

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 533 923**

51 Int. Cl.:

C12N 9/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.06.2002** **E 10179521 (9)**

97 Fecha y número de publicación de la concesión europea: **07.01.2015** **EP 2298868**

54 Título: **Polipéptidos con actividad de celobiohidrolasa I y polinucleótidos que codifican los mismos**

30 Prioridad:

26.06.2001 DK 200101000

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

16.04.2015

73 Titular/es:

**NOVOZYMES A/S (100.0%)
Krogshøjvej 36
2880 Bagsvaerd, DK**

72 Inventor/es:

**LANGE, LENE;
WU, WENPING;
AUBERT, DOMINIQUE;
LANDVIK, SARA;
SCHNORR, KIRK MATTHEW y
CLAUSEN, IB GROTH**

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

Observaciones :

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

ES 2 533 923 T3

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

DESCRIPCIÓN

Polipéptidos con actividad de celobiohidrolasa I y polinucleótidos que codifican los mismos

5 Campo de la invención

[0001] La presente invención se refiere a polipéptidos con actividad de celobiohidrolasa I (también llamada CBH I o CBH 1) y polinucleótidos con una secuencia nucleotídica que codifica para los polipéptidos. La invención también se refiere a vectores y células huésped que incluyen constructos de ácidos nucleicos al igual que métodos para producir y usar los polipéptidos.

Antecedentes de la invención

[0002] Celulosa es una materia prima industrial importante y una fuente de energía renovable. La estructura física y la morfología de celulosa nativa son complejas y los detalles finos de su estructura han sido difíciles de determinar experimentalmente. No obstante, la composición química de la celulosa es simple, consiste en residuos de D-glucosa enlazados por enlaces beta-1,4-glicosídicos para formar polímeros lineales con longitud de cadenas de más de 10.000 residuos glicosídicos.

[0003] Para ser eficaz, la digestión de celulosa requiere diferentes tipos de enzimas que actúan cooperativamente. Al menos tres categorías de enzimas son necesarias para convertir celulosa en glucosa: endo (1,4)-beta-D-glucanasas (EC 3,2,1,4) que cortan al azar las cadenas de celulosa; celobiohidrolasas (EC 3,2,1,91) que disocian unidades de celobiosil de las extremidades de cadena de celulosa y beta-glucosidasas (EC 3,2,1,21) que convierten celobiosa y celodextrinas solubles en glucosa. Entre estas tres categorías de enzimas implicadas en la biodegradación de celulosa, celobiohidrolasas son las enzimas clave para la degradación de celulosa cristalina nativa.

[0004] Exo-celobiohidrolasas (celobiohidrolasa I o CBH I) se refieren a las celobiohidrolasas que degradan celulosa mediante hidrolización de la celobiosa del extremo no-reducido de las cadenas de polímero de celulosa.

[0005] de Oliviera Alzevedo et al han divulgado un polipéptido de gen CBH-1 de *Humicola grisea* var. *thermoidea* en Nucleic Acids Res. 1990, 1990, 18:668-668 (SWISSPROT: P15828, EBI: X17258), que tiene un 81,3 % de identidad con la SEC ID N° 4. Kvachadze et al han publicado el efecto del tratamiento de una cepa de *Chaetomium thermophile* con luz uv y etilenimina en Mikrobiologiya. 1997, 66(5):644-649. Ramesh et al. describen en Biochimica, Biophysica Acta 1989, 993: 266-274 la purificación y caracterización de dos celobiohidrolasas a partir de *Chaetomium thermophile* var. *coprophile*.

[0006] Es un objeto de la presente invención proporcionar polipéptidos mejorados con actividad de celobiohidrolasa I y polinucleótidos que codifiquen los polipéptidos. Los polipéptidos mejorados pueden tener actividad específica mejorada y/o estabilidad mejorada, en particular termoestabilidad mejorada. Los polipéptidos pueden también tener una capacidad mejorada para resistir inhibición por celobiosa.

Resumen de la invención

[0007] En un primer aspecto, la presente invención se refiere a un polipéptido con actividad de celobiohidrolasa I, seleccionado del grupo que consiste en:

- (a) un polipéptido que comprende una secuencia de aminoácidos que tiene como mínimo 95% de identidad con aminoácidos 1 a 529 de SEC ID n° 4, y
- (b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 95% de identidad con el polipéptido codificado por nucleótidos 1 a 1578 de SEC ID n° 3,

[0008] En un segundo aspecto, la presente invención se refiere a un polinucleótido con una secuencia de nucleótidos que codifica para el polipéptido de la invención.

[0009] En un tercer aspecto, la presente invención se refiere a un constructo de ácidos nucleicos que comprende la secuencia de nucleótidos, que codifica para el polipéptido de la invención, operativamente enlazado a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.

[0010] En un cuarto aspecto, la presente invención se refiere a un vector de expresión recombinante que comprende el constructo de ácidos nucleicos de la invención.

[0011] En un quinto aspecto, la presente invención se refiere a una célula huésped recombinante que comprende el constructo de ácidos nucleicos de la invención.

[0012] En un sexto aspecto, la presente invención se refiere a un método para la producción de un polipéptido de la invención, el método incluye:

- (a) cultivo de una célula huésped recombinante de la invención bajo condiciones propicias para la producción del

polipéptido; y
(b) recuperación el polipéptido.

[0013] En un séptimo aspecto, la presente invención se refiere a un método para producción in situ de un polipéptido de la invención, el método incluye:

- (a) cultivo de una célula huésped recombinante de la invención bajo condiciones propicias para la producción del polipéptido; y
- (b) contacto del polipéptido con un sustrato deseado sin recuperación previa del polipéptido.

Definiciones

[0014] Antes de discutir la presente invención en detalles adicionales, los siguientes términos y convenios serán definidos primeramente:

Polipéptido substancialmente puro: en el presente contexto, el término "polipéptido substancialmente puro" significa una preparación de polipéptido que contiene como mucho 10% en peso de otro material polipéptido con el cual es originalmente asociado (porcentajes inferiores de otro material polipéptido son preferidos, por ejemplo como mucho 8% en peso, como mucho 6% en peso, como mucho 5% en peso, como mucho 4%, como mucho 3% en peso, como mucho 2% en peso, como mucho 1% en peso y como mucho ½% en peso). Así, se prefiere que el polipéptido substancialmente puro sea al menos 92% puro, es decir que el polipéptido constituya al menos 92% en peso del material polipéptido total presente en la preparación, y porcentajes más altos son preferidos tal como al menos 94% puro, al menos 95% puro, al menos 96% puro, al menos 96% puro, al menos 97% puro, al menos 98% puro, al menos 99% y como mucho 99.5% puro. Los polipéptidos descritos aquí están preferiblemente en una forma substancialmente pura. En particular, se prefiere que los polipéptidos descritos aquí estén en "forma esencialmente pura", es decir, que la preparación de polipéptido esté esencialmente libre de otro material polipéptido con el cual es originalmente asociado. Esto puede ser realizado, por ejemplo, por preparación del polipéptido mediante bien conocidos métodos recombinantes. Aquí, el término "polipéptido substancialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada".

[0015] Actividad de celobiohidrolasa I: el término "actividad de celobiohidrolasa I" es definido aquí como una actividad de celulosa 1,4-beta-celobiosidasa (también llamada exo-glucanasa, exo-celobiohidrolasa o 1,4-beta-celobiohidrolasa), como se define en la clase enzimática EC 3.2.1.91, que cataliza la hidrólisis de enlaces de 1,4-beta-D-glucosídico en celulosa y celotetraosa, liberando celobiosa de las extremidades no-reducidas de las cadenas.

[0016] Para fines de la presente invención, actividad de celobiohidrolasa I puede ser determinada según al procedimiento descrito en el ejemplo 2.

[0017] En una forma de realización, actividad de celobiohidrolasa I puede ser determinada según el procedimiento descrito en Deshpande MV et al., Methods in Enzymology, pp. 126-130 (1988): "Selective Assay for Exo-1,4-Beta-Glucanases". Según este procedimiento, una unidad de actividad de celobiohidrolasa I (actividad de escisión de enlace aglucónico) es definida como 1,0 µmol de p-nitrofenol producido por minuto a 50°C, pH 5,0.

[0018] Los polipéptidos de la presente invención deberían preferiblemente tener al menos 20% de la actividad de celobiohidrolasa I de un polipéptido que consiste en una secuencia de aminoácidos seleccionados del grupo que consiste en SEC ID nº 4. En una forma de realización particular preferida, los polipéptidos deberían tener al menos 40%, tal como al menos 50%, preferiblemente al menos 60%, tal como al menos 70%, más preferiblemente al menos 80%, tal como al menos 90%, muchos preferiblemente al menos 95%, tal como al rededor de o al menos 100% de la actividad de celobiohidrolasa I del polipéptido que consiste en la secuencia de aminoácidos seleccionados de los aminoácidos 1 a 526 de SEC ID nº 4.

[0019] Identidad: en el presente contexto, la homología entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos es descrita por el parámetro "identidad".

[0020] Para fines de la presente invención, el grado de identidad entre dos secuencias de aminoácidos es determinado usando el programa FASTA incluido en la versión 2.0x del paquete del programa FASTA (ver W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448 ; and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA", Methods in Enzymology 183:63-98). La matriz de marcado usada fue BLOSUM50, penalización del espacio fue -12, y penalización de extensión del espacio fue -2.

[0021] El grado de identidad entre dos secuencias de nucleótidos es determinado usando el mismo algoritmo y paquete de software, como se ha descrito anteriormente. La matriz de marcado usada fue la matriz de identidad, la penalización de gap fue -16 y penalización de extensión de gap fue -4.

[0022] Fragmento: cuando se usa aquí, un "fragmento" de una secuencia seleccionada de SEC ID nº 2 es un polipéptido con uno o más aminoácidos eliminados del terminal amino y/o carboxilo de esta secuencia de aminoácidos. Preferiblemente, un fragmento es un polipéptido con la secuencia de aminoácidos eliminada correspondiente al "dominio de enlace de celulosa" y/o el "dominio enlazador" de Celobiohidrolasa I de *Trichoderma reesei* como está

descrito en el número de accesión de SWISS-PROT P00725. Más preferiblemente, un fragmento incluye la secuencia de aminoácidos que corresponde al "dominio catálico" de Celobiohidrolasa I de *Trichoderma reesei* como está descrito en el número de accesión de SWISS-PROT P00725. De la forma más preferible, un fragmento contiene al menos 434 residuos de aminoácidos, por ejemplo, los residuos de aminoácido seleccionados de aminoácidos 1 a 434 de SEC ID n° 4. En particular, un fragmento contiene al menos 215 residuos de aminoácidos, por ejemplo, los residuos de aminoácidos seleccionados de los aminoácidos 200 a 434 de SEC ID n° 4.

[0023] Variante alélica: en el presente contexto, el término "variante alélica" denota dos o más formas cualquiera alternativas de un gen que ocupa la misma localización cromosómica. Variante alélica surge naturalmente a través de mutación y puede suponer polimorfismo dentro de poblaciones. Mutaciones de gen pueden ser silenciosas (ningún cambio en el polipéptido codificado) o pueden codificar polipéptidos con secuencias de aminoácidos alteradas. Una variante alélica de un polipéptido es un polipéptido codificado por una variante alélica de un gen.

[0024] Polinucleótido substancialmente puro: el término "polinucleótido substancialmente puro" como se utiliza en este caso se refiere a una preparación polinucleótida, donde el polinucleótido ha sido quitado de su ambiente natural genético, y está así libre de otras secuencias de codificación indeseadas o extrañas y está en una forma adecuada para su uso dentro de sistemas de producción de proteína creada genéticamente. Así, un polinucleótido substancialmente puro contiene como mucho 10% en peso de otro material polinucleótido con el cual es originalmente asociado (porcentajes inferiores de otro material polinucleótido son preferidos, por ejemplo como mucho 8% en peso, como mucho 6% en peso, como mucho 5% en peso, como mucho 4% como mucho 3% en peso, como mucho 2% en peso, como mucho 1% en peso, y como mucho ½% por peso). Un polinucleótido substancialmente puro puede, no obstante, incluir regiones no traducidas 5' y 3' de origen natural, tales como promotores y terminadores. Se prefiere que el polinucleótido substancialmente puro sea al menos 92% puro, es decir que el polinucleótido constituya al menos 92% en peso del material de polinucleótido total presente en la preparación, y porcentajes más altos son preferidos tal como al menos 94% puro, al menos 95% puro, al menos 96% puro, al menos 96% puro, al menos 97% puro, al menos 98% puro, al menos 99% y, como mucho, 99,5% puro. Los polinucleótidos descritos aquí están preferiblemente en una forma substancialmente pura. En particular, se prefiere que los polinucleótidos descritos aquí estén en "forma esencialmente pura", es decir que la preparación polinucleótida esté esencialmente libre de otro material polinucleótido con el cual sea originalmente asociada. Aquí, el término "polinucleótido substancialmente puro" es sinónimo de los términos "polinucleótido aislado" y "polinucleótido en forma aislada".

[0025] Modificación(es): en el contexto de la presente invención el término "modificación(es)" significa cualquier modificación química de un polipéptido que consista en una secuencia de aminoácidos seleccionado de SEC ID n° 4, al igual que manipulación genética del ADN que codifica ese polipéptido. La modificación(es) puede ser sustitución(es) de la cadena(s) lateral de aminoácido, sustitución(es), delección(es) y/o inserciones(s) en el o los aminoácido(s) de interés.

[0026] Variante artificial: cuando se usa aquí, el término "variante artificial" significa un polipéptido con actividad de celobiohidrolasa I, que ha sido producido por un organismo que está expresando un gen modificado en comparación con la SEC ID n° 3. El gen modificado, donde dicha variante se produce cuando expresada en un huésped adecuado, se obtiene a través de intervención humana por modificación de una secuencia de nucleótido seleccionada de la SEC ID n° 3.

[0027] ADNc: el término "ADNc" cuando se usa en el presente contexto, se destina a cubrir una molécula de ADN que se puede preparar por transcripción inversa de una molécula de ARNm dividida madura derivada de una célula eucariota. A ADNc le faltan las secuencias de intrón que están normalmente presentes en el ADN genómico correspondiente. La transcripción de ARN primario inicial es un precursor para ARNm y va a través de una serie de eventos de tratamiento antes de aparecer como ARNm maduro dividido. Estos eventos incluyen la eliminación de secuencias de intrón por un proceso llamado empalme. Cuando ADNc se deriva de ARNm, por lo tanto, le faltan secuencias de intrón.

[0028] Constructo de ácidos nucleicos: cuando se usa aquí, el término "constructo de ácidos nucleicos" significa una molécula de ácido nucleico, mono o bicatenario, que se aísla de un gen de origen natural o que ha sido modificada para contener segmentos de ácidos nucleicos de manera que de otra forma no existirían en la naturaleza. El término constructo de ácidos nucleicos es sinónimos del término "cassette de expresión" cuando el constructo de ácidos nucleicos contiene las secuencias de control requerida para la expresión de una secuencia codificante de la presente invención.

[0029] Secuencia de control: el término "secuencias de control" es definido aquí para incluir todos los componentes que son necesarios o ventajosos para la expresión de un polipéptido de la presente invención. Cada secuencia de control puede ser nativa o foránea a la secuencia de nucleótidos que codifica el polipéptido. Tales secuencias de control incluyen, pero de forma no limitativa, una secuencia de poliadenilación guía, secuencia de propéptido, promotor, secuencia de péptido señal, y terminador de transcripción. Como mínimo, las secuencias de control incluyen un promotor y señales de parada traduccional y transcripcional. Las secuencias de control se pueden proporcionar con enlaces para introducir sitios de restricción específica facilitando la ligadura de las secuencias de control con la región de codificación de la secuencia de nucleótidos que codifica un polipéptido.

[0030] Operativamente enlazado: el término "operativamente enlazado" es definido aquí como una configuración en la que una secuencia de control es apropiadamente colocada en una posición en relación a la secuencia de codificación de la secuencia de ADN de tal forma que la secuencia de control dirige la expresión de un polipéptido.

[0031] Secuencia codificante: cuando se usa aquí el término "secuencia codificante" se destina a cubrir una secuencia de nucleótidos que especifica directamente la secuencia de aminoácidos de su producto proteínico. Los bordes de la secuencia codificante están generalmente determinados por un marco de lectura abierto, que comienza normalmente con el codón de inicio ATG. La secuencia codificante típicamente incluye ADN, ADNc y secuencias de nucleótidos recombinantes.

[0032] Expresión: en el presente contexto, el término "expresión" incluye cualquier paso implicado en la producción del polipéptido incluyendo, pero no limitado a, transcripción, modificación postranscripcional, traducción, modificación postraduccional y secreción.

[0033] Vector de expresión: en el presente contexto, el término "vector de expresión" cubre una molécula de ADN, lineal o circular, que incluye un segmento que codifica un polipéptido de la invención y que es operativamente enlazado a segmentos adicionales que proveen para su transcripción.

[0034] Célula huésped: el término "célula huésped", como se utiliza en este caso, incluye cualquier tipo de célula que sea susceptible de transformación con un constructo de ácidos nucleicos.

[0035] Los términos "sonda polinucleótida", "hibridación", al igual que las varias condiciones de astringencia se definen en la sección titulada "polipéptidos con actividad de celobiohidrolasa I".

[0036] Termoestabilidad: el término "termoestabilidad", como se utiliza en este caso, es medido como está descrito en el ejemplo 2.

Descripción detallada de la invención

Polipéptidos con actividad de celobiohidrolasa I

[0037] En una primera forma de realización, la presente invención se refiere a polipéptidos con actividad de celobiohidrolasa I y donde los polipéptidos incluyen, preferiblemente, una secuencia de aminoácidos que tiene un grado de identidad con una secuencia de aminoácido de SEC ID nº4, (es decir, el polipéptido maduro) de al menos 95%, como por ejemplo al menos 96%, tal como al menos 97%, incluso más preferiblemente al menos 98%, más preferiblemente al menos 95%, por ejemplo al menos 96%, tal como al menos 97% e incluso más preferiblemente al menos 98%, tal como al menos 99% (de ahora en adelante "polipéptidos homólogos"). La secuencia de aminoácidos puede diferir en, como mucho, diez aminoácidos (p. ej. en diez aminoácidos), en particular en como mucho cinco aminoácidos (p. ej. en cinco aminoácidos), tal como en como mucho cuatro aminoácidos (p. ej. en cuatro aminoácidos), por ejemplo en como mucho tres aminoácidos (p. ej. en tres aminoácidos) de una secuencia de aminoácidos seleccionada de SEC ID nº 4. La secuencia de aminoácidos puede diferir en como mucho dos aminoácidos (p. ej. en dos aminoácidos), tal como en un aminoácido de una secuencia de aminoácidos seleccionados de SEC ID nº 4.

[0038] Preferiblemente, los polipéptidos de la presente invención comprenden la secuencia de aminoácidos SEC ID nº 4; una variante alélica de la misma; o un fragmento de la misma que tenga una actividad de celobiohidrolasa I. En otra forma de realización preferida, el polipéptido de la presente invención consiste en la secuencia de aminoácidos SEC ID nº 4.

[0039] El polipéptido de la invención puede ser una celobiohidrolasa I de tipo salvaje identificada y aislada de una fuente natural. Tales polipéptidos de tipo salvaje pueden ser específicamente seleccionados por técnicas estándar conocidas en la técnica, tal como selección molecular como está descrito en el ejemplo 1. Además, el polipéptido de la invención se puede preparar por la técnica de redistribución de ADN, tal como descrito en J.E. Ness et al. Nature Biotechnology 17, 893-896 (1999). Tales variantes artificiales pueden ser construidas por técnicas estándar conocidas en la técnica, tales como mutagénesis dirigida al sitio/aleatoria del polipéptido que comprende la secuencia de aminoácidos SEC ID nº 4. Los cambios de aminoácidos (en la variante artificial al igual que en los polipéptidos de tipo salvaje) son de naturaleza menor, esto es, sustituciones prudentes de aminoácido que no afectan significativamente al pliegue y/o actividad de la proteína; delecciones pequeñas, típicamente de uno a aproximadamente 30 aminoácidos; pequeñas extensiones amino terminal o carboxilo terminal, tales como un residuo de metionina amino terminal; un pequeño péptido enlazador de hasta aproximadamente 20-25 residuos; o una pequeña extensión que facilita purificación cambiando la carga de red u otra función, tal como un tracto de polihistidina, un epítipo antigénico o un dominio de unión.

[0040] Ejemplos de sustituciones menores están dentro del grupo de aminoácidos básicos (arginina, lisina e histidina), aminoácidos ácidos (ácido glutámico y ácido aspártico), aminoácidos polares (glutamina y asparagina), aminoácidos hidrofóbicos (leucina, isoleucina, valina y metionina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina) y pequeños aminoácidos (glicina, alanina, serina y treonina). Sustituciones de aminoácidos que generalmente no alteran

la actividad específica se conocen en la técnica y son descritas, por ejemplo, por H. Neurath and R.L. Hill, 1979, In, The Proteins, Academic Press, New York . Los intercambios que más frecuentemente ocurren son Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, y Asp/Glyal igual que éstos a la inversa.

[0041] En una forma de realización interesante de la descripción, los cambios de aminoácidos son de tal naturaleza que las propiedades físico químicas de los polipéptidos son alteradas. Por ejemplo, cambios de aminoácidos pueden ser realizados, lo que mejora la termoestabilidad del polipéptido, lo que altera la especificidad de sustrato, cambia el pH óptimo y similares.

[0042] Preferiblemente, el número de tales sustituciones, deleciones y/o inserciones comparadas con una secuencia de aminoácidos de SEC ID nº 4 es como máximo 10, tal como máximo 9, por ejemplo como máximo 8, más preferiblemente como máximo 7, por ejemplo como máximo 6, tal como máximo 5, de la forma más preferible como máximo 4, por ejemplo como máximo 3, tal como máximo 2, en particular como máximo 1.

[0043] Los presentes inventores han aislado secuencias de nucleótidos que codifican polipéptidos con actividad de celobiohidrolasa I del microorganismo *Acremonium thermophilum*. Así, la presente divulgación se refiere a polipéptidos que incluyen una secuencia de aminoácido que tiene al menos 95% de identidad con el polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en el organismo de *Chaetoniium thermophilum*. En una forma de realización interesante de la invención, el polipéptido comprende una secuencia de aminoácido que tiene al menos 95%, por ejemplo al menos 96%, tal como al menos 97%, e incluso más preferiblemente al menos 98%, tal como al menos 99% de identidad con el polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en el organismo de *Chaetoniium thermophilum* (de aquí en adelante "polipéptidos homólogos". En una forma de realización interesante, la secuencia de aminoácidos difiere en como mucho diez aminoácidos (p. ej. en diez aminoácidos), en particular en como mucho cinco aminoácidos (p. ej. en cinco aminoácidos), tal como en como mucho cuatro aminoácidos (p. ej. en cuatro aminoácidos), por ejemplo en como mucho tres aminoácidos (p. ej. en tres aminoácidos) del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en un organismo seleccionado del grupo que consiste en *Chaetoniium thermophilum*. En una forma de realización particular interesante, la secuencia de aminoácidos difiere en como mucho dos aminoácidos (p. ej. en dos aminoácidos), tal como en un aminoácido del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presentes en un organismo seleccionado del grupo que consiste en *Chaetoniium thermophilum*.

[0044] Preferiblemente, los polipéptidos de la presente invención comprenden la secuencia de aminoácidos del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótido insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC nº 0581. En otra forma de realización preferida, el polipéptido de la presente invención consiste en la secuencia de aminoácidos del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC nº 0581.

[0045] De una forma similar a como se ha descrito anteriormente, el polipéptido de la invención puede ser una variante artificial que comprende, preferiblemente consiste en una secuencia de aminoácidos que tiene al menos una sustitución, deleción y/o inserción de un aminoácido en comparación con la secuencia de aminoácidos codificada por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en el plásmido presente en el microorganismo depositado CGMCC nº. 0581.

[0046] Una secuencia de nucleótidos seleccionada de SEC ID nº 3 o una subsecuencia de la misma, al igual que una secuencia de aminoácidos seleccionada de SEC ID nº 4 o un fragmento de la misma se puede utilizar para diseñar una sonda polinucleótida para identificar y clonar polipéptidos de codificación de ADN con actividad de celobiohidrolasa I de cepas de diferentes géneros o especies, según métodos bien conocidos en la técnica. En particular, tales sondas se pueden usar para hibridación con el genómico o ADNc del género o especie de interés, siguiendo procedimientos estándar de Southern blotting, para identificar y aislar el gen correspondiente en este. Tales sondas puede ser considerablemente más cortas que la secuencia entera, pero deberían ser al menos 15, preferiblemente al menos 25, más preferiblemente al menos 35 nucleótidos de longitud, tal como al menos 70 nucleótidos de longitud. Es, no obstante, preferido que la sonda polinucleótida sea de al menos 100 nucleótidos de longitud. Por ejemplo, la sonda polinucleótida puede ser de al menos 200 nucleótidos de longitud, al menos 300 nucleótidos de longitud, al menos 400 nucleótidos de longitud o al menos 500 nucleótidos de longitud. Incluso sondas más largas pueden ser utilizadas, por ejemplo, sondas polinucleótidas que son de al menos 600 nucleótidos de longitud, al menos 700 nucleótidos de longitud, al menos 800 nucleótidos de longitud, o al menos 900 nucleótidos de longitud. Ambas sondas de ADN y ARN pueden ser usadas. Las sondas son típicamente etiquetadas para detectar del gen correspondiente (por ejemplo, con ³²P, ³H, ³⁵S, biotina o avidina).

[0047] Así, un ADN genómico o genoteca de ADNc obtenida a partir de tales otros organismos se pueden seleccionar para ADN que hibridiza con las sondas anteriormente descritas y que codifica un polipéptido con actividad de celobiohidrolasa I. ADN genómico u otro de tales otros organismos se puede separar por agarosa o electroforesis en gel de poliacrilamida u otras técnicas de separación. ADN de las genotecas o ADN separado puede ser transferido a e

inmovilizado en nitrocelulosa u otros materiales portadores adecuados. Para identificar un clon o ADN que es homólogo el material portador con el ADN inmovilizado es usado en un *Southern blot*.

[0048] Las moléculas a las que la sonda polinucleótida hibrida bajo estas condiciones pueden ser detectadas usando una película radiográfica o por cualquier otro método conocido en la técnica. Siempre que el término "sonda polinucleótida" se use en el presente contexto, debe entenderse que tal sonda contiene al menos 15 nucleótidos.

[0049] En una forma de realización interesante, la sonda polinucleótida es la cadena complementaria de los nucleótidos seleccionados del grupo que consiste en:

nucleótidos 1 a 1587 de SEC ID nº 3,

nucleótidos 1 a 1302 de SEC ID nº 3;

o la cadena complementaria de los nucleótidos seleccionados del grupo que consiste en:

nucleótidos 1 a 500 de SEC ID nº 3;

o la cadena complementaria de los nucleótidos seleccionados de:

nucleótidos 1 a 200 de SEC ID nº 3.

[0050] En otra forma de realización interesante, la sonda polinucleótida es la cadena complementaria de la secuencia de nucleótido que codifica un polipéptido de SEC ID nº 4. En una forma de realización aún más interesante, la sonda polinucleótida es la cadena complementaria del nucleótido seleccionado de SEC ID nº 3. En otra forma de realización interesante, la sonda polinucleótida es la cadena complementaria de la secuencia de nucleótido contenida en un plásmido que se contiene en el microorganismo depositado CGMCC nº. 0581.

[0051] Para sondas largas de al menos 100 nucleótidos de longitud, condiciones de astringencia de muy baja a muy alta son definidas como prehibridación e hibridación a 42°C en 5X SSPE, 1,0% SDS, 5X solución de Denhardt, 100 µg/ml ADN de esperma de salmón desnaturalizado y cortado, siguiendo procedimientos estándar de Southern blotting. Preferiblemente, las sondas largas de al menos 100 nucleótidos no contienen más de 1.000 nucleótidos. Para sondas largas de al menos 100 nucleótidos de longitud, el material portador es finalmente lavado tres veces cada uno durante 15 minutos usando 2 x SSC, 0,1% SDS a 42°C (astringencia muy baja), preferiblemente lavado tres veces cada uno durante 15 minutos usando 0,5 x SSC, 0,1% SDS a 42°C (astringencia baja), más preferiblemente lavado tres veces cada uno durante 15 minutos usando 0,2 x SSC, 0,1% SDS a 42°C (astringencia media), incluso más preferiblemente lavado tres veces cada uno durante 15 minutos usando 0,2 x SSC, 0,1% SDS a 55°C (astringencia media alta), de la forma más preferible lavado tres veces cada uno durante 15 minutos usando 0,1 x SSC, 0,1% SDS a 60°C (astringencia alta), en particular lavado tres veces cada uno durante 15 minutos usando 0,1 x SSC, 0,1% SDS a 68°C (astringencia muy alta).

[0052] Aunque no particularmente preferido, se contempla que sondas más cortas, por ejemplo, sondas que son de aproximadamente 15 a 99 nucleótidos de longitud, tal como de aproximadamente 15 a aproximadamente 70 nucleótidos de longitud, pueden también ser usadas. Para este tipo de sondas cortas, condiciones de astringencia son definidas como prehibridación, hibridación y post-hibridación de lavado de 5°C a 10°C por debajo de la T_m calculada usando el cálculo según Bolton and McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) en 0,9 M NaCl, 0,09 M Tris-HCl pH 7,6, 6 mM EDTA, 0,5% NP-40, 1X solución de Denhardt, 1 mM pirofosfato de sodio, 1 mM fosfato monobásico de sodio, 0,1 mM ATP y 0,2 mg de ARN de levadura por ml siguiendo los procedimientos de *Southern blot* estándar.

[0053] Para sondas cortas que son de aproximadamente 15 nucleótidos a 99 nucleótidos de longitud, el material portador se lava una vez en 6X SCC más 0,1% SDS durante 15 minutos y dos veces cada uno durante 15 minutos usando 6X SSC a 5°C a 10°C por debajo del calculado T_m.

Fuentes para polipéptidos con actividad de celobiohidrolasa I

[0054] Un polipéptido de la presente invención se puede obtener de microorganismos de cualquier género. Para los propósitos de la presente invención, el término "obtenido de" como se usa aquí significará que el polipéptido codificado por la secuencia de nucleótidos es producido por una célula en la que está presente naturalmente la secuencia de nucleótidos o en la que ha sido insertada la secuencia de nucleótidos. En una forma de realización preferida, el polipéptido es secretado naturalmente.

[0055] Un polipéptido de la presente invención puede ser un polipéptido bacteriano. Por ejemplo, el polipéptido puede ser un polipéptido bacteriano, tal como un polipéptido *Bacillus*, por ejemplo, un polipéptido *Bacillus*, por ejemplo, un *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, o polipéptido *Bacillus thuringiensis*; o un polipéptido *Streptomyces*, por ejemplo un polipéptido *Streptomyces lividans* o polipéptido *Streptomyces murinus*; o un

[0056] Un polipéptido de la presente invención puede ser un polipéptido fúngico y más preferiblemente un polipéptido de levadura, tal como un polipéptido de *Candida*, *Kluyveromyces*, *Neocallimastix*, *Pichia*, *Piromyces*, *Saccharomyces*, *Schizosaccharomyces*, o *Yarrowia*; o más preferiblemente un polipéptido fúngico filamentoso, tan como un polipéptido

Acremonium, Aspergillus, Aureobasidium, Cryptococcus, Filibasidium, Fusarium, Humicola, Magnaporthe, Mucor, Myceliophthora, Neurospora, Paecilomyces, Penicillium, Schizophyllum, Talaromyces, Thermoascus, Thielavia, Tolypocladium, o Trichoderma.

5 [0057] En una forma de realización más preferida, el polipéptido es un polipéptido *Chaetomium thermophilum*, por ejemplo el polipéptido que consiste en la secuencia de aminoácidos SEC ID n° 4

10 [0058] Se entenderá que para las especies mencionadas anteriormente, la invención comprende tanto los estados perfectos como imperfectos, y otros equivalentes taxonómicos, por ejemplo, anamorfos, con independencia del nombre de especie por el que se les conoce. Los expertos en la técnica reconocerán fácilmente la identidad de equivalentes apropiados.

15 [0059] Cepas de estas especies son fácilmente accesibles para el público en varias colecciones de cultivo, tal como la American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), China General Microbiological Culture Collection Center (CGMCC), Centraalbureau Voor Schimmelcultures (CBS), y Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

20 [0060] Además, tales polipéptidos se pueden identificar y obtener de otras fuentes que incluyen microorganismos aislados de la naturaleza (p. ej., tierra, agua, plantas, animales, etc.) usando las sondas arriba mencionadas. Técnicas para aislar microorganismos de hábitats naturales son bien conocidas en la técnica. La secuencia de nucleótidos puede luego ser derivada mediante la selección de forma similar de un genómico o genoteca de ADNc de otro microorganismo. Una vez una secuencia de nucleótidos que codifica un polipéptido ha sido detectada con la sonda(s), la secuencia puede ser aislada o clonada utilizando de técnicas que se conocen por aquellos de habilidad común en la técnica (ver, por ejemplo, Sambrook et al., 1989, supra).

25 [0061] Polipéptidos codificados por secuencias de nucleótidos de la presente divulgación también incluyen polipéptidos fundidos o polipéptidos divisibles de fusión en los que otro polipéptido se funde al N-término o al C-término del polipéptido o fragmento del mismo. Un polipéptido fusionado se produce por fundición de una secuencia de nucleótidos (o una parte de la misma) que codifica otro polipéptido a una secuencia de nucleótidos (o una parte de la misma) de la presente invención. Técnicas para producir polipéptidos de fusión se conocen en la técnica e incluyen ligar las secuencias de codificación que codifican los polipéptidos de modo que éstas están en bastidor y que la expresión del polipéptido fusionado está bajo control del mismo promotor(es) y terminador.

35 Polinucleótidos y secuencias de nucleótidos

[0062] La presente invención también se refiere a polinucleótidos con una secuencia de nucleótidos que codifica un polipéptido de la invención. En particular, la presente invención se refiere a polinucleótidos que consisten en una secuencia de nucleótidos que codifica un polipéptido de la invención. En una forma de realización preferida, la secuencia de nucleótidos es SEC ID n.º: 3. En una forma de realización más preferida, la secuencia de nucleótidos es el polipéptido maduro que codifica la región contenida en un plásmido que está contenido en un microorganismo depositado seleccionado de CGMCC n.º 0581. La presente invención también abarca polinucleótidos que comprenden, preferiblemente que consisten en, secuencias de nucleótidos que codifican un polipéptido que consiste en una secuencia de aminoácidos seleccionada de SEC ID n.º: 4, que difiere de una secuencia de nucleótidos seleccionada de SEC ID n.º: 3 en virtud de la degeneración del código genético.

45 [0063] La presente invención también se refiere a polinucleótidos que comprenden, preferiblemente que consisten en, una subsecuencia de una secuencia de nucleótidos SEC ID n° 3 que codifica fragmentos de una secuencia de aminoácidos de SEC ID n° 4 que tienen actividad de celobiohidrolasa I. Una subsecuencia de una secuencia de nucleótidos de SEC ID n° 3 es una secuencia de nucleótidos encerrada por una secuencia de SEC ID n° 3 excepto que uno o más nucleótidos del final 5' y/o 3' han sido eliminados.

50 [0064] La presente invención también se refiere a polinucleótidos que tienen, preferiblemente que consisten en, una secuencia de nucleótidos modificada que comprende como mínimo una modificación en el polipéptido maduro que codifica la secuencia seleccionada que consiste en SEC ID n° 3, y donde la secuencia de nucleótido modificada codifica un polipéptido que consiste en una secuencia de aminoácido seleccionada del grupo que consiste en SEC ID n° 4.

60 [0065] Las técnicas usadas para aislar o clonar una secuencia de nucleótidos que codifican un polipéptido se conocen en la técnica e incluyen aislamiento de ADN genómico, preparación de ADNc o una combinación de las mismas. La clonación de las secuencias de nucleótidos de la presente invención de tal ADN genómico puede ser efectuada, por ejemplo, usando la bien conocida reacción en cadena de polimerasa (PCR) o selección de anticuerpos de genotecas de expresión para detectar fragmentos de ADN clonados con características estructurales compartidas. Ver, por ejemplo, Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Otros procedimientos de amplificación tales como reacción en cadena de ligasa (LCR), transcripción activada ligada (LAT) y amplificación basada en secuencias de nucleótidos (NASBA) puede ser utilizados. La secuencia de nucleótidos se puede clonar de una cepa seleccionada del grupo que consiste en *Acremonium, Scytalidium, Thermoascus, Thielavia, Verticillium, Neoterme*,
65 *Melanocarpus, Poitrasia, Coprinus, Trichotecium, Humicola, Cladorrhinum, Diplodia, Myceliophthora, Rhizomucor,*

Meripilus, *Exidia*, *Xylaria*, *Trichophaea*, *Chaetomium*, *Chaetomidium*, *Sporotrichum*, *Thielavia*, *Aspergillus*, *Scopulariopsis*, *Fusarium*, *Pseudoplectania* y *Phytophthora* u otros u organismo relacionado y así, por ejemplo, puede ser un alélico o variante de especies del polipéptido que codifica la región de la secuencia de nucleótidos.

5 [0066] La secuencia de nucleótidos se puede obtener por procedimientos de clonación estándar usados en ingeniería genética para recolocar la secuencia de nucleótidos de su ubicación natural a un sitio diferente donde ésta será reproducida. Los procedimientos de clonación pueden implicar escisión y aislamiento de un fragmento deseado que incluye la secuencia de nucleótidos que codifica el polipéptido, inserción del fragmento en una molécula de vector e incorporación del vector recombinante en una célula huésped donde copias múltiples o clones de la secuencia de
10 nucleótidos serán replicadas. La secuencia de nucleótidos puede ser de origen genómico, ADNc, ARN, semisintético sintético o cualquier combinación de las mismas.

[0067] La presente invención también se refiere a un polinucleótido que comprende, preferiblemente que consiste en, una secuencia de nucleótidos que tiene un grado de identidad con una secuencia de nucleótidos que consiste en los
15 nucleótidos 1 a 1587 de SEC ID nº 3, de al menos 95% de identidad, tal como al menos 96% de identidad, por ejemplo al menos 97% de identidad, incluso más preferiblemente al menos 98% de identidad, tal como al menos 99%. Preferiblemente, la secuencia de nucleótidos codifica un polipéptido con actividad de celobiohidrolasa I. El grado de identidad entre dos secuencias de nucleótidos se determina como se ha descrito previamente (ver la sección titulada "Definiciones").

20 [0068] En otro aspecto interesante la presente invención se refiere a un polinucleótido que tiene, preferiblemente consiste en una secuencia de nucleótidos que tiene al menos 95% de identidad con la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en el microorganismo depositado CGMCC nº. 0581. En una forma de realización preferida, el grado de identidad con la celobiohidrolasa I que codifica parte de la
25 secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado CGMCC nº. 0581 es al menos del 95%, tal como al menos del 96%, por ejemplo al menos del 97%, incluso más preferiblemente al menos del 98%, tal como al menos del 99%. Preferiblemente, la secuencia de nucleótidos comprende la celobiohidrolasa I que codifica la parte de la secuencia de nucleótidos insertada en un plásmido presente en el depositado CGMCC nº. 0581. En una forma de realización aún más preferida, la secuencia de nucleótidos consiste en la celobiohidrolasa I que
30 codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en el microorganismo depositado CGMCC nº. 0581.

[0069] La modificación de una secuencia de nucleótidos que codifica un polipéptido de la presente invención puede ser necesaria para la síntesis de un polipéptido, que comprende una secuencia de aminoácidos que tiene al menos una
35 sustitución, delección y/o inserción al compararla con una secuencia de aminoácidos de SEC ID nº 4. Estas variantes artificiales puede diferir en alguna forma creada genéticamente del polipéptido aislado de su fuente nativa, por ejemplo, variantes que difieren en la actividad específica, termoestabilidad, pH óptimo o similares.

[0070] Será evidente para los expertos en la técnica que tales modificaciones pueden ser hechas fuera de las regiones críticas a la función de la molécula y todavía suponer un polipéptido activo. Residuos de aminoácidos esenciales para la actividad del polipéptido codificado por la secuencia de nucleótidos de la invención, y por lo tanto preferiblemente no sujetos a modificación, tal como sustitución, se pueden identificar según procedimientos conocidos en la técnica, tal como mutagénesis dirigida o mutagénesis de escaneado de alanina (ver, por ejemplo, Cunningham and Wells, 1989, Science 244: 1081-1085). En la anterior técnica, mutaciones se introducen en cada residuo positivamente cargado de la
45 molécula, y las moléculas mutantes resultantes se evalúan para actividad de celobiohidrolasa I para identificar residuos de aminoácidos que son críticos para la actividad de la molécula. Sitios de interacción de enzima-sustrato pueden también ser determinados por análisis de la estructura tridimensional como determinado por tales técnicas como análisis de resonancia magnética nuclear, cristalografía o etiquetado de fotoafinidad (ver, por ejemplo, de Vos et al., 1992, Science 255: 306-312 ; Smith et al., 1992, Journal of Molecular Biology 224: 899-904; Wlodaver et al., 1992, FEBS Letters 309: 59-64).
50

[0071] Por otra parte, una secuencia de nucleótidos que codifica un polipéptido de la presente invención se puede modificar por introducción de sustituciones de nucleótidos que no dan lugar a otra secuencia de aminoácidos del polipéptido codificado por la secuencia de nucleótidos, pero que corresponden al uso de codón del organismo huésped destinado a la producción de la enzima.
55

[0072] La introducción de una mutación en la secuencia de nucleótidos para intercambiar un nucleótido por otro nucleótido se puede realizar por mutagénesis dirigida usando cualquiera de los métodos conocidos en la técnica. Particularmente útil es el procedimiento que utiliza un vector de ADN bicatenario superenrollado con un inserto de interés y dos cebadores sintéticos que contienen la mutación deseada. Los cebadores oligonucleótidos, cada uno complementario de las cepas opuestas del vector, se extienden durante la variación cíclica de la temperatura mediante *Pfu* ADN polimerasa. En la incorporación de los cebadores, un plásmido mutado con cortes en bisel es generado. Después del ciclo de temperatura, el producto se trata con *DpnI* que es específico para ADN hemimetilado y metilado para digerir el molde de ADN progenitor y para seleccionar para ADN sintetizado con mutación. Otros procedimientos conocidos en la técnica pueden también ser usados. Para una descripción general de sustitución de nucleótidos, véase, por ejemplo, Ford et al., 1991, Protein Expression and Purification 2: 95-107.
60
65

[0073] Como será entendido, detalles y particulares en lo que se refiere a hibridación de las secuencias de nucleótidos serán iguales o análogos a los aspectos de hibridación discutidos en la sección titulada "Polipéptidos con actividad de celobiohidrolasa I" aquí expuesta.

Constructos de Ácidos Nucleicos

[0074] La presente invención también se refiere a constructos de ácidos nucleicos que comprenden una secuencia de nucleótidos de la presente invención operativamente enlazados a una o más secuencias de control que dirigen la expresión de la secuencia codificante en una célula huésped adecuada bajo condiciones compatibles con las secuencias de control.

[0075] Una secuencia de nucleótidos que codifica un polipéptido de la presente invención se puede manipular en una variedad de vías para proveer a expresión del polipéptido. La manipulación de la secuencia de nucleótidos antes de su inserción en un vector puede ser deseable o necesaria dependiendo del vector de expresión. Las técnicas para modificar secuencias de nucleótidos utilizando métodos de ADN recombinante se conocen bien en la técnica.

[0076] La secuencia de control puede ser una secuencia del promotor apropiada, una secuencia de nucleótidos que se reconoce por una célula huésped para la expresión de la secuencia de nucleótidos. La secuencia del promotor contiene secuencias de control transcripcionales, que median la expresión del polipéptido. El promotor puede ser cualquier secuencia de nucleótidos que muestre actividad transcripcional en la célula huésped de elección, incluidos promotores mutantes truncados e híbridos, y se pueden obtener de genes que codifican polipéptidos intracelulares o extracelulares ya sean heterólogos u homólogos de la célula huésped.

[0077] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de ácidos nucleicos de la presente invención, especialmente en una célula huésped bacteriana, son los promotores obtenidos del operón lac de *E. coli*, gen de agarasa de *Streptomyces coelicolor* (dagA), gen de levansucrasa de *Bacillus subtilis* (sacB), gen de alfa-amilasa de *Bacillus licheniformis* (amyL), gen de amilasa maltogénica de *Bacillus stearothermophilus* (amyM), gen de alfa-amilasa de *Bacillus amyloliquefaciens* (amyQ), gen de penicilinas de *Bacillus licheniformis* (penP), genes xylA y xylB de *Bacillus subtilis*, y gen de beta-lactamasa procariótica (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75: 3727-3731), así como el promotor tac (DeBoer et al., 1983, Proceedings of the National Academy of Sciences USA 80: 21-25). Más promotores son descritos en "Useful proteins from recombinant bacteria" en Scientific American, 1980, 242: 74-94; y en Sambrook et al., 1989, supra.

[0078] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de ácidos nucleicos de la presente invención en una célula huésped filamentosa fúngica son promotores obtenidos de los genes para TAKA amilasa de *Aspergillus oryzae*, proteinasa aspártica de *Rhizomucor miehei*, alfa-amilasa neutra de *Aspergillus niger*, alfa-amilasa estable en medio ácido de *Aspergillus niger*, glucoamilasa de *Aspergillus niger* o de *Aspergillus awamori* (glaA), lipasa de *Rhizomucor miehei*, proteasa alcalina de *Aspergillus oryzae*, triosa fosfato isomerasa de *Aspergillus oryzae*, acetamidasa de *Aspergillus nidulans* y proteasa de tipo tripsina de *Fusarium oxysporum* (WO 96/00787), al igual que el promotor NA2-tpi (un híbrido de los promotores de los genes para alfa-amilasa neutra de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae*) y promotores mutantes, truncados e híbridos de los mismos.

[0079] En un huésped de levadura, promotores útiles se obtienen de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactocinasa de *Saccharomyces cerevisiae* (GAL1), alcohol deshidrogenasa /gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH2/GAP) y 3-fosfoglicerato-cinasa de *Saccharomyces cerevisiae*. Otros promotores útiles para células huésped de levadura son descritos por Romanos et al., 1992, Yeast 8: 423-488.

[0080] La secuencia de control puede también ser una secuencia del terminador de transcripción adecuada, una secuencia reconocida por una célula huésped para terminar transcripción. La secuencia del terminador es operativamente enlazada al terminal 3' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier terminador que es funcional en la célula huésped de elección se puede utilizar en la presente invención.

[0081] Terminadores preferidos para células huésped fúngicas filamentosas se obtienen de los genes para TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, sintasa de antranilato de *Aspergillus nidulans*, alfa-glucosidasa de *Aspergillus niger* y proteasa de tipo tripsina de *Fusarium oxysporum*.

[0082] Terminadores preferidos para células huésped de levadura se obtienen de los genes para enolasa de *Saccharomyces cerevisiae*, citocromo C de *Saccharomyces cerevisiae* (CYC1), y de gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae*. Otros terminadores útiles para células huésped de levadura son descritas por Romanos et al., 1992, supra.

[0083] La secuencia de control puede también ser una secuencia líder adecuada, una región no traducida de un ARNm que es importante para traducción por la célula huésped. La secuencia líder está operativamente enlazada al terminal 5' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier secuencia líder que es funcional en la célula

huésped de elección se puede utilizar en la presente invención.

[0084] Los líderes preferidos para células huésped fúngicas filamentosas se obtienen de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.

[0085] Líderes adecuados para células huésped de levadura se obtienen de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), quinasa de 3-fosfoglicerato de *Saccharomyces cerevisiae*, alfa-factor de *Saccharomyces cerevisiae* y de alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH2/GAP).

[0086] La secuencia de control puede también ser una secuencia de poliadenilación, una secuencia operativamente enlazada al terminal 3' de la secuencia de nucleótidos y que, cuando se transcribe, es reconocida por la célula huésped como una señal para añadir residuos de poliadenosina al ARNm transcrito. Cualquier secuencia de poliadenilación que es funcional en la célula huésped de elección se puede utilizar en la presente invención.

[0087] Secuencias de poliadenilación preferida para células huésped filamentosas fúngicas se obtienen de los genes para TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, sintasa de antranilato de *Aspergillus nidulans*, proteasa de tipo tripsina de *Fusarium oxysporum*, y alfa-glucosidasa de *Aspergillus niger*.

[0088] Secuencias de poliadenilación útil para células huésped de levadura son descritas por Guo and Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.

[0089] La secuencia de control puede también ser una región de codificación de péptido de señal que codifica para una secuencia de aminoácidos enlazada al amino terminal de un polipéptido y que dirige el polipéptido codificado en la vía secretora de la célula. El extremo 5' de la secuencia codificante de la secuencia de nucleótidos puede intrínsecamente contener una región de codificación de péptido señal naturalmente enlazado en el marco de lectura de traducción con el segmento de la región de codificación que codifica el polipéptido segregado. Alternativamente, el extremo 5' de la secuencia codificante puede contener una región de codificación de péptido señal que es foránea a la secuencia codificante. La región de codificación de péptido señal foránea puede ser requerida donde la secuencia codificante no contenga naturalmente una región de codificación de péptido señal. Alternativamente, la región de codificación de péptido señal foráneo puede simplemente reemplazar la región de codificación de péptido señal natural para mejorar secreción del polipéptido. No obstante, cualquier región de codificación de péptido señal que dirige el polipéptido expresado en la vía secretora de una célula huésped de elección se puede utilizar en la presente invención.

[0090] Regiones de codificación de péptido señal eficaces para células huésped bacterianas son las regiones de codificación de péptido señal obtenidas de los genes para amilasa maltogénica de *Bacillus* NCIB 11837, alfa-amilasa de *Bacillus stearothermophilus*, subtilisina de *Bacillus licheniformis*, beta-lactamasa de *Bacillus licheniformis*, proteasas neutras de *Bacillus stearothermophilus* (*nprT*, *nprS*, *nprM*), y *Bacillus subtilis* *prxA*. Más péptidos señal son descritos por Simonen and Palva, 1993, Microbiological Reviews 57: 109-137 .

[0091] Regiones de codificación de péptido señal eficaces para células huésped filamentosas fúngicas son las regiones de codificación de péptido señal obtenidas de los genes para TAKA amilasa de *Aspergillus oryzae*, amilasa neutra de *Aspergillus niger*, glucoamilasa de *Aspergillus niger*, proteinasa aspártica de *Rhizomucor miehei*, celulasa de *Humicola insolens* y lipasa de *Humicola lanuginosa*.

[0092] Péptidos señal útiles para células huésped de levadura se obtienen de los genes para factor alfa de *Saccharomyces cerevisiae* e invertasa de *Saccharomyces cerevisiae*. Otras regiones de codificación de péptido señal útil son descritas por Romanos et al., 1992, supra.

[0093] La secuencia de control puede también ser una región de codificación de propéptido que codifica para una secuencia de aminoácidos situados en el terminal de amino de un polipéptido. El polipéptido resultante es conocido como una proenzima o propolipéptido (o un zimógeno en algunos casos). Un propolipéptido está generalmente inactivo y se puede convertir a un polipéptido maduro activo por escisión autocatalítica o catalítica del propéptido del propolipéptido. La región de codificación de propéptido se puede obtener de los genes para proteasa alcalina de *Bacillus subtilis* (*aprE*), proteasa neutra de *Bacillus subtilis* (*nprT*), alfa-factor de *Saccharomyces cerevisiae*, proteinasa aspártica de *Rhizomucor miehei*, y lacasa de *Myceliophthora thermophila* (WO 95/33836).

[0094] Donde ambos péptidos señal y regiones de propéptido están presentes en el terminal amino de un polipéptido, la región de propéptido está situada junto al terminal amino de un polipéptido y la región de péptido señal está situada junto al terminal amino de la región de propéptido.

[0095] Puede también ser deseable añadir secuencias reguladoras que permitan la regulación de la expresión del polipéptido relativa al crecimiento de la célula huésped. Ejemplos de sistemas reguladores son los que causan que la expresión del gen sea activada o desactivada en respuesta a un estímulo físico o químico, incluida la presencia de un compuesto regulador. Sistemas reguladores en sistemas procarióticos incluyen sistemas de operador *lac*, *tac* y *trp*. En la levadura, el sistema ADH2 o el sistema GAL1 pueden ser utilizados. En hongos filamentosos, el promotor de alfa-amilasa TAKA, promotor de glucoamilasa de *Aspergillus niger* y promotor de glucoamilasa de *Aspergillus oryzae* se

pueden utilizar como secuencias reguladoras. Otros ejemplos de secuencias reguladoras son los que permiten amplificación génica. En sistemas eucarióticos, estos incluyen el gen de dihidrofolato-reductasa que se amplifica en presencia de metotrexato y los genes de metalotioneína que se amplifican con metales pesados. En estos casos, la secuencia de nucleótidos que codifica el polipéptido sería operativamente enlazada con la secuencia reguladora.

5

Vectores de expresión

10

[0096] La presente invención también se refiere a vectores de expresión recombinantes que incluyen el constructo de ácidos nucleicos de la descripción. Las varias secuencias de nucleótidos y de control anteriormente descritas pueden ser juntadas para producir un vector de expresión recombinante que puede incluir uno o más sitios de restricción convenientes para permitir la inserción o sustitución de la secuencia de nucleótidos que codifica el polipéptido en tales sitios. Alternativamente, la secuencia de nucleótidos de la presente invención se puede expresar por inserción de la secuencia de nucleótidos o un constructo de ácidos nucleicos que comprende la secuencia en un vector apropiado para la expresión. Al crear el vector de expresión, la secuencia codificante se localiza en el vector de modo que la secuencia codificante es operativamente enlazada con las secuencias de control apropiadas para la expresión.

15

20

[0097] El vector de expresión recombinante puede ser cualquier vector (p. ej., un plásmido o virus) que puede ser convenientemente sometido a procedimientos de ADN recombinante y puede provocar la expresión de la secuencia de nucleótidos. La elección del vector típicamente dependerá de la compatibilidad del vector con la célula huésped en la que el vector va a ser introducido. Los vectores puede ser plásmidos lineales o circulares cerrados.

25

[0098] El vector puede ser un vector de replicación autónoma, es decir, un vector que existe como una entidad extracromosómica, la replicación del cual es independiente de replicación cromosómica, por ejemplo, un plásmido, un elemento extracromosómico, un minicromosoma o un cromosoma artificial.

30

[0099] El vector puede contener cualquier medio para asegurar autorreplicación. Alternativamente, el vector puede ser uno que, cuando es introducido en la célula huésped, se integra en el genoma y es replicado con el cromosoma(s) en que ha sido integrado. Además, un único vector o plásmido o dos o más vectores o plásmidos que juntos contienen el ADN total a ser introducido en el genoma de la célula huésped o un transposón pueden ser utilizados.

35

[0100] Los vectores de la presente invención preferiblemente contienen uno o más marcadores seleccionables que permitir selección fácil de células transformadas. Un marcador seleccionable es un gen cuyo producto proporciona biocida o resistencia vírica, resistencia a metales pesados, prototrofia para auxótrofos y similares.

40

[0101] Ejemplos de marcadores seleccionables bacterianos son los genes *dal* de *Bacillus subtilis* o *Bacillus licheniformis*, o marcadores que confieren resistencia antibiótica tal como resistencia a ampicilina, canamicina, cloranfenicol o resistencia. Marcadores adecuados para células huésped de levadura son ADE2, HIS3, LEU2, LYS2, MET3, TRP1 y URA3. Marcadores seleccionables para uso en una célula huésped filamentosa fúngica incluyen, pero de forma no limitativa, *amdS* (acetamidasa), *argB* (ornitina-carbamoyltransferasa), *bar* (fosfonitrilina acetiltransferasa), *hygB* (higromicina fosfotransferasa), *niaD* (nitrato-reductasa), *pyrG* (orotidina-5'-fosfato-descarboxilasa), *sC* (adeniltransferasa de sulfato), *trpC* (sintasa de antranilato), al igual que equivalentes de los mismos.

45

[0102] Preferidos para uso en una célula de *Aspergillus* son los genes de *amdS* y *pyrG* de *Aspergillus nidulans* o *Aspergillus oryzae* y el gen *bar* de *Streptomyces hygroscopicus*.

50

[0103] Los vectores de la presente invención preferiblemente contienen un elemento(s) que permite integración estable del vector en el genoma de la célula huésped o replicación autónoma del vector en la célula independiente del genoma.

55

[0104] Para integración en el genoma de la célula huésped, el vector puede contar con la secuencia de nucleótidos que codifica el polipéptido o cualquier otro elemento del vector para integración estable del vector en el genoma por recombinación homóloga o no homóloga. Alternativamente, el vector puede contener secuencias de nucleótidos adicionales para dirigir integración por recombinación homóloga en el genoma de la célula huésped. Las secuencias de nucleótidos adicionales permiten que el vector sea integrado en el genoma de la célula huésped en una ubicación(es) precisa del cromosoma(s). Para aumentar la probabilidad de integración en una ubicación precisa, los elementos integracionales deberían preferiblemente contener un número suficiente de nucleótidos, tal como 100 a 1.500 pares de bases, preferiblemente 400 a 1.500 pares de bases y, de la forma más preferible 800 a 1.500 pares de bases, que son altamente homólogos de la secuencia objetivo correspondiente para mejorar la probabilidad de recombinación homóloga. Los elementos integracionales pueden ser cualquier secuencia que sea homóloga de la secuencia objetivo del genoma de la célula huésped. Además, los elementos integracionales pueden ser secuencias de nucleótidos codificantes o no codificantes. Por otro lado, el vector se puede integrar en el genoma de la célula huésped por recombinación no-homóloga.

60

65

[0105] Para replicación autónoma, el vector puede comprender además un origen de replicación permitiendo al vector replicar de manera autónoma en la célula huésped en cuestión. Ejemplos de orígenes bacterianos de replicación son los orígenes de replicación de plásmidos pBR322; pUC19; pACYC177 y pACYC184 que permiten replicación en *E. coli*, y pUB110; pE194; pTA1060, y pAM β 1 que permiten replicación en *Bacillus*. Ejemplos de orígenes de replicación para uso

en una célula huésped de levadura son el origen de 2 micras de replicación, ARS1, ARS4, la combinación de ARS1 y CEN3 y la combinación de ARS4 y CEN6. El origen de replicación puede ser uno que tiene una mutación que hace sensible su temperatura de funcionamiento en la célula huésped (ver, por ejemplo, Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75: 1433).

[0106] Más de una copia de una secuencia de nucleótidos de la presente invención se puede insertar en la célula huésped para aumentar producción del producto genético. Un aumento en el número de copias de la secuencia de nucleótidos se puede obtener por integración de al menos una copia adicional de la secuencia en el genoma de la célula huésped o incluyendo un gen marcador seleccionable amplificable con la secuencia de nucleótidos donde células que contienen copias amplificadas del gen marcador seleccionable, y así copias adicionales de la secuencia de nucleótidos se puede seleccionar para por cultivo de las células en presencia del agente apropiado seleccionable.

[0107] Los procedimientos usados para enlazar los elementos anteriormente descritos para construir los vectores de expresión recombinantes de la presente invención son bien conocidos por un experto en la técnica (ver, por ejemplo, Sambrook et al., 1989, supra).

Células huésped

[0108] La presente invención también se refiere a recombinar una célula huésped que incluye el constructo de ácidos nucleicos de la descripción, que es ventajosamente usado en la producción recombinante de los polipéptidos. Un vector que comprende una secuencia de nucleótidos de la presente invención se introduce en una célula huésped de modo que el vector es mantenido como un integrante cromosómico o como un vector extracromosómico que se duplica como se describe anteriormente.

[0109] La célula huésped puede ser un microorganismo unicelular, por ejemplo, un microorganismo procariota o uno no unicelular, por ejemplo, un eucariota.

[0110] Células útiles unicelulares son células bacterianas tales como bacterias gram positivas incluidas, pero no limitadas a, una célula de *Bacillus*, por ejemplo, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, y *Bacillus thuringiensis*; o una célula de *Streptomyces*, por ejemplo, *Streptomyces lividans* o *Streptomyces murinus* o bacterias gram negativas tales como *E. coli* y *Pseudomonas sp.* En una forma de realización preferida, la célula huésped bacteriana es una célula de *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus*, o *Bacillus subtilis*. En otra forma de realización preferida, la célula de *Bacillus* es un *Bacillus* alcalofílico.

[0111] La introducción de un vector en una célula huésped bacteriana puede, por ejemplo, ser efectuada por transformación de protoplasto (ver, por ejemplo, Chang and Cohen, 1979, Molecular General Genetics 168: 111-115), usando células competentes (ver, por ejemplo, Young and Spizizin, 1961, Journal of Bacteriology 81: 823-829, o Dubnau and Davidoff-Abelson, 1971, Journal of Molecular Biology 56: 209-221), electroporación (ver, por ejemplo, Shigekawa and Dower, 1988, Biotechniques 6: 742-751), o conjugación (ver, por ejemplo, Koehler and Thorne, 1987, Journal of Bacteriology 169: 5771-5278).

[0112] La célula huésped puede ser un eucariota, tal como un mamífero, insecto, planta o célula fúngica.

[0113] En una forma de realización preferida, la célula huésped es una célula fúngica. "Fúngica", como se utiliza en este caso, incluye filo *Ascomycota*, *Basidiomycota*, *Chytridiomycota* y *Zygomycota* (como definido por Hawksworth et al., In, Ainsworth and Bisby's Dictionary of The Fungi, 8ª edición, 1995, CAB International, University Press, Cambridge, UK) así como la *Oomycota* (como citado en Hawksworth et al., 1995, supra, página 171) y todos los hongos mitospóricos (Hawksworth et al., 1995, supra).

[0114] En una forma de realización más preferida, la célula huésped fúngica es una célula de levadura. "Yeast" como se utiliza en este caso incluye levadura ascoesporógena (Endomicetales), levadura basidioesporogénea, y levadura de los hongos imperfectos (Blastomicetos). Dado que la clasificación de levadura puede cambiar en el futuro, para los fines de esta invención, la levadura debe ser definida como se describe en Biology and Activities of Yeast (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

[0115] En una forma de realización aún más preferida, la célula huésped de levadura es una célula de *Candida*, *Aschbyii*, *Hansenula*, *Kluyveromices*, *Pichia*, *Saccharomyces*, *Schizosaccharomices* o *Yarrowia*.

[0116] En la forma de realización más preferida, la célula huésped de levadura es una célula de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* o *Saccharomyces oviformis*. En otra forma de realización más preferida, la célula huésped de levadura es una célula de *Kluyveromyces lactis*. En otra forma de realización más preferida, la célula huésped de levadura es una célula de *Yarrowia lipolytica*.

- [0117] En otra forma de realización más preferida, la célula huésped fúngica es una célula micótica filamentosa. "Hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión *Eumycota* y *Oomycota* (como definido por Hawksworth et al., 1995, supra). Los hongos filamentosos se caracterizan por una pared micelial compuesta por quitina, celulosa, glucano, quitosano, manano y otros polisacáridos complejos. Crecimiento vegetativo es por alargamiento hifal y catabolismo de carbono es estrictamente aeróbico. En cambio, crecimiento vegetativo por levaduras tal como *Saccharomyces cerevisiae* es por injerto de un talo unicelular y catabolismo de carbono puede ser fermentativo.
- [0118] En una forma de realización aún más preferida, la célula huésped filamentosa fúngica es una célula de una especie de, pero no limitado a, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolipocladium* o *Trichoderma*.
- [0119] En la forma de realización más preferida, la célula huésped filamentosa fúngica es una célula de *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*. En otra forma de realización más preferida, la célula huésped filamentosa fúngica es una célula de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochromum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, o *Fusarium venenatum*. En una forma de realización incluso más preferida, la célula madre filamentosa fúngica es una célula de *Fusarium venenatum* (Nirenberg sp. nov.). En otra forma de realización más preferida, la célula huésped filamentosa fúngica es una célula de *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.
- [0120] Células fúngicas se pueden transformar por un proceso implicando formación de protoplasto, transformación de los protoplastos y regeneración de la pared celular de una manera conocida per se. Procedimientos adecuados para transformación de células huésped de *Aspergillus* son descritos en EP 238 023 y Yelton et al., 1984, Proceedings of the National Academy of Sciences USA 81: 1470-1474. Métodos adecuados para transformar especies de *Fusarium* son descritas por Malardier et al., 1989, Gene 78: 147-156 and WO 96/00787. Levadura puede ser transformada usando los procedimientos descritos por Becker and Guarente, In Abelson, J.N. and Simon, M.I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito et al., 1983, Journal of Bacteriology 153: 163; and Hinnen et al., 1978, Proceedings of the National Academy of Sciences USA 75: 1920.
- Métodos de producción
- [0121] La presente divulgación también se refiere a métodos para producir un polipéptido de la presente invención que incluye (a) cultivo de una cepa, que en su forma de tipo salvaje es capaz de producir el polipéptido; y (b) recuperación del polipéptido. Preferiblemente, la cepa es seleccionada del grupo que consiste en infestantes de *Acremonium*, *Scytalidium*, *Thermoascus*, *Thielavia*, *Verticillium*, *Neotermes*, *Melanocarpus*, *Poitrasia*, *Coprinus*, *Trichothecium*, *Humicola*, *Cladorrhinum*, *Diplodia*, *Myceliophthora*, *Rhizomucor*, *Meripilus*, *Exidia*, *Xylaria*, *Trichophaea*, *Chaetomium*, *Chaetomidium*, *Sporotrichum*, *Thielavia*, *Aspergillus*, *Scopulariopsis*, *Fusarium*, *Pseudoplectania* y *Phytophthora*; más preferiblemente la cepa es seleccionada del grupo que consiste en *Acremonium thermophilum*, *Chaetomium thermophilum*, *Scytalidium thermophilum*, *Thermoascus aurantiacus*, *Thielavia australiensis*, *Verticillium tenerum*, *Neotermes castaneus*, *Melanocarpus albomyces*, *Poitrasia circinans*, *Coprinus cinereus*, *Trichothecium roseum*, *Humicola nigrescens*, *Cladorrhinum foecundissimum*, *Diplodia gossypina*, *Myceliophthora thermophila*, *Rhizomucor pusillus*, *Meripilus giganteus*, *Exidia glandulosa*, *Xylaria hypoxylon*, *Trichophaea saccata*, *Chaetomidium pingtungium*, *Myceliophthora thermophila*, *Myceliophthora hinnulea*, *Sporotrichum pruinosum*, *Thielavia* cf. *microspora*, *Pseudoplectania nigrella* y *Phytophthora*.
- [0122] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que incluye (a) cultivo de una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.
- [0123] La presente invención también se refiere a métodos para producción in situ de un polipéptido de la presente invención que incluye (a) cultivo de una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) contacto del polipéptido con un sustrato deseado, tal como un sustrato celulósico, sin recuperación previa del polipéptido. El término "producción in situ" significa que el polipéptido se produce directamente en la localización en la que está destinado a ser usado, tal como en un proceso de fermentación para la producción de etanol.
- [0124] En los métodos de producción de la presente invención, las células se cultivan en un medio nutritivo adecuado para la producción de los polipéptidos usando métodos conocidos en la técnica. Por ejemplo, la célula se puede cultivar por cultivo en matraz vibrante, fermentación a pequeña o gran escala (incluida fermentación continua, de lote, de lote alimentado o de estado sólido) en el laboratorio o fermentadores industriales realizados en un medio adecuado y bajo condiciones que permiten al polipéptido ser expresado y/o aislado. El cultivo se desarrolla en un medio nutritivo adecuado que incluye fuentes de nitrógeno y carbono y sales inorgánicas, utilizando procedimientos conocidos en la

técnica. Medios adecuados están disponibles de proveedores comerciales o se pueden preparar según las composiciones publicadas (p. ej., en catálogos de la American Type Culture Collection). Si el polipéptido se segrega en el medio nutritivo, el polipéptido se puede recuperar directamente del medio. Si el polipéptido no está segregado, se puede recuperar de lisatos celulares.

[0125] Los polipéptidos pueden ser detectados usando métodos conocidos en la técnica que son específicos para los polipéptidos. Estos métodos de detección pueden incluir uso de anticuerpos específicos, formación de un producto enzimático o desaparición de un sustrato enzimático. Por ejemplo, un ensayo enzimático se puede utilizar para determinar la actividad del polipéptido como se describe en este caso.

[0126] El polipéptido resultante se puede recuperar por métodos conocidos en la técnica. Por ejemplo, el polipéptido puede ser recuperado del medio nutritivo por procedimientos convencionales, que incluyen, pero no están limitados a, centrifugado, filtración, extracción, secado por pulverización, evaporación o precipitación.

[0127] Los polipéptidos de la presente invención se pueden purificar por una variedad de procedimientos conocidos en la técnica, incluidos, pero no limitados a, cromatografía (p. ej., de intercambio iónico, de afinidad, hidrofóbica, de cromatografía de exclusión de tamaño), procedimientos electroforéticos (p. ej., isoelectroenfoque preparativo), solubilidad diferencial (p. ej., precipitación de sulfato amónico), SDS-PAGE o extracción (ver, por ejemplo, Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Plantas

[0128] La presente divulgación también se refiere a una planta transgénica, parte de planta o célula vegetal que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido con actividad de celobiohidrolasa I de la presente invención para expresar y producir el polipéptido en cantidades recuperables. El polipéptido se puede recuperar de la planta o parte de planta. Alternativamente, la planta o parte de planta con el polipéptido recombinante se puede utilizar como tal para mejorar la calidad de un alimento o pienso, por ejemplo, mejorar el valor nutritivo, palatabilidad y propiedades reológicas, o para destruir un factor antinutritivo.

[0129] La planta transgénica puede ser dicotiledónea (un dicto) o monocotiledónea (un monocot). Ejemplos de plantas monocotiledóneas son hierbas, tales como césped de pradera (*Poa pratensis*, *Poa*), hierba forrajera tal como *Festuca*, *Lolium*, césped templado, tal como *Agrostis* y cereales, por ejemplo, trigo, avena, centeno, cebada, arroz, sorgo, mijo y maíz.

[0130] Ejemplos de plantas dicot son tabaco, altramuces, patata, remolacha azucarera, leguminosas, tales como guisante, judía y semilla de soja y plantas de crucífero (de familia *Brassicaceae*) tales como coliflor, colza, canola y el organismo de modelo cercanamente relacionado *Arabidopsis thaliana*.

[0131] Ejemplos de partes de planta son tallos, callos, hojas, raíces, frutas, semillas, y tubérculos. También tejidos de planta específicos, tales como cloroplasto, apoplasto, mitocondria, vacuola, peroxisomas y citoplasma se consideran una parte de planta. Además, cualquier célula vegetal, sea cual sea el origen del tejido, se considera una parte de planta.

[0132] También están incluidas dentro del campo de la presente invención la progenie (clonal o semilla) de tales plantas, partes de planta y células vegetales.

[0133] La planta transgénica o célula vegetal que expresa un polipéptido de la presente invención se puede construir conforme a métodos conocidos en la técnica. Brevemente, la planta o célula vegetal se construye por incorporación de uno o más constructos de expresión que codifican un polipéptido de la presente invención en el genoma de huésped de planta y propagando la resultante planta modificada o célula vegetal en una planta transgénica o célula vegetal.

[0134] Convenientemente, el constructo de expresión es un constructo de ácidos nucleicos que comprende una secuencia de nucleótidos que codifica un polipéptido de la presente invención operativamente enlazado con secuencias reguladoras apropiadas requerido para la expresión de la secuencia de nucleótidos en la planta o parte de planta elegida. Además, el constructo de expresión puede incluir una etiqueta seleccionable útil para identificar células huésped en las que el constructo de expresión ha sido integrado y secuencias de ADN necesarias para introducción del constructo en la planta en cuestión (esto depende del método de introducción de ADN que va a ser usado).

[0135] La elección de secuencias reguladoras, tales como secuencias promotoras y terminadoras y, opcionalmente, secuencias de señal o de tránsito es determinada, por ejemplo, basándose en cuándo, dónde y cómo se desea que sea expresado el polipéptido. Por ejemplo, la expresión del gen que codifica un polipéptido de la presente invención puede ser constitutiva o inducible, o puede ser desarrollable, fase o tejido específica, y el producto genético puede ser previsto para un tejido específico o parte de planta, tales como semillas u hojas. Secuencias reguladoras están, por ejemplo, descritas por Tague et al., 1988, Plant Physiology 86: 506.

[0136] Para expresión constitutiva, se puede usar el promotor 35S-CaMV (Franck et al., 1980, Cell 21: 285-294).

- Promotores específicos de un órgano pueden ser, por ejemplo, un promotor de tejidos de depósito de almacenamiento, tal como semillas, tubérculos de patata y frutas (Edwards & Coruzzi, 1990, Ann. Rev. Genet. 24: 275-303), o de tejidos sumidero metabólicos tales como meristemas (Ito et al., 1994, Plant Mol. Biol. 24: 863-878), un promotor específico de semilla tal como la glutelina, prolamina, globulina o promotor de albúmina de arroz (Wu et al., 1998, Plant and Cell Physiology 39: 885-889), un promotor de *Vicia faba* de la legumina B4 y el gen de proteína de semilla desconocida de *Vicia faba* (Conrad et al., 1998, Journal of Plant Physiology 152: 708-711), un promotor de una proteína del cuerpo oleoso de la semilla (Chen et al., 1998, Plant and Cell Physiology 39: 935-941), el promotor de la proteína de almacenamiento napA de *Brassica napus* o cualquier otro promotor específico de semilla conocido en la técnica, por ejemplo, como se describe en WO 91/14772. Además, el promotor puede ser un promotor específico de hoja tal como el promotor *rbcS* de arroz o tomate (Kyoizuka et al., 1993, Plant Physiology 102: 991-1000), el promotor genético de metiltransferasa de adenina del virus de *chlorella* (Mittra and Higgins, 1994, Plant Molecular Biology 26: 85-93) o el promotor genético *aldP* de arroz (Kagaya et al., 1995, Molecular and General Genetics 248: 668-674) o un promotor inducible por lesión tal como el promotor de patata *pin2* (Xu et al., 1993, Plant Molecular Biology 22: 573-588).
- [0137] Un elemento promotor intensificador puede también ser usado para conseguir expresión más alta de la enzima en la planta. Por ejemplo, el elemento promotor intensificador puede ser un intrón que es colocado entre el promotor y la secuencia de nucleótidos que codifica un polipéptido de la presente invención. Por ejemplo, Xu et al., 1993, supra revela el uso del primer intrón del gen de actina 1 de arroz para mejorar expresión.
- [0138] El gen marcador seleccionable y cualquier otra parte del constructo de expresión se puede elegir de aquellas disponibles en la técnica.
- [0139] El constructo de ácidos nucleicos se incorpora en el genoma de la planta según técnicas convencionales conocidas en la técnica, incluyendo transformación mediada por *Agrobacterium*, transformación mediada por virus, microinyección, bombardeo de partícula, transformación biolística y electroporación (Gasser et al., 1990, Science 244: 1293; Potrykus, 1990, Bio/Technology 8: 535; Shimamoto et al., 1989, Nature 338: 274).
- [0140] Actualmente, transferencia genética por medio de *Agrobacterium tumefaciens* es el método de elección para generar dicotiledóneas transgénicas (para una revisión, véase Hooykas and Schilperoort, 1992, Plant Molecular Biology 19: 15-38). No obstante esto puede también ser usado para monocotiledóneas de transformación, aunque otros métodos de transformación son generalmente preferidos para estas plantas. Actualmente, el método de elección para generar monocotiledóneas transgénicas es bombardeo de partícula (oro microscópico o partículas de tungsteno revestidas con el ADN transformante) de callos embrionarios o embriones en desarrollo (Christou, 1992, Plant Journal 2: 275-281; Shimamoto, 1994, Current Opinion Biotechnology 5: 158-162; Vasil et al., 1992, Bio/Technology 10: 667-674). Un método alternativo para transformación de monocotiledóneas se basa en transformación de protoplasto como se describe por Omirulleh et al., 1993, Plant Molecular Biology 21: 415-428.
- [0141] Tras la transformación, los transformantes que tienen el constructo de expresión incorporado se seleccionan y son regenerados en plantas enteras según métodos bien conocidos en la técnica.
- [0142] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que consisten en (a) cultivar una planta transgénica o una célula vegetal que comprende una secuencia de nucleótidos que codifican un polipéptido con actividad de celobiohidrolasa I de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido.
- [0143] La presente invención también se refiere a métodos para producción in situ de un polipéptido de la presente invención que consisten en (a) cultivar una planta transgénica o una célula vegetal que incluye una secuencia de nucleótidos que codifica un polipéptido con actividad de celobiohidrolasa I de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) contactar el polipéptido con un sustrato deseado, tal como un sustrato celulósico, sin recuperación previa del polipéptido.
- Composiciones
- [0144] En otro aspecto ulterior, la presente invención se refiere a composiciones que comprenden un polipéptido de la presente invención.
- [0145] La composición puede comprender un polipéptido de la invención como el componente enzimático mayor, por ejemplo, una composición monocomponente. Alternativamente, la composición puede comprender múltiples actividades enzimáticas, tales como aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, glicosiltransferasa de ciclodextrina, desoxiribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, glucosidasa alfa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolítica, peptidoglutaminasa, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa o xilanasas.
- [0146] Las composiciones se pueden preparar conforme a métodos conocidos en la técnica y pueden ser en forma de un líquido o una composición seca. Por ejemplo, la composición del polipéptido puede ser en forma de un granulado o

un microgranulado. El polipéptido a ser incluido en la composición se puede estabilizar conforme a métodos conocidos en la técnica.

- 5 [0147] Más abajo se dan ejemplos de usos preferidos de las composiciones de polipéptidos de la invención. La dosificación de la composición de polipéptidos de la invención y otras condiciones bajo las cuales la composición es usada pueden ser determinadas basándose en métodos conocidos en la técnica.

Composiciones detergentes

- 10 [0148] El polipéptido de la invención se puede adicionar a y así convertirse en un componente de una composición detergente.

- 15 [0149] La composición detergente de la invención puede por ejemplo ser formulada como una composición de detergente de lavado de ropa a máquina o a mano que incluye una composición de aditivo de lavandería adecuada para pretratamiento de tejidos manchados y una composición de suavizante adicionado para el enjuague, o ser formulado como una composición detergente para usar en operaciones de limpieza de superficies duras del hogar generales, o ser formulado para operaciones de lavado de la vajilla a máquina o a mano.

- 20 [0150] En un aspecto específico, la invención proporciona un aditivo detergente que incluye el polipéptido de la invención. El aditivo detergente al igual que la composición detergente puede comprender una o más enzimas, tales como una proteasa, una lipasa, una cutinasa, una amilasa, una carbohidrasa, una celulasa, una pectinasa, una mananasa, una arabinasa, una galactanasa, una xilanas, una oxidasas, por ejemplo, una lacasa y/o una peroxidasa.

- 25 [0151] En general las propiedades de la enzima(s) elegida deberían ser compatibles con el detergente seleccionado, (es decir, pH óptimo, compatibilidad con otros ingredientes no enzimáticos y enzimáticos, etc.), y la enzima(s) debería estar presente en cantidades eficaces.

- 30 Proteasas: proteasas adecuadas incluyen aquellas de origen animal, vegetal o microbiano. El origen microbiano es preferido. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. La proteasa puede ser una serina proteasa o una metalo proteasa, preferiblemente una proteasa alcalina microbiana o una proteasa de tipo tripsina. Ejemplos de proteasas alcalinas son subtilisin, especialmente aquellas derivadas de *Bacillus*, por ejemplo, subtilisina Novo, subtilisina Carlsberg, subtilisina 309, subtilisina 147 y subtilisina 168 (descritas en WO 89/06279). Ejemplos de proteasas de tipo tripsina son tripsina (p. ej. de origen porcino o bovino) y la proteasa de *Fusarium* descrita en WO 89/06270 y WO 94/25583.

- 35 [0152] Ejemplos de proteasas útiles son las variantes descritas en WO 92/19729, WO 98/20115, WO 98/20116, y WO 98/34946, especialmente las variantes con sustituciones en una o más de las posiciones siguientes: 27, 36, 57, 76, 87, 97, 101, 104, 120, 123, 167, 170, 194, 206, 218, 222, 224, 235 y 274.

- 40 Lipasas: lipasas adecuadas incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. Ejemplos de lipasas útiles incluyen lipasas de *Humicola* (sinónimo *Thermomyces*), por ejemplo de *H. lanuginosa* (*T. Lanuginosus*) como descrito en EP 258 068 y EP 305 216 o de *H. insolens* como descrito en WO 96/13580, una lipasa de *Pseudomonas*, por ejemplo de *P. alcaligenes* o *P. pseudoalcaligenes* (EP 218 272), *P. cepacia* (EP 331 376), *P. stutzeri* (GB 1,372,034), *P. fluorescens*, cepas de *Pseudomonas* sp. SD 705 (WO 95/06720 y WO 96/27002), *P. wisconsinensis* (WO 96/12012), una lipasa de *Bacillus*, por ejemplo de *B. subtilis* (Dartois et al. (1993), Biochemica et Biophysica Acta, 1131, 253-360), *B. stearothermophilus* (JP 64/744992) o *B. pumilus* (WO 91/16422).

- 45 [0153] Otros ejemplos son variantes de lipasa tales como aquellos descritos en WO 92/05249, WO 94/01541, EP 407 225, EP 260 105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 y WO 97/07202.

Amilasas: amilasas adecuadas (beta y/o alfa) incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. Amilasas incluyen, por ejemplo, alfa-amilasas obtenidas de *Bacillus*, por ejemplo una cepa especial de *B. licheniformis*, descrita con más detalle en GB 1,296,839.

- 55 [0154] Ejemplos de amilasas útiles son las variantes descritas en WO 94/02597, WO 94/18314, WO 96/23873, y WO 97/43424, especialmente las variantes con sustituciones en una o más de las posiciones siguientes: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408, y 444.

- 60 Celulasas: celulasas adecuadas incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. Celulasas adecuadas incluyen celulasas del género *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium*, por ejemplo, las celulasas fúngicas producidas de *Humicola insolens*, *Myceliophthora thermophila* y *Fusarium oxysporum* descritas en US 4,435,307, US 5,648,263, US 5,691,178, US 5,776,757 y WO 89/09259.

- 65 [0155] Celulasas especialmente adecuadas son las celulasas neutrales o alcalinas con beneficios de cuidado de color. Ejemplos de tales celulasas son celulasas descritas en EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son variantes de celulasa tales como aquellas descritas en WO 94/07998, EP 0 531 315, US

5,457,046, US 5,686,593, US 5,763,254, WO 95/24471, WO 98/12307 y PCT/DK98/00299.

Peroxidasas/oxididasas: peroxidasas/oxididasas adecuadas incluyen aquellas de origen vegetal, fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. Ejemplos de peroxidasas útiles incluyen peroxidasas de *Coprinus*, por ejemplo de *C. cinereus*, y variantes de las mismas como aquellas descritas en WO 93/24618, WO 95/10602, y WO 98/15257.

[0156] La enzima(s) detergente se puede incluir en una composición detergente añadiendo aditivos separados con una o más enzimas, o por adición de un aditivo combinado que incluya todas estas enzimas. Un aditivo detergente de la invención, es decir, un aditivo separado o un aditivo combinado, se puede formular por ejemplo como un granulado, un líquido, un compuesto acuoso, etc. Las formulaciones de aditivo detergente preferidas son granulados, en particular granulados no polvorientos, líquidos, en particular líquidos estabilizados, o suspensiones acuosas.

[0157] Granulados no polvorientos pueden ser producidos, por ejemplo, como descritos en US 4,106,991 y 4,661,452 y pueden opcionalmente ser revestidos por métodos conocido en la técnica. Ejemplos de materiales de recubrimiento ceroso son productos de poli(etileno óxido) (polietilenglicol; PEG) con pesos molares medios de 1.000 a 20.000 nonilfenoles etoxilados con de 16 a 50 unidades de óxido de etileno; alcoholes grasos etoxilados en los que el alcohol contiene de 12 a 20 átomos de carbono y en el que hay de 15 a 80 unidades de óxido de etileno; alcoholes grasos; ácidos grasos; y mono- y di- y triglicéridos de ácidos grasos. Ejemplos de materiales de recubrimiento que forman película adecuados para aplicación por técnicas de lecho fluido se dan en GB 1483591. Preparaciones enzimáticas líquidas pueden, por ejemplo, ser estabilizadas añadiendo un poliol tal como propilenglicol, un azúcar o alcohol de azúcar, ácido láctico o ácido bórico según métodos establecidos. Enzimas protegidas se pueden preparar según al método descrito en EP 238,216.

[0158] La composición detergente de la invención puede ser en cualquier forma conveniente, por ejemplo, una barra, una pastilla, un polvo, un gránulo, una pasta o un líquido. Un detergente líquido puede ser acuoso, típicamente con hasta 70 % agua y 0-30 % solvente orgánico o no acuoso.

[0159] La composición detergente comprende uno o más tensioactivos, que pueden ser no iónicos, incluyendo semipolar y/o aniónico y/o zwitteriónico y/o catiónico. Los tensioactivos están típicamente presentes a un nivel de 0,1 % a 60% en peso.

[0160] Cuando se incluye en esto, el detergente normalmente contiene aproximadamente de 1% a aproximadamente 40% de un tensioactivo aniónico tal como alquilbencenosulfonato lineal, alfa-olefinsulfonato, sulfato de alquilo (sulfato de alcohol graso), etoxisulfato alcohólico, alcanosulfonato secundario, éster metílico de ácido graso alfa-sulfo, alquil- o ácido alquenilsuccínico o jabón.

[0161] Cuando se incluye en esto, el detergente normalmente contiene aproximadamente de 0,2% a aproximadamente 40% de un tensioactivo no iónico tal como alcohol etoxilado, nonilfenol etoxilado, alquilpoliglicósido, óxido de alquildimetilamina, monoetanolamida de ácido graso etoxilado, monoetanolamida de ácido graso, amida de ácido graso de polihidroxi-alquilo, o derivados de n-alquilo de n-acilo de glucosamina ("glucamidas").

[0162] El detergente puede contener 0-65 % de un constructor de detergente o agente complejante tal como zeolita, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilenodiaminatetraacético, ácido dietilenotriaminopentaacético, ácido alquenilsuccínico o de alquilo, silicatos solubles o silicatos estratificados (p. ej. SKS-6 de Hoechst).

[0163] El detergente puede comprender uno o más polímeros. Ejemplos son carboximetilcelulosa, poli(vinilpirrolidona), poli(etilenglicol), alcohol polivinílico, poli(vinilpiridina-N-óxido), poli(vinilimidazol), policarboxilatos tal como poliacrilatos, copolímeros de ácido maléico/acrílico y copolímeros de ácido acrílico/lauril metacrilato.

[0164] El detergente puede contener un sistema blanqueante que puede comprender una fuente de H₂O₂ tal como perborato o percarbonato, que se puede combinar con un activador blanqueante de formación de perácido tal como tetraacetilietilenodiamina o nonanoiloxibencenosulfonato. Alternativamente, el sistema blanqueante puede comprender peroxiácidos de, por ejemplo, amida, imida o tipo sulfona.

[0165] La enzima(s) de la composición detergente de la invención puede ser estabilizada usando agentes estabilizantes convencionales, por ejemplo, un poliol tal como propilenglicol o glicerol, un azúcar o alcohol de azúcar, ácido láctico, ácido bórico o un derivado de ácido bórico, por ejemplo, un éster de borato aromático o un derivado de ácido borónico de fenilo, tal como ácido borónico de 4-formilfenil, y la composición se puede formular como se describe en, por ejemplo, WO 92/19709 y WO 92/19708.

[0166] El detergente puede también contener otros ingredientes de detergentes convencionales tales como por ejemplo acondicionadores de tejidos, incluidas arcillas, potenciadores de espuma, supresores de espuma, agentes anticorrosivos, agentes de suspensión de suciedad, agentes de reposición antisuciedad, colorantes, bactericidas, blanqueadores ópticos, hidrotropos, inhibidores de decoloración o perfumes.

[0167] Es actualmente contemplado que en las composiciones detergentes cualquier enzima, en particular el polipéptido de la invención, se puede adicionar en una cantidad correspondiente a 0,01-100 mg de proteína enzimática por litro de solución de lavado, preferiblemente 0,05-5 mg de proteína enzimática por litro de solución de lavado, en particular 0,1-1 mg de proteína enzimática por litro de solución de lavado.

[0168] El polipéptido de la invención puede adicionalmente ser incorporado en las formulaciones detergentes descritas en WO 97/07202 que es por la presente incorporada como referencia.

Recombinación de ADN (redistribución)

[0169] Las secuencias de nucleótidos de SEC ID nº 1 se pueden usar en un proceso de recombinación (o de redistribución) de ADN. Las nuevas secuencias polinucleótidas obtenidas en tal proceso pueden codificar polipéptidos nuevos con actividad de celobiasa con propiedades mejoradas, tales como estabilidad mejorada (estabilidad de almacenamiento, termoestabilidad), actividad específica mejorada, pH óptimo mejorado y/o tolerancia mejorada a compuestos específicos.

[0170] Redistribución entre dos o más polinucleótidos de entrada homólogos (polinucleótidos de punto de inicio) implica fragmentación de los polinucleótidos y recombinación de los fragmentos para obtener polinucleótidos de salida (es decir, polinucleótidos que han sido sometidos a un ciclo de redistribución) donde un número de fragmentos de nucleótidos se cambian en comparación con los polinucleótidos de entrada.

[0171] Recombinación o redistribución de ADN puede ser un proceso (parcial) aleatorio en el que una genoteca de genes quiméricos se genera de dos o más genes de inicio. Varios formatos conocidos pueden utilizarse para efectuar este proceso de redistribución o de recombinación.

[0172] El proceso puede implicar fragmentación aleatoria de ADN progenitor seguida de reensamblaje por PCR para nuevos genes en toda su longitud, por ejemplo como presentado en US5605793, US5811238, US5830721, US6117679. Recombinación in vitro de genes puede ser realizada, por ejemplo como descrito en US6159687, WO98/41623, US6159688, US5965408, US6153510. El proceso de recombinación puede ocurrir in vivo en una célula viva, por ejemplo como descrito en WO 97/07205 y WO 98/28416.

[0173] El ADN progenitor se puede fragmentar por tratamiento de ADNasa I o por digeridos de endonucleasa de restricción como descrito por Kikuchi et al (2000a, Gene 236:159-167). La redistribución de dos progenitores puede realizarse por redistribución del ADN progenitor monocatenario de los dos progenitores como descrito en Kikuchi et al (2000b, Gene 243:133-137).

[0174] Un método particular de redistribución es seguir los métodos descritos en Cramer et al, 1998, Nature, 391: 288-291 y Ness et al. Nature Biotechnology 17: 893-896. Otro formato serían los métodos descritos en US 6159687: Ejemplos 1 y 2.

Producción de etanol a partir de biomasa

[0175] La presente invención también se refiere a métodos para producir etanol a partir de biomasa, tal como materiales celulósicos, incluyendo contacto de la biomasa con los polipéptidos de la invención. Etanol puede posteriormente ser recuperado. Los polipéptidos de la invención pueden ser producidos "in-situ", es decir, como parte de o directamente en un proceso de producción de etanol, por cultivo de una célula huésped o una cepa, que en su forma de tipo salvaje es capaz de producir los polipéptidos, bajo condiciones propicias para la producción de los polipéptidos.

[0176] Etanol se puede producir por degradación enzimática de biomasa y conversión de los polisacáridos liberados en etanol. Esta especie de etanol es frecuentemente llamada bioetanol o biocombustible. Esto se puede usar como un aditivo de combustible o suplemento en mezclas de menos de 1% y hasta 100% (un sustituto de combustible). En algunos países, tal como Brasil, etanol está sustituyendo a la gasolina muy ampliamente.

[0177] El polisacárido predominante en la pared celular primaria de biomasa es celulosa, el segundo más abundante es hemicelulosa y el tercero es pectina. La pared celular secundaria, producida después de que la célula haya detenido su crecimiento, también contiene polisacáridos y se refuerza a través de lignina polimérica de manera covalente reticulada para hemicelulosa. Celulosa es un homopolímero de anhidrocelobiosa y así un beta-(1-4)-D-glucano lineal, mientras hemicelulosas incluyen una variedad de compuestos, tales como xilanos, xiloglucanos, arabinoxilanas y mananos en estructuras ramificadas complejas con un espectro de sustituyentes. Aunque generalmente poliforme, se encuentra celulosa en tejido vegetal principalmente como una matriz de cristalino insoluble de cadenas de glucano paralelo. Hemicelulosas normalmente enlazan por hidrógeno a celulosa, al igual que a otras hemicelulosas, lo que ayuda a estabilizar la matriz de pared celular.

[0178] Tres clases principales de enzimas de celulasa se utilizan para descomponer biomasa:
- las "endo-1,4-beta-glucanasas" o 1,4-beta-D-glucano-4-glucanohidrolasas (EC 3.2.1.4), que actúan de forma aleatoria en sustratos de 1,4-beta-glucano insoluble y soluble.

- las "exo-1,4-beta-D-glucanasas" que incluye 1,4-beta-D-glucano glucohidrolasas (EC 3.2.1.74), que liberan D-glucosa de 1,4-beta-D-glucanos e hidrolizan D-celobiosa lentamente, y 1,4-beta-D-glucano celobiohidrolasa (EC 3.2.1.91), también llamada celobiohidrolasa I, que libera D-celobiosa de 1,4- beta-glucanos.
- las "beta-D-glucosidasas" o glucohidrolasas de beta-d-glucósido (EC 3.2.1.21), que actúan para liberar unidades de D-glucosa de celobiosa y celodextrinas solubles, al igual que un conjunto de glucósidos.

[0179] Estas tres clases de enzimas trabajan juntas sinérgicamente en una interacción compleja que produce decristalización eficaz e hidrólisis de celulosa nativa de biomasa para producir los azúcares reductores que se convierten en etanol por fermentación.

[0180] La presente invención es posteriormente descrita por los siguientes ejemplos.

EJEMPLOS

[0181] Productos químicos usados como tampones y sustratos fueron productos comerciales de al menos calidad reactiva.

Ejemplo 1

Clonación de una secuencia de ADN de celobiohidrolasa I parcial y en toda su longitud (CBH1)

[0182] Una genoteca de ADNc de *Diplodia gossypina* fue seleccionada mediante PCR para presencia del gen CBH1. Para este propósito conjuntos de cebadores fueron contruidos, basados en alineación de secuencia e identificación de regiones conservadas entre proteínas de CBH1. La banda de PCR de una electrofóresis en gel fue usada para obtener una secuencia parcial del gen CBH1 de *Diplodia gossypina*. La búsqueda de homología confirmó que la secuencia parcial era una secuencia parcial del gen CBH1 (EC 3.2.1.91).

[0183] El gen CBH1 en toda su longitud de *Diplodia gossypina* se obtiene por acceso del depósito de patente CBS 247,96, hacer una preparación de ADN o ADNc, usar la secuencia parcial como base para construcción de cebadores específicos y usar técnicas de clonación de PCR estándar para paso a paso obtener el gen entero.

[0184] Otros métodos diferentes puede ser tomados:

- Podría realizarse una selección PCR de la genoteca de ADNc o de los ADNc que fueron usados para la construcción de la genoteca. Para hacer así, cebadores específicos de gen (GSP) y cebadores de vector/adaptador se construyen de la secuencia de ADNc parcial del gen CBH1 y de secuencia de vector/adaptador respectivamente; ambos conjuntos de cebadores diseñados para ir externos en las regiones faltantes 5' y 3' del ADNc de CBH1. Los productos de PCR más largos obtenidos usando combinaciones de GSP y cebador de vector/adaptador representan las regiones finales 5' y 3' del ADNc de CBH1 de *Diplodia gossypina*. Búsqueda de homología y comparación con la secuencia de ADNc parcial confirman que los productos de PCR 5' y 3' pertenecen al mismo ADNc de CBH1 de *Diplodia gossypina*. El ADNc de longitud completa puede entonces ser obtenido por PCR usando un conjunto de cebadores contruidos de las extremidades 5' y 3'.
- Alternativamente, la genoteca de ADNc podría ser seleccionada para el ADNc en toda su longitud usando técnicas de hibridación estándar y la secuencia de ADNc parcial como una sonda. Los clones que dan una señal de hibridación positiva con la sonda son entonces purificados y ordenados para determinar la secuencia de ADNc más larga. Búsqueda de homología y comparación confirma que el ADNc en toda su longitud corresponde a la secuencia parcial de ADNc de CBH1 que fue originalmente usada como una sonda.

[0185] Los dos métodos anteriormente descritos cuentan con la presencia del ADNc de CBH1 en toda su longitud en la genoteca de ADNc o en los ADNc usados para su construcción. Alternativamente, las técnicas de RACE (amplificación rápida de extremidades de ADNc) 5' y 3' o técnicas derivadas podrían ser usadas para identificar las regiones 5' y 3' faltantes. Para este propósito, se aíslan preferiblemente ADNm de *Diplodia gossypina* y se utilizan para sintetizar primero cepas de ADNc usando cebador de adaptador conteniendo oligo(dT)- o un 5'- cebador específico de gen (GSP).

[0186] El ADNc en toda su longitud del gen CBH1 de *Diplodia gossypina* puede también ser obtenido usando ADN genómico de *Diplodia gossypina*. El gen CBH1 se puede identificar por técnicas de PCR tales como la descrita anteriormente o por selección de genoteca genómica estándar usando técnicas de hibridación y el ADNc de CBH1 parcial como una sonda. Búsqueda de homología y comparación con el ADNc de CBH1 parcial confirma que la secuencia genómica corresponde al gen CBH1 de *Diplodia gossypina*. Identificación de secuencias de consenso tales como sitio de iniciación de transcripción, inicio y parada de codones o sitios poliA podrían ser usados para definir la región que comprende el ADNc en toda su longitud. Cebadores contruidos de las extremidades 5' y 3' de esta región podrían entonces ser usados para amplificar el ADNc en toda su longitud de ADNm o la genoteca de ADNc de *Diplodia gossypina* (ver arriba).

[0187] Por expresión del gen en toda su longitud en un constructo huésped de expresión adecuada la enzima CBH1 se cosecha como una enzima intra celular o extra celular del caldo de cultivo.

[0188] Los métodos anteriormente descritos recurren a la clonación de secuencias de ADN de celobiohidrolasa I de todos organismos y no sólo de *Diplodia gossypina*.

5 Ejemplo 2

Actividad de celobiohidrolasa I (CBH I)

10 [0189] Una celobiohidrolasa I se caracteriza por la capacidad para hidrolizar celulosa altamente cristalina muy eficazmente en comparación con otras celulasas. Celobiohidrolasa I puede tener una actividad catalítica más alta usando PASC (celulosa hinchada de ácido fosfórico) como sustrato que usando CMC como sustrato. Para los fines de la presente invención, cualquiera de los siguientes ensayos puede utilizarse para identificar una celobiohidrolasa I:

15 Actividad en Azo-Avicel

[0190] Azo-Avicel (Megazyme, Bray Business Park, Bray, Wicklow, Ireland) fue usado según las instrucciones de los fabricantes.

20 Actividad en PNP-beta-celobiosa

[0191]

Solución de sustrato: 5 mM PNP beta-D-celobiosa (p-nitrofenil β -D-Celobiosida Sigma N-5759) en 0,1 M tampón de Na-acetato, pH 5,0;

25 [0192] Reactivo de parada: 0.1 M Na-carbonato, pH 11.5.

[0193] Solución de 50 μ L CBH I fue mezclada con 1 mL de solución de sustrato e incubada 20 minutos a 40°C. La reacción fue detenida por adición de 5 mL de reactivo de parada. Absorbancia fue medida a 404 nm.

30 Actividad en PASC y CMC

[0194] El sustrato se degrada con celobiohidrolasa I (CBH I) para formar azúcares reductores. Un Microdochium nivale (rMnO) de carbohidrato oxidasa u otra oxidasa equivalente actúa en los azúcares reductores para formar H_2O_2 en presencia de O_2 . El formado H_2O_2 activa en presencia de peroxidasa en exceso la condensación oxidante de 4-aminoantipirina (AA) y N-etil-N-sulfopropil-m-toluidina (TOPS) para formar un producto morado que se puede cuantificar por su absorbancia a 550 nm.

35 [0195] Cuando todos los componentes excepto CBH I están en exceso, el índice de aumento en la absorbancia es proporcional a la actividad de CBH I. La reacción es una reacción de un paso cinético y se puede realizar automáticamente en un analizador por centrifugación Cobas Fara (Hoffmann La Roche) u otro espectrofotómetro equivalente que pueda medir cinética estable. Tampón: 50 mM tampón de Na-acetato (pH 5,0);

40 [0196] Reactivos: rMnO oxidasa, carbohidrato oxidasa purificada de Microdochium nivale, 2 mg/L (concentración final); Peroxidasa, SIGMA P-8125 (96 U/mg), 25 mg/L (concentración final); 4-aminoantipirina, SIGMA A-4382, 200 mg/L (concentración final); TOPS, SIGMA E-8506, 600 mg/L (concentración final); PASC o CMC (ver más abajo), 5 g/L (concentración final).

45 [0197] Todos los reactivos fueron adicionados al tampón en las concentraciones indicadas por encima y esta solución reactiva fue mezclada íntegramente.

50 [0198] Muestra de 50 μ L de celobiohidrolasa I (en una dilución adecuada) fue mezclada con 300 μ L de solución reactiva e incubada 20 minutos a 40°C. Formación de color morado fue detectada y medida como absorbancia a 550 nm.

55 [0199] El coeficiente de absorción de AA-TOPS-condensado es 0,01935 $A_{550}/(\mu M \text{ cm})$. El índice se calcula como μ moles que reducen azúcar producido por minuto de OD550 /minuto y el coeficiente de absorción.

PASC:

[0200]

60 Materiales: 5 g de Avicel® (Art. 2331 Merck);
150 mL 85% de Orto-fosfórico-ácido (Art. 573 Merck);
800 mL de acetona (Art. 14 Merck);
Aprox. 2 litros de agua desionizada (Milli-Q);
65 Vaso de precipitado de vidrio de 1 litro;
Embudo de filtro de vidrio de 1 litro;

Matraz de succión de 2 litros;
Homogeneizador UltraTurrax.

- 5 [0201] Acetona y orto-fosfórico-ácido se enfría en el hielo. Avicel® se humedece con agua y luego 150 mL 85% orto-fosfórico-ácido helado se añade. La mezcla se coloca en un baño de hielo con agitación débil durante una hora.
- 10 [0202] Añadir 500 mL de acetona helada con agitación y transferir la mezcla a un embudo de filtro de vidrio y lavar con 3 x 100 mL de acetona helada, aspirar lo más seco posible en cada lavado. Lavar con 2 x 500 mL de agua (o hasta no haya olor de acetona), aspirar lo más seco posible en cada lavado.
- 15 [0203] Resuspender los sólidos en el agua a un volumen total de 500 mL y mezclar para homogeneidad usando un homogeneizador Ultra Turrax. Guardar húmedo en el frigorífico y equilibrar con tampón por centrifugado y resuspensión antes de uso.
- 20 CMC:
- [0204] Microfibrillas de celulosa bacteriana en una forma impura fueron obtenidas del producto alimenticio japonés "nata de coco" (Fujico Company, Japón). La celulosa en 350 g de este producto fue purificada por suspensión del producto en aproximadamente 4 L de agua del grifo. Esta agua fue sustituida por agua dulce dos veces al día durante 4 días.
- 25 [0205] Luego 1% (p/v) de NaOH fue usado en vez de agua y el producto fue resuspendido en la solución de alcali dos veces al día durante 4 días. Neutralización fue hecha por aclarado de la celulosa purificada con agua destilada hasta que el pH de la superficie del producto fue neutral (pH 7).
- 30 [0206] La celulosa fue microfibrilada y una suspensión de microfibrillas de celulosa bacteriana individual se obtuvo por homogenización de las microfibrillas de celulosa purificada en una mezcladora Waring durante 30 min. Las microfibrillas de celulosa fueron de nuevo purificadas por dialización de esta suspensión a través de una membrana de poro contra agua destilada y las microfibrillas de celulosa aisladas y purificadas fueron almacenadas en una suspensión de agua a 4°C.
- Depósito de material biológico
- China General Microbiological Culture Collection Center (CGMCC)
- 35 [0207] El siguiente material biológico ha sido depositado según las condiciones del tratado de Budapest con el China General Microbiological Culture Collection Center (CGMCC), Institute of Microbiology, Chinese Academy of Sciences, Haidian, Beijing 100080, China:
- 40 Número de accesión: CGMCC n°. 0584
Referencia de solicitantes: ND000575
Fecha de depósito: 29-05-2001
Descripción: gen de CBH I de *Acremonium thermophilum* en plásmido
Clasificación: *Ascomycota*; *Sordariomycetes*; *Hypocrales*; *Hypocreaceae*
Origen: China, 1999
- 45 Secuencia(s) relacionada: SEC ID n° 1 y SEC ID n° 2 (secuencia de ADN que codifica una celobiohidrolasa I de *Acremonium thermophilum* y la secuencia de proteína correspondiente)
- 50 Número de registro: **CGMCC n° 0581**
Referencia de solicitantes: ND000548
Fecha de depósito: 29-05-2001
Descripción: gen de CBH I de *Chaetomium termophilum* en el plásmido
- 55 Clasificación: *Ascomycota*; *Sordariomycetes*; *Sordariales*; *Chaetomiaceae*
Origen: China, 1999
Secuencia(s) relacionada: SEC ID n° 3 y SEC ID n° 4 (secuencia de ADN que codifica una celobiohidrolasa I de *Chaetomium termophilum* y la secuencia de proteína correspondiente)
- 60 Número de registro: **CGMCC n° 0585**
Referencia de solicitantes: ND001223
Fecha de depósito: 29-05-2001
Descripción: Gen de sp. CBH I de *Scytalidium sp.* en el plásmido
- 65 Clasificación: *Ascomycota*; *Mitosporic*
Origen: China, 1999
Secuencia(s) relacionada: SEC ID n° 5 y SEC ID n° 6 (secuencia de ADN que codifica una celobiohidrolasa I de

Scytalidium sp. y la secuencia de proteína correspondiente)

Número de registro: **CGMCC nº 0582**

Referencia de solicitantes: N D000549

5 Fecha de depósito: 29-05-2001

Descripción: gen de CBH I de *Thermoascus aurantiacus* en el plásmido

Clasificación: *Eurotiomycetes*; *Eurotiales*; *Trichocomaceae*

Origen: China

10 Secuencia(s) relacionada: SEC ID nº 7 y SEC ID nº 8 (secuencia de ADN que codifica una celobiohidrolasa I de *Thermoascus aurantiacus* y la secuencia de proteína correspondiente)

Número de registro: **CGMCC nº 0583**

Referencia de solicitantes: ND001182

Fecha de depósito: 29-05-2001

15 Descripción: gen de CBH I de *Thielavia australiensis* en el plásmido

Clasificación: *Ascomycota*; *Sordariomycetes*; *Sordariales*; *Chaetomiaceae*

Origen: China, 1998

20 Secuencia(s) relacionada: SEC ID nº 9 y SEC ID nº 10 (secuencia de ADN que codifica una celobiohidrolasa I de *Thielavia australiensis* y la secuencia de proteína correspondiente)

Número de registro: **CGMCC nº 0580**

Referencia de solicitantes: N D000562

Fecha de depósito: 29-05-2001

Descripción: Gen de CBH I de *Melanocarpus albomyces* en el plásmido

25 Clasificación: *Ascomycota*; *Sordariomycetes*; *Sordariales*

Origen: China, 1999

Secuencia(s) relacionada(s): SEC ID nº 15 y SEC ID nº 16 (secuencia de ADN que codifica una celobiohidrolasa I de *Melanocarpus albomyces* y la secuencia de proteína correspondiente)

30 Número de registro: **CGMCC nº 0748**

Referencia de solicitantes: ND001181

Fecha de depósito: 07-06-2002

Descripción: Gen de CBH I de *Acremonium* sp. en el plásmido

Clasificación: *Ascomycetes* mitospóricos

35 Origen: China, 2000

Secuencia(s) relacionada: SEC ID nº 53 y SEC ID nº 54

Número de registro: **CGMCC nº 0749**

Referencia de solicitantes: N D000577

40 Fecha de depósito: 2002-06-07

Descripción: Gen de CBH I de *Chaetomidium pingtungium* en el plásmido

Clasificación: *Chaetomiaceae*, *Sordariales*, *Ascomycota*

Origen: China, 2000

Secuencia(s) relacionada(s): SEC ID nº 55 y SEC ID nº 56

45 Número de registro: CGMCC nº 0747

Referencia de solicitantes: ND001175

Fecha de depósito: 07-06-2002

Descripción: Gen de CBH I de *Sporotrichum pruinosum* en el plásmido

Clasificación: *Meruliaceae*, *Stereales*, *Basidiomycota*

50 Origen: China, 2000

Secuencia(s) relacionada: SEC ID nº 57 y SEC ID nº 58

Número de registro: CGMCC nº 0750

Referencia de solicitantes: ND000571

55 Fecha de depósito: 07-06-2002

Descripción: Gen de CBH I de *Scytalidium thermophilum* en el plásmido

Clasificación: *Ascomycota*; *Mitosporic*

Origen: China, 2000

Secuencia(s) relacionada: SEC ID nº 59 y SEC ID nº 60

60

Centraalbureau Voor Schimmelcultures (CBS)

[0208] El siguiente material biológico ha sido depositado según las condiciones del Tratado de Budapest en el Centraalbureau Voor Schimmelcultures (CBS), Uppsalalaan 8,3584 CT Utrecht, Países Bajos (alternativamente P.O.Box 85167,3508 AD Utrecht, Países Bajos):

65

- Número de acceso: **CBS 109513**
Referencia de solicitantes: ND000538
Fecha de depósito: 01-06-2001
Descripción: *Verticillium tenerum*
5 Clasificación: *Ascomycota, Hypocreales, Pyrenomycetes* (mitospórico)
Origen:-
Secuencia(s) relacionada(s): SEC ID nº 11 y SEC ID nº 12 (secuencia de ADN que codifica una celobiohidrolasa I de *Verticillium tenerum* y la secuencia de proteína correspondiente)
- 10 Número de acceso: **CBS 819.73**
Referencia de solicitantes: ND000533
Fecha de depósito: públicamente disponible (no depositado por solicitante)
Descripción: *Humicola nigrescens*
Clasificación: *Sordariaceae, Sordariales, Sordariomycetes; Ascomycota*
15 Origen:-
Secuencia(s) relacionada: SEC ID nº 18 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Humicola nigrescens*)
- Número de acceso: **CBS 427.97**
20 Referencia de solicitantes: ND000530
Fecha de depósito: 23-01-1997
Descripción: *Cladorrhinum foecundissimum*
Clasificación: *Sordariaceae, Sordariales, Sordariomycetes; Ascomycota*
Origen: Jamaica
25 Secuencia(s) relacionada: SEC ID nº 19 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Cladorrhinum foecundissimum*)
- Número de acceso: **CBS 247.96**
Referencia de solicitantes: ND000534 y ND001231
30 Fecha de depósito: 12-03-1996
Descripción: *Diplodia gossypina*
Clasificación: *Dothideaceae, Dothideales, Dothidomycetes; Ascomycota*
Origen: Indonesia, 1992
Secuencia(s) relacionada(s): SEC ID NO:20 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Diplodia gossypina*), SEC ID NO:37 (secuencia de ADN completa que codifica una celobiohidrolasa I de *Diplodia gossypina*) y SEC ID NO:38 (secuencia de proteína completa de celobiohidrolasa I de *Diplodia gossypina*)
35
- Número de acceso: **CBS 117.65**
Referencia de solicitantes: ND000536
40 Fecha de depósito: públicamente disponible
Descripción: *Myceliophthora thermophila*
Clasificación: *Sordariaceae, Sordariales, Sordariomycetes; Ascomycota*
Origen:-
Secuencia(s) relacionada(s): SEC ID nº 21 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Myceliophthora thermophila*)
45
- Número de acceso: **CBS 109471**
Referencia de solicitantes: ND000537
Fecha de depósito: 29-05-2001
50 Descripción: *Rhizomucor pusillus*
Clasificación: *Mucoraceae, Mucorales, Zygomycota*
Origen: Dinamarca
Secuencia(s) relacionada: SEC ID nº 22 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Rhizomucor pusillus*)
55
- Número de registro: **CBS 521.95**
Referencia de solicitantes: ND000542
Fecha de depósito: 04-07-1995
Descripción: *Meripilus giganteus*
60 Clasificación: *Rigidiporaceae, Hymenomycetales, Basidiomycota*
Origen: Dinamarca, 1993
Secuencia(s) relacionada: SEC ID nº 23 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Meripilus giganteus*)
- 65 Número de registro: **CBS 277.96**
Referencia de solicitantes: ND000543; ND001346 y ND001243

- Fecha de depósito: 12-03-1996
 Descripción: *Exidia glandulosa*
 Clasificación: *Exidiaceae, Auriculariales, Hymenomycetes, Basidiomycota*
 Origen: Dinamarca, 1993
- 5 Secuencia(s) relacionada(s): SEC ID nº 24 (secuencia de ADN parcial que codifica parcialmente una celobiohidrolasa I de *Exidia glandulosa*), SEC ID nº 45 (secuencia de ADN completa que codifica una celobiohidrolasa I con CBD de *Exidia glandulosa*), SEC ID nº 46 (secuencia completa de proteína de celobiohidrolasa I con CBD de *Exidia glandulosa*), SEC ID nº 47 (secuencia de ADN completa que codifica una celobiohidrolasa I de *Exidia glandulosa*) y SEC ID nº 48 (secuencia de proteína completa de celobiohidrolasa I de *Exidia glandulosa*)
- 10 Número de acceso: **CBS 284.96**
 Referencia de solicitantes: ND000544 y ND001235
 Fecha de depósito: 12-3-1996
 Descripción: *Xylaria hypoxylon*
- 15 Clasificación: *Sordariaceae, Sordariales, Sordariomycetes, Ascomycota*
 Origen: Dinamarca, 1993
 Secuencia(s) relacionada(s): SEC ID nº 25 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Xylaria hypoxylon*), SEC ID nº 43 (secuencia de ADN completa que codifica una celobiohidrolasa I de *Xylaria hypoxylon*) y SEC ID nº 44 (secuencia de proteína completa de celobiohidrolasa I de *Xylaria hypoxylon*)
- 20 Número de acceso: **CBS 804.70**
 Referencia de solicitantes: ND001227
 Fecha de depósito: públicamente disponible
 Descripción: *Trichophaea saccata*
- 25 Clasificación: *Ascomycota; Pezizomycetes; Pezizales; Pyronemataceae*
 Secuencia(s) relacionada: SEC ID nº 36 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Trichophaea saccata*)
- Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ)
- 30 [0209] El siguiente material biológico ha sido depositado según las condiciones del Tratado de Budapest en el Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Mascheroder Weg 1 b, 38124 Braunschweig, Alemania:
- Número de acceso: DSM 14348
 Referencia de solicitantes: ND000551
 Fecha de depósito: 13-06-2001
 Descripción: gen de CBH I de termitas de *Neotermes castaneus*, en el plásmido
 Clasificación:-
 Origen: Cultivos de larvas de termitas comprados de BAM, Alemania, 1999
- 40 Secuencia(s) relacionada(s): SEC ID nº 13 y SEC ID nº 14 (secuencia de ADN que codifica una celobiohidrolasa I de células de intestino o microbios intestinales de *Neotermes castaneus* y la secuencia de proteína correspondiente)
- Número de registro: **DSM 15066**
 Referencia de solicitantes: ND001349
 Fecha de depósito: 21-06-2002
 Descripción: Gen de CBH I de *Poitrasia circinans* gen en el plásmido
 Clasificación: *Choanephoraceae, Zygomycota, Mucorales*
 Origen:-
- 50 Secuencia(s) relacionada(s): SEC ID nº 49 (secuencia de ADN que codifica una celobiohidrolasa I de *Poitrasia circinans*) y SEC ID NO:50 (secuencia de proteína de celobiohidrolasa I de *Poitrasia circinans*)
- Número de registro: **DSM 15065**
 Referencia de solicitantes: ND001339
 Fecha de depósito: 21-06-2002
 Descripción: Gen de CBH I de *Coprinus cinereus* en el plásmido
 Clasificación: *Basidiomycota, Hymenomycetes; Agaricales, Agaricaceae*
 Origen: Dinamarca
- 55 Secuencia(s) relacionada(s): SEC ID nº 51 (secuencia de ADN que codifica una celobiohidrolasa I de *Coprinus cinereus*) y SEC ID nº 52 (secuencia de proteína de celobiohidrolasa I de *Coprinus cinereus*)
- 60 Número de registro: **DSM 15064**
 Referencia de solicitantes: ND001264
 Fecha de depósito: 21-06-2002
 Descripción: gen de CBH I de *Trichophaea saccata* en el plásmido
- 65 Clasificación: *Ascomycota; Pezizomycetes; Pezizales; Pyronemataceae*

Origen:-

Secuencia(s) relacionada(s): SEC ID nº 39 (secuencia de ADN que codifica una celobiohidrolasa I de *Trichofaea saccata*) y SEC ID nº 40 (secuencia de proteína de celobiohidrolasa I de *Trichophaea saccata*)

5 Número de registro: **DSM 15067**

Referencia de solicitantes: N D001232

Fecha de depósito: 21-06-21

Descripción: Gen de CBH I de *Myceliophthora thermophila* en el plásmido

Clasificación: *Sordariaceae*, *Sordariales*, *sordariomycetes*; *Ascomycota*

10 Origen:-

Secuencia(s) relacionada(s): SEC ID nº 41 (secuencia de ADN que codifica una celobiohidrolasa I de *Miceliophthora thermophila*) y SEC ID nº 42 (secuencia de proteína de celobiohidrolasa I de *Myceliophthora thermophila*)

Instituto para fermentación, Osaka (IFO)

15

[0210] El siguiente material biológico ha sido depositado según las condiciones del Tratado de Budapest en el Instituto para fermentación, Osaka (IFO), 17- 85, Juso-honmachi 2-chome, Yodogawa-ku, Osaka 532-8686, Japón:

Número de registro: IFO 5372

20 Referencia del solicitante: ND000531

Fecha de depósito: públicamente disponible (no depositado por solicitante)

Descripción: *Trichothecium roseum*

Clasificación: *ascomycetes* mitospóricos

Origen:-

25 Secuencia(s) relacionada(s): SEC ID nº 17 (secuencia de ADN parcial que codifica una celobiohidrolasa I de *Trichothecium roseum*)

[0211] El depósito de CBS 427.97, CBS 247.96, CBS 521.95, CBS 284.96, CBS 274.96 fue realizado por Novo Nordisk A/S y fue más tarde cedido a Novozymes A/S.

30

0-1	Formulario - PCT/RO/134 (fácil) indicaciones en relación a Microorganismo(s) depositado(s) y otro material biológico (PCT Regla 13bis)	
0-1-1	Preparado usando	PCT-EASY Versión 2.92 (actualizado 01.06.2002)
0-2	Número de solicitud internacional	
0.3	Referencia de fichero del solicitante o del agente	10129.204-WO

1	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
1-1	página	63-64
1-2	línea	31-2
1-3	Identificación de depósito	
1-3-1	Nombre de institución depositaria	China General Microbiological Culture Collection Center
1-3-2	Dirección Institución depositaria	China Committee for Culture Collection of Microorganisms, P.O. Box 2714, Pekin 100080, China
1-3-3	Fecha de depósito	29 mayo 2001 (29.05.2001)
1-3-4	Número de registro	CGMCC 0584
1-4	Indicaciones adicionales	NINGUNA
1-5	Estados designados que se han indicado	todos los estados designados
1-6	provisión separada de indicaciones Estas indicaciones serán presentadas más tarde ante la Oficina Internacional	NINGUNA
2	Las indicaciones hechas debajo se refieren al microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
2-1	página	64
2-2	línea	4-12
2-3	Identificación de depósito	
2-3-1	Nombre de institución depositaria	China General Microbiological Culture Collection Center

2-3-2	Dirección de Institución depositaria	China Committee for Culture Collection of Microorganisms, P.O. Bos 2714, Pekin 100080, China
2-3-3	Fecha de depósito	29 mayo 2001 (29.05.2001)
2-3-4	Número de registro	CGMCC 0581
2-4	Indicaciones adicionales	NINGUNA
2-5	Estados designados que se han indicado	Todos los Estados designados
2-6	Suministro separado de indicaciones Estas indicaciones serán presentadas más tarde ante la Oficina Internacional	NINGUNA
3	Las indicaciones hechas abajo se refieren al microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
3-1	página	64
3-2	línea	14-22
3-3	Identificación de depósito	
3-3-1	Nombre de Institución depositaria	China General Microbiological Culture Collection
3-3-2	Dirección de Institución depositaria	China Committee for Culture Collection of Microorganisms, P.O. Box 2714, Pekin 100080, China
3-3-3	Fecha de depósito	29 mayo 2001 (29.05.2001)
3-3-4	Número de registro	CGMCC 0585
3-4	Indicaciones adicionales	NINGUNA
3-5	Estados designados que se han indicado	todos los estados designados
3-6	Suministro separado de indicaciones	NINGUNA
	Estas indicaciones se presentarán más tarde ante la Oficina Internacional	
4	Las indicaciones hechas debajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
4-1	página	64
4-2	línea	24-32
4-3	identificación de depósito	
4-3-1	Nombre de Institución depositaria	China General Microbiological Culture Collection Center
4-3-2	Dirección de Institución depositaria	China Committe for Culture Collection of Microorganisms, P.O. Box 2714, Pekin 00080, China
4-3-3	Fecha de depósito	29 mayo 2001 (29.05.2001)
4-3-4	Número de registro	CGMCC 0582
4-4	Indicaciones adicionales	NINGUNA
4-5	Estados designados que se han indicado	todos los estados designados
4-6	Suministro separado de indicaciones	NINGUNA
	Estas indicaciones se presentarán más tarde ante la Oficina Internacional	
5	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en descripción en:	
5-1	Página	64-65
5-2	Línea	34-5
5-3	Identificación de depósito	
5-3-1	Nombre de Institución depositaria	China General Microbiological Culture Collection Center
5-3-2	Dirección de Institución depositaria	China Committe for Culture Collection of Microorganisms, P.O. Box 2714, Pekin 100080, China
5-3-3	Fecha de depósito	29 mayo 2001 (29.05.2001)
5-3-4	Número de registro	CGMCC 0 5 8 3
5-4	Indicaciones adicionales	NINGUNA
5-5	Estados designados que se han indicado	todos los Estados designados
5-6	Suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
6	Las indicaciones hechas debajo se refieren al microorganismo(s) depositado(s) u otro material biológico	

6-1	referido en la descripción en: página	65
6-2.	línea	7-15
6-3	Identificación de depósito	China General Microbiological Culture Collection Center China Committee for Culture Collection of Microorganisms, P.O. Box 2714, Pekin 100080, China 29 mayo 2001 (29.05.2001) CGMCC 0580
6-3-1	Nombre de Institución depositaria	
6-3-2	Dirección de Institución depositaria	
6-3-3	Fecha de depósito	
6-3-4	Número de registro	CGMCC 0580
6-4	Indicaciones adicionales	NINGUNA
6-5	Estados designados para los que se han hecho indicaciones	todos los estados designados
6-6	Suministro separado de indicaciones Estas indicaciones serán sometidas más tarde ante la Oficina Internacional	NINGUNA
7	Las indicaciones hechas abajo se refieren a los microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
7-1	página	65
7-2	línea	17-23
7-3	Identificación de depósito	Général Chinois de Cultures Microbiologiques Chine- Comité pour la collection de cultures de micro-organismes, P.O. Box 2714, Pekin 100080 07 junio 2002 (07.06.2002) CGCCM 0748
7-3-1	Nombre de Institución depositaria	
7-3-2	Dirección de Institución depositaria	
7-3-3	Fecha de depósito	
7-3-4	Número de registro	CGCCM 0748
7-4	Indicaciones adicionales	NINGUNA
7-5	Estados designados que se han indicado	Todos los Estados designados
7-6	Suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
8	Las indicaciones hechas abajo se refieren a los microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
8-1	página	65
8-2	línea	25-31
8-3	Identificación de depósito	Centre Général Chinois de Cultures Microbiologiques Chine- Comité pour la collection de cultures de micro-organismes, P.O. Box 2714, Pekin 100080 07 junio 2002 (07.06.2002) CGCCM 0749
8-3-1	Nombre de Institución depositaria	
8-3-2	Dirección de Institución depositaria	
8-3-3	Fecha de depósito	
8-3-4	Número de registro	CGCCM 0749
8-4	Indicaciones adicionales	NINGUNA
8-5	Estados designados para los que se han hecho indicaciones	todos los estados designados
8-6	Suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
9	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
9-1	página	65-66
9-2	línea	33-2
9-3	Identificación de depósito	Centre Général Chinois de Cultivos Microbiologiques Chine - Comité pour la collection de cultures de micro-organismes, P.O. Box 2714, Pekin 100080 07 junio 2002 (07.06.2002) CGCCM 0747
9-3-1	Nombre de Institución depositaria	
9-3-2	Dirección de Institución depositaria	
9-3-3	Fecha de depósito	
9-3-4	Número de registro	CGCCM 0747
9-4	Indicaciones adicionales	NINGUNA
9-5	Estados designados para que se han hecho indicaciones	Todos los Estados designados

9-6	Suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
10	Las indicaciones hechas debajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
10- 1	página	66
10- 2	línea	4-10
10- 3	identificación de depósito	
10- 3-1	nombre de Institución depositaria	Centre Général Chinois de Cultures Microbiologiques
10- 3-2	dirección de Institución depositaria	Chine –Comité pour la collection de cultures de micro-organismes, P.O. Box 2714, Pekin 100080
10- 3-3	Fecha de depósito	07 junio 2002 (07.06.2002)
10- 3-4	Número de registro	CGCCM 0750
10- 4	Indicaciones adicionales	NINGUNA
10- 5	Estados designados para los que se han hecho indicaciones	Todos los Estados designados
10- 6	Suministro separado de indicaciones Estas indicaciones serán presentadas más tarde ante la Oficina Internacional	NINGUNA
11	Las indicaciones hechas debajo se refieren al microorganismo(s) depositado u otro material biológico referido en la descripción en:	
11- 1	página	66
11- 2	línea	18-26
11- 3	Identificación de depósito	
11- 3-1	Nombre de Institución depositaria	Centraalbureau voor Schimmelcultures
11- 3-2	Dirección de Institución de depositaria	Uppsalalaan 8; NL-3584 CT Utrecht, Países Bajos / P.O. Box 85167; NL-3508 AD Utrecht, Países Bajos
11- 3-3	fecha de depósito	01 junio 2001 (01.06.2001)
11- 3-4	número de registro	CBS 109513
11- 4	indicaciones adicionales	NINGUNA
11- 5	Estados designados para los que se han hecho indicaciones	Todos los Estados designados
11- 6	suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
12	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
12- 1	página	66-67
12- 2	línea	37-7
12- 3	identificación de depósito	
12- 3-1	Nombre de Institución de depositaria	Centraalbureau voor Schimmelcultures
12- 3-2	dirección de Institución depositaria	Uppsalalaan 8; NL-3584 CT Utrecht, Países Bajos / P.O. Box 85167; NL-3508 AD Utrecht, Países Bajos
12- 3-3	fecha de depósito	23 enero 1997 (23.01.1997)
12- 3-4	número de registro	CBS 427.97
12- 4	indicaciones adicionales	NINGUNA
12- 5	Estados designados para los que se han hecho indicaciones	todos los Estados designados
12- 6	suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
13	Las indicaciones hechas abajo se refieren al microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
13- 1	página	67
13- 2	línea	9-19
13- 3	identificación de depósito	
13- 3-1	nombre de Institución de depositaria	Centraalbureau voor Schimmelcultures
13- 3-2	dirección de Institución depositaria	Uppsalalaan 8; NL-3584 CT Utrecht, Países

		Bajos / P.O. Box 85167; NL-3508 AD Utrecht, Países Bajos
13- 3-3	fecha de depósito	12 marzo 1996 (12.03.1996)
13- 3-4	número de registro	CBS 247.96
13- 4	indicaciones adicionales	NINGUNA
13- 5	Estados designados para los que se han hecho indicaciones	todos los Estados designados
13- 6	suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina internacional	NINGUNA
14	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
14- 1	página	67
14- 2	línea	30-37
14- 3	identificación de depósito	
14- 3-1	nombre de Institución depositaria	Centraalbureau voor Schimmelcultures
14- 3-2	dirección de Institución depositaria	Uppsalalaan 8; NL-3584 CT Utrecht, Países Bajos / P.O. Box 85167; NL-3508 AD Utrecht, Países Bajos
14- 3-3	fecha de depósito	29 may 2001 (29.05.2001)
14- 3-4	número de registro	CBS 109471
14- 4	indicaciones adicionales	NINGUNA
14- 5	Estados designados para los que se han hecho indicaciones	Todos los Estados designados
14- 6	suministro separado de indicaciones	NINGUNA
	Estas indicaciones se presentarán más tarde ante la Oficina Internacional	
15	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
15- 1	página	68
15- 2	línea	2-9
15- 3	identificación de depósito	
15- 3-1	nombre de Institución depositaria	Centraalbureau voor Schimmelcultures
15- 3-2	dirección de Institución depositaria	Uppsalalaan 8; NL-3584 CT Utrecht, Países Bajos / P.O. Box 85167; NL-3508 AD Utrecht, Países Bajos
15- 3-3	fecha de depósito	04 julio 1995 (04.07.1995)
15- 3-4	número de registro	CBS 521.95
15- 4	indicaciones adicionales	NINGUNA
15- 5	Estados designados para los que se han hecho indicaciones	todos los Estados designados
15- 6	suministro separado de indicaciones	NINGUNA
	Estas indicaciones se presentarán más tarde ante la Oficina Internacional	
16	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
16- 1	página	68
16- 2	línea	26-36
16- 3	identificación de depósito	
16- 3-1	nombre de Institución depositaria	Centraalbureau voor Schimmelcultures
16- 3-2	dirección de institución depositaria	Uppsalalaan 8; NL-3584 CT Utrecht, Países Bajos / P.O. Box 85167; NL-3508 AD Utrecht, Países Bajos
16- 3-3	fecha de depósito	12 marzo 1996 (12.03.1996)
16- 3-4	número de registro	CBS 284.96
16- 4	indicaciones adicionales	NINGUNA
16- 5	Estados designados para los que se han hecho indicaciones	Todos los Estados designados
16- 6	suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA

17	Las indicaciones hechas abajo se refieren al microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
17- 1	página	68
17- 2	línea	11-24
17- 3	identificación de depósito	
17- 3-1	nombre de Institución depositaria	Centraalbureau voor Schimmelcultures
17- 3-2	dirección de Institución depositaria	Uppsalalaan 8; NL-3584 CT Utrecht, Países Bajos / P.O. Caja 85167; NL-3508 AD Utrecht, Países Bajos
17- 3-3	fecha de depósito	12 marzo 1996 (12.03.1996)
17- 3-4	número de registro	CBS 277.96
17- 4	indicaciones adicionales	NINGUNA
17- 5	Estados designados para los que se han hecho indicaciones	Todos los Estados designados
17- 6	suministro separado de indicaciones Estas indicaciones serán presentadas más tarde ante la Oficina Internacional	NINGUNA
18	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
18- 1	página	69
18- 2	línea	15-23
18- 3	identificación de depósito	
183-1	Nombre de Institución depositaria	DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH
18- 3-2	dirección de Institución depositaria	Mascheroder Weg 1b, D-38124 Braunschweig, Alemania
18- 3-3	fecha de depósito	13 junio 2001 (13.06.2001)
18- 3-4	número de registro	DSMZ 14348
18- 4	indicaciones adicionales	NINGUNA
18- 5	Estados designados para los que se han hecho indicaciones	Todos los Estados designados
18- 6	suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
19	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
19- 1	página	69
19- 2	línea	25-33
19- 3	identificación de depósito	
19- 3-1	nombre de institución depositaria	DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH
19- 3-2	dirección de Institución depositaria	Mascheroder Weg 1b, D-38124 Braunschweig, Alemania
19- 3-3	fecha de depósito	21 junio 2002 (21.06.2002)
19- 3-4	número de registro	DSMZ 15066
19- 4	indicaciones adicionales	NINGUNA
19- 5	Estados designados para los que se han hecho indicaciones	todos los Estados designados
19- 6	almacenamiento separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
20	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
20- 1	página	69-70
20- 2	línea	35-6
20- 3	identificación de depósito	
20- 3-1	nombre de Institución depositaria	DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH
20- 3-2	dirección de la Institución depositaria	Mascheroder Weg 1b, D-38124 Braunschweig, Alemania
20- 3-3	fecha de depósito	21 junio 2002 (21.06.2002)

20- 3-4	número de registro	DSMZ 15065
	Indicaciones adicionales	NINGUNA
20- 5	Estados designados para los que se han hecho indicaciones	todos los Estados designados
20- 6	almacenamiento separado de indicaciones	Ninguna
	Estas indicaciones serán presentadas más tarde ante la Oficina Internacional	
21	Las indicaciones hechas abajo se refieren a microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
21- 1	página	70
21- 2	línea	8-16
21- 3	identificación de depósito	
21- 3-1	nombre de Institución depositaria	DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH Mascheroder Weg 1b, D-38124 Braunschweig, Alemania
21- 3-2	dirección de Institución depositaria	
21- 3-3	fecha de depósito	21 junio 2002 (21.06.2002)
21- 3-4	número de registro	DSMZ 15064
21- 4	indicaciones adicionales	NINGUNA
21- 5	Estados designados para los que se han hecho indicaciones	Todos los Estados designados
21- 6	suministro separado de indicaciones Estas indicaciones se presentarán más tarde ante la Oficina Internacional	NINGUNA
22	Las indicaciones hechas abajo se refieren al microorganismo(s) depositado(s) u otro material biológico referido en la descripción en:	
22- 1	página	70
22- 2	línea	18-26
22- 3	identificación de depósito	
22- 3-1	nombre de Institución depositaria	DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH Mascheroder Weg 1b, D-38124 Braunschweig, Alemania
22- 3-2	dirección de Institución depositaria	
22- 3-3	fecha de depósito	21 junio 2002 (21.06.2002)
22- 3-4	número de registro	DSMZ 15067
22- 4	indicaciones adicionales	NINGUNA
22- 5	Estados designados para los que se han hecho indicaciones	todos los Estados designados

Forma de realización preferida

5

[0212]

10

1. Un polipéptido con actividad celobiohidrolasa I, seleccionado del grupo consistente en:

- (a) un polipéptido que incluye una secuencia de aminoácidos que tiene al menos 95% identidad con aminoácidos 1 a 529 de SEC ID n° 4,
- (b) un polipéptido que incluye una secuencia de aminoácidos que tiene al menos 95% identidad con el polipéptido codificado por nucleótidos 1 a 1587 de SEC ID n° 3.

15

2. Polipéptido según forma de realización 1, que consiste de aminoácidos 1 a 529 de SEC ID n° 4.

20

3. Polipéptido según la forma de realización 1, donde el polipéptido es una variante artificial que comprende una secuencia de aminoácidos que tiene al menos una sustitución, eliminación y/o inserción de un aminoácido en comparación con aminoácidos 1 a 529 de SEC ID n° 4.

25

4. Polipéptido según la forma de realización 1, que incluye un aminoácido que tiene al menos 95% identidad, con el polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en el microorganismo depositado CGMCC n° 0581.

5. Polipéptido según la forma de realización 4, que comprende la secuencia de aminoácidos codificada por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en

microorganismo depositado CGMCC nº 0581.

6. Polipéptido según las formas de realización 5 o 6, que consiste en la secuencia de aminoácidos codificada por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en el microorganismo depositado CGMCC nº 0581.

7. Polipéptido según las formas de realización 4 o 5, donde el polipéptido es una variante artificial que comprende una secuencia de aminoácidos que tiene al menos una sustitución, eliminación y/o inserción de un aminoácido en comparación con la secuencia de aminoácidos codificada por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en el microorganismo depositado CGMCC nº 0581.

8. Polinucleótido con una secuencia de nucleótidos que codifica el polipéptido definido en cualquiera de formas de realización 1-7.

9. Constructo de ácidos nucleicos que comprende la secuencia de nucleótidos definida en la forma de realización 8 operativamente enlazada a uno o varias secuencias de control que persiguen la producción del polipéptido en un huésped adecuado.

10. Vector de expresión recombinante que comprende el constructo de ácidos nucleicos definido en la forma de realización 9.

11. Célula huésped recombinante que comprende el constructo de ácidos nucleicos definido en la forma de realización 10.

12. Método para producir un polipéptido tal y como se define en cualquiera de formas de realización 1-7, comprendiendo el método:

(a) cultivo de una célula huésped recombinante tal y como se define en la forma de realización 11 bajo condiciones propicias para la producción del polipéptido; y
(b) recuperación del polipéptido.

13. Método para producción in situ de un polipéptido tal y como se define en cualquiera de formas de realización 1-7, el método comprendiendo:

(a) cultivo de una célula huésped recombinante tal y como se define en la forma de realización 11 bajo condiciones propicias para la producción del polipéptido; y
(b) contactar el polipéptido con un sustrato deseado sin recuperación previa del polipéptido.

14. Método para producir etanol de biomasa, que comprende contactar la biomasa con el polipéptido tal y como se define en cualquiera de formas de realización 1-7.

15. Uso del polipéptido tal y como se define en cualquiera de formas de realización 1-7 para producir etanol.

16. Planta transgénica, parte de planta o célula vegetal, que ha sido transformado con una secuencia de nucleótidos que codifica un polipéptido con actividad de celobiohidrolasa I tal y como se define en cualquiera de las formas de realización 1-7, donde la secuencia de nucleótidos está presente en un vector de expresión que permite integración estable del vector en el genoma de la célula huésped.

17. Composición de detergente que comprende un tensioactivo y el polipéptido según cualquier de formas de realización 1-7.

Listado de secuencias

[0213]

<110> Novozymes A/S

<120> Polipéptidos con actividad de celobiohidrolasa I y polinucleótidos que codifican la misma

<130> 10129-WO

<160> 67

<170> Versión de patentIn 3.1

<210> 1

<211> 1581

<212> ADN

<213> Acremonium thermophilum

<220>
 <221> CDS
 <222> (1)..(1581)
 <223>

5

<400> 1

atg cac gcc aag ttc ggg acc ctc gcc gcc ctt gtg ggg tcc gcc ggg	48
Met His Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Ser Ala Ala	
1 5 10 15	
gcc cag cag gcc tgc aca ctc acg gct gag aac cac ccc acc ctg tgg	96
Ala Gln Gln Ala Cys Thr Leu Thr Ala Glu Asn His Pro Thr Leu Ser	
20 25 30	
tgg tcc aag tgc acg tcc ggc ggc agc tgc acc agc gtc tgg ggc tcc	144
Trp Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Ser Gly Ser	
35 40 45	
gtc acc atc gat gcc aac tgg cgg tgg act cac cag gtc tgg agc tgg	192
Val Thr Ile Asp Ala Asn Trp Arg Trp Thr His Gln Val Ser Ser Ser	
50 55 60	
acc aac tgc tac acg ggc aat gag tgg gac acg tcc atc tgc acc gac	240
Thr Asn Cys Tyr Thr Gly Asn Glu Trp Asp Thr Ser Ile Cys Thr Asp	
65 70 75 80	
ggc gct tgg tgc gcc gcc gcc tgc tgc ctc gat ggc gcc gac tac tgg	288
Gly Ala Ser Cys Ala Ala Ala Cys Cys Leu Asp Gly Ala Asp Tyr Ser	
85 90 95	
ggc acc tat ggc atc acc acc agc ggc aac gcc ctc agc ctc cag ttc	336
Gly Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ala Leu Ser Leu Gln Phe	
100 105 110	
gtc act cag ggc ccc tac tgg acc aac att ggc tgg cgt acc tac ctg	384
Val Thr Gln Gly Pro Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu	
115 120 125	
atg gcc tgg gac acc aag tac cag atg ttc act ctg ctc ggc aac gag	432
Met Ala Ser Asp Thr Lys Tyr Gln Met Phe Thr Leu Leu Gly Asn Glu	
130 135 140	
ttc acc ttc gac gtg gac gtc aca ggc ctc ggc tgc ggt ctg aac ggc	480

10

Phe Thr Phe Asp Val Asp Val Thr Gly Leu Gly Cys Gly Leu Asn Gly			
145	150	155	160
ggc ctc tac ttc gtc tcc atg gac gag gac ggt ggt ctt tcc aag tac	528		
Ala Leu Tyr Phe Val Ser Met Asp Glu Asp Gly Gly Leu Ser Lys Tyr			
165	170	175	
tgc ggc aac aag gct ggc gcc aag tac ggc acc ggc tac tgc gac tgc	576		
Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser			
180	185	190	
cag tgc ccc cgc gac ctc aag ttc atc aac ggc gag gct aac aac gtt	624		
Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Asn Val			
195	200	205	
ggc tgg acc ccg tgc tcc aac gac aag aac gcc ggc ttg ggc aac tac	672		
Gly Trp Thr Pro Ser Ser Asn Asp Lys Asn Ala Gly Leu Gly Asn Tyr			
210	215	220	
ggc agc tgc tgc tcc gag atg gat gtc tgg gag gcc aac agc atc tgc	720		
Gly Ser Cys Cys Ser Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser			
225	230	235	240
gcg gcc tac acg ccc cat cct tgc act acc atc ggc cag acg cgc tgc	768		
Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Thr Arg Cys			
245	250	255	
gag ggc gac gac tgc ggt ggt acc tac agc act gac cgc tac gcc ggc	816		
Glu Gly Asp Asp Cys Gly Gly Thr Tyr Ser Thr Asp Arg Tyr Ala Gly			
260	265	270	
gag tgc gac cct gac gga tgc gac ttc aac tgc tac cgc atg ggc aac	864		
Glu Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asn			
275	280	285	
acg acc ttc tac ggc aag ggc atg acc gtc gac acc agc aag aag ttc	912		
Thr Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Ser Lys Lys Phe			
290	295	300	
acg gtg gtg acc cag ttc ctg acg gac tgc tct ggc aac ctg tcc gag	960		
Thr Val Val Thr Gln Phe Leu Thr Asp Ser Ser Gly Asn Leu Ser Glu			
305	310	315	320
atc aag cgc ttc tac gtc cag aac ggc gtc gtc att ccc aac tgc aac	1008		
Ile Lys Arg Phe Tyr Val Gln Asn Gly Val Val Ile Pro Asn Ser Asn			
325	330	335	
tcc aac atc gcg ggc gtc tgc ggc aac tcc atc acc cag gcc ttc tgc	1056		
Ser Asn Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Gln Ala Phe Cys			
340	345	350	
gat gct cag aag acc gct ttc ggc gac acc aac gtc ttc gac caa aag	1104		
Asp Ala Gln Lys Thr Ala Phe Gly Asp Thr Asn Val Phe Asp Gln Lys			
355	360	365	
ggc ggc ctg gcc cag atg ggc aag gct ctt gcc cag ccc atg gtc ctc	1152		
Gly Gly Leu Ala Gln Met Gly Lys Ala Leu Ala Gln Pro Met Val Leu			
370	375	380	
gtc atg tcc ctc tgg gac gac cac gcc gtc aac atg ctc tgg ctc gac	1200		
Val Met Ser Leu Trp Asp Asp His Ala Val Asn Met Leu Trp Leu Asp			
385	390	395	400

tcg acc tac ccg acc aac gag gcc ggc aag ccg ggc gcc gcc cgc ggt	1248
Ser Thr Tyr Pro Thr Asn Ala Ala Gly Lys Pro Gly Ala Ala Arg Gly	
405 410 415	
acc tgc ccc acc acc tcg ggc gtc ccc gcc gac gtc gag tcc cag gcg	1296
Thr Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu Ser Gln Ala	
420 425 430	
ccc aac tcc aag gtc atc tac tcc aac atc cgc ttc ggc ccc atc ggc	1344
Pro Asn Ser Lys Val Ile Tyr Ser Asn Ile Arg Phe Gly Pro Ile Gly	
435 440 445	
tcc acc gtc tcc ggc ctg ccc ggc ggc ggc agc aac ccc ggc ggc ggc	1392
Ser Thr Val Ser Gly Leu Pro Gly Gly Gly Ser Asn Pro Gly Gly Gly	
450 455 460	
tcc agc tcc acc acc acc acc acc aga ccc gcc acc tcc acc acc tcc	1440
Ser Ser Ser Thr Thr Thr Thr Thr Arg Pro Ala Thr Ser Thr Thr Ser	
465 470 475 480	
tcg gcc agc tcc ggc ccg acc ggc ggt ggc acg gct gcc cac tgg ggc	1488
Ser Ala Ser Ser Gly Pro Thr Gly Gly Gly Thr Ala Ala His Trp Gly	
485 490 495	
cag tgc ggc ggc atc ggc tgg acc ggc ccg acc gtc tgc gcc tcg ccc	1536
Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro Thr Val Cys Ala Ser Pro	
500 505 510	
tac acc tgc cag aag ctg aac gac tgg tac tac cag tgc ctc taa	1581
Tyr Thr Cys Gln Lys Leu Asn Asp Trp Tyr Tyr Gln Cys Leu	
515 520 525	

<210> 2

<211> 526

<212> PRT

<213> Acremonium thermophilum

<400> 2

Met His Ala Lys phe Ala Thr Leu Ala Ala Leu Val Ala Ser Ala Ala	
1 5 10 15	
Ala Gln Gln Ala Cys Thr Leu Thr Ala Glu Asn His Pro Thr Leu Ser	
20 25 30	
Trp Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Ser Gly Ser	
35 40 45	
Val Thr Ile Asp Ala Asn Trp Arg Trp Thr His Gln Val Ser Ser Ser	
50 55 60	

Thr Asn Cys Tyr Thr Gly Asn Glu Trp Asp Thr Ser Ile Cys Thr Asp
 65 70 75 80
 Gly Ala Ser Cys Ala Ala Ala Cys Cys Leu Asp Gly Ala Asp Tyr Ser
 85 90 95
 Gly Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ala Leu Ser Leu Gln Phe
 100 105 110
 Val Thr Gln Gly Pro Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu
 115 120 125
 Met Ala Ser Asp Thr Lys Tyr Gln Met Phe Thr Leu Leu Gly Asn Glu
 130 135 140
 Phe Thr Phe Asp Val Asp Val Thr Gly Leu Gly Cys Gly Leu Asn Gly
 145 150 155 160
 Ala Leu Tyr Phe Val Ser Met Asp Glu Asp Gly Gly Leu Ser Lys Tyr
 165 170 175
 Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser
 180 185 190
 Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Asn Val
 195 200 205
 Gly Trp Thr Pro Ser Ser Asn Asp Lys Asn Ala Gly Leu Gly Asn Tyr
 210 215 220
 Gly Ser Cys Cys Ser Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser
 225 230 235 240
 Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Thr Arg Cys
 245 250 255
 Glu Gly Asp Asp Cys Gly Gly Thr Tyr Ser Thr Asp Arg Tyr Ala Gly
 260 265 270
 Glu Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asn
 275 280 285
 Thr Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Ser Lys Lys Phe
 290 295 300
 Thr Val Val Thr Gln Phe Leu Thr Asp Ser Ser Gly Asn Leu Ser Glu
 305 310 315 320
 Ile Lys Arg Phe Tyr Val Gln Asn Gly Val Val Ile Pro Asn Ser Asn
 325 330 335
 Ser Asn Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Gln Ala Phe Cys

Thr Asn Cys Tyr Thr Gly Asn Glu Trp Asp Thr Ser Ile Cys Thr Asp
 65 70 75 80
 Gly Ala Ser Cys Ala Ala Ala Cys Cys Leu Asp Gly Ala Asp Tyr Ser
 85 90
 Gly Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ala Leu Ser Leu Gln Phe
 100 105 110
 Val Thr Gln Gly Pro Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu
 115 120 125
 Met Ala Ser Asp Thr Lys Tyr Gln Met Phe Thr Leu Leu Gly Asn Glu
 130 135 140
 Phe Thr Phe Asp Val Asp Val Thr Gly Leu Gly Cys Gly Leu Asn Gly
 145 150 155 160
 Ala Leu Tyr Phe Val Ser Met Asp Glu Asp Gly Gly Leu Ser Lys Tyr
 165 170 175
 Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser
 180 185 190
 Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Asn Val
 195 200 205
 Gly Trp Thr Pro Ser Ser Asn Asp Lys Asn Ala Gly Leu Gly Asn Tyr
 210 215 220
 Gly Ser Cys Cys Ser Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser
 225 230 235 240
 Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Thr Arg Cys
 245 250 255
 Glu Gly Asp Asp Cys Gly Gly Thr Tyr Ser Thr Asp Arg Tyr Ala Gly
 260 265 270
 Glu Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asn
 275 280 285
 Thr Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Ser Lys Lys Phe
 290 295 300
 Thr Val Val Thr Gln Phe Leu Thr Asp Ser Ser Gly Asn Leu Ser Glu
 305 310 315 320
 Ile Lys Arg Phe Tyr Val Gln Asn Gly Val Val Ile Pro Asn Ser Asn
 325 330 335
 Ser Asn Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Gln Ala Phe Cys

	340		345		350										
Asp	Ala	Gln	Lys	Thr	Ala	Phe	Gly	Asp	Thr	Asn	Val	Phe	Asp	Gln	Lys
		355					360					365			
Gly	Gly	Leu	Ala	Gln	Met	Gly	Lys	Ala	Leu	Ala	Gln	Pro	Met	Val	Leu
	370					375					380				
Val	Met	Ser	Leu	Trp	Asp	Asp	His	Ala	Val	Asn	Met	Leu	Trp	Leu	Asp
385					390					395					400
Ser	Thr	Tyr	Pro	Thr	Asn	Ala	Ala	Gly	Lys	Pro	Gly	Ala	Ala	Arg	Gly
				405					410					415	
Thr	Cys	Pro	Thr	Thr	Ser	Gly	Val	Pro	Ala	Asp	Val	Glu	Ser	Gln	Ala
			420					425					430		
Pro	Asn	Ser	Lys	Val	Ile	Tyr	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	Gly
		435					440					445			
Ser	Thr	Val	Ser	Gly	Leu	Pro	Gly	Gly	Gly	Ser	Asn	Pro	Gly	Gly	Gly
	450					455					460				
Ser	Ser	Ser	Thr	Thr	Thr	Thr	Thr	Arg	Pro	Ala	Thr	Ser	Thr	Thr	Ser
465					470					475					480
Ser	Ala	Ser	Ser	Gly	Pro	Thr	Gly	Gly	Gly	Thr	Ala	Ala	His	Trp	Gly
				485					490					495	
Gln	Cys	Gly	Gly	Ile	Gly	Trp	Thr	Gly	Pro	Thr	Val	Cys	Ala	Ser	Pro
			500					505					510		
Tyr	Thr	Cys	Gln	Lys	Leu	Asn	Asp	Trp	Tyr	Tyr	Gln	Cys	Leu		
		515					520					525			

<210> 3
 <211> 1590
 <212> DNA
 <213> Chaetomium thermophilum

<220>
 <221> CDS
 <222> (1)..(1590)
 <223>

<400> 3

atg atg tac aag aag ttc gcc gct ctc gcc gcc ctc gtg gct ggc gcc

48

Met 1	Met	Tyr	Lys	Lys 5	Phe	Ala	Ala	Leu	Ala 10	Ala	Leu	Val	Ala	Gly 15	Ala		
gcc Ala	gcc Ala	cag Gln	cag Gln 20	gct Ala	tgc Cys	tcc Ser	ctc Leu	acc Thr 25	act Thr	gag Glu	acc Thr	cac His	ccc Pro 30	aga Arg	ctc Leu		96
act Thr	tgg Trp	aag Lys 35	cgc Arg	tgc Cys	acc Thr	tct Ser	ggc Gly 40	ggc Gly	aac Asn	tgc Cys	tgc Ser	acc Thr 45	gtg Val	aac Asn	ggc Gly		144
gcc Ala	gtc Val 50	acc Thr	atc Ile	gat Asp	gcc Ala	aac Asn 55	tgg Trp	cgc Arg	tgg Trp	act Thr	cac His 60	acc Thr	gtt Val	tcc Ser	ggc Gly		192
tgc Ser 65	acc Thr	aac Asn	tgc Cys	tac Tyr	acc Thr 70	ggc Gly	aac Asn	gag Glu	tgg Trp	gat Asp 75	acc Thr	tcc Ser	atc Ile	tgc Cys	tct Ser 80		240
gat Asp	ggc Gly	aag Lys	agc Ser	tgc Cys 85	gcc Ala	cag Gln	acc Thr	tgc Cys	tgc Cys 90	gtc Val	gac Asp	ggc Gly	gct Ala	gac Asp 95	tac Tyr		288
tct Ser	tgc Ser	acc Thr	tat Tyr 100	ggt Gly	atc Ile	acc Thr	acc Thr	agc Ser 105	ggt Gly	gac Asp	tcc Ser	ctg Leu	aac Asn 110	ctc Leu	aag Lys		336
ttc Phe	gtc Val	acc Thr 115	aag Lys	cac His	cag Gln	tac Tyr	ggc Gly 120	acc Thr	aat Asn	gtc Val	ggc Gly	tct Ser 125	cgt Arg	gtc Val	tac Tyr		384
ctg Leu	atg Met 130	gag Glu	aac Asn	gac Asp	acc Thr	aag Lys 135	tac Tyr	cag Gln	atg Met	ttc Phe	gag Glu 140	ctc Leu	ctc Leu	ggc Gly	aac Asn		432
gag Glu 145	ttc Phe	acc Thr	ttc Phe	gat Asp	gtc Val 150	gat Asp	gtc Val	tct Ser	aac Asn	ctg Leu 155	ggc Gly	tgc Cys	ggt Gly	ctc Leu	aac Asn 160		480
ggt Gly	gcc Ala	ctc Leu	tac Tyr	ttc Phe 165	gtc Val	tcc Ser	atg Met	gac Asp	gct Ala 170	gat Asp	ggt Gly	ggt Gly	atg Met	agc Ser 175	aag Lys		528
tac Tyr	tct Ser	ggc Gly	aac Asn 180	aag Lys	gct Ala	ggc Gly	gcc Ala	aag Lys 185	tac Tyr	ggg Gly	acg Thr	ggg Gly	tac Tyr 190	tgt Cys	gat Asp		576
gct Ala	cag Gln	tgc Cys 195	cgc Pro	cgc Arg	gac Asp	ctt Leu	aag Lys 200	ttc Phe	atc Ile	aac Asn	ggc Gly	gag Glu 205	gcc Ala	aac Asn	att Ile		624
gag Glu	aac Asn 210	tgg Trp	acc Thr	cct Pro	tgc Ser	acc Thr 215	aat Asn	gat Asp	gcc Ala	aac Asn	gcc Ala 220	ggt Gly	ttc Phe	ggc Gly	cgc Arg		672
tat Tyr 225	ggc Gly	agc Ser	tgc Cys	tgc Cys	tct Ser 230	gag Glu	atg Met	gat Asp	atc Ile	tgg Trp 235	gag Glu	gcc Ala	aac Asn	aac Asn	atg Met 240		720
gct Ala	act Thr	gcc Ala	ttc Phe	act Thr 245	cct Pro	cac His	cct Pro	tgc Cys	acc Thr 250	att Ile	atc Ile	ggc Gly	cag Gln	agc Ser 255	cgc Arg		768
tgc Cys	gag Glu	ggc Gly	aac Asn 260	agc Ser	tgc Cys	ggt Gly	ggc Gly	acc Thr 265	tac Tyr	agc Ser	tct Ser	gag Glu	cgc Arg 270	tat Tyr	gct Ala		816
ggt Gly	gtt Val	tgc Cys	gat Asp	cct Pro	gat Asp	ggc Gly	tgc Cys	gac Asp	ttc Phe	aac Asn	gcc Ala	tac Tyr	cgc Arg	cag Gln	ggc Gly		864

Met 1	Met	Tyr	Lys	Lys 5	Phe	Ala	Ala	Leu	Ala 10	Ala	Leu	Val	Ala	Gly 15	Ala		
gcc Ala	gcc Ala	cag Gln	cag Gln 20	gct Ala	tgc Cys	tcc Ser	ctc Leu	acc Thr 25	act Thr	gag Glu	acc Thr	cac His	ccc Pro 30	aga Arg	ctc Leu		96
act Thr	tgg Trp	aag Lys 35	cgc Arg	tgc Cys	acc Thr	tct Ser	ggc Gly 40	ggc Gly	aac Asn	tgc Cys	tgc Ser	acc Thr 45	gtg Val	aac Asn	ggc Gly		144
gcc Ala	gtc Val 50	acc Thr	atc Ile	gat Asp	gcc Ala	aac Asn 55	tgg Trp	cgc Arg	tgg Trp	act Thr	cac His 60	acc Thr	gtt Val	tcc Ser	ggc Gly		192
tgc Ser 65	acc Thr	aac Asn	tgc Cys	tac Tyr	acc Thr 70	ggc Gly	aac Asn	gag Glu	tgg Trp	gat Asp 75	acc Thr	tcc Ser	atc Ile	tgc Cys	tct Ser 80		240
gat Asp	ggc Gly	aag Lys	agc Ser	tgc Cys 85	gcc Ala	cag Gln	acc Thr	tgc Cys	tgc Cys 90	gtc Val	gac Asp	ggc Gly	gct Ala	gac Asp 95	tac Tyr		288
tct Ser	tgc Ser	acc Thr	tat Tyr 100	ggt Gly	atc Ile	acc Thr	acc Thr	agc Ser 105	ggt Gly	gac Asp	tcc Ser	ctg Leu	aac Asn 110	ctc Leu	aag Lys		336
ttc Phe	gtc Val	acc Thr 115	aag Lys	cac His	cag Gln	tac Tyr	ggc Gly 120	acc Thr	aat Asn	gtc Val	ggc Gly	tct Ser 125	cgt Arg	gtc Val	tac Tyr		384
ctg Leu	atg Met 130	gag Glu	aac Asn	gac Asp	acc Thr	aag Lys 135	tac Tyr	cag Gln	atg Met	ttc Phe	gag Glu 140	ctc Leu	ctc Leu	ggc Gly	aac Asn		432
gag Glu 145	ttc Phe	acc Thr	ttc Phe	gat Asp	gtc Val 150	gat Asp	gtc Val	tct Ser	aac Asn	ctg Leu 155	ggc Gly	tgc Cys	ggt Gly	ctc Leu	aac Asn 160		480
ggt Gly	gcc Ala	ctc Leu	tac Tyr	ttc Phe 165	gtc Val	tcc Ser	atg Met	gac Asp	gct Ala 170	gat Asp	ggt Gly	ggt Gly	atg Met	agc Ser 175	aag Lys		528
tac Tyr	tct Ser	ggc Gly	aac Asn 180	aag Lys	gct Ala	ggc Gly	gcc Ala	aag Lys 185	tac Tyr	ggg Gly	acg Thr	ggg Gly	tac Tyr 190	tgt Cys	gat Asp		576
gct Ala	cag Gln	tgc Cys 195	cgc Pro	cgc Arg	gac Asp	ctt Leu	aag Lys 200	ttc Phe	atc Ile	aac Asn	ggc Gly	gag Glu 205	gcc Ala	aac Asn	att Ile		624
gag Glu	aac Asn 210	tgg Trp	acc Thr	cct Pro	tgc Ser	acc Thr 215	aat Asn	gat Asp	gcc Ala	aac Asn	gcc Ala 220	ggt Gly	ttc Phe	ggc Gly	cgc Arg		672
tat Tyr 225	ggc Gly	agc Ser	tgc Cys	tgc Cys	tct Ser 230	gag Glu	atg Met	gat Asp	atc Ile	tgg Trp 235	gag Glu	gcc Ala	aac Asn	aac Asn	atg Met 240		720
gct Ala	act Thr	gcc Ala	ttc Phe	act Thr 245	cct Pro	cac His	cct Pro	tgc Cys	acc Thr 250	att Ile	atc Ile	ggc Gly	cag Gln	agc Ser 255	cgc Arg		768
tgc Cys	gag Glu	ggc Gly	aac Asn 260	agc Ser	tgc Cys	ggt Gly	ggc Gly	acc Thr 265	tac Tyr	agc Ser	tct Ser	gag Glu	cgc Arg 270	tat Tyr	gct Ala		816
ggt Gly	gtt Val	tgc Cys	gat Asp	cct Pro	gat Asp	ggc Gly	tgc Cys	gac Asp	ttc Phe	aac Asn	gcc Ala	tac Tyr	cgc Arg	cag Gln	ggc Gly		864

Met 1	Met	Tyr	Lys	Lys 5	Phe	Ala	Ala	Leu	Ala 10	Ala	Leu	Val	Ala	Gly 15	Ala		
gcc Ala	gcc Ala	cag Gln	cag Gln 20	gct Ala	tgc Cys	tcc Ser	ctc Leu	acc Thr 25	act Thr	gag Glu	acc Thr	cac His	ccc Pro 30	aga Arg	ctc Leu		96
act Thr	tgg Trp	aag Lys 35	cgc Arg	tgc Cys	acc Thr	tct Ser	ggc Gly 40	ggc Gly	aac Asn	tgc Cys	tgc Ser	acc Thr 45	gtg Val	aac Asn	ggc Gly		144
gcc Ala	gtc Val 50	acc Thr	atc Ile	gat Asp	gcc Ala	aac Asn 55	tgg Trp	cgc Arg	tgg Trp	act Thr	cac His 60	acc Thr	gtt Val	tcc Ser	ggc Gly		192
tgc Ser 65	acc Thr	aac Asn	tgc Cys	tac Tyr	acc Thr 70	ggc Gly	aac Asn	gag Glu	tgg Trp	gat Asp 75	acc Thr	tcc Ser	atc Ile	tgc Cys	tct Ser 80		240
gat Asp	ggc Gly	aag Lys	agc Ser	tgc Cys 85	gcc Ala	cag Gln	acc Thr	tgc Cys	tgc Cys 90	gtc Val	gac Asp	ggc Gly	gct Ala	gac Asp 95	tac Tyr		288
tct Ser	tgc Ser	acc Thr	tat Tyr 100	ggt Gly	atc Ile	acc Thr	acc Thr	agc Ser 105	ggt Gly	gac Asp	tcc Ser	ctg Leu	aac Asn 110	ctc Leu	aag Lys		336
ttc Phe	gtc Val	acc Thr 115	aag Lys	cac His	cag Gln	tac Tyr	ggc Gly 120	acc Thr	aat Asn	gtc Val	ggc Gly	tct Ser 125	cgt Arg	gtc Val	tac Tyr		384
ctg Leu	atg Met 130	gag Glu	aac Asn	gac Asp	acc Thr	aag Lys 135	tac Tyr	cag Gln	atg Met	ttc Phe	gag Glu 140	ctc Leu	ctc Leu	ggc Gly	aac Asn		432
gag Glu 145	ttc Phe	acc Thr	ttc Phe	gat Asp	gtc Val 150	gat Asp	gtc Val	tct Ser	aac Asn	ctg Leu 155	ggc Gly	tgc Cys	ggt Gly	ctc Leu	aac Asn 160		480
ggt Gly	gcc Ala	ctc Leu	tac Tyr	ttc Phe 165	gtc Val	tcc Ser	atg Met	gac Asp	gct Ala 170	gat Asp	ggt Gly	ggt Gly	atg Met	agc Ser 175	aag Lys		528
tac Tyr	tct Ser	ggc Gly	aac Asn 180	aag Lys	gct Ala	ggc Gly	gcc Ala	aag Lys 185	tac Tyr	ggg Gly	acg Thr	ggg Gly	tac Tyr 190	tgt Cys	gat Asp		576
gct Ala	cag Gln	tgc Cys 195	cgc Pro	cgc Arg	gac Asp	ctt Leu	aag Lys 200	ttc Phe	atc Ile	aac Asn	ggc Gly	gag Glu 205	gcc Ala	aac Asn	att Ile		624
gag Glu	aac Asn 210	tgg Trp	acc Thr	cct Pro	tgc Ser	acc Thr 215	aat Asn	gat Asp	gcc Ala	aac Asn	gcc Ala 220	ggt Gly	ttc Phe	ggc Gly	cgc Arg		672
tat Tyr 225	ggc Gly	agc Ser	tgc Cys	tgc Cys	tct Ser 230	gag Glu	atg Met	gat Asp	atc Ile	tgg Trp 235	gag Glu	gcc Ala	aac Asn	aac Asn	atg Met 240		720
gct Ala	act Thr	gcc Ala	ttc Phe	act Thr 245	cct Pro	cac His	cct Pro	tgc Cys	acc Thr 250	att Ile	atc Ile	ggc Gly	cag Gln	agc Ser 255	cgc Arg		768
tgc Cys	gag Glu	ggc Gly	aac Asn 260	agc Ser	tgc Cys	ggt Gly	ggc Gly	acc Thr 265	tac Tyr	agc Ser	tct Ser	gag Glu	cgc Arg 270	tat Tyr	gct Ala		816
ggt Gly	gtt Val	tgc Cys	gat Asp	cct Pro	gat Asp	ggc Gly	tgc Cys	gac Asp	ttc Phe	aac Asn	gcc Ala	tac Tyr	cgc Arg	cag Gln	ggc Gly		864

275					280					285									
gac Asp	aag Lys 290	acc Thr	ttc Phe	tac Tyr	ggc Gly	aag Lys 295	ggc Gly	atg Met	acc Thr	gtc Val	gac Asp 300	acc Thr	acc Thr	aag Lys	aag Lys	912			
atg Met 305	acc Thr	gtc Val	gtc Val	acc Thr	cag Gln 310	ttc Phe	cac His	aag Lys	aac Asn	tcg Ser 315	gct Ala	ggc Gly	gtc Val	ctc Leu	agc Ser 320	960			
gag Glu	atc Ile	aag Lys	cgc Arg	ttc Phe 325	tac Tyr	gtt Val	cag Gln	gac Asp	ggc Gly 330	aag Lys	gtc Val	att Ile	gcc Ala	aac Asn 335	gcc Ala	1008			
gag Glu	tcc Ser	aag Lys	atc Ile 340	ccc Pro	ggc Gly	aac Asn	ccc Pro	ggc Gly 345	aac Asn	tcc Ser	atc Ile	acc Thr	cag Gln 350	gag Glu	tgg Trp	1056			
tgc Cys	gat Asp	gcc Ala 355	cag Gln	aag Lys	gtc Val	gcc Ala	ttc Phe 360	ggt Gly	gac Asp	atc Ile	gat Asp	gac Asp 365	ttc Phe	aac Asn	cgc Arg	1104			
aag Lys	ggc Gly 370	ggt Gly	atg Met	gct Ala	cag Gln	atg Met 375	agc Ser	aag Lys	gcc Ala	ctc Leu	gaa Glu 380	ggc Gly	cct Pro	atg Met	gtc Val	1152			
ctg Leu 385	gtc Val	atg Met	tcc Ser	gtc Val	tgg Trp 390	gat Asp	gac Asp	cac His	tac Tyr	gcc Ala 395	aac Asn	atg Met	ctc Leu	tgg Trp	ctc Leu 400	1200			
gac Asp	tcg Ser	acc Thr	tac Tyr	ccc Pro 405	atc Ile	gac Asp	aag Lys	gcc Ala	ggc Gly 410	acc Thr	ccc Pro	ggc Gly	gcc Ala	gag Glu 415	cgc Arg	1248			
ggt Gly	gct Ala	tgc Cys	ccg Pro 420	acc Thr	acc Thr	tcc Ser	ggt Gly	gtc Val 425	cct Pro	gcc Ala	gag Glu	att Ile	gag Glu 430	gcc Ala	cag Gln	1296			
gtc Val	ccc Pro	aac Asn 435	agc Ser	aac Asn	gtc Val	atc Ile	ttc Phe 440	tcc Ser	aac Asn	atc Ile	cgc Arg	ttc Phe 445	ggc Gly	ccc Pro	atc Ile	1344			
ggc Gly	tcg Ser 450	acc Thr	gtc Val	cct Pro	ggc Gly	ctc Leu 455	gac Asp	ggc Gly	agc Ser	act Thr	ccc Pro 460	agc Ser	aac Asn	ccg Pro	acc Thr	1392			
gcc Ala 465	acc Thr	gtt Val	gct Ala	cct Pro	ccc Pro 470	act Thr	tct Ser	acc Thr	acc Thr	agc Ser 475	gtg Val	aga Arg	agc Ser	agc Ser	act Thr 480	1440			
act Thr	cag Gln	att Ile	tcc Ser	acc Thr 485	ccg Pro	act Thr	agc Ser	cag Gln	ccc Pro 490	ggc Gly	ggc Gly	tgc Cys	acc Thr	acc Thr 495	cag Gln	1488			
aag Lys	tgg Trp	ggc Gly	cag Gln 500	tgc Cys	ggt Gly	ggt Gly	atc Ile	ggc Gly 505	tac Tyr	acc Thr	ggc Gly	tgc Cys	act Thr 510	aac Asn	tgc Cys	1536			
gtt Val	gct Ala	ggc Gly 515	act Thr	acc Thr	tgc Cys	act Thr	gag Glu 520	ctc Leu	aac Asn	ccc Pro	tgg Trp	tac Tyr 525	agc Ser	cag Gln	tgc Cys	1584			
ctg Leu	taa															1590			

5

<210> 4

<211> 529

<212> PRT

<213> Chaetomium thermophilum

<400> 4

```

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

```

Ala Thr Ala Phe Thr 245 Pro His Pro Cys Thr 250 Ile Ile Gly Gln Ser 255 Arg
Cys Glu Gly Asn 260 Ser Cys Gly Gly Thr 265 Tyr Ser Ser Glu Arg 270 Tyr Ala
Gly Val Cys 275 Asp Pro Asp Gly Cys 280 Asp Phe Asn Ala Tyr 285 Arg Gln Gly
Asp Lys 290 Thr Phe Tyr Gly Lys 295 Gly Met Thr Val Asp 300 Thr Thr Lys Lys
Met Thr Val Val Thr 310 Gln Phe His Lys Asn 315 Ser Ala Gly Val Leu Ser 320
Glu Ile Lys Arg Phe 325 Tyr Val Gln Asp Gly 330 Lys Val Ile Ala Asn 335 Ala
Glu Ser Lys Ile 340 Pro Gly Asn Pro Gly 345 Asn Ser Ile Thr Gln Glu Trp
Cys Asp Ala 355 Gln Lys Val Ala Phe 360 Gly Asp Ile Asp 365 Asp Phe Asn Arg
Lys Gly 370 Gly Met Ala Gln Met 375 Ser Lys Ala Leu Glu 380 Gly Pro Met Val
Leu Val Met Ser Val Trp 390 Asp Asp His Tyr Ala 395 Asn Met Leu Trp Leu 400
Asp Ser Thr Tyr Pro 405 Ile Asp Lys Ala Gly 410 Thr Pro Gly Ala Glu Arg 415
Gly Ala Cys Pro 420 Thr Thr Ser Gly Val 425 Pro Ala Glu Ile Glu Ala Gln 430
Val Pro Asn 435 Ser Asn Val Ile Phe 440 Ser Asn Ile Arg Phe 445 Gly Pro Ile
Gly Ser Thr Val Pro Gly 455 Leu Asp Gly Ser Thr Pro 460 Ser Asn Pro Thr
Ala Thr Val Ala Pro 470 Pro Thr Ser Thr Thr Ser 475 Val Arg Ser Ser Thr 480
Thr Gln Ile Ser Thr 485 Pro Thr Ser Gln Pro 490 Gly Gly Cys Thr Thr Gln 495
Lys Trp Gly Gln 500 Cys Gly Gly Ile Gly 505 Tyr Thr Gly Cys Thr 510 Asn Cys

Val Ala Gly Thr Thr Cys Thr Glu Leu Asn Pro Trp Tyr Ser Gln Cys
 515 520 525

Leu

<210> 5
 <211> 1356
 <212> DNA
 <213> Scytalidium sp.

<220>
 <221> CDS
 <222> (1)..(1356)
 <223>
 <400>5

atg cag atc aag agc tac atc cag tac ctg gcc gcg gct ctg ccg ctc	48
Met Gln Ile Lys Ser Tyr Ile Gln Tyr Leu Ala Ala Ala Leu Pro Leu	
1 5 10 15	
ctg agc agc gtc gct gcc cag cag gcc ggc acc atc acc gcc gag aac	96
Leu Ser Ser Val Ala Ala Gln Gln Ala Gly Thr Ile Thr Ala Glu Asn	
20 25 30	
cac ccc agg atg acc tgg aag agg tgc tgc ggc ccc ggc aac tgc cag	144
His Pro Arg Met Thr Trp Lys Arg Cys Ser Gly Pro Gly Asn Cys Gln	
35 40 45	
acc gtg cag ggc gag gtc gtc atc gac gcc aac tgg cgc tgg ctg cac	192
Thr Val Gln Gly Glu Val Val Ile Asp Ala Asn Trp Arg Trp Leu His	
50 55 60	
aac aac ggc cag aac tgc tat gag ggc aac aag tgg acc agc cag tgc	240
Asn Asn Gly Gln Asn Cys Tyr Glu Gly Asn Lys Trp Thr Ser Gln Cys	
65 70 75 80	
agc tgc gcc acc gac tgc gcg cag agg tgc gcc ctc gac ggt gcc aac	288
Ser Ser Ala Thr Asp Cys Ala Gln Arg Cys Ala Leu Asp Gly Ala Asn	
85 90 95	
tac cag tgc acc tac ggc gcc tgc acc agc ggc gac tcc ctg acg ctc	336
Tyr Gln Ser Thr Tyr Gly Ala Ser Thr Ser Gly Asp Ser Leu Thr Leu	
100 105 110	
aag ttc gtc acc aag cac gag tac ggc acc aac atc ggc tgc cgc ttc	384
Lys Phe Val Thr Lys His Glu Tyr Gly Thr Asn Ile Gly Ser Arg Phe	
115 120 125	
tac ctc atg gcc aac cag aac aag tac cag atg ttc acc ctg atg aac	432
Tyr Leu Met Ala Asn Gln Asn Lys Tyr Gln Met Phe Thr Leu Met Asn	
130 135 140	
aac gag ttc gcc ttc gat gtc gac ctc tcc aag gtt gag tgc ggt atc	480
Asn Glu Phe Ala Phe Asp Val Asp Leu Ser Lys Val Glu Cys Gly Ile	
145 150 155 160	

aac Asn	agc Ser	gct Ala	ctg Leu	tac Tyr 165	ttc Phe	gtc Val	gcc Ala	atg Met	gag Glu 170	gag Glu	gat Asp	ggt Gly	ggc Gly	atg Met 175	gcc Ala	528
agc Ser	tac Tyr	ccg Pro	agc Ser 180	aac Asn	cgt Arg	gct Ala	ggt Gly	ggc Ala 185	aag Lys	tac Tyr	ggc Gly	acg Thr	ggc Gly 190	tac Tyr	tgc Cys	576
gat Asp	gcc Ala	caa Gln 195	tgc Cys	gcc Ala	cgt Arg	gac Asp	ctc Leu 200	aag Lys	ttc Phe	att Ile	ggc Gly	ggc Gly 205	aag Lys	gcc Ala	aac Asn	624
att Ile	gag Glu 210	ggc Gly	tgg Trp	cgc Arg	ccg Pro	tcc Ser 215	acc Thr	aac Asn	gac Asp	ccc Pro	aac Asn 220	ggc Ala	ggt Gly	gtc Val	ggt Gly	672
ccc Pro 225	atg Met	ggt Gly	gcc Ala	tgc Cys	tgc Cys 230	gct Ala	gag Glu	atc Ile	gac Asp	gtt Val 235	tgg Trp	gag Glu	tcc Ser	aac Asn	gcc Ala 240	720
tat Tyr	gct Ala	tat Tyr	gcc Ala	ttc Phe 245	acc Thr	ccc Pro	cac His	gcc Ala	tgc Cys 250	ggc Gly	agc Ser	aag Lys	aac Asn	cgc Arg 255	tac Tyr	768
cac His	atc Ile	tgc Cys	gag Glu 260	acc Thr	aac Asn	aac Asn	tgc Cys	ggt Gly 265	ggt Gly	acc Thr	tac Tyr	tgc Ser	gat Asp 270	gac Asp	cgc Arg	816
ttc Phe	gcc Ala	ggc Gly 275	tac Tyr	tgc Cys	gac Asp	gcc Ala	aac Asn 280	ggc Gly	tgc Cys	gac Asp	tac Tyr	aac Asn 285	ccc Pro	tac Tyr	cgc Arg	864
atg Met	ggc Gly 290	aac Asn	aag Lys	gac Asp	ttc Phe 295	ggc Gly	aag Lys	ggc Gly	aag Lys	acc Thr 300	gtc Val	gac Asp	acc Thr	aac Asn		912
cgc Arg 305	aag Lys	ttc Phe	acc Thr	gtt Val 310	gtc Val	tcc Ser	cgc Arg	ttc Phe	gag Glu 315	cgt Arg 315	aac Asn	agg Arg	ctc Leu	tct Ser	cag Gln 320	960
ttc Phe	ttc Phe	gtc Val	cag Gln 325	gac Asp	ggc Gly	cgc Arg	aag Lys	atc Ile	gag Glu 330	gtg Val	ccc Pro	cct Pro	ccg Pro	acc Thr 335	tgg Trp	1008
ccc Pro	ggc Gly	ctc Leu	ccg Pro 340	aac Asn	agc Ser	gcc Ala	gac Asp	atc Ile 345	acc Thr	cct Pro	gag Glu	ctc Leu	tgc Cys 350	gat Asp	gct Ala	1056
cag Gln	ttc Phe	cgc Arg 355	gtc Val	ttc Phe	gat Asp	gac Asp	cgc Arg 360	aac Asn	cgc Arg	ttc Phe	ggc Ala	gag Glu 365	acc Thr	ggt Gly	ggc Gly	1104
ttc Phe	gat Asp 370	gct Ala	ctg Leu	aac Asn	gag Glu	gcc Ala 375	ctc Leu	acc Thr	att Ile	ccc Pro	atg Met 380	gtc Val	ctt Leu	gtc Val	atg Met	1152
tcc Ser 385	atc Ile	tgg Trp	gat Asp	gac Asp	cac His 390	cac His	tcc Ser	aac Asn	atg Met	ctc Leu 395	tgg Trp	ctc Leu	gac Asp	tcc Ser	agc Ser 400	1200
tac Tyr	ccg Pro	ccc Pro	gag Glu	aag Lys 405	gcc Ala	ggc Gly	ctc Leu	ccc Pro	ggt Gly 410	ggc Gly	gac Asp	cgt Arg	ggc Gly 415	ccg Pro	tgc Cys	1248
ccg Pro	acc Thr	acc Thr	tct Ser 420	ggt Gly	gtc Val	cct Pro	gcc Ala	gag Glu 425	gtc Val	gag Glu	gct Ala	cag Gln	tac Tyr 430	ccc Pro	gat Asp	1296
gct Cag	gtc Gtc	gtc Gtc	tgg Tgg	tcc Tcc	aac Tcc	atc Tcc	cgc Tcc	ttc Tcc	ggc Tcc	ccc Tcc	atc Tcc	ggc Tcc	tgc Tcc	acc Tcc		1344

Ala Gln Val Val Trp Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser Thr
 435 440 445

gtc aac gtc taa
 Val Asn Val
 450

1356

- 5 <210> 6
- <211> 451
- <212> PRT
- <213> scytalidium sp.
- <212> 6
- 10 <400> 6

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

195	200	205
Ile Glu Gly Trp Arg Pro Ser Thr Asn Asp Pro Asn Ala Gly Val Gly	210	215 220
Pro Met Gly Ala Cys Cys Ala Glu Ile Asp Val Trp Glu Ser Asn Ala	225	230 235 240
Tyr Ala Tyr Ala Phe Thr Pro His Ala Cys Gly Ser Lys Asn Arg Tyr	245	250 255
His Ile Cys Glu Thr Asn Asn Cys Gly Gly Thr Tyr Ser Asp Asp Arg	260	265 270
Phe Ala Gly Tyr Cys Asp Ala Asn Gly Cys Asp Tyr Asn Pro Tyr Arg	275	280 285
Met Gly Asn Lys Asp Phe Tyr Gly Lys Gly Lys Thr Val Asp Thr Asn	290	295 300
Arg Lys Phe Thr Val Val Ser Arg Phe Glu Arg Asn Arg Leu Ser Gln	305	310 315 320
Phe Phe Val Gln Asp Gly Arg Lys Ile Glu Val Pro Pro Pro Thr Trp	325	330 335
Pro Gly Leu Pro Asn Ser Ala Asp Ile Thr Pro Glu Leu Cys Asp Ala	340	345 350
Gln Phe Arg Val Phe Asp Asp Arg Asn Arg Phe Ala Glu Thr Gly Gly	355	360 365
Phe Asp Ala Leu Asn Glu Ala Leu Thr Ile Pro Met Val Leu Val Met	370	375 380
Ser Ile Trp Asp Asp His His Ser Asn Met Leu Trp Leu Asp Ser Ser	385	390 395 400
Tyr Pro Pro Glu Lys Ala Gly Leu Pro Gly Gly Asp Arg Gly Pro Cys	405	410 415
Pro Thr Thr Ser Gly Val Pro Ala Glu Val Glu Ala Gln Tyr Pro Asp	420	425 430
Ala Gln Val Val Trp Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser Thr	435	440 445
Val Asn Val	450	

<210> 7
 <211> 1374
 <212> DNA
 <213> *Thermoascus aurantiacus*

<220>
 <221> CDS
 <222> (1)..(1374)
 <223>

<400>7

atg tat cag cgc gct ctt ctc ttc tct ttc ttc ctc tcc gcc gcc cgc	48
Met Tyr Gln Arg Ala Leu Leu Phe Ser Phe Phe Leu Ser Ala Ala Arg	
1 5 10 15	
gcg cag cag gcc ggt acc cta acc gca gag aat cac cct tcc ctg acc	96
Ala Gln Gln Ala Gly Thr Leu Thr Ala Glu Asn His Pro Ser Leu Thr	
20 25 30	
tgg cag caa tgc tcc agc ggc ggt agt tgt acc acg cag aat gga aaa	144
Trp Gln Gln Cys Ser Ser Gly Gly Ser Cys Thr Thr Gln Asn Gly Lys	
35 40 45	
gtc gtt atc gat gcg aac tgg cgt tgg gtc cat acc acc tct gga tac	192
Val Val Ile Asp Ala Asn Trp Arg Trp Val His Thr Thr Ser Gly Tyr	
50 55 60	
acc aac tgc tac acg ggc aat acg tgg gac acc agt atc tgt ccc gac	240
Thr Asn Cys Tyr Thr Gly Asn Thr Trp Asp Thr Ser Ile Cys Pro Asp	
65 70 75 80	
gac gtg acc tgc gct cag aat tgt gcc ttg gat gga gcg gat tac agt	288
Asp Val Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr Ser	
85 90 95	
ggc acc tat ggt gtt acg acc agt ggc aac gcc ctg aga ctg aac ttt	336
Gly Thr Tyr Gly Val Thr Thr Ser Gly Asn Ala Leu Arg Leu Asn Phe	
100 105 110	
gtc acc caa agc tca ggg aag aac att ggc tgc cgc ctg tac ctg ctg	384
Val Thr Gln Ser Ser Gly Lys Asn Ile Gly Ser Arg Leu Tyr Leu Leu	
115 120 125	
cag gac gac acc act tat cag atc ttc aag ctg ctg ggt cag gag ttt	432
Gln Asp Asp Thr Thr Tyr Gln Ile Phe Lys Leu Leu Gly Gln Glu Phe	
130 135 140	
acc ttc gat gtc gac gtc tcc aat ctc cct tgc ggg ctg aac ggc gcc	480
Thr Phe Asp Val Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala	
145 150 155 160	
ctc tac ttt gtg gcc atg gac gcc gac ggc gga ttg tcc aaa tac cct	528
Leu Tyr Phe Val Ala Met Asp Ala Asp Gly Gly Leu Ser Lys Tyr Pro	
165 170 175	
ggc aac aag gca ggc gct aag tat ggc act ggt tac tgc gac tct cag	576
Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln	
180 185 190	
tgc cct cgg gat ctc aag ttc atc aac ggt cag gcc aac gtt gaa ggc	624
Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly	

195					200					205						
tgg	cag	ccg	tct	gcc	aac	gac	cca	aat	gcc	ggc	gtt	ggt	aac	cac	ggt	672
Trp	Gln	Pro	Ser	Ala	Asn	Asp	Pro	Asn	Ala	Gly	Val	Gly	Asn	His	Gly	
	210					215					220					
tcc	tgc	tgc	gct	gag	atg	gat	gtc	tgg	gaa	gcc	aac	agc	atc	tct	act	720
Ser	Cys	Cys	Ala	Glu	Met	Asp	Val	Trp	Glu	Ala	Asn	Ser	Ile	Ser	Thr	
225					230				235						240	
gcg	gtg	acg	cct	cac	cca	tgc	gac	acc	ccc	ggc	cag	acc	atg	tgc	cag	768
Ala	Val	Thr	Pro	His	Pro	Cys	Asp	Thr	Pro	Gly	Gln	Thr	Met	Cys	Gln	
				245					250					255		
gga	gac	gac	tgt	ggt	gga	acc	tac	tcc	tcc	act	cga	tat	gct	ggt	acc	816
Gly	Asp	Asp	Cys	Gly	Gly	Thr	Tyr	Ser	Ser	Thr	Arg	Tyr	Ala	Gly	Thr	
			260					265					270			
tgc	gac	cct	gat	ggc	tgc	gac	ttc	aat	cct	tac	cgc	cag	ggc	aac	cac	864
Cys	Asp	Pro	Asp	Gly	Cys	Asp	Phe	Asn	Pro	Tyr	Arg	Gln	Gly	Asn	His	
		275					280					285				
tcg	ttc	tac	ggc	ccc	ggg	aag	atc	gtc	gac	act	agc	tcc	aaa	ttc	acc	912
Ser	Phe	Tyr	Gly	Pro	Gly	Lys	Ile	Val	Asp	Thr	Ser	Ser	Lys	Phe	Thr	
	290					295					300					
gtc	gtc	acc	cag	ttc	atc	acc	gac	gac	ggg	acc	ccc	tcc	ggc	acc	ctg	960
Val	Val	Thr	Gln	Phe	Ile	Thr	Asp	Asp	Gly	Thr	Pro	Ser	Gly	Thr	Leu	
305					310				315						320	
acg	gag	atc	aaa	cgc	ttc	tac	gtc	cag	aac	ggc	aag	gtg	atc	ccc	cag	1008
Thr	Glu	Ile	Lys	Arg	Phe	Tyr	Val	Gln	Asn	Gly	Lys	Val	Ile	Pro	Gln	
				325					330					335		
tcg	gag	tcg	acg	atc	agc	ggc	gtc	acc	ggc	aac	tca	atc	acc	acc	gag	1056
Ser	Glu	Ser	Thr	Ile	Ser	Gly	Val	Thr	Gly	Asn	Ser	Ile	Thr	Thr	Glu	
			340					345					350			
tat	tgc	acg	gcc	cag	aag	gcc	gcc	ttc	ggc	gac	aac	acc	ggc	ttc	ttc	1104
Tyr	Cys	Thr	Ala	Gln	Lys	Ala	Ala	Phe	Gly	Asp	Asn	Thr	Gly	Phe	Phe	
		355				360						365				
acg	cac	ggc	ggg	ctt	cag	aag	atc	agt	cag	gct	ctg	gct	cag	ggc	atg	1152
Thr	His	Gly	Gly	Leu	Gln	Lys	Ile	Ser	Gln	Ala	Leu	Ala	Gln	Gly	Met	
	370					375					380					
gtc	ctc	gtc	atg	agc	ctg	tgg	gac	gat	cac	gcc	gcc	aac	atg	ctc	tgg	1200
Val	Leu	Val	Met	Ser	Leu	Trp	Asp	Asp	His	Ala	Ala	Asn	Met	Leu	Trp	
385					390					395					400	
ctg	gac	agc	acc	tac	ccg	act	gat	gcg	gac	ccg	gac	acc	cct	ggc	gtc	1248
Leu	Asp	Ser	Thr	Tyr	Pro	Thr	Asp	Ala	Asp	Pro	Asp	Thr	Pro	Gly	Val	
				405					410					415		
gcg	cgc	ggt	acc	tgc	ccc	acg	acc	tcc	ggc	gtc	ccg	gcc	gac	gtt	gag	1296
Ala	Arg	Gly	Thr	Cys	Pro	Thr	Thr	Ser	Gly	Val	Pro	Ala	Asp	Val	Glu	
			420					425				430				
tcg	cag	aac	ccc	aat	tca	tat	gtt	atc	tac	tcc	aac	atc	aag	gtc	gga	1344
Ser	Gln	Asn	Pro	Asn	Ser	Tyr	Val	Ile	Tyr	Ser	Asn	Ile	Lys	Val	Gly	
		435					440					445				
ccc	atc	aac	tcg	acc	ttc	acc	gcc	aac	taa							1374
Pro	Ile	Asn	Ser	Thr	Phe	Thr	Ala	Asn								
	450					455										

<210> 8

<211> 457

<212> PRT

<213> *Thermoascus aurantiacus*

5

<400>8

Met Tyr Gln Arg Ala Leu Leu Phe Ser Phe Phe Leu Ser Ala Ala Arg
 1 5 10 15

Ala Gln Gln Ala Gly Thr Leu Thr Ala Glu Asn His Pro Ser Leu Thr
 20 25 30

Trp Gln Gln Cys Ser Ser Gly Gly Ser Cys Thr Thr Gln Asn Gly Lys
 35 40 45

Val Val Ile Asp Ala Asn Trp Arg Trp Val His Thr Thr Ser Gly Tyr
 50 55 60

Thr Asn Cys Tyr Thr Gly Asn Thr Trp Asp Thr Ser Ile Cys Pro Asp
 65 70 75 80

Asp Val Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr Ser
 85 90 95

Gly Thr Tyr Gly Val Thr Thr Ser Gly Asn Ala Leu Arg Leu Asn Phe
 100 105 110

Val Thr Gln Ser Ser Gly Lys Asn Ile Gly Ser Arg Leu Tyr Leu Leu
 115 120 125

Gln Asp Asp Thr Thr Tyr Gln Ile Phe Lys Leu Leu Gly Gln Glu Phe
 130 135 140

Thr Phe Asp Val Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala
 145 150 155 160

Leu Tyr Phe Val Ala Met Asp Ala Asp Gly Gly Leu Ser Lys Tyr Pro
 165 170 175

Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln
 180 185 190

Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly
 195 200 205

Trp Gln Pro Ser Ala Asn Asp Pro Asn Ala Gly Val Gly Asn His Gly
 210 215 220

Ser Cys Cys Ala Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser Thr
 225 230 235 240

Ala Val Thr Pro His Pro Cys Asp Thr Pro Gly Gln Thr Met Cys Gln
245 250 255
Gly Asp Asp Cys Gly Gly Thr Tyr Ser Ser Thr Arg Tyr Ala Gly Thr
260 265 270
Cys Asp Pro Asp Gly Cys Asp Phe Asn Pro Tyr Arg Gln Gly Asn His
275 280 285
Ser Phe Tyr Gly Pro Gly Lys Ile Val Asp Thr Ser Ser Lys Phe Thr
290 295 300
Val Val Thr Gln Phe Ile Thr Asp Asp Gly Thr Pro Ser Gly Thr Leu
305 310 315 320
Thr Glu Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Gln
325 330 335
Ser Glu Ser Thr Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Thr Glu
340 345 350
Tyr Cys Thr Ala Gln Lys Ala Ala Phe Gly Asp Asn Thr Gly Phe Phe
355 360 365
Thr His Gly Gly Leu Gln Lys Ile Ser Gln Ala Leu Ala Gln Gly Met
370 375 380
Val Leu Val Met Ser Leu Trp Asp Asp His Ala Ala Asn Met Leu Trp
385 390 395 400
Leu Asp Ser Thr Tyr Pro Thr Asp Ala Asp Pro Asp Thr Pro Gly Val
405 410 415
Ala Arg Gly Thr Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu
420 425 430
Ser Gln Asn Pro Asn Ser Tyr Val Ile Tyr Ser Asn Ile Lys Val Gly
435 440 445
Pro Ile Asn Ser Thr Phe Thr Ala Asn
450 455

5 <210> 9
<211> 1617
<212> DNA
<213> Thielavia australiensis

10 <220>
<221> CDS
<222> (1)..(1617)

<223>

<400> 9

atg tat gcc aag ttc gcg acc ctc gcc gcc ctc gtg gct ggc gcc tcc Met Tyr Ala Lys Phe Ala Thr Leu Ala Ala 10 Leu Val Ala Gly Ala Ser 1 5 15	48
gcc cag gcc gtc tgc agc ctt acc gct gag acg cac cct tcc ctg acg Ala Gln Ala Val Cys Ser Leu Thr Ala Glu Thr His Pro Ser Leu Thr 20 25 30	96
tgg cag aag tgc acg gcc ccc ggc agc tgc acc aac gtc gcc ggc tcc Trp Gln Lys Cys Thr Ala Pro Gly Ser Cys Thr Asn Val Ala Gly Ser 35 40 45	144
atc acc atc gac gcc aac tgg cgc tgg act cac cag acc tcg tcc gcg Ile Thr Ile Asp Ala Asn Trp Arg Trp Thr His Gln Thr Ser Ser Ala 50 55 60	192
acc aac tgc tac agc ggc agc aag tgg gac tcg tcc atc tgc acg acc Thr Asn Cys Tyr Ser Gly Ser Lys Trp Asp Ser Ser Ile Cys Thr Thr 65 70 75 80	240
ggc acc gac tgc gcc tcc aag tgc tgc att gat ggc gcc gag tac tcg Gly Thr Asp Cys Ala Ser Lys Cys Cys Ile Asp Gly Ala Glu Tyr Ser 85 90 95	288
agc acc tac ggc atc acc acc agc ggc aat gcc ctg aac ctc aag ttc Ser Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ala Leu Asn Leu Lys Phe 100 105 110	336
gtc acc aag ggc cag tac tcg acc aac att ggc tcg cgt acc tac ctc Val Thr Lys Gly Gln Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu 115 120 125	384
atg gag tcg gac acc aag tac cag atg ttc aag ctc ctt ggc aac gag Met Glu Ser Asp Thr Lys Tyr Gln Met Phe Lys Leu Leu Gly Asn Glu 130 135 140	432
ttc acc ttc gac gtc gat gtc tcc aac ctc ggc tgc ggc ctc aac ggc Phe Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn Gly 145 150 155 160	480
gcc ctg tac ttc gtc tcc atg gat gcc gac ggt ggc atg tcc aag tac Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Met Ser Lys Tyr 165 170 175	528
tcg ggc aac aag gcc ggt gcc aag tac ggt acc ggc tac tgc gat gct Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala 180 185 190	576
cag tgc ccc cgc gac ctc aag ttc atc aac ggc gag gcc aac gtt gag Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Val Glu 195 200 205	624
ggc tgg gag agc tcg acc aac gac gcc aac gcc ggc tcg ggc aag tac Gly Trp Glu Ser Ser Thr Asn Asp Ala Asn Ala Gly Ser Gly Lys Tyr 210 215 220	672
ggc agc tgc tgc acc gag atg gac gtc tgg gag gcc aac aac atg gcg Gly Ser Cys Cys Thr Glu Met Asp Val Trp Glu Ala Asn Asn Met Ala 225 230 235 240	720

act	gcc	ttc	act	cct	cac	cct	tgc	acc	acc	att	ggc	cag	act	cgc	tgc	768
Thr	Ala	Phe	Thr	Pro	His	Pro	Cys	Thr	Thr	Ile	Gly	Gln	Thr	Arg	Cys	
				245					250					255		
gag	ggc	gac	acc	tgc	ggc	ggc	acc	tac	agc	tca	gac	cgc	tac	gcc	ggc	816
Glu	Gly	Asp	Thr	Cys	Gly	Gly	Thr	Tyr	Ser	Ser	Asp	Arg	Tyr	Ala	Gly	
			260					265					270			
gtc	tgc	gac	ccc	gac	gga	tgc	gac	ttc	aac	tgc	tac	cgc	cag	ggc	aac	864
Val	Cys	Asp	Pro	Asp	Gly	Cys	Asp	Phe	Asn	Ser	Tyr	Arg	Gln	Gly	Asn	
		275					280					285				
aag	acc	ttc	tac	ggc	aag	ggc	atg	acc	gtc	gac	acc	acc	aag	aag	atc	912
Lys	Thr	Phe	Tyr	Gly	Lys	Gly	Met	Thr	Val	Asp	Thr	Thr	Lys	Lys	Ile	
	290					295					300					
acg	gtc	gtc	acc	cag	ttc	ctc	aag	aac	tgc	ggc	ggc	gag	ctc	tcc	gag	960
Thr	Val	Val	Thr	Gln	Phe	Leu	Lys	Asn	Ser	Ala	Gly	Glu	Leu	Ser	Glu	
305				310						315					320	
atc	aag	cgc	ttc	tac	ggc	cag	gac	ggc	aag	gtc	atc	ccg	aac	agt	gag	1008
Ile	Lys	Arg	Phe	Tyr	Ala	Gln	Asp	Gly	Lys	Val	Ile	Pro	Asn	Ser	Glu	
				325					330					335		
tct	acc	att	ggc	ggc	atc	ccc	ggc	aac	tcc	atc	acc	aag	ggc	tac	tgc	1056
Ser	Thr	Ile	Ala	Gly	Ile	Pro	Gly	Asn	Ser	Ile	Thr	Lys	Ala	Tyr	Cys	
			340					345					350			
gac	ggc	cag	aag	acc	gtc	ttc	cag	aac	acc	gac	gac	ttc	acc	ggc	aag	1104
Asp	Ala	Gln	Lys	Thr	Val	Phe	Gln	Asn	Thr	Asp	Asp	Phe	Thr	Ala	Lys	
		355					360					365				
ggc	ggc	ctc	gtc	cag	atg	ggc	aag	ggc	ctc	ggc	ggc	gac	atg	gtc	ctc	1152
Gly	Gly	Leu	Val	Gln	Met	Gly	Lys	Ala	Leu	Ala	Gly	Asp	Met	Val	Leu	
	370					375					380					
gtc	atg	tcc	gtc	tgg	gac	gac	cac	ggc	gtc	aac	atg	ctc	tgg	cta	gac	1200
Val	Met	Ser	Val	Trp	Asp	Asp	His	Ala	Val	Asn	Met	Leu	Trp	Leu	Asp	
385					390					395					400	
tgc	acc	tac	ccg	acc	gac	cag	gtc	ggc	gtt	ggc	ggc	gct	gag	cgc	ggc	1248
Ser	Thr	Tyr	Pro	Thr	Asp	Gln	Val	Gly	Val	Ala	Gly	Ala	Glu	Arg	Gly	
			405					410					415			
ggc	tgc	ccc	acc	acc	tgc	ggc	gtc	ccc	tgc	gat	gtt	gag	ggc	aac	ggc	1296
Ala	Cys	Pro	Thr	Thr	Ser	Gly	Val	Pro	Ser	Asp	Val	Glu	Ala	Asn	Ala	
			420					425					430			
ccc	aac	tcc	aac	gtc	atc	ttc	tcc	aac	atc	cgc	ttc	ggc	ccc	atc	ggc	1344
Pro	Asn	Ser	Asn	Val	Ile	Phe	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	Gly	
		435					440					445				
tcc	acc	gtc	cag	ggc	ctg	ccc	agc	tcc	ggc	ggc	acc	tcc	agc	agc	tgc	1392
Ser	Thr	Val	Gln	Gly	Leu	Pro	Ser	Ser	Gly	Gly	Thr	Ser	Ser	Ser	Ser	
	450					455					460					
agc	ggc	gct	ccc	cag	tgc	acc	agc	acc	aag	ggc	tgc	acc	acc	acc	tca	1440
Ser	Ala	Ala	Pro	Gln	Ser	Thr	Ser	Thr	Lys	Ala	Ser	Thr	Thr	Thr	Ser	
465					470					475					480	
gct	gtc	cgc	acc	acc	tgc	act	ggc	acc	acc	aag	acc	acc	tcc	tgc	gct	1488
Ala	Val	Arg	Thr	Thr	Ser	Thr	Ala	Thr	Thr	Lys	Thr	Thr	Ser	Ser	Ala	
				485					490					495		
ccc	ggc	cag	ggc	acc	aac	act	ggc	aag	cat	tgg	cag	caa	tgc	ggc	ggc	1536
Pro	Ala	Gln	Gly	Thr	Asn	Thr	Ala	Lys	His	Trp	Gln	Gln	Cys	Gly	Gly	
			500					505					510			
aac	ggc	tgg	acc	ggc	ccg	acg	gtg	tgc	gag	tct	ccc	tac	aag	tgc	acc	1584

<210> 10
 <211> 538
 <212> PRT
 <213> Thielavia australiensis

5

<400> 10
 Met Tyr Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Gly Ala Ser
 1 5 10 15
 Ala Gln Ala Val Cys Ser Leu Thr Ala Glu Thr His Pro Ser Leu Thr
 20 25 30
 Trp Gln Lys Cys Thr Ala Pro Gly Ser Cys Thr Asn Val Ala Gly Ser
 35 40 45
 Ile Thr Ile Asp Ala Asn Trp Arg Trp Thr His Gln Thr Ser Ser Ala
 50 55 60
 Thr Asn Cys Tyr Ser Gly Ser Lys Trp Asp Ser Ser Ile Cys Thr Thr
 65 70 75 80
 Gly Thr Asp Cys Ala Ser Lys Cys Cys Ile Asp Gly Ala Glu Tyr Ser
 85 90 95
 Ser Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ala Leu Asn Leu Lys Phe
 100 105 110
 Val Thr Lys Gly Gln Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu
 115 120 125
 Met Glu Ser Asp Thr Lys Tyr Gln Met Phe Lys Leu Leu Gly Asn Glu
 130 135 140
 Phe Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn Gly
 145 150 155 160
 Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Met Ser Lys Tyr
 165 170 175
 Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala
 180 185 190
 Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Val Glu

195	200	205
Gly Trp Glu Ser Ser Thr	Asn Asp Ala Asn Ala	Gly Ser Gly Lys Tyr
210	215	220
Gly Ser Cys Cys Thr	Glu Met Asp Val Trp	Glu Ala Asn Asn Met Ala
225	230	235
Thr Ala Phe Thr	Pro His Pro Cys Thr	Thr Ile Gly Gln Thr Arg Cys
	245	250
Glu Gly Asp Thr	Cys Gly Gly Thr Tyr	Ser Ser Asp Arg Tyr Ala Gly
	260	265
Val Cys Asp	Pro Asp Gly Cys Asp	Phe Asn Ser Tyr Arg Gln Gly Asn
	275	280
Lys Thr Phe Tyr Gly Lys	Gly Met Thr Val Asp	Thr Thr Lys Lys Ile
	295	300
Thr Val Val Thr Gln	Phe Leu Lys Asn Ser	Ala Gly Glu Leu Ser Glu
	310	315
Ile Lys Arg Phe	Tyr Ala Gln Asp Gly	Lys Val Ile Pro Asn Ser Glu
	325	330
Ser Thr Ile	Ala Gly Ile Pro Gly	Asn Ser Ile Thr Lys Ala Tyr Cys
	340	345
Asp Ala Gln Lys Thr Val	Phe Gln Asn Thr Asp Asp	Phe Thr Ala Lys
	355	360
Gly Gly Leu Val Gln Met	Gly Lys Ala Leu Ala	Gly Asp Met Val Leu
	375	380
Val Met Ser Val Trp	Asp Asp His Ala Val	Asn Met Leu Trp Leu Asp
	390	395
Ser Thr Tyr Pro	Thr Asp Gln Val Gly	Val Ala Gly Ala Glu Arg Gly
	405	410
Ala Cys Pro	Thr Thr Ser Gly Val	Pro Ser Asp Val Glu Ala Asn Ala
	420	425
Pro Asn Ser	Asn Val Ile Phe Ser	Asn Ile Arg Phe Gly Pro Ile Gly
	435	440
Ser Thr Val Gln Gly Leu	Pro Ser Ser Gly Gly	Thr Ser Ser Ser Ser
	455	460
Ser Ala Ala Pro Gln	Ser Thr Ser Thr Lys	Ala Ser Thr Thr Thr Ser
	470	475
		480

Ala Val Arg Thr Thr Ser Thr Ala Thr Thr Lys Thr Thr Ser Ser Ala
485 490 495
Pro Ala Gln Gly Thr Asn Thr Ala Lys His Trp Gln Gln Cys Gly Gly
500 505 510
Asn Gly Trp Thr Gly Pro Thr Val Cys Glu Ser Pro Tyr Lys Cys Thr
515 520 525
Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu
530 535

<210> 11

<211> 1248

<212> DNA

<213> verticillium tenerum

<220>

<221> CDS

<222> (1)..(1248)

<223>

<400>11

atg aag aag gct ctc atc acc agc ctc tcc ctg ctg gcc acg gcc atg	48
Met Lys Lys Ala Leu Ile Thr Ser Leu Ser Leu Leu Ala Thr Ala Met	
1 5 10 15	
ggc cag cag gcc ggt acc ctc gag acc gag acg cat ccc aag ctg acc	96
Gly Gln Gln Ala Gly Thr Leu Glu Thr Glu Thr His Pro Lys Leu Thr	
20 25 30	
tgg cag cgc tgc acc acc tcc ggc tgt acc aac gtc aac ggc gag gtc	144
Trp Gln Arg Cys Thr Thr Ser Gly Cys Thr Asn Val Asn Gly Glu Val	
35 40 45	
gtc atc gac gcc aac tgg cgt tgg gcc cac gac atc aac ggc tac gag	192
Val Ile Asp Ala Asn Trp Arg Trp Ala His Asp Ile Asn Gly Tyr Glu	
50 55 60	
aac tgc ttc gag ggc aac acc tgg acc ggc acc tgc agc ggc gcc gac	240
Asn Cys Phe Glu Gly Asn Thr Trp Thr Gly Thr Cys Ser Gly Ala Asp	
65 70 75 80	
ggc tgc gcg aag aac tgc gcc gtc gag gga gcc aac tac cag tcg acc	288
Gly Cys Ala Lys Asn Cys Ala Val Glu Gly Ala Asn Tyr Gln Ser Thr	
85 90 95	
tac ggt gtc tgc acc agc ggc aac gcc ctc tcc ctg cgc ttc gtc acc	336
Tyr Gly Val Ser Thr Ser Gly Asn Ala Leu Ser Leu Arg Phe Val Thr	
100 105 110	
gag cac gag cac ggc gtc aac acc ggt tcg cgc acg tac ctc atg gag	384
Glu His Glu His Gly Val Asn Thr Gly Ser Arg Thr Tyr Leu Met Glu	

115					120					125						
agc Ser	gcc Ala 130	acc Thr	aag Lys	tac Tyr	cag Gln	atg Met 135	ttc Phe	acc Thr	ctg Leu	atg Met	aac Asn 140	aac Asn	gag Glu	ctc Leu	gcc Ala	432
ttc Phe 145	gac Asp	gtc Val	gac Asp	ctg Leu	tcc Ser 150	aag Lys	gtc Val	gcc Ala	tgc Cys	ggc Gly 155	atg Met	aac Asn	agc Ser	gcc Ala	ctc Leu 160	480
tac Tyr	ctc Leu	gtc Val	ccc Pro	atg Met 165	aag Lys	gcc Ala	gac Asp	ggc Gly	ggt Gly 170	ctc Leu	tgc Ser	tcc Ser	gag Glu	acc Thr 175	aac Asn	528
aac Asn	aac Asn	gcc Ala	ggc Gly 180	gcc Ala	aag Lys	tac Tyr	ggt Gly	acc Thr 185	ggt Gly	tac Tyr	tgc Cys	gac Asp	gcc Ala 190	cag Gln	tgc Cys	576
gct Ala	cgc Arg	gat Asp 195	ctc Leu	aag Lys	ttc Phe	gtc Val	aac Asn 200	ggc Gly	aag Lys	gcc Ala	aac Asn	atc Ile 205	gag Glu	ggc Gly	tgg Trp	624
caa Gln 210	gcc Ala	tcc Ser	aag Lys	acc Thr	gac Asp	gag Glu 215	aac Asn	tct Ser	ggc Gly	gtc Val	ggt Gly 220	aac Asn	atg Met	ggc Gly	tcc Ser	672
tgc Cys 225	tgt Cys	gct Ala	gag Glu	att Ile	gac Asp 230	gtt Val	tgg Trp	gag Glu	tcc Ser	aac Asn 235	cgc Arg	gag Glu	tct Ser	ttc Phe	gcc Ala 240	720
ttc Phe	acc Thr	cct Pro	cac His	gct Ala 245	tgc Cys	tgc Ser	cag Gln	aac Asn	gag Glu 250	tac Tyr	cac His	gtc Val	tgc Cys	acc Thr 255	ggc Gly	768
gcc Ala	aac Asn	tgc Cys	ggc Gly 260	ggt Gly	acc Thr	tac Tyr	tgc Ser	gac Asp 265	gac Asp	cgc Arg	ttc Phe	gcc Ala	ggc Gly 270	aag Lys	tgc Cys	816
gat Asp	gcc Ala	aac Asn 275	ggt Gly	tgc Cys	gac Asp	tac Tyr	aac Asn 280	ccc Pro	ttc Phe	cgc Arg	gtg Val	ggc Gly 285	aac Asn	cag Gln	aac Asn	864
ttc Phe	tac Tyr 290	ggc Gly	ccc Pro	ggc Gly	atg Met	acc Thr 295	gtc Val	aac Asn	acc Thr	aac Asn	tcc Ser 300	aag Lys	ttc Phe	act Thr	gtc Val	912
atc Ile 305	tct Ser	cgc Arg	ttc Phe	cgg Arg	gag Glu 310	aac Asn	gag Glu	gcc Ala	tac Tyr	cag Gln 315	gtc Val	ttc Phe	atc Ile	cag Gln	aac Asn 320	960
ggc Gly	cgc Arg	acc Thr	atc Ile	gag Glu 325	gtc Val	ccc Pro	cgt Arg	ccc Pro	acc Thr 330	ctc Leu	tcc Ser	ggc Gly	atc Ile	acc Thr 335	cag Gln	1008
ttc Phe	gag Glu	gcc Ala	aag Lys 340	atc Ile	acc Thr	ccc Pro	gag Glu	ttc Phe 345	tgc Cys	tgc Ser	acc Thr	tac Tyr	ccc Pro 350	acc Thr	gtc Val	1056
ttc Phe	ggc Gly	gac Asp 355	cgc Arg	gac Asp	cgc Arg	cac His	ggc Gly 360	gag Glu	atc Ile	ggc Gly	ggc Gly	cac His 365	acc Thr	gcc Ala	ctc Leu	1104
aac Asn	gcg Ala 370	gcc Ala	ctc Leu	cgc Arg	atg Met	ccc Pro 375	atg Met	gtc Val	ctc Leu	gtc Val	atg Met 380	tcc Ser	atc Ile	tgg Trp	gcc Ala	1152
gac Asp 385	cac His	tac Tyr	gcc Ala	aac Asn	atg Met 390	ctc Leu	tgg Trp	ctc Leu	gac Asp	tcc Ser 395	atc Ile	tac Tyr	cgc Pro	cca Pro	gag Glu 400	1200

aag agg ggc cag ccc ggc gcc cac cgc ggc cgc aga tct aga ggg tga 1248
 Lys Arg Gly Gln Pro Gly Ala His Arg Gly Arg Arg Ser Arg Gly
 405 410 415

<210> 12

<211> 415

<212> PRT

<213> verticillium tenerum

<400> 12

Met Lys Lys Ala Leu Ile Thr Ser Leu Ser Leu Leu Ala Thr Ala Met
 1 5 10 15

Gly Gln Gln Ala Gly Thr Leu Glu Thr Glu Thr His Pro Lys Leu Thr
 20 25 30

Trp Gln Arg Cys Thr Thr Ser Gly Cys Thr Asn Val Asn Gly Glu Val
 35 40 45

Val Ile Asp Ala Asn Trp Arg Trp Ala His Asp Ile Asn Gly Tyr Glu
 50 55 60

Asn Cys Phe Glu Gly Asn Thr Trp Thr Gly Thr Cys Ser Gly Ala Asp
 65 70 75 80

Gly Cys Ala Lys Asn Cys Ala Val Glu Gly Ala Asn Tyr Gln Ser Thr
 85 90 95

Tyr Gly Val Ser Thr Ser Gly Asn Ala Leu Ser Leu Arg Phe Val Thr
 100 105 110

Glu His Glu His Gly Val Asn Thr Gly Ser Arg Thr Tyr Leu Met Glu
 115 120 125

Ser Ala Thr Lys Tyr Gln Met Phe Thr Leu Met Asn Asn Glu Leu Ala
 130 135 140

Phe Asp Val Asp Leu Ser Lys Val Ala Cys Gly Met Asn Ser Ala Leu
 145 150 155 160

Tyr Leu Val Pro Met Lys Ala Asp Gly Gly Leu Ser Ser Glu Thr Asn
 165 170 175

Asn Asn Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala Gln Cys
 180 185 190

Ala Arg Asp Leu Lys Phe Val Asn Gly Lys Ala Asn Ile Glu Gly Trp
 195 200 205

Gln Ala Ser Lys Thr Asp Glu Asn Ser Gly Val Gly Asn Met Gly Ser
 210 215 220
 Cys Cys Ala Glu Ile Asp Val Trp Glu Ser Asn Arg Glu Ser Phe Ala
 225 230 235 240
 Phe Thr Pro His Ala Cys Ser Gln Asn Glu Tyr His Val Cys Thr Gly
 245 250 255
 Ala Asn Cys Gly Gly Thr Tyr Ser Asp Asp Arg Phe Ala Gly Lys Cys
 260 265 270
 Asp Ala Asn Gly Cys Asp Tyr Asn Pro Phe Arg Val Gly Asn Gln Asn
 275 280 285
 Phe Tyr Gly Pro Gly Met Thr Val Asn Thr Asn Ser Lys Phe Thr Val
 290 295 300
 Ile Ser Arg Phe Arg Glu Asn Glu Ala Tyr Gln Val Phe Ile Gln Asn
 305 310 315 320
 Gly Arg Thr Ile Glu Val Pro Arg Pro Thr Leu Ser Gly Ile Thr Gln
 325 330 335
 Phe Glu Ala Lys Ile Thr Pro Glu Phe Cys Ser Thr Tyr Pro Thr Val
 340 345 350
 Phe Gly Asp Arg Asp Arg His Gly Glu Ile Gly Gly His Thr Ala Leu
 355 360 365
 Asn Ala Ala Leu Arg Met Pro Met Val Leu Val Met Ser Ile Trp Ala
 370 375 380
 Asp His Tyr Ala Asn Met Leu Trp Leu Asp Ser Ile Tyr Pro Pro Glu
 385 390 395 400
 Lys Arg Gly Gln Pro Gly Ala His Arg Gly Arg Arg Ser Arg Gly
 405 410 415

5 <210> 13
 <211> 1341
 <212> DNA
 <213> Neotermes castaneus

10 <220>
 <221> CDS
 <222> (1)..(1341)
 <223>

<400> 13

gca Ala 1	cga Arg	ggg Gly	ctc Leu	gct Ala 5	gct Ala	gca Ala	tgg Leu	ttc Phe	acc Thr 10	ttt Phe	gca Ala	tgt Cys	agc Ser	gtt Val 15	ggt Gly	48
atc Ile	ggc Gly	acc Thr	aaa Lys 20	acg Thr	gcc Ala	gag Glu	aac Asn	cac His 25	ccg Pro	aag Lys	ctg Leu	aac Asn	tgg Trp 30	cag Gln	aac Asn	96
tgc Cys	gcc Ala	tcc Ser 35	aag Lys	ggc Gly	agc Ser	tgc Cys	tca Ser 40	caa Gln	gtg Val	tcc Ser	ggc Gly	gaa Glu 45	gtg Val	aca Thr	atg Met	144
gac Asp 50	tgc Ser	aac Asn	tgg Trp	cgg Arg	tgg Trp	acc Thr 55	cac His	gat Asp	ggc Gly	aac Asn	ggc Gly 60	aag Lys	aac Asn	tgc Cys	tac Tyr	192
gac Asp 65	ggc Gly	aac Asn	acc Thr	tgg Trp	atc Ile 70	tcc Ser	agc Ser	ctc Leu	tgc Cys	cca Pro 75	gac Asp	ggc Gly	aag Lys	acc Thr	tgc Cys 80	240
tct Ser	gac Asp	aag Lys	tgc Cys	gtc Val 85	ctc Leu	gat Asp	ggc Gly	gcc Ala	gaa Glu 90	tac Tyr	caa Gln	gcg Ala	acc Thr	tac Tyr 95	ggc Gly	288
atc Ile	acc Thr	tgc Ser	aac Asn 100	ggg Gly	acc Thr	gcg Ala	gtc Val	acc Thr 105	ctc Leu	aag Lys	ttc Phe	gtc Val	acc Thr 110	cac His	ggc Gly	336
tgc Ser	tac Tyr	tgc Ser 115	acg Thr	aac Asn	atc Ile	ggc Gly	tcc Ser 120	cgc Arg	ctg Leu	tat Tyr	ctc Leu	ctc Leu 125	aag Lys	gac Asp	gaa Glu	384
aac Asn 130	act Thr	tac Tyr	tac Tyr	atc Ile	ttc Phe	aag Lys 135	gtg Val	aac Asn	aac Asn	aag Lys	gaa Glu 140	ttc Phe	aca Thr	ttc Phe	agc Ser	432
gtc Val 145	gat Asp	gtg Val	tgc Ser	aag Lys	ctc Leu 150	ccg Pro	tgc Cys	ggc Gly	ctg Leu	aac Asn 155	ggt Gly	gcc Ala	ctc Leu	tac Tyr	ttc Phe 160	480
gtc Val	tgc Ser	atg Met	gac Asp	gcc Ala 165	gac Asp	ggt Gly	ggc Gly	gca Ala	gga Gly 170	aag Lys	tat Tyr	tca Ser	ggt Gly	gcg Ala 175	aag Lys	528
cca Pro	ggc Gly	gcg Ala	aag Lys 180	tac Tyr	ggc Gly	ctc Leu	ggc Gly	tac Tyr 185	tgc Cys	gat Asp	gcg Ala	caa Gln	tgc Cys 190	ccg Pro	agc Ser	576
gat Asp	ctg Leu	aag Lys 195	ttc Phe	atc Ile	aac Asn	ggc Gly	gaa Glu 200	gcg Ala	aac Asn	agc Ser	gat Asp	ggc Gly 205	tgg Trp	aag Lys	ccc Pro	624
cag Gln 210	gcg Ala	aac Asn	gac Asp	aag Lys	aac Asn	gag Ala 215	gga Gly	aac Asn	ggc Gly	aaa Lys	tac Tyr 220	gga Gly	tgc Ser	tgc Cys	tgc Cys	672
tgc Ser 225	gaa Glu	atg Met	gac Asp	gtt Val	tgg Trp 230	gag Glu	gcg Ala	aac Asn	tgc Ser	cag Gln 235	gca Ala	aca Thr	gct Ala	tac Tyr	act Thr 240	720
ccg Pro	cac His	gtc Val	tgc Cys	aag Lys 245	acc Thr	acg Thr	ggc Gly	cag Gln	cag Gln 250	cgc Arg	tgc Cys	tgc Ser	ggc Gly	aca Thr 255	tgc Ser	768

gaa tgc ggc ggc cag gat ggc gca gcg cgt ttc cag gga ctg tgc gac	816
Glu Cys Gly Gly Gln Asp Gly Ala Ala Arg Phe Gln Gly Leu Cys Asp	
260 265 270	
gag gac ggt tgc gac ttc aac agc tgg cgc cag ggc gac aag acg ttc	864
Glu Asp Gly Cys Asp Phe Asn Ser Trp Arg Gln Gly Asp Lys Thr Phe	
275 280 285	
tac ggc ccg gga ttg act gtt gac acg aag tcg ccg ttc aca gtc gtc	912
Tyr Gly Pro Gly Leu Thr Val Asp Thr Lys Ser Pro Phe Thr Val Val	
290 295 300	
aca caa ttc gtc gga agt ccg gtg aag gaa atc cgc agg aag tac gtc	960
Thr Gln Phe Val Gly Ser Pro Val Lys Glu Ile Arg Arg Lys Tyr Val	
305 310 315 320	
cag aac gga aag gtg att gag aac tcg aag aac aag att tcg gga att	1008
Gln Asn Gly Lys Val Ile Glu Asn Ser Lys Asn Lys Ile Ser Gly Ile	
325 330 335	
gac gag acg aac gca gtg agt gat act ttc tgc gat cag caa aag aag	1056
Asp Glu Thr Asn Ala Val Ser Asp Thr Phe Cys Asp Gln Gln Lys Lys	
340 345 350	
gcc ttc ggt gat acg aac gat ttc aag aac aag ggc ggt ttc gct aag	1104
Ala Phe Gly Asp Thr Asn Asp Phe Lys Asn Lys Gly Gly Phe Ala Lys	
355 360 365	
ttg ggt cag gtg ttc gag act ggt cag gtt ctc gtg ctg tcg ctg tgg	1152
Leu Gly Gln Val Phe Glu Thr Gly Gln Val Leu Val Leu Ser Leu Trp	
370 375 380	
gat gac cac tcg gtt gca atg ctg tgg ttg gac tcg gcc tac cca acg	1200
Asp Asp His Ser Val Ala Met Leu Trp Leu Asp Ser Ala Tyr Pro Thr	
385 390 395 400	
aac aag gat aag agc agc cca ggt gtt gac cgt ggc cct tgc ccg acg	1248
Asn Lys Asp Lys Ser Ser Pro Gly Val Asp Arg Gly Pro Cys Pro Thr	
405 410 415	
act tcc ggg aag ccg gat gat gtt gaa agc caa tct ccc gat gca acc	1296
Thr Ser Gly Lys Pro Asp Asp Val Glu Ser Gln Ser Pro Asp Ala Thr	
420 425 430	
gtc att tat ggc aac atc aag ttc ggt gca ctg gac tcc act tac	1341
Val Ile Tyr Gly Asn Ile Lys Phe Gly Ala Leu Asp Ser Thr Tyr	
435 440 445	

<210> 14

<211> 447

<212> PRT

<213> Neotermes castaneus

<400> 14

Ala Arg Gly Leu Ala Ala Ala Leu Phe Thr Phe Ala Cys Ser Val Gly
1 5 10 15

Ile Gly Thr Lys Thr Ala Glu Asn His Pro Lys Leu Asn Trp Gln Asn
20 25 30

Cys Ala Ser₃₅ Lys Gly Ser Cys₄₀ Ser Gln Val Ser Gly₄₅ Glu Val Thr Met
 Asp Ser₅₀ Asn Trp Arg Trp₅₅ His Asp Gly Asn₆₀ Gly Lys Asn Cys Tyr
 Asp Gly Asn Thr Trp₆₅ Ile₇₀ Ser Ser Leu Cys₇₅ Pro Asp Gly Lys Thr Cys₈₀
 Ser Asp Lys Cys Val₈₅ Leu Asp Gly Ala₉₀ Glu Tyr Gln Ala Thr Tyr₉₅ Gly
 Ile Thr Ser₁₀₀ Asn Gly Thr Ala Val₁₀₅ Thr Leu Lys Phe Val₁₁₀ Thr His Gly
 Ser Tyr Ser₁₁₅ Thr Asn Ile Gly Ser₁₂₀ Arg Leu Tyr Leu₁₂₅ Lys Asp Glu
 Asn Thr Tyr Tyr Ile Phe₁₃₀ Lys₁₃₅ Val Asn Asn Lys₁₄₀ Glu Phe Thr Phe Ser
 Val₁₄₅ Asp Val Ser Lys₁₅₀ Leu Pro Cys Gly Leu₁₅₅ Asn Gly Ala Leu Tyr Phe₁₆₀
 Val Ser Met Asp₁₆₅ Ala Asp Gly Gly Ala₁₇₀ Gly Lys Tyr Ser Gly₁₇₅ Ala Lys
 Pro Gly Ala₁₈₀ Lys Tyr Gly Leu Gly₁₈₅ Tyr Cys Asp Ala Gln₁₉₀ Cys Pro Ser
 Asp Leu Lys₁₉₅ Phe Ile Asn Gly Glu₂₀₀ Ala Asn Ser Asp Gly₂₀₅ Trp Lys Pro
 Gln Ala₂₁₀ Asn Asp Lys Asn₂₁₅ Ala Gly Asn Gly Lys Tyr₂₂₀ Gly Ser Cys Cys
 Ser Glu Met Asp Val₂₂₅ Trp₂₃₀ Glu Ala Asn Ser Gln₂₃₅ Ala Thr Ala Tyr Thr₂₄₀
 Pro His Val Cys₂₄₅ Lys Thr Thr Gly Gln₂₅₀ Arg Cys Ser Gly₂₅₅ Thr Ser
 Glu Cys Gly₂₆₀ Gly Gln Asp Gly Ala₂₆₅ Arg Phe Gln Gly₂₇₀ Leu Cys Asp
 Glu Asp Gly₂₇₅ Cys Asp Phe Asn₂₈₀ Ser Trp Arg Gln Gly₂₈₅ Asp Lys Thr Phe
 Tyr Gly₂₉₀ Pro Gly Leu Thr Val₂₉₅ Asp Thr Lys Ser Pro₃₀₀ Phe Thr Val Val
 Thr Gln Phe Val Gly Ser Pro Val Lys Glu Ile Arg Arg Lys Tyr Val

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

5 <210> 15
 <211> 1359
 <212> DNA
 <213> Melanocarpus albomyces

```
10      <220>
      <221> CDS
      <222> (1)..(1359)
      <223>
```

15 <400> 15

atg	atg	atg	aag	cag	tac	ctc	cag	tac	ctc	gcg	gcc	gcg	ctg	ccg	ctc	48
Met	Met	Met	Lys	Gln	Tyr	Leu	Gln	Tyr	Leu	Ala	Ala	Ala	Leu	Pro	Leu	
1				5					10					15		
gtc	ggc	ctc	gcc	gcc	ggc	cag	cgc	gct	ggt	aac	gag	acg	ccc	gag	agc	96
Val	Gly	Leu	Ala	Ala	Gly	Gln	Arg	Ala	Gly	Asn	Glu	Thr	Pro	Glu	Ser	
			20					25					30			
cac	ccc	ccg	ctc	acc	tgg	cag	agg	tgc	acg	gcc	ccg	ggc	aac	tgc	cag	144
His	Pro	Pro	Leu	Thr	Trp	Gln	Arg	Cys	Thr	Ala	Pro	Gly	Asn	Cys	Gln	
		35					40					45				
acc	gtg	aac	gcc	gag	gtc	gta	att	gac	gcc	aac	tgg	cgc	tgg	ctg	cac	192

Thr	Val	Asn	Ala	Glu	Val	Val	Ile	Asp	Ala	Asn	Trp	Arg	Trp	Leu	His	
50						55					60					
gac	gac	aac	atg	cag	aac	tgc	tac	gac	ggc	aac	cag	tgg	acc	aac	gcc	240
Asp	Asp	Asn	Met	Gln	Asn	Cys	Tyr	Asp	Gly	Asn	Gln	Trp	Thr	Asn	Ala	
65					70					75					80	
tgc	agc	acc	gcc	acc	gac	tgc	gct	gag	aag	tgc	atg	atc	gag	ggc	gcc	288
Cys	Ser	Thr	Ala	Thr	Asp	Cys	Ala	Glu	Lys	Cys	Met	Ile	Glu	Gly	Ala	
				85					90					95		
ggc	gac	tac	ctg	ggc	acc	tac	ggc	gcc	tgc	acc	agc	ggc	gac	gcc	ctg	336
Gly	Asp	Tyr	Leu	Gly	Thr	Tyr	Gly	Ala	Ser	Thr	Ser	Gly	Asp	Ala	Leu	
			100					105					110			
acg	ctc	aag	ttc	gtc	acg	aag	cac	gag	tac	ggc	acc	aac	gtc	ggc	tgc	384
Thr	Leu	Lys	Phe	Val	Thr	Lys	His	Glu	Tyr	Gly	Thr	Asn	Val	Gly	Ser	
		115					120					125				
cgc	ttc	tac	ctc	atg	aac	ggc	cgc	gac	aag	tac	cag	atg	ttc	gac	ctc	432
Arg	Phe	Tyr	Leu	Met	Asn	Gly	Pro	Asp	Lys	Tyr	Gln	Met	Phe	Asp	Leu	
	130					135					140					
ctg	ggc	aac	gag	ctt	ggc	ttt	gac	gtc	gac	ctc	tgc	acc	gtc	gag	tgc	480
Leu	Gly	Asn	Glu	Leu	Ala	Phe	Asp	Val	Asp	Leu	Ser	Thr	Val	Glu	Cys	
145					150					155					160	
ggc	atc	aac	agc	gcc	ctg	tac	ttc	gtc	gcc	atg	gag	gag	gac	ggc	ggc	528
Gly	Ile	Asn	Ser	Ala	Leu	Tyr	Phe	Val	Ala	Met	Glu	Glu	Asp	Gly	Gly	
				165					170					175		
atg	gcc	agc	tac	cgc	agc	aac	cag	gcc	ggc	gcc	cgc	tac	ggc	act	ggg	576
Met	Ala	Ser	Tyr	Pro	Ser	Asn	Gln	Ala	Gly	Ala	Arg	Tyr	Gly	Thr	Gly	
			180					185					190			
tac	tgc	gat	gcc	caa	tgc	gct	cgt	gac	ctc	aag	ttc	gtt	ggc	ggc	aag	624
Tyr	Cys	Asp	Ala	Gln	Cys	Ala	Arg	Asp	Leu	Lys	Phe	Val	Gly	Gly	Lys	
		195					200					205				
ggc	aac	att	gag	ggc	tgg	aag	cgc	tcc	acc	aac	gac	ccc	aac	gct	ggc	672
Ala	Asn	Ile	Glu	Gly	Trp	Lys	Pro	Ser	Thr	Asn	Asp	Pro	Asn	Ala	Gly	
	210					215					220					
gtc	ggc	cgc	tac	ggc	ggc	tgc	tgc	gct	gag	atc	gac	gtc	tgg	gag	tgc	720
Val	Gly	Pro	Tyr	Gly	Gly	Cys	Cys	Ala	Glu	Ile	Asp	Val	Trp	Glu	Ser	
225					230					235					240	
aac	ggc	tat	gcc	ttc	gct	ttc	acg	cgc	cac	gcg	tgc	acg	acc	aac	gag	768
Asn	Ala	Tyr	Ala	Phe	Ala	Phe	Thr	Pro	His	Ala	Cys	Thr	Thr	Asn	Glu	
				245					250					255		
tac	cac	gtc	tgc	gag	acc	acc	aac	tgc	ggc	ggc	acc	tac	tgc	gag	gac	816
Tyr	His	Val	Cys	Glu	Thr	Thr	Asn	Cys	Gly	Gly	Thr	Tyr	Ser	Glu	Asp	
			260					265					270			
cgc	ttc	ggc	ggc	aag	tgc	gac	ggc	aac	ggc	tgc	gac	tac	aac	ccc	tac	864
Arg	Phe	Ala	Gly	Lys	Cys	Asp	Ala	Asn	Gly	Cys	Asp	Tyr	Asn	Pro	Tyr	
		275					280					285				
cgc	atg	ggc	aac	ccc	gac	ttc	tac	ggc	aag	ggc	aag	acg	ctc	gac	acc	912
Arg	Met	Gly	Asn	Pro	Asp	Phe	Tyr	Gly	Lys	Gly	Lys	Thr	Leu	Asp	Thr	
	290					295					300					
agc	cgc	aag	ttc	acc	gtc	gtc	tcc	cgc	ttc	gag	gag	aac	aag	ctc	tcc	960
Ser	Arg	Lys	Phe	Thr	Val	Val	Ser	Arg	Phe	Glu	Glu	Asn	Lys	Leu	Ser	
305					310					315					320	
cag	tac	ttc	atc	cag	gac	ggc	cgc	aag	atc	gag	atc	cgc	cgc	cgc	acg	1008
Gln	Tyr	Phe	Ile	Gln	Asp	Gly	Arg	Lys	Ile	Glu	Ile	Pro	Pro	Pro	Thr	

325								330				335				
tgg	gag	ggc	atg	ccc	aac	agc	agc	gag	atc	acc	ccc	gag	ctc	tgc	tcc	1056
Trp	Glu	Gly	Met	Pro	Asn	Ser	Ser	Glu	Ile	Thr	Pro	Glu	Leu	Cys	Ser	
340								345				350				
acc	atg	ttc	gat	gtg	ttc	aac	gac	cgc	aac	cgc	ttc	gag	gag	gtc	ggc	1104
Thr	Met	Phe	Asp	Val	Phe	Asn	Asp	Arg	Asn	Arg	Phe	Glu	Glu	Val	Gly	
355								360				365				
ggc	ttc	gag	cag	ctg	aac	aac	gcc	ctc	cgg	gtt	ccc	atg	gtc	ctc	gtc	1152
Gly	Phe	Glu	Gln	Leu	Asn	Asn	Ala	Leu	Arg	Val	Pro	Met	Val	Leu	Val	
370								380								
atg	tcc	atc	tgg	gac	gac	cac	tac	gcc	aac	atg	ctc	tgg	ctc	gac	tcc	1200
Met	Ser	Ile	Trp	Asp	Asp	His	Tyr	Ala	Asn	Met	Leu	Trp	Leu	Asp	Ser	
385								395				400				
atc	tac	ccg	ccc	gag	aag	gag	ggc	cag	ccc	ggc	gcc	gcc	cgt	ggc	gac	1248
Ile	Tyr	Pro	Pro	Glu	Lys	Glu	Gly	Gln	Pro	Gly	Ala	Ala	Arg	Gly	Asp	
405								410				415				
tgc	ccc	acg	gac	tcg	ggt	gtc	ccc	gcc	gag	gtc	gag	gct	cag	ttc	ccc	1296
Cys	Pro	Thr	Asp	Ser	Gly	Val	Pro	Ala	Glu	Val	Glu	Ala	Gln	Phe	Pro	
420								425				430				
gac	gcc	cag	gtc	gtc	tgg	tcc	aac	atc	cgc	ttc	ggc	ccc	atc	ggc	tcg	1344
Asp	Ala	Gln	Val	Val	Trp	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	Gly	Ser	
435								440				445				
acc	tac	gac	ttc	taa												1359
Thr	Tyr	Asp	Phe													450

<210> 16

<211> 452

<212> PRT

<213> Melanocarpus albomyces

<400> 16

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

Gly Asp Tyr Leu Gly Thr Tyr Gly Ala Ser Thr Ser Gly Asp Ala Leu
 100 105 110
 Thr Leu Lys Phe Val Thr Lys His Glu Tyr Gly Thr Asn Val Gly Ser
 115 120 125
 Arg Phe Tyr Leu Met Asn Gly Pro Asp Lys Tyr Gln Met Phe Asp Leu
 130 135 140
 Leu Gly Asn Glu Leu Ala Phe Asp Val Asp Leu Ser Thr Val Glu Cys
 145 150 155 160
 Gly Ile Asn Ser Ala Leu Tyr Phe Val Ala Met Glu Glu Asp Gly Gly
 165 170 175
 Met Ala Ser Tyr Pro Ser Asn Gln Ala Gly Ala Arg Tyr Gly Thr Gly
 180 185 190
 Tyr Cys Asp Ala Gln Cys Ala Arg Asp Leu Lys Phe Val Gly Gly Lys
 195 200 205
 Ala Asn Ile Glu Gly Trp Lys Pro Ser Thr Asn Asp Pro Asn Ala Gly
 210 215 220
 Val Gly Pro Tyr Gly Gly Cys Cys Ala Glu Ile Asp Val Trp Glu Ser
 225 230 235 240
 Asn Ala Tyr Ala Phe Ala Phe Thr Pro His Ala Cys Thr Thr Asn Glu
 245 250 255
 Tyr His Val Cys Glu Thr Thr Asn Cys Gly Gly Thr Tyr Ser Glu Asp
 260 265 270
 Arg Phe Ala Gly Lys Cys Asp Ala Asn Gly Cys Asp Tyr Asn Pro Tyr
 275 280 285
 Arg Met Gly Asn Pro Asp Phe Tyr Gly Lys Gly Lys Thr Leu Asp Thr
 290 295 300
 Ser Arg Lys Phe Thr Val Val Ser Arg Phe Glu Glu Asn Lys Leu Ser
 305 310 315 320
 Gln Tyr Phe Ile Gln Asp Gly Arg Lys Ile Glu Ile Pro Pro Pro Thr
 325 330 335
 Trp Glu Gly Met Pro Asn Ser Ser Glu Ile Thr Pro Glu Leu Cys Ser
 340 345 350
 Thr Met Phe Asp Val Phe Asn Asp Arg Asn Arg Phe Glu Glu Val Gly
 355 360 365

Gly Phe Glu Gln Leu Asn Asn Ala Leu Arg Val Pro Met Val Leu Val
 370 375 380
 Met Ser Ile Trp Asp Asp His Tyr Ala Asn Met Leu Trp Leu Asp Ser
 385 390 395 400
 Ile Tyr Pro Pro Glu Lys Glu Gly Gln Pro Gly Ala Ala Arg Gly Asp
 405 410 415
 Cys Pro Thr Asp Ser Gly Val Pro Ala Glu Val Glu Ala Gln Phe Pro
 420 425 430
 Asp Ala Gln Val Val Trp Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser
 435 440 445
 Thr Tyr Asp Phe
 450

<210> 17
 <211> 221
 <212> DNA
 <213> Trichothecium roseum

<220>
 <221> misc_feature
 <222> (1)..(221)
 <223> Secuencia de codificación parcial de CBH1

<400> 17
 tacgcccagt ggcgccgtga cctcaagttc ctcggcggca cttccaacta cgacggctgg 60
 aagccctcgg acactgacga cagcgccggt gtcggcaacc gcggatcctg ctgcgccgag 120
 attgacatct gggagtccaa ctcgcacgcc ttgccttca cccccacgc ctgcgagaac 180
 aacgagtacc acatctgcga gaccaccgac tgcggcggca c 221

<210> 18
 <211> 239
 <212> DNA
 <213> Humicola nigrescens

<220>
 <221> misc_feature
 <222> (1)..(239)
 <223> Secuencia de codificación parcial de CBH1
 <400> 18

tacggcacgg ggtactgcga cgcccaatgc gcccgcgatc tcaagttcgt tggcggcaag 60
gccaatgttg agggctggaa acagtccacc aacgatgcca atgccggcgt gggcccgatg 120
ggcgggttgct gcgccgaaat tgacgtctgg gaatcgaacg cccatgcctt cgccttcacg 180
ccgcacgcgt gcgagaacaa caagtaccac atctgcgaga ctgacggatg cggcggcac 239

5 <210> 19
<211> 199
<212> DNA
<213> Cladorrhinum foecundissimum

10 <220>
<221> misc_feature
<222> (1)..(199)
<223> Secuencia de codificación parcial de CBH1

15 <400> 19
tacataaacg gtatcggcaa cgttgagggg tggctctcct ctaccaacga tcccaacgct 60
gggtgtcggta accrcggtac ttgctgctcc gagaatggat atctgggagg ccaacaagat 120
ctcgaccgcc tacactcccc acccctgcac caccatcgac cagcacatgt gcgagggcaa 180
ctcgtgcggc ggaacctac 199

20 <210> 20
<211> 191
<212> DNA
<213> Diplodia gossypina

<220>
<221> misc_feature
25 <222> (1)..(191)
<223> Secuencia de codificación parcial de CBH1

<400> 20
30 gttgatccga cggcaaggcc caacgtcgag ggctgggtcc cgtccgagaa cgactccaac 60
gctggtgtcg gcaaccttgg ctcttgctgt gctgagatgg atatctggga ggccaactcc 120
atctcgaccg cctacacccc ccacagctgc aagacggtcg cccagcactc ttgcactggc 180
gacgactgcg g 191

35 <210> 21
<211> 232
<212> DNA
<213> Myceliophthora thermophila

<220>
<221> misc_feature

<222> (1)..(232)

<223> Secuencia de codificación parcial de CBH1

<400> 21

```

gggtactgcg acgccaatg cgcacgcgac ctcaagttcg tcggcggcaa gggcaacatc      60
gagggctgga agccgtccac caacgatgcc aatgccggtg tcggtcetta tggcgggtgc      120
tgcgctgaga tcgacgtctg ggagtcgaac aagtatgctt tcgctttcac cccgcacggt      180
tgcgagaacc ctaaatacca cgtctgcgag accaccaact gcggcggcac ct                232

```

<210> 22

<211> 467

<212> DNA

<213> Rhizomucor pusillus

<220>

<221> misc_feature

<222> (1)..(467)

<223> Secuencia de codificación parcial de CBH1

<400> 22

```

tccttcgcct ttacccccca cgcttgctcg cagnaacgag taccacgtct gcaccaccaa      60
caactgcggc ggcacctact cggacgaccg cttcgccggc aagtgcgacg ccaacggttg      120
cgactacaac ccgttcgcgc tgggcaacca ggacttctac ggcccgggca tgaccgtcga      180
caccaactcc aagttcaccc tcattctccg cttcagggag aacgaggcct accaggtctt      240
catgcagggc ggccggacca tcgaggtccc ggccccgcag ctgtccgggc tcaccagtt      300
cgacgccaag atcacccccg agttctgcga cactaccgc accgtcttcg acgaccgcaa      360
ccgccacggc gagatcggcg gccacaccgc cctcaacgcc gccctgcgca tgcccatggt      420
cctcgtcatg tccattctggg ctgaccacta cgccagctgc tagtgtc                467

```

<210> 23

<211> 534

<212> DNA

<213> Meripilus giganteus

<220>

<221> misc_feature

<222> (1)..(534)

<223> Secuencia de codificación parcial de CBH1

<400> 23

```

gggaggggctc cccgaacgac ccgaacgcgg gaagcggcca gtacggaacg tgctgcaacg      60
agatggacat ctgggaggcg aaccagaacg gcgcggcggt caccgcgcac gtctgctccg      120
tcgacggcca gacgcgctgc gagggcacgg actgcggcga cggcgacgag cggtagcagc      180
gcattctgca caaggacggc tgcgacttca actcgtaccg catggggegac cagtccctcc      240
tcggcctcgg caagaccgct gacacctcga agaagttcac cgtcgtcacc cagttcctca      300
ccgcggacaa caccgacgtc ggccagctca cggagatccg ccggtgttac gtgcaggacg      360
gcaagggtcat cgcgaactcg aagacgaaca tccccggcct cgactcgttc gactccatca      420
ccgacgactt ctgcaacgcg cagaaggagg tcttcggcga caccaactcg ttcgagaagc      480
tcggcggcct cgcggagatg ggcaaggcct tccagaaggg catggtcctc gtca          534

```

<210> 24
 <211> 563
 <212> DNA
 <213> Exidia glandulosa

<220>
 <221> misc_feature
 <222> (1)..(563)
 <223> Secuencia de codificación parcial de CBH1
 <400> 24

```

gccacgtcga gggctggact ccttcgccaa cgatgccaac gccggcattg gcacccacgg      60
ctcctgctgt tcggagatgg acatctggga ggctaacaat gttgccgctg cgtacacccc      120
ccatccttgc acaactatcg gccagtcgat ctgctcgggc gattcttgcg gaggaaccta      180
cagctctgac cgttacgccg gtgtctgca tccagacggg tgcgatttca acagctaccg      240
catgggcgac acgggcttct acggcaaggg cctgacagtc gacacgagct ccaagttcac      300
cgtcgtcacc cagttcctca ccggtccga cggcaacctt tccgagatca agcgcttcta      360
cgtccagaac ggcaagggtc ttcccaactc gcagtcgaag attgccggcg tcagcggcaa      420
ctccatcacc accgacttct gctccgccca gaagaccgcc ttcggcgaca ccaacgtctt      480
cgcgcaaaag ggaggtactc gccgggatgg gcgcgcgcct caaggccggc atggtcctcg      540
tcattgtccat ctgggacgac cac          563

```

<210> 25
 <211> 218
 <212> DNA
 <213> xylaria hypoxylon

<220>
 <221> misc_feature
 <222> (1)..(218)
 <223> Secuencia de codificación parcial de CBH1
 <400> 25

gacgctcagt gtgcccgtga cttgaagttc gtcgggtggca agggcaacgt tgaggggatgg 60
gagccatcca ccaacgacga caacgcccgt gttggccctt acggwgccctg ctgtgccgaa 120
atsgatgtst gggagtccaa ctstcactct ttcgctttca cccctcacc wtgcaccacc 180
aacgaatacc acgtctgtga gcaggacgag tgtggcgg 218

<210> 26

<211> 492

<212> DNA

5 <213> Acremonium sp.

<220>

<221> misc_feature

<222> (1)..(492)

10 <223> Secuencia de codificación parcial de CBH1

<400> 26

gggacggggt actgcgacgc ccaatgcgcc cgtgatctca agttcgtcgg cggcaaggcc 60
aacattgagg gctggaggcc gtccaccaac gacgcgaacg ccggcgtcgg cccgatgggc 120
ggctgctgcg cggaaatcga tgtctgggag tccaacgcc acgcttttgc cttcacgccg 180
cacgcgtgcg agaacaacaa ctaccacatc tgcgagacct ccaactgcgg cggtacctac 240
tccgacgacc gcttcgccgg cctctgcgac gccaacggct gcgactacaa cccgtaccgc 300
atgggcaacc ccgactttcta cggcaagggc aagactcttg acacctcgcg gaagttcacc 360
gtcgtcaccg gctttcagga gaacgacctc tcgcagtact tcgtccagga cggcccgaag 420
atcgagatcc cgccccgac ctgggacggc ctcccgaaga gcagcacata cgccgagctg 480
tgcgcgaccc ag 492

<210> 27

<211> 481

15 <212> DNA

<213> Acremonium sp.

<220>

<221> misc_feature

20 <222> (1)..(481)

<223> Secuencia de codificación parcial de CBH1

<400> 27

```

ggctccggtt actcctaccc ttgcacggaa atcggccaga gccgctgcga gggcgacagc    60
tgcggcggta cctacagcac cgaccgctac gctggcgctt gcgaccccgga tggatgcgac    120
ttcaactcgt accgccaggg caacaagacc ttctatggca agggcatgac cgtcgacacc    180
accaagaaga ttaccgtcgt caccaggttc ctcaccgact cgtccggcaa cctgtccgag    240
atcaagcgtt tctacgccc gaacggcgct gtcattccca actccgagtc caccattgct    300
ggcgtccctg gcaactcgat caccagggac tactgcgaca agcagaagac cgcctttggt    360
gacaacaacg acttcgacaa gaagggtggt ctgcccaga tgggtaaggc cctggcccaa    420
cccattggtc tcgtcatgtc cgtctgggat gaccatgccg tcaacatgct ctgcttcgaa    480
a                                                                 481

```

<210> 28

<211> 463

<212> DNA

<213> Chaetomium sp.

<220>

<221> misc_feature

<222> (1)..(463)

<223> Secuencia de codificación parcial de CBH1

<400> 28

```

ctccccgtct tcacgccgca cgcgtgcaag aacatcaagt accacgtctg cgagacgtcg    60
ggatgcggcg gcacctactc ggaggaccgc ttgcggggcg actgcgacgc caacgggtgc    120
gactacaacc cctaccgcat gggcaacacc gacttctacg gcaagggcat gacggctcgac    180
accagcaaga agttcacctg cgtgacccaa ttccaggaga acaagctcac ccagttcttc    240
gtccagaacg gcaagaagat cgagatccct ggcccgaagt gggacggcat tgagggcgac    300
agcgccgcca tcacgcccc gctgtgcact tccatgttca aggccttcga cgaccgcgat    360
cgtttctcgg aggtcggcgg cttcaccag atcaaccagg ccctctcggg gcccatggtg    420
ctcgtcatgt ccatctggga cgaccactac gccaacatgc ttg                                463

```

<210> 29

<211> 513

<212> DNA

<213> Chaetomidium pingtungium

<220>

<221> misc_feature

<222> (1)..(513)

<223> Secuencia de codificación parcial de CBH1

<400> 29

```

gaagpggtggc agccctcctc caacgatgcc aatgcgggta ccggcaacca cgggtcctgc      60
tgcgcggaga tggatatctg ggaggccaac agcatctcca cggccttcac ccccatccg      120
tgcgacacgc ccggccaggt gatgtgcacc ggtgatgcct gcggtggcac ctacagctcc      180
gaccgctacg gcggcacctg cgaccccgac ggatgtgatt tcaactcctt ccgccagggc      240
aacaagacct tctacggccc tggcatgacc gtcgacacca agagcaagtt taccgtcgtc      300
accagttca tcaccgacga cggcacctcc agcggcaccc tcaaggagat caagcgcttc      360
tacgtgcaga acggcaaggt gatccccaac tcggagtcca cctggaccgg cgtcagcggc      420
aactccatca ccaccgagta ctgcaccgcc cagaagagcc tgttcaggga ccagaacgtc      480
ttcgaaaagc acggtggcct cgagggcatg ggt                                     513

```

<210> 30

<211> 579

<212> DNA

<213> Myceliophthora thermophila

<220>

<221> misc_feature

<222> (1)..(579)

<223> Secuencia de codificación parcial de CBH1

<400> 30

```

gagatggata tttgggaggc caacaacatg gccgccgctt tcaactccca cccttgaccc      60
gtgatcggcc agtcgcgctg cgagggcgac tcgtgcggcg gtacctacag caccgaccgc      120
tatgccggca tctgcgaccc cgacgggatg gacttcaact cgtaccgcca gggcaacaag      180
accttctacg gcaagggcat gacggtcgac acgaccaaga agatcacggg cgtcaccag      240
ttcctcaaga actcggccgg cgagctctcc gagatcaagc ggttctacgt ccagaacggc      300
aaggatcatc ccaactccga gtccaccatc ccgggcgctg agggcaactc cattaccag      360
gactgggtgc accgccagaa ggccgctttc ggcgacgtga ccgactttca ggacaagggc      420
ggcatggctc agatgggcaa ggccctcgcg ggcccaatgg tcctcgtcat gtccatctgg      480
gacgaccacg ccgtcaacat gctctggctc gaaatcacta gtgcggccgc tgcaggtcga      540
ccatatggga gagctccacg cgttgggatg atagcttga                                     579

```

<210> 31

<211> 514

<212> DNA

<213> Myceliophthora hinnulea

<220>

<221> misc_feature

<222> (1)..(514)

<223> Secuencia de codificación parcial de CBH1

<400> 31

```

cgtgagggct gggagagctc gaccaacgat gccaacgccg gcacgggacg gtacggcagc      60
tgctgctccg agatggacgt ctgggaggcc aacaacatgg ccaccgcctt cccccccat      120
ccttgacca tcacgggcca gtgcgcctgc gagggcgaga cgtgcggcgg cacctacagc      180
tcggaccgct acgccggcgt ctgcgacccc gacggctgcg acttcaactc gtaccgccag      240
ggcaacaaga ctttctacgg caagggcatg acggtcgaca cgaccaagaa gctcacggtc      300
gtcacgcagt tcctcaagaa ctccggccggc gagctgtccg agatcaagcg gttctacgtc      360
caggacggca aggtgatccc caactccgag tccaccatcc ccggcgtcga gggcaactcg      420
atcacgcagg actggtgcga ccgccagaag gccgccttcg gcgacgtcac cgacttccag      480
gacaagggcg gcatgggtcca gatggcaagg cgct                                     514

```

5
 <210> 32
 <211> 477
 <212> DNA
 <213> Sporotrichum pruinosum

10
 <220>
 <221> misc_feature
 <222> (1)..(477)
 <223> Secuencia de codificación parcial de CBH1
 <400> 32
 cacccttgcc gcaccacgaa cgacgggtggc taccaacgct gccaaaggacg tgactgcaac 60
 cagcctcggt atgaggggtct ttgcgatcct gacggttgcg actacaaccc tttccgtatg 120
 ggtaaccgcg aattctacgg ccctggaaag accgtcgaca ccaacaggaa gttcactggt 180
 gtgacccaat tcattaccga caacaactct gacactggta ccctcgtcga catccgccgc 240
 ctctacgtcc aagacggccg tgtcattgcc aaccctccca ccaacttccc cgggtctcatg 300
 cccgcccacg actccatcac ttagcaattc tgtgacgacg ccaagcgagc attcgaggac 360
 aacgacagct ttggcaggaa cgggtggtctt gtcacatgg gtcgctccct tgccaagggc 420
 catgtcctcg ccctttccat ttggaatgat cacactgcca acatgctctg gctcgaa 477

15
 <210> 33
 <211> 500
 <212> DNA
 <213> Thielavia cf. microspora

20
 <220>
 <221> misc_feature
 <222> (1)..(500)
 <223> Secuencia de codificación parcial de CBH1
 25
 <400> 33

gagatagatg tctgggagtc caactcgcac tcgtttgcct tcacgccgca cgcgtgcaag 60
aacaacaagt accacgtctg ccagacgacc ggggtgcggcg gcacctactc ggaggaccgc 120
ttcgccggcg actgcgacgc caacggctgc gactacaacc cctaccgcat gggcaacacc 180
gactttttacg gcaaggggcaa gacggctgcac acgagcaaga agttttaccat ggtgacccag 240
ttccaaaaga acaagctcgt ccagttcttt gtccaggacg gcaagaagat cgacatcccc 300
ggccccaagt gggacggcct gccgcagggc agcgcgcgca tcaccccgga gctgtgcacc 360
ttcatgttca aggccttcaa cgaccgcgac cgcttctcag aggttggcgg cttcgaccag 420
atcaacacgg cctctcgggt gccaatgggt ctcgtcatgt ccatctggga tgatcactac 480
gccaacatgc tctggcttga 500

<210> 34

<211> 470

<212> DNA

<213> Scytalidium sp.

<220>

<221> misc_feature

<222> (1)..(470)

<223> Secuencia de codificación parcial de CBH1

<400> 34

cgtnngccc gcgtcgcatg ctcccggccg catggcccg cggatttcca gccagagcat 60
gttggagtgg tggatcatcc agatggacat gacaaggacc atgggaatgg tgagggcctc 120
gttcagagca tcgaagccac cggctctcggc gaagcggttg cggatcatga agacgcggaa 180
ctgagcatcg cagagctcag ggggtgatgtc ggcgctgttc gggaggccgg gccaggtcgg 240
agggggcacc tcgatcttgc ggccgtcctg gacgaagaac tgagagagcc tgttacgctc 300
gaagcgggag acaacgggtga acttgcggtt ggtgtcgacg gtcttgccct tgccatagaa 360
gtccttggtg cccatgcggg aggggttgta gtcgcagccg ttggcatcgc agtagccggc 420
gaagcgggtca tccgagtagg taccaccgca gttgttggtc tccagatgtg 470

<210> 35

<211> 491

<212> DNA

<213> Scytalidium sp.

<220>

<221> misc_feature

<222> (1)..(491)

<223> Secuencia de codificación parcial de CBH1

<400> 35

gaaatcgacg tctgggagtc gaacgcctat gcctatgcct tacccegcac gcttgcgcca 60
gccagaaccg ctaccacgtc tgcgagacca acaactgcgg tggtacctac tcggatgacc 120
gcttcgcccg ttactgcgat gccaacggct gcgactacaa cccgtaccgc atgggcaaca 180
gggactttcta cggcaagggc ctgcaggctc acaccagccg gaagttcacc gtcgtgagcc 240
gcttcgagcg caacaagctc acccagttct tcgttcagga cggccgcaag atcgagcccc 300
ctgcgcccgc ctgggacggc atcccgaaga gcgccgacat ccccccgag ttctgcagcg 360
cccagttcaa ggtcttcgac gaccgtgacc gcttcgcgga gactggcggc ttcgatgccc 420
tgaacgatgc tctcagcatt cccatgggtc ttgtcatgtc catctgggat taccactact 480
ccaacataat c 491

<210> 36
<211> 221
<212> DNA
<213> Trichophaea saccata

<220>
<221> misc_feature
<222> (1)..(221)
<223> Secuencia de codificación parcial de CBH1
<400> 36

tgcgactccc agtgtccccg cgatctcaag ttcattcaatg gacagggcaa cggtgaaggc 60
tggaagccat cctcaaatga tgccaacgca ggcgtcgggg gacacggttc ctgctgcgca 120
gagatggatg tttgggaggg caattccatc tccgcggccg taacaccgca ctcgtgctcc 180
acaaccagcc agacgatgtg caacggcgac tcctgcggcg g 221

<210> 37
<211> 1365
<212> DNA
<213> Diplodia gossypina

<220>
<221> CDS
<222> (1)..(1365)
<223>
<400> 37

atg Met 1	ctt Leu	acc Thr	cag Gln 5	gca Ala	gtt Val	ctc Leu	gct Ala	act Thr	ctc Leu 10	gcc Ala	acc Thr	ctg Leu	gcc Ala 15	gcc Ala	agc Ser	48
cag Gln	cag Gln	gtc Val	ggc Gly 20	acc Thr	cag Gln	aag Lys	gag Glu	gag Glu 25	gtc Val	cac His	ccc Pro	tcc Ser	atg Met 30	acc Thr	tgg Trp	96
cag Gln	act Thr	tgc Cys 35	acc Thr	agc Ser	agc Ser	ggc Gly 40	tgc Cys	acc Thr	acc Thr	aac Asn	cag Gln	ggc Gly 45	tcc Ser	atc Ile	gtc Val	144
gtt Val	gac Asp 50	gcc Ala	aac Asn	tgg Trp	cgc Arg	tgg Trp 55	gtc Val	cac His	aac Asn	acc Thr	gag Glu 60	ggc Gly	tac Tyr	acc Thr	aac Asn	192
tgc Cys 65	tac Tyr	acg Thr	ggc Gly	aac Asn	acc Thr 70	tgg Trp	aac Asn	gcc Ala	gac Asp	tac Tyr 75	tgc Cys	acc Thr	gac Asp	aac Asn	acc Thr 80	240
gag Glu	tgc Cys	gcc Ala	tcc Ser	aac Asn 85	tgc Cys	gcc Ala	ctc Leu	gac Asp	ggc Gly 90	gcc Ala	gac Asp	tac Tyr	tct Ser	ggc Gly 95	acc Thr	288
tac Tyr	ggc Gly	gct Ala	acc Thr 100	acc Thr	tcc Ser	ggc Gly	gac Asp	tgc Ser 105	ctg Leu	cgc Arg	ctg Leu	aac Asn	ttc Phe 110	atc Ile	acc Thr	336
aac Asn	ggc Gly	cag Gln 115	cag Gln	aag Lys	aac Asn	att Ile	ggc Gly 120	tcc Ser	cgc Arg	atg Met	tac Tyr	ctc Leu 125	atg Met	cag Gln	gat Asp	384
gac Asp	gag Glu 130	acc Thr	tac Tyr	gcc Ala	gtc Val	cac His 135	aag Lys	ctc Leu	ctc Leu	aac Asn	aag Lys 140	gag Glu	ttc Phe	acc Thr	ttc Phe	432
gac Asp 145	gtc Val	gac Asp	acc Thr	tcc Ser	aag Lys 150	ctg Leu	cct Pro	tgc Cys	ggc Gly 155	ctc Leu 155	aac Asn	ggt Gly	gcc Ala	gtc Val	tac Tyr 160	480
ttc Phe	gtc Val	tcc Ser	atg Met	gac Asp 165	gct Ala	gac Asp	ggt Gly	ggc Gly	atg Met 170	gcc Ala	aag Lys	ttc Phe	ccc Pro	gac Asp 175	aac Asn	528
aag Lys	gcc Ala	ggc Gly 180	ggc Ala	aag Lys	tac Tyr	ggt Gly	acc Thr	ggt Gly 185	tac Tyr	tgc Cys	gac Asp	tgc Ser	cag Gln 190	tgc Cys	ccc Pro	576
cgt Arg	gac Asp	ctc Leu 195	aag Lys	ttc Phe	atc Ile	gac Asp	ggc Gly 200	aag Lys	gcc Ala	aac Asn	gtc Val	gag Glu 205	ggc Gly	tgg Trp	gtc Val	624
ccg Pro	tcc Ser 210	gag Glu	aac Asn	gac Asp	tcc Ser	aac Asn 215	gct Ala	ggt Gly	gtc Val	ggc Gly 220	aac Asn	ctt Leu	ggc Gly	tct Ser	tgc Cys	672

<210> 38

<211> 454

<212> PRT

<213> *Diplodia gossypina*

<400> 38

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

	<210> 39
	<211> 1377
	<212> DNA
5	<213> Trichophaea saccata
	<220>
	<221> CDS
	<222> (1)..(1377)
10	<223>
	<400> 39

atg Met 1	caa Gln	cgc Arg	ctt Leu	ctc Leu 5	gtt Val	ctt Leu	ctc Leu	acc Thr 10	tcc Ser	ctt Leu	ctc Leu	gct Ala	ttc Phe 15	acc Thr	tat Tyr	48
ggc Gly	caa Gln	caa Gln	gtt Val 20	ggc Gly	act Thr	caa Gln	cag Gln	gcc Ala 25	gaa Glu	gtc Val	cac His	ccc Pro	tgc Ser 30	atg Met	acc Thr	96
tgg Trp	cag Gln	cag Gln 35	tgt Cys	aca Thr	aag Lys	tcc Ser	ggc Gly 40	ggc Gly	tgc Cys	acc Thr	acg Thr	aag Lys 45	aac Asn	ggc Gly	aaa Lys	144
gtc Val 50	gtg Val	atc Ile	gat Asp	gcc Ala	aac Asn	tgg Trp 55	cgt Arg	tgg Trp	gta Val	cac His	aat Asn 60	gtc Val	ggc Gly	ggc Gly	tac Tyr	192
acc Thr 65	aat Asn	tgc Cys	tac Tyr	act Thr	ggc Gly 70	aac Asn	acc Thr	tgg Trp	gac Asp	agt Ser 75	tgc Ser	ctt Leu	tgt Cys	ccc Pro	gac Asp 80	240
gat Asp	gtc Val	acc Thr	tgc Cys	gcg Ala 85	aag Lys	aat Asn	tgc Cys	gct Ala	ctt Leu 90	gat Asp	ggc Gly	gcg Ala	gac Asp	tac Tyr 95	tct Ser	288
ggc Gly	act Thr	tat Tyr	gga Gly 100	gtt Val	act Thr	gcg Ala	ggc Gly	ggg Gly 105	aat Asn	tgc Ser	ttg Leu	aag Lys	ctc Leu 110	acc Thr	ttc Phe	336
gtc Val	act Thr	aag Lys 115	ggt Gly	caa Gln	tac Tyr	tct Ser	act Thr 120	aat Asn	gtg Val	ggc Gly	tgc Ser	cga Arg 125	ttg Leu	tat Tyr	atg Met	384
ctc Leu	gcc Ala 130	gac Asp	gac Asp	agc Ser	aca Thr	tac Tyr 135	cag Gln	atg Met	tat Tyr	aat Asn	ctg Leu 140	ctg Leu	aac Asn	cag Gln	gag Glu	432
ttt Phe 145	acg Thr	ttc Phe	gac Asp	gtt Val	gat Asp 150	gtt Val	tct Ser	aat Asn	ctt Leu	cct Pro 155	tgt Cys	ggg Gly	ctt Leu	aac Asn	ggg Gly 160	480
gct Ala	ctg Leu	tat Tyr	ttc Phe	gtc Val 165	tgc Ser	atg Met	gat Asp	aag Lys	gat Asp 170	ggt Gly	ggg Gly	atg Met	tgc Ser	aag Lys 175	tac Tyr	528
tct Ser	ggg Gly	aac Asn	aag Lys 180	gct Ala	ggt Gly	gcc Ala	aag Lys	tat Tyr 185	gga Gly	act Thr	ggg Gly	tac Tyr	tgc Cys 190	gac Asp	tcc Ser	576
cag Gln	tgt Cys	ccc Pro 195	cgc Arg	gat Asp	ctc Leu	aag Lys	ttc Phe 200	atc Ile	aat Asn	gga Gly	cag Gln	ggc Gly 205	aac Asn	gtt Val	gaa Glu	624
ggc Gly	tgg Trp 210	aag Lys	cca Pro	tcc Ser	tca Ser	aat Asn 215	gat Asp	gcc Ala	aac Asn	gca Ala	ggc Gly 220	gtc Val	ggg Gly	gga Gly	cac His	672
ggt Gly 225	tcc Ser	tgc Cys	tgc Cys	gca Ala	gag Glu 230	atg Met	gat Asp	gtt Val	tgg Trp	gag Glu 235	gcc Ala	aat Asn	tcc Ser	atc Ile	tcc Ser 240	720
gcg Ala	gcc Ala	gta Val	aca Thr	ccg Pro 245	cac His	tgc Ser	tgc Cys	tcc Ser	aca Thr 250	acc Thr	agc Ser	cag Gln	acg Thr	atg Met 255	tgc Cys	768
aac Asn	ggc Gly	gac Asp	tcc Ser	tgc Cys	ggc Gly	ggt Gly	acc Thr	tac Tyr	tca Ser	gcc Ala	aca Thr	cga Arg	tac Tyr	gct Ala	ggt Gly	816

260	265	270	
gtc tgc gat ccc gat ggc tgc gac ttc aac tcc tac cgt atg ggc gac Val Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asp	275	280	285
acg acc ttc tac ggc aag gga aag acg gtc gat acc agc tcc aag ttc Thr Thr Phe Tyr Gly Lys Gly Lys Thr Val Asp Thr Ser Ser Lys Phe	290	295	300
acg gtc gtg acc cag ttc atc acc gac act gga acc gcc tcc ggc tcg Thr Val Val Thr Gln Phe Ile Thr Asp Thr Gly Thr Ala Ser Gly Ser	305	310	315
ctc acg gag atc cgc cgc ttc tac gtc cag aac gga aag ttg atc ccc Leu Thr Glu Ile Arg Arg Phe Tyr Val Gln Asn Gly Lys Leu Ile Pro	325	330	335
aac tcc cag tcg aag atc tcg ggc gtc act ggc aac tcc atc acc tct Asn Ser Gln Ser Lys Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Ser	340	345	350
gct ttc tgc gac gct cag aag gcg gct ttc ggc gat aac tac acg ttc Ala Phe Cys Asp Ala Gln Lys Ala Ala Phe Gly Asp Asn Tyr Thr Phe	355	360	365
aag gac aag ggc ggc ttc gca tcc atg act act gct atg aag aac gga Lys Asp Lys Gly Gly Phe Ala Ser Met Thr Thr Ala Met Lys Asn Gly	370	375	380
atg gtc ctg gtt atg agt ctt tgg gat gac cac tac gcc aat atg ctc Met Val Leu Val Met Ser Leu Trp Asp Asp His Tyr Ala Asn Met Leu	385	390	395
tgg ctt gat agc gac tat ccc act aac gcg gac tcc tcc aag ccg ggt Trp Leu Asp Ser Asp Tyr Pro Thr Asn Ala Asp Ser Ser Lys Pro Gly	405	410	415
gtt gct cgt ggc acc tgc ccg act tct tcc ggc gtg ccc tcg gat gtc Val Ala Arg Gly Thr Cys Pro Thr Ser Ser Gly Val Pro Ser Asp Val	420	425	430
gag act aac aat gca agc gct tcg gtc acg tac tcc aac att aga ttt Glu Thr Asn Asn Ala Ser Ala Ser Val Thr Tyr Ser Asn Ile Arg Phe	435	440	445
gga gat ctc aat tcc act tac acc gcc cag taa Gly Asp Leu Asn Ser Thr Tyr Thr Ala Gln	450	455	

<210> 40

<211> 458

<212> PRT

<213> Trichophaea saccata

<400> 40

Met Gln Arg Leu Leu Val Leu Leu Thr Ser Leu Leu Ala Phe Thr Tyr
1 5 10 15

Gly Gln Gln Val Gly Thr Gln Gln Ala Glu Val His Pro Ser Met Thr
20 25 30

<210> 41

<211> 1353

<212> DNA

<213> Myceliophthora thermophila

<220>

<221> CDS

<222> (1)..(1353)

<223>

<400> 41

atg	aag	cag	tac	ctc	cag	tac	ctc	gcg	gcg	acc	ctg	ccc	ctg	gtg	ggc	48
Met	Lys	Gln	Tyr	Leu	Gln	Tyr	Leu	Ala	Ala	Thr	Leu	Pro	Leu	Val	Gly	
1				5				10					15			
ctg	gcc	acg	gcc	cag	cag	gcg	ggt	aac	ctg	cag	acc	gag	act	cac	ccc	96
Leu	Ala	Thr	Ala	Gln	Gln	Ala	Gly	Asn	Leu	Gln	Thr	Glu	Thr	His	Pro	
			20				25					30				

agg	ctc	act	tgg	tcc	aag	tgc	acg	gcc	ccg	gga	tcc	tgc	caa	cag	gtc	144
Arg	Leu	Thr	Trp	Ser	Lys	Cys	Thr	Ala	Pro	Gly	Ser	Cys	Gln	Gln	Val	
		35					40					45				
aac	ggc	gag	gtc	gtc	atc	gac	tcc	aac	tgg	cgc	tgg	gtg	cac	gac	gag	192
Asn	Gly	Glu	Val	Val	Ile	Asp	Ser	Asn	Trp	Arg	Trp	Val	His	Asp	Glu	
	50					55					60					
aac	gcg	cag	aac	tgc	tac	gac	ggc	aac	cag	tgg	acc	aac	gct	tgc	agc	240
Asn	Ala	Gln	Asn	Cys	Tyr	Asp	Gly	Asn	Gln	Trp	Thr	Asn	Ala	Cys	Ser	
	65				70					75					80	
tct	gcc	acc	gac	tgc	gcc	gag	aat	tgc	gcg	ctc	gag	ggt	gcc	gac	tac	288
Ser	Ala	Thr	Asp	Cys	Ala	Glu	Asn	Cys	Ala	Leu	Glu	Gly	Ala	Asp	Tyr	
				85					90					95		
cag	ggc	acc	tat	ggc	gcc	tgc	acc	agc	ggc	aat	gcc	ctg	acg	ctc	acc	336
Gln	Gly	Thr	Tyr	Gly	Ala	Ser	Thr	Ser	Gly	Asn	Ala	Leu	Thr	Leu	Thr	
			100					105					110			
ttc	gtc	act	aag	cac	gag	tac	ggc	acc	aac	att	ggc	tgc	cgc	ctc	tac	384
Phe	Val	Thr	Lys	His	Glu	Tyr	Gly	Thr	Asn	Ile	Gly	Ser	Arg	Leu	Tyr	
		115					120					125				
ctc	atg	aac	ggc	gcg	aac	aag	tac	cag	atg	ttc	acc	ctc	aag	ggc	aac	432
Leu	Met	Asn	Gly	Ala	Asn	Lys	Tyr	Gln	Met	Phe	Thr	Leu	Lys	Gly	Asn	
	130					135					140					
gag	ctg	gcc	ttc	gac	gtc	gac	ctc	tgc	gcc	gtc	gag	tgc	ggc	ctc	aac	480
Glu	Leu	Ala	Phe	Asp	Val	Asp	Leu	Ser	Ala	Val	Glu	Cys	Gly	Leu	Asn	
	145				150					155					160	
agc	gcc	ctc	tac	ttc	gtg	gcc	atg	gag	gag	gat	ggc	ggt	gtg	tgc	agc	528
Ser	Ala	Leu	Tyr	Phe	Val	Ala	Met	Glu	Glu	Asp	Gly	Gly	Val	Ser	Ser	
				165				170						175		
tac	ccg	acc	aac	acg	gcc	ggt	gct	aag	ttc	ggc	act	ggg	tac	tgc	gac	576
Tyr	Pro	Thr	Asn	Thr	Ala	Gly	Ala	Lys	Phe	Gly	Thr	Gly	Tyr	Cys	Asp	
			180					185					190			
gcc	caa	tgc	gca	cgc	gac	ctc	aag	ttc	gtc	ggc	ggc	aag	ggc	aac	atc	624
Ala	Gln	Cys	Ala	Arg	Asp	Leu	Lys	Phe	Val	Gly	Gly	Lys	Gly	Asn	Ile	
		195					200					205				
gag	ggc	tgg	aag	ccg	tcc	acc	aac	gat	gcc	aat	gcc	ggt	gtc	ggt	cct	672
Glu	Gly	Trp	Lys	Pro	Ser	Thr	Asn	Asp	Ala	Asn	Ala	Gly	Val	Gly	Pro	
	210					215					220					
tat	ggc	ggg	tgc	tgc	gct	gag	atc	gac	gtc	tgg	gag	tgc	aac	aag	tat	720
Tyr	Gly	Gly	Cys	Cys	Ala	Glu	Ile	Asp	Val	Trp	Glu	Ser	Asn	Lys	Tyr	
	225				230					235					240	
gct	ttc	gct	ttc	acc	ccg	cac	ggt	tgc	gag	aac	cct	aaa	tac	cac	gtc	768
Ala	Phe	Ala	Phe	Thr	Pro	His	Gly	Cys	Glu	Asn	Pro	Lys	Tyr	His	Val	
				245					250					255		
tgc	gag	acc	acc	aac	tgc	ggt	ggc	acc	tac	tcc	gag	gac	cgc	ttc	gct	816
Cys	Glu	Thr	Thr	Asn	Cys	Gly	Gly	Thr	Tyr	Ser	Glu	Asp	Arg	Phe	Ala	
			260					265					270			
ggt	gac	tgc	gat	gcc	aac	ggc	tgc	gac	tac	aac	ccc	tac	cgc	atg	ggc	864
Gly	Asp	Cys	Asp	Ala	Asn	Gly	Cys	Asp	Tyr	Asn	Pro	Tyr	Arg	Met	Gly	
		275					280					285				
aac	cag	gac	ttc	tac	ggt	ccc	ggc	ttg	acg	gtc	gat	acc	agc	aag	aag	912
Asn	Gln	Asp	Phe	Tyr	Gly	Pro	Gly	Leu	Thr	Val	Asp	Thr	Ser	Lys	Lys	
	290					295					300					

ttc acc gtc gtc agc cag ttc gag gag aac aag ctc acc cag ttc ttc Phe Thr Val Val Ser Gln Phe Glu Glu Asn Lys Leu Thr Gln Phe Phe 305 310 315 320	960
gtc cag gac ggc aag aag att gag atc ccc ggc ccc aag gtc gag ggc Val Gln Asp Gly Lys Lys Ile Glu Ile Pro Gly Pro Lys Val Glu Gly 325 330 335	1008
atc gat gcg gac agc gcc gct atc acc cct gag ctg tgc agt gcc ctg Ile Asp Ala Asp Ser Ala Ala Ile Thr Pro Glu Leu Cys Ser Ala Leu 340 345 350	1056
ttc aag gcc ttc gat gac cgt gac cgc ttc tcg gag gtt ggc ggc ttc Phe Lys Ala Phe Asp Asp Arg Asp Arg Phe Ser Glu Val Gly Gly Phe 355 360 365	1104
gat gcc atc aac acg gcc ctg agc act ccc atg gtc ctc gtc atg tcc Asp Ala Ile Asn Thr Ala Leu Ser Thr Pro Met Val Leu Val Met Ser 370 375 380	1152
atc tgg gat gat cac tac gcc aat atg ctc tgg ctc gac tcg agc tac Ile Trp Asp Asp His Tyr Ala Asn Met Leu Trp Leu Asp Ser Ser Tyr 385 390 395 400	1200
ccc cct gag aag gct ggc cag cct ggc ggt gac cgt ggc ccg tgt cct Pro Pro Glu Lys Ala Gly Gln Pro Gly Gly Asp Arg Gly Pro Cys Pro 405 410 415	1248
cag gac tct ggc gtc ccg gcc gac gtt gag gct cag tac cct aat gcc Gln Asp Ser Gly Val Pro Ala Asp Val Glu Ala Gln Tyr Pro Asn Ala 420 425 430	1296
aag gtc atc tgg tcc aac atc cgc ttc ggc ccc atc ggc tcg act gtc Lys Val Ile Trp Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser Thr Val 435 440 445	1344
aac gtc taa Asn Val 450	1353

<210> 42

<211> 450

<212> PRT

<213> Myceliophthora thermophila

<400> 42

Met	Lys	Gln	Tyr	Leu	Gln	Tyr	Leu	Ala	Ala	Thr	Leu	Pro	Leu	Val	Gly
1				5					10					15	
Leu	Ala	Thr	Ala	Gln	Gln	Ala	Gly	Asn	Leu	Gln	Thr	Glu	Thr	His	Pro
			20					25					30		
Arg	Leu	Thr	Trp	Ser	Lys	Cys	Thr	Ala	Pro	Gly	Ser	Cys	Gln	Gln	Val
		35					40					45			
Asn	Gly	Glu	Val	Val	Ile	Asp	Ser	Asn	Trp	Arg	Trp	Val	His	Asp	Glu
	50					55					60				

Asn Ala Gln Asn Cys Tyr Asp Gly Asn Gln Trp Thr Asn Ala Cys Ser
 65 70 75 80
 Ser Ala Thr Asp Cys Ala Glu Asn Cys Ala Leu Glu Gly Ala Asp Tyr
 85 90 95
 Gln Gly Thr Tyr Gly Ala Ser Thr Ser Gly Asn Ala Leu Thr Leu Thr
 100 105 110
 Phe Val Thr Lys His Glu Tyr Gly Thr Asn Ile Gly Ser Arg Leu Tyr
 115 120 125
 Leu Met Asn Gly Ala Asn Lys Tyr Gln Met Phe Thr Leu Lys Gly Asn
 130 135 140
 Glu Leu Ala Phe Asp Val Asp Leu Ser Ala Val Glu Cys Gly Leu Asn
 145 150 155 160
 Ser Ala Leu Tyr Phe Val Ala Met Glu Glu Asp Gly Gly Val Ser Ser
 165 170 175
 Tyr Pro Thr Asn Thr Ala Gly Ala Lys Phe Gly Thr Gly Tyr Cys Asp
 180 185 190
 Ala Gln Cys Ala Arg Asp Leu Lys Phe Val Gly Gly Lys Gly Asn Ile
 195 200 205
 Glu Gly Trp Lys Pro Ser Thr Asn Asp Ala Asn Ala Gly Val Gly Pro
 210 215 220
 Tyr Gly Gly Cys Cys Ala Glu Ile Asp Val Trp Glu Ser Asn Lys Tyr
 225 230 235 240
 Ala Phe Ala Phe Thr Pro His Gly Cys Glu Asn Pro Lys Tyr His Val
 245 250 255
 Cys Glu Thr Thr Asn Cys Gly Gly Thr Tyr Ser Glu Asp Arg Phe Ala
 260 265 270
 Gly Asp Cys Asp Ala Asn Gly Cys Asp Tyr Asn Pro Tyr Arg Met Gly
 275 280 285
 Asn Gln Asp Phe Tyr Gly Pro Gly Leu Thr Val Asp Thr Ser Lys Lys
 290 295 300
 Phe Thr Val Val Ser Gln Phe Glu Glu Asn Lys Leu Thr Gln Phe Phe
 305 310 315 320
 Val Gln Asp Gly Lys Lys Ile Glu Ile Pro Gly Pro Lys Val Glu Gly
 325 330 335
 Ile Asp Ala Asp Ser Ala Ala Ile Thr Pro Glu Leu Cys Ser Ala Leu

	340		345		350										
Phe	Lys	Ala	Phe	Asp	Asp	Arg	Asp	Arg	Phe	Ser	Glu	Val	Gly	Gly	Phe
		355					360					365			
Asp	Ala	Ile	Asn	Thr	Ala	Leu	Ser	Thr	Pro	Met	Val	Leu	Val	Met	Ser
	370					375					380				
Ile	Trp	Asp	Asp	His	Tyr	Ala	Asn	Met	Leu	Trp	Leu	Asp	Ser	Ser	Tyr
385					390					395					400
Pro	Pro	Glu	Lys	Ala	Gly	Gln	Pro	Gly	Gly	Asp	Arg	Gly	Pro	Cys	Pro
				405					410					415	
Gln	Asp	Ser	Gly	Val	Pro	Ala	Asp	Val	Glu	Ala	Gln	Tyr	Pro	Asn	Ala
			420					425					430		
Lys	Val	Ile	Trp	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	Gly	Ser	Thr	Val
		435					440					445			
Asn	Val														
	450														

<210> 43
 <211> 1341
 <212> DNA
 <213> xylaria hypoxylon

<220>
 <221> CDS
 <222> (1)..(1341)
 <223>

<400> 43

atg	ttg	tcc	ctc	gcc	gtg	tcg	gcc	gcc	ctt	ctc	ggg	ctc	gcg	tct	gcc	48
Met	Leu	Ser	Leu	Ala	Val	Ser	Ala	Ala	Leu	Leu	Gly	Leu	Ala	Ser	Ala	
1				5					10					15		
cag	cag	gtt	gga	aag	gag	caa	tct	gag	act	cac	cct	aag	ctg	tct	tgg	96
Gln	Gln	Val	Gly	Lys	Glu	Gln	Ser	Glu	Thr	His	Pro	Lys	Leu	Ser	Trp	
			20					25					30			
aag	aag	tgc	acc	agc	ggt	ggt	tcc	tgc	acc	cag	acc	aac	gct	gag	gtg	144
Lys	Lys	Cys	Thr	Ser	Gly	Gly	Ser	Cys	Thr	Gln	Thr	Asn	Ala	Glu	Val	
		35					40					45				
acc	atc	gac	tct	aac	tgg	cga	tgg	ctt	cac	tct	ctc	gaa	ggc	act	gag	192
Thr	Ile	Asp	Ser	Asn	Trp	Arg	Trp	Leu	His	Ser	Leu	Glu	Gly	Thr	Glu	
	50					55					60					
aac	tgc	tac	gat	ggt	aac	aag	tgg	acc	tcg	cag	tgc	agc	act	ggc	gag	240

Asn 65	Cys	Tyr	Asp	Gly	Asn 70	Lys	Trp	Thr	Ser	Gln 75	Cys	Ser	Thr	Gly	Glu 80	
gac Asp	tgc Cys	gcc Ala	acc Thr	aag Lys 85	tgc Cys	gcc Ala	atc Ile	gag Glu	ggg Gly 90	gcc Ala	gac Asp	tac Tyr	agc Ser	aag Lys 95	acc Thr	288
tac Tyr	ggg Gly	gcc Ala	tct Ser 100	act Thr	agc Ser	ggc Gly	gat Asp	gct Ala 105	ctt Leu	acc Thr	ctc Leu	aag Lys	ttc Phe 110	ctg Leu	acc Thr	336
aag Lys	cac His	gag Glu 115	tac Tyr	gga Gly	acc Thr	aac Asn	atc Ile 120	ggc Gly	tcc Ser	cga Arg	ttc Phe	tac Tyr 125	ctt Leu	atg Met	aat Asn	384
ggg Gly	gcc Ala 130	gac Asp	aag Lys	tac Tyr	cag Gln	acc Thr 135	ttc Phe	gac Asp	ctc Leu	aag Lys	ggg Gly 140	aac Asn	gag Glu	ttc Phe	acc Thr	432
ttc Phe 145	gat Asp	gtc Val	gac Asp	ctg Leu	tcc Ser 150	acc Thr	gtc Val	gac Asp	tgt Cys	ggg Gly 155	ctt Leu	aac Asn	gcc Ala	gct Ala	ctt Leu 160	480
tac Tyr	ttc Phe	gtc Val	gcc Ala	atg Met 165	gag Glu	gaa Glu	gac Asp	ggg Gly 170	atg Met	gct Ala	agc Ser	tac Tyr	ccc Pro 175	aac Asn		528
aac Asn	aag Lys	gcc Ala	ggg Gly 180	gcc Ala	aag Lys	tac Tyr	ggg Gly 185	ggg Gly	tac Tyr	tgt Cys	gac Asp	gct Ala 190	cag Gln	tgt Cys		576
gcc Ala	cgt Arg	gac Asp 195	tgg Leu	aag Lys	ttc Phe	gtc Val	ggg Gly 200	ggc Gly	aag Lys	ggc Gly	aac Asn	ggt Val 205	gag Glu	gga Gly	tgg Trp	624
gag Glu 210	cca Pro	tcc Ser	acc Thr	aac Asn	gac Asp	gac Asp 215	aac Asn	ggc Ala	ggg Gly	ggt Val	ggc Gly 220	cct Pro	tac Tyr	ggg Gly	gcc Ala	672
tgc Cys 225	tgt Cys	gcc Ala	gaa Glu	atc Ile	gat Asp 230	gtc Val	tgg Trp	gag Glu	tcc Ser	aac Asn 235	tct Ser	cac His	tct Ser	ttc Phe	gct Ala 240	720
ttc Phe	acc Thr	cct Pro	cac His	cct Pro 245	tgc Cys	acc Thr	acc Thr	aac Asn	gaa Glu 250	tac Tyr	cac His	gtc Val	tgt Cys	gag Glu 255	cag Gln	768
gac Asp	gag Glu	tgt Cys	ggg Gly 260	ggg Gly	acc Thr	tac Tyr	tct Ser	gag Glu 265	gac Asp	cga Arg	ttc Phe	gct Ala	ggc Gly 270	aag Lys	tgt Cys	816
gat Asp	gcc Ala	aac Asn 275	ggg Gly	tgt Cys	gac Asp	tac Tyr	aac Asn 280	cct Pro	tac Tyr	cgc Arg	atg Met	ggg Gly 285	aac Asn	acc Thr	gac Asp	864
ttc Phe 290	tac Tyr	ggc Gly	cag Gln	ggc Gly	aag Lys	acc Thr 295	gtc Val	gac Asp	acc Thr	agc Ser	aag Lys 300	aaa Lys	ttc Phe	act Thr	ggt Val	912
gtc Val 305	acc Thr	cag Gln	ttc Phe	gcc Ala	gaa Glu 310	aac Asn	aag Lys	tgg Leu	act Thr	cag Gln 315	ttc Phe	ttc Phe	gtc Val	cag Gln	gac Asp 320	960
ggg Gly	aag Lys	aag Lys	att Ile	gag Glu 325	atc Ile	ccc Pro	ggg Gly	ccc Pro	aag Lys 330	att Ile	gac Asp	ggg Gly	ttc Phe	cct Pro 335	acc Thr	1008
gat Asp	agc Ser	gcc Ala	atc Ile	acc Thr	ccc Pro	gag Glu	tac Tyr	tgc Cys	act Thr	gcc Ala	gaa Glu	ttc Phe	aac Asn	ggt Val	cta Leu	1056

340										345					350					
gga	gac	cgt	gac	cgc	ttc	agt	gaa	ggt	ggt	ggc	ttc	gac	cag	ctc	aac	1104				
Gly	Asp	Arg	Asp	Arg	Phe	Ser	Glu	Val	Gly	Gly	Phe	Asp	Gln	Leu	Asn					
		355					360					365								
aac	gct	ctt	gac	gta	ccc	atg	gtc	ctt	gtc	atg	tcc	atc	tgg	gac	gac	1152				
Asn	Ala	Leu	Asp	Val	Pro	Met	Val	Leu	Val	Met	Ser	Ile	Trp	Asp	Asp					
	370					375					380									
cac	tac	gcc	aac	atg	ctt	tgg	ctc	gac	tcc	agc	tac	ccc	cct	gag	aag	1200				
His	Tyr	Ala	Asn	Met	Leu	Trp	Leu	Asp	Ser	Ser	Tyr	Pro	Pro	Glu	Lys					
385					390					395					400					
gct	ggc	cag	ccc	ggt	ggt	gac	cgt	ggt	gac	tgt	gcc	ccc	gac	tcc	ggt	1248				
Ala	Gly	Gln	Pro	Gly	Gly	Asp	Arg	Gly	Asp	Cys	Ala	Pro	Asp	Ser	Gly					
				405				410						415						
gtc	ccc	tcc	gac	gtc	gag	gcc	agc	atc	ccc	gat	gcc	aag	gtc	gtc	tgg	1296				
Val	Pro	Ser	Asp	Val	Glu	Ala	Ser	Ile	Pro	Asp	Ala	Lys	Val	Val	Trp					
			420					425					430							
tcc	aac	atc	cgc	ttc	ggt	ccc	atc	ggc	tct	act	gtc	gag	gtt	taa		1341				
Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	Gly	Ser	Thr	Val	Glu	Val							
		435					440					445								

<210> 44

<211> 446

<212> PRT

<213> Xylaria hypoxylon

<400> 44

Met Leu Ser Leu Ala Val Ser Ala Ala Leu Leu Gly Leu Ala Ser Ala
 1 5 10 15
 Gln Gln Val Gly Lys Glu Gln Ser Glu Thr His Pro Lys Leu Ser Trp
 20 25 30
 Lys Lys Cys Thr Ser Gly Gly Ser Cys Thr Gln Thr Asn Ala Glu Val
 35 40 45
 Thr Ile Asp Ser Asn Trp Arg Trp Leu His Ser Leu Glu Gly Thr Glu
 50 55 60
 Asn Cys Tyr Asp Gly Asn Lys Trp Thr Ser Gln Cys Ser Thr Gly Glu
 65 70 75 80
 Asp Cys Ala Thr Lys Cys Ala Ile Glu Gly Ala Asp Tyr Ser Lys Thr
 85 90 95
 Tyr Gly Ala Ser Thr Ser Gly Asp Ala Leu Thr Leu Lys Phe Leu Thr
 100 105 110
 Lys His Glu Tyr Gly Thr Asn Ile Gly Ser Arg Phe Tyr Leu Met Asn
 115 120 125

Gly Ala Asp Lys Tyr Gln Thr Phe Asp Leu Lys Gly Asn Glu Phe Thr
 130 135 140
 Phe Asp Val Asp Leu Ser Thr Val Asp Cys Gly Leu Asn Ala Ala Leu
 145 150 155 160
 Tyr Phe Val Ala Met Glu Glu Asp Gly Gly Met Ala Ser Tyr Pro Asn
 165 170 175
 Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala Gln Cys
 180 185 190
 Ala Arg Asp Leu Lys Phe Val Gly Gly Lys Gly Asn Val Glu Gly Trp
 195 200 205
 Glu Pro Ser Thr Asn Asp Asp Asn Ala Gly Val Gly Pro Tyr Gly Ala
 210 215 220
 Cys Cys Ala Glu Ile Asp Val Trp Glu Ser Asn Ser His Ser Phe Ala
 225 230 235 240
 Phe Thr Pro His Pro Cys Thr Thr Asn Glu Tyr His Val Cys Glu Gln
 245 250 255
 Asp Glu Cys Gly Gly Thr Tyr Ser Glu Asp Arg Phe Ala Gly Lys Cys
 260 265 270
 Asp Ala Asn Gly Cys Asp Tyr Asn Pro Tyr Arg Met Gly Asn Thr Asp
 275 280 285
 Phe Tyr Gly Gln Gly Lys Thr Val Asp Thr Ser Lys Lys Phe Thr Val
 290 295 300
 Val Thr Gln Phe Ala Glu Asn Lys Leu Thr Gln Phe Phe Val Gln Asp
 305 310 315 320
 Gly Lys Lys Ile Glu Ile Pro Gly Pro Lys Ile Asp Gly Phe Pro Thr
 325 330 335
 Asp Ser Ala Ile Thr Pro Glu Tyr Cys Thr Ala Glu Phe Asn Val Leu
 340 345 350
 Gly Asp Arg Asp Arg Phe Ser Glu Val Gly Gly Phe Asp Gln Leu Asn
 355 360 365
 Asn Ala Leu Asp Val Pro Met Val Leu Val Met Ser Ile Trp Asp Asp
 370 375 380
 His Tyr Ala Asn Met Leu Trp Leu Asp Ser Ser Tyr Pro Pro Glu Lys
 385 390 395 400

Ala Gly Gln Pro Gly Gly Asp Arg Gly Asp Cys Ala Pro Asp Ser Gly
405 410 415
Val Pro Ser Asp Val Glu Ala Ser Ile Pro Asp Ala Lys Val Val Trp
420 425 430
Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser Thr Val Glu Val
435 440 445

<210> 45
<211> 1584
<212> DNA
<213> Exidia glandulosa

<220>
<221> CDS
<222> (1)..(1584)
<223>

<400> 45

atg tac gcc aag ttc gct acc ctc gct gcc ctc gtg gca gct gcc agc Met Tyr Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Ala Ala Ser 1 5 10 15	48
gcc cag cag gca tgc aca ctc acc gcc gag aac cat ccc tcc atg act Ala Gln Gln Ala Cys Thr Leu Thr Ala Glu Asn His Pro Ser Met Thr 20 25 30	96
tgg tct aag tgt gcc gcc gga ggt agc tgc act tcg gtt tct ggt tca Trp Ser Lys Cys Ala Ala Gly Gly Ser Cys Thr Ser Val Ser Gly Ser 35 40 45	144
gtc acc atc gat gcc aac tgg cga tgg ctt cac cag ctc aac agc gcc Val Thr Ile Asp Ala Asn Trp Arg Trp Leu His Gln Leu Asn Ser Ala 50 55 60	192
acc aac tgc tac gac ggc aac aag tgg aac acc acc tac tgc agc aca Thr Asn Cys Tyr Asp Gly Asn Lys Trp Asn Thr Thr Tyr Cys Ser Thr 65 70 75 80	240
gat gct act tgc gct gct cag tgc tgt gtt gat ggc tca gac tat gct Asp Ala Thr Cys Ala Ala Gln Cys Cys Val Asp Gly Ser Asp Tyr Ala 85 90 95	288
ggc acc tac ggt gcc acc act agc ggt aac gct ctg aac ctc aag ttc Gly Thr Tyr Gly Ala Thr Thr Ser Gly Asn Ala Leu Asn Leu Lys Phe 100 105 110	336
gtc acc caa ggg tcc tat tct aag aac atc ggt tcc cgg ttg tac ctc Val Thr Gln Gly Ser Tyr Ser Lys Asn Ile Gly Ser Arg Leu Tyr Leu 115 120 125	384
atg gag tcg gat acc aag tat cag atg ttt caa ctg ctc ggc cag gag Met Glu Ser Asp Thr Lys Tyr Gln Met Phe Gln Leu Leu Gly Gln Glu 130 135 140	432

ttc Phe 145	act Thr	ttc Phe	gac Asp	gta Val 150	gat Asp 150	gtc Val	tcc Ser	aac Asn	ttg Leu	ggc Gly 155	tgc Cys	ggt Gly	ctc Leu	aac Asn	ggt Gly 160	480
gcc Ala	ctc Leu	tac Tyr	ttc Phe	gtc Val 165	agc Ser	atg Met	gac Asp	gct Ala	gac Asp 170	ggt Gly	ggc Gly	acg Thr	tcc Ser	aag Lys 175	tat Tyr	528
acc Thr	ggc Gly	aac Asn	aag Lys 180	gcc Ala	ggc Gly	gcc Ala	aag Lys	tat Tyr 185	ggc Gly	act Thr	ggc Gly	tac Tyr	tgc Cys 190	gac Asp	agc Ser	576
cag Gln	tgc Cys	ccg Pro 195	cgc Arg	gac Asp	ctg Leu	aag Lys	ttc Phe 200	atc Ile	aat Asn	ggt Gly	cag Gln	gcc Ala 205	aac Asn	gtc Val	gag Glu	624
ggc Gly	tgg Trp 210	act Thr	cct Pro	tcc Ser	acc Thr	aac Asn 215	gat Asp	gcc Ala	aac Asn	gcc Ala	ggc Gly 220	att Ile	ggc Gly	acc Thr	cac His	672
ggc Gly 225	tcc Ser	tgc Cys	tgt Cys	tcg Ser	gag Glu 230	atg Met	gac Asp	atc Ile	tgg Trp	gag Glu 235	gct Ala	aac Asn	aat Asn	gtt Val	gcc Ala 240	720
gct Ala	gcg Ala	tac Tyr	acc Thr	ccc Pro 245	cat His	cct Pro	tgc Cys	aca Thr	act Thr 250	atc Ile	ggc Gly	cag Gln	tcg Ser	atc Ile 255	tgc Cys	768
tcg Ser	ggc Gly	gat Asp	tct Ser 260	tgc Cys	gga Gly	gga Gly	acc Thr	tac Tyr 265	agc Ser	tct Ser	gac Asp	cgt Arg	tac Tyr 270	gcc Ala	ggt Gly	816
gtc Val	tgc Cys	gat Asp 275	cca Pro	gac Asp	ggt Gly	tgc Cys	gat Asp 280	ttc Phe	aac Asn	agc Ser	tac Tyr	cgc Arg 285	atg Met	ggc Gly	gac Asp	864
acg Thr	ggc Gly 290	ttc Phe	tac Tyr	ggc Gly	aag Lys	ggc Gly 295	ctg Leu	aca Thr	gtc Val	gac Asp 300	acg Thr	agc Ser	tcc Ser	aag Lys	ttc Phe	912
acc Thr 305	gtc Val	gtc Val	acc Thr	cag Gln	ttc Phe 310	ctc Leu	acc Thr	ggc Gly	tcc Ser	gac Asp 315	ggc Gly	aac Asn	ctt Leu	tcc Ser	gag Glu 320	960
atc Ile	aag Lys	cgc Arg	ttc Phe 325	tac Tyr	gtc Val	cag Gln	aac Asn	ggc Gly	aag Lys 330	gtc Val	att Ile	ccc Pro	aac Asn	tcg Ser 335	cag Gln	1008
tcc Ser	aag Lys	att Ile	gcc Ala 340	ggc Gly	gtc Val	agc Ser	ggc Gly	aac Asn 345	tcc Ser	atc Ile	acc Thr	acc Thr	gac Asp 350	ttc Phe	tgc Cys	1056
tcc Ser	gcc Ala	cag Gln 355	aag Lys	acc Thr	gcc Ala	ttc Phe	ggc Gly 360	gac Asp	acc Thr	aac Asn	gtc Val	ttc Phe 365	gcg Ala	caa Gln	aag Lys	1104
gga Gly 370	ggt Gly	ctc Leu	gcc Ala	ggg Gly	atg Met	ggc Gly 375	gcc Ala	gcc Ala	ctc Leu	aag Lys	gcc Ala 380	ggc Gly	atg Met	gtc Val	ctc Leu	1152
gtc Val 385	atg Met	tcc Ser	atc Ile	tgg Trp	gac Asp 390	gac Asp	cac His	gca Ala	gtc Val	aac Asn 395	atg Met	ctg Leu	tgg Trp	ctg Leu	gac Asp 400	1200
tcg Ser	acc Thr	tac Tyr	ccg Pro	acc Thr 405	gac Asp	agc Ser	acc Thr	aag Lys	ccc Pro 410	ggc Gly	gcg Ala	gcc Ala	cgc Arg	ggc Gly 415	acc Thr	1248

tgc	ccg	acc	acc	tcc	ggc	gtc	ccc	gcc	gac	gtc	gag	gcc	cag	gtc	ccc	1296
Cys	Pro	Thr	Thr	Ser	Gly	Val	Pro	Ala	Asp	Val	Glu	Ala	Gln	Val	Pro	
			420					425					430			
aac	tcg	aac	gtc	atc	tac	tcc	aac	atc	aag	gtc	ggc	ccc	atc	aac	tcg	1344
Asn	Ser	Asn	Val	Ile	Tyr	Ser	Asn	Ile	Lys	Val	Gly	Pro	Ile	Asn	Ser	
		435					440					445				
act	ttc	acc	ggc	ggc	act	tcc	ggc	ggc	ggc	ggc	agg	agg	agg	agg	tcc	1392
Thr	Phe	Thr	Gly	Gly	Thr	Ser	Gly	Gly	Gly	Gly	Ser	Ser	Ser	Ser	Ser	
	450					455					460					
acc	acc	atc	cga	acc	agg	acc	acc	agg	act	cgc	acc	acc	agg	acc	agg	1440
Thr	Thr	Ile	Arg	Thr	Ser	Thr	Thr	Ser	Thr	Arg	Thr	Thr	Ser	Thr	Ser	
					470					475					480	
acc	ggc	ccc	ggc	ggc	ggc	tcc	act	ggc	agg	gcc	ggc	gcc	gat	cac	tgg	1488
Thr	Ala	Pro	Gly	Gly	Gly	Ser	Thr	Gly	Ser	Ala	Gly	Ala	Asp	His	Trp	
			485						490					495		
gcg	caa	tgc	ggc	ggc	atc	ggc	tgg	act	ggc	ccc	acc	acc	tgc	aag	agg	1536
Ala	Gln	Cys	Gly	Gly	Ile	Gly	Trp	Thr	Gly	Pro	Thr	Thr	Cys	Lys	Ser	
			500					505					510			
ccg	tac	acc	tgc	aca	gcc	tcc	aac	ccg	tac	tac	tcc	cag	tgc	ttg	taa	1584
Pro	Tyr	Thr	Cys	Thr	Ala	Ser	Asn	Pro	Tyr	Tyr	Ser	Gln	Cys	Leu		
		515					520					525				

<210> 46

<211> 527

<212> PRT

<213> Exidia glandulosa

<400> 46

Met Tyr Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Ala Ala Ser
 1 5 10 15
 Ala Gln Gln Ala Cys Thr Leu Thr Ala Glu Asn His Pro Ser Met Thr
 20 25 30
 Trp Ser Lys Cys Ala Ala Gly Gly Ser Cys Thr Ser Val Ser Gly Ser
 35 40 45
 Val Thr Ile Asp Ala Asn Trp Arg Trp Leu His Gln Leu Asn Ser Ala
 50 55 60
 Thr Asn Cys Tyr Asp Gly Asn Lys Trp Asn Thr Thr Tyr Cys Ser Thr
 65 70 75 80
 Asp Ala Thr Cys Ala Ala Gln Cys Cys Val Asp Gly Ser Asp Tyr Ala
 85 90 95
 Gly Thr Tyr Gly Ala Thr Thr Ser Gly Asn Ala Leu Asn Leu Lys Phe
 100 105 110

Val Thr Gln Gly Ser Tyr Ser Lys Asn Ile Gly Ser Arg Leu Tyr Leu
 115 120 125
 Met Glu Ser Asp Thr Lys Tyr Gln Met Phe Gln Leu Leu Gly Gln Glu
 130 135 140
 Phe Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn Gly
 145 150 155 160
 Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Thr Ser Lys Tyr
 165 170 175
 Thr Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser
 180 185 190
 Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu
 195 200 205
 Gly Trp Thr Pro Ser Thr Asn Asp Ala Asn Ala Gly Ile Gly Thr His
 210 215 220
 Gly Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Asn Val Ala
 225 230 235 240
 Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Ser Ile Cys
 245 250 255
 Ser Gly Asp Ser Cys Gly Gly Thr Tyr Ser Ser Asp Arg Tyr Ala Gly
 260 265 270
 Val Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asp
 275 280 285
 Thr Gly Phe Tyr Gly Lys Gly Leu Thr Val Asp Thr Ser Ser Lys Phe
 290 295 300
 Thr Val Val Thr Gln Phe Leu Thr Gly Ser Asp Gly Asn Leu Ser Glu
 305 310 315 320
 Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Asn Ser Gln
 325 330 335
 Ser Lys Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Thr Asp Phe Cys
 340 345 350
 Ser Ala Gln Lys Thr Ala Phe Gly Asp Thr Asn Val Phe Ala Gln Lys
 355 360 365
 Gly Gly Leu Ala Gly Met Gly Ala Ala Leu Lys Ala Gly Met Val Leu
 370 375 380
 Val Met Ser Ile Trp Asp Asp His Ala Val Asn Met Leu Trp Leu Asp

385 390 395 400
 Ser Thr Tyr Pro Thr Asp Ser Thr Lys Pro Gly Ala Ala Arg Gly Thr
 405 410 415
 Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu Ala Gln Val Pro
 420 425 430
 Asn Ser Asn Val Ile Tyr Ser Asn Ile Lys Val Gly Pro Ile Asn Ser
 435 440 445
 Thr Phe Thr Gly Gly Thr Ser Gly Gly Gly Gly Ser Ser Ser Ser Ser
 450 455 460
 Thr Thr Ile Arg Thr Ser Thr Thr Ser Thr Arg Thr Thr Ser Thr Ser
 465 470 475 480
 Thr Ala Pro Gly Gly Gly Ser Thr Gly Ser Ala Gly Ala Asp His Trp
 485 490 495
 Ala Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro Thr Thr Cys Lys Ser
 500 505 510
 Pro Tyr Thr Cys Thr Ala Ser Asn Pro Tyr Tyr Ser Gln Cys Leu
 515 520 525

<210> 47

<211> 1368

<212> DNA

<213> Exidia glandulosa

<220>

<221> CDS

<222> (1)..(1368)

<223>

<400> 47

atg	tac	gcc	aag	ttc	gct	acc	ctc	gct	gcc	ctc	gtg	gca	gct	gcc	agc	48
Met	Tyr	Ala	Lys	Phe	Ala	Thr	Leu	Ala	Ala	Leu	Val	Ala	Ala	Ala	Ser	
1				5				10					15			
gcc	cag	cag	gca	tgc	aca	ctc	acc	gcc	gag	aac	cat	ccc	tcc	atg	act	96
Ala	Gln	Gln	Ala	Cys	Thr	Leu	Thr	Ala	Glu	Asn	His	Pro	Ser	Met	Thr	
			20					25					30			
tgg	tct	aag	tgt	gcc	gcc	gga	ggt	agc	tgc	act	tcg	gtt	tct	ggt	tca	144
Trp	Ser	Lys	Cys	Ala	Ala	Gly	Gly	Ser	Cys	Thr	Ser	Val	Ser	Gly	Ser	
		35				40						45				
gtc	acc	atc	gat	gcc	aac	tgg	cga	tgg	ctt	cac	cag	ctc	aac	agc	gcc	192

Val	Thr	Ile	Asp	Ala	Asn	Trp	Arg	Trp	Leu	His	Gln	Leu	Asn	Ser	Ala	
50						55					60					
acc	aac	tgc	tac	gac	ggc	aac	aag	tgg	aac	acc	acc	tac	tgc	agc	aca	240
Thr	Asn	Cys	Tyr	Asp	Gly	Asn	Lys	Trp	Asn	Thr	Thr	Tyr	Cys	Ser	Thr	
65					70					75					80	
gat	gct	act	tgc	gct	gct	cag	tgc	tgt	gtt	gat	ggc	tca	gac	tat	gct	288
Asp	Ala	Thr	Cys	Ala	Ala	Gln	Cys	Cys	Val	Asp	Gly	Ser	Asp	Tyr	Ala	
				85					90					95		
ggc	acc	tac	ggt	gcc	acc	act	agc	ggt	aac	gct	ctg	aac	ctc	aag	ttc	336
Gly	Thr	Tyr	Gly	Ala	Thr	Thr	Ser	Gly	Asn	Ala	Leu	Asn	Leu	Lys	Phe	
			100					105					110			
gtc	acc	caa	ggg	tcc	tat	tct	aag	aac	atc	ggt	tcc	cgg	ttg	tac	ctc	384
Val	Thr	Gln	Gly	Ser	Tyr	Ser	Lys	Asn	Ile	Gly	Ser	Arg	Leu	Tyr	Leu	
		115					120					125				
atg	gag	tgc	gat	acc	aag	tat	cag	atg	ttt	caa	ctg	ctc	ggc	cag	gag	432
Met	Glu	Ser	Asp	Thr	Lys	Tyr	Gln	Met	Phe	Gln	Leu	Leu	Gly	Gln	Glu	
	130					135					140					
ttc	act	ttc	gac	gta	gat	gtc	tcc	aac	ttg	ggc	tgc	ggc	ctc	aac	ggt	480
Phe	Thr	Phe	Asp	Val	Asp	Val	Ser	Asn	Leu	Gly	Cys	Gly	Leu	Asn	Gly	
				150						155					160	
gcc	ctc	tac	ttc	gtc	agc	atg	gac	gct	gac	ggt	ggc	acg	tcc	aag	tat	528
Ala	Leu	Tyr	Phe	Val	Ser	Met	Asp	Ala	Asp	Gly	Gly	Thr	Ser	Lys	Tyr	
				165				170						175		
acc	ggc	aac	aag	gcc	ggc	gcc	aag	tat	ggc	act	ggc	tac	tgc	gac	agc	576
Thr	Gly	Asn	Lys	Ala	Gly	Ala	Lys	Tyr	Gly	Thr	Gly	Tyr	Cys	Asp	Ser	
			180					185					190			
cag	tgc	ccg	cgc	gac	ctg	aag	ttc	atc	aat	ggt	cag	gcc	aac	gtc	gag	624
Gln	Cys	Pro	Arg	Asp	Leu	Lys	Phe	Ile	Asn	Gly	Gln	Ala	Asn	Val	Glu	
		195					200					205				
ggc	tgg	act	cct	tcc	acc	aac	gat	gcc	aac	gcc	ggc	att	ggc	acc	cac	672
Gly	Trp	Thr	Pro	Ser	Thr	Asn	Asp	Ala	Asn	Ala	Gly	Ile	Gly	Thr	His	
	210					215					220					
ggc	tcc	tgc	tgt	tgc	gag	atg	gac	atc	tgg	gag	gct	aac	aat	gtt	gcc	720
Gly	Ser	Cys	Cys	Ser	Glu	Met	Asp	Ile	Trp	Glu	Ala	Asn	Asn	Val	Ala	
	225				230					235					240	
gct	gcg	tac	acc	ccc	cat	cct	tgc	aca	act	atc	ggc	cag	tgc	atc	tgc	768
Ala	Ala	Tyr	Thr	Pro	His	Pro	Cys	Thr	Thr	Ile	Gly	Gln	Ser	Ile	Cys	
				245					250					255		
tgc	ggc	gat	tct	tgc	gga	gga	acc	tac	agc	tct	gac	cgt	tac	gcc	ggt	816
Ser	Gly	Asp	Ser	Cys	Gly	Gly	Thr	Tyr	Ser	Ser	Asp	Arg	Tyr	Ala	Gly	
			260					265					270			
gtc	tgc	gat	cca	gac	ggt	tgc	gat	ttc	aac	agc	tac	cgc	atg	ggc	gac	864
Val	Cys	Asp	Pro	Asp	Gly	Cys	Asp	Phe	Asn	Ser	Tyr	Arg	Met	Gly	Asp	
		275					280					285				
acg	ggc	ttc	tac	ggc	aag	ggc	ctg	aca	gtc	gac	acg	agc	tcc	aag	ttc	912
Thr	Gly	Phe	Tyr	Gly	Lys	Gly	Leu	Thr	Val	Asp	Thr	Ser	Ser	Lys	Phe	
	290					295					300					
acc	gtc	gtc	acc	cag	ttc	ctc	acc	ggc	tcc	gac	ggc	aac	ctt	tcc	gag	960
Thr	Val	Val	Thr	Gln	Phe	Leu	Thr	Gly	Ser	Asp	Gly	Asn	Leu	Ser	Glu	
	305			310				315							320	
atc	aag	cgc	ttc	tac	gtc	cag	aac	ggc	aag	gtc	att	ccc	aac	tgc	cag	1008
Ile	Lys	Arg	Phe	Tyr	Val	Gln	Asn	Gly	Lys	Val	Ile	Pro	Asn	Ser	Gln	

325								330				335					
tcc	aag	att	gcc	ggc	gtc	agc	ggc	aac	tcc	atc	acc	acc	gac	ttc	tgc	1056	
Ser	Lys	Ile	Ala	Gly	Val	Ser	Gly	Asn	Ser	Ile	Thr	Thr	Asp	Phe	Cys		
			340													350	
tcc	gcc	cag	aag	acc	gcc	ttc	ggc	gac	acc	aac	gtc	ttc	gcg	caa	aag	1104	
Ser	Ala	Gln	Lys	Thr	Ala	Phe	Gly	Asp	Thr	Asn	Val	Phe	Ala	Gln	Lys		
			355													365	
gga	ggt	ctc	gcc	ggg	atg	ggc	gcc	gcc	ctc	aag	gcc	ggc	atg	gtc	ctc	1152	
Gly	Gly	Leu	Ala	Gly	Met	Gly	Ala	Ala	Leu	Lys	Ala	Gly	Met	Val	Leu		
			370													380	
gtc	atg	tcc	atc	tgg	gac	gat	cac	tac	gcc	aac	atg	ctg	tgg	ctc	gac	1200	
Val	Met	Ser	Ile	Trp	Asp	Asp	His	Tyr	Ala	Asn	Met	Leu	Trp	Leu	Asp		
			385													400	
tcg	acc	tac	ccg	act	gac	gcc	tct	ccc	gat	gag	ccc	ggc	aag	ggc	cgc	1248	
Ser	Thr	Tyr	Pro	Thr	Asp	Ala	Ser	Pro	Asp	Glu	Pro	Gly	Lys	Gly	Arg		
			405													415	
ggc	acc	tgc	gac	acc	agc	tcg	ggt	gtt	cct	gct	gac	atc	gag	acc	agc	1296	
Gly	Thr	Cys	Asp	Thr	Ser	Ser	Gly	Val	Pro	Ala	Asp	Ile	Glu	Thr	Ser		
			420													430	
cag	gcc	agc	aac	tca	gtc	atc	tac	tcg	aac	atc	aag	ttc	gga	ccc	atc	1344	
Gln	Ala	Ser	Asn	Ser	Val	Ile	Tyr	Ser	Asn	Ile	Lys	Phe	Gly	Pro	Ile		
			435													445	
aac	tcg	acc	ttc	aag	gcg	tcc	taa										1368
Asn	Ser	Thr	Phe	Lys	Ala	Ser											
			450													455	

<210> 48

<211> 455

<212> PRT

<213> Exidia glandulosa

<400> 48

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

Gly Thr Tyr Gly Ala Thr Thr Ser Gly Asn Ala Leu Asn Leu Lys Phe
 100 105 110
 Val Thr Gln Gly Ser Tyr Ser Lys Asn Ile Gly Ser Arg Leu Tyr Leu
 115 120 125
 Met Glu Ser Asp Thr Lys Tyr Gln Met Phe Gln Leu Leu Gly Gln Glu
 130 135 140
 Phe Thr Phe Asp Val Asp Val Ser Asn Leu Gly Cys Gly Leu Asn Gly
 145 150 155 160
 Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Thr Ser Lys Tyr
 165 170 175
 Thr Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser
 180 185 190
 Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu
 195 200 205
 Gly Trp Thr Pro Ser Thr Asn Asp Ala Asn Ala Gly Ile Gly Thr His
 210 215 220
 Gly Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Asn Val Ala
 225 230 235 240
 Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Ser Ile Cys
 245 250 255
 Ser Gly Asp Ser Cys Gly Gly Thr Tyr Ser Ser Asp Arg Tyr Ala Gly
 260 265 270
 Val Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asp
 275 280 285
 Thr Gly Phe Tyr Gly Lys Gly Leu Thr Val Asp Thr Ser Ser Lys Phe
 290 295 300
 Thr Val Val Thr Gln Phe Leu Thr Gly Ser Asp Gly Asn Leu Ser Glu
 305 310 315 320
 Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Asn Ser Gln
 325 330 335
 Ser Lys Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Thr Asp Phe Cys
 340 345 350
 Ser Ala Gln Lys Thr Ala Phe Gly Asp Thr Asn Val Phe Ala Gln Lys
 355 360 365

Gly Gly Leu Ala Gly Met Gly Ala Ala Leu Lys Ala Gly Met Val Leu
370 375 380
Val Met Ser Ile Trp Asp Asp His Tyr Ala Asn Met Leu Trp Leu Asp
385 390 395 400
Ser Thr Tyr Pro Thr Asp Ala Ser Pro Asp Glu Pro Gly Lys Gly Arg
405 410 415
Gly Thr Cys Asp Thr Ser Ser Gly Val Pro Ala Asp Ile Glu Thr Ser
420 425 430
Gln Ala Ser Asn Ser Val Ile Tyr Ser Asn Ile Lys Phe Gly Pro Ile
435 440 445
Asn Ser Thr Phe Lys Ala Ser
450 455

<210> 49
<211> 1395
<212> DNA
<213> Poitrasia circinans

<220>
<221> CDS
<222> (1)..(1395)
<223>

<400> 49

atg cat cag act tcc gtt ctt tct tcg ctc tct ttg ctc ctc gca gcc Met His Gln Thr Ser Val Leu Ser Ser Leu Ser Leu Leu Leu Ala Ala 1 5 10 15	48
tcc ggt gcc cag cag gtc ggc acc cag aat gct gag act cac ccg agt Ser Gly Ala Gln Gln Val Gly Thr Gln Asn Ala Glu Thr His Pro Ser 20 25 30	96
ctg acc acc cag aag tgt acc acc gac ggc ggc tgc acc gac cag tcc Leu Thr Thr Gln Lys Cys Thr Thr Asp Gly Gly Cys Thr Asp Gln Ser 35 40 45	144
act gcc atc gtg ctt gac gcc aac tgg cgc tgg ctg cac acc acc gag Thr Ala Ile Val Leu Asp Ala Asn Trp Arg Trp Leu His Thr Thr Glu 50 55 60	192
ggc tac acc aac tgc tac act ggc cag gaa tgg gac acc gac atc tgc Gly Tyr Thr Asn Cys Tyr Thr Gly Gln Glu Trp Asp Thr Asp Ile Cys 65 70 75 80	240
tcc tcc ccg gag gct tgc gcc acc ggc tgc gct ctt gac ggt gcc gac Ser Ser Pro Glu Ala Cys Ala Thr Gly Cys Ala Leu Asp Gly Ala Asp 85 90 95	288

tac	gag	ggc	act	tac	ggc	att	acg	act	gac	ggc	aac	gct	ctt	tcc	atg	336
Tyr	Glu	Gly	Thr	Tyr	Gly	Ile	Thr	Thr	Asp	Gly	Asn	Ala	Leu	Ser	Met	
			100					105					110			
aag	ttt	gtc	acc	cag	ggc	tcg	cag	aag	aac	gtc	ggc	ggt	cgt	ggt	tac	384
Lys	Phe	Val	Thr	Gln	Gly	Ser	Gln	Lys	Asn	Val	Gly	Gly	Arg	Val	Tyr	
		115					120					125				
ctg	ctt	gct	ccc	gac	tcc	gaa	gat	gcg	tac	gag	ctc	ttc	aag	ttg	aag	432
Leu	Leu	Ala	Pro	Asp	Ser	Glu	Asp	Ala	Tyr	Glu	Leu	Phe	Lys	Leu	Lys	
		130				135					140					
aac	cag	gag	ttc	act	ttc	gac	gtt	gac	gtc	tcc	gac	ctc	ccc	tgc	ggc	480
Asn	Gln	Glu	Phe	Thr	Phe	Asp	Val	Asp	Val	Ser	Asp	Leu	Pro	Cys	Gly	
145					150					155					160	
ctg	aac	ggc	gcc	ctg	tac	ttc	tcc	gag	atg	gat	gaa	gat	ggt	ggc	atg	528
Leu	Asn	Gly	Ala	Leu	Tyr	Phe	Ser	Glu	Met	Asp	Glu	Asp	Gly	Gly	Met	
				165					170					175		
tcc	aag	tac	gag	aac	aac	aag	gcc	ggc	gcc	aag	tac	ggc	act	ggc	tac	576
Ser	Lys	Tyr	Glu	Asn	Asn	Lys	Ala	Gly	Ala	Lys	Tyr	Gly	Thr	Gly	Tyr	
			180					185					190			
tgc	gac	acg	cag	tgc	ccc	cac	gac	gtc	aag	ttc	atc	aac	ggc	gag	gcc	624
Cys	Asp	Thr	Gln	Cys	Pro	His	Asp	Val	Lys	Phe	Ile	Asn	Gly	Glu	Ala	
		195					200					205				
aac	att	ctc	aac	tgg	acc	aag	tcc	gag	acc	gac	gtc	aac	gcc	ggc	act	672
Asn	Ile	Leu	Asn	Trp	Thr	Lys	Ser	Glu	Thr	Asp	Val	Asn	Ala	Gly	Thr	
	210					215					220					
ggc	caa	tac	ggc	tcc	tgc	tgc	aac	gag	atg	gat	atc	tgg	gag	gcc	aac	720
Gly	Gln	Tyr	Gly	Ser	Cys	Cys	Asn	Glu	Met	Asp	Ile	Trp	Glu	Ala	Asn	
225					230					235					240	
tcg	cag	gcc	acc	gcc	gtc	act	ccc	cac	gtc	tgc	aac	gcc	gat	gtc	atc	768
Ser	Gln	Ala	Thr	Ala	Val	Thr	Pro	His	Val	Cys	Asn	Ala	Asp	Val	Ile	
				245					250					255		
ggc	cag	gtc	cgt	tgc	aac	ggc	acc	gac	tgc	ggt	gac	ggc	gac	aac	cgc	816
Gly	Gln	Val	Arg	Cys	Asn	Gly	Thr	Asp	Cys	Gly	Asp	Gly	Asp	Asn	Arg	
			260					265				270				
tac	ggc	ggc	gtc	tgc	gac	aag	gat	ggc	tgc	gac	tac	aac	ccc	tac	cgc	864
Tyr	Gly	Gly	Val	Cys	Asp	Lys	Asp	Gly	Cys	Asp	Tyr	Asn	Pro	Tyr	Arg	
		275					280					285				
atg	ggc	aac	gag	tcg	ttc	tac	ggc	tcc	aac	ggc	agc	acc	atc	gac	acc	912
Met	Gly	Asn	Glu	Ser	Phe	Tyr	Gly	Ser	Asn	Gly	Ser	Thr	Ile	Asp	Thr	
	290					295					300					
act	gcc	aag	ttc	acc	gtc	att	acg	cag	ttc	atc	acc	tcg	gac	aac	act	960
Thr	Ala	Lys	Phe	Thr	Val	Ile	Thr	Gln	Phe	Ile	Thr	Ser	Asp	Asn	Thr	
305					310					315					320	
tcg	act	ggc	gac	ctc	gtt	gag	atc	cgc	cgc	aag	tac	gtc	cag	gac	ggc	1008
Ser	Thr	Gly	Asp	Leu	Val	Glu	Ile	Arg	Arg	Lys	Tyr	Val	Gln	Asp	Gly	
				325				330						335		
acc	gtc	atc	gag	aac	tcg	ttc	gcc	gac	tac	gac	acc	ctg	gcc	acg	ttc	1056
Thr	Val	Ile	Glu	Asn	Ser	Phe	Ala	Asp	Tyr	Asp	Thr	Leu	Ala	Thr	Phe	
			340				345						350			
aac	tcc	atc	tcg	gac	gac	ttc	tgc	gac	gcc	cag	aag	acg	ctc	ttc	ggc	1104
Asn	Ser	Ile	Ser	Asp	Asp	Phe	Cys	Asp	Ala	Gln	Lys	Thr	Leu	Phe	Gly	
		355					360					365				

gac	gag	aac	gac	ttc	aag	acc	aag	ggc	ggc	att	gcc	cgc	atg	ggc	gag	1152
Asp	Glu	Asn	Asp	Phe	Lys	Thr	Lys	Gly	Gly	Ile	Ala	Arg	Met	Gly	Glu	
	370					375					380					
tcc	ttc	gag	cgc	ggc	atg	gtc	ctc	gtc	atg	agc	atc	tgg	gat	gac	cac	1200
Ser	Phe	Glu	Arg	Gly	Met	Val	Leu	Val	Met	Ser	Ile	Trp	Asp	Asp	His	
	385				390					395					400	
gcg	gcc	aac	gcc	ctc	tgg	ctc	gac	tcg	acc	tac	ccc	gtc	gac	ggc	gac	1248
Ala	Ala	Asn	Ala	Leu	Trp	Leu	Asp	Ser	Thr	Tyr	Pro	Val	Asp	Gly	Asp	
				405					410					415		
gcg	acc	aag	cct	ggc	atc	aag	cgc	ggc	cct	tgc	ggc	acc	gac	act	ggt	1296
Ala	Thr	Lys	Pro	Gly	Ile	Lys	Arg	Gly	Pro	Cys	Gly	Thr	Asp	Thr	Gly	
			420					425					430			
gtt	ccc	gcc	gac	gtc	gag	tcg	gag	tcg	ccc	gat	tcg	acc	gtc	atc	tac	1344
Val	Pro	Ala	Asp	Val	Glu	Ser	Glu	Ser	Pro	Asp	Ser	Thr	Val	Ile	Tyr	
		435					440					445				
tcc	aac	att	cgc	tac	gga	gac	att	ggc	tcc	acc	ttc	aac	gcc	acc	gct	1392
Ser	Asn	Ile	Arg	Tyr	Gly	Asp	Ile	Gly	Ser	Thr	Phe	Asn	Ala	Thr	Ala	
	450					455					460					
tag																1395

<210> 50

<211> 464

<212> PRT

<213> Poitrasia circinans

<400> 50

Met His Gln Thr Ser Val Leu Ser Ser Leu Ser Leu Leu Leu Ala Ala
 1 5 10 15
 Ser Gly Ala Gln Gln Val Gly Thr Gln Asn Ala Glu Thr His Pro Ser
 20 25 30
 Leu Thr Thr Gln Lys Cys Thr Thr Asp Gly Gly Cys Thr Asp Gln Ser
 35 40 45
 Thr Ala Ile Val Leu Asp Ala Asn Trp Arg Trp Leu His Thr Thr Glu
 50 55 60
 Gly Tyr Thr Asn Cys Tyr Thr Gly Gln Glu Trp Asp Thr Asp Ile Cys
 65 70 75 80
 Ser Ser Pro Glu Ala Cys Ala Thr Gly Cys Ala Leu Asp Gly Ala Asp
 85 90 95
 Tyr Glu Gly Thr Tyr Gly Ile Thr Thr Asp Gly Asn Ala Leu Ser Met
 100 105 110
 Lys Phe Val Thr Gln Gly Ser Gln Lys Asn Val Gly Gly Arg Val Tyr
 115 120 125

Leu Leu Ala Pro Asp Ser Glu Asp Ala Tyr Glu Leu Phe Lys Leu Lys
 130 135 140
 Asn Gln Glu Phe Thr Phe Asp Val Asp Val Ser Asp Leu Pro Cys Gly
 145 150 155 160
 Leu Asn Gly Ala Leu Tyr Phe Ser Glu Met Asp Glu Asp Gly Gly Met
 165 170 175
 Ser Lys Tyr Glu Asn Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr
 180 185 190
 Cys Asp Thr Gln Cys Pro His Asp Val Lys Phe Ile Asn Gly Glu Ala
 195 200 205
 Asn Ile Leu Asn Trp Thr Lys Ser Glu Thr Asp Val Asn Ala Gly Thr
 210 215 220
 Gly Gln Tyr Gly Ser Cys Cys Asn Glu Met Asp Ile Trp Glu Ala Asn
 225 230 235 240
 Ser Gln Ala Thr Ala Val Thr Pro His Val Cys Asn Ala Asp Val Ile
 245 250 255
 Gly Gln Val Arg Cys Asn Gly Thr Asp Cys Gly Asp Gly Asp Asn Arg
 260 265 270
 Tyr Gly Gly Val Cys Asp Lys Asp Gly Cys Asp Tyr Asn Pro Tyr Arg
 275 280 285
 Met Gly Asn Glu Ser Phe Tyr Gly Ser Asn Gly Ser Thr Ile Asp Thr
 290 295 300
 Thr Ala Lys Phe Thr Val Ile Thr Gln Phe Ile Thr Ser Asp Asn Thr
 305 310 315 320
 Ser Thr Gly Asp Leu Val Glu Ile Arg Arg Lys Tyr Val Gln Asp Gly
 325 330 335
 Thr Val Ile Glu Asn Ser Phe Ala Asp Tyr Asp Thr Leu Ala Thr Phe
 340 345 350
 Asn Ser Ile Ser Asp Asp Phe Cys Asp Ala Gln Lys Thr Leu Phe Gly
 355 360 365
 Asp Glu Asn Asp Phe Lys Thr Lys Gly Gly Ile Ala Arg Met Gly Glu
 370 375 380
 Ser Phe Glu Arg Gly Met Val Leu Val Met Ser Ile Trp Asp Asp His
 385 390 395 400

Ala Ala Asn Ala Leu Trp Leu Asp Ser Thr Tyr Pro Val Asp Gly Asp
405 410 415
Ala Thr Lys Pro Gly Ile Lys Arg Gly Pro Cys Gly Thr Asp Thr Gly
420 425 430
Val Pro Ala Asp Val Glu Ser Glu Ser Pro Asp Ser Thr Val Ile Tyr
435 440 445
Ser Asn Ile Arg Tyr Gly Asp Ile Gly Ser Thr Phe Asn Ala Thr Ala
450 455 460

<210> 51

<211> 1383

<212> DNA

<213> Coprinus cinereus

<220>

<221> CDS

<222> (1)..(1383)

<223>

<400> 51

atg ttc aag aaa gtc gcc ctc acc gct ctc tgc ttc ctc gcc gtc gca Met Phe Lys Lys Val Ala Leu Thr Ala Leu Cys Phe Leu Ala Val Ala 1 5 10 15	48
cag gcc caa cag gtc ggt cgc gaa gtc gct gaa aac cac ccc cgt ctc Gln Ala Gln Val Gly Arg Glu Val Ala Glu Asn His Pro Arg Leu 20 25 30	96
ccg tgg cag cgt tgc act cgc aac ggc gga tgc cag act gtc tcc aac Pro Trp Gln Arg Cys Thr Arg Asn Gly Gly Cys Gln Thr Val Ser Asn 35 40 45	144
ggt cag gtc gtc ctc gac gcc aac tgg cga tgg ctc cac gtc acc gac Gly Gln Val Val Leu Asp Ala Asn Trp Arg Trp Leu His Val Thr Asp 50 55 60	192
ggc tac acc aac tgc tac acc ggt aac tcc tgg aac agc acc gtc tgc Gly Tyr Thr Asn Cys Tyr Thr Gly Asn Ser Trp Asn Ser Thr Val Cys 65 70 75 80	240
tcc gac ccc acc acc tgc gct cag cga tgc gct ctc gag ggt gcc aac Ser Asp Pro Thr Thr Cys Ala Gln Arg Cys Ala Leu Glu Gly Ala Asn 85 90 95	288
tac cag caa acc tac ggt atc acc acc aac gga gac gcc ctc acc atc Tyr Gln Gln Thr Tyr Gly Ile Thr Thr Asn Gly Asp Ala Leu Thr Ile 100 105 110	336
aag ttc ctc acc cga tcc caa caa acc aac gtc ggt gct cgt gtc tac Lys Phe Leu Thr Arg Ser Gln Gln Thr Asn Val Gly Ala Arg Val Tyr 115 120 125	384

ctc Leu	atg Met 130	gag Glu	aac Asn	gag Glu	aac Asn	cga Arg 135	tac Tyr	cag Gln	atg Met	ttc Phe	aac Asn 140	ctc Leu	ctc Leu	aac Asn	aag Lys	432
gag Glu 145	ttc Phe	acc Thr	ttc Phe	gac Asp	gtt Val 150	gac Asp	gtc Val	tcc Ser	aag Lys	gtt Val 155	cct Pro	tgc Cys	ggt Gly	atc Ile	aac Asn 160	480
ggt Gly	gcc Ala	ctc Leu	tac Tyr	ttc Phe 165	atc Ile	cag Gln	atg Met	gac Asp	gcc Ala 170	gat Asp	ggt Gly	ggt Gly	atg Met	agc Ser 175	aag Lys	528
caa Gln	ccc Pro	aac Asn	aac Asn 180	agg Arg	gct Ala	ggt Gly	gct Ala	aag Lys 185	tac Tyr	ggt Gly	acc Thr	ggc Gly	tac Tyr 190	tgc Cys	gac Asp	576
tct Ser	cag Gln	tgc Cys 195	ccc Pro	cgt Arg	gac Asp	atc Ile	aag Lys 200	ttc Phe	att Ile	gac Asp	ggc Gly	gtg Val 205	gcc Ala	aac Asn	agc Ser	624
gcc Ala	gac Asp 210	tgg Trp	act Thr	cca Pro	tcc Ser	gag Glu 215	acc Thr	gat Asp	ccc Pro	aat Asn	gcc Ala 220	gga Gly	agg Arg	ggt Gly	cgc Arg	672
tac Tyr 225	ggc Gly	att Ile	tgc Cys	tgc Cys	gcc Ala 230	gag Glu	atg Met	gat Asp	atc Ile	tgg Trp 235	gag Glu	gcc Ala	aac Asn	tcc Ser	atc Ile 240	720
tcc Ser	aat Asn	gcc Ala	tac Tyr	acc Thr 245	ccc Pro	cac His	cct Pro	tgc Cys	cga Arg 250	acc Thr	cag Gln	aac Asn	gat Asp	ggt Gly 255	ggc Gly	768
tac Tyr	cag Gln	cgc Arg	tgc Cys 260	gag Glu	ggc Gly	cgc Arg	gac Asp	tgc Cys 265	aac Asn	cag Gln	cct Pro	cgc Arg	tat Tyr 270	gag Glu	ggt Gly	816
ctt Leu	tgc Cys	gat Asp 275	cct Pro	gat Asp	ggc Gly	tgt Cys	gac Asp 280	tac Tyr	aac Asn	ccc Pro	ttc Phe	cgc Arg 285	atg Met	ggt Gly	aac Asn	864
aag Lys	gac Asp 290	ttc Phe	tac Tyr	gga Gly	ccc Pro	gga Gly 295	aag Lys	acc Thr	gtc Val	gac Asp	acc Thr 300	aac Asn	agg Arg	aag Lys	atg Met	912
acc Thr 305	gtc Val	gtc Val	acc Thr	caa Gln	ttc Phe 310	atc Ile	acc Thr	cac His	gac Asp	aac Asn 315	acc Thr	gac Asp	act Thr	ggc Gly	acc Thr 320	960
ctc Leu	gtt Val	gac Asp	atc Ile	cgc Arg 325	cgc Arg	ctc Leu	tac Tyr	gtt Val	caa Gln 330	gac Asp	ggc Gly	cgt Arg	gtc Val	att Ile 335	gcc Ala	1008
aac Asn	cct Pro	ccc Pro	acc Thr 340	aac Asn	ttc Phe	ccc Pro	ggt Gly	ctc Leu 345	atg Met	ccc Pro	gcc Ala	cac His	gac Asp 350	tcc Ser	atc Ile	1056
acc Thr	gag Glu	cag Gln 355	ttc Phe	tgc Cys	act Thr	gac Asp	cag Gln 360	aag Lys	aac Asn	ctc Leu	ttc Phe	ggc Gly 365	gac Asp	tac Tyr	agc Ser	1104
agc Ser	ttc Phe 370	gct Ala	cgt Arg	gac Asp	ggt Gly	ggt Gly 375	ctc Leu	gct Ala	cac His	atg Met	ggt Gly 380	cgc Arg	tcc Ser	ctc Leu	gcc Ala	1152
aag Lys 385	ggt Gly	cac His	gtc Val	ctc Leu	gct Ala 390	ctc Leu	tcc Ser	atc Ile	tgg Trp	aac Asn 395	gac Asp	cac His	ggt Gly	gcc Ala	cac His 400	1200

atg	ttg	tgg	ctc	gac	tcc	aac	tac	ccc	acc	gac	gct	gac	ccc	aac	aag	1248
Met	Leu	Trp	Leu	Asp	Ser	Asn	Tyr	Pro	Thr	Asp	Ala	Asp	Pro	Asn	Lys	
				405					410					415		
ccc	ggt	att	gct	cgt	ggt	acc	tgc	ccg	acc	act	ggt	ggc	acc	ccc	cgt	1296
Pro	Gly	Ile	Ala	Arg	Gly	Thr	Cys	Pro	Thr	Thr	Gly	Gly	Thr	Pro	Arg	
			420					425					430			
gaa	acc	gaa	caa	aac	cac	cct	gat	gcc	cag	gtc	atc	ttc	tcc	aac	att	1344
Glu	Thr	Glu	Gln	Asn	His	Pro	Asp	Ala	Gln	Val	Ile	Phe	Ser	Asn	Ile	
		435					440					445				
aaa	ttc	ggt	gac	atc	ggc	tgc	act	ttc	tct	ggt	tac	taa				1383
Lys	Phe	Gly	Asp	Ile	Gly	Ser	Thr	Phe	Ser	Gly	Tyr					
	450					455					460					

<210> 52

<211> 460

<212> PRT

<213> Coprinus cinereus

<400> 52

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

Gly Ala Leu Tyr Phe₁₆₅ Ile Gln Met Asp Ala₁₇₀ Asp Gly Gly Met Ser₁₇₅ Lys
 Gln Pro Asn Asn₁₈₀ Arg Ala Gly Ala Lys₁₈₅ Tyr Gly Thr Gly Tyr₁₉₀ Cys Asp
 Ser Gln Cys₁₉₅ Pro Arg Asp Ile Lys₂₀₀ Phe Ile Asp Gly Val₂₀₅ Ala Asn Ser
 Ala Asp₂₁₀ Trp Thr Pro Ser Glu₂₁₅ Thr Asp Pro Asn Ala₂₂₀ Gly Arg Gly Arg
 Tyr Gly Ile Cys Cys Ala₂₃₀ Glu Met Asp Ile Trp₂₃₅ Glu Ala Asn Ser Ile₂₄₀
 Ser Asn Ala Tyr Thr₂₄₅ Pro His Pro Cys Arg₂₅₀ Thr Gln Asn Asp Gly₂₅₅ Gly
 Tyr Gln Arg Cys₂₆₀ Glu Gly Arg Asp Cys₂₆₅ Asn Gln Pro Arg Tyr₂₇₀ Glu Gly
 Leu Cys Asp₂₇₅ Pro Asp Gly Cys Asp₂₈₀ Tyr Asn Pro Phe Arg₂₈₅ Met Gly Asn
 Lys Asp₂₉₀ Phe Tyr Gly Pro Gly₂₉₅ Lys Thr Val Asp₃₀₀ Thr Asn Arg Lys Met
 Thr Val Val Thr Gln Phe₃₁₀ Ile Thr His Asp Asn₃₁₅ Thr Asp Thr Gly Thr₃₂₀
 Leu Val Asp Ile Arg₃₂₅ Arg Leu Tyr Val Gln₃₃₀ Asp Gly Arg Val Ile₃₃₅ Ala
 Asn Pro Pro Thr₃₄₀ Asn Phe Pro Gly Leu₃₄₅ Met Pro Ala His Asp₃₅₀ Ser Ile
 Thr Glu Gln₃₅₅ Phe Cys Thr Asp Gln₃₆₀ Lys Asn Leu Phe Gly₃₆₅ Asp Tyr Ser
 Ser Phe₃₇₀ Ala Arg Asp Gly Gly₃₇₅ Leu Ala His Met Gly₃₈₀ Arg Ser Leu Ala
 Lys₃₈₅ Gly His Val Leu Ala₃₉₀ Leu Ser Ile Trp Asn₃₉₅ Asp His Gly Ala His₄₀₀
 Met Leu Trp Leu Asp₄₀₅ Ser Asn Tyr Pro Thr₄₁₀ Asp Ala Asp Pro Asn₄₁₅ Lys
 Pro Gly Ile Ala₄₂₀ Arg Gly Thr Cys Pro Thr Thr Gly Gly Thr₄₃₀ Pro Arg
 Glu Thr Glu Gln Asn His Pro Asp Ala Gln Val Ile Phe Ser Asn Ile

435

440

-

445

Lys Phe Gly Asp Ile Gly Ser Thr Phe Ser Gly Tyr
450 455 460

<210> 53

<211> 1353

<212> DNA

<213> Acremonium sp.

<220>

<221> CDS

<222> (1)..(1353)

<223>

<400> 53

atg Met 1	atg Met	aag Lys	cag Gln	tat Tyr 5	ctt Leu	cag Gln	tac Tyr	ctg Leu	gcg Ala 10	gcg Ala	gct Ala	ctg Leu	ccc Pro	cta Leu 15	atg Met	48
ggc Gly	ctt Leu	gcc Ala 20	gcg Ala	ggc Gly	cag Gln	caa Gln	gcc Ala	ggc Gly 25	cgg Arg	gag Glu	acg Thr	ccc Pro	gaa Glu 30	aac Asn	cac His	96
ccc Pro	cgg Arg	ctc Leu 35	acc Thr	tgg Trp	aag Lys	aag Lys	tgc Cys 40	tcg Ser	ggc Gly	cag Gln	ggg Gly 45	tcc Ser	tgc Cys	cag Gln	acc Thr	144
gtc Val	aac Asn 50	ggc Gly	gag Glu	gtc Val	gtc Val	att Ile 55	gat Asp	gcc Ala	aac Asn	tgg Trp	cgc Arg 60	tgg Trp	ctc Leu	cac His	gac Asp	192
tcc Ser 65	aac Asn	atg Met	cag Gln	aac Asn	tgc Cys 70	tac Tyr	gac Asp	ggc Gly	aac Asn	cag Gln 75	tgg Trp	acc Thr	agc Ser	gcg Ala	tgc Cys 80	240
agc Ser	tcg Ser	gcc Ala	acc Thr	gac Asp 85	tgc Cys	gcc Ala	tcc Ser	aag Lys	tgc Cys 90	tac Tyr	atc Ile	gag Glu	ggc Gly 95	gcc Ala	gac Asp	288
tac Tyr	ggc Gly	agg Arg	acc Thr 100	tac Tyr	ggc Gly	gct Ala	tcg Ser	acg Thr 105	agc Ser	ggc Gly	gac Asp	tcc Ser	ctc Leu 110	acg Thr	ctc Leu	336
aag Lys	ttt Phe 115	gtc Val	act Thr	cag Gln	cac His	gag Glu	tac Tyr 120	ggt Gly	acc Thr	aac Asn	atc Ile	ggc Gly 125	tcg Ser	cgc Arg	ttc Phe	384
tac Tyr	ctg Leu 130	atg Met	agc Ser	agc Ser	ccg Pro	acc Thr 135	cgg Arg	tac Tyr	cag Gln	atg Met	ttc Phe 140	acc Thr	ctc Leu	atg Met	aac Asn	432
aac Asn 145	gaa Glu	ttt Phe	gct Ala	ttc Phe	gat Asp 150	gtc Val	gac Asp	ctc Leu	tcg Ser	acc Thr 155	gtc Val	gag Glu	tgc Cys	ggc Gly	atc Ile 160	480
aac	agc	gcc	ctg	tac	ttc	gtc	gcc	atg	gag	gag	gac	ggc	ggc	atg	gcc	528

Asn	Ser	Ala	Leu	Tyr 165	Phe	Val	Ala	Met	Glu 170	Glu	Asp	Gly	Gly	Met 175	Ala	
agc	tac	ccc	acc	aac	aag	gcc	gga	gcc	aag	tac	ggc	acg	ggt	tac	tgc	576
Ser	Tyr	Pro	Thr 180	Asn	Lys	Ala	Gly	Ala 185	Lys	Tyr	Gly	Thr	Gly 190	Tyr	Cys	
gac	gcc	caa	tgc	gcc	cgt	gat	ctc	aag	ttc	gtc	ggc	ggc	aag	gcc	aac	624
Asp	Ala	Gln 195	Cys	Ala	Arg	Asp	Leu 200	Lys	Phe	Val	Gly	Gly 205	Lys	Ala	Asn	
att	gag	ggc	tgg	agg	ccg	tcc	acc	aac	gac	gcg	aac	gcc	ggc	gtc	ggc	672
Ile	Glu 210	Gly	Trp	Arg	Pro	Ser 215	Thr	Asn	Asp	Ala	Asn 220	Ala	Gly	Val	Gly	
ccg	atg	ggc	ggc	tgc	tgc	gcg	gaa	atc	gat	gtt	tgg	gag	tcc	aac	gcc	720
Pro	Met	Gly	Gly	Cys	Cys 230	Ala	Glu	Ile	Asp	Val 235	Trp	Glu	Ser	Asn	Ala 240	
cac	gct	ttt	gcc	ttc	acg	ccg	cac	gcg	tgc	gag	aac	aac	aac	tac	cac	768
His	Ala	Phe	Ala	Phe 245	Thr	Pro	His	Ala	Cys 250	Glu	Asn	Asn	Asn	Tyr 255	His	
atc	tgc	gag	acc	tcc	aac	tgc	ggc	ggt	acc	tac	tcc	gac	gac	cgc	ttc	816
Ile	Cys	Glu	Thr 260	Ser	Asn	Cys	Gly	Gly 265	Thr	Tyr	Ser	Asp	Asp 270	Arg	Phe	
gcc	ggc	ctc	tgc	gac	gcc	aac	ggc	tgc	gac	tac	aac	ccg	tac	cgc	atg	864
Ala	Gly	Leu 275	Cys	Asp	Ala	Asn	Gly 280	Cys	Asp	Tyr	Asn	Pro 285	Tyr	Arg	Met	
ggc	aac	ccc	gac	ttc	tac	ggc	aag	ggc	aag	act	ctt	gac	acc	tcg	cgg	912
Gly	Asn 290	Pro	Asp	Phe	Tyr	Gly 295	Lys	Gly	Lys	Thr	Leu 300	Asp	Thr	Ser	Arg	
aag	ttc	acc	gtc	gtc	acc	cgc	ttc	cag	gag	aac	gac	ctc	tcg	cag	tac	960
Lys	Phe	Thr	Val	Val	Thr 310	Arg	Phe	Gln	Glu	Asn 315	Asp	Leu	Ser	Gln	Tyr 320	
ttc	atc	cag	gac	ggc	cgc	aag	atc	gag	atc	ccg	ccc	ccg	acc	tgg	gac	1008
Phe	Ile	Gln	Asp	Gly 325	Arg	Lys	Ile	Glu	Ile 330	Pro	Pro	Pro	Thr	Trp 335	Asp	
ggc	ctc	ccg	aag	agc	agc	cac	atc	acg	ccc	gag	ctg	tgc	gcg	acc	cag	1056
Gly	Leu	Pro	Lys 340	Ser	Ser	His	Ile	Thr 345	Pro	Glu	Leu	Cys	Ala 350	Thr	Gln	
ttc	gac	gtc	ttc	gac	gac	cgc	aac	cgc	ttc	gag	gag	gtc	ggc	ggc	ttc	1104
Phe	Asp	Val 355	Phe	Asp	Asp	Arg	Asn 360	Arg	Phe	Glu	Glu	Val 365	Gly	Gly	Phe	
ccc	gcc	ctc	aac	gcc	gct	ctc	cgc	atc	ccc	atg	gtc	ctt	gtc	atg	tcc	1152
Pro	Ala 370	Leu	Asn	Ala	Ala	Leu 375	Arg	Ile	Pro	Met	Val 380	Leu	Val	Met	Ser	
atc	tgg	gac	gac	cac	tac	gcc	aac	atg	ctc	tgg	ctc	gac	tcc	gtc	tac	1200
Ile	Trp	Asp	Asp	His 390	Tyr	Ala	Asn	Met	Leu	Trp 395	Leu	Asp	Ser	Val	Tyr 400	
ccg	ccc	gag	aag	gag	ggc	acc	ccc	ggc	gcc	gag	cgt	ggc	cct	tgc	ccc	1248
Pro	Pro	Glu	Lys	Glu 405	Gly	Thr	Pro	Gly	Ala 410	Glu	Arg	Gly	Pro	Cys 415	Pro	
cag	acc	tct	ggt	gtc	ccc	gcc	gaa	gtc	gag	gcc	cag	tac	ccc	aac	gcc	1296
Gln	Thr	Ser	Gly 420	Val	Pro	Ala	Glu	Val 425	Glu	Ala	Gln	Tyr	Pro 430	Asn	Ala	
aag	gtc	gtc	tgg	tcc	aac	atc	cgc	ttc	ggc	ccc	atc	ggc	tcg	acc	tac	1344
Lys	Val	Val	Trp	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	Gly	Ser	Thr	Tyr	

	435	440	445	
aac atg taa				1353
Asn Met				
450				

<210> 54
 <211> 450
 <212> PRT
 <213> Acremonium sp.

<400> 54

5

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

Ile Glu Gly Trp Arg Pro Ser Thr Asn Asp Ala Asn Ala Gly Val Gly
 210 215 220

Pro Met Gly Gly Cys Cys Ala Glu Ile Asp Val Trp Glu Ser Asn Ala
 225 230 235 240

His Ala Phe Ala Phe Thr Pro His Ala Cys Glu Asn Asn Asn Tyr His
 245 250 255

Ile Cys Glu Thr Ser Asn Cys Gly Gly Thr Tyr Ser Asp Asp Arg Phe
 260 265 270

Ala Gly Leu Cys Asp Ala Asn Gly Cys Asp Tyr Asn Pro Tyr Arg Met
 275 280 285

Gly Asn Pro Asp Phe Tyr Gly Lys Gly Lys Thr Leu Asp Thr Ser Arg
 290 295 300

Lys Phe Thr Val Val Thr Arg Phe Gln Glu Asn Asp Leu Ser Gln Tyr
 305 310 315 320

Phe Ile Gln Asp Gly Arg Lys Ile Glu Ile Pro Pro Pro Thr Trp Asp
 325 330 335

Gly Leu Pro Lys Ser Ser His Ile Thr Pro Glu Leu Cys Ala Thr Gln
 340 345 350

Phe Asp Val Phe Asp Asp Arg Asn Arg Phe Glu Glu Val Gly Gly Phe
 355 360 365

Pro Ala Leu Asn Ala Ala Leu Arg Ile Pro Met Val Leu Val Met Ser
 370 375 380

Ile Trp Asp Asp His Tyr Ala Asn Met Leu Trp Leu Asp Ser Val Tyr
 385 390 395 400

Pro Pro Glu Lys Glu Gly Thr Pro Gly Ala Glu Arg Gly Pro Cys Pro
 405 410 415

Gln Thr Ser Gly Val Pro Ala Glu Val Glu Ala Gln Tyr Pro Asn Ala
 420 425 430

Lys Val Val Trp Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser Thr Tyr
 435 440 445

Asn Met
 450

<210> 55

<211> 1599

<212> DNA

<213> Chaetomidium pingtungium

5

<220>
<221> CDS
<222> (1)..(1599)
<223>

<400> 55

atg Met 1	ctg Leu	gcc Ala	tcc Ser	acc Thr 5	ttc Phe	tcc Ser	tac Tyr	cgc Arg	atg Met 10	tac Tyr	aag Lys	acc Thr	gcg Ala	ctc Leu 15	atc Ile	48
ctg Leu	gcc Ala	gcc Ala	ctt Leu 20	ctg Leu	ggc Gly	tct Ser	ggc Gly	cag Gln 25	gct Ala	cag Gln	cag Gln	gtc Val	ggg Gly 30	act Thr	tcc Ser	96
cag Gln	gcg Ala	gaa Glu 35	gtg Val	cat His	ccg Pro	tcc Ser	atg Met 40	acc Thr	tgg Trp	cag Gln	agc Ser	tgc Cys 45	acc Thr	gct Ala	ggc Gly	144
ggc Gly 50	agc Ser	tgc Cys	acc Thr	acc Thr	aac Asn	aac Asn 55	ggc Gly	aag Lys	gtg Val	gtc Val	atc Ile 60	gac Asp	gcg Ala	aac Asn	tgg Trp	192
cgt Arg 65	tgg Trp	gtg Val	cac His	aaa Lys	gtc Val 70	ggc Gly	gac Asp	tac Tyr	acc Thr	aac Asn 75	tgc Cys	tac Tyr	acc Thr	ggc Gly	aac Asn 80	240
acc Thr	tgg Trp	gac Asp	acg Thr	act Thr 85	atc Ile	tgc Cys	cct Pro	gac Asp	gat Asp 90	gcg Ala	acc Thr	tgc Cys	gca Ala	tcc Ser 95	aac Asn	288
tgc Cys	gcc Ala	ctt Leu	gag Glu 100	ggg Gly	gcc Ala	aac Asn	tac Tyr	gaa Glu 105	tcc Ser	acc Thr	tat Tyr	ggg Gly	gtg Val 110	acc Thr	gcc Ala	336
agc Ser	ggc Gly	aat Asn 115	tcc Ser	ctc Leu	cgc Arg	ctc Leu	aac Asn 120	ttc Phe	gtc Val	acc Thr	acc Thr	agc Ser 125	cag Gln	cag Gln	aag Lys	384
aac Asn	att Ile 130	ggc Gly	tcg Ser	cgt Arg	ctg Leu	tac Tyr 135	atg Met	atg Met	aag Lys	gac Asp	gac Asp 140	tcg Ser	acc Thr	tac Tyr	gag Glu	432
atg Met 145	ttt Phe	aag Lys	ctg Leu	ctg Leu	aac Asn 150	cag Gln	gag Glu	ttc Phe	acc Thr	ttc Phe 155	gat Asp	gtc Val	gat Asp	gtc Val	tcc Ser 160	480
aac Asn	ctc Leu	ccc Pro	tgc Cys	ggg Gly 165	ctc Leu	aac Asn	ggg Gly	gct Ala	ctg Leu 170	tac Tyr	ttt Phe	gtc Val	gcc Ala	atg Met 175	gac Asp	528
gcc Ala	ggc Gly	ggg Gly	ggc Gly 180	atg Met	tcc Ser	aag Lys	tac Tyr	cca Pro 185	acc Thr	aac Asn	aag Lys	gcc Ala	ggg Gly 190	gcc Ala	aag Lys	576
tac Tyr	ggg Gly	act Thr 195	gga Gly	tac Tyr	tgt Cys	gac Asp	tcg Ser 200	cag Gln	tgc Cys	cct Pro	cgc Arg	gac Asp 205	ctc Leu	aag Lys	ttc Phe	624

atc Ile	aac Asn 210	ggt Gly	cag Gln	gcc Ala	aac Asn	gtt Val 215	gaa Glu	ggg Gly	tgg Trp	cag Gln	ccc Pro 220	tcc Ser	tcc Ser	aac Asn	gat Asp	672
gcc Ala 225	aat Asn	gcg Ala	ggt Gly	acc Thr	ggc Gly 230	aac Asn	cac His	ggg Gly	tcc Ser	tgc Cys 235	tgc Cys	gcg Ala	gag Glu	atg Met	gat Asp 240	720
atc Ile	tgg Trp	gag Glu	gcc Ala	aac Asn 245	agc Ser	atc Ile	tcc Ser	acg Thr	gcc Ala 250	ttc Phe	acc Thr	ccc Pro	cat His 255	ccg Pro	tgc Cys	768
gac Asp	acg Thr	ccc Pro	ggc Gly 260	cag Gln	gtg Val	atg Met	tgc Cys	acc Thr 265	ggt Gly	gat Asp	gcc Ala	tgc Cys	ggt Gly 270	ggc Gly	acc Thr	816
tac Tyr	agc Ser	tcc Ser 275	gac Asp	cgc Arg	tac Tyr	ggc Gly	ggc Gly 280	acc Thr	tgc Cys	gac Asp	ccc Pro	gac Asp 285	gga Gly	tgt Cys	gat Asp	864
ttc Phe	aac Asn 290	tcc Ser	ttc Phe	cgc Arg	cag Gln	ggc Gly 295	aac Asn	aag Lys	acc Thr	ttc Phe 300	tac Tyr 300	ggc Gly	cct Pro	ggc Gly	atg Met	912
acc Thr 305	gtc Val	gac Asp	acc Thr	aag Lys	agc Ser 310	aag Lys	ttt Phe	acc Thr	gtc Val	gtc Val 315	acc Thr	cag Gln	ttc Phe	atc Ile	acc Thr 320	960
gac Asp	gac Asp	ggc Gly	acc Thr	tcc Ser 325	agc Ser	ggc Gly	acc Thr	ctc Leu	aag Lys 330	gag Glu	atc Ile	aag Lys	cgc Arg	ttc Phe 335	tac Tyr	1008
gtg Val	cag Gln	aac Asn	ggc Gly 340	aag Lys	gtg Val	atc Ile	ccc Pro	aac Asn 345	tcg Ser	gag Glu	tcg Ser	acc Thr	tgg Trp 350	acc Thr	ggc Gly	1056
gtc Val	agc Ser	ggc Gly 355	aac Asn	tcc Ser	atc Ile	acc Thr	acc Thr 360	gag Glu	tac Tyr	tgc Cys	acc Thr	gcc Ala 365	cag Gln	aag Lys	agc Ser	1104
ctg Leu 370	ttc Phe	cag Gln	gac Asp	cag Gln	aac Asn	gtc Val 375	ttc Phe	gaa Glu	aag Lys	cac His	ggc Gly 380	ggc Gly	ctc Leu	gag Glu	ggc Gly	1152
atg Met 385	ggt Gly	gct Ala	gcc Ala	ctc Leu	gcc Ala 390	cag Gln	ggc Gly	atg Met	gtt Val	ctc Leu 395	gtc Val	atg Met	tcc Ser	ctg Leu	tgg Trp 400	1200
gat Asp	gat Asp	cac His	tcg Ser	gcc Ala 405	aac Asn	atg Met	ctc Leu	tgg Trp	ctc Leu 410	gac Asp	agc Ser	aac Asn	tac Tyr	ccg Pro 415	acc Thr	1248
act Thr	gcc Ala	tct Ser	tcc Ser 420	acc Thr	act Thr	ccc Pro	ggc Gly	gtc Val 425	gcc Ala	cgt Arg	ggt Gly	acc Thr	tgc Cys 430	gac Asp	atc Ile	1296
tcc Ser	tcc Ser	ggc Gly 435	gtc Val	cct Pro	gcg Ala	gat Asp	gtc Val 440	gag Glu	gcg Ala	aac Asn	cac His	ccc Pro 445	gac Asp	gcc Ala	tac Tyr	1344
gtc Val 450	gtc Val	tac Tyr	tcc Ser	aac Asn	atc Ile	aag Lys 455	gtc Val	ggc Gly	ccc Pro	atc Ile	ggc Gly 460	tcg Ser	acc Thr	ttc Phe	aac Asn	1392
agc Ser 465	ggt Gly	ggc Gly	tcg Ser	aac Asn	ccc Pro 470	ggt Gly	ggc Gly	gga Gly	acc Thr	acc Thr 475	acg Thr	aca Thr	act Thr	acc Thr	acc Thr 480	1440

cag cct act acc acc acg acc acg gct gga aac cct ggc ggc acc gga	1488
Gln Pro Thr Thr Thr Thr Thr Thr Ala Gly Asn Pro Gly Gly Thr Gly	
485 490 495	
gtc gca cag cac tat ggc cag tgt ggt gga atc gga tgg acc gga ccc	1536
Val Ala Gln His Tyr Gly Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro	
500 505 510	
aca acc tgt gcc agc cct tat acc tgc cag aag ctg aat gat tat tac	1584
Thr Thr Cys Ala Ser Pro Tyr Thr Cys Gln Lys Leu Asn Asp Tyr Tyr	
515 520 525	
tct cag tgc ctg tag	1599
Ser Gln Cys Leu	
530	

<210> 56

<211> 532

<212> PRT

<213> Chaetomidium pingtungium

<400> 56

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

Asn Leu Pro Cys Gly₁₆₅ Leu Asn Gly Ala Leu₁₇₀ Tyr Phe Val Ala Met₁₇₅ Asp
 Ala Gly Gly Gly₁₈₀ Met Ser Lys Tyr Pro₁₈₅ Thr Asn Lys Ala Gly₁₉₀ Ala Lys
 Tyr Gly Thr₁₉₅ Gly Tyr Cys Asp Ser₂₀₀ Gln Cys Pro Arg Asp₂₀₅ Leu Lys Phe
 Ile Asn₂₁₀ Gly Gln Ala Asn Val₂₁₅ Glu Gly Trp Gln Pro₂₂₀ Ser Ser Asn Asp
 Ala Asn₂₂₅ Ala Gly Thr Gly₂₃₀ Asn His Gly Ser Cys₂₃₅ Cys Ala Glu Met Asp₂₄₀
 Ile Trp Glu Ala Asn₂₄₅ Ser Ile Ser Thr Ala₂₅₀ Phe Thr Pro His Pro₂₅₅ Cys
 Asp Thr Pro Gly₂₆₀ Gln Val Met Cys Thr₂₆₅ Gly Asp Ala Cys Gly₂₇₀ Gly Thr
 Tyr Ser Ser₂₇₅ Asp Arg Tyr Gly Gly₂₈₀ Thr Cys Asp Pro Asp₂₈₅ Gly Cys Asp
 Phe Asn₂₉₀ Ser Phe Arg Gln Gly₂₉₅ Asn Lys Thr Phe Tyr₃₀₀ Gly Pro Gly Met
 Thr Val Asp Thr Lys Ser₃₁₀ Lys Phe Thr Val Val₃₁₅ Thr Gln Phe Ile Thr₃₂₀
 Asp Asp Gly Thr Ser₃₂₅ Ser Gly Thr Leu Lys₃₃₀ Glu Ile Lys Arg Phe Tyr₃₃₅
 Val Gln Asn Gly₃₄₀ Lys Val Ile Pro Asn₃₄₅ Ser Glu Ser Thr Trp₃₅₀ Thr Gly
 Val Ser Gly₃₅₅ Asn Ser Ile Thr Thr₃₆₀ Glu Tyr Cys Thr Ala₃₆₅ Gln Lys Ser
 Leu Phe₃₇₀ Gln Asp Gln Asn Val₃₇₅ Phe Glu Lys His Gly₃₈₀ Gly Leu Glu Gly
 Met₃₈₅ Gly Ala Ala Leu Ala₃₉₀ Gln Gly Met Val Leu₃₉₅ Val Met Ser Leu Trp₄₀₀
 Asp Asp His Ser Ala₄₀₅ Asn Met Leu Trp Leu₄₁₀ Asp Ser Asn Tyr Pro₄₁₅ Thr
 Thr Ala Ser Ser₄₂₀ Thr Thr Pro Gly Val₄₂₅ Ala Arg Gly Thr Cys₄₃₀ Asp Ile
 Ser Ser Gly Val Pro Ala Asp Val Glu Ala Asn His Pro Asp Ala Tyr

	435		440		445										
Val	Val	Tyr	Ser	Asn	Ile	Lys	Val	Gly	Pro	Ile	Gly	Ser	Thr	Phe	Asn
	450					455					460				
Ser	Gly	Gly	Ser	Asn	Pro	Gly	Gly	Gly	Thr	Thr	Thr	Thr	Thr	Thr	Thr
465					470					475					480
Gln	Pro	Thr	Thr	Thr	Thr	Thr	Thr	Ala	Gly	Asn	Pro	Gly	Gly	Thr	Gly
				485					490					495	
Val	Ala	Gln	His	Tyr	Gly	Gln	Cys	Gly	Gly	Ile	Gly	Trp	Thr	Gly	Pro
			500					505					510		
Thr	Thr	Cys	Ala	Ser	Pro	Tyr	Thr	Cys	Gln	Lys	Leu	Asn	Asp	Tyr	Tyr
		515					520					525			
Ser	Gln	Cys	Leu												
	530														

<210> 57
 <211> 1383
 <212> DNA
 <213> Sporotrichum pruinosum

<220>
 <221> CDS
 <222> (1)..(1383)
 <223>

<400> 57

atg ttc aag aaa gtc gcc ctc acc gct ctc tgc ttc ctc gcc gtc gca Met Phe Lys Lys Val Ala Leu Thr Ala Leu Cys Phe Leu Ala Val Ala 1 5 10 15	48
cag gcc caa cag gtc ggt cgc gaa gtc gct gaa aac cac ccc cgt ctc Gln Ala Gln Gln Val Gly Arg Glu Val Ala Glu Asn His Pro Arg Leu 20 25 30	96
ccg tgg cag cgt tgc act cgc aac ggc gga tgc cag act gtc tct aac Pro Trp Gln Arg Cys Thr Arg Asn Gly Gly Cys Gln Thr Val Ser Asn 35 40 45	144
ggt cag gtc gtc ctc gac gcc aac tgg cga tgg ctc cac gtc acc gat Gly Gln Val Val Leu Asp Ala Asn Trp Arg Trp Leu His Val Thr Asp 50 55 60	192
ggc tac acc aac tgc tac acc ggt aac tcc tgg aac agc acc gtc tgc Gly Tyr Thr Asn Cys Tyr Thr Gly Asn Ser Trp Asn Ser Thr Val Cys 65 70 75 80	240
tcc gac ccc acc acc tgc gct cag cga tgc gct ctc gag ggt gcc aac	288

Ser	Asp	Pro	Thr	Thr 85	Cys	Ala	Gln	Arg	Cys 90	Ala	Leu	Glu	Gly	Ala 95	Asn		
tac	cag	caa	acc	tac	ggt	atc	acc	acc	aac	gga	gac	gcc	ctc	acc	atc		336
Tyr	Gln	Gln	Thr 100	Tyr	Gly	Ile	Thr	Thr 105	Asn	Gly	Asp	Ala	Leu 110	Thr	Ile		
aag	ttc	ctc	acc	cga	tcc	caa	caa	acc	aac	gtc	ggt	gct	cgt	gtc	tac		384
Lys	Phe	Leu 115	Thr	Arg	Ser	Gln	Gln 120	Thr	Asn	Val	Gly	Ala 125	Arg	Val	Tyr		
ctc	atg	gag	aac	gag	aac	cga	tac	cag	atg	ttc	aac	ctc	ctc	aac	aag		432
Leu	Met 130	Glu	Asn	Glu	Asn	Arg 135	Tyr	Gln	Met	Phe	Asn 140	Leu	Leu	Asn	Lys		
gag	ttc	acc	ttc	gac	gtt	gac	gtc	tcc	aag	gtt	cct	tgc	ggt	atc	aac		480
Glu 145	Phe	Thr	Phe	Asp	Val 150	Asp	Val	Ser	Lys	Val 155	Pro	Cys	Gly	Ile	Asn 160		
ggt	gcc	ctc	tac	ttc	atc	cag	atg	gac	gcc	gat	ggt	ggt	atg	agc	aag		528
Gly	Ala	Leu	Tyr	Phe 165	Ile	Gln	Met	Asp	Ala 170	Asp	Gly	Gly	Met	Ser 175	Lys		
caa	ccc	aac	aac	agg	gct	ggt	gct	aag	tac	ggt	acc	ggc	tac	tgc	gac		576
Gln	Pro	Asn	Asn 180	Arg	Ala	Gly	Ala	Lys 185	Tyr	Gly	Thr	Gly	Tyr 190	Cys	Asp		
tct	cag	tgc	ccc	cgt	gac	atc	aag	ttc	att	gac	ggc	gtg	gcc	aac	agc		624
Ser	Gln	Cys 195	Pro	Arg	Asp	Ile	Lys 200	Phe	Ile	Asp	Gly	Val 205	Ala	Asn	Ser		
gcc	gac	tgg	act	cca	tcc	gag	acc	gat	ccc	aat	gcc	gga	agg	ggt	cgc		672
Ala 210	Asp	Trp	Thr	Pro	Ser	Glu 215	Thr	Asp	Pro	Asn	Ala 220	Gly	Arg	Gly	Arg		
tac	ggc	att	tgc	tgc	gcc	gag	atg	gat	atc	tgg	gag	gcc	aac	tcc	atc		720
Tyr 225	Gly	Ile	Cys	Cys	Ala 230	Glu	Met	Asp	Ile	Trp 235	Glu	Ala	Asn	Ser	Ile 240		
tcc	aat	gcc	tac	acc	ccc	cac	cct	tgc	cga	acc	cag	aac	gat	ggt	ggc		768
Ser	Asn	Ala	Tyr	Thr 245	Pro	His	Pro	Cys	Arg 250	Thr	Gln	Asn	Asp	Gly 255	Gly		
tac	cag	cgc	tgc	gag	ggc	cgc	gac	tgc	aac	cag	cct	cgc	tat	gag	ggt		816
Tyr	Gln	Arg	Cys 260	Glu	Gly	Arg	Asp	Cys 265	Asn	Gln	Pro	Arg	Tyr 270	Glu	Gly		
ctt	tgc	gat	cct	gat	ggc	tgt	gac	tac	aac	ccc	ttc	cgc	atg	ggt	aac		864
Leu	Cys	Asp 275	Pro	Asp	Gly	Cys	Asp 280	Tyr	Asn	Pro	Phe	Arg 285	Met	Gly	Asn		
aag	gac	ttc	tac	gga	ccc	gga	aag	acc	atc	gac	acc	aac	agg	aag	atg		912
Lys	Asp 290	Phe	Tyr	Gly	Pro	Gly 295	Lys	Thr	Ile	Asp	Thr 300	Asn	Arg	Lys	Met		
acc	gtc	gtc	acc	caa	ttc	atc	acc	cac	gac	aac	acc	gac	act	ggc	acc		960
Thr 305	Val	Val	Thr	Gln 310	Phe	Ile	Thr	His	Asp	Asn 315	Thr	Asp	Thr	Gly	Thr 320		
ctc	gtt	gac	atc	cgc	cgc	ctc	tac	gtt	caa	gac	ggc	cgt	gtc	att	gcc		1008
Leu	Val	Asp	Ile	Arg 325	Arg	Leu	Tyr	Val	Gln 330	Asp	Gly	Arg	Val	Ile 335	Ala		
aac	cct	ccc	acc	aac	ttc	ccc	ggt	ctc	atg	ccc	gcc	cac	gac	tcc	atc		1056
Asn	Pro	Pro	Thr 340	Asn	Phe	Pro	Gly	Leu 345	Met	Pro	Ala	His	Asp 350	Ser	Ile		
acc	gag	cag	ttc	tgc	act	gac	cag	aag	aac	ctc	ttc	ggc	gac	tac	agc		1104
Thr	Glu	Gln	Phe	Cys	Thr	Asp	Gln	Lys	Asn	Leu	Phe	Gly	Asp	Tyr	Ser		

355					360					365						
agc	ttc	gct	cgt	gac	ggt	ggt	ctc	gct	cac	atg	ggt	cgc	tcc	ctc	gcc	1152
Ser	Phe	Ala	Arg	Asp	Gly	Gly	Leu	Ala	His	Met	Gly	Arg	Ser	Leu	Ala	
	370					375					380					
aag	ggt	cac	gtc	ctc	gct	ctc	tcc	atc	tgg	aac	gac	cac	ggt	gcc	cac	1200
Lys	Gly	His	Val	Leu	Ala	Leu	Ser	Ile	Trp	Asn	Asp	His	Gly	Ala	His	
	385				390					395					400	
atg	ttg	tgg	ctc	gac	tcc	aac	tac	ccc	acc	gac	gct	gac	ccc	aac	aag	1248
Met	Leu	Trp	Leu	Asp	Ser	Asn	Tyr	Pro	Thr	Asp	Ala	Asp	Pro	Asn	Lys	
				405					410					415		
ccc	ggt	att	gct	cgt	ggt	acc	tgc	ccg	acc	act	ggt	ggc	acc	ccc	cgt	1296
Pro	Gly	Ile	Ala	Arg	Gly	Thr	Cys	Pro	Thr	Thr	Gly	Gly	Thr	Pro	Arg	
			420				425						430			
gaa	acc	gaa	caa	aac	cac	cct	gat	gcc	cag	gtc	atc	ttc	tcc	aac	att	1344
Glu	Thr	Glu	Gln	Asn	His	Pro	Asp	Ala	Gln	Val	Ile	Phe	Ser	Asn	Ile	
		435					440					445				
aaa	ttc	ggt	gac	atc	ggc	tcg	act	ttc	tct	ggt	tac	taa				1383
Lys	Phe	Gly	Asp	Ile	Gly	Ser	Thr	Phe	Ser	Gly	Tyr					
	450				455					460						

<210> 58

<211> 460

<212> PRT

<213> Sporotrichum pruinosum

<400> 58

Met Phe Lys Lys Val Ala Leu Thr Ala Leu Cys Phe Leu Ala Val Ala
 1 5 10 15
 Gln Ala Gln Gln Val Gly Arg Glu Val Ala Glu Asn His Pro Arg Leu
 20 25 30
 Pro Trp Gln Arg Cys Thr Arg Asn Gly Gly Cys Gln Thr Val Ser Asn
 35 40 45
 Gly Gln Val Val Leu Asp Ala Asn Trp Arg Trp Leu His Val Thr Asp
 50 55 60
 Gly Tyr Thr Asn Cys Tyr Thr Gly Asn Ser Trp Asn Ser Thr Val Cys
 65 70 75 80
 Ser Asp Pro Thr Thr Cys Ala Gln Arg Cys Ala Leu Glu Gly Ala Asn
 85 90 95
 Tyr Gln Gln Thr Tyr Gly Ile Thr Thr Asn Gly Asp Ala Leu Thr Ile
 100 105 110
 Lys Phe Leu Thr Arg Ser Gln Gln Thr Asn Val Gly Ala Arg Val Tyr
 115 120 125

Leu Met Glu Asn Glu Asn Arg Tyr Gln Met Phe Asn Leu Leu Asn Lys
 130 135 140
 Glu Phe Thr Phe Asp Val Asp Val Ser Lys Val Pro Cys Gly Ile Asn
 145 150 155 160
 Gly Ala Leu Tyr Phe Ile Gln Met Asp Ala Asp Gly Gly Met Ser Lys
 165 170 175
 Gln Pro Asn Asn Arg Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp
 180 185 190
 Ser Gln Cys Pro Arg Asp Ile Lys Phe Ile Asp Gly Val Ala Asn Ser
 195 200 205
 Ala Asp Trp Thr Pro Ser Glu Thr Asp Pro Asn Ala Gly Arg Gly Arg
 210 215 220
 Tyr Gly Ile Cys Cys Ala Glu Met Asp Ile Trp Glu Ala Asn Ser Ile
 225 230 235 240
 Ser Asn Ala Tyr Thr Pro His Pro Cys Arg Thr Gln Asn Asp Gly Gly
 245 250 255
 Tyr Gln Arg Cys Glu Gly Arg Asp Cys Asn Gln Pro Arg Tyr Glu Gly
 260 265 270
 Leu Cys Asp Pro Asp Gly Cys Asp Tyr Asn Pro Phe Arg Met Gly Asn
 275 280 285
 Lys Asp Phe Tyr Gly Pro Gly Lys Thr Ile Asp Thr Asn Arg Lys Met
 290 295 300
 Thr Val Val Thr Gln Phe Ile Thr His Asp Asn Thr Asp Thr Gly Thr
 305 310 315 320
 Leu Val Asp Ile Arg Arg Leu Tyr Val Gln Asp Gly Arg Val Ile Ala
 325 330 335
 Asn Pro Pro Thr Asn Phe Pro Gly Leu Met Pro Ala His Asp Ser Ile
 340 345 350
 Thr Glu Gln Phe Cys Thr Asp Gln Lys Asn Leu Phe Gly Asp Tyr Ser
 355 360 365
 Ser Phe Ala Arg Asp Gly Gly Leu Ala His Met Gly Arg Ser Leu Ala
 370 375 380
 Lys Gly His Val Leu Ala Leu Ser Ile Trp Asn Asp His Gly Ala His
 385 390 395 400

Met Leu Trp Leu Asp Ser Asn Tyr Pro Thr Asp Ala Asp Pro Asn Lys
405 410 415
Pro Gly Ile Ala Arg Gly Thr Cys Pro Thr Thr Gly Gly Thr Pro Arg
420 425 430
Glu Thr Glu Gln Asn His Pro Asp Ala Gln Val Ile Phe Ser Asn Ile
435 440 445
Lys Phe Gly Asp Ile Gly Ser Thr Phe Ser Gly Tyr
450 455 460

<210> 59
<211> 1578
<212> DNA
<213> Scytalidium thermophilum

<220>
<221> CDS
<222> (1)..(1578)
<223>

<400> 59

atg Met 1	cgt Arg	acc Thr	gcc Ala	aag Lys 5	ttc Phe	gcc Ala	acc Thr	ctc Leu	gcc Ala 10	gcc Ala	ctt Leu	gtg Val	gcc Ala	tcg Ser 15	gcc Ala	48
gcc Ala	gcc Ala	cag Gln	cag Gln 20	gcg Ala	tgc Cys	agt Ser	ctc Leu	acc Thr 25	acc Thr	gag Glu	agg Arg	cac His	cct Pro 30	tcc Ser	ctc Leu	96
tct Ser	tgg Trp	aag Lys 35	aag Lys	tgc Cys	acc Thr	gcc Ala	ggc Gly 40	ggc Gly	cag Gln	tgc Cys	cag Gln	acc Thr 45	gtc Val	cag Gln	gct Ala	144
tcc Ser 50	atc Ile	act Thr	ctc Leu	gac Asp	tcc Ser	aac Asn 55	tgg Trp	cgc Arg	tgg Trp	act Thr	cac His 60	cag Gln	gtg Val	tct Ser	ggc Gly	192
tcc Ser 65	acc Thr	aac Asn	tgc Cys	tac Tyr	acg Thr 70	ggc Gly	aac Asn	aag Lys	tgg Trp	gat Asp 75	act Thr	agc Ser	atc Ile	tgc Cys	act Thr 80	240
gat Asp	gcc Ala	aag Lys	tcg Ser	tgc Cys 85	gct Ala	cag Gln	aac Asn	tgc Cys	tgc Cys 90	gtc Val	gat Asp	ggc Gly	gcc Ala	gac Asp 95	tac Tyr	288
acc Thr	agc Ser	acc Thr	tat Tyr 100	ggc Gly	atc Ile	acc Thr	acc Thr	aac Asn 105	ggc Gly	gat Asp	tcc Ser	ctg Leu	agc Ser 110	ctc Leu	aag Lys	336
ttc Phe	gtc Val	acc Thr 115	aag Lys	ggc Gly	cag Gln	cac His	tcg Ser 120	acc Thr	aac Asn	gtc Val	ggc Gly	tcg Ser 125	cgt Arg	acc Thr	tac Tyr	384

ctg Leu	atg Met 130	gac Asp	ggc Gly	gag Glu	gac Asp	aag Lys 135	tat Tyr	cag Gln	acc Thr	ttc Phe 140	gag Glu 140	ctc Leu	ctc Leu	ggc Gly	aac Asn	432
gag Glu 145	ttc Phe	acc Thr	ttc Phe	gat Asp	gtc Val 150	gat Asp	gtc Val	tcc Ser	aac Asn	atc Ile 155	ggc Gly	tgc Cys	ggt Gly	ctc Leu	aac Asn 160	480
ggc Gly	gcc Ala	ctg Leu	tac Tyr	ttc Phe 165	gtc Val	tcc Ser	atg Met	gac Asp	gcc Ala 170	gat Asp	ggt Gly	ggt Gly	ctc Leu	agc Ser 175	cgc Arg	528
tat Tyr	cct Pro	ggc Gly	aac Asn 180	aag Lys	gct Ala	ggt Gly	gcc Ala	aag Lys 185	tac Tyr	ggt Gly	acc Thr	ggc Gly	tac Tyr 190	tgc Cys	gat Asp	576
gct Ala	cag Gln	tgc Cys 195	ccc Pro	cgt Arg	gac Asp	atc Ile	aag Lys 200	ttc Phe	atc Ile	aac Asn	ggc Gly	gag Glu 205	gcc Ala	aac Asn	att Ile	624
gag Glu	ggc Gly 210	tgg Trp	acc Thr	ggc Gly	tcc Ser	acc Thr 215	aac Asn	gac Asp	ccc Pro	aac Asn	gcc Ala 220	ggc Gly	gcg Ala	ggc Gly	cgc Arg	672
tat Tyr 225	ggt Gly	acc Thr	tgc Cys	tgc Cys	tct Ser 230	gag Glu	atg Met	gat Asp	atc Ile	tgg Trp 235	gaa Glu	gcc Ala	aac Asn	aac Asn	atg Met 240	720
gct Ala	act Thr	gcc Ala	ttc Phe	act Thr 245	cct Pro	cac His	cct Pro	tgc Cys	acc Thr 250	atc Ile	att Ile	ggc Gly	cag Gln	agc Ser 255	cgc Arg	768
tgc Cys	gag Glu	ggc Gly	gac Asp 260	tgc Ser	tgc Cys	ggt Gly	ggc Gly	acc Thr 265	tac Tyr	agc Ser	aac Asn	gag Glu 270	cgc Arg	tac Tyr	gcc Ala	816
ggc Gly	gtc Val	tgc Cys 275	gac Asp	ccc Pro	gat Asp	ggc Gly	tgc Cys 280	gac Asp	ttc Phe	aac Asn	tgc Ser 285	tac Tyr	cgc Arg	cag Gln	ggc Gly	864
aat Asn	aag Lys 290	acc Thr	ttc Phe	tac Tyr	ggc Gly	aag Lys 295	ggc Gly	atg Met	acc Thr	gtc Val	gac Asp 300	acc Thr	acc Thr	aag Lys	aag Lys	912
atc Ile 305	act Thr	gtc Val	gtc Val	acc Thr	cag Gln 310	ttc Phe	ctc Leu	aag Lys	gat Asp	gcc Ala 315	aac Asn	ggc Gly	gat Asp	ctc Leu	ggc Gly 320	960
gag Glu	gtc Val	aag Lys	cgc Arg	ttc Phe 325	tac Tyr	gtc Val	cag Gln	gat Asp	ggc Gly 330	aag Lys	atc Ile	atc Ile	ccc Pro	aac Asn 335	tcc Ser	1008
gag Glu	tcc Ser	acc Thr	atc Ile 340	ccc Pro	ggc Gly	gtc Val	gag Glu	ggc Gly 345	aac Asn	tcc Ser	atc Ile	acc Thr	cag Gln 350	gac Asp	tgg Trp	1056
tgc Cys	gac Asp	cgc Arg 355	cag Gln	aag Lys	gtt Val	gcc Ala	ttt Phe 360	ggc Gly	gac Asp	att Ile	gac Asp	gac Asp 365	ttc Phe	aac Asn	cgc Arg	1104
aag Lys	ggc Gly 370	ggc Gly	atg Met	aag Lys	cag Gln	atg Met 375	ggc Gly	aag Lys	gcc Ala	ctc Leu	gcc Ala 380	ggc Gly	ccc Pro	atg Met	gtc Val	1152
ctg Leu 385	gtc Val	atg Met	tcc Ser	atc Ile	tgg Trp 390	gat Asp	gac Asp	cac His	gcc Ala	tcc Ser 395	aac Asn	atg Met	ctc Leu	tgg Trp	ctc Leu 400	1200

gac	tgc	acc	ttc	cct	gtc	gat	gcc	gct	ggc	aag	ccc	ggc	gcc	gag	cgc	1248
Asp	Ser	Thr	Phe	Pro	Val	Asp	Ala	Ala	Gly	Lys	Pro	Gly	Ala	Glu	Arg	
				405					410					415		
ggt	gcc	tgc	ccg	acc	acc	tgc	ggt	gtc	cct	gct	gag	gtt	gag	gcc	gag	1296
Gly	Ala	Cys	Pro	Thr	Thr	Ser	Gly	Val	Pro	Ala	Glu	Val	Glu	Ala	Glu	
			420					425					430			
gcc	ccc	aac	agc	aac	gtc	gtc	ttc	tcc	aac	atc	cgc	ttc	ggc	ccc	atc	1344
Ala	Pro	Asn	Ser	Asn	Val	Val	Phe	Ser	Asn	Ile	Arg	Phe	Gly	Pro	Ile	
		435					440					445				
ggc	tgc	acc	gtt	gct	ggt	ctc	ccc	ggc	gcg	ggc	aac	ggc	ggc	aac	aac	1392
Gly	Ser	Thr	Val	Ala	Gly	Leu	Pro	Gly	Ala	Gly	Asn	Gly	Gly	Asn	Asn	
	450					455					460					
ggc	ggc	aac	ccc	ccg	ccc	ccc	acc	acc	acc	acc	tcc	tgc	gct	ccg	gcc	1440
Gly	Gly	Asn	Pro	Pro	Pro	Pro	Thr	Thr	Thr	Thr	Ser	Ser	Ala	Pro	Ala	
465					470					475					480	
acc	acc	acc	acc	gcc	agc	gct	ggc	ccc	aag	gct	ggc	cac	tgg	cag	cag	1488
Thr	Thr	Thr	Thr	Ala	Ser	Ala	Gly	Pro	Lys	Ala	Gly	His	Trp	Gln	Gln	
				485					490					495		
tgc	ggc	ggc	atc	ggc	ttc	act	ggc	ccg	acc	cag	tgc	gag	gag	ccc	tac	1536
Cys	Gly	Gly	Ile	Gly	Phe	Thr	Gly	Pro	Thr	Gln	Cys	Glu	Glu	Pro	Tyr	
			500					505					510			
act	tgc	acc	aag	ctc	aac	gac	tgg	tac	tct	cag	tgc	ctg	taa			1578
Thr	Cys	Thr	Lys	Leu	Asn	Asp	Trp	Tyr	Ser	Gln	Cys	Leu				
		515					520					525				

<210> 60

<211> 525

<212> PRT

<213> scytalidium thermophilum

<400> 60

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

Thr Ser Thr Tyr Gly Ile Thr Thr Asn Gly Asp Ser Leu Ser Leu Lys
 100 105 110
 Phe Val Thr Lys Gly Gln His Ser Thr Asn Val Gly Ser Arg Thr Tyr
 115 120 125
 Leu Met Asp Gly Glu Asp Lys Tyr Gln Thr Phe Glu Leu Leu Gly Asn
 130 135 140
 Glu Phe Thr Phe Asp Val Asp Val Ser Asn Ile Gly Cys Gly Leu Asn
 145 150 155 160
 Gly Ala Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Leu Ser Arg
 165 170 175
 Tyr Pro Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp
 180 185 190
 Ala Gln Cys Pro Arg Asp Ile Lys Phe Ile Asn Gly Glu Ala Asn Ile
 195 200 205
 Glu Gly Trp Thr Gly Ser Thr Asn Asp Pro Asn Ala Gly Ala Gly Arg
 210 215 220
 Tyr Gly Thr Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Asn Met
 225 230 235 240
 Ala Thr Ala Phe Thr Pro His Pro Cys Thr Ile Ile Gly Gln Ser Arg
 245 250 255
 Cys Glu Gly Asp Ser Cys Gly Gly Thr Tyr Ser Asn Glu Arg Tyr Ala
 260 265 270
 Gly Val Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Gln Gly
 275 280 285
 Asn Lys Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Thr Lys Lys
 290 295 300
 Ile Thr Val Val Thr Gln Phe Leu Lys Asp Ala Asn Gly Asp Leu Gly
 305 310 315 320
 Glu Val Lys Arg Phe Tyr Val Gln Asp Gly Lys Ile Ile Pro Asn Ser
 325 330 335
 Glu Ser Thr Ile Pro Gly Val Glu Gly Asn Ser Ile Thr Gln Asp Trp
 340 345 350
 Cys Asp Arg Gln Lys Val Ala Phe Gly Asp Ile Asp Asp Phe Asn Arg
 355 360 365
 Lys Gly Gly Met Lys Gln Met Gly Lys Ala Leu Ala Gly Pro Met Val

370 375 380
 Leu Val Met Ser Ile Trp Asp Asp His Ala Ser Asn Met Leu Trp Leu
 385 390 395 400
 Asp Ser Thr Phe Pro Val Asp Ala Ala Gly Lys Pro Gly Ala Glu Arg
 405 410 415
 Gly Ala Cys Pro Thr Thr Ser Gly Val Pro Ala Glu Val Glu Ala Glu
 420 425 430
 Ala Pro Asn Ser Asn Val Val Phe Ser Asn Ile Arg Phe Gly Pro Ile
 435 440 445
 Gly Ser Thr Val Ala Gly Leu Pro Gly Ala Gly Asn Gly Gly Asn Asn
 450 455 460
 Gly Gly Asn Pro Pro Pro Pro Thr Thr Thr Thr Ser Ser Ala Pro Ala
 465 470 475 480
 Thr Thr Thr Thr Ala Ser Ala Gly Pro Lys Ala Gly His Trp Gln Gln
 485 490 495
 Cys Gly Gly Ile Gly Phe Thr Gly Pro Thr Gln Cys Glu Glu Pro Tyr
 500 505 510
 Thr Cys Thr Lys Leu Asn Asp Trp Tyr Ser Gln Cys Leu
 515 520 525

<210> 61

<211> 519

<212> DNA

<213> Aspergillus sp.

<220>

<221> misc_feature

<222> (1)..(519)

<223> Secuencia de codificación parcial de CBH1

<400> 61

gagatggaca tatgggaggc caacagcatc tccacggcct tcacgccccca cccctgcat 60
gtccccggcc aggtgatgtg cgagggcgac tcctgcggtg gcacctacag cagcgaccgc 120
tatggcggca cctgcatcc cgatggatgt gacttcaact cctaccgcca gggcaacaag 180
tccttctacg gccccggcat gaccgtcgac accaacagca aggtcaccgt cgtgactcag 240
ttcctcaccg acgacggcac tgccaccggc accctgtcgg agatcaagcg gttctacgtg 300
cagaacggca aggtcatccc caactccgag tcgacctggc ccggcgctcg cggaactcc 360
atcaccaccg actactgtct ggcccagaag agcctcttcg gcgataccga cgtcttcacc 420
aagcacggcg gtatggaggg catgggcgcc gccctcgccg agggcatggt cctcgtcctg 480
agtctctggg acgaccacca ctccaacatg ctctggctg 519

<210> 62
<211> 497
<212> DNA
<213> Scopulariopsis sp.

<220>
<221> misc_feature
<222> (1)..(497)
<223> Secuencia de codificación parcial de CBH1

<400> 62
gagatcgatg tgtgggagtc gaacgcctat gccttcgttt tcacgcccga cgcgtgcacg 60
accaacgagt accacgtctg cgagaccacc aactgcggtg gcacctactc ggaggaccgc 120
ttcaccggca agtgcgacgc caacggctgc gactacaacc cctaccgcat gggcaacccc 180
gacttctacg gcaagggcaa gacgctcgac accagccgca agttcaccgt cgtctcccgc 240
ttcgaggaga acaagctctc ccagtacttc atccaggacg gccgcaagat cgagatcccg 300
ccgcccacgt gggagggcat gcccacacagc agcgagatca ccccgagct ctgctccacc 360
atgttcgatg tgctcgacga ccgcaaccgc ttgcaggagg tcggcggtt cgagcagctg 420
aacaacgccc tccgggttcc catggtcctc gtcattgtcca tctgggacga ccactacgcc 480
aacatgctct ggctcga 497

<210> 63
<211> 498
<212> DNA
<213> Fusarium sp.

<220>
<221> misc_feature
<222> (1)..(498)
<223> Secuencia de codificación parcial de CBH1

<400> 63

```

gagatggata tctgggaggc caacaagatc tccactgcct acactcecca cccctgcaag    60
agcctcacc cagcagtcctg cgagggcgat gcctgcggtg gcacctactc tactaccgcg    120
tatgctggaa cttgcgaccc cgatgggtgc gatttcaacc cttaccgcca gggcaacaag    180
accttctacg gccccggctc cggcttcaac gttgatacca ccaagaagg t gactgtcgtg    240
acccagttca tcaagggcag cgacggcaag ctttccgaga tcaagcgtct ctatgttcag    300
aatggcaagg tcattggcaa cccccagtct gagattgcca gcaaccctgg cagcagcgtc    360
accgacagct tctgcaaggc ccagaagggt gccttcaacg accccgatga cttcaacaag    420
aaggggtggc ggagcgggat gagcgacgcc ctgcgaagc ccatgggttct cgtcatgagc    480
ttgtggcacg acgtgagt                                     498

```

<210> 64
 <211> 525
 <212> DNA
 <213> Verticillium sp.

<220>
 <221> misc_feature
 <222> (1)..(525)
 <223> Secuencia de codificación parcial de CBH1

```

<400> 64
gagatggata tctgggaggc caacaagatc tccacggcct acactcecca tccctgcaag    60
agcctcacc cagcagtcctg tgagggcgat gcctgcggtg gcacctactc ttccaccgcg    120
tatgctggaa cttgcgaccc cgatgggtgc gatttcaacc cttaccgcca gggcaaccac    180
accttctacg gtccccggctc cggcttcaac gtcgatacca ccaagaagg t gactgtcgtg    240
acccagttca tcaagggcag cgacggcaag ctttccgaga tcaagcgtct ctatgttcag    300
aatggcaagg tcattggcaa cccccagtcc gagattgcaa acaaccccg g cagctccgtc    360
accgacagct tctgcaaggc ccagaagggt gccttcaacg accccgatga cttcaacaag    420
aaggggtggc ggagcgggat gaacgacgcc ctgcgaagc ccatgggttct cgtcatgagc    480
ctgtggcacg acgtgagtaa tctaaccctt gagtctcgga caaga                                     525

```

<210> 65
 <211> 1371
 <212> DNA
 <213> Pseudoplectania nigrella

<220>
 <221> CDS
 <222> (1)..(1371)
 <223>

<400> 65

atg Met 1	cta Leu	tcc Ser	aat Asn	ctc Leu 5	ctt Leu	ctc Leu	tca Ser	ctc Leu	tct Ser 10	ttc Phe	ctt Leu	tcc Ser	cta Leu	gcc Ala 15	tcc Ser	48
ggg Gly	caa Gln	aac Asn	atc Ile 20	ggg Gly	acc Thr	aac Asn	acc Thr	gcc Ala 25	gaa Glu	agc Ser	cac His	ccc Pro	caa Gln 30	ctt Leu	cgt Arg	96
tct Ser	caa Gln	acc Thr 35	tgc Cys	acc Thr	aaa Lys	ggc Gly	aac Asn 40	gga Gly	tgc Cys	agc Ser	acc Thr 45	caa Gln	tcc Ser	acc Thr	tcc Ser	144
gta Val	gtc Val 50	ctg Leu	gac Asp	tcc Ser	aac Asn	tgg Trp 55	cgc Arg	tgg Trp	ctg Leu	cac His	aat Asn 60	aat Asn	gga Gly	ggg Gly	tca Ser	192
acg Thr 65	aac Asn	tgc Cys	tac Tyr	acc Thr	ggc Gly 70	aat Asn	tcc Ser	tgg Trp	gac Asp	tct Ser 75	aca Thr	tta Leu	tgt Cys	ccc Pro	gac Asp 80	240
cca Pro	gtt Val	acc Thr	tgc Cys	gcc Ala 85	aag Lys	aac Asn	tgt Cys	gct Ala 90	ctc Leu	gac Asp	ggg Gly	gcc Ala	gac Asp	tat Tyr 95	tct Ser	288
ggg Gly	aca Thr	tac Tyr	gga Gly 100	atc Ile	acc Thr	tct Ser	acg Thr	gga Gly 105	gat Asp	gct Ala	ttg Leu	acg Thr	ttg Leu 110	aag Lys	ttt Phe	336
gtt Val	act Thr	cag Gln 115	ggg Gly	cct Pro	tat Tyr	tcg Ser	act Thr 120	aat Asn	att Ile	gga Gly	tct Ser	cgg Arg 125	gta Val	tac Tyr	cta Leu	384
atg Met	gcg Ala 130	agt Ser	gat Asp	act Thr	cag Gln	tat Tyr 135	aag Lys	atg Met	ttc Phe	cag Gln	ctc Leu 140	aag Lys	aac Asn	aag Lys	gag Glu	432
ttt Phe 145	acg Thr	ttt Phe	gat Asp	gtt Val	gat Asp 150	gtc Val	tct Ser	aat Asn	ctt Leu	cct Pro 155	tgt Cys	gga Gly	tta Leu	aac Asn	gga Gly 160	480
gcg Ala	ttg Leu	tat Tyr	ttt Phe	gtg Val 165	gag Glu	atg Met	gat Asp	gcg Ala 170	gat Asp	gga Gly	gga Gly	atg Met	tcg Ser	aaa Lys 175	tac Tyr	528
ccg Pro	tct Ser	aat Asn	aaa Lys 180	gcc Ala	ggg Gly	gca Ala	aaa Lys	tat Tyr 185	gga Gly	acc Thr	ggg Gly	tat Tyr	tgt Cys 190	gat Asp	gcg Ala	576
cag Gln	tgt Cys	cca Pro 195	cat His	gat Asp	atc Ile	aaa Lys	ttt Phe 200	atc Ile	aac Asn	ggg Gly	gag Glu	gca Ala 205	aat Asn	ctc Leu	cta Leu	624
gac Asp	tgg Trp 210	acg Thr	cct Pro	tca Ser	acc Thr	agc Ser 215	gac Asp	aaa Lys	aat Asn	gcc Ala	ggc Gly 220	tcc Ser	gga Gly	cgt Arg	tac Tyr	672

ggg Gly 225	acc Thr	tgt Cys	tgt Cys	caa Gln	gaa Glu 230	atg Met	gac Asp	atc Ile	tgg Trp	gaa Glu 235	gcc Ala	aac Asn	agc Ser	atg Met	gca Ala 240	720
acc Thr	gcc Ala	tat Tyr	aca Thr	ccg Pro 245	cat His	ccc Pro	tgt Cys	agt Ser	gtc Val 250	tca Ser	gga Gly	cct Pro	acc Thr	cga Arg 255	tgc Cys	768
tca Ser	gga Gly	acc Thr	caa Gln 260	tgt Cys	ggg Gly	gat Asp	ggt Gly	tct Ser 265	aac Asn	cgt Arg	cat His	aac Asn	gga Gly 270	att Ile	tgc Cys	816
gat Asp	aaa Lys	gat Asp 275	ggc Gly	tgc Cys	gat Asp	ttc Phe	aat Asn 280	tcc Ser	tac Tyr	cgt Arg	atg Met	ggc Gly 285	aat Asn	acg Thr	aca Thr	864
ttc Phe	ttc Phe 290	ggc Gly	aag Lys	gga Gly	gca Ala	acg Thr 295	gtt Val	aac Asn	acc Thr	aac Asn	tcc Ser 300	aaa Lys	ttt Phe	act Thr	gtt Val	912
gta Val 305	acg Thr	caa Gln	ttc Phe	atc Ile	acc Thr 310	tcc Ser	gac Asp	aac Asn	acc Thr	tca Ser 315	act Thr	gga Gly	gcg Ala	cta Leu	aag Lys 320	960
gag Glu	att Ile	cgt Arg	cgt Arg	ctt Leu 325	tat Tyr	att Ile	cag Gln	aat Asn	gga Gly 330	aaa Lys	gtc Val	atc Ile	cag Gln	aac Asn 335	tgc Ser	1008
aaa Lys	agt Ser	aat Asn 340	atc Ile	tcc Ser	ggc Gly	atg Met	tca Ser	gct Ala 345	tac Tyr	gac Asp	tct Ser	ata Ile	acc Thr 350	gag Glu	gat Asp	1056
ttc Phe	tgt Cys	ggc Ala 355	gct Ala	caa Gln	aaa Lys	acc Thr	gca Ala 360	ttt Phe	gga Gly	gac Asp	aca Thr	aat Asn 365	gac Asp	ttt Phe	aag Lys	1104
gca Ala 370	aag Lys	ggc Gly	gga Gly	ttt Phe	aca Thr	aac Asn 375	ctt Leu	ggg Gly	aat Asn	gcg Ala	ttg Leu 380	caa Gln	aag Lys	gga Gly	atg Met	1152
gtt Val 385	ttg Leu	gcg Ala	ttg Leu	agt Ser	att Ile 390	tgg Trp	gat Asp	gat Asp	cat His	gct Ala 395	gcg Ala	cag Gln	atg Met	ctt Leu	tgg Trp 400	1200
ttg Leu	gat Asp	agt Ser	tct Ser	tac Tyr 405	ccg Pro	ctc Leu	gat Asp	aaa Lys	gac Asp 410	cct Pro	tct Ser	caa Gln	cca Pro	ggt Gly 415	gtt Val	1248
aag Lys	agg Arg	ggc Gly	gcg Ala 420	tgt Cys	gct Ala	acc Thr	tct Ser	tct Ser 425	ggt Gly	aaa Lys	ccg Pro	tgc Ser	gat Asp 430	gtc Val	gag Glu	1296
aac Asn	cag Gln	tct Ser 435	ccg Pro	aat Asn	gcg Ala	tgc Ser	gtg Val 440	act Thr	ttt Phe	tgc Ser	aac Asn	att Ile 445	aag Lys	ttt Phe	ggg Gly	1344
gat Asp 450	att Ile	gga Gly	tgc Ser	act Thr	tat Tyr	tcc Ser 455	tct Ser	tag								1371

<210> 66

<211> 456

<212> PRT

<213> Pseudoplectania nigrella

<400> 66

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

	260		265		270										
Asp	Lys	Asp	Gly	Cys	Asp	Phe	Asn	Ser	Tyr	Arg	Met	Gly	Asn	Thr	Thr
		275					280					285			
Phe	Phe	Gly	Lys	Gly	Ala	Thr	Val	Asn	Thr	Asn	Ser	Lys	Phe	Thr	Val
	290					295					300				
Val	Thr	Gln	Phe	Ile	Thr	Ser	Asp	Asn	Thr	Ser	Thr	Gly	Ala	Leu	Lys
305					310					315					320
Glu	Ile	Arg	Arg	Leu	Tyr	Ile	Gln	Asn	Gly	Lys	Val	Ile	Gln	Asn	Ser
				325					330					335	
Lys	Ser	Asn	Ile	Ser	Gly	Met	Ser	Ala	Tyr	Asp	Ser	Ile	Thr	Glu	Asp
			340					345					350		
Phe	Cys	Ala	Ala	Gln	Lys	Thr	Ala	Phe	Gly	Asp	Thr	Asn	Asp	Phe	Lys
		355					360					365			
Ala	Lys	Gly	Gly	Phe	Thr	Asn	Leu	Gly	Asn	Ala	Leu	Gln	Lys	Gly	Met
	370					375					380				
Val	Leu	Ala	Leu	Ser	Ile	Trp	Asp	Asp	His	Ala	Ala	Gln	Met	Leu	Trp
385					390					395					400
Leu	Asp	Ser	Ser	Tyr	Pro	Leu	Asp	Lys	Asp	Pro	Ser	Gln	Pro	Gly	Val
				405					410					415	
Lys	Arg	Gly	Ala	Cys	Ala	Thr	Ser	Ser	Gly	Lys	Pro	Ser	Asp	Val	Glu
			420					425					430		
Asn	Gln	Ser	Pro	Asn	Ala	Ser	Val	Thr	Phe	Ser	Asn	Ile	Lys	Phe	Gly
		435					440					445			
Asp	Ile	Gly	Ser	Thr	Tyr	Ser	Ser								
	450					455									

<210> 67

<211> 951

<212> DNA

<213> Phytophthora infestans

<220>

<221> misc_feature

<222> (1)..(951)

<223> Secuencia de codificación parcial de CBH1

<400> 67

tgcgatgctg atggttgga cttcaactct taccgccagg gtaacacctc tttctatggt	60
gcagggtctta ccgtgaacac caacaaagtt ttcaccggtg taaccaatt catcaccaac	120
gatggaacag cttcaggtac cttgaaagaa atccgacgat tctatgttca gaatggcgtc	180
gtgattccaa actcgcaatc cacaatcgct ggagttccag gaaattccat caccgactct	240
ttctgtgccg cacaaaagac tgcttttggt gacaccaacg aattcgctac taagggaggt	300
cttgccacaa tgagcaaagc tttggcaaag ggtatggtag ttgtcatgtc catttgggat	360
gaccataccg ccaacatggt gtggctcgat gcccttacc cagcaaccaa atccccaagc	420
gccccagggtg tcaactgagg atcatgcagt gctacttcag gtaaccccggt tgatgttgaa	480
gccaattctc caggttcttc cgtcaccttc tcaaacatca agtgggggtcc catcaactct	540
acctacactg gatctggagc cgcaccaagt gttccaggca ctacaaccgt tagctcggca	600
cccgcatcga ctgcaacttc aggagctggt ggtgtcgcta agtatgccca atgtggaggt	660
actggataca gtggagctac cgcttgcggt tcaggcagca cctgtgttgc cctcaaccct	720
tactactccc aatgccaata gattgtttcc ctcaggagca attaggtttc caacctaagg	780
ggagagatct tcacaagtct gtacataggg tcagctaaat gttgatcatt catattcttt	840
catgtattta gttgttgaca atttgaagtt gcaagtcaag acgggaaaac agaagcagga	900
aatatatggg acataacaaa gtcaatcggt tacataagaa cttcttttaa a	951

REIVINDICACIONES

1. Polipéptido con actividad de celobiohidrolasa I, seleccionado del grupo que consiste en:
 - (a) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 95% de identidad con los aminoácidos 1 a 529 de SEC ID nº 4, y
 - (b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 95% de identidad con el polipéptido codificado por los nucleótidos 1 a 1.587 de SEC ID nº 3.
2. Polipéptido según la reivindicación 1, que comprende una secuencia de aminoácidos que tiene al menos 95% con los aminoácidos 1 a 529 de SEC ID nº 4.
3. Polipéptido según cualquiera de las reivindicaciones 1 ó 2, que consiste en los aminoácidos 1 a 529 de SEC ID nº 4.
4. Polinucleótido con una secuencia de nucleótidos que codifica para el polipéptido definido en cualquiera de las reivindicaciones 1-3.
5. Constructo de ácidos nucleicos que comprende la secuencia de nucleótidos definida en la reivindicación 4, ligada de forma operativa a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.
6. Vector de expresión recombinante que comprende el constructo de ácidos nucleicos definido en la reivindicación 5.
7. Célula huésped recombinante que comprende el constructo de ácidos nucleicos definido en la reivindicación 5.
8. Método para la producción de un polipéptido como está definido en cualquiera de las reivindicaciones 1-3, el método comprendiendo:
 - (a) cultivar una célula huésped recombinante tal y como se define en la reivindicación 8 bajo condiciones propicias para la producción del polipéptido; y
 - (b) recuperar el polipéptido.
9. Método para la producción in situ de un polipéptido como está definido en cualquiera de las reivindicaciones 1-3, el método comprendiendo:
 - (a) cultivar una célula huésped recombinante tal y como se define en la reivindicación 8 bajo condiciones propicias para la producción del polipéptido; y
 - (b) contactar el polipéptido con un sustrato deseado sin recuperación previa del polipéptido.
10. Método para la producción de etanol de biomasa, que incluye contacto de la biomasa con el polipéptido como se define en cualquiera de las reivindicaciones 1-3.
11. Uso del polipéptido tal y como se define en cualquiera de las reivindicaciones 1-3 para producir etanol.
12. Una planta transgénica, parte de planta o célula vegetal, que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de celobiohidrolasa I tal y como se define en cualquiera de las reivindicaciones 1-3, donde la secuencia de nucleótidos está presente en un vector de expresión que permite integración estable del vector en el genoma de la célula huésped.
13. Composición detergente que comprende un tensioactivo y el polipéptido según cualquiera de las reivindicaciones 1-3.