly
275 280 285

```

ES 2 371 362 T3

Ala Phe Glu Lys Arg Ile Val Leu Thr Leu Pro Gln Glu Ser Gly Ala
 290 295 300
 Gly Gly Ser Ala Cys Trp Arg Ile Ser Asp Tyr Leu Lys Ser Gln Gly
 305 310 315 320
 5 Ala Asn Ile Gln Thr Ala Ser Glu Ile Ile Gly Glu Asp Pro Trp Phe
 325 330 335
 Ser Thr Thr Phe Asn Ser Tyr Val Asn Gln Val Pro Val Leu Pro Phe
 340 345 350
 10 Asp His His Ser Leu Ala Ala Leu Ile Ala Pro Arg Gly Leu Phe Val
 355 360 365
 Ile Asp Asn Asn Ile Asp Trp Leu Gly Pro Gln Ser Cys Phe Gly Cys
 370 375 380
 15 Met Thr Ala Ala His Met Ala Trp Gln Ala Leu Gly Val Ser Asp His
 385 390 395 400
 Met Gly Tyr Ser Gln Ile Gly Ala His Ala His Cys Ala Phe Pro Ser
 405 410 415
 20 Asn Gln Gln Ser Gln Leu Thr Ala Phe Val Gln Lys Phe Leu Leu Gly
 420 425 430 435
 Gln Ser Thr Asn Thr Ala Ile Phe Gln Ser Asp Phe Ser Ala Asn Gln
 440 445 450
 Ser Gln Trp Ile Asp Trp Thr Thr Pro Thr Leu Ser
 455 460

<210> 8
 <211> 17
 <212> PRT
 <213> Trichoderma reesei

<400> 8
 30 Met Ala Ser Arg Phe Phe Ala Leu Leu Leu Leu Ala Ile Pro Ile Gln
 1 5 10 15
 Ala

<210> 9
 <211> 443
 <212> PRT
 <213> Trichoderma reesei

<400> 9
 40 Gln Ser Pro Val Trp Gly Gln Cys Gly Gly Ile Gly Trp Ser Gly Pro
 1 5 10 15
 Thr Thr Cys Val Gly Gly Ala Thr Cys Val Ser Tyr Asn Pro Tyr Tyr
 20 25 30
 45 Ser Gln Cys Ile Pro Ser Thr Gln Ala Ser Ser Ser Ile Ala Ser Thr
 35 40 45
 Thr Leu Val Thr Ser Phe Thr Thr Thr Thr Ala Thr Arg Thr Ser Ala
 50 55 60 65
 Ser Thr Pro Pro Ala Ser Ser Thr Gly Ala Gly Gly Ala Thr Cys Ser
 70 75 80
 55 Ala Leu Pro Gly Ser Ile Thr Leu Arg Ser Asn Ala Lys Leu Asn Asp
 85 90 95
 Leu Phe Thr Met Phe Asn Gly Asp Lys Val Thr Thr Lys Asp Lys Phe
 100 105 110
 Ser Cys Arg Gln Ala Glu Met Ser Glu Leu Ile Gln Arg Tyr Glu Leu
 115 120 125
 60 Gly Thr Leu Pro Gly Arg Pro Ser Thr Leu Thr Ala Ser Phe Ser Gly
 130 135 140
 Asn Thr Leu Thr Ile Asn Cys Gly Glu Ala Gly Lys Ser Ile Ser Phe
 145 150 155 160
 Thr Val Thr Ile Thr Tyr Pro Ser Ser Gly Thr Ala Pro Tyr Pro Ala
 165 170 175
 Ile Ile Gly Tyr Gly Gly Gly Ser Leu Pro Ala Pro Ala Gly Val Ala

ES 2 371 362 T3

5 Met Ile Asn Phe Asn Asn Asp Asn Ile Ala Ala Gln Val Asn Thr Gly
 Ser Arg Gly Gln Gly Lys Phe Tyr Asp Leu Tyr Gly Ser Ser His Ser
 Ala Gly Ala Met Thr Ala Trp Ala Trp Gly Val Ser Arg Val Ile Asp
 Ala Leu Glu Leu Val Pro Gly Ala Arg Ile Asp Thr Thr Lys Ile Gly
 10 Val Thr Gly Cys Ser Arg Asn Gly Lys Gly Ala Met Val Ala Gly Ala
 Phe Glu Lys Arg Ile Val Leu Thr Leu Pro Gln Glu Ser Gly Ala Gly
 Gly Ser Ala Cys Trp Arg Ile Ser Asp Tyr Leu Lys Ser Gln Gly Ala
 15 Asn Ile Gln Thr Ala Ser Glu Ile Ile Gly Glu Asp Pro Trp Phe Ser
 Thr Thr Phe Asn Ser Tyr Val Asn Gln Val Pro Val Leu Pro Phe Asp
 20 His His Ser Leu Ala Ala Leu Ile Ala Pro Arg Gly Leu Phe Val Ile
 Asp Asn Asn Ile Asp Trp Leu Gly Pro Gln Ser Cys Phe Gly Cys Met
 Thr Ala Ala His Met Ala Trp Gln Ala Leu Gly Val Ser Asp His Met
 25 Gly Tyr Ser Gln Ile Gly Ala His Ala His Cys Ala Phe Pro Ser Asn
 Gln Gln Ser Gln Leu Thr Ala Phe Val Gln Lys Phe Leu Leu Gly Gln
 30 Ser Thr Asn Thr Ala Ile Phe Gln Ser Asp Phe Ser Ala Asn Gln Ser
 Gln Trp Ile Asp Trp Thr Thr Pro Thr Leu Ser

35 <210> 10
 <211> 966
 <212> ADN
 <213> Trichoderma reesei

40 <400> 10
 atggagctta aagcactcag tgccgttgtg ctgagctttg taactcttgt cgcggcagca 60
 ccggcgacct gcacgcttcc gtccacatac cgctggaatt cgaccggtgc tttagccagc 120
 ccgaaatcag gctgggtctc gctgaaagac tictcccatg tcatttataa tggccagcat 180
 cttgtatggg gctcagacta tgacacagga acaatctggg gttcaatgaa ctttggctctg 240
 45 ttcagtgact ggtccaatat ggcaacggca agccagaaca aaatgactcc cggcactggt 300
 gctcctaccg tcttctactt tgccccgaag aatatttggg tactcgccta tcaatggggc 360
 ccgaccacgt tttcctacct gacgtcaagc aaccctcca gcgtcaatgg atggtcgtca 420
 ccacagcctc tcttctccgg cagtatctca ggctccagcc cgctggatca gacggtcatt 480
 ggcgacagca cgaacatgta tctgttcttc gcgggggacg acgggaaaat ctacagggcg 540
 50 agcatgccta tcggttaact ccccggaagc ttcggttcga cgtcaacggt ggtcctgagc 600
 gatgaaagga acaatctggt tgaggcagtt caggtctata ccgtctcagg gcagaagcaa 660
 tatctcatga ttgtcgaggc aataggcgca aatggccggt atttccggtc cttcacagcg 720
 acaaacctcg gcggcacatg gactccgcaa gccaccagcg aaagtcagcc gtttgccggt 780
 aaggcaaaca gtggcgctac ctggacaaac gacatcagtc atggtgatct aattcgtagc 840
 55 aaccctgac agacaatgac tatcgacctc tgcaatctgc agttcttcta ccaggggaga 900
 gcgacaaact ctggcggcga ctacggcctc ttgccctatc gaccagggct gctaactctc 960
 cagcgc 966

60 <210> 11
 <211> 322
 <212> PRT
 <213> Trichoderma reesei

ES 2 371 362 T3

<400> 11
 Met Glu Leu Lys Ala Leu Ser Ala Val Val Leu Ser Phe Val Thr Leu
 1 5 10 15
 5 Val Ala Ala Ala Pro Ala Thr Cys Thr Leu Pro Ser Thr Tyr Arg Trp
 20 25 30
 Asn Ser Thr Gly Ala Leu Ala Ser Pro Lys Ser Gly Trp Val Ser Leu
 35 40 45
 10 Lys Asp Phe Ser His Val Ile Tyr Asn Gly Gln His Leu Val Trp Gly
 50 55 60
 65 Ser Thr His Asp Thr Gly Thr Ile Trp Gly Ser Met Asn Phe Gly Leu
 70 75 80
 Phe Ser Asp Trp Ser Asn Met Ala Thr Ala Ser Gln Asn Lys Met Thr
 85 90 95
 15 Pro Gly Thr Val Ala Pro Thr Val Phe Tyr Phe Ala Pro Lys Asn Ile
 100 105 110
 Trp Val Leu Ala Tyr Gln Trp Gly Pro Thr Thr Phe Ser Tyr Leu Thr
 115 120 125
 20 Ser Ser Asn Pro Ser Ser Val Asn Gly Trp Ser Ser Pro Gln Pro Leu
 130 135 140
 Phe Ser Gly Ser Ile Ser Gly Ser Ser Pro Leu Asp Gln Thr Val Ile
 145 150 155 160
 Gly Asp Ser Thr Asn Met Tyr Leu Phe Phe Ala Gly Asp Asp Gly Lys
 165 170 175
 25 Ile Tyr Arg Ala Ser Met Pro Ile Gly Asn Phe Pro Gly Ser Phe Gly
 180 185 190
 Ser Thr Ser Thr Val Val Leu Ser Asp Glu Arg Asn Asn Leu Phe Glu
 195 200 205
 30 Ala Val Gln Val Tyr Thr Val Ser Gly Gln Lys Gln Tyr Leu Met Ile
 210 215 220
 Val Glu Ala Ile Gly Ala Asn Gly Arg Tyr Phe Arg Ser Phe Thr Ala
 225 230 235 240
 Thr Asn Leu Gly Gly Thr Trp Thr Pro Gln Ala Thr Ser Glu Ser Gln
 245 250 255
 35 Pro Phe Ala Gly Lys Ala Asn Ser Gly Ala Thr Trp Thr Asn Asp Ile
 260 265 270
 Ser His Gly Asp Leu Ile Arg Ser Asn Pro Asp Gln Thr Met Thr Ile
 275 280 285
 40 Asp Pro Cys Asn Leu Gln Phe Leu Tyr Gln Gly Arg Ala Thr Asn Ser
 290 295 300
 Gly Gly Asp Tyr Gly Leu Leu Pro Tyr Arg Pro Gly Leu Leu Thr Leu
 305 310 315 320
 Gln Arg

45 <210> 12
 <211> 19
 <212> PRT
 <213> Trichoderma reesei

50 <400> 12
 Met Glu Leu Lys Ala Leu Ser Ala Val Val Leu Ser Phe Val Thr Leu
 1 5 10 15
 Val Ala Ala

55 <210> 13
 <211> 303
 <212> PRT
 <213> Trichoderma reesei

60 <400> 13
 Ala Pro Ala Thr Cys Thr Leu Pro Ser Thr Tyr Arg Trp Asn Ser Thr
 1 5 10 15

ES 2 371 362 T3

Gly Ala Leu Ala Ser Pro Lys Ser Gly Trp Val Ser Leu Lys Asp Phe
 Ser His Val Ile Tyr Asn Gly Gln His Leu Val Trp Gly Ser Thr His
 5 Asp Thr Gly Thr Ile Trp Gly Ser Met Asn Phe Gly Leu Phe Ser Asp
 Trp Ser Asn Met Ala Thr Ala Ser Gln Asn Lys Met Thr Pro Gly Thr
 65 Val Ala Pro Thr Val Phe Tyr Phe Ala Pro Lys Asn Ile Trp Val Leu
 10 Ala Tyr Gln Trp Gly Pro Thr Thr Phe Ser Tyr Leu Thr Ser Ser Asn
 Pro Ser Ser Val Asn Gly Trp Ser Ser Pro Gln Pro Leu Phe Ser Gly
 15 Ser Ile Ser Gly Ser Ser Pro Leu Asp Gln Thr Val Ile Gly Asp Ser
 Thr Asn Met Tyr Leu Phe Ala Gly Asp Asp Gly Lys Ile Tyr Arg
 145 Ala Ser Met Pro Ile Gly Asn Phe Pro Gly Ser Phe Gly Ser Thr Ser
 20 Thr Val Val Leu Ser Asp Glu Arg Asn Asn Leu Phe Glu Ala Val Gln
 Val Tyr Thr Val Ser Gly Gln Lys Gln Tyr Leu Met Ile Val Glu Ala
 195 Ile Gly Ala Asn Gly Arg Tyr Phe Arg Ser Phe Thr Ala Thr Asn Leu
 25 Gly Gly Thr Trp Thr Pro Gln Ala Thr Ser Glu Ser Gln Pro Phe Ala
 225 Gly Lys Ala Asn Ser Gly Ala Thr Trp Thr Asn Asp Ile Ser His Gly
 30 Asp Leu Ile Arg Ser Asn Pro Asp Gln Thr Met Thr Ile Asp Pro Cys
 Asn Leu Gln Phe Leu Tyr Gln Gly Arg Ala Thr Asn Ser Gly Gly Asp
 275 Tyr Gly Leu Leu Pro Tyr Arg Pro Gly Leu Leu Thr Leu Gln Arg
 290 300

<210> 14
 <211> 900
 <212> ADN
 <213> Trichoderma reesei

<400> 14
 45 atgCGcGccc tctcactctc cctccccctc tccctctcgc tgctcGccgc cagctcaaca 60
 gCGgcaacga catgCGcaaa gggcctctac atggctGttg cccGcGgcag cGaggagccc 120
 gccGgCacgg gcgtgacggg caacctcAcg agccaaatcG cCGcaaaggT gccCGgcagc 180
 gaggtcGtgG cGgtgGacta cccGgCCagc tttgacgact acgaggattc cGagggcGac 240
 gGcgtcaagg cGatgCggca gctGctcaac agctacGccg agGcctgtcc gGgaaacaag 300
 attGcGgtGc tgggatactc tcagggcGcc caagtcGcaa cagacaccat ctGcGgcGgt 360
 50 gccGgCgatc cGtttaccag cGacaagggc atgtctgacg atgtcatgga cGacgtcGtt 420
 gccGtggcca ttttcGgaga cccaaccat gtcGccaaca tgacgtacga cCGaggcacc 480
 agcattcaca acgggctctt caaccGgagc tcgtccagca tcGaggctG caagtGgtac 540
 gccagccGca tcgtctcGta ctGcGacagc gGcGacatct actGcGacGc cGgCagcaac 600
 tcGaccgttc accacatgta catccagcGc tacGgcGacg aaatcGtcga ctttGtcGtc 660
 55 agccagtttG agaagagcac cagctcGgga tcGgggGtcGg gtactaatGc caccagacc 720
 acggctccGg ctcccaccgt gtctcctacc accaccagcG gtggcaacag cacagtGcct 780
 acGcgaaccg gtggcccGac gacgagttcG acGcaaggat cGggtgCgag tgctttgacg 840
 agcagtttga tgctgggagg tcttttgacg gttttgacGg cGgtgtctca gatGctgtga 900

60 <210> 15
 <211> 299
 <212> PRT

ES 2 371 362 T3

<213> Trichoderma reesei

<400> 15

5 Met Arg Ala Leu Ser Leu Ser Leu Pro Leu Ser Leu Ser Leu Leu Ala
 1 5 10 15
 Ala Ser Ser Thr Ala Ala Thr Thr Cys Ala Lys Gly Leu Tyr Met Val
 20 25 30
 Val Ala Arg Gly Ser Glu Glu Pro Ala Gly Thr Gly Val Thr Gly Asn
 35 40 45
 10 Leu Thr Ser Gln Ile Ala Ala Lys Val Pro Gly Ser Glu Val Val Ala
 50 55 60
 Val Asp Tyr Pro Ala Ser Phe Asp Asp Tyr Glu Asp Ser Glu Gly Asp
 65 70 75 80
 Gly Val Lys Ala Met Arg Gln Leu Leu Asn Ser Tyr Ala Glu Ala Cys
 85 90 95
 15 Pro Gly Asn Lys Ile Ala Val Leu Gly Tyr Ser Gln Gly Ala Gln Val
 100 105 110
 Ala Thr Asp Thr Ile Cys Gly Gly Ala Gly Asp Pro Phe Thr Ser Asp
 115 120 125
 20 Lys Gly Met Ser Asp Asp Val Met Asp Asp Val Val Ala Val Ala Ile
 130 135 140
 Phe Gly Asp Pro Thr His Val Ala Asn Met Thr Tyr Asp Arg Gly Thr
 145 150 155 160
 25 Ser Ile His Asn Gly Leu Phe Asn Arg Ser Ser Ser Ile Glu Val
 165 170 175
 Cys Lys Ser Tyr Ala Ser Arg Ile Val Ser Tyr Cys Asp Thr Gly Asp
 180 185 190
 Ile Tyr Cys Asp Ala Gly Ser Asn Ser Thr Val His His Met Tyr Ile
 195 200 205
 30 Gln Arg Tyr Gly Asp Glu Ile Val Asp Phe Val Val Ser Gln Phe Glu
 210 215 220
 Lys Ser Thr Ser Ser Gly Ser Gly Ser Gly Thr Asn Ala Thr Thr Thr
 225 230 235 240
 35 Thr Ala Pro Ala Pro Thr Val Ser Pro Thr Thr Thr Ser Gly Gly Asn
 245 250 255
 Ser Thr Val Pro Thr Arg Thr Gly Gly Pro Thr Thr Ser Ser Thr Gln
 260 265 270
 Gly Ser Gly Ala Ser Ala Leu Thr Ser Ser Leu Met Leu Gly Gly Leu
 275 280 285
 40 Leu Thr Val Leu Thr Ala Val Ser Gln Met Leu
 290 295

<210> 16

<211> 21

45 <212> PRT

<213> Trichoderma reesei

<400> 16

50 Met Arg Ala Leu Ser Leu Ser Leu Pro Leu Ser Leu Ser Leu Leu Ala
 1 5 10 15
 Ala Ser Ser Thr Ala
 20

<210> 17

<211> 278

<212> PRT

<213> Trichoderma reesei

<400> 17

60 Ala Thr Thr Cys Ala Lys Gly Leu Tyr Met Val Val Ala Arg Gly Ser
 1 5 10 15
 Glu Glu Pro Ala Gly Thr Gly Val Thr Gly Asn Leu Thr Ser Gln Ile

ES 2 371 362 T3

				20					25				30			
	Ala	Ala	Lys	Val	Pro	Gly	Ser	Glu	Val	Val	Ala	Val	Asp	Tyr	Pro	Ala
			35					40					45			
5	Ser	Phe	Asp	Asp	Tyr	Glu	Asp	Ser	Glu	Gly	Asp	Gly	Val	Lys	Ala	Met
		50					55					60				
	Arg	Gln	Leu	Leu	Asn	Ser	Tyr	Ala	Glu	Ala	Cys	Pro	Gly	Asn	Lys	Ile
	65					70					75				80	
	Ala	Val	Leu	Gly	Tyr	Ser	Gln	Gly	Ala	Gln	Val	Ala	Thr	Asp	Thr	Ile
				85						90					95	
10	Cys	Gly	Gly	Ala	Gly	Asp	Pro	Phe	Thr	Ser	Asp	Lys	Gly	Met	Ser	Asp
				100					105					110		
	Asp	Val	Met	Asp	Asp	Val	Val	Ala	Val	Ala	Ile	Phe	Gly	Asp	Pro	Thr
			115					120					125			
	His	Val	Ala	Asn	Met	Thr	Tyr	Asp	Arg	Gly	Thr	Ser	Ile	His	Asn	Gly
		130					135					140				
15	Leu	Phe	Asn	Arg	Ser	Ser	Ser	Ser	Ile	Glu	Val	Cys	Lys	Ser	Tyr	Ala
	145					150					155					160
	Ser	Arg	Ile	Val	Ser	Tyr	Cys	Asp	Thr	Gly	Asp	Ile	Tyr	Cys	Asp	Ala
					165					170					175	
20	Gly	Ser	Asn	Ser	Thr	Val	His	His	Met	Tyr	Ile	Gln	Arg	Tyr	Gly	Asp
				180					185					190		
	Glu	Ile	Val	Asp	Phe	Val	Val	Ser	Gln	Phe	Glu	Lys	Ser	Thr	Ser	Ser
			195					200						205		
25	Gly	Ser	Gly	Ser	Gly	Thr	Asn	Ala	Thr	Thr	Thr	Thr	Ala	Pro	Ala	Pro
		210					215					220				
	Thr	Val	Ser	Pro	Thr	Thr	Thr	Ser	Gly	Gly	Asn	Ser	Thr	Val	Pro	Thr
	225					230					235				240	
	Arg	Thr	Gly	Gly	Pro	Thr	Thr	Ser	Ser	Thr	Gln	Gly	Ser	Gly	Ala	Ser
					245					250					255	
30	Ala	Leu	Thr	Ser	Leu	Met	Leu	Gly	Gly	Leu	Leu	Thr	Val	Leu	Thr	
				260				265					270			
	Ala	Val	Ser	Gln	Met	Leu										
			275													
35	<210> 18															
	<211> 768															
	<212> PRT															
	<213> Ruminococcus flavefaciens															
40	<400> 18															
	Met	Lys	Lys	His	Phe	Val	Val	Gly	Glu	Thr	Ile	Lys	Arg	Phe	Leu	Arg
	1				5					10					15	
	Ile	Gly	Thr	Ser	Leu	Ala	Leu	Ser	Ile	Ser	Thr	Leu	Ser	Leu	Leu	Pro
				20					25					30		
45	Ser	Ala	Pro	Arg	Leu	Ser	Ser	Ala	Ala	Gly	Thr	Ile	Lys	Ile	Met	Pro
			35					40					45			
	Leu	Gly	Asp	Ser	Ile	Thr	Tyr	Gly	Met	Ala	Asp	Glu	Gly	Gly	Tyr	Arg
		50					55					60				
50	Lys	Tyr	Leu	Ser	Tyr	Phe	Leu	Gln	Gln	Lys	Gly	Tyr	Thr	Asn	Val	Asp
	65					70					75					80
	Leu	Val	Gly	Pro	Glu	Gly	Lys	Asp	Ser	Ala	Ser	Phe	Asn	Tyr	Asn	Gly
					85					90					95	
	Gln	Ser	Val	Lys	Tyr	Asp	Asp	Asn	His	Ala	Gly	Tyr	Ser	Gly	Tyr	Thr
				100					105					110		
55	Ile	Thr	Asn	Leu	Pro	Gly	Gly	Trp	Phe	Gly	Gln	Leu	Asn	Gly	Ile	Leu
			115					120					125			
	Glu	Thr	Met	Gln	Gly	Gly	Asp	Tyr	Ile	Lys	Lys	Tyr	Ser	Pro	Asp	Ile
		130					135					140				
	Ile	Leu	Leu	Gln	Ile	Gly	Thr	Asn	Asp	Val	Ser	Asn	Gly	His	Leu	Asp
	145					150					155					160
60	Gly	Ser	Glu	Glu	Arg	Leu	His	Lys	Leu	Leu	Asp	Tyr	Leu	Arg	Glu	Asn
					165					170					175	

ES 2 371 362 T3

	Met	Pro	Ser	Asn	Gly	Lys	Val	Phe	Leu	Thr	Thr	Ile	Pro	Asp	Leu	Gly
				180					185					190		
	Asn	Ser	Gly	Trp	Gly	Gly	Asn	Ser	Asn	Gly	Asp	Ile	Ala	Lys	Tyr	Asn
			195				200					205				
5	Glu	Leu	Ile	Lys	Lys	Val	Ala	Asn	Asp	Tyr	Ser	Ser	Lys	Asn	Val	Ile
		210					215					220				
	Tyr	Ala	Asp	Ile	His	Ser	Val	Ile	Asp	Ala	Ser	Lys	Asp	Leu	Ala	Asp
	225				230						235					240
10	Gly	Val	His	Pro	Asn	Ala	Gly	Gly	Tyr	Glu	Lys	Met	Gly	Lys	Tyr	Trp
					245					250					255	
	Leu	Glu	Gln	Ile	Glu	Gly	Tyr	Leu	Lys	Ala	Ser	Asp	Gly	Pro	Gln	Gln
			260						265					270		
	Thr	Gln	Pro	Thr	Gln	Pro	Ser	Gln	Gly	Asp	Ser	Gly	Pro	Glu	Leu	Ile
			275					280					285			
15	Tyr	Gly	Asp	Leu	Asp	Gly	Asp	Lys	Thr	Ile	Thr	Ser	Phe	Asp	Ala	Val
		290					295					300				
	Ile	Met	Arg	Lys	Gly	Leu	Ile	Asn	Asp	Phe	Lys	Asp	Asn	Asn	Val	Lys
	305					310					315					320
	Lys	Ala	Ala	Asp	Ile	Asp	Gln	Asn	Gly	Lys	Ala	Glu	Val	Ala	Asp	Leu
					325					330					335	
20	Val	Gln	Leu	Gln	Ser	Phe	Ile	Ile	Gly	Lys	Ile	Lys	Glu	Phe	Thr	Val
			340						345					350		
	Ala	Glu	Lys	Thr	Val	Thr	Glu	Lys	Pro	Val	Phe	Glu	Lys	Ser	Tyr	Asn
			355					360					365			
25	Phe	Pro	Ala	Val	Asn	Gln	Leu	Lys	Ser	Ser	Lys	Asp	Ile	Pro	Asp	Pro
		370					375					380				
	Phe	Ile	Phe	Met	Asp	Gly	Ser	Lys	Val	Glu	Ser	Thr	Asp	Asp	Trp	Trp
	385					390					395					400
	Lys	Arg	Gln	Ser	Glu	Ile	Ser	Cys	Met	Tyr	Glu	Tyr	Tyr	Met	Tyr	Gly
					405					410					415	
30	Lys	Trp	Ile	Asp	Gly	Ser	Asp	Asp	Glu	Thr	Thr	Tyr	Ser	Ile	Ser	Gly
									425					430		
	Asn	Ser	Met	Thr	Ile	Asn	Val	Lys	Arg	Lys	Ser	Thr	Gly	Lys	Thr	Ala
			435					440					445			
35	Ser	Phe	Lys	Ala	Val	Ile	Asn	Leu	Pro	Lys	Asn	Val	Arg	His	Glu	Gly
							455					460				
	Gly	Ala	Pro	Val	Ile	Leu	Gly	Met	His	Lys	Gly	Ile	Ser	Glu	Ser	Thr
	465					470					475					480
40	Ala	Thr	Ser	Asn	Gly	Tyr	Ala	Val	Ile	Thr	Tyr	Asp	Ser	Asp	Gly	Met
					485					490					495	
	Phe	Ser	Ala	Pro	Gly	Thr	Ala	Gln	Asp	Asn	Asn	Gln	His	Lys	Gly	Ala
								505						510		
	Phe	Tyr	Asp	Leu	Tyr	Pro	Tyr	Gly	Arg	Asn	Trp	Asp	Glu	Gln	Thr	Gly
			515					520					525			
45	Asp	Leu	Met	Ala	Trp	Ser	Trp	Gly	Ile	Ser	Arg	Ile	Leu	Asp	Ala	Leu
		530					535					540				
	Tyr	Asn	Gly	Ala	Ala	Lys	Glu	Leu	Asn	Ile	Asn	Pro	Asp	Ser	Ser	Ile
	545					550					555					560
	Val	Thr	Gly	Val	Ser	Arg	Tyr	Gly	Lys	Ala	Ala	Ser	Val	Cys	Gly	Ala
					565					570					575	
50	Phe	Asp	Thr	Arg	Ile	Lys	Met	Cys	Ala	Pro	Ser	Cys	Ser	Gly	Ala	Gly
				580					585					590		
	Gly	Leu	Ala	Leu	Tyr	Arg	Tyr	Ser	Ser	Val	Gly	Lys	Thr	Tyr	Asp	Phe
								600					605			
55	Ser	Ser	Lys	Gly	Gly	Ser	Ser	Ser	Tyr	Thr	Tyr	Lys	Glu	Asn	Glu	Pro
		610					615					620				
	Leu	Gly	Ser	Leu	Gln	Ala	Ser	Gly	Glu	Gln	Gly	Trp	Phe	Asn	Gly	Arg
	625					630					635					640
60	Phe	Met	Glu	Phe	Arg	Asn	Ala	Glu	Gln	Phe	Pro	Met	Asp	Gln	His	Met
					645					650					655	
	Leu	Gly	Ala	Leu	Cys	Cys	Asp	Pro	Asp	Arg	Tyr	Leu	Phe	Ile	Ile	Gly
				660					665					670		

ES 2 371 362 T3

Ser Cys Glu Ser Glu Asp Trp Val Asn Ala Pro Ser Val Trp Met Ala
 Tyr Leu Gly Met Lys His Val Trp Asp Tyr Val Gly Ile Ser Asp His
 5 Leu Ala Ile Asn Ile His Lys Ser Gly His Ala Val Ile Ala Glu Asp
 705 690 710 725 730 715 735 720
 Ile Glu Lys Met Val Gln Tyr Phe Asp Tyr His Val Tyr Gly Ile Gln
 10 Pro Lys Met Asn Leu Glu Glu Leu Gln Thr Ser Val Phe Ala Leu Pro
 740 745 750 760 765
 Lys Asn Lys Asp Ser Phe Ala Asp Thr Phe Ala Ser Lys Trp Leu Tyr
 755 765

<210> 19
 15 <211> 511
 <212> PRT
 <213> Secuencia Artificial

<220>
 20 <223> secuencia consenso

<220>
 <221> VARIANTE
 <222> (1)...(511)
 25 <223> Xaa = Cualquier aminoácido

<400> 19
 Xaa Xaa Xaa Xaa Xaa Xaa Ala Xaa Leu Xaa Xaa Ala Xaa Xaa Xaa Gln
 1 5 10 15
 30 Xaa Xaa Xaa Pro Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Gly Xaa Xaa Xaa
 20 25 30
 Xaa Phe Xaa Ala
 35 35 40 45
 Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Xaa Xaa
 50 50 55 60
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Ala Xaa Ala Xaa Ile Ala Xaa
 65 70 75 80
 Xaa Xaa Xaa Leu Xaa Ser Phe Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser
 85 90 95
 40 Xaa Ala Xaa Xaa Xaa Xaa Ser Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser Xaa
 100 105 110
 Xaa Xaa Xaa Pro Ala Xaa Xaa Xaa Leu Lys Ser Xaa Xaa Xaa Ile Xaa
 115 120 125
 45 Asp Xaa Phe Xaa Xaa Xaa Xaa Gly Xaa Lys Val Xaa Ser Xaa Asp Xaa
 130 135 140
 Phe Xaa Xaa Arg Gln Ala Glu Ile Ser Xaa Leu Xaa Xaa Xaa Tyr Xaa
 145 150 155 160
 Xaa Gly Xaa Thr Xaa Ser Xaa
 165 170 175
 50 Ser Gly Asn Ser Leu Thr Ile Asn Xaa Xaa Xaa Xaa Xaa Gly Lys
 180 185 190
 Ser Xaa Ser Phe Xaa Xaa Xaa Ile Xaa Xaa Pro Xaa Xaa Xaa Xaa Xaa
 195 200 205
 55 Xaa Xaa Gly Xaa Xaa Xaa Xaa Ala Ile Xaa Xaa Xaa Xaa Xaa Xaa
 210 215 220
 Ser Xaa Xaa Xaa Xaa Xaa Gly Xaa Ala Met Ile Xaa Phe Xaa Xaa Asp
 225 230 235 240
 Xaa Ile Xaa Ala Xaa Xaa Xaa Thr Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 245 250 255
 60 Gly Xaa Phe Tyr Asp Leu Tyr Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 260 265 270
 Xaa Gly Xaa Leu Xaa Ala Trp Ala Trp Gly Ile Ser Arg Ile Ile Asp

ES 2 371 362 T3

Asn Leu Phe Glu Ala Val Gln Val Tyr Thr Val Lys Gly Gly Gly Tyr
 210 215 220
 Leu Met Ile Val Glu Ala Val Gly Ser Gly Gly Arg Tyr Phe Arg Ser
 225 230 235 240
 5 Phe Thr Ala Ser Ser Leu Ser Gly Ser Trp Thr Pro Asn Ala Ala Thr
 245 250 255
 Glu Ser Asn Pro Phe Ala Gly Lys Ala Asn Ser Gly Ala Thr Trp Thr
 260 265 270
 10 Asn Asp Ile Ser His Gly Asp Leu Val Lys Val Thr Asn Asp Glu Thr
 275 280 285
 Met Thr Val Asp Pro Cys Asn Leu Gln Leu Leu Tyr Gln Gly Arg Ala
 290 295 300
 Pro Asn Ser Gly Gly Asp Tyr Asp Arg Leu Pro Tyr Arg Pro Gly Leu
 305 310 315 320
 15 Leu Thr

<210> 21

<211> 363

<212> PRT

20 <213> Streptomyces thermoviolaceus

<400> 21

Met Ser Phe His Arg Ser Leu Pro Phe Arg Pro Lys Arg Leu Phe Gly
 1 5 10 15
 25 Val Leu Ala Pro Leu Leu Leu Ala Gly Val Met Ser Thr Gln Pro Ala
 20 25 30
 Gly Ala Ala Thr Val Val Pro Ser Asp Asp Val Gln Gly Thr Gly Arg
 35 40 45
 Gln Ser Gln Leu Thr Asp Gly Phe Gly Thr Arg Ala Ser Cys Glu Leu
 50 55 60
 Pro Ser Thr Tyr Arg Trp Thr Ser Thr Gly Ala Leu Ala Gln Pro Arg
 65 70 75 80
 Ser Gly Trp Val Ser Leu Lys Asp Phe Thr Val Val Pro Tyr Asn Gly
 85 90 95
 35 Gln His Leu Val Tyr Ala Thr Thr His Asp Thr Gly Thr Arg Trp Gly
 100 105 110
 Ser Met Asn Phe Glu Pro Phe Gly Asp Trp Ser Gln Met Ala Thr Ala
 115 120 125
 40 Arg Gln Asn Ala Met Asn Ser Pro Thr Val Ala Pro Thr Leu Phe Tyr
 130 135 140
 Phe Ala Pro Lys Asp Ile Trp Val Leu Ala Tyr Gln Trp Gly Gly Ser
 145 150 155 160
 Ala Phe Ser Tyr Arg Thr Ser His Asp Pro Thr Asp Pro Asn Gly Trp
 165 170 175
 45 Ser Ser Glu Gln Val Leu Phe Ser Gly Ser Ile Ala Asp Ser Ala Thr
 180 185 190
 Gly Pro Ile Asp Gln Thr Leu Ile Gly Asp Asp Thr His Met Tyr Leu
 195 200 205
 50 Phe Phe Ala Gly Asp Asn Gly Lys Ile Tyr Arg Ala Ser Met Pro Ile
 210 215 220
 Gly Asp Phe Pro Gly Ser Phe Gly Ser Thr Ala Thr Val Val Met Ser
 225 230 235 240
 Asp Thr Arg Asn Asn Leu Phe Glu Ala Pro Gln Val Tyr Lys Leu Gln
 245 250 255
 55 Gly Gln Asn Arg Tyr Leu Met Ile Val Glu Ala Ile Gly Ala Gln Gly
 260 265 270
 Gln Arg Tyr Phe Arg Ser Phe Thr Ala Thr Ser Leu Asp Gly Glu Trp
 275 280 285
 60 Thr Pro Gln Ala Thr Ser Glu Ser Asn Pro Phe Ala Gly Lys Ala Asn
 290 295 300
 Ser Gly Ala Thr Trp Thr Asp Asp Ile Ser His Gly Glu Leu Ile Arg
 305 310 315 320

ES 2 371 362 T3

Thr Thr Ala Asp Gln Thr Met Thr Val Asp Pro Cys Asn Leu Gln Leu
 Leu Tyr Gln Gly Arg Asp Pro Gly Ser Gly Gly Thr Tyr Asp Leu Leu
 5 Pro Tyr Arg Pro Gly Leu Leu Thr Leu Gln Arg
 355 360

<210> 22
 <211> 24
 10 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> secuencia consenso

15 <220>
 <221> VARIANTE
 <222> (1)...(24)
 <223> Xaa = Cualquier aminoácido

20 <400> 22
 Leu Xaa Phe Met Pro Xaa Lys Ala Phe Ser Ala Leu Ala Leu Ala Leu
 1 5 10 15
 25 Leu Ala Xaa Val Ala Ser Ala Gln
 20

<210> 23
 <211> 304
 <212> PRT
 30 <213> Secuencia Artificial

<220>
 <223> secuencia consenso

35 <220>
 <221> VARIANTE
 <222> (1)...(304)
 <223> Xaa = Cualquier aminoácido

40 <400> 23
 Ala Ser Cys Xaa Leu Pro Ser Thr Tyr Arg Trp Thr Ser Thr Gly Ala
 1 5 10 15
 Leu Ala Gln Pro Lys Ser Gly Trp Val Ser Leu Lys Asp Phe Thr Ile
 20 25 30
 45 Val Xaa Tyr Asn Gly Gln His Leu Val Tyr Ala Thr Thr His Asp Thr
 35 40 45
 Gly Thr Lys Trp Gly Ser Met Asn Phe Xaa Pro Phe Gly Asp Trp Ser
 50 55 60
 Asn Met Ala Thr Ala Ser Gln Asn Ala Met Xaa Xaa Xaa Thr Val Ala
 65 70 75 80
 Pro Thr Leu Phe Tyr Phe Ala Pro Lys Asn Ile Trp Val Leu Ala Tyr
 85 90 95
 Gln Trp Gly Pro Thr Thr Phe Ser Tyr Arg Thr Ser Ser Asp Pro Thr
 100 105 110
 55 Xaa Pro Asn Gly Trp Ser Ser Xaa Gln Xaa Leu Phe Ser Gly Ser Ile
 115 120 125
 Ser Gly Ser Ala Thr Gly Pro Ile Asp Gln Thr Val Ile Gly Asp Ala
 130 135 140
 60 Thr Asn Met Tyr Leu Phe Ala Gly Asp Asn Gly Lys Ile Tyr Arg
 145 150 155 160
 Ala Ser Met Pro Ile Gly Asn Phe Pro Gly Ser Phe Gly Ser Thr Ser

ES 2 371 362 T3

					165					170					175		
	Thr	Val	Val	Met	Ser	Asp	Ser	Arg	Asn	Asn	Leu	Phe	Glu	Ala	Val	Gln	
				180					185					190			
5	Val	Tyr	Thr	Val	Xaa	Gly	Gln	Xaa	Xaa	Tyr	Leu	Met	Ile	Val	Glu	Ala	
			195					200					205				
	Ile	Gly	Ala	Asn	Gly	Xaa	Arg	Tyr	Phe	Arg	Ser	Phe	Thr	Ala	Thr	Ser	
		210					215					220					
	Leu	Xaa	Gly	Ser	Trp	Thr	Pro	Gln	Ala	Thr	Ser	Glu	Ser	Asn	Pro	Phe	
						230					235				240		
10	Ala	Gly	Lys	Ala	Asn	Ser	Gly	Ala	Thr	Trp	Thr	Asn	Asp	Ile	Ser	His	
					245					250					255		
	Gly	Asp	Leu	Ile	Arg	Ser	Thr	Xaa	Asp	Gln	Thr	Met	Thr	Val	Asp	Pro	
				260					265					270			
	Cys	Asn	Leu	Gln	Leu	Leu	Tyr	Gln	Gly	Arg	Ala	Pro	Asn	Ser	Gly	Gly	
			275					280					285				
15	Asp	Tyr	Asp	Leu	Leu	Pro	Tyr	Arg	Pro	Gly	Leu	Leu	Thr	Leu	Gln	Arg	
		290					295					300					
	<210> 24																
20	<211> 302																
	<212> PRT																
	<213> Trichoderma reesei																
	<400> 24																
25	Met	Pro	Ser	Val	Lys	Glu	Thr	Leu	Thr	Leu	Leu	Leu	Ser	Gln	Ala	Phe	
	1				5					10					15		
	Leu	Ala	Thr	Gly	Ser	Pro	Val	Asp	Gly	Glu	Thr	Val	Val	Lys	Arg	Gln	
				20					25					30			
30	Cys	Pro	Ala	Ile	His	Val	Phe	Gly	Ala	Arg	Glu	Thr	Thr	Val	Ser	Gln	
			35					40					45				
	Gly	Tyr	Gly	Ser	Ser	Ala	Thr	Val	Val	Asn	Leu	Val	Ile	Gln	Ala	His	
		50					55					60					
	Pro	Gly	Thr	Thr	Ser	Glu	Ala	Ile	Val	Tyr	Pro	Ala	Cys	Gly	Gly	Gln	
						70				75				80			
35	Ala	Ser	Cys	Gly	Gly	Ile	Ser	Tyr	Ala	Asn	Ser	Val	Val	Asn	Gly	Thr	
					85					90				95			
	Asn	Ala	Ala	Ala	Ala	Ala	Ile	Asn	Asn	Phe	His	Asn	Ser	Cys	Pro	Asp	
				100					105					110			
40	Thr	Gln	Leu	Val	Leu	Val	Gly	Tyr	Ser	Gln	Gly	Ala	Gln	Ile	Phe	Asp	
				115				120					125				
	Asn	Ala	Leu	Cys	Gly	Gly	Gly	Asp	Pro	Gly	Glu	Gly	Ile	Thr	Asn	Thr	
		130					135					140					
	Ala	Val	Pro	Leu	Thr	Ala	Gly	Ala	Val	Ser	Ala	Val	Lys	Ala	Ala	Ile	
						150					155					160	
45	Phe	Met	Gly	Asp	Pro	Arg	Asn	Ile	His	Gly	Leu	Pro	Tyr	Asn	Val	Gly	
					165					170				175			
	Thr	Cys	Thr	Thr	Gln	Gly	Phe	Asp	Ala	Arg	Pro	Ala	Gly	Phe	Val	Cys	
				180					185					190			
50	Pro	Ser	Ala	Ser	Lys	Ile	Lys	Ser	Tyr	Cys	Asp	Ala	Ala	Asp	Pro	Tyr	
			195					200					205				
	Cys	Cys	Thr	Gly	Asn	Asp	Pro	Asn	Val	His	Gln	Gly	Tyr	Gly	Gln	Glu	
		210					215					220					
	Tyr	Gly	Gln	Gln	Ala	Leu	Ala	Phe	Ile	Asn	Ser	Gln	Leu	Ser	Ser	Gly	
						230					235					240	
55	Gly	Ser	Gln	Pro	Pro	Gly	Gly	Gly	Pro	Thr	Ser	Thr	Ser	Arg	Pro	Thr	
					245					250					255		
	Ser	Thr	Arg	Thr	Gly	Ser	Ser	Pro	Gly	Pro	Thr	Gln	Thr	His	Trp	Gly	
				260					265					270			
60	Gln	Cys	Gly	Gly	Gln	Gly	Trp	Thr	Gly	Pro	Thr	Gln	Cys	Glu	Ser	Gly	
			275					280					285				
	Thr	Thr	Cys	Gln	Val	Ile	Ser	Gln	Trp	Tyr	Ser	Gln	Cys	Leu			
		290					295					300					

ES 2 371 362 T3

<210> 25
 <211> 310
 <212> PRT
 5 <213> Secuencia Artificial

 <220>
 <223> secuencia consenso

 10 <220>
 <221> VARIANTE
 <222> (1)...(310)
 <223> Xaa = Cualquier aminoácido

 15 <400> 25
 Ser Leu Ser Leu Xaa Leu Ser Xaa Ala Xaa Leu Ala Xaa Xaa Ser Xaa
 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
 Xaa Xaa Xaa Xaa Thr Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ala Ile His Met
 20 Xaa Xaa Ala Arg Xaa Ser Xaa Xaa Xaa Xaa Gly Xaa Gly Xaa Ser Ala
 35 40 45 50 55 60 65 70 75 80 85 90 95
 Xaa Leu Xaa Xaa Xaa Ile Xaa Xaa Xaa Xaa Pro Gly Ser Xaa Xaa Xaa
 50 55 60 65 70 75 80 85 90 95
 Ala Ile Xaa Tyr Pro Ala Xaa
 25 65 70 75 80 85 90 95
 Xaa Tyr Xaa Xaa Ser Xaa Xaa Xaa Gly Xaa Xaa Ala Xaa Xaa Xaa Xaa Xaa
 100 105 110 115 120 125 130 135 140 145 150 155 160
 Ile Asn Xaa Phe Xaa Xaa Ala Cys Pro Xaa Xaa Xaa Ile Xaa Leu Leu
 165 170 175 180 185 190 195 200 205 210 215 220
 30 Gly Tyr Ser Gln Gly Ala Gln Ile Xaa Xaa Xaa Xaa Ile Cys Gly Gly
 115 120 125 130 135 140 145 150 155 160 165 170
 Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser Xaa Xaa Xaa Leu Ser Xaa
 175 180 185 190 195 200 205 210 215 220 225 230 235 240
 Xaa Xaa Met Xaa Xaa Val Xaa Ala Xaa Xaa Xaa Xaa Gly Asp Pro Xaa
 35 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300
 Xaa Ile Xaa Xaa Leu Xaa Tyr Xaa Xaa Gly Thr Xaa Xaa Xaa Xaa Asn Gly
 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240
 Xaa Xaa Xaa Arg Xaa Ala Xaa Ala
 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255
 40 Ser Lys Ile Xaa Ser Tyr Cys Asp Xaa Ala Asp Xaa Tyr Cys Xaa Xaa
 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270
 Gly Xaa Xaa Xaa Xaa Val His Xaa Xaa Tyr Xaa Gln Xaa Tyr Gly Xaa
 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285
 Xaa Xaa Leu Xaa Phe Ile Xaa Ser Gln Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser Xaa
 45 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300
 Xaa Xaa Gly Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ala Pro Xaa Xaa
 245 250 255 260 265 270 275 280 285 290 295 300
 Thr Xaa Xaa Pro Thr Ser Thr Xaa Xaa Gly Xaa Ser Xaa Xaa Pro Thr
 260 265 270 275 280 285 290 295 300
 50 Xaa Thr Xaa Gly Xaa Gly Xaa Ser
 275 280 285 290 295 300
 Ala Xaa Thr Xaa Ile Xaa Xaa
 290 295 300
 Xaa Xaa Ser Gln Xaa Leu
 55 305 310

 <210> 26
 <211> 400
 <212> PRT
 60 <213> Streptomyces coelicor

 <400> 26

ES 2 371 362 T3

Met Arg Thr Arg Val Leu Arg Leu Leu Arg Arg Pro Trp Thr Ala Ala
 1 5 10 15
 Val Ala Ala Val Ala Leu Val Val Ser Val Leu Val Ala Met Pro Ala
 20 25 30
 5 Ser Gly Ala Ala Ala Ala Cys Arg Val Asp Tyr Gly Val Asp Ala
 35 40 45
 Trp Ala Gly Gly Tyr Thr Ala Arg Val Arg Ile Thr Asn Leu Gly Pro
 50 55 60
 10 Ala Val Ser Asp Trp Arg Leu Thr Trp Thr Tyr Thr Gly Asp Gln Gln
 65 70 75 80
 Val Thr Ser Ala Trp Asn Ala Thr Val Thr Gln Thr Gly Ala Ser Val
 85 90 95
 Val Ala Val Asp Ala Gly Trp Asn Gly Ala Val Ser Thr Gly Gly Thr
 100 105 110
 15 Ala Glu Phe Gly Leu Gln Gly Thr Trp Arg Ser Ala Asp Pro Ala Pro
 115 120 125
 Asp Asp Phe Ala Leu Asn Gly Thr Ser Cys Gly Asp Gly Thr Pro
 130 135 140
 20 Thr Ala Thr Pro Thr Thr Ser Pro Thr Ala Pro Pro Thr Thr Pro Pro
 145 150 155 160
 Thr Thr Pro Pro Pro Thr Thr Pro Pro Pro Ala Ala Glu Cys Gly Asp
 165 170 175
 Ala Val Ile Cys Ser Gly Phe Glu Asp Gln Ala Gly Pro Glu Pro Ser
 180 185 190
 25 Gly Asp Trp Arg Phe Thr Ala Pro Asp Cys Gln Gly Thr Gly Thr Ala
 195 200 205
 Ala Val Asp Ser Ala Val Ser His Ala Gly Gly Arg Ser Leu Arg Val
 210 215 220
 30 Asp Gly Arg Ala Gly Tyr Cys Asn His Ala Phe Val Ala His Thr Ala
 225 230 235 240
 Asp Leu Ser Ser Val Gly Pro Val Met Tyr Val Arg Met Trp Val Arg
 245 250 255
 His Thr Thr Ala Leu Pro Thr Ser His Val Thr Phe Val Ser Met Pro
 260 265 270
 35 Asp Ser Ala Gln Gly Gly Arg Ala Leu Arg Val Gly Gly Gln Asn Gly
 275 280 285
 Ala Leu Gln Trp Asn Arg Glu Ser Asp Asp Ala Thr Leu Pro Ala Gln
 290 295 300
 40 Ser Pro Ala Gly Val Ala Leu Ser Arg Pro Leu Pro Thr Asp Gly Trp
 305 310 315 320
 Gln Cys Leu Arg Phe Ala Ile Asp Thr Ser Ala Ala Gly Leu Asp Thr
 325 330 335
 Trp Leu Gly Asp Glu Gln Val Pro Gly Leu His Ala Asp Gly Val Pro
 340 345 350
 45 Thr Gln Asp Val Asp Gln Gln Trp Leu Thr Arg Gly Thr Ala Pro Arg
 355 360 365
 Pro Thr Ala Leu Arg Leu Gly Trp Glu Ser Tyr Ala Thr Gly Asp Asp
 370 375 380
 50 Thr Val Trp Phe Asp Asp Val Ala Val Gly Ser Ala Pro Ile Gly Cys
 385 390 395 400

<210> 27

<211> 236

<212> PRT

55 <213> Secuencia Artificial

<220>

<223> secuencia consenso

60 <220>

<221> VARIANTE

<222> (1)...(236)

ES 2 371 362 T3

<223> Xaa = Cualquier aminoácido

<400> 27

5 Thr Xaa Xaa Xaa Xaa Xaa Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ile Xaa Xaa
 1 5 10 15
 Xaa Xaa Asp Asp Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Trp Xaa Xaa
 20 25 30
 Ser Ala Pro Asp Cys Asn Xaa Xaa Gly Thr Xaa Ala Leu Asp Ser Xaa
 35 40 45
 10 Val Ala His Ala Gly Xaa Xaa Ser Leu Lys Val Xaa Gly Xaa Xaa Xaa
 50 55 60
 Gly Tyr Cys Xaa His Xaa Phe Xaa Ala Xaa Thr Xaa Xaa Xaa Ser Xaa
 65 70 75 80
 Xaa Xaa Xaa Xaa Met Tyr Val Arg Xaa Trp Ile Arg Xaa Xaa Thr Ala
 85 90 95
 15 Leu Xaa Ser Xaa His Val Thr Phe Ile Xaa Met Pro Asp Ser Ala Gln
 100 105 110
 Gly Gly Lys Xaa Leu Arg Ile Gly Gly Gln Xaa Xaa Xaa Leu Xaa Trp
 115 120 125
 20 Asn Arg Glu Ser Asp Asp Ala Thr Leu Pro Xaa Xaa Ser Pro Xaa Gly
 130 135 140
 Ile Ala Xaa Ser Xaa Xaa Leu Pro Thr Xaa Ala Phe Gln Cys Xaa Xaa
 145 150 155 160
 25 Phe Xaa Ile Xaa Thr Xaa Ala Xaa Xaa Ile Asp Thr Trp Leu Xaa Xaa
 165 170 175
 Xaa Xaa Ile Pro Gly Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Asn Xaa Xaa
 180 185 190
 Asp Xaa Xaa Trp Xaa Xaa Xaa Xaa Xaa Xaa Pro Xaa Xaa Thr Ala Leu
 195 200 205
 30 Xaa Xaa Gly Trp Glu Ala Tyr Ala Xaa Xaa Xaa Xaa Thr Val Trp Phe
 210 215 220
 Asp Asp Ile Ala Ile Ala Ser Xaa Xaa Ile Gly Cys
 225 230 235

35 <210> 28
 <211> 18
 <212> ADN
 <213> Secuencia Artificial

40 <220>
 <223> cebador

45 <400> 28
 gacaatccaa acgacgct

18

50 <210> 29
 <211> 20
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> cebador

55 <400> 29
 caatcgagat gtcgtcgaac

20

60 <210> 30
 <211> 20
 <212> ADN
 <213> Secuencia Artificial

ES 2 371 362 T3

	<220>		
	<223> cebador		
5	<400> 30	ctcctccaca cccggtgccg	20
	<210> 31		
	<211> 17		
	<212> ADN		
10	<213> Secuencia Artificial		
	<220>		
	<223> cebador		
15	<400> 31	tgctgccaat ggggccg	17
	<210> 32		
	<211> 17		
20	<212> ADN		
	<213> Secuencia Artificial		
	<220>		
	<223> cebador		
25	<400> 32	acgtattcag gcaacct	17
	<210> 33		
30	<211> 17		
	<212> ADN		
	<213> Secuencia Artificial		
	<220>		
35	<223> cebador		
	<400> 33	gcagtggcca tggctcc	17
40	<210> 34		
	<211> 17		
	<212> ADN		
	<213> Secuencia Artificial		
45	<220>		
	<223> cebador		
	<400> 34	ccagtacatg aactggc	17
50	<210> 35		
	<211> 17		
	<212> ADN		
	<213> Secuencia Artificial		
55	<220>		
	<223> cebador		
	<400> 35	agacctaatg tctcccc	17
60			

ES 2 371 362 T3

<210> 36
 <211> 17
 <212> ADN
 <213> Secuencia Artificial
 5
 <220>
 <223> cebador
 <400> 36
 10 **cgaattgtgc tcctggc** 17
 <210> 37
 <211> 17
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> cebador
 20 <400> 37
gtggttgac cggatgg 17
 <210> 38
 <211> 17
 25 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> cebador
 30 <400> 38
cctaccgtgg tadcagg 17
 <210> 39
 <211> 17
 35 <212> ADN
 <213> Secuencia Artificial
 <220>
 40 <223> cebador
 <400> 39
tggttctgct ggtcggg 17
 45 <210> 40
 <211> 17
 <212> ADN
 <213> Secuencia Artificial
 50 <220>
 <223> cebador
 <400> 40
 55 **catttcgaca tcatggc** 17
 <210> 41
 <211> 17
 <212> ADN
 60 <213> Secuencia Artificial

	<220> <223> cebador	
5	<400> 41 ctgtcccacg cagaggc	17
10	<210> 42 <211> 17 <212> ADN <213> Secuencia Artificial	
15	<220> <223> cebador <400> 42 ccggctggct tcgtctg	17
20	<210> 43 <211> 17 <212> ADN <213> Secuencia Artificial	
25	<220> <223> cebador <400> 43 tggccgtaac cttggtg	17
30	<210> 44 <211> 17 <212> ADN <213> Secuencia Artificial	
35	<220> <223> cebador <400> 44 cctctctcac gactcgc	17
40	<210> 45 <211> 17 <212> ADN <213> Secuencia Artificial	
45	<220> <223> cebador <400> 45 gttcgatgag ttgtacc	17
50	<210> 46 <211> 21 <212> ADN <213> Secuencia Artificial	
55	<220> <223> cebador <400> 46 cccccaaacg gaacaacttc c	21

	<210> 47	
	<211> 21	
	<212> ADN	
	<213> Secuencia Artificial	
5	<220>	
	<223> cebador	
	<400> 47	
10	ctgtatctgt ggttgtag g	21
	<210> 48	
	<211> 14	
	<212> ADN	
15	<213> Secuencia Artificial	
	<220>	
	<223> cebador	
20	<220>	
	<221> misc_feature	
	<222> (1)...(14)	
	<223> n = A,T,C o G	
25	<400> 48	
	ggncartgyg gngg	14
	<210> 49	
	<211> 14	
30	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
	<223> cebador	
35	<220>	
	<221> misc_feature	
	<222> (1)...(14)	
	<223> n = A,T,C o G	
40	<400> 49	
	adrcaytgng arta	14
	<210> 50	
45	<211> 14	
	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
50	<223> cebador	
	<400> 50	
	adrcaytgrc trta	14
55	<210> 51	
	<211> 14	
	<212> ADN	
	<213> Secuencia Artificial	
60	<220>	

	<223> cebador	
	<220>	
	<221> misc_feature	
5	<222> (1)...(14)	
	<223> n = A,T,C o G	
	<400> 51	
10	adrcaytgng crta	14
	<210> 52	
	<211> 25	
	<212> ADN	
	<213> Secuencia Artificial	
15	<220>	
	<223> cebador	
	<400> 52	
20	caccatgggtt cgccggactg ctctg	25
	<210> 53	
	<211> 24	
	<212> ADN	
25	<213> Secuencia Artificial	
	<220>	
	<223> cebador	
30	<400> 53	
	ttataagcac tgggagtagt atgg	24
	<210> 54	
	<211> 23	
35	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
	<223> cebador	
40	<400> 54	
	caccatggct tcccgcttct ttg	23
	<210> 55	
45	<211> 20	
	<212> ADN	
	<213> Secuencia Artificial	
	<220>	
50	<223> cebador	
	<400> 55	
	tcaactcagc gttgggggttg	20
55	<210> 56	
	<211> 54	
	<212> ADN	
	<213> Secuencia Artificial	
60	<220>	

ES 2 371 362 T3

<223> cebador
 <400> 56
 5 ggggacaagt ttgtacaaaa aagcaggcta tggagcttaa agcactcagt gccg 54
 <210> 57
 <211> 51
 <212> ADN
 <213> Secuencia Artificial
 10 <220>
 <223> cebador
 <400> 57
 15 ggggaccact ttgtacaaga aagctggggtt cagcgctgga gagttagcag c 51
 <210> 58
 <211> 50
 <212> ADN
 20 <213> Secuencia Artificial
 <220>
 <223> cebador
 25 <400> 58
 ggggacaagt ttgtacaaaa aagcaggcta tgcgcgccct ctcactctcc 50
 <210> 59
 <211> 51
 30 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> cebador
 35 <400> 59
 ggggaccact ttgtacaaga aagctggggtt cacagcatct gagacaccgc c 51

REIVINDICACIONES

1. Polinucleótido aislado seleccionado del grupo que consiste en:

- 5 (a) una secuencia de ácidos nucleicos que codifica un polipéptido CIP2 que tiene actividad de unión a celulosa y que tiene por lo menos un 85% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:9, o el complemento de la misma;
- 10 (b) una secuencia de ácidos nucleicos que codifica un polipéptido CIP2 que tiene actividad de unión a celulosa y que tiene por lo menos un 90% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:9, o el complemento de la misma;
- (c) una secuencia de ácidos nucleicos que codifica un polipéptido CIP2 que tiene actividad de unión a celulosa y que tiene por lo menos un 95% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:9, o el complemento de la misma;
- 15 (d) una secuencia de ácidos nucleicos que codifica un polipéptido CIP2 que tiene actividad de unión a celulosa y que tiene la secuencia de aminoácidos presentada como SEC ID NO:9, o el complemento de la misma;
- (e) una secuencia de ácidos nucleicos que codifica un polipéptido CIP2 que tiene actividad de unión a celulosa y que tiene por lo menos un 95% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:7, o el complemento de la misma;
- 20 (f) una secuencia de ácidos nucleicos que codifica un polipéptido CIP2 que tiene actividad de unión a celulosa y que tiene la secuencia de aminoácidos presentada como SEC ID NO:7, o el complemento de la misma; y
- (g) una secuencia de ácidos nucleicos presentada como SEC ID NO:6, o el complemento de la misma,

25 donde el % de identidad se calcula utilizando el programa CLUSTAL-W en MacVector versión 6.5, operado con los parámetros por defecto, incluyendo una penalización por la apertura del espacio de 10,0, una penalización por la extensión del espacio de 0,1 y una matriz de similitud de BLOSUM30.

30 2. Polinucleótido aislado según la reivindicación 1, en el que dicho polinucleótido es una molécula de ARN.

3. Polinucleótido aislado según la reivindicación 1, en el que la proteína deriva de una fuente de Trichoderma.

4. Polinucleótido aislado según la reivindicación 3, en el que la proteína deriva de Trichoderma reesei.

35 5. Construcción de expresión que comprende una secuencia de polinucleótidos (i) que codifica un polipéptido que tiene afinidad de unión a celulosa y que tiene por lo menos un 85% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:9, o el complemento de la misma; o (ii) que codifica un polipéptido que tiene afinidad de unión a celulosa y que tiene por lo menos un 85% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:7, o el complemento de la misma.

40 6. Vector que comprende la construcción de expresión según la reivindicación 5.

7. Vector que comprende un polinucleótido aislado según la reivindicación 1, unido operativamente a secuencias de control reconocidas por una célula huésped transformada con el vector.

45 8. Célula huésped que comprende el vector según la reivindicación 6.

9. Célula huésped que comprende el vector según la reivindicación 7.

50 10. Célula huésped según la reivindicación 9, que es una célula procariota.

11. Célula huésped según la reivindicación 9, que es una célula eucariota.

55 12. Célula huésped procariota recombinante que comprende un polinucleótido según la reivindicación 1.

13. Polipéptido CIP2 sustancialmente purificado que tiene actividad de unión a celulosa, que comprende una secuencia seleccionada del grupo que consiste en:

- 60 (a) una secuencia de aminoácidos que tiene por lo menos un 85% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:9;

- 5 (b) una secuencia de aminoácidos que tiene por lo menos un 90% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:9;
- (c) una secuencia de aminoácidos que tiene por lo menos un 95% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:9;
- (d) una secuencia de aminoácidos presentada como SEC ID NO:9;
- (e) una secuencia de aminoácidos que tiene por lo menos un 95% de identidad en la secuencia con la secuencia de aminoácidos presentada como SEC ID NO:7;
- (f) una secuencia de aminoácidos presentada como SEC ID NO:7; y
- 10 (g) un fragmento biológicamente activo sustancialmente purificado de la secuencia de aminoácidos presentada como SEC ID NO: 9, en la que el fragmento tiene actividad de unión a celulosa y en la que el fragmento comprende por lo menos 200 aminoácidos.
14. Método de producción de una proteína que tiene actividad de unión a celulosa, que comprende:
- 15 (a) transformar de manera estable una célula huésped con un vector de expresión que comprende un polinucleótido tal como se define en la reivindicación 1;
- (b) cultivar dicha célula huésped transformada bajo condiciones adecuadas para dicha célula huésped para producir dicha proteína de unión a celulosa; y
- 20 (c) recuperar dicha proteína que tiene actividad de unión a celulosa.
15. Método según la reivindicación 16, en el que la célula huésped es un hongo filamentoso o una célula de levadura.
- 25 16. Método de preparación de una célula huésped recombinante que comprende proporcionar una célula parental que tiene un gen *cip2* que comprende un polinucleótido tal como se define en la reivindicación 1, y realizar una delección o inserción u otra alteración en dicho gen *cip2* que inactiva el gen, para producir una célula huésped recombinante que es incapaz de producir un polipéptido CIP2 tal como se define en la reivindicación 13.
- 30 17. Oligonucleótido antisentido complementario a un ARN mensajero que codifica un polipéptido CIP2 que tiene la secuencia presentada como SEC ID NO:9, en el que tras la exposición a una célula huésped productora de CIP2, dicho oligonucleótido disminuye o inhibe la producción de CIP2 por dicha célula huésped.
18. Oligonucleótido antisentido según la reivindicación 17, en el que la célula huésped es un hongo filamentoso.

ES 2 371 362 T3

Figura 1: secuencia de ADNc de *cip1* (SEC Id No. 1)

<u>GACTAGTTCA</u>	<u>TAATACAGTA</u>	<u>GTTGAGTTCA</u>	<u>TAGCAACTTC</u>	<u>ACTCTCTAGC</u>	50
<u>TGAACAAATT</u>	<u>ATCTGCGCAA</u>	<u>ACATGGTTCG</u>	<u>CCGGACTGCT</u>	<u>CTGCTGGCCC</u>	100
<u>TTGGGGCTCT</u>	<u>CTCAACGCTC</u>	<u>TCTATGGCCC</u>	<u>AAATCTCAGA</u>	<u>CGACTTCGAG</u>	150
<u>TCGGGCTGGG</u>	<u>ATCAGACTAA</u>	<u>ATGGCCCATT</u>	<u>TCGGCACCAG</u>	<u>ACTGTAACCA</u>	200
<u>GGGCGGCACC</u>	<u>GTCAGCCTCG</u>	<u>ACACCACAGT</u>	<u>AGCCCACAGC</u>	<u>GGCAGCAACT</u>	250
<u>CCATGAAGGT</u>	<u>CGTTGGTGGC</u>	<u>CCCAATGGCT</u>	<u>ACTGTGGACA</u>	<u>CATCTTCTTC</u>	300
<u>GGCACTACCC</u>	<u>AGGTGCCAAC</u>	<u>TGGGGATGTA</u>	<u>TATGTCAGAG</u>	<u>CTTGGATTCC</u>	350
<u>GCTTCAGACT</u>	<u>GCTCTCGGCA</u>	<u>GCAACCACGT</u>	<u>CACATTCATC</u>	<u>ATCATGCCAG</u>	400
<u>ACACCGCTCA</u>	<u>GGGAGGGAAG</u>	<u>CACCTCCGAA</u>	<u>TTGGTGGCCA</u>	<u>AAGCCAAGTT</u>	450
<u>CTCGACTACA</u>	<u>ACCGCGAGTC</u>	<u>CGACGATGCC</u>	<u>ACTCTTCCGG</u>	<u>ACCTGTCTCC</u>	500
<u>CAACGGCATT</u>	<u>GCCTCCACCG</u>	<u>TCACTCTGCC</u>	<u>TACCGGCGCG</u>	<u>TTCCAGTGCT</u>	550
<u>TCGAGTACCA</u>	<u>CCTGGGCACT</u>	<u>GACGGAACCA</u>	<u>TCGAGACGTG</u>	<u>GCTCAACGGC</u>	600
<u>AGCCTCATCC</u>	<u>CGGGCATGAC</u>	<u>CGTGGGCCCT</u>	<u>GGCGTCGACA</u>	<u>ATCCAAACGA</u>	650
<u>CGCTGGCTGG</u>	<u>ACGAGGGCCA</u>	<u>GCTATATTCC</u>	<u>GGAGATCACC</u>	<u>GGTGTCAACT</u>	700
<u>TTGGCTGGGA</u>	<u>GGCCTACAGC</u>	<u>GGAGACGTCA</u>	<u>ACACCGTCTG</u>	<u>GTTCGACGAC</u>	750
<u>ATCTCGATTG</u>	<u>CGTCGACCCG</u>	<u>CGTGGGATGC</u>	<u>GGCCCCGGCA</u>	<u>GCCCCGGCGG</u>	800
<u>TCCTGGAAGC</u>	<u>TCGACGACTG</u>	<u>GGCGTAGCAG</u>	<u>CACCTCGGGC</u>	<u>CCGACGAGCA</u>	850
<u>CTTCGAGGCC</u>	<u>AAGCACCACC</u>	<u>ATTCCGCCAC</u>	<u>CGACTTCCAG</u>	<u>GACAACGACC</u>	900
<u>GCCACGGGTC</u>	<u>CGACTCAGAC</u>	<u>ACACTATGGC</u>	<u>CAGTGCGGAG</u>	<u>GGATTGGTTA</u>	1000
<u>CAGCGGGCCT</u>	<u>ACGGTCTGCG</u>	<u>CGAGCGGCAC</u>	<u>GACCTGCCAG</u>	<u>GTCCTGAACC</u>	1050
<u>CATACTACTC</u>	<u>CCAGTGCTTA</u>	<u>TAAGGGGATG</u>	<u>AGCATGGAGT</u>	<u>GAAGTGAAGT</u>	1100
<u>GAAGTGGAGA</u>	<u>GAGTTGAAAGT</u>	<u>GGCATTGCGC</u>	<u>TGGGCTGGGT</u>	<u>AGATAAAAGT</u>	1150
<u>CAGCAGCTAT</u>	<u>GAATACTCTA</u>	<u>TGTGATGCTC</u>	<u>ATTGGCGTGT</u>	<u>ACGTTTTTAA</u>	1200
<u>AAAAAAAAAA</u>	<u>AAAAAAAAAA</u>	<u>AAAAAAAAAA</u>	<u>AAAAAAAAAA</u>	<u>AAAAAAAAAA</u>	1250
<u>AAAAAAAAAAG</u>	<u>GGGCGGCCG</u>	<u>C</u>			1271

Figura 2: secuencia codificante de *cip1* (SEC Id No. 2)

ATGGTTTCGCC	GGACTGCTCT	GCTGGCCCTT	GGGGCTCTCT	CAACGCTCTC	50
TATGGCCCAA	ATCTCAGACG	ACTTCGAGTC	GGGCTGGGAT	CAGACTAAAT	100
GGCCCATTTT	GGCACCAGAC	TGTAACCAGG	GCGGCACCGT	CAGCCTCGAC	150
ACCACAGTAG	CCCACAGCGG	CAGCAACTCC	ATGAAGGTCG	TTGGTGGCCC	200
CAATGGCTAC	TGTGGACACA	TCTTCTTCGG	CACTACCCAG	GTGCCAACTG	250
GGGATGTATA	TGTCAGAGCT	TGGATTTCGGC	TTCAGACTGC	TCTCGGCAGC	300
AACCACGTCA	CATTCATCAT	CATGCCAGAC	ACCGCTCAGG	GAGGGAAGCA	350
CCTCCGAATT	GGTGGCCAAA	GCCAAGTTCT	CGACTACAAC	CGCGAGTCCG	400
ACGATGCCAC	TCTTCCGGAC	CTGTCTCCCA	ACGGCATTGC	CTCCACCGTC	450
ACTCTGCCTA	CCGGCGCGTT	CCAGTGCTTC	GAGTACCACC	TGGGCACTGA	500
CGGAACCATC	GAGACGTGGC	TCAACGGCAG	CCTCATCCCG	GGCATGACCG	550
TGGGCCCTGG	CGTCGACAAT	CCAAACGACG	CTGGCTGGAC	GAGGGCCAGC	600
TATATTCCGG	AGATCACCGG	TGTCAACTTT	GGCTGGGAGG	CCTACAGCGG	650
AGACGTCAAC	ACCGTCTGGT	TCGACGACAT	CTCGATTGCG	TCGACCCGCG	700
TGGGATGCGG	CCCCGGCAGC	CCCGGCGGTC	CTGGAAGCTC	GACGACTGGG	750
CGTAGCAGCA	CCTCGGGCCC	GACGAGCACT	TCGAGGCCAA	GCACCACCAT	800
TCCGCCACCG	ACTTCCAGGA	CAACGACCGC	CACGGGTCCG	ACTCAGACAC	850
ACTATGGCCA	GTGCGGAGGG	ATTGGTTACA	GCGGGCCTAC	GGTCTGCGCG	900
AGCGGCACGA	CCTGCCAGGT	CCTGAACCCA	TACTACTCCC	AGTGCTTATA	950
A					951

Figura 3A: secuencia de la proteína CIP1 (SEC Id No. 3)

MVRR TALLAL	GALSTLS MAQ	ISDDFESGWD	QTKWPISAPD	CNQGGTVSLD	50
TTVAHSGSNS	MKVVGGPNGY	CGHIFFGTTQ	VPTGDVYVRA	WIRLQTALGS	100
NHVTFIIMPD	TAQGGKHLRI	GGQSQVLDYN	RESDDATLPD	LSPNGIASTV	150
TLPTGAFQCF	EYHLGTDGTI	ETWLNGLSLIP	GMTVGPGVDN	PNDAGWTRAS	200
YIPEITGVNF	GWEAYSGDVN	TVWFDDISIA	STRVGCPGPS	PGGPGSSSTG	250
RSSTSGPTST	SRPSTTIPPP	TSRTTTATGP	TQTHYQCGG	IGYSGPTVCA	300
SGTTCQVLNP	YYSQCL				316

Figura 3B: secuencia señal CIP1 (SEC Id No. 4)

MVRR TALLAL	GALSTLS MA	19
--------------------	-------------------	----

Figura 3C: secuencia de la proteína madura CIP1 (SEC Id No. 5)

QISDDFESGW	DQTKWPISAP	DCNQGGTVSL	DTTVAHSGSN	SMKVVGGPNG	50
YCGHIFFGTT	QVPTGDVYVR	AWIRLQTALG	SNHVTFIIMP	DTAQGGKHLR	100
IGGQSQVLDY	NRESDDATLP	DLSPNGIAST	VTLPTGAFQC	FEYHLGTDGT	150
IETWLNGLSLI	PGMTVGPGDV	NPNDAGWTRA	SYIPEITGVN	FGWEAYSGDV	200
NTVWFDISI	ASTRVGCPGP	SPGGPGSSTT	GRSSTSGPTS	TSRPSTTIPP	250
PTSRTTTATG	PTQTHYQCG	GIGYSGPTVC	ASGTTCQLN	PYYSQCL	297

Figura 4: secuencia de longitud completa de *cip2*
(SEC Id No. 6)

```

1   ATGGCTTCCC GCTTCTTTGC TCTTCTCCTT TTAGCGATCC CAATCCAGGC
51  CCAATCTCCA GTCTGGGGAC AATGTGGTGG AATTGGTTGG TCTGGCCCAA
101 CAACTTGTGT TGGAGGTGCG ACTTGTGTAT CATATAACCC TTATTACTCG
151 CAATGTATTC CCAGTACACA GGCTTCATCG AGCATAGCCT CTACAACGCT
201 GGTCACATCA TTTACGACCA CCACTGCTAC GAGGACTTCG GCATCAACGC
251 CTCCAGCGAG CAGTACAGGT GCAGGCGGGC CAACATGCTC AGCACTGCCG
301 GGCTCCATTA CCCTGAGATC CAACGCAAAG CTCAACGATC TGTTTACAAT
351 GTTCAATGGA GATAAGGTCA CCACGAAAGA CAAATTCTCG TGCCGCCAGG
401 CAGAGATGTC GGAGCTAATA CAACGATATG AGCTCGGCAC CCTGCCCGGA
451 CGACCAAGCA CTCTCACAGC CTCATTCTCG GGCAATACGT TGACCATCAA
501 TTGCGGAGAG GCCGAAAGT CAATTTTATT CACAGTCACG ATCACTTATC
551 CATCTTCCGG AACAGCACCA TACCCTGCGA TTATCGGCTA TGGAGGCGGC
601 AGTCTTCCAG CTCCCGCCGG GGTGCCATG ATCAACTTTA ACAATGACAA
651 CATAGCAGCC CAAGTTAATA CAGGCAGCCG CGGACAGGGC AAGTTCTACG
701 ATCTCTACGG GAGCTCGCAC TCCGCGGGCG CCATGACCCG ATGGGCCTGG
751 GGAGTAAGCC GAGTCATTGA TGCTCTTGAG CTTGTACCAG GCGCAAGAAT
801 AGACACCAAC AAGATTGGCG TGACGGGGTG TTCACGAAAT GGCAAAGGCG
851 CAATGGTTCG AGGTGCTTTC GAGAAACGAA TCGTTCTGAC ACTTCCCAG
901 GAGTCGGGCG CCGGTGGCTC TGCGTGCTGG AGGATTTTCA ACTACTTAAA
951 GTCCCAAGGA GCCAATATCC AGACCGCGTC TGAGATCATT GGCGAAGACC
1001 CCTGGTTCTC GACTACTTTC AACAGCTACG TCAACCAAGT GCCGGTGTG
1051 CCGTTTGACC ACCATTGCTT TGCTGCCTTG ATAGCCCCGA GAGGATTATT
1101 CGTCATCGAC AACCAATATT ACTGGCTCGG CCCACAAAGC TGCTTTGGCT
1151 GTATGACAGC TGCTCACATG GCATGGCAAG CTTTGGGTGT CTCGGACCAC
1201 ATGGGCTATT CGCAGATTGG AGCTCACGCA CACTGCGCGT TCCCATCAAA
1251 CCAGCAATCG CAACTTACTG CCTTGTGTTA GAAATTCCTG CTGGGCCAGT
1301 CCACAAATAC GCGGATTTTC CAAAGCGACT TTTCCGCCAA TCAAAGCCAA
1351 TGGATCGAQT GGACAACCCC AACGCTGAGT TGA

```

Figura 5A: secuencia de la proteína CIP2 (SEC Id No. 7)

```

1  MASRFFALLL LAIPIQAQSP VWGQCGGIGW SGPTTCVGGG TCVSYNPYYS
51  QCIPSTQASS SIASTTLVTS FTTTTATRIS ASTPPASSTG AGGATCSALP
101 GSITLRSNAK LNDLFTMFNG DKVTTKDKFS CRQAEMSELI QRYELGTLPG
151 RPSTLTASF S GNTLTINCGE AGKSISFTVT ITYPSSGTAP YPAIIGYGGG
201 SLPAPAGVAM INFNNDNIAA QVNTGSRGQG KFYDLYGSSH SAGAMTAWAW
251 GVSRVIDALE LVPGARIDTT KIGVTGCSRN GKGAMVAGAF EKRIVLTLPQ
301 ESGAGGSACW RISDYLSQAG ANIQTASEII GEDPWFSTTF NSYVNQVPVL
351 PFDHHSALAL IAPRGLFVID NNIDWLG PQS CFCGCMATAHM AWQALGVSDH
401 MGYSQIGAHA HCAFPNSQQS QLTAQVQKFL LGQSTNTAIF QSDFSANQSQ
451 WIDWTTPTLS

```

Figura 5B: secuencia señal de CIP2 (SEC Id No. 8)

```

1  MASRFFALLL LAIPIQA

```

Figura 5C: secuencia de la proteína madura (SEC Id No. 9)

```

1  QSPVWGQCGG IGWSGPTTCV GGATCVSYNP YYSQCIPSTQ ASSSIASTTL
51  VTSFTTTTAT RTSASTPPAS STGAGGATCS ALPGSITLRS NAKLNDLFTM
101 FNGDKVTTKD KFSCRQAEMS ELIQRYELGT LPGRPSTLTA SPSGNTLTIN
151 CGEAGKSI SF TVTITYPSSG TAPYPAIIGY GGGSLPAPAG VAMINFNNDN
201 IAAQVNTGSR GQGFYDLYG SSHSAGAMTA WAWGVSRVID ALELVPGARI
251 DTTKIGVTGC SRNGKAMVA GAFEKRIVLT LPQESGAGGS ACWRISDYLK
301 SQGANIQTAS EIIGEDPWFS TTFNSYVNQV PVL PFDHHSAL AALIAPRGLF
351 VIDNNIDWLG PQSCFCGCMTA AHMAWQALGV SDHMGYSQIG AHAHCAFPNS
401 QSQQLTAFVQ KFL LGQSTNT AIFQSDFSAN QSQWIDWTTPTLS

```

ES 2 371 362 T3

Figura 6: Alineación de Cip2 con R. Flavefaciens cesA CAB55348
 Cip2 tiene una secuencia señal N-terminal prevista de 17 aminoácidos seguida de 36 aminoácidos que comprenden un módulo de unión a carbohidrato de la familia CBM1 y una región enlazadora que termina en aproximadamente el aminoácido 95

		1	50
R.flavefaciens cesA	(1)	MKKHFPVVGETIKRFLRIGTSLALSISTLSLLPSAPRLSSAAGTIKIMPLG	
CIP2 040302	(1)	-----	
Consensus	(1)	-----	
		51	100
R.flavefaciens cesA	(51)	DSITYGMADEGGYRKLYSYFLQQKGYTNVDLVGPEGKDSASFNYMGQSVK	
CIP2 040302	(1)	-----	
Consensus	(51)	-----	
		101	150
R.flavefaciens cesA	(101)	YDDNHAGYSGYTITNLPFGWFGQLNGILETMQGGDYIKKYSFDIILLQIG	
CIP2 040302	(1)	-----	
Consensus	(101)	-----	
		151	200
R.flavefaciens cesA	(151)	TNDVSNGLDGSERLHKLLDYLRENMPNSGKVFLLTTFDLGNSGWGGNS	
CIP2 040302	(1)	-----	
Consensus	(151)	-----	
		201	250
R.flavefaciens cesA	(201)	NGDIKYNELIKKVANDYSSKNVIYADIHSVIDASKDLADGVHPNAGGYE	
CIP2 040302	(1)	-----	
Consensus	(201)	-----	
		251	300
R.flavefaciens cesA	(251)	KMGKYWLEQIEGYKASDGPVQTCVTPQSQDSEFELIYGDLGDKTITS	
CIP2 040302	(1)	-----MASRPFLLLLAIPVLAQSVWQQCGI-----	
Consensus	(251)	-----A L A Q P G G-----	
		301	350
R.flavefaciens cesA	(301)	FDVIMRKLLINDPKDINVKKAADIDQNGKREVLDVQLQMIIGKIKKFP	
CIP2 040302	(30)	WSGPTTCVIGATCVSYIPYYSQCIPSTQASSEISTTLATITTTTATRT	
Consensus	(301)	F A G N N A A L A L S F	
		351	400
R.flavefaciens cesA	(351)	TVAEKTVREKPVFEKSYNFAAVNOGKSKDIPPEIFMDSAVESTDW	
CIP2 040302	(80)	SASTPPASSTGAGGATCSALGSSITRINAKINLITMPNDAVITKQRF	
Consensus	(351)	S A S S P A L K S I D P G K V S D F	
		401	450
R.flavefaciens cesA	(400)	WKAISLTCMYEYMYKWDGSDDETYYHLSSEMGLVKKRSTPTA	
CIP2 040302	(130)	SCQAWWELIQRELTLPGRPSTLAAPELITLTCG--EALSL	
Consensus	(401)	RQAEIS L Y G T S SGNSLTIN GKS S	
		451	500
R.flavefaciens cesA	(450)	AKAVNLKKNVRHEGLAPVILGMHKGISETATSNYVITXDSGMFSA	
CIP2 040302	(177)	STVLTYS-----STAPYPAIIGYGGGLPAPAVMNEFNNAIAQ	
Consensus	(451)	F I P G AI S G AMI F D I A	
		501	550
R.flavefaciens cesA	(500)	PGAQDNOHKAVLPPYGRNWDEQTDLMSPTIILDNYNGAA	
CIP2 040302	(222)	VNDSRG---QKAGS---SHSAAMTAAALVNVILDELVPG	
Consensus	(501)	TA G FYDLY G L AWAAGISRITDAL A	
		551	600
R.flavefaciens cesA	(550)	KELININPDSSIVWYMAVAVSVDTEKMCASCSLILLY	
CIP2 040302	(265)	ARDPTTKIG--KICINFGAMAAEKAVLTLAQEASGSCNLI	
Consensus	(551)	I V T G SR GKAA V GAED RI L P SGAGG A WR	
		601	650
R.flavefaciens cesA	(600)	ISVGTYDFSSKGGSSYTYKENEPLGSLQASVQGGNGRMEFR--AE	
CIP2 040302	(313)	DYLSQGANIOTA-----EILGDFLSTTNSYVQVP	
Consensus	(601)	S KS AS GE WF F FN	
		651	700
R.flavefaciens cesA	(649)	QPMQMGTECCDPDRYLPFIIGSCSEAVNAPVVMAYLGMKHVVDY	
CIP2 040302	(349)	VLFRHSLAIAAPRGLEVIDN---NIIIGPQCEGCMTAAHMAQA	
Consensus	(651)	P D H LAAL FL DWL S F A W	
		701	750
R.flavefaciens cesA	(699)	VILKLAINIKSGHAVIAEDIEKMQVQYFDYHVYGLQPKMLEELTSV	
CIP2 040302	(395)	LVLKMGYSQIGAHACAPPSNQSQSLTAFVQKFLIGSTTAIFSDP	
Consensus	(701)	LGISDHLA A F I N QS	
		751	770
R.flavefaciens cesA	(749)	FLLPKNKDSFADTFKWLNY	
CIP2 040302	(445)	SNQSQWIDWTTPTL----	
Consensus	(751)	A N F S	

Figura 7: secuencia de ácido nucleico de abf2 (SEC
Id No. 10)

```

1   ATGGAGCTTA AAGCACTCAG TGCCGTTGTG CTGAGCTTTG TAACTCTTGT
51  CGCGGCAGCA CCGGCACCTT GCACGCTTCC GTCCACATAC CGCTGGAATT
101 CGACCGGTGC TTTAGCCAGC CCGAAATCAG GCTGGGTCTC GCTGAAAGAC
151 TTCTCCCATG TCATTTATAA TGGCCAGCAT CTTGTATGGG GCTCGACTCA
201 TGACACAGGA ACAATCTGGG GTTCAATGAA CTTTGGTCTG TTCAGTGAAT
251 GGTCCAATAT GGCAACGGCA AGCCAGAACA AAATGACTCC CGGCACTGTT
301 GCTCCTACCG TCTTCTACTT TGCCCCGAAG AATATTGGG TACTCGCCTA
351 TCAATGGGGC CCGACCACGT TTTCTACCT GACGTCAAGC AACCCCTCCA
401 GCGTCAATGG ATGGTCGTCA CCACAGCCTC TCTTCTCCGG CAGTATCTCA
451 GGCTCCAGCC CGCTGGATCA GACGGTCATT GGCAGACGCA CGAACATGTA
501 TCTGTTCTTC GCGGGGGACG ACGGGAAAAT CTACAGGGCG AGCATGCCTA
551 TCGGTAACCT CCCCAGGAGC TTCGGTTCGA CGTCAACGGT GGTCTCTGAGC
601 GATGAAAGGA ACAATCTGTT TGAGGCAGTT CAGGTCTATA CCGTCTCAGG
651 GCAGAAGCAA TATCTCATGA TTGTCGAGGC AATAGGCGCA AATGGCCGGT
701 ATTTCCGGTC CTTACACAGC ACAAACCTCG GCGGCACATG GACTCCGCAA
751 GCCACCAGCG AAAGTCAGCC GTTGGCCGGT AAGGCAAACA GTGGCGCTAC
801 CTGGACAAAC GACATCAGTC ATGGTGATCT AATTCGTAGC AACCCCTGATC
851 AGACAATGAC TATCGACCCT TGCAATCTGC AGTTCCTGTA CCAGGGGAGA
901 GCGACAAACT CTGGCGGCGA CTACGGCCTC TTGCCCTATC GACCAGGGCT
951 GCTAACTCTC CAGCGC

```

Figura 8A: secuencia de aminoácidos de Abf2 (SEC Id No. 11)

```

1  MELKALSAVV LSFVTLVAAA PATCTLPSTY RWNSTGALAS PKSGWVSLKD
51  FSHVIYNGQH LVWGSTHDTG TIWGS MNFGL FSDWSNMATA SQNKMTPGTV
101 APTVFYFAPK NIWVLAYQWG PTTFSYLTSS NPSSVNGWSS PQPLFSGSIS
151 GSSPLDQTVI GDSTNMYLFF AGDDGKIYRA SMPIGNFPPG FGSTSTVVLS
201 DERNNLFEAV QVYTVSGQKQ YLMIVEAIGA NGRYFRSFTA TNLGGTWTTPQ
251 ATSESQPFAG KANS GATWTN DISHGDLIRS NPDQMTIDP CNLQFLYQGR
301 ATNSGGDYGL LPYR PGLLTL QR

```

Figura 8B: secuencia señal de Abf2 (SEC Id No. 12)

```

1  MELKALSAVV LSFVTLVAAA

```

Figura 8C: secuencia de proteína madura de Abf2 (SEC Id No. 13)

3)

```

1  APATCTLPST YRWNSTGALA SPKSGWVSLK DFSHVIYNGQ HLWVGSTHDT
51  GTIWGS MNFG LFS DWSNMAT ASQNKMTPGT VAPT VFYFAP KNIWVLAYQW
101 GP TTF SYLTS SNPSSVNGWS SPQPLFSGSI SGSSPLDQTV IGDSTNMYLF
151 FAGDDGKIYR ASMPIGNFPPG SFGSTSTVVL SDERNNLFEA VQVYTVSGQK
201 QYLMIVEAIG ANGRYFRSFT ATNLGGTWTTP QATSESQPFA GKANS GATWT
251 NDISHGDLIR SNP DQMTID PCNLQFLYQG RATNSGGDYG LLPYR PGLLTL
301 LQR

```

Figura 9: alineación de la familia de GH62:
 Abf2 es un miembro de la familia de glicosil hidrolasas 62.
 Se prevé que tiene una secuencia señal N-terminal de 19 aminoácidos.

		1	50
C.carbonum ARP1	(1)	-----MRFVFDLSFSAAAVALLASTASAO-----	
S. thermoviolaceus stxIV	(1)	MSFHRSEPPFRPKRLFGVIAPELLLAGVMSTQPAGAATVVPDSDVQGTGRQS	
Abf2	(1)	-----MELKRLSAVVLGFVTLVAAP-----	
Consensus	(1)	L FMP KAFSALALALLA VASAO	
		51	100
C.carbonum ARP1	(25)	-----SILKLTSSKLTISGILQKLNANLITSSLSKRII-----	
S. thermoviolaceus stxIV	(51)	QLTDGPGTRASLSTPRSTITLQRELVSSKAMTVPYKQILG	
Abf2	(22)	-----ATLHLSSTPRNITLALSKYVSSKSHVIYKQILG	
Consensus	(51)	ASC LPSTYRWTSTGALAQPKSGWVSLKDFITV YNGQHLV	
		101	150
C.carbonum ARP1	(65)	YATDHTA SKYKMASSPQSGHEIASSSITATPFTAVGFLPRKAVIN	
S. thermoviolaceus stxIV	(101)	YATTHVTRWLNDEPFGDWQATLQINAMNSPTVQMLVYQPFID	
Abf2	(63)	WGSTYVITLWLNGLASDWLNWATPSLNKMTPGTVAIVYVQLN	
Consensus	(101)	YATTHDTGTKWGSNMF PFGDWSNMATASQNAM TVAPTLPFYFAPKNI	
		151	200
C.carbonum ARP1	(115)	WVLRGKNIPTTTPMRLSDLTNPISMGGVYIIRKNSGISTGALIPAV	
S. thermoviolaceus stxIV	(151)	WVLRGKNIPTTTPMRLSDLTNPISMGGVYIIRKNSGISTGALIPAV	
Abf2	(113)	WVLRGKNIPTTTPMRLSDLTNPISMGGVYIIRKNSGISTGALIPAV	
Consensus	(151)	WVLRGKNIPTTTPMRLSDLTNPISMGGVYIIRKNSGISTGALIPAV	
		201	250
C.carbonum ARP1	(165)	GDATNMYLFPAGDNGKIYRAMPIGNPFGSFGSTSTVVMGSRNMLPEAV	
S. thermoviolaceus stxIV	(201)	GDATNMYLFPAGDNGKIYRAMPIGNPFGSFGSTSTVVMGSRNMLPEAV	
Abf2	(161)	GDATNMYLFPAGDNGKIYRAMPIGNPFGSFGSTSTVVMGSRNMLPEAV	
Consensus	(201)	GDATNMYLFPAGDNGKIYRAMPIGNPFGSFGSTSTVVMGSRNMLPEAV	
		251	300
C.carbonum ARP1	(215)	QVYTV GQ YLMIVEAIGANG RYFRSPTATSL GSWTPQATSESNPFA	
S. thermoviolaceus stxIV	(251)	QVYTV GQ YLMIVEAIGANG RYFRSPTATSL GSWTPQATSESNPFA	
Abf2	(211)	QVYTV GQ YLMIVEAIGANG RYFRSPTATSL GSWTPQATSESNPFA	
Consensus	(251)	QVYTV GQ YLMIVEAIGANG RYFRSPTATSL GSWTPQATSESNPFA	
		301	350
C.carbonum ARP1	(263)	GKANSATWINDISHGDLIRST DQMTVDPCNLQLLYQGRAPNSGGDYD	
S. thermoviolaceus stxIV	(301)	GKANSATWINDISHGDLIRST DQMTVDPCNLQLLYQGRAPNSGGDYD	
Abf2	(260)	GKANSATWINDISHGDLIRST DQMTVDPCNLQLLYQGRAPNSGGDYD	
Consensus	(301)	GKANSATWINDISHGDLIRST DQMTVDPCNLQLLYQGRAPNSGGDYD	
		351	363
C.carbonum ARP1	(313)	LLPYRPGLLTLQR	
S. thermoviolaceus stxIV	(351)	LLPYRPGLLTLQR	
Abf2	(310)	LLPYRPGLLTLQR	
Consensus	(351)	LLPYRPGLLTLQR	

Figura 10: ADNc de *axe2* (SEC Id No. 14)

ATGCGCGCCC	TCTCACTCTC	CCTCCCCCTC	TCCCTCTCGC	TGCTCGCCGC	50
CAGCTCAACA	GCGGCAACGA	CATGCGCAA	GGCCTCTAC	ATGGTCGTTG	100
CCCGCGGCAG	CGAGGAGCCC	GCCGGCACGG	GCGTGACGGG	CAACCTCACG	150
AGCCAAATCG	CCGCAAAGGT	GCCCGGCAGC	GAGGTCGTGG	CGGTGGACTA	200
CCCGGCCAGC	TTTGACGACT	ACGAGGATTC	CGAGGGCGAC	GGCGTCAAGG	250
CGATGCGGCA	GCTGCTCAAC	AGCTACGCCG	AGGCCTGTCC	GGGAAACAAG	300
ATTGCGGTGC	TGGGATACTC	TCAGGGCGCC	CAAGTCGCAA	CAGACACCAT	350
CTGCGGCGGT	GCCGGCGATC	CGTTTACCAG	CGACAAGGGC	ATGTCTGACG	400
ATGTCATGGA	CGACGTCGTT	GCCGTGGCCA	TTTTCGGAGA	CCCAACCCAT	450
GTCGCCAACA	TGACGTACGA	CCGAGGCACC	AGCATTCA	ACGGGCTCTT	500
CAACCGGAGC	TCGTCCAGCA	TCGAGGTCTG	CAAGTCGTAC	GCCAGCCGCA	550
TCGTCTCGTA	CTGCGACACG	GGCGACATCT	ACTGCGACGC	CGGCAGCAAC	600
TCGACCGTTC	ACCACATGTA	CATCCAGCGC	TACGGCGACG	AAATCGTCGA	650
CTTTGTCTGTC	AGCCAGTTTG	AGAAGAGCAC	CAGCTCGGGA	TCGGGGTCGG	700
GTACTAATGC	CACCACGACC	ACGGCTCCGG	CTCCCACCGT	GTCTCCTACC	750
ACCACCAGCG	GTGGCAACAG	CACAGTGCCT	ACGCGAACCG	GTGGCCCGAC	800
GACGAGTTCG	ACGCAAGGAT	CGGGTGCGAG	TGCTTTGACG	AGCAGTTTGA	850
TGCTGGGAGG	TCTTTTGACG	GTTTTGACGG	CGGTGTCTCA	GATGCTGTGA	900

Figura 11A: secuencia de aminoácidos de AXE2 (SEC Id No. 15)

1	MRALSLSLPL SLSLLAASST AATTCAGLY	MVVARGSEEP	AGTGVTGNLT	50
51	SQIAAKVPGS EVVAVDYPAS FDDYEDSEGD	GVKAMRQLLN	SYAEACPGNK	100
101	IAVLGYSQGA QVATDTICGG AGDPFTSDKG	MSDDVMDDVV	AVAIFGDPTH	150
151	VANMTYDRGT SIHNGLFNRS SSSIEVCKSY	ASRIVSYCDT	GDIYCDAGSN	200
201	STVHHMYIQR YGDEIVDFVV SQFEKSTSSG	SGSGTNATT	TAPAPTVSPT	250
251	TTSGGNSTVP TRTGGPTTSS TQSGASALT	SSLMLGLLT	VLTAVSQML	299

Figura 11B: secuencia señal de AXE2 (SEC Id No. 16)

1	MRALSLSLPL SLSLLAASST A	21
---	--------------------------------	----

Figura 11C: secuencia de proteína madura de AXE2 (SEC Id No. 17)

1	ATTCAKGLYM VVARGSEEP	GTGVTGNLTS	QIAAKVPGSE	VVAVDYPASF	50
51	DDYEDSEGDG VKAMRQLLNS	YAEACPGNKI	AVLGYSQGAQ	VATDTICGGA	100
101	GDPFTSDKGM SDDVMDDVVA	VAIFGDPTHV	ANMTYDRGTS	IHNGLFNRSS	150
151	SSIEVCKSYA SRIVSYCDTG	DIYCDAGSNS	TVHHMYIQR	GDEIVDFVVS	200
201	QFEKSTSSGS GSGTNATTTT	APAPTVSPTT	TSGGNSTVPT	RTGGPTTSST	250
251	QSGASALTS SLMLGLLTV	LTAVSQML			278

Figura 12: alineación de axe2
 Miembro de la familia CE5

Axe2 es un miembro de la familia 5 de carbohidrato esterasas (CE5). Se prevé que tiene una secuencia señal N-terminal de 21 aminoácidos. Presenta un potencial sitio de unión al ancla de GPI en el aminoácido número 274, correspondiente al residuo de serina en la posición 291 en la alineación (Udenfriend, S. y K. Kodukula, 1995. Prediction of ω site in nascent precursor of glycosphatidylinositol proteína. Methods in Enzymology, 250: 57-82)

		1		50
T. reesei	axe1	(1)	MPSVKETITLAIQNFVTCIPVDGEVVKRCCPAIHVFCHEITVSOIY	
	axe2	(1)	--MRALS-SHPILSLVAS-----AATTCARGLYVVMGSEEPAT	
	Consensus	(1)	SLSL LS A LA S T AIHM AR S G	
		51		100
T. reesei	axe1	(51)	SSATVNLVIQAHITSEIVVLCGGQASCGGISVAMVVMITNAA	
	axe2	(44)	INTGNLTSQTAARKVSEVVVDLIS-----FDDIEDEGDVKQMR	
	Consensus	(51)	G SA L I PGS AI YPA Y S G A	
		101		150
T. reesei	axe1	(101)	AAIINPHNSDTOLVLPVDFDNDVYGGPGEGITNTAVPIA	
	axe2	(87)	QLISYAENINKNIAVLVAVATDTIAG---DFPDSKQMSD	
	Consensus	(101)	IN F ACP I LIGYSQGAQI ICGGA S LS	
		151		200
T. reesei	axe1	(151)	GAVSAKGAIFMQLRNHGLPNTVICTIQFDAPAG---PVCPSLR	
	axe2	(134)	DVMDDVVAIFLTHVANNTDRVSIHMLFNSSSIEVCKSYLR	
	Consensus	(151)	M V A GDP I L Y GT NG R A ASK	
		201		250
T. reesei	axe1	(198)	IKKTAAPVCTNDPNHQCCEQQALAVDELLSSGQPPPIG	
	axe2	(184)	VVAVTGLIDALSNSTLHMILRDEIVDVMPEKSTISGSLIS	
	Consensus	(201)	I SYCD AD YC G VH Y Q YG L PI SQ S G	
		251		300
T. reesei	axe1	(248)	-----GITSISRISRTISPGVQHWGQCG---IQMTGPIQCE	
	axe2	(234)	GINATTTTAPVSTSGNFWRRGGPTSSSTQSSASALSSL	
	Consensus	(251)	AP T PST G S PT T G G SA T	
		301		316
T. reesei	axe1	(287)	SGITCOVTSOWYIC	
	axe2	(284)	MLGGLT/LTAVEM	
	Consensus	(301)	I SQ L	

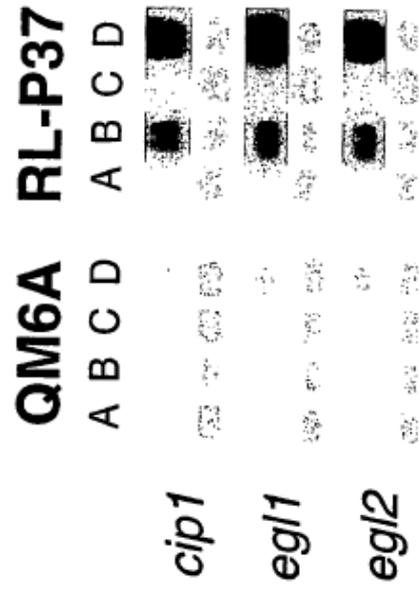


FIGURA 13: transferencia Northern

FIGURA 14: Análisis por microarray

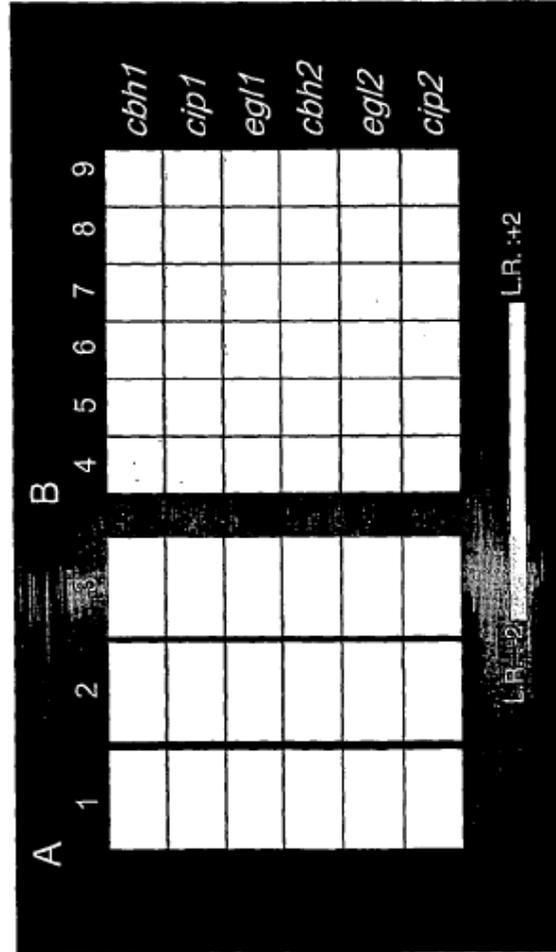
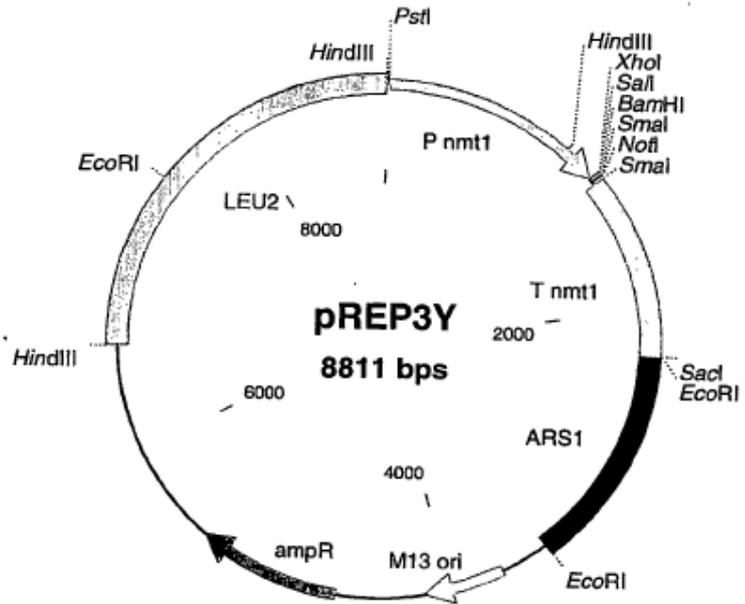


FIGURA 15



ES 2 371 362 T3

FIGURA 16: Alineación de Cip1 con la hidrolasa secretada potencial de Streptomyces coelicolor A3 (número de acceso CAA18323)

```

1 50
cip1 (1) -----
Streptomyces coelicolor A3 (1) MRTRVLRLLRRPWTAAVAVALVVSVLVAMPASGAAAAACRVVDYGVDAWA
Consensus (1) -----
51 100
cip1 (1) -----
Streptomyces coelicolor A3 (51) GGYTARVRIITNLGPAVSDWRLTWTYTGDDQVTSANNATVTTQIGASVVAVD
Consensus (51) -----
101 150
cip1 (1) -----
Streptomyces coelicolor A3 (101) AGWNGAVSTGGTAEFGLQGTWRSADPAPDDFALNGTSCGGGGTPTATPTT
Consensus (101) -----
151 200
cip1 (1) -----MVRRLALLALGALSTLSMAQISDI FESGWDQTKMPISG
Streptomyces coelicolor A3 (151) SPTAPPTTPPTTPPTTPPPAECCDAVICSGFEI QAGPEPSGDLRPTAI
Consensus (151) T A I DD W SAP
201 250
cip1 (40) LK NQGIVSLITTAASISIMKVVGPNSVGIIPGTAQVPIGD--V
Streptomyces coelicolor A3 (201) LKGGTAAAVSASIAIGRLKQDRAIPNDAVAHADLSSVGPVM
Consensus (201) DCN GT ALDS VAHAG SLKV G GYC H F A T S M
251 300
cip1 (87) LVAWIHLQVGGSNVYTIHETASIKHPTASISQVLDY
Streptomyces coelicolor A3 (250) LVAWIHLQVGGSNVYTIHETASIKHPTASISQVLDY
Consensus (251) YVR WIR TAL S HVTFI MPDSAQGGK LRIGGQ L WNRESDDA
301 350
cip1 (137) FTRDLINLTASTVTVAGAPFEMHLGDDGTTEWINGSLITMTVG
Streptomyces coelicolor A3 (300) FTRDLINLTASTVTVAGAPFEMHLGDDGTTEWINGSLITMTVG
Consensus (301) TLP SF GIA S LPT AFQC F I T A IDTWL IPGL
351 400
cip1 (186) PGVDNPNAGRTRASYINBIQWNEPAAVSGDVNPHHGLISIAITRV
Streptomyces coelicolor A3 (350) PGVDNPNAGRTRASYINBIQWNEPAAVSGDVNPHHGLISIAITRV
Consensus (351) N D W P TAL GWEAYA TVWFDDIALAS IG
401 450
cip1 (236) GPGSPGGPGSSTTGRSSTSGPTSTSRPSTTIIPPTSRITTTATGPTQTHY
Streptomyces coelicolor A3 (400) -----
Consensus (401) C
451 481
cip1 (286) GQCGGIGYSGPTVCASGTTCOVLNPPYYSQCL
Streptomyces coelicolor A3 (401) -----
Consensus (451)

```

FIGURA 17

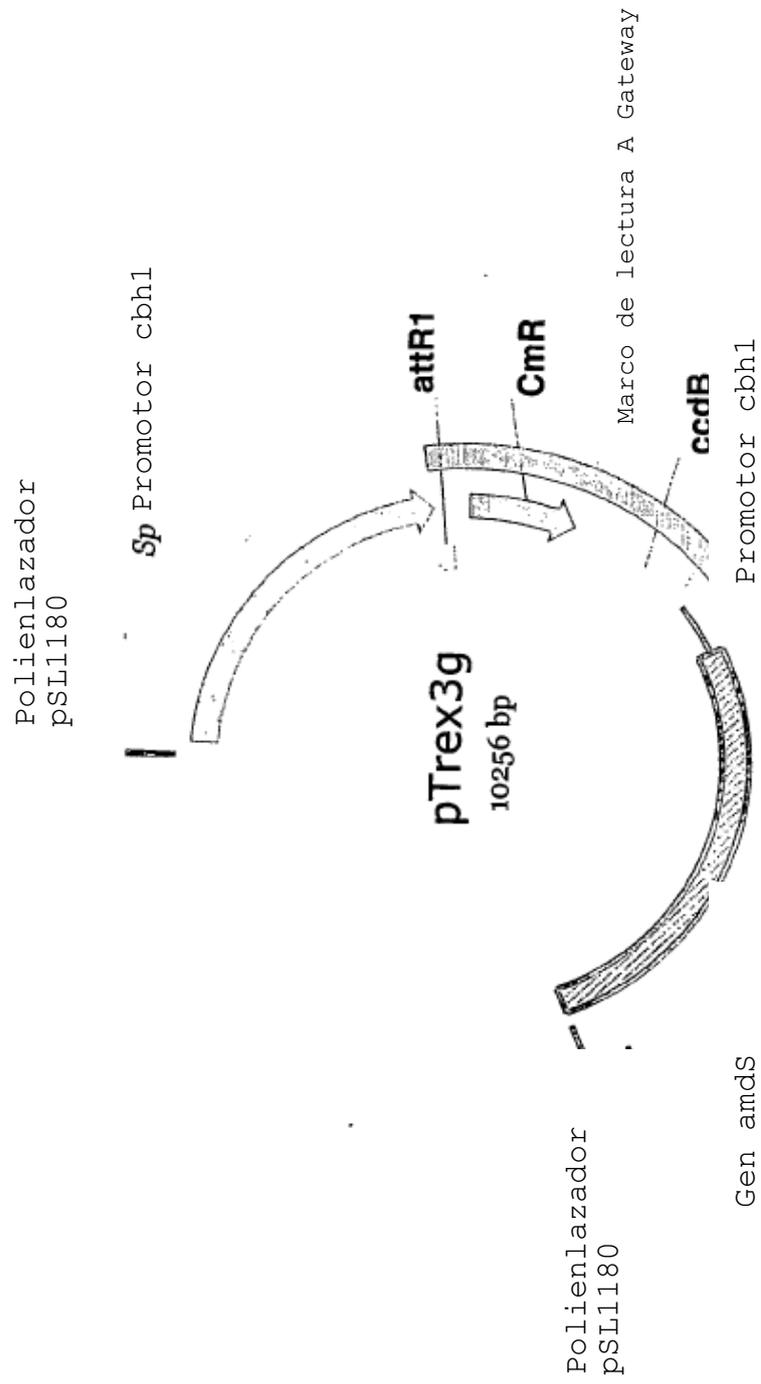
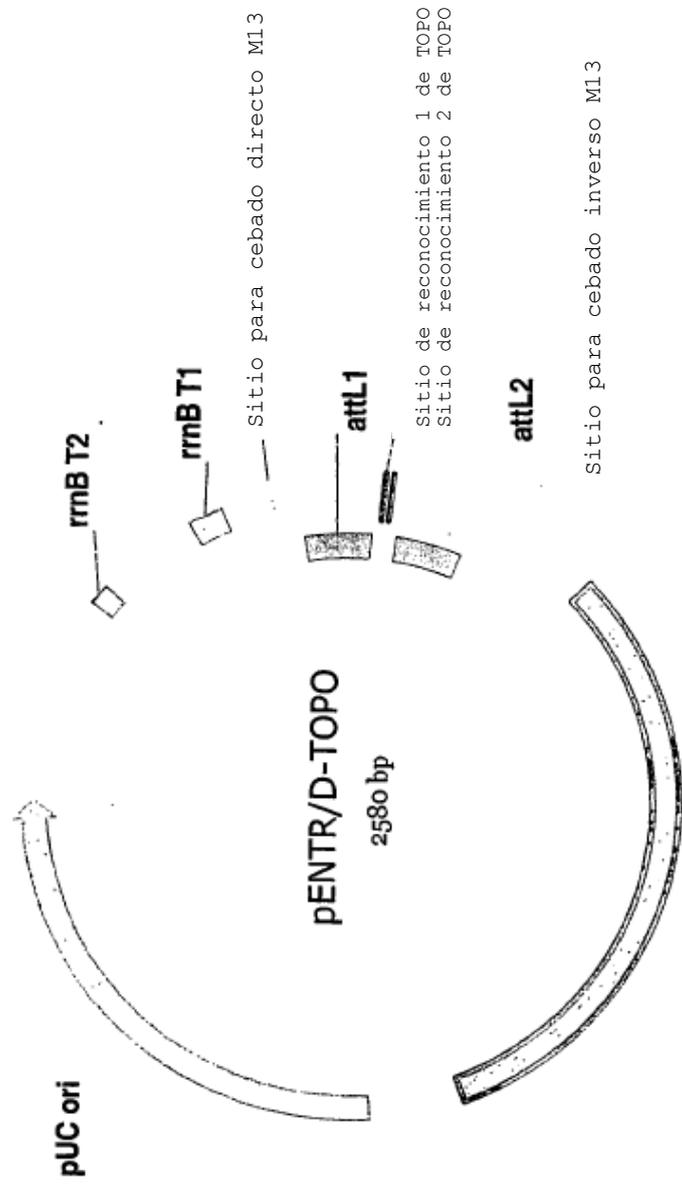


FIGURA 18



Gen de resistencia a kanamicina

FIGURA 19

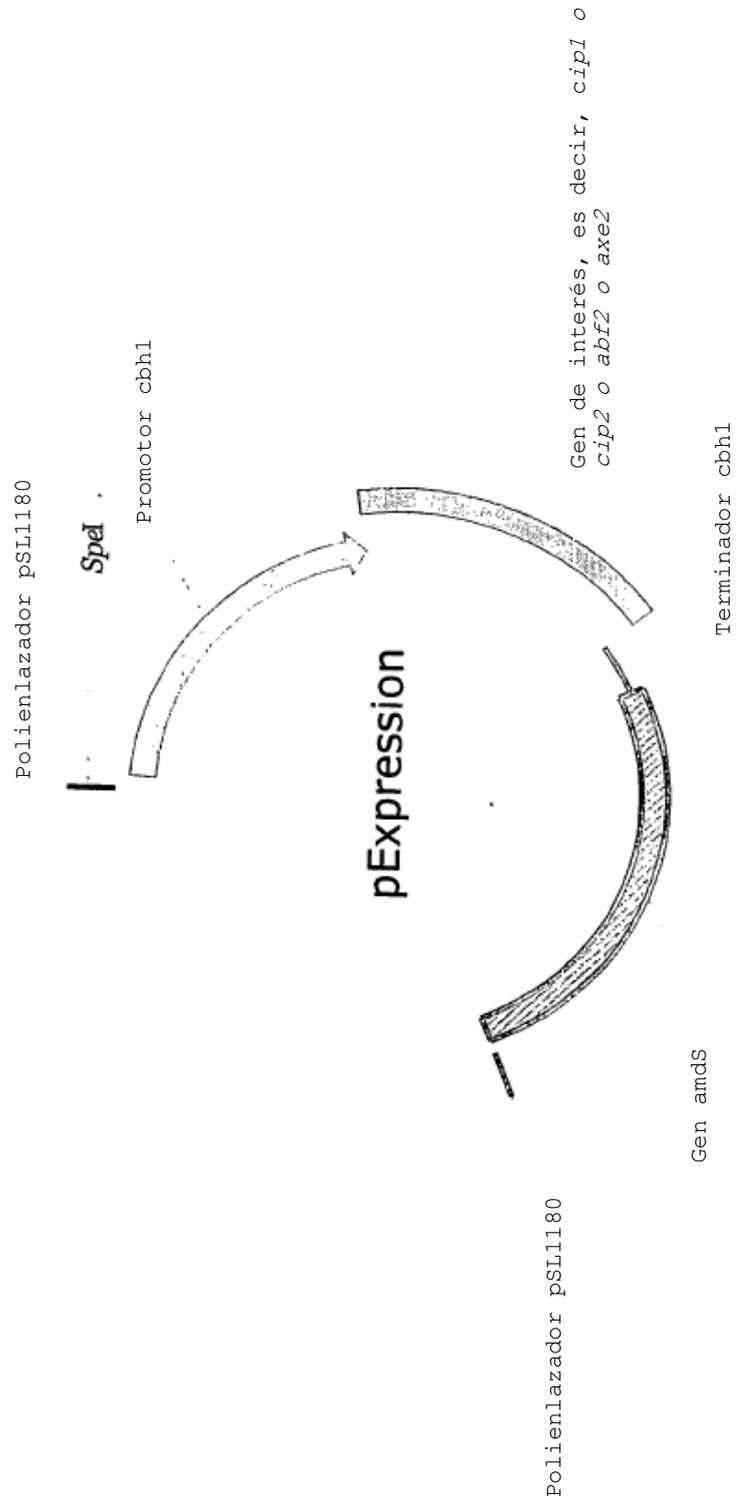


FIGURA 20

Muestras con Cipl en
matraz de agitación

- Muestras:
- 1) Marcadores de peso molecular
 - 2) Transformante #1
 - 3) Transformante #2
 - 4) Transformante #3
 - 5) Transformante #4
 - 6) Transformante #5
 - 7) Transformante #6
 - 8) Transformante #7
 - 9) Transformante #8
 - 10) Transformante #9
 - 11) Transformante #10
 - 12) Transformante #11

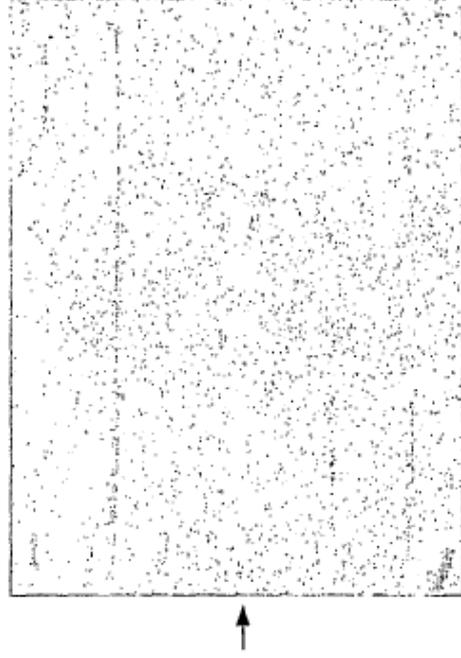


FIGURA 21

Muestras con Cip2 en
matraz de agitación

- Muestras:
- 1) Marcadores de peso molecular
 - 2) Cepa con quad eliminado
 - 3) Transformante #1
 - 4) Transformante #2
 - 5) Transformante #3
 - 6) Transformante #4
 - 7) Transformante #5
 - 8) Transformante #6
 - 9) Transformante #7
 - 10) Transformante #8
 - 11) Transformante #9
 - 12) Transformante #10

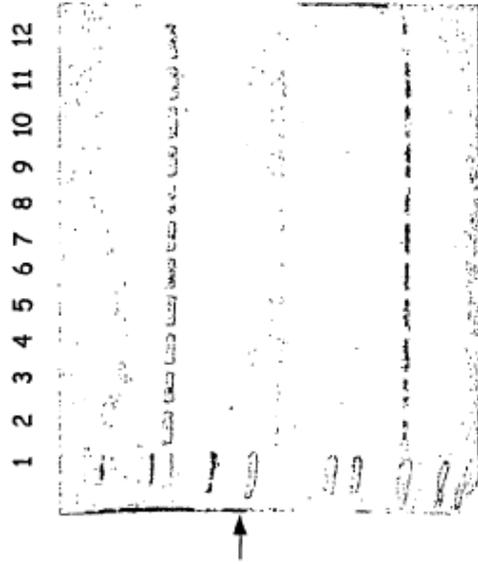


FIGURA 22

Muestras con Abf2 en
matraz de agitación



Muestras:

- 1) Cepa con quad eliminado
- 2) Marcadores de peso molecular
- 3) Transformante Abf2