

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 787 225**

51 Int. Cl.:

**C12N 1/15** (2006.01)  
**C12N 15/80** (2006.01)  
**C07K 14/37** (2006.01)  
**C12P 21/02** (2006.01)  
**C12N 9/02** (2006.01)  
**C12P 23/00** (2006.01)  
**C12N 15/113** (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **14.12.2015 PCT/SG2015/050491**
- 87 Fecha y número de publicación internacional: **23.06.2016 WO16099401**
- 96 Fecha de presentación y número de la solicitud europea: **14.12.2015 E 15870471 (8)**
- 97 Fecha y número de publicación de la concesión europea: **15.04.2020 EP 3234105**

54 Título: **Métodos para ajustar los niveles de producción de carotenoides y composiciones en géneros de Rhodosporidium y Rhodotorula**

30 Prioridad:

**15.12.2014 US 201462091913 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**15.10.2020**

73 Titular/es:

**TEMASEK LIFE SCIENCES LABORATORY LIMITED (100.0%)**  
**1 Research Link The National University Of Singapore**  
**Singapore 117604, SG**

72 Inventor/es:

**LIU, YANBIN;**  
**KOH, CHONG MEI y**  
**JI, LIANGHUI**

74 Agente/Representante:

**SÁEZ MAESO, Ana**

ES 2 787 225 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

Métodos para ajustar los niveles de producción de carotenoides y composiciones en géneros de *Rhodospiridium* y *Rhodotorula*

Referencia cruzada a solicitudes relacionadas

La presente solicitud se relaciona con y reivindica la prioridad a la solicitud de patente provisional de los Estados Unidos núm. de serie 62/091,913, presentada el 15 de diciembre de 2014. Esta solicitud se incorpora en la presente descripción como referencia en su totalidad.

Envío de secuencia

La presente solicitud se presenta junto con una lista de secuencias en formato electrónico. El Listado de secuencias se titula 2577244PCTSequenceListing.txt, creado el 03 de noviembre de 2015 y tiene un tamaño de 184 kb. La información de la Lista de Secuencias en formato electrónico se incorpora en la presente descripción como referencia en su totalidad.

Antecedentes de la invención

La presente invención se refiere al campo de la biotecnología fúngica, más particularmente a métodos de ingeniería genética para la producción de carotenoides en huéspedes fúngicos seleccionados de los géneros *Rhodospiridium* y *Rhodotorula*.

Las publicaciones y otros materiales usados en la presente descripción para ilustrar los antecedentes de la invención, y en particular, los casos que proporcionan detalles adicionales con respecto a la práctica, se incorporan como referencia, y por conveniencia se referencian en el texto siguiente por el autor y la fecha y se enumeran alfabéticamente por el autor en la bibliografía anexa.

Está bien documentado que la producción de carotenoides se inicia con la biosíntesis de geranilgeranil difosfato (GGPP) catalizada por la GGPP sintasa para la condensación de C<sub>15</sub> farnesil difosfato (FPP) y C<sub>5</sub> isopentenil difosfato (IPP). Posteriormente, dos moléculas de GGPP se condensan adicionalmente para formar el precursor incoloro fitoeno, que es catalizado por la fitoeno sintasa. En hongos y eubacterias, la fitoeno desaturasa cataliza las 4 etapas de desaturación del fitoeno para producir el licopeno de color rojo, mientras que esto se realiza por fitoeno desaturasa y  $\gamma$ -caroteno o turasa y desaturasa separadas en plantas, algas y cianobacterias. El licopeno se cicla por la caroteno ciclasa para formar  $\gamma$ -caroteno y  $\delta$ -caroteno monocíclicos, y  $\alpha$ -caroteno y  $\beta$ -caroteno dicíclicos [1, 2]. Aguas arriba de la vía biosintética, el FPP se produce por la condensación catalizada por farnesil difosfato sintasa (FPS), de IPP y C<sub>10</sub> geranil difosfato (GPP), este último producido por la condensación catalizada por la GPP sintasa, de IPP y dimetilalil difosfato (DMAPP), el producto de la isomerasa IPP (IPI). El IPP y el DMAPP pueden sintetizarse a través de la vía del mevalonato (MVP) o de la vía de 2-C-metil-D-eritritol 4-fosfato/1-desoxi-D-xilulosa 5-fosfato (MEP/DOXP) [3, 4].

Los carotenoides son tetraterpenoides de 40 carbonos (C<sub>40</sub>) [5]. Los carotenoides no oxigenados, tales como el  $\gamma$ -caroteno, el  $\beta$ -caroteno y el licopeno se conocen como carotenos. Las modificaciones enzimáticas adicionales de los carotenos producen moléculas que contienen oxígeno, tales como luteína, retinol (vitamina A), zeaxantina y astaxantina [6, 7]. La biosíntesis de carotenoides se produce en todos los organismos fotosintéticos [8] y en muchos microorganismos no fotosintéticos, tales como bacterias y hongos [1, 5, 9, 10] y algunos insectos [11].

Los carotenoides desempeñan un papel importante en la salud y el desarrollo humano y animal [12-15]. Por ejemplo, una mayor ingestión de carotenoides en la dieta se asoció con un riesgo menor de degeneración macular relacionada con la edad (ADM) [13]; la deficiencia de vitamina A se asocia con un crecimiento anormal del esqueleto y los dientes y la infertilidad en ratas [14]; el retinal (retinaldehído) es esencial para la visión, mientras que el ácido retinoico es esencial para la salud de la piel, la remineralización de los dientes y el crecimiento óseo [16]; la ingestión de licopeno se relaciona con riesgo menor de cáncer de próstata [17]. Los carotenoides son colorantes naturales con muchos colores disponibles [18-20]. Los carotenoides son precursores para la producción de compuestos aromáticos valiosos [21]. El  $\beta$ -caroteno puede escindirse por la citocromo oxidasa P450 para producir retinal (retinaldehído) [16], que es esencial para la visión y cuando se convierte en ácido retinoico, es esencial para la salud cutánea, la remineralización de los dientes y el crecimiento óseo. Por lo tanto, los carotenoides son valiosos aditivos de alimentos y piensos, nutracéuticos y cosmeceúticos.

El retinol, el retinal y el ácido retinoico se conocen como retinoides, que se derivan del fallo de la salud cutánea, la remineralización de los dientes y el crecimiento óseo. El retinol y el retinal son interconvertibles y se catalizan por la alcohol deshidrogenasa y las deshidrogenasa/reductasas de cadena corta, mientras que las familias de enzimas aldehído deshidrogenasa y citocromo P450 catalizan la oxidación irreversible de retinal a ácido retinoico. La identificación de enzimas que catalizan la oxidación del retinol *in vivo* ha sido controversial, lo que se debe en parte a la dificultad por la naturaleza reversible de esta reacción [22].

*Rhodosporidium* y *Rhodotorula* son dos géneros fúngicos que pertenecen al subfilum *Pucciniomycotina*. Pueden cultivarse en forma de células individuales en una densidad celular muy alta en fermentadores a una velocidad de crecimiento rápida y acumular niveles altos de triacilglicéridos [23-26]. *Rhodosporidium* y *Rhodotorula* son capaces de producir niveles altos de carotenoides [27-30], con beta-caroteno, gamma-caroteno, torularodina y toruleno como componentes principales [31]. Torularodina y toruleno son colorantes potenciales e inductores de la expresión génica. Aparte de la identificación de un supuesto homólogo de *CAR2* [32], no hay ningún informe sobre la vía biosintética de carotenoides en *Rhodosporidium* y *Rhodotorula*. Cualquier método que mejore la productividad y la pureza del producto de los carotenoides y sus derivados tiene valor y significado comercial.

## Resumen de la invención

La presente invención se refiere al campo de la biotecnología fúngica, más particularmente a métodos de ingeniería genética para la producción de carotenoides en huéspedes fúngicos seleccionados de los géneros *Rhodosporidium* y *Rhodotorula*.

Así, en un aspecto, la presente invención proporciona un método para ajustar el nivel de producción y composición de carotenoides en un huésped fúngico. De acuerdo con este aspecto, el método comprende la manipulación genética de uno o más de los polinucleótidos implicados en la biosíntesis de carotenoides. En una modalidad, los carotenoides son licopeno, beta-caroteno, gamma-caroteno, toruleno, torularodina o derivados de estos. En algunas modalidades, un derivado es un derivado hidroxilado, un derivado glicosilado o un derivado oxidado. En otra modalidad, el huésped fúngico es *Rhodosporidium* o *Rhodotorula*. En algunas modalidades, el método comprende manipular genéticamente uno o más polinucleótidos implicados en la biosíntesis de carotenoides en un huésped fúngico y hacer crecer el huésped fúngico para producir los carotenoides, por lo que se ajusta el nivel de producción o la composición de los carotenoides.

En una modalidad, la manipulación genética comprende la regulación negativa de uno o más polinucleótidos seleccionados de SEQ ID NOs: 1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 19 y 20, o un homólogo que comparte al menos el 75 % de identidad de nucleótidos con los mismos en un huésped fúngico. En otra modalidad, la manipulación genética comprende la regulación negativa de uno o más polinucleótidos que codifican polipéptidos seleccionados de SEQ ID NOs: 2, 4, 6, 12, 15, 18 y 21 o un homólogo que comparte al menos el 75 % de identidad con los mismos en un huésped fúngico. En algunas modalidades, la regulación negativa se compara con un huésped fúngico sin la manipulación genética. En otras modalidades, uno o más polinucleótidos se regulan negativamente por ARNi, un represor transcripcional artificial o un promotor débil.

En una modalidad, la manipulación genética es la inactivación total de la función enzimática en un huésped fúngico. En algunas modalidades, la inactivación se logra mediante la delección de todo o una parte de uno o más polinucleótidos seleccionados de las SEQ ID NOs: 10, 11, 16, 17, 19 y 20 o un homólogo que comparte al menos el 75 % de identidad de nucleótidos con los mismos. En otras modalidades, la inactivación se logra mediante la delección de todo o una parte de uno o más polinucleótidos que codifican polipéptidos seleccionados entre las SEQ ID NOs: 12, 18 y 21 o un homólogo que comparte al menos el 75 % de identidad de nucleótidos con los mismos. En una modalidad, la delección se realiza mediante el uso de una técnica de recombinación homóloga. En otra modalidad, la delección se ayuda mediante el uso de una nucleasa artificial. En una modalidad adicional, la nucleasa artificial es una nucleasa de dedos de zinc (ZFN) o un complejo Cas9-ARNg.

En una modalidad, la manipulación genética implica la sobreexpresión de uno o más polinucleótidos seleccionados de SEQ ID NOs: 1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 19 y 20, o un homólogo que comparte al menos el 75 % de identidad de nucleótidos con los mismos. En otra modalidad, la manipulación genética implica la sobreexpresión de uno o más polinucleótidos que codifican polipéptidos seleccionados de SEQ ID NOs: 2, 4, 6, 12, 15, 18 y 21; o un homólogo que comparte al menos el 75 % de identidad con los mismos. En algunas modalidades, la sobreexpresión está mediada por la introducción de un casete genético sintético en una célula fúngica huésped. En otra modalidad, el casete comprende un promotor heterólogo unido operativamente al polinucleótido, opcionalmente unido operativamente a un terminador transcripcional. En algunas modalidades, la sobreexpresión se compara con un huésped fúngico sin manipulación genética.

En algunas modalidades, la manipulación genética implica una combinación de las manipulaciones genéticas descritas previamente. En una modalidad, la manipulación genética comprende la regulación negativa de uno o más de los polinucleótidos y la sobreexpresión de uno o más polinucleótidos diferentes. En otra modalidad, la manipulación genética comprende la inactivación total de la función enzimática de uno o más polipéptidos codificados por uno o más polinucleótidos y la sobreexpresión de uno o más de diferentes polinucleótidos. En una modalidad adicional, la manipulación genética comprende la inactivación total de la función enzimática de uno o más polipéptidos codificados por uno o más polinucleótidos y la regulación negativa de uno o más polinucleótidos diferentes. En una modalidad adicional, la manipulación genética comprende la inactivación total de la función enzimática de uno o más polipéptidos codificados por uno o más polinucleótidos, la sobreexpresión de uno o más de diferentes polinucleótidos y la regulación negativa de uno o más polinucleótidos diferentes.

En algunas modalidades, los polinucleótidos descritos en la presente descripción que se han incorporado de manera estable en el genoma fúngico están unidos operativamente a un promotor que permite la expresión eficiente en especies

- del género *Rhodospiridium* y el género *Rhodotorula*. Los promotores para cada polinucleótido incorporado pueden ser iguales o diferentes. En algunas modalidades, los promotores son promotores encontrados en especies del género *Rhodospiridium* y el género *Rhodotorula*. Los ejemplos de promotores adecuados incluyen, pero no se limitan a, promotores de los genes siguientes que codifican las proteínas siguientes: gliceraldehído 3-fosfato deshidrogenasa (GPD), proteína transportadora de acil-CoA (ACP), ácido graso desaturasa, factor de elongación de traducción (TEF), piruvato descarboxilasa (PDC), enolasa (2-fosfoglicerato deshidratasa) (ENO), peptidilprolil isomerasa (PPI), acetil-CoA carboxilasa (ACC) o transaldolasa. En otras modalidades, los genes descritos en la presente descripción también incluyen un terminador transcripcional de ARNm que puede encontrarse en cualquier especie eucariota y sus virus de ADN.
- En otra modalidad, la presente invención proporciona un método para producir carotenoides que comprende cultivar una célula fúngica huésped descrita en la presente descripción en condiciones adecuadas para producir carotenoides. Puede usarse cualquier medio con al menos el 5 % de fuente de carbono. En algunas modalidades, la fuente de carbono es glucosa, manosa, glicerol, sacarosa, xilosa o sus combinaciones. En una modalidad, el medio es MinCAR que contiene 30-100 g de glucosa, 1,5 g de extracto de levadura, 0,5 g de  $(\text{NH}_4)_2\text{SO}_4$ , 2,05 g de  $\text{K}_2\text{HPO}_4$ , 1,45 g de  $\text{KH}_2\text{PO}_4$ , 0,6 g de  $\text{MgSO}_4$ , 0,3 g de  $\text{NaCl}$ , 10 mg de  $\text{CaCl}_2$ , 1 mg de  $\text{FeSO}_4$ , 0,5 mg de  $\text{ZnSO}_4$ , 0,5 mg de  $\text{CuSO}_4$ , 0,5 mg de  $\text{H}_3\text{BO}_3$ , 0,5 mg de  $\text{MnSO}_4$ , 0,5 mg de  $\text{NaMoO}_4$  (por litro). El medio se ajusta preferentemente a pH 5-7. En algunas modalidades, el cultivo celular se realiza preferentemente a 25 °-35 °C. En otras modalidades, el cultivo se realiza preferentemente en una condición con iluminación.
- Breve descripción de las figuras
- La figura 1 muestra la organización de pRH201. LB: borde izquierdo de T-DNA; RB: borde derecho de T-DNA; *P<sub>gpd</sub>*: Promotor de 595 pb de *Um<sub>gpd1</sub>*; *P<sub>GPD1</sub>*: Promotor de 795 pb de *Rt<sub>GPD1</sub>*; *hpt-3*: gen de resistencia a la higromicina optimizado por codones que se basa en el sesgo de uso de codones en *R. toruloides*; *Tnos*: terminador del gen de la nopalina sintasa de *A. tumefaciens*. Los sitios de corte de enzimas de restricción únicos se muestran en rojo. *loxP-RE* y *loxP-LE* y los sitios de reconocimiento de mutantes para Cre recombinasa. *Sp/Str* son genes de resistencia para espectinomomicina y estreptomomicina; *eGFP-His6* es un gen de codón adaptado que codifica la proteína de fusión de etiqueta eGFP-histidina de *R. toruloides*; *35S*: terminador del gen *35S* del virus del mosaico de la coliflor.
- Las Figuras 2A-2E muestran la identificación de mutantes RCM y la caracterización del gen *CAR1*. Figura 2A: Fenotipos de color de colonias de mutantes RCM. Todas las cepas se inocularon en estrías en placas PDA y se incubaron a 28 °C durante 2 días. Figura 2B: Diagrama esquemático de *CAR1* y su estrategia de delección. Figura 2C: Análisis de Southern blot del mutante nulo del candidato *CAR1* ( $\Delta$  *car1*). Las secuencias homólogas usadas para la delección de *CAR1* fueron de 1036 pb (*CAR1L*) y de 830 pb (*CAR1R*) de longitud, que oscilan de -89 a +947 y de +2098 a +2928 del codón de inicio traduccional. El fragmento de ADN marcado con digoxigenina *CAR1R* se usó como sonda para la detección de ADN total *HindIII*. Figura 2D: Colores de colonias de WT, mutante nulo ( $\Delta$  *car1*) y cepa de complementación ( $\Delta$  *car1C*) cultivadas en placa PDA. Figura 2E: Perfiles de carotenoides en la cepa de tipo salvaje de *R. toruloides*,  $\Delta$  *car1* y  $\Delta$  *car1C*. Se analizó el contenido de cuatro componentes principales de carotenoides.
- Las Figuras 3A-3D muestran el grupo de genes biosintéticos de carotenoides en *R. toruloides*. Figura 3A: Organización genómica de grupos de genes biosintéticos de carotenoides en 5 hongos carotenogénicos, *Blakeslea trispora*, *Fusarium fujikuroi*, *Phycomyces blakesleeanus* y *Sporobolomyces roseus*. Figura 3B: Delección de *CAR3*, *CCD1* y *CDS1*. El panel superior muestra los esquemas de delección y el panel inferior muestra el análisis de Southern blot de mutantes de desactivación génica. Las barras negras indican las sondas usadas para la hibridación mediante Southern blot. Figura 3C: Fenotipo de color de colonias de mutantes nulos implicados en la vía biosintética de carotenoides en *R. toruloides*. Figura 3D: Perfiles de carotenoides en cepa de tipo salvaje y mutantes nulos de *R. toruloides* de genes *CCD1* y *CDS1*.
- La figura 4 muestra los perfiles de carotenoides en cepa de tipo salvaje de *R. toruloides*, mutante nulo *Ald1* (*ald1*) y sobreexpresión de *Ald1OE*.
- La figura 5 muestra los niveles relativos de ARNm de genes biosintéticos de carotenoides después de cambiar a iluminación. El nivel de expresión de cada gen se realizó mediante qRT-PCR y se normalizó para el gen Actin (*ACT1*).
- Las Figuras 6A-6C muestran la caracterización de *Roc1*. Figura 6A: Estructura esquemática de *ROC1* y estrategia de delección de genes. Figura 6B: Análisis de árbol filogenético de reguladores negativos de la biosíntesis de carotenoides fúngicos. El número de registro de NCBI GenBank se siguió del nombre del gen. Figura 6C: Comparación de dominios de dedo RING y dominios LON. Las secuencias de consenso se indican en la línea inferior de cada una. Los números de registro de GenBank (con secuencias en la Figura 6C establecidas en las secuencias que se indican): *B. trispora* *crgA*: CAE51310.1 (SEQ ID NO:26); *M. circinelloides* *crgA*: CAB61339.2 (SEQ ID NO:28); *A. fumigatus* *crgA*: XP\_755380.1 (SEQ ID NO:25); *F. fujikuroi* *carS*: CCP50075.1 (SEQ ID NO:27); *P. blakesleeanus* *carS*: ADU04395.1 (SEQ ID NO:29); *U. maydis*: EAK85777.1 (SEQ ID NO:31); *R. toruloides*: (SEQ ID NO:30)
- Las Figuras 7A-7E muestra la delección de *ROC1*. Figura 7A: Diferencias de color de colonias entre recombinación ectópica y homóloga de *ROC1* de desactivación génica. Figura 7B: Verificación mediante Southern blot de mutantes de delección génica. Se usó Digoxigenin-labeled DAN Molecular Weight Marker VII (Roche Diagnosis, Estados Unidos) como marcador. Figura 7C: Morfología celular del mutante nulo y tipo salvaje de *roc1*. La barra representa 10  $\mu\text{m}$ . Figura 7D: transcritos de ARNm de genes de carotenoides en la cepa mutante y de tipo salvaje de *roc1* en condiciones de iluminación.

Figura 7E: Comparación de la producción de carotenoides en la cepa de tipo salvaje, mutante de desactivación génica ( $\Delta roc1$ ) y su cepa de complementación ( $\Delta roc1C$ ).

Descripción detallada de la invención

5

La presente invención se refiere al campo de la biotecnología fúngica, más particularmente a métodos de ingeniería genética para la producción de carotenoides en huéspedes fúngicos seleccionados de los géneros *Rhodospiridium* y *Rhodotorula*.

10

A menos que se defina de cualquier otra manera, todos los términos técnicos y científicos usados en la presente descripción tienen el mismo significado que el que se conoce comúnmente por el experto en la técnica a la que pertenece la invención.

15

El término "aproximadamente" significa dentro de un intervalo de un valor estadísticamente significativo. Dicho intervalo puede estar dentro de un orden de magnitud, preferentemente dentro del 50 %, más preferentemente dentro del 20 %, aún más preferentemente dentro del 10 %, y aún más preferentemente dentro del 5 % de un valor o intervalo dado. La variación permisible abarcada por el término "aproximadamente" depende del sistema particular en estudio, y puede apreciarse fácilmente por un experto en la técnica.

20

Como se usa en la presente descripción, "alelo" se refiere a cualquiera de una o más formas alternativas de un locus génico, todos los alelos se relacionan con un rasgo o característica. En una célula u organismo diploide, los dos alelos de un gen dado ocupan los loci correspondientes en un par de cromosomas homólogos.

25

El término "regulación negativa" se refiere a la disminución del nivel de expresión de un polinucleótido, como un gen, en comparación con un control que usa cualquier método conocido en la técnica, como ARNi, un represor transcripcional artificial para atacar específicamente el gen si es de interés, como ZFN y Cas9 que se fusionan a un dominio represor transcripcional y se unen a una secuencia de ADN específica en un promotor de un gen o secuencia codificante para lograr la regulación negativa; o el uso de un promotor más débil para impulsar la expresión del gen de interés. El término "regulado negativamente" se usa en la presente descripción para indicar que la expresión del gen objetivo se reduce en 1-100 %. Por ejemplo, la expresión puede reducirse en aproximadamente 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 %.

30

Como se usa en la presente descripción, una "ventana de comparación" se refiere a un segmento conceptual de al menos 6 posiciones contiguas, generalmente de aproximadamente 50 a aproximadamente 100, más generalmente de aproximadamente 100 a aproximadamente 150, en el que una secuencia se compara con una secuencia de referencia del mismo número de posiciones contiguas después que las dos secuencias se alinean de manera óptima. La ventana de comparación puede comprender adiciones o deleciones (es decir, brechas) de aproximadamente el 20 % o menos en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para la alineación óptima de las dos secuencias. Los expertos en la técnica deben hacer referencia a los métodos detallados usados para la alineación de secuencias, como en Wisconsin Genetics Software Package Release 7.0 (Genetics Computer Group, 575 Science Drive Madison, Wisconsin, Estados Unidos).

35

40

Una molécula "dsRNA" or "ARNi", como se usa en la presente descripción en el contexto de ARNi, se refiere a un compuesto, el cual es capaz de regular negativamente o reducir la expresión de un gen o la actividad del producto de dicho gen en una medida suficiente para lograr un efecto biológico o fisiológico deseado. El término "dsRNA" o "molécula de ARNi", como se usa en la presente descripción, se refiere a uno o más de un dsRNA, siRNA, miRNA, hpRNA, ihpRNA.

45

El término "expresión" con respecto a una secuencia génica se refiere a la transcripción del gen y, según sea apropiado, a la traducción del transcrito de ARNm resultante a una proteína. Por lo tanto, como quedará claro por el contexto, la expresión de una secuencia codificante de proteínas es el resultado de la transcripción y traducción de la secuencia codificante.

50

Como se usa en la presente descripción, "genotipo" se refiere a la constitución genética de una célula u organismo.

55

El término "homólogo" como se usa en la presente descripción se refiere a un gen relacionado con un segundo gen por descendencia de una secuencia de ADN ancestral común. El término, homólogo, puede aplicarse a la relación entre genes separados por el evento de especiación (ortólogo) o a la relación entre genes separados por el evento de duplicación genética (parálogo). El término homólogo se usa genéricamente para referirse a todas las especies.

60

"Enlace operable" o "unido operativamente" como se usa en la presente descripción, se entiende que significa, por ejemplo, la disposición secuencial de un promotor y el ácido nucleico a expresarse y, si es apropiado, elementos reguladores adicionales tales como, para ejemplo, un terminador, de tal manera que cada uno de los elementos reguladores pueda cumplir su función en la expresión recombinante del ácido nucleico para formar dsRNA. Esto no requiere necesariamente un enlace directo en el sentido químico. Las secuencias de control genético como, por ejemplo, las secuencias potenciadoras, también pueden ejercer su función en la secuencia objetivo desde posiciones que están algo distantes, o incluso desde otras moléculas de ADN (localización cis o trans). Las disposiciones preferidas son

65

aquellas en las que la secuencia de ácido nucleico que va a expresarse de forma recombinante se coloca aguas abajo de la secuencia que actúa como promotor, de modo que las dos secuencias se unen covalentemente entre sí. Las secuencias reguladoras o de control pueden colocarse en el lado 5' de la secuencia de nucleótidos o en el lado 3' de la secuencia de nucleótidos como se conoce bien en la técnica.

5

El término "sobrexpresión" se refiere al aumento en el nivel de expresión de un polinucleótido, como un gen, en comparación con un control que usa cualquier método conocido en la técnica, como un activador transcripcional artificial para atacar específicamente el gen si es de interés, tales como ZFN y Cas9 que se fusionan a un dominio activador transcripcional y se unen a una secuencia de ADN específica en un promotor de un gen; o el uso de un promotor más fuerte para impulsar la expresión del gen de interés. El término "sobrexpresión" se usa en la presente descripción para indicar que la expresión del gen objetivo aumenta en un 25 % o más en comparación con un control. Por ejemplo, la expresión puede aumentarse en aproximadamente el 25 %, 50 %, 100 %, 200 %, 500 %, 1000 % y así sucesivamente.

10

Como se usa en la presente descripción, el "fenotipo" se refiere a las características detectables de una célula u organismo, cuyas características son la manifestación de la expresión génica.

15

Los términos "polinucleótido", "ácido nucleico" y "molécula de ácido nucleico" se usan indistintamente en la presente descripción para referirse a un polímero de nucleótidos que puede ser un arreglo lineal natural o sintético y secuencial de nucleótidos y/o nucleósidos, que incluyen el ácido desoxirribonucleico, el ácido ribonucleico y derivados de estos. Incluye ADN cromosómico, plásmidos autorreplicantes, polímeros infecciosos de ADN o ARN y ADN o ARN que desempeñan un papel primordialmente estructural. A menos que se indique lo contrario, los ácidos nucleicos o el polinucleótido se escriben de izquierda a derecha en una orientación de 5'a 3', los nucleótidos se denominan por sus códigos de una sola letra comúnmente aceptados. Los intervalos numéricos incluyen los números que definen el intervalo.

20

Los términos "polipéptido", "péptido" y "proteína" se usan indistintamente en la presente descripción para referirse a un polímero de residuos de aminoácidos. Los términos se aplican a polímeros de aminoácidos en los que uno o más residuos de aminoácidos es un análogo químico artificial de un aminoácido de origen natural correspondiente, así como a polímeros de aminoácidos de origen natural. Los aminoácidos pueden denominarse por sus símbolos de tres letras o una letra comúnmente conocidos. Las secuencias de aminoácidos se escriben de izquierda a derecha en orientación de amino a carboxilo, respectivamente. Los intervalos numéricos incluyen los números que definen el intervalo.

25

30

Como se usa en la presente descripción, el término "identidad de secuencia", "similitud de secuencia" u "homología" se usa para describir relaciones de secuencia entre dos o más secuencias de nucleótidos. El porcentaje de "identidad de secuencia" entre dos secuencias se determina mediante la comparación de dos secuencias alineadas óptimamente sobre una ventana de comparación como la longitud total de una SEQ ID NO: referenciada, en donde la porción de la secuencia en la ventana de comparación puede comprender adiciones o deleciones (es decir, brechas) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para la alineación óptima de las dos secuencias. El porcentaje se calcula mediante la determinación del número de posiciones en las que el residuo de aminoácido o de base de ácido nucleico idéntico aparece en ambas secuencias para producir el número de posiciones coincidentes, tras dividir el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación, y multiplicar el resultado por 100 para obtener el porcentaje de identidad de secuencia. Se plantea que una secuencia que es idéntica en cada posición en comparación con una secuencia de referencia es idéntica a la secuencia de referencia y viceversa. Se plantea que una primera secuencia de nucleótidos cuando se observa en la dirección 5' a 3' es un "complemento" de, o es complementaria a, una segunda secuencia o de referencia de nucleótidos observada en la dirección 3' a 5' si la primera secuencia de nucleótidos muestra complementariedad completa con la segunda secuencia o de referencia. Tal como se usa en la presente descripción, se plantea que las moléculas de secuencias de ácido nucleico presentan "complementariedad completa" cuando cada nucleótido de una de las secuencias que se leen en la dirección 5' a 3' es complementaria a cada nucleótido de la otra secuencia cuando se lee en la dirección 3' a 5'. Una secuencia de nucleótidos que es complementaria a una secuencia de nucleótidos de referencia presentará una secuencia idéntica a la secuencia complementaria inversa de la secuencia de nucleótidos de referencia. Estos términos y descripciones están bien definidos en la técnica y se conocen fácilmente por los expertos en la técnica.

35

40

45

50

El término "inactivación total de la función enzimática", como se usa en la presente descripción, se refiere a la pérdida completa de la función o actividad enzimática para una enzima dada o un polipéptido dado que tiene función o actividad enzimática en comparación con un control en el que la enzima no ha sido inactivada. La función enzimática puede inactivarse totalmente mediante la deleción de todo o parte de un polinucleótido que codifica el polipéptido que tiene función enzimática. "Parte de un polinucleótido", como se usa en la presente descripción con respecto a la inactivación de la función enzimática, se refiere a una porción del polinucleótido, cuya deleción es suficiente para causar la inactivación total de la función enzimática. Una parte del polinucleótido podría ser un solo nucleótido o múltiples nucleótidos siempre que la deleción de como resultado la terminación prematura del polipéptido o una mutación de desplazamiento de marco, cada uno de los cuales resultaría en un polipéptido inactivo. Una parte del polinucleótido podría ser una parte que codifica al menos el 10 % pero menos del 90 % del polipéptido, lo que también da como resultado la pérdida de la función enzimática.

55

60

65

El término "ajustar", como se usa en una modalidad de la presente descripción, se refiere a aumentar o disminuir el rendimiento de un carotenoide en al menos el 25 %, en al menos el 30 %, en al menos el 35 %, en al menos el 40 %, en

al menos el 45 %, o aproximadamente el 50 % o más en un huésped fúngico manipulado genéticamente en comparación con un huésped fúngico control no manipulado genéticamente. El término "ajustado", como se usa en una modalidad de la presente descripción, se refiere a un rendimiento aumentado o disminuido de un carotenoide en al menos el 25 %, en al menos el 30 %, en al menos el 35 %, en al menos el 40 %, en al menos el 45 %, o aproximadamente el 50 % o más en un huésped fúngico manipulado genéticamente en comparación con un huésped fúngico control no manipulado genéticamente.

El término "ajustar", como se usa en otra modalidad de la presente descripción, se refiere a un cambio en la composición de carotenoides en un huésped fúngico manipulado genéticamente en comparación con un huésped fúngico control no manipulado genéticamente. El término "ajustado", como se usa en otra modalidad de la presente descripción, se refiere a una composición cambiada de carotenoides en un huésped fúngico manipulado genéticamente en comparación con un huésped fúngico control no manipulado genéticamente.

Así, en un aspecto, la presente invención proporciona un método para ajustar el nivel de producción y composición de carotenoides en un huésped fúngico. De acuerdo con este aspecto, el método comprende la manipulación genética de uno o más de los polinucleótidos implicados en la biosíntesis de carotenoides. En una modalidad, los carotenoides son licopeno, beta-caroteno, gamma-caroteno, toruleno, torularodina o derivados de estos. En algunas modalidades, un derivado es un derivado hidroxilado, un derivado glicosilado o un derivado oxidado. En otra modalidad, el huésped fúngico es *Rhodospiridium* o *Rhodotorula*. En algunas modalidades, el método comprende manipular genéticamente uno o más polinucleótidos implicados en la biosíntesis de carotenoides en un huésped fúngico y hacer crecer el huésped fúngico para producir los carotenoides, por lo que se ajusta el nivel de producción o la composición de los carotenoides.

En una modalidad, la manipulación genética comprende la regulación negativa de uno o más polinucleótidos seleccionados de SEQ ID NOs:1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 19 y 20, o un homólogo que comparte al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o al menos el 99 % de identidad de nucleótidos con los mismos en un huésped fúngico. En otra modalidad, la manipulación genética comprende la regulación negativa de uno o más polinucleótidos que codifican polipéptidos seleccionados de SEQ ID NO:2, 4, 6, 12, 15, 18 y 21, o un homólogo que comparte al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o al menos el 99 % de identidad de aminoácidos con los mismos en un huésped fúngico. En algunas modalidades, la regulación negativa se compara con un huésped fúngico sin manipulación genética. En otras modalidades, uno o más polinucleótidos se regulan negativamente por ARNi, un represor transcripcional artificial o un promotor débil.

La regulación negativa de un polinucleótido de la presente invención puede lograrse mediante el uso de técnicas bien conocidas, que incluyen, pero no se limitan a, técnicas de ARNi, como dsRNA, miRNA, siRNA, smRNA, hpRNA o ihpRNA (denominadas colectivamente moléculas de ARNi), supresión de sentido (co-supresión), antisentido, y similares. Dichas técnicas se describen en la Patente de los Estados Unidos núm. 7,312,323 y las referencias citadas allí. Por ejemplo, la reducción podría lograrse, por ejemplo, con la transformación de una célula fúngica huésped para comprender un promotor y otras regiones reguladoras 5' y/o 3' descritas en la presente descripción unidas a una secuencia de nucleótidos antisentido, horquilla, molécula interferente de ARN, ARN bicatenario, microARN u otra molécula de ácido nucleico, de manera que la expresión de la molécula preferida por los tejidos interfiera con la traducción del ARNm de la secuencia de ADN nativa o inhíba la expresión de la secuencia de ADN nativa en las células vegetales. Para una descripción más detallada de las técnicas de ARNi o técnicas de microARN, véase, por ejemplo, las Patentes de los Estados Unidos núms. 5,034,323; 6,326,527; 6,452,067; 6,573,099; 6,753,139; y 6,777,588. Véase también las publicaciones internacionales núms. WO 97/01952, WO 98/36083, WO 98/53083, WO 99/32619 y WO 01/75164; y la Publicación de la Solicitud de Patente de los Estados Unidos núms. 2003/0175965, 2003/0175783, 2003/0180945, 2004/0214330, 2005/0244858, 2005/0277610, 2006/0130176, 2007/0265220, 2008/0313773, 2009/0094711, 2009/0215860, 2009/0308041, 2010/0058498 y 2011/0091975. Las moléculas de ARNi o las moléculas de microARN (denominadas colectivamente en la presente descripción como moléculas de ARNi) pueden prepararse por un experto en la técnica mediante el uso de técnicas bien conocidas en la técnica, que incluyen técnicas para la selección y prueba de moléculas de ARNi y moléculas de microARN que son útiles para la regulación negativa de un polinucleótido de la presente invención. Véase, por ejemplo, Wesley y otros (2001), Mysara y otros (2011) y Yan y otros (2012).

Típicamente se ha encontrado que el dsRNA de 200-700 pb es particularmente adecuado para inducir ARNi en vegetales. También se ha encontrado que los ARN de horquilla que contienen un intrón, por ejemplo, un constructo que comprende una secuencia codificante de ARN en una dirección de sentido operativamente unida a un intrón operativamente unido a una secuencia codificante de ARN en una dirección antisentido o *viceversa* que es capaz de formar un ARN de intrón-horquilla (ihpRNA), es adecuado para inducir ARNi en vegetales. Véase, por ejemplo, Wang y otros (2000), Fuentes y otros (2006), Bonfim y otros (2007) Vanderschuren y otros (2007a, 2007b), Zrachya y otros (2007). Por ejemplo, puede prepararse un constructo de ácido nucleico que incluye un ácido nucleico que se transcribe en un ARN que puede hibridarse consigo mismo, por ejemplo, un ARN bicatenario que tiene una estructura de horquilla. Además, las estructuras de horquilla pueden prepararse como se describe por Guo y otros. (2003).

Por ejemplo, puede prepararse un constructo de ácido nucleico que incluye un nucleico que se transcribe en un ARN que puede hibridarse consigo mismo, por ejemplo, un ARN bicatenario que tiene una estructura de horquilla. En algunas modalidades, una hebra de la porción del tallo de un ARN bicatenario comprende una secuencia que es similar o idéntica

a la secuencia codificante de sentido, o un fragmento de esta, de un polinucleótido como se describe en la presente descripción, y que es de aproximadamente 10 nucleótidos a aproximadamente 1800 nucleótidos de longitud. La longitud de la secuencia que es similar o idéntica a la secuencia codificante de sentido puede ser de 10 nucleótidos a 1000 nucleótidos, de 15 nucleótidos a 600 nucleótidos, de 20 nucleótidos a 500 nucleótidos, o de 25 nucleótidos a 100 nucleótidos, o cualquier longitud dentro de los 10 nucleótidos a 2500 nucleótidos. La otra hebra de la porción del tallo de un ARN bicatenario comprende una secuencia que es similar o idéntica a la hebra antisentido, o un fragmento de esta, de la secuencia codificante del polipéptido de interés, y puede tener una longitud que es más corta, igual o mayor que la longitud correspondiente de la secuencia de sentido. En algunos casos, una hebra de la porción del tallo de un ARN bicatenario comprende una secuencia que es similar o idéntica a la región no traducida 3' o 5', o un fragmento de esta, del ARNm que codifica un polipéptido descrito en la presente descripción, y la otra la hebra de la porción del tallo del ARN bicatenario comprende una secuencia que es similar o idéntica a la secuencia que es complementaria a la región no traducida 3' o 5', respectivamente, o un fragmento de esta, del ARNm que codifica un polipéptido descrito en la presente descripción. En otras modalidades, una hebra de la porción del tallo de un ARN bicatenario comprende una secuencia que es similar o idéntica a la secuencia de un intrón o un fragmento de este en el pre-ARNm transcrito a partir de un polinucleótido descrito en la presente descripción, y la otra hebra de la porción del tallo comprende una secuencia que es similar o idéntica a la secuencia que es complementaria a la secuencia del intrón o fragmento de este en el pre-ARNm.

La porción del bucle de un ARN bicatenario puede ser de 3 nucleótidos a 5000 nucleótidos, por ejemplo, de 3 nucleótidos a 2500 nucleótidos, de 15 nucleótidos a 1000 nucleótidos, de 20 nucleótidos a 500 nucleótidos, o de 25 nucleótidos a 200 nucleótidos, o cualquier longitud dentro de los 3 nucleótidos a 5000 nucleótidos. La porción de bucle del ARN puede incluir un intrón o un fragmento de este. Un ARN bicatenario puede tener cero, uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, o más estructuras de horquilla.

En una modalidad, la manipulación genética es la inactivación total de la función enzimática en un huésped fúngico. En algunas modalidades, la inactivación se logra mediante la delección de todo o parte de uno o más polinucleótidos seleccionados de SEQ ID NOs:10, 11, 16, 17, 19 y 20, o un homólogo que comparte al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o al menos el 99 % de identidad de nucleótidos con los mismos. En otras modalidades, la inactivación se logra mediante la delección de todo o parte de uno o más polinucleótidos que codifican polipéptidos seleccionados de SEQ ID NO:12, 18 y 21 o un homólogo que comparte al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o al menos el 99 % de identidad de aminoácidos con los mismos. En una modalidad, la delección se realiza mediante el uso de técnicas de recombinación homóloga. Las técnicas de recombinación homóloga se conocen bien por el experto en la técnica. En otra modalidad, la delección se ayuda mediante el uso de una nucleasa artificial. En una modalidad adicional, la nucleasa artificial es una nucleasa de dedos de zinc (ZFN) o un complejo Cas9-ARNg. Las tecnologías de nucleasa artificial se conocen bien por el experto en la técnica. Véase, por ejemplo, Durai y otros (2005), Makarova y otros (2011) y Mali y otros (2013).

En una modalidad, la manipulación genética implica la sobreexpresión de uno o más polinucleótidos seleccionados de SEQ ID NOs:1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 16, 17, 19 y 20, o un homólogo que comparte al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o al menos el 99 % de identidad de nucleótidos con los mismos. En otra modalidad, la manipulación genética implica la sobreexpresión de uno o más polinucleótidos que codifican polipéptidos seleccionados de SEQ ID NOs:2, 4, 6, 12, 15, 18 y 21, o un homólogo que comparte al menos el 75 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o al menos el 99 % de identidad de aminoácidos con los mismos. En algunas modalidades, la sobreexpresión está mediada por la introducción de un casete genético sintético en una célula fúngica huésped. En otra modalidad, el casete comprende un promotor heterólogo unido operativamente al polinucleótido, opcionalmente unido operativamente a un terminador transcripcional. En algunas modalidades, la sobreexpresión se compara con un huésped fúngico sin manipulación genética.

En algunas modalidades, la manipulación genética implica una combinación de las manipulaciones genéticas descritas previamente. En una modalidad, la manipulación genética comprende la regulación negativa de uno o más de los polinucleótidos y la sobreexpresión de uno o más polinucleótidos diferentes. En otra modalidad, la manipulación genética comprende la inactivación total de la función enzimática de uno o más polipéptidos codificados por uno o más polinucleótidos y la sobreexpresión de uno o más polinucleótidos diferentes. En una modalidad adicional, la manipulación genética comprende la inactivación total de la función enzimática de uno o más polipéptidos codificados por uno o más polinucleótidos y la regulación negativa de uno o más polinucleótidos diferentes. En una modalidad adicional, la manipulación genética comprende la inactivación total de la función enzimática de uno o más polipéptidos codificados por uno o más polinucleótidos, la sobreexpresión de uno o más polinucleótidos diferentes y la regulación negativa de uno o más polinucleótidos diferentes.

La Tabla 1 muestra ejemplos representativos de manipulación genética en huéspedes fúngicos de la presente invención y el efecto con respecto al ajuste de la producción o composición de carotenoides.

Tabla 1

65

Ejemplos de ajuste de la producción o composición de carotenoides		
SEQ ID NO: (Nombre del Gen)	Manipulación genética	Efecto de ajuste
1 ( <i>CAR1</i> )	regulación negativa	nivel disminuído de carotenoides
3 ( <i>CAR2</i> )	sobreexpresión	nivel aumentado de carotenoides
1 ( <i>CAR1</i> )	sobreexpresión	nivel aumentado de carotenoides totales
19 ( <i>ROC1</i> )	regulación negativa	
10 ( <i>CCD1</i> )	delección	nivel aumentado de torularodina y/o sus derivados
17( <i>ALD1</i> )	sobreexpresión	niveles aumentados de carotenoides totales, toruleno y torularodina

En algunas modalidades, los polinucleótidos descritos en la presente descripción se han incorporado de manera estable en el genoma fúngico. En estas modalidades, los polinucleótidos están unidos operativamente a un promotor que permite la expresión eficiente en especies del género *Rhodosporidium* y el género *Rhodotorula*. Los promotores para cada polinucleótido incorporado pueden ser iguales o diferentes. En algunas modalidades, los promotores son promotores encontrados en especies del género *Rhodosporidium* y el género *Rhodotorula*. En otras modalidades, los promotores son promotores que se encuentran en otras especies fúngicas. Los ejemplos de promotores adecuados incluyen, pero no se limitan a, promotores de los genes siguientes que codifican las proteínas siguientes: gliceraldehído 3-fosfato deshidrogenasa (GPD), proteína transportadora de acil-CoA (ACP), ácido graso desaturasa, factor de elongación de traducción (TEF), piruvato descarboxilasa (PDC), enolasa (2-fosfoglicerato deshidratasa) (ENO), peptidilprolil isomerasa (PPI), acetil-CoA carboxilasa (ACC) o transaldolasa. En otras modalidades, los genes descritos en la presente descripción también incluyen un terminador transcripcional de ARNm que puede encontrarse en cualquier especie eucariota y sus virus de ADN.

En algunas modalidades, un promotor adecuado se describe en la Publicación de Solicitud de Patente Internacional núm. WO 2012/169969, incorporado por referencia en la presente descripción en su totalidad. Esta solicitud publicada describe varias secuencias de polinucleótidos derivadas de la región aguas arriba del gen de la *gliceraldehído fosfato deshidrogenasa (GPD1)*, el gen del *factor de iniciación de la traducción (TEF1)* y el gen de la *estearoil-CoA-delta 9-desaturasa (FAD1)* que funcionan como promotores en hongos. Los promotores descritos en esta solicitud publicada se establecen en las SEQ ID NOs:94-101. En otras modalidades, se describen promotores adicionales en la Publicación de Solicitud de Patente Internacional núm. WO 2014/142747, incorporado por referencia en la presente descripción en su totalidad. Los promotores descritos en esta solicitud publicada se establecen en las SEQ ID NOs:102-118.

Además, los fragmentos operables de las secuencias promotoras descritas en la presente descripción pueden aislarse mediante el uso de ensayos convencionales de tamizaje de promotores y pueden tamizarse para la selección eficiente de las células fúngicas transformadas mediante el uso de las técnicas descritas en la presente descripción. En una modalidad, un fragmento operable, también denominado una porción promotora en la presente descripción, tiene aproximadamente 400 pares de bases hasta aproximadamente 1100 pares de bases de longitud comenzando desde la posición -1 del codón ATG. Como se usa en la presente descripción "hasta" se refiere a la longitud de la porción promotora de los promotores expuestos en las SEQ ID NOs descritas. Por lo tanto, "hasta" se refiere a la longitud máxima de la secuencia promotora si hay menos de 1100 nucleótidos de los promotores de las SEQ ID NOs descritas.

En una modalidad, se proporciona una secuencia promotora que tiene al menos el 60 % de identidad con cualquiera de estas secuencias promotoras. En otra modalidad, se proporciona una secuencia promotora que tiene al menos el 70 % de identidad con cualquiera de estas secuencias promotoras. En una modalidad adicional, se proporciona una secuencia promotora que tiene al menos el 80 % de identidad con cualquiera de estas secuencias promotoras. En una modalidad adicional, se proporciona una secuencia promotora que tiene al menos el 90 % de identidad con cualquiera de estas secuencias promotoras. En otra modalidad, se proporciona una secuencia promotora que tiene al menos el 95 % de identidad con cualquiera de estas secuencias promotoras. En otra modalidad, se proporciona una secuencia promotora que tiene al menos el 98 % de identidad con cualquiera de estas secuencias promotoras.

Los genes que se incorporarán de manera estable en el genoma fúngico están típicamente en forma de un constructo de ADN o polinucleótido que comprende las secuencias promotoras descritas en la presente descripción, una secuencia codificante de polipéptidos unidos operativamente descrita en la presente descripción y una secuencia terminadora transcripcional de ARN unida operativamente. En una modalidad, puede usarse cualquier terminador transcripcional operable en especies de hongos. Los terminadores se ubican típicamente aguas abajo (3') del gen, después del codón de parada (TGA, TAG o TAA). Los terminadores desempeñan un papel importante en el procesamiento y la estabilidad del ARN así como también en la traducción.

Un constructo de ADN o ácido nucleico que comprende un promotor operable en hongos, una secuencia de ADN que codifica proteínas y un terminador operable en hongos también puede denominarse en la presente descripción como un casete de expresión. El casete de expresión puede incluir otras regiones reguladoras transcripcionales como se conocen

bien en la técnica. En otras modalidades, el constructo de ADN o ácido nucleico o casete de expresión comprende además un marcador de selección. Los marcadores de selección se conocen bien por los expertos en la técnica como casetes de expresión que incorporan dichos marcadores y promotores seleccionables para impulsar su expresión, tal como se describe en la Publicación de Solicitud de Patente Internacional núm. WO 2012/169969. Cualquier promotor adecuado unido operativamente a cualquier marcador de selección adecuado puede usarse en la presente invención. En algunas modalidades, pueden usarse una o más moléculas de ADN en las que cada molécula de ADN tiene uno o más constructos de ácido nucleico.

En otra modalidad, la presente invención proporciona un método para producir carotenoides que comprende cultivar una célula fúngica huésped descrita en la presente descripción en condiciones adecuadas para producir carotenoides. Puede usarse cualquier medio con al menos el 5 % de fuente de carbono. En algunas modalidades, la fuente de carbono es glucosa, manosa, glicerol, sacarosa, xilosa o sus combinaciones. En una modalidad, el medio es MinCAR que contiene 30-100 g de glucosa, 1,5 g de extracto de levadura, 0,5 g de  $(\text{NH}_4)_2\text{SO}_4$ , 2,05 g de  $\text{K}_2\text{HPO}_4$ , 1,45 g de  $\text{KH}_2\text{PO}_4$ , 0,6 g de  $\text{MgSO}_4$ , 0,3 g de  $\text{NaCl}$ , 10 mg de  $\text{CaCl}_2$ , 1 mg de  $\text{FeSO}_4$ , 0,5 mg de  $\text{ZnSO}_4$ , 0,5 mg de  $\text{CuSO}_4$ , 0,5 mg de  $\text{H}_3\text{BO}_3$ , 0,5 mg de  $\text{MnSO}_4$ , 0,5 mg de  $\text{NaMoO}_4$  (por litro). El medio se ajusta preferentemente a pH 5-7. En algunas modalidades, el cultivo celular se realiza preferentemente a 25 ° - 35 °C. En otras modalidades, el cultivo se realiza preferentemente en una condición con iluminación.

En la preparación del constructo de ácido nucleico o un casete de expresión, los diversos fragmentos de ADN pueden manipularse, para proporcionar las secuencias de ADN en la orientación adecuada y, según sea apropiado, en el marco de lectura apropiado. Con este fin, pueden emplearse adaptadores o enlazadores para unir los fragmentos de ADN u otras manipulaciones pueden estar implicadas para proporcionar sitios de restricción convenientes, eliminación de ADN superfluo, eliminación de sitios de restricción, o similares. Para este propósito, pueden estar involucradas las mutagénesis *in vitro*, reparación de cebadores, restricción, hibridación, resustituciones, por ejemplo, transiciones y transversiones.

Los ácidos nucleicos de la presente invención también pueden sintetizarse, ya sea totalmente o en parte, especialmente cuando se desea proporcionar secuencias preferidas en vegetales, mediante métodos que se conocen en la técnica. Por lo tanto, todos o una porción de los ácidos nucleicos de la presente invención pueden sintetizarse mediante el uso de codones preferidos por un huésped seleccionado. Los codones preferidos por especie pueden determinarse, por ejemplo, a partir de los codones usados con mayor frecuencia en las proteínas que se expresan en una especie huésped particular. Otras modificaciones de las secuencias de nucleótidos pueden dar como resultado mutantes con actividad ligeramente alterada.

Puede ser útil generar un número de hongos individuales transformados con cualquier constructo recombinante para recuperar hongos libres de cualquier efecto posicional. También puede ser preferible seleccionar los hongos que contienen más de una copia del constructo de polinucleótido introducido de manera que se obtengan niveles altos de expresión de la molécula recombinante.

La práctica de la presente invención emplea, a menos que se indique lo contrario, técnicas convencionales de química, biología molecular, microbiología, ADN recombinante, genética, inmunología, biología celular, cultivo celular y biología transgénica, que están dentro del estado de la técnica. Véase, por ejemplo, Maniatis y otros, 1982, Molecular Cloning (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York); Sambrook y otros, 1989, Molecular Cloning, 2da Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York); Sambrook y Russell, 2001, Molecular Cloning, 3ra Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York); Ausubel y otros, 1992), Current Protocols in Molecular Biology (John Wiley & Sons, incluidas las actualizaciones periódicas); Glover, 1985, ADN Cloning (IRL Press, Oxford); Russell, 1984, Molecular biology of plants: a laboratory course manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.); Anand, Techniques for the Analysis of Complex Genomes, (Academic Press, Nueva York, 1992); Guthrie y Fink, Guide to Yeast Genetics and Molecular Biology (Academic Press, Nueva York, 1991); Harlow y Lane, 1988, Antibodies, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York); Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Methods In Enzymology, Vols. 154 y 155 (Wu y otros eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer y Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volúmenes I-IV (D. M. Weir y C. C. Blackwell, eds., 1986); Riott, Essential Immunology, 6ta Edición, Blackwell Scientific Publications, Oxford, 1988; Fire y otros, ARN Interference Technology: From Basic Science to Drug Development, Cambridge University Press, Cambridge, 2005; Schepers, ARN Interference in Practice, Wiley-VCH, 2005; Engelke, RNA Interference (ARNi): The Nuts & Bolts of siRNA Technology, DNA Press, 2003; Gott, RNA Interference, Editing, and Modification: Methods and Protocols (Methods in Molecular Biology), Human Press, Totowa, NJ, 2004; Sohail, Gene Silencing by RNA Interference: Technology and Application, CRC, 2004.

## EJEMPLOS

La presente invención se describe con referencia a los siguientes Ejemplos, que se ofrecen a modo de ilustración y no pretenden limitar la invención de ninguna manera. Se utilizaron técnicas estándar bien conocidas en la técnica o técnicas específicamente descritas a continuación.

## Ejemplo 1

Cepas, productos químicos, medios y condiciones de cultivo:

Las cepas ATCC 10657, ATCC 10788, ATCC 204091 de *Rhodospiridium toruloides* (anteriormente conocido como *Rhodotorula glutinis*), la cepa ATCC 90781 de *Rhodotorula glutinis* se adquirieron de ATCC (Estados Unidos). Las cepas WP1 de *Rhodotorula glutinis graminis* y FGSC 10293 (IAM13481) de *Sporobolomyces roseus* se obtuvieron de Fungal Genetics Stock Center (Universidad de Missouri, Estados Unidos). Las cepas AGL1 [33] y AGL2 [34] de *A. tumefaciens* se usaron para ATMT.

Las cepas de *Rhodospiridium* se cultivaron a 28 °-30 °C en YPD Broth (1 % de extracto de levadura, 2 % de peptona, 2 % de glucosa) o en agar patata-dextrosa sólido (PDA). *A. tumefaciens* se cultivó a 28 °C en medio 2YT líquido o sólido (1,6 % de triptona, 1 % de extracto de levadura, 0,5 % de NaCl). *Escherichia coli* XL1-Blue se cultivó en caldo Luria-Bertani (LB) o en agar LB y se usó para la manipulación rutinaria del ADN. Para acumular la producción de carotenoides y lípidos, *R. toruloides* se cultivó en medio productor de carotenoides (MinCAR) y medio de acumulación de lípidos (MinRL2), respectivamente. MinCAR y MinRL2 se modificaron del medio de carotenoides [35] y el medio lipídico [36]. El medio MinCAR contiene (por litro) 70 g de glucosa, 1,5 g de extracto de levadura, 0,5 g de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2,05 g de K<sub>2</sub>HPO<sub>4</sub>, 1,45 g de KH<sub>2</sub>PO<sub>4</sub>, 0,6 g de MgSO<sub>4</sub>, 0,3 g de NaCl, 10 mg de CaCl<sub>2</sub>, 1 mg de FeSO<sub>4</sub>, 0,5 mg de ZnSO<sub>4</sub>, 0,5 mg de CuSO<sub>4</sub>, 0,5 mg de H<sub>3</sub>BO<sub>4</sub>, 0,5 mg de MnSO<sub>4</sub>, 0,5 mg de NaMoO<sub>4</sub> (pH 6,1). El medio MinRL2 contiene (por litro) 100 g de glucosa, 1,5 g de extracto de levadura, 0,5 g de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2,05 g de K<sub>2</sub>HPO<sub>4</sub>, 1,45 g de KH<sub>2</sub>PO<sub>4</sub>, 0,6 g de MgSO<sub>4</sub>, 0,3 g de NaCl, 10 mg de CaCl<sub>2</sub>, 1 mg de FeSO<sub>4</sub>, 0,5 mg de ZnSO<sub>4</sub>, 0,5 mg de CuSO<sub>4</sub>, 0,5 mg de H<sub>3</sub>BO<sub>4</sub>, 0,5 mg de MnSO<sub>4</sub>, 0,5 mg de NaMoO<sub>4</sub> (pH 6,1).

## Ejemplo 2

Aislamiento de ADN genómico y ARN total

El ADN genómico y el ARN de *R. toruloides* se extrajeron como se describió anteriormente [37]. El ADN extraído se analizó cualitativamente mediante electroforesis en gel de agarosa y se cuantificó con el Espectrofotómetro NanoDrop® ND-1000 (Nanodrop Technologies, Estados Unidos).

## Ejemplo 3

Transformación mediada por *Agrobacterium tumefaciens*(ATMT):

La transformación de los hongos a través de ATMT se realizó como se describió anteriormente, a menos que se indique lo contrario [37]. Los vectores binarios se transformaron mediante electroporación en *A. tumefaciens* AGL1 o AGL2 mediante el uso de una cubeta al 0,2 mM acoplada a un electroporador BioRad que se fijó a 2,5 kV, 25 µF, 400 Ω. Los transformantes se seleccionaron en placas de agar 2YT suplementadas con estreptomycin (100 µg/ml).

## Ejemplo 4

Constructos de ADN

Los oligonucleótidos usados se enumeran en la Tabla 2. Todas las enzimas de restricción y modificación se adquirieron de New England Biolabs (NEB, Massachusetts, Estados Unidos).

Tabla 2

Secuencias de oligonucleótidos			
Nombre	Secuencia (5'-3') (SEQ ID NO:)	Sitio de restricción	Descripción
PgpdR2-Sf	TTTactagtGGACGGCTTGTCTCTCTCTG (32)	SpeI	Promotor <i>gpdI</i> de <i>U. maydis</i>
Rt012N	TTTccatggTGAGTGATCTGGTGTTGTTC (33)	NcoI	
PgpdR2-Sf	TTTactagtGGACGGCTTGTCTCTCTCTG (34)		Promotor <i>RtGPD1</i>
Rt012N	TTTccatggTGAGTGATCTGGTGITGTTC (35)		
Rt127-2	GGAACATCCGCTCGATCG (36)		Delección de <i>CAR1</i>
Rt128-2	CAGGCCTTCGCCATCGGATT (37)		
Rt129	TCCTCTCCGACTGGGACAA (38)		PCR de colonias para <i>Δcar1</i>
Rt130	CCCAAACAACCCGAGAGGA (39)		

ES 2 787 225 T3

CAR3Lf	AAACACTGATAGTTTTTGGAGGGTGACGCACCTC (40)		Delección de <i>CAR3</i>
CAR3Rr2	TCGAGCTCGGTACCCAGGAGGAGAAGAAGGTGATGG (41)		
Rt138	TCGCTGGATTGGTACGACAAC(42)		PCR de colonias para $\Delta car3$
Rt139	CCACCAGTGACCATCTCTTCG (43)		
CCD1L-Kf	AAAggtaccGACTTGTCCGAGCGAGAGAC (44)	KpnI	Delección de <i>CCD1</i>
CCD1L-Hr	AAAaagcttAGACTCCAGAACCCGACCGTA (45)	HindIII	
CCD1R-Bf	TTTggatccCGAGTCTCAATCCCTCCA (46)	BamHI	
CCD1R-Str	TTTaggcttGGAGGACGGGCGATACAACCTC (47)	StuI	
CCD1f	GTCTTTTCGCGCCCTCTTCCTC (48)		PCR de colonias para $\Delta ccd1$
CCD1r	CGTAGGAGATGACGGGCTTGC (49)		
CDS1L-Stf	TTTaggcttCTCGCTCTCCTGCACACTTCG (50)	StuI	Delección de <i>CDS1</i>
CDS1L-Hr	TTTaaagcttCGCATTTCCAGTCCCATCGC (51)	HindIII	
CDS1R-Bf	TTTggatccACCCTCTACGTCCCCTTACC (52)	BamHI	
CDS1R-Sr	TTTgagctcAACGCCTCGATCCTGACTTGC (53)	SacI	
CDS1f	GTCCTGCTCGCAACCCTCAC (54)		PCR de colonias para $\Delta cds1$
CDS1r2	GAGACGAAGGATGGAGTGGCG (55)		
ALD1Lf	CACCCGTCCTCTCCGCTTC (56)		Delección de <i>ALD1</i>
ALD1Rr	CCTCGCTCTTTGCTGGTTC (57)		
Rt134	CAGCCACATTCGTTCTTCAGG (58)		PCR de colonias para $\Delta ald1$
Rt135	TGGATGATGCGGATATTGAGG (59)		
Rt203Nf	TTTccatggAGGACACTCCCATCGACAGC (60)	NcoI	Expresión de <i>ALD1</i>
Rt204Br	TTTggatccCCTGTCCCGTCAACTTCTGC (61)	BamHI	
CARSL-Stf	TTTaggcttCAGCCAAGTTCAAGCACAACC (62)	StuI	Delección de <i>ROC1</i>
CARSL-Hr	TTTaaagcttCGACCGATCTCGAGGAGACAT (63)	HindIII	
CARSR-Bf	AAAggatccGGAACGATACCCTCCAAGACG (64)	BamHI	
CARSR-Sr	AAAagagctcTGGGAGTTGCGAGGTCATAGA (65)	SacI	
CARSf	TTGTTCTCGGATGTGCGATTGG (66)		PCR de colonias para $\Delta roc1$
CARSr	ATAATCTTGGTGAAGCGGATGTT (67)		
Rt301Nf	TTTccatggCGACTCTAGCCATCAGACC (68)	NcoI	Expresión de <i>ROC1</i>
Rt302Evr	TTTgatatcGAGGCTAGGCGATGTTGCAG (69)	EcoRV	
Rt303Sf	TTTactagtCAAGATCTACGAGGCGAC (70)	SpeI	Complementación de <i>ROC1</i>
Rt304Pmr	TTTgttaaacGAGTGCCCAACGACTTTCTAC (71)	PmeI	
Rt140	CGCTGACCTCCCAATCTTTC (72)		Sonda DIG para <i>CAR1</i>
Rt141	CTTTCCGACCGACTTCTTGCT (73)		
Rt146	GAACCGCAGGTGAAGGTCAAT (74)		Sonda DIG para <i>CAR3</i>
Rt147	TATCGGCAAGGTACGTCTCTCTTC (75)		
Rt148	CAGGTTTCATCGCAACTACATTGA (76)		Sonda DIG para <i>ALD1</i>
Rt149	AACAGAGCGAGTTGAAGAGTAGCC (77)		
qCAR3f	GCGACGACTACGTGAACCTG (78)		qPCR de <i>CAR1</i>
qCAR3r	CGATGGGGAAGGAGAATTTG (79)		
qCAR2f	GCACACTGCACGCCTTACTC (80)		qPCR de <i>CAR2</i>
qCAR2r	ACGAGCTGAAGAGCCTGTCC (81)		

5	qCAR1f	GCAAGATACCCCAGCTCGAC (82)		qPCR de <i>CAR1</i>
	qCAR1r	GGGGACGTTGACGTAGAAGG (83)		
	qCCD1f	GGCTGGATGAAGGAGTGGAC (84)		qPCR de <i>CCD1</i>
	qCCD1r	AGGAGGAGCGTGAGTGGAAG (85)		
10	qCDS1f	ATGGGACTGGAAATGCGAAC (86)		qPCR de <i>CDS1</i>
	qCDS1r	GGGAGACGAAGGATGGAGTG (87)		
	qALD1f	TCGTGCACAACCCGAACTAC (88)		qPCR de <i>ALD1</i>
	qALD1r	ATCTTGCGCTCCTTCTCGTC (89)		
15	qROC1f	ACCAGCTTCAGACCACGTCTC (90)		qPCR de <i>ROC1</i>
	qROC1r	AGAAGTTGGAGGAAGGGATGG (91)		
	qACT1f	CGACAACCTTTGACGACCCTTC (92)		qPCR de <i>ACT1</i>
20	qACT1r	CAGGTTGGGACAAGTTGGGTA (93)		

Diversos promotores, como el promotor de *U. maydis gpd1* (*Pgpd*, 595 pb de longitud) [38, 39] y *RtGPD1* (795 pb) [37], se han descrito previamente y se amplificaron mediante el uso de la plantilla del plásmido pEX1 [42] y el ADN genómico de *R. toruloides* ATCC 10657, mediante el uso de pares de cebadores *Pgpd-Sf/Pgpd-Nr* y *Rt011S/Rt012N*, respectivamente. Los productos resultantes de PCR se digirieron con *SpeI* y *NcoI* y se clonaron mediante ligación de 3 fragmentos con el fragmento de ADN sintético *Bsp de 1030 pbHl/SmaI* del gen *hpt-3* [37] y el fragmento de ADN de pEC3GPD-GUS de 8855 pb *SpeI/SacI* (extremo romo), que crea pEC3GPD-HPT3 y pEC3GPD-R-HPT3, respectivamente.

Para crear constructos de delección para *CAR1*, *CAR3* y *ALD1*, el fragmento de ADN que cubre las regiones codificantes completas del gen (3,0 kb, 2,8 kb y 3,0 kb, respectivamente) se amplificó mediante el uso de ADN genómico de *R. toruloides* ATCC 10657 como plantilla y el par de oligos *Rt127-2/Rt128-2*, *CAR3Lf/CAR3Rr2* y *ALD1Lf/ALD1Rr* como pares de cebadores, respectivamente. El producto de PCR de extremos romos se ligó al vector pEX2 *PmeI/SacI* (extremos romos) para crear el vector intermedio pEX2*CAR1*, pEX2*CAR3* y pEX2*ALD1*, respectivamente. Posteriormente, el casete de resistencia a la higromicina (*RtGPD1::hpt-3::Tnos* con la versión de 795 pb del promotor *RtGPD1* que impulsa la expresión de *hpt-3*) amplificado a partir del plásmido pRH2034 se ligó al pEX2*CAR1* cortado con *SmaI/MfeI*, pEX2*CAR3* cortado con *PvuII/BglII*, y pEX2*ALD1* cortado con *XhoI/BspHI* (extremo romo) para crear el plásmido específico del gen, pKOCAR1, pKOCAR3 y pKOALD1, respectivamente.

Para la delección de *CCD1*, el brazo derecho e izquierdo (0,9 kb cada uno) para recombinación homóloga se amplificó mediante el uso de ADN genómico de *R. toruloides* ATCC 10657 como plantilla y el par de oligos específico *CCD1L-Kf/CCD1L-Hr* y *CCD1R-Bf/CCD1R-Str*, respectivamente. El plásmido de delección génica pKOCOD1 se creó mediante la ligación de cuatro fragmentos que consisten en brazo derecho digerido por *KpnI/HindIII*, brazo izquierdo digerido por *BamHI/StuI*, casete de resistencia a la higromicina *HindIII/BamHI* de pDXP795hptR [32] y el vector pEX2 digerido por *KpnI/SacI*. Las cepas de *E. coli* recombinante con los fragmentos correctos se identificaron mediante PCR de colonias seguido de secuenciación del ADN de todo el casete de recombinación. Se aplicó una estrategia similar para la delección de *CDS1*. Los pares de oligos *CDS1L-Stf/CDS1L-Hr* y *CDS1R-Bf/CDS1R-Sr* se usaron para amplificar los brazos de homología derecha e izquierda de *CDS1* (0,5 kb cada uno), respectivamente. Los fragmentos de ADN se digirieron mediante el uso de *StuI/HindIII* para el brazo izquierdo y *BamHI/SacI* para el brazo derecho, que se ligaron con el casete de resistencia a la higromicina *HindIII/BamHI* y el vector pEX2 digerido con *SmaI/SacI*.

Para la expresión *in vitro* de *ALD1*, las secuencias de ADNc se amplificaron mediante RT-PCR con la plantilla de ARN total y los oligos *Rt203Nf* y *Rt204Br*. Los productos de PCR de doble digestión con *NcoI-BamHI* se clonaron en el vector pRH2034 en los mismos sitios que se crea el plásmido pRH*ALD1*, en el que una fusión *Ald1-eGFP* se impulsó por el promotor *RtGPD1*.

Para la delección de *ROC1*, se usaron pares de los oligos *CARSL-Stf/CARSL-Hr* y *CARSR-Bf/CARSR-Sr* para amplificar los fragmentos flanqueantes homólogos 5' y 3' (0,9 kb cada uno). Se realizó una ligación de cuatro fragmentos con el vector binario pEX2 *SacI-PmeI*, los lados 5' de *StuI-HindIII*, el casete resistente a la higromicina optimizado con el codón *BamHI-HindIII* de pDXP795hptR (*P<sub>GPD1</sub>::hpt-3::Tnos*) y los lados 3' de *BamHI-SacI* para generar el plásmido de delección génica pKOROC1, donde *P<sub>GPD1</sub>* es el promotor de gliceraldehído 3-fosfato de *R. toruloides* ATCC 10657 con el número de registro de GenBank de JN208861, *hpt-3* es el gen optimizado con el codón que codifica la higromicina fosfotransferasa (JQ806387), y *Tnos* es el terminador de agrobacterium (Liu, Koh y otros 2013).

Para estudios de complementación de *Δcar1*, el locus del genoma de *CAR1* que oscila de 389556 a 393649 nt del andamio WGS # 18 se amplificó mediante el uso de la plantilla de *R. toruloides* y los oligos *Rt127-2* y *CDSL1*. Los productos de PCR de 4,1 kb fueron de extremos romos y se ligaron con el vector pRH2034 linealizado con *SpeI* (extremo romo)-*PmeI* para crear el plásmido de complementación pRH*CAR1*. Para estudios de complementación de *Δroc1*, el locus del genoma

de *ROC1* que oscila entre 622910 y 627480 nt del andamio WGS # 9 se amplificó mediante el uso de la plantilla de *R. toruloides* y los oligos Rt303Sf y Rt304Pmr. Los productos de PCR de 4,6 kb se digirieron doblemente con Spel y Pmel, y se ligaron con el vector pRH2034 linealizado con Spel-Pmel para crear el plásmido de complementación pRHROC1.

## 5 Ejemplo 5

### PCR de colonias y análisis de Southern blot

10 El análisis por PCR de colonias fúngicas se usó para la detección de candidatos de mutantes de delección génica. Brevemente, se cultivaron colonias individuales de transformantes en 150 µl de YPD Broth suplementado con cefotaxima (300 mg/ml) e higromicina (150 mg/ml) durante varias horas. Se usó un microlitro de cultivo celular para el análisis de PCR de colonias con un par de oligos apropiado dentro de la región de selección génica (Tabla 2). El PCR se realizó mediante el uso de i-Taq polimerasa (i-DNA Biotech, Singapur) y el programa siguiente: iniciación a 95 °C durante 5 minutos, seguido de 35 ciclos a 94 °C durante 30 segundos, a 58 °C durante 30 segundos y a 72 °C durante 45 segundos, y una mayor extensión a 72 °C durante 5 minutos. Se usó una electroforesis para la identificación de mutantes candidatos que carecen de fragmentos de ADN que podrían amplificarse mediante el uso de la plantilla de ADN genómico de WT.

15 Para verificar los verdaderos mutantes de delección génica sin ninguna integración ectópica, los ADN genómicos se digirieron con HindIII, PstI, PvuI, PvuI, HincII y PvuI para los supuestos mutantes de desactivación génica  $\Delta car1$ ,  $\Delta car3$ ,  $\Delta ccd1$ ,  $\Delta cds1$ ,  $\Delta ald1$  y  $\Delta roc1$ , respectivamente. Los fragmentos de ADN que contienen los brazos derechos de *CAR1*, *CAR3*, *ALD1*, *ROC1* y los brazos izquierdos de *CCD1*, *CDS1* se marcaron con digoxigenina mediante el uso de DIG-High prime DNA labeling and detection starter Kit II (Roche Diagnostics, Estados Unidos). La hibridación Southern blot se realizó de acuerdo con las instrucciones del fabricante.

## 25 Ejemplo 6

### PCR de transcripción inversa cuantitativa

30 El ARN total se extrajo como se describió anteriormente [37]. Antes del PCR de transcripción inversa cuantitativa (qRT-PCR), el ARN total se trató con DNasa I (Roche Diagnostics) para eliminar las trazas de ADN y se recuperó mediante precipitación con etanol. El ADNc se sintetizó mediante el uso de iScript™ Reverse Transcription Supermix (Bio-Rad, Estados Unidos) y el q-PCR se realizó en el Sistema de detección de secuencia ABI PRISM 7900HT mediante el uso de ABI SYBR® Select Master Mix (Life Technologies, Estados Unidos). Las condiciones del qRT-PCR fueron las siguientes: una etapa inicial de desnaturalización a 50 °C durante 2 min y a 95 °C durante 10 min, seguida de 40 ciclos de desnaturalización a 95 °C durante 15 segundos, hibridación a 60 °C durante 1 minuto. Las muestras se analizaron por triplicado y los datos se obtuvieron mediante el uso del programa SDS 2.4 (Life Technologies, Estados Unidos). Los niveles relativos de expresión génica se calcularon contra el gen de referencia *ACT1* (SEQ ID NOs: 22, 23, 24 para genómica, ADNc y proteína, respectivamente) mediante el uso del método 2- $\Delta\Delta C_t$  con el programa RQ Manager v1.2.1 (Applied Biosystems, Estados Unidos).

## 40 Ejemplo 7

### Detección de genes involucrados en la biosíntesis de carotenoides

45 *La cepa haploide de R. toruloides* ATCC 10657 se mutagenizó mediante inserciones aleatorias de T-DNA mediante la transformación del vector binario de T-DNA pRH201 mediada por *Agrobacterium tumefaciens* (Figura 1). Los transformantes se seleccionaron en medio de agar YPD suplementado con 300 µg/ml de cefotaxima y 150 µg/ml de higromicina. Después de incubarse a 28 °C durante 5 días, los transformantes que muestran colores albinos o pálidos se transfirieron a medio YPD líquido (300 µg/ml de cefotaxima, 150 µg/ml de higromicina) para la propagación. Después de 50 rayar en las placas de PDA suplementadas con los antibióticos anteriores, las colonias individuales que mostraron un fenotipo de color estable se nombraron *Rhodospordium* *Carotenoide* *Mutant* (RCM).

## Ejemplo 8

### 55 Identificación de posiciones de etiquetado de T-DNA

Las posiciones de la etiqueta T-DNA en el genoma se identificaron mediante PCR termal de entrelazado asimétrico de alta eficacia (hiTAIL-PCR) [43, 44]. Se usaron cebadores específicos (HRSP1, HRSP2 y HRSP3) y el cebador arbitrario LAD1-4 para las secuencias de flanqueo del borde izquierdo (LB) de la T-DNA mientras que los cebadores específicos (HRRSP1, HRRSP2 y HRRSP3) y el cebador arbitrario LAD1-4 para las secuencias flanqueantes del borde derecho (RB). Las reacciones de PCR se llevaron a cabo con i-Taq DNA polymerase (i-DNA, Singapur) en un controlador térmico programable PTC-200™ (Bio-Rad, Estados Unidos). Los productos del PCR se purificaron mediante el uso del kit de extracción de gel (Qiagen, Estados Unidos) y se secuenciaron directamente mediante el uso del kit terminador BigDye (Applied Biosystems, Estados Unidos) con oligos HRRSP3 (para secuencias flanqueantes RB) o HRSP3 (secuencias 60 flanqueantes LB). Para las muestras que dieron resultados de secuenciación pobres, los productos del PCR se clonaron en el vector pGTM-T easy (Promega, Estados Unidos) y se secuenciaron mediante el uso de los oligos M13FP y M13RP.

## Ejemplo 9

## Extracción de carotenoides

5

Las células se cultivaron en 50 ml de medio MinCAR en matraces de agitación a 30 °C y se sedimentaron por centrifugación. Después de lavar dos veces con agua, se determinó la masa celular húmeda mediante el pesaje y mezclado con la misma masa de perlas de vidrio lavadas con ácido (0,4-0,6 mm de diámetro, Sigma-Aldrich) y 5 ml de DMSO. Las células se lisaron mediante agitación vigorosa en vórtex durante 10 minutos, 1 hora de incubación a 65 °C seguida de congelación a -20 °C. Después de descongelar, la suspensión se centrifugó a 10 000 g, y el sobrenadante que contenía carotenoides solubles en DMSO se transfirió a un tubo nuevo mientras que el residuo celular insoluble se volvió a extraer con 30 ml de éter de petróleo ligero - acetato de etilo (36:19) durante 10 minutos a temperatura ambiente. Los contenidos de las dos extracciones se combinaron y se extrajeron con 2 ml de NaCl saturado. La fase de disolvente se recogió después de la centrifugación y se secó bajo un flujo de gas nitrógeno. Las muestras se redisolviéron en hexano y se almacenaron a -20 °C antes de un análisis posterior.

15

## Ejemplo 10

## Métodos de cuantificación

20

La biomasa celular (peso seco de la célula) se determinó mediante la liofilización del sedimento celular recogido por centrifugación hasta que se alcanzó un peso constante.

25

La concentración de glucosa en los medios se cuantificó mediante HPLC. El medio se separó de las células mediante centrifugación y se filtró a través de una membrana de nylon de 0,2 µm. Se inyectaron 10 µl de la muestra y se pasaron a través de una columna Aminex 87H de 300 x 7,0 mm (Bio-Rad) a una tasa de flujo constante de 0,7 ml/min mediante el uso de ácido sulfúrico 5 mM como fase móvil. La columna se mantuvo a 50 °C y se detectó la glucosa con un detector de índice de refracción (RID, Shimadzu, Japón). La concentración de glucosa en el cultivo celular se determinó mediante el uso de una curva de calibración construida con la solución acuosa de glucosa estándar.

30

Los picos principales se determinaron mediante los espectros de absorción en hexano y espectrometría de masas. La técnica de ionización química a presión atmosférica (APCI) es como se describió anteriormente [46, 47] con algunas modificaciones. Brevemente, las muestras (x µl) se analizaron en un sistema Shimadzu UPLC-MS (APCI) (Shimadzu, Japón) equipado con una columna de YMC-carotenoide (fase inversa C30, Φ3 µm, 150 mm x 3,0 mm I.D., YMC, Japón) a una tasa de flujo de 0,3 ml/min en un gradiente lineal en 3 minutos, desde la mezcla A al 100 % (MeOH/terc-butilmetiléter/agua, 30:1:10, v/v/v) hasta la mezcla B al 50 % (MeOH/terc-butilmetiléter, 1:1, v/v) seguido de la mezcla B al 100 % durante 0,5 min y después la columna se mantuvo en las condiciones durante 15 min a una tasa de flujo de 0,6 ml/min. Se usó APCI en modo positivo para la identificación de componentes de carotenoides con 15 L/s de gas nitrógeno como envoltura y gas auxiliar. La temperatura del vaporizador y el capilar se fijó a 350 °C y 150 °C, respectivamente, y los voltajes de las lentes de los capilares y los tubos se fijaron a 50 V y 125 V, respectivamente.

40

## Ejemplo 11

## Caracterización de mutantes biosintéticos de carotenoides

45

La detección de aproximadamente 20 000 transformantes de T-DNA mediante la identificación visual de los colores de las colonias conduce a la identificación de seis mutantes de carotenoides, que se denominan RAM1-5 (mutantes de carotenoides de *R. toruloides*) (Figs. 2a-2E, Tabla 3). La búsqueda mediante hiTAIL-PCR y BLAST de las etiquetas de secuencia reveló que los T-DNA se insertaron en genes que codifican un supuesto transportador de riboflavina, aldehído hidrogenasa, transportador de hexosa, factor asociado a la proteína de unión a TATA, fitoeno desaturasa y aldehído deshidrogenasa grasa, respectivamente (Tabla 3). Se conoce bien el papel de la fitoeno desaturasa, o CarI en hongos (EC 1.3.99.30) en la producción de licopeno [48].

50

Tabla 3

55

Caracterización de <i>mutantes albinos</i> de <i>R. toruloides</i> (RAM)						
Número de secuencia <sup>a</sup>	Sitio génico <sup>b,c</sup>	Mejor acierto <sup>d</sup>	Anotación <sup>e</sup>	Organismo <sup>f</sup>	Identidad <sup>g</sup>	
<b>Secuencias RB</b>						
RAM1	Secuencia génica	XP_003032296	Transportador de riboflavina MCH5	<i>Schizophyllum commune</i>	52 %	
RAM2	Aguas arriba-0,5 kb	YP_001220603	resolvasa	<i>Aeromonas bestiarum</i>	95 %	

60

65

5	RAM3	Secuencia génica	XP_571856	proteína relacionada con el transporte de hexosa	<i>Cryptococcus neoformans</i>	36 %
10	RAM4	Secuencia génica	XP_758766	Factor Taf2 asociado a la proteína de unión a TATA (TBP) (MTCC 457 contig458_1:18376-18377 +)	<i>Ustilago maydis</i>	35 %
15	RAM5	Secuencia génica	KF601426.1	fitoeno sintasa	<i>Rhodosporidium diobovatum</i>	98 %
20	<p><sup>a</sup> Secuencia flanqueante obtenida del número correspondiente de transformante de T-DNA  <sup>b</sup> Los genes etiquetados con T-DNA se determinaron de acuerdo con los resultados de BLASTx  <sup>c</sup> Aguas arriba 1,0 kb, Aguas arriba 0,5 kb y aguas abajo 0,3 kb denota inserciones de T-DNA dentro de aguas arriba 501~1000 pb, 500 pb y aguas abajo 300 pb del gen etiquetado correspondiente, respectivamente  <sup>d</sup> El mejor acierto denota el resultado de BLASTx con el puntaje E más alto  <sup>e</sup> Las anotaciones se determinaron de acuerdo con los resultados de BLASTx  <sup>f</sup> El microorganismo denota el receptor del mejor acierto  <sup>g</sup> Los valores de identidad fueron de los resultados de BLASTx  <sup>h</sup> No disponible debido al resultado de una mala secuenciación</p>					

Ejemplo 12

25 Análisis del grupo de genes de biosíntesis de carotenoides en *R. toruloides* mediante Genética inversa

El hiTAIL-PCR reveló que el fenotipo albino RCM5 es el resultado de un T-DNA insertado entre 391802 nt y 391803 nt en el andamio genómico # 18 (AEVR02000018) de *R. glutinis* ATCC 204091, ubicado en el 3<sup>er</sup> exón del supuesto gen de la fitoeno desaturasa (*CAR1*, locus del genoma RTG\_00274) (Figura 2B), una enzima implicada en la biosíntesis de carotenoides. Para confirmar la función de este gen que se interrumpió por la inserción de T-DNA, el supuesto *CAR1* CDS se eliminó por desactivación génica específica mediante el uso de ATMT de pKOCARI, en el que la secuencia de nucleótidos que oscila entre +948 y +2097 del CDS se reemplazó por el casete resistente a la higromicina (*P<sub>GPD1</sub>::hpt-3::Tnos*, Figura 2B). El mutante nulo correcto ( $\Delta car1$ ) se verificó mediante análisis de Southern blot (Figura 2C). Como se esperaba, la colonia  $\Delta car1$  mostró un color cremoso en lugar del color naranja observado en WT. El color cremoso de  $\Delta car1$  podría restaurarse a naranja mediante la integración ectópica del alelo del gen *CAR1* en el genoma (cepa  $\Delta car1C$  en la Figura 2D). Los análisis de los perfiles de carotenoides confirmaron la pérdida de picos de  $\beta$ -caroteno,  $\gamma$ -caroteno, toruleno y torularodina en  $\Delta car1$  y todos los picos perdidos se restauraron en la cepa complementada,  $\Delta car1C$ , con la integración del T-DNA del vector binario pRHCAR1, donde el alelo completo de *CAR1* que oscila desde -1166 aguas arriba hasta +517 aguas abajo de los codones transcripcionales de inicio y parada, respectivamente (Figura 2E). Los resultados respaldan fuertemente que *CAR1* codifica una de las enzimas clave involucradas en la vía biosintética de carotenoides. Para identificar más genes en la vía biosintética de carotenoides, se realizaron búsquedas de tBLASTn mediante el uso de la GGPP sintasa de *U. maydis* (*CAR3*) (XP\_760606, GenBank) y fitoeno sintasa/caroteno ciclasa (*CAR2*) (XP\_762434) y las secuencias ortólogas correspondientes en *R. toruloides* se identificaron exitosamente. *CAR3* CDS se encontró ubicado en nt 849806-851310 en el andamio # 13 (locus del genoma RTG\_00457, AVER02000013) mientras *CAR2* en nt 396838-399094 en el andamio # 18.

Nosotros hemos informado que la desactivación génica *CAR2* conduce al fenotipo albino [32]. Sin embargo, los detalles sobre su estructura génica continúan siendo desconocidos. Mediante el uso de la amplificación rápida de los extremos de ADNc (RACE) y las técnicas de PCR de transcripción inversa (RT-PCR), se obtuvieron las secuencias de ADNc para *CAR1* (SEQ ID NO:1), *CAR2* (SEQ ID NO:3), y *CAR3* (SEQ ID NO:5). Los ADNc de *CAR1*, *CAR2* y *CAR3* abarcan 2430, 2334, y 1546 nt genómico de longitud, que contienen 10, 8 y 6 exones y codifican proteínas de 554 (Car1, SEQ ID NO:2), 608 (Car2, SEQ ID NO:4) y 359 (Car3, SEQ ID NO:6) aa, con 19, 77 y 41 nt 5'UTR en los ADNc, respectivamente. Las secuencias genómicas correspondientes se enumeran en las SEQ ID NOs: 7, 8 y 9, respectivamente. El empalme de los 3 ARNm sigue estrictamente la regla canónica GU-AG. Los resultados también revelaron que el T-DNA de RAM5 se integró entre +681 y +682 desde el codón de inicio de *CAR1*, lo que da como resultado la terminación prematura de la traducción del ARNm de *CAR1* después del 158<sup>vo</sup> aa (Figura 2B).

*CAR1* y *CAR2* se ubican en el mismo andamio # 18, separados por una secuencia de ADN de 4354 pb. Esta organización es análoga a varios otros hongos carotenogénicos, tales como *Blakeslea trispora*, *Fusarium fujikuroi*, *Phycomyces blakesleeanus* y *Sporobolomyces roseus* [49-52] (Figura 3A). Una búsqueda homóloga (BLASTx, NCBI) de la secuencia de ADN entre el locus genómico *CAR1* y *CAR2* descubrió dos genes supuestos que codifican una dioxigenasa de escisión de carotenoides (*Ccd1*) y una carotenoide desaturasa (*Cds1*) (Figura 3A). Las RACE 5' y 3' y el RT-PCR revelaron que los ADNc de *CCD1* (SEQ ID NO:11) y *CDS1* (SEQ ID NO:14) abarcan 2079 y 793 nt genómico en longitud, que contienen 4 y 3 exones con 41 y 19 nt 5'UTR que codifican proteínas con 636 (*Ccd1*, SEQ ID NO:12) y 224 aa (*Cds1*, SEQ ID NO:14), respectivamente. Nuevamente, el empalme sigue estrictamente la regla canónica GU-AG. Las secuencias genómicas correspondientes se enumeran en SEQ ID NOs: 10 y 13, respectivamente.

La organización divergente de *CAR1* y *CAR2* son análogas a *Mucor circinelloides* (andamio # 1), *Phycomyces blakesleeanus* (andamio # 5), *Blakeslea trispora* y *F. fujikuroi* (cromosoma # 11) excepto que los genes *CCD1* y *CDS1* se encuentran solo en *R. toruloides*. El genoma de *S. roseus* parece que sufrió una recombinación entre *CAR1* y *CAR2*, que dio como resultado la pérdida de *CDS1* y la translocación de *CAR1* (Figura 5A). *F. fujikuroi* *CarX* se ubica fuera del *CarRA* (*CAR2* ortólogo) y *CarB* (*CAR1* ortólogo) y se considera como el ortólogo de *CCD1* porque su producto proteico exhibe una secuencia de aa altamente homóloga a *Ccd1* (43 % de identidad). La sintenia genética compartida entre estos hongos carotenogénicos sugiere un origen evolutivo común de los grupos de genes carotenoides.

Para confirmar sus funciones en la biosíntesis de carotenoides, se crearon mutantes nulos. Similar a  $\Delta car1$  y  $\Delta car2$ , las colonias de  $\Delta car3$  exhibieron un fenotipo de color cremoso mientras que  $\Delta ccd1$  y  $\Delta cds1$  mostraron también colores diferentes significativamente a WT (Figura 3C). El análisis de sus carotenoides mediante HPLC reveló que la producción de carotenoides se abolió totalmente en  $\Delta car3$ ,  $\Delta car2$  y  $\Delta car1$  (datos no mostrados). Los niveles de producción total de carotenoides aumentaron ligeramente en  $\Delta ccd1$  en un 18 %, pero disminuyeron drásticamente en  $\Delta cds1$  en un 69 % (Figura 3D). Excepto la ligera disminución de  $\beta$ -caroteno (6 %), con respecto a los componentes de los carotenoides acumulados, la delección de *CCD1* podría dar como resultado un aumento del 86 %, el 14 % y el 65 % en torularodina, toruleno y  $\gamma$ -caroteno, respectivamente (Figura 3D). En el mutante *cds1*, la cuantificación de todos los componentes de carotenoides disminuyó, especialmente más de la mitad disminuyó en torularodina y toruleno (Figura 3D).

#### Ejemplo 13

##### Efecto de la delección de *ALD1* sobre la producción de carotenoides

El mutante RCM6 se identificó inicialmente en un examen de detección de genes que afectan la biosíntesis de ácidos grasos y se encontró que el gen representativo *ALD1*, gen que codifica la aldehído deshidrogenasa grasa, está involucrado en la degradación de un ácido graso poliinsaturado (ácido alfa-linolénico, C18:3n=9) (Solicitud de patente provisional de Estados Unidos núm. 62/047,300 presentada el 8 de septiembre de 2014, incorporada como referencia en la presente descripción). En cuanto a la diferencia significativa en el color de la célula frente a WT, también se estudió el papel de *ALD1* en la biosíntesis de carotenoides. La delección génica mutante de  $\Delta ald1$  se generó por recombinación homóloga a través de ATMT mediante el uso del vector binario pKOALD1. Para obtener la sobreexpresión génica mutante, la proteína de fusión Ald1-RtGFP se fusionó con el promotor *RtGPD1* y se integró ectópicamente en  $\Delta ald1$  a través de ATMT (mutante OEALD1). Después de una fermentación de 5 días en el medio productor de carotenoides MinCAR, el rendimiento total de carotenoides en  $\Delta ald1$  fue similar a WT, sin embargo, el nivel de beta-caroteno aumentó en aproximadamente el 36 %. Por el contrario, los niveles de toruleno y de torularodina aumentaron significativamente en la cepa de sobreexpresión de *ALD1* (Figura 4). Una secuencia genómica, secuencia de ADNc y secuencia de proteína para *ALD1* se establecen en las SEQ ID NOs: 16, 17 y 18, respectivamente.

#### Ejemplo 14

##### Perfiles de ARNm de *CAR1*, *CAR2*, *CAR3*, *ALD1*, *CCD1* y *CDS1*

Se inocularon colonias individuales de *R. toruloides* ATCC 10657 y mutantes de carotenoides en 50 ml de YPD Broth en matraces cónicos de 250 ml y se cultivaron a 28 °C y 250 rpm hasta la saturación. Los cultivos celulares se separaron en mitades y se continuó el cultivo para cada uno en una plataforma de agitación durante 5 horas más, donde la iluminación se transmitió mediante el uso de una luz fluorescente (Philips TLD 36 W/865, luz blanca de 4 W/m<sup>2</sup>, a 75 cm de distancia) o se evitó tras cubrirlo con papel de aluminio. Se tomaron muestras de las células en varios puntos de tiempo y se extrajeron los ARN totales. El qRT-PCR se realizó por triplicado mediante el uso de pares de oligos qCAR3f/qCAR3r, qCAR2f/qCAR2r, qCAR1f/qCAR1r, qCCD1f/qCCD1r, qCDS1f/qCDS1r y qALD1f/qALD1r para *CAR3*, *CAR2*, *CAR1*, *CCD1*, *CDS1*, *ALD1*, respectivamente (Tabla 2). Los niveles relativos de ARNm se calcularon frente al gen de referencia de actina (par de oligos qACT1f/qACT1r, Tabla 2) mediante el uso del método 2- $\Delta\Delta$ Ct. Como se muestra en la Figura 5, los niveles de ARNm de *CAR1* y *CAR2* fueron mucho más bajos que los de *CAR3* y probablemente fueron el cuello de botella para la redirección del flujo de carbono hacia la producción de carotenoides. El ARNm de *CAR1*, *CAR2*, *CAR3* y *ALD1* se indujo mediante la luz.

#### Ejemplo 15

##### Caracterización de *ROC1*

En *Fusarium fujikuroi*, la interrupción de *carS* mostró una mejor biosíntesis de carotenoides independientemente de la iluminación con la luz (Avalos y Cerdà-Olmedo 1987). Sin embargo, los genes involucrados en las etapas anteriores de la vía biosintética de terpenoides como FPP sintasa y HMGR, no se vieron afectadas por la delección de *carS* (Rodríguez-Ortiz, Limón y otros 2009).

Mediante el uso de la secuencia de proteínas *carS* de *Fusarium fujikuroi* (NCBI GenBank número de registro CCP50075.1) como consulta para buscar las secuencias de escopeta de genoma completo de *R. glutinis* ATCC 204091 a través del programa en línea tBLASTn (NCBI, Estados Unidos), se encontró un ortólogo supuesto en el andamio de secuencia # 9

(valor E y cobertura de consulta de 4E-37 y 53 %, respectivamente). Para ser consistente con la nomenclatura génica del gen supuesto de *R. toruloides* se denominó *ROC1* (Regulador de la biosíntesis de carotenoides).

La RACE 5' y 3' produjo un par de fragmentos de ADNc de aproximadamente 0,9 y 0,7 kb para *ROC1* (datos no mostrados). Mediante el uso del par de oligos Rt301Nf y Rt302Evr (Tabla 2), el ADNc de longitud completa de *ROC1* podría amplificarse exitosamente mediante RT-PCR (datos no mostrados). La comparación entre el ADNc y las secuencias genómicas reveló 5 exones para *ROC1* (Figura 6A), donde las uniones de empalme se atienen estrictamente a la regla cónica GU-AG.

*ROC1* abarca una región del genoma de 3136 nt (SEQ ID NO:19) que codifica un ARNm de 2805 nt con un 5'UTR de 84 nt (SEQ ID NO:20). *ROC1* codifica una proteína de 934-aa (SEQ ID NO:21) que contiene un dominio de dedo RING (tipo C3HC4, dominio conservado NCBI cd00162) y un dominio La (LON) de proteasa dependiente de ATP (pfam02190). La secuencia muestra el 96 % y el 97 % de identidad de aa con un homólogo en *R. toruloides* cepa NP11 (EMS19961.1) y CECT1137 (CDR44527.1, respectivamente).

Rocl comparte menos del 20 % de identidad con los ortólogos de varios hongos carotenogénicos, excepto los de dos zigomicetos, *B. trispora* y *M. cricinelloides* (Figura 6B). Además, las secuencias de proteínas dentro del dominio de dedo RING central también muestran una homología muy baja, algunas incluso carecen del motivo C3HC4 (Figura 6C).

## Ejemplo 16

### Efectos de la delección de *ROC1*

Para comprender genéticamente el papel de *ROC1* en la biosíntesis de carotenoides, el gen *ROC1* se eliminó mediante recombinación homóloga (Figura 7A). La transformación con el vector de delección génica (pKOROC1) mostró dos tipos de transformantes con diferencias de color obvias, naranja claro y fuerte (Figura 7A), y se encontró que los transformantes naranja fuerte son los verdaderos mutantes de desactivación génica mediante el análisis de Southern blot (Figura 7B). Los mutantes nulos *roc1* mostraron una morfología y crecimiento celular similar a WT (Figuras 7C y 7D). Sin embargo, pudo observarse una mejora significativa de la biosíntesis de caroteno en mutantes *roc1* y esto podría restaurarse completamente mediante la introducción de un alelo génico WT *ROC1* en un mutante *roc1* (Figura 7D). Sorprendentemente, la producción de biomasa celular disminuyó significativamente en el mutante de complementación (Figura 7D). El análisis cuantitativo reveló que el cultivo de 5 días en medio MinCAR produjo aproximadamente 1,5 veces más carotenoides en el mutante *roc1* en comparación con los de WT (Figura 7E).

El uso de los términos " un", " una" y "el/la" y referentes similares en el contexto de la descripción de la invención (especialmente en el contexto de las siguientes reivindicaciones) se debe interpretar que abarca tanto el singular como el plural, a menos que se indique de cualquier otra manera en la presente descripción o se contradiga evidentemente por el contexto. Los términos "que comprende", "que tiene", "que incluye" y "que contiene" deben interpretarse como términos abiertos (es decir, que significan "que incluye pero sin limitarse a") a menos que se indique lo contrario. La mención de intervalos de valores en la presente descripción tiene meramente la intención de servir como un método abreviado para referirse individualmente a cada valor por separado que cae dentro del intervalo, a menos que se indique de cualquier otra forma en la presente, y cada valor por separado se incorpora en la descripción como si se mencionara individualmente en la presente. Todos los métodos que se describen en la presente descripción pueden llevarse a cabo en cualquier orden adecuado a menos que se indique de otra manera en la presente descripción o que el contexto lo contradiga claramente de otra manera. El uso de cualquiera y todos los ejemplos, o lenguaje ilustrativo (por ejemplo, "tal como") proporcionados en la presente descripción simplemente tiene el propósito de ilustrar mejor la invención y no representa una limitación en el alcance de la invención al menos que se reivindique de cualquier otra forma. Ningún lenguaje en la descripción debe interpretarse como que indica cualquier elemento no reivindicado como esencial para la práctica de la invención.

## Bibliografía

1. Hirschberg J: Carotenoid biosynthesis in flowering plants. Current opinion in plant biology 2001, 4:210-218.
2. Armstrong GA, Hearst JE: Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis. FASEB J 1996,10:228-237.
3. Eisenreich W, Bacher A, Arigoni D, Rohdich F: Biosynthesis of isoprenoids via the non-mevalonate pathway. Cellular and Molecular Life Sciences CMLS 2004, 61:1401-1426.
4. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nature biotechnology 2003, 21:796-802.
5. Armstrong GA, Hearst JE: Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis. The FASEB Journal 1996, 10:228-237.

6. Kajiwara S, Kakizono T, Saito T, Kondo K, Ohtani T, Nishio N, Nagai S, Misawa N: Isolation and functional identification of a novel cDNA for astaxanthin biosynthesis from *Haematococcus pluvialis*, and astaxanthin synthesis in *Escherichia coli*. *Plant molecular biology* 1995, 29:343-352.
- 5 7. Fraser PD, Miura Y, Misawa N: In vitro characterization of astaxanthin biosynthetic enzymes. *Journal of Biological Chemistry* 1997, 272:6128-6135.
8. Cunningham FX, Gantt E: Genes and Enzymes of Carotenoid Biosynthesis in Plants. *Annu Rev Plant Physiol Plant Mol Biol* 1998, 49:557-583.
- 10 9. Mpari SA, Nielsen KF, Larsen TO, Frisvad JC, Meyer AS, Thrane U: Exploring fungal biodiversity for the production of water-soluble pigments as potential natural food colorants. *Current Opinion in Biotechnology* 2005, 16:231-238.
- 15 10. Takaichi S: Carotenoids and carotenogenesis in anoxygenic photosynthetic bacteria. In *The photochemistry of carotenoids*. Springer; 1999: 39-69
11. Cobbs C, Heath J, Stireman III JO, Abbot P: Carotenoids in unexpected places: Gall midges, lateral gene transfer, and carotenoid biosynthesis in animals. *Molecular phylogenetics and evolution* 2013, 68:221-228.
- 20 12. Sporn M, Dunlop N, Newton D, Smith J: Prevention of chemical carcinogenesis by vitamin A and its synthetic analogs (retinoids). In *Federation proceedings*. 1976: 1332-1338.
13. Seddon JM, Ajani UA, Sperduto RD, Hiller R, Blair N, Burton TC, Farber MD, Gragoudas ES, Haller J, Miller DT: Dietary carotenoids, vitamins A, C, and E, and advanced age-related macular degeneration. *Jama* 1994, 272:1413-1420.
- 25 14. Wolbach SB, Howe PR: Tissue changes following deprivation of fat-soluble A vitamin. *The Journal of experimental medicine* 1925, 42:753-777.
15. Miki W: Biological functions and activities of animal carotenoids. *Pure and Applied Chemistry* 1991, 63:141-146.
- 30 16. Estrada AF, Brefort T, Mengel C, Díaz-Sánchez V, Alder A, Al-Babili S, Avalos J: *Ustilago maydis* accumulates isofort T, Mengel C, Díaz-Sánchez V, Alder A, Al-Babili S, Avalos J: *J: io Fungal Genetics and Biology* 2009, 46:803-813.
17. Giovannucci E, Ascherio A, Rimm EB, Stampfer MJ, Colditz GA, Willett WC: Intake of carotenoids and retino in relation to risk of prostate cancer. *Journal of the national cancer institute* 1995, 87:1767-1776.
- 35 18. Spencer KG: Pigmentation supplements for animal feed compositions. Google Patents; 1989.
19. Herring P: The carotenoid pigments of *Daphnia magna* and pure carotenoids. *Comparative biochemistry and physiology* 1968, 24:187-203.
- 40 20. Mortensen A: Carotenoids and other pigments as natural colorants. *Pure and Applied Chemistry* 2006, 78:1477-1491.
21. Peter F, Andrea L-R, Peter W, Naoharu W: Carotenoid Cleavage Enzymes in Animals and Plants. In *Carotenoid-Derived Aroma Compounds*. Volume 802: American Chemical Society; 2001: 76-88: ACS Symposium Series].
- 45 22. Duester G: Involvement of alcohol dehydrogenase, short-chain dehydrogenase/reductase, aldehyde dehydrogenase, and cytochrome P450 in the control of retinoid signaling by activation of retinoic acid synthesis. *Biochemistry* 1996, 35:12221-12227.
- 50 23. Li Y, Zhao ZK, Bai F: High-density cultivation of oleaginous yeast *Rhodospiridium toruloides* Y4 in fed-batch culture. *Enzyme and Microbial Technology* 2007, 41:312-317.
24. Zhao X, Hu C, Wu S, Shen H, Zhao ZK: Lipid production by *Rhodospiridium toruloides* Y4 using different substrate feeding strategies. *J Ind Microbiol Biotechnol* 2011, 38:627-632.
- 55 25. Pan JG, Kwak MY, Rhee JS: High density cell culture of *Rhodotorula glutinis* using oxygen-enriched air. *Biotechnology letters* 1986, 8:715-718.
26. Frengova GI, Beshkova DM: Carotenoids from *Rhodotorula* and *Phaffia*: yeasts of biotechnological importance. *J Ind Microbiol Biotechnol* 2009, 36:163-180.
- 60 27. Cerdá-Olmedo E: Production of carotenoids with fungi. In *Biotechnology of vitamins, pigments and growth factors*. Springer; 1989: 27-42
- 65

28. Buzzini P, Martini A: Production of carotenoids by strains of *Rhodotorula glutinis* cultured in raw materials of agro-industrial origin. *Bioresource technology* 2000, 71:41-44.
29. Frengova GI, Beshkova DM: Carotenoids from *Rhodotorula* and *Phaffia*: yeasts of biotechnological importance. *Journal of industrial microbiology & biotechnology* 2009, 36:163-180.
30. Bhosale P, Gadre R:  $\beta$ -Bhosale P production in sugarcane molasses by a *Rhodotorula glutinis* mutant. *Journal of Industrial Microbiology and Biotechnology* 2001, 26:327-332.
31. Sperstad S, Lutnæs B, Stormo S, Liaaen-Jensen S, Landfald B: Torularhodin and torulene are the major contributors to the carotenoid pool of marine *Rhodospiridium babjevae* (Golubev). *Journal of Industrial Microbiology and Biotechnology* 2006, 33:269-273.
32. Koh CM, Liu Y, Du M, Ji L: Molecular characterization of KU70 and KU80 homologues and exploitation of a KU70-deficient mutant for improving gene deletion frequency in *Rhodospiridium toruloides*. *BMC Microbiology* 2014, 14:50.
33. Lazo GR, Stein PA, Ludwig RA: A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Biotechnology (N Y)* 1991, 9:963-967.
34. Cai L, Sun L, Fu L, Ji L: media compositions, selection methods and *agrobacterium* strains for transformation of plants. *Google Patents*; 2009.
35. Buzzini P, Martini A: Production of carotenoids by strains of *Rhodotorula glutinis* cultured in raw materials of agro-industrial origin. *Bioresource Technology* 1999, 71:41-44.
36. Wu S, Hu C, Jin G, Zhao X, Zhao ZK: Phosphate-limitation mediated lipid production by *Rhodospiridium toruloides*. *Bioresour Technol* 2010, 101:6124-6129.
37. Liu Y, Koh CM, Sun L, Hlaing MM, Du M, Peng N, Ji L: Characterization of glyceraldehyde-3-phosphate dehydrogenase gene *RtGPD1* and development of genetic transformation method by dominant selection in oleaginous yeast *Rhodospiridium toruloides*. *Appl Microbiol Biotechnol* 2013, 97:719-729.
38. Smith TL, Leong SA: Isolation and characterization of a *Ustilago maydis* glyceraldehyde-3-phosphate dehydrogenase-encoding gene. *Gene* 1990, 93:111-117.
39. Ji L, Jiang ZD, Liu Y, Koh CM, Zhang LH: A Simplified and efficient method for transformation and gene tagging of *Ustilago maydis* using frozen cells. *Fungal Genet Biol* 2010, 47:279-287.
40. Punt PJ, Dingemans MA, Kuyvenhoven A, Soede RD, Pouwels PH, van den Hondel CA: Functional elements in the promoter region of the *Aspergillus nidulans* *gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene* 1990, 93:101-109.
41. Steiner S, Philippsen P: Sequence and promoter analysis of the highly expressed *TEF* gene of the filamentous fungus *Ashbya gossypii*. *Mol Gen Genet* 1994, 242:263-271.
42. Liu Y, Koh CMJ, Sun L, Hlaing MM, Du M, Peng N, Ji L: Characterization of glyceraldehyde-3-phosphate dehydrogenase gene *RtGPD1* and development of genetic transformation method by dominant selection in oleaginous yeast *Rhodospiridium toruloides*. *Applied microbiology and biotechnology* 2013, 97:719-729.
43. Liu YG, Whittier RF: Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* 1995, 25:674-681.
44. Liu YG, Chen Y: High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *Biotechniques* 2007, 43:649-650, 652, 654 passim.
45. Saelices L, Youssar L, Holdermann I, Al-Babili S, Avalos J: Identification of the gene responsible for torulene cleavage in the *Neurospora* carotenoid pathway. *Mol Genet Genomics* 2007, 278:527-537.
46. Bijttebier S, D'Hondt E, Noten B, Hermans N, Apers S, Voorspoels S: Ultra high performance liquid chromatography versus high performance liquid chromatography: stationary phase selectivity for generic carotenoid screening. *J Chromatogr A* 2014, 1332:46-56.
47. Bijttebier SK, D'Hondt E, Hermans N, Apers S, Voorspoels S: Unravelling ionization and fragmentation pathways of carotenoids using orbitrap technology: a first step towards identification of unknowns. *J Mass Spectrom* 2013, 48:740-754.

48. Hausmann A, Sandmann G: A single five-step desaturase is involved in the carotenoid biosynthesis pathway to beta-carotene and torulene in *Neurospora crassa*. *Fungal Genet Biol* 2000, 30:147-153.
- 5 49. Keller NP, Turner G, Bennett JW: Fungal secondary metabolism-from biochemistry to genomics. *Nature Reviews Microbiology* 2005, 3:937-947.
50. Linnemannstons P, Prado MM, Fernandez-Martin R, Tudzynski B, Avalos J: A carotenoid biosynthesis gene cluster in *Fusarium fujikuroi*: the genes *carB* and *carRA*. *Mol Genet Genomics* 2002, 267:593-602.
- 10 51. Thewes S, Prado-Cabrero A, Prado MM, Tudzynski B, Avalos J: Characterization of a gene in the *car* cluster of *Fusarium fujikuroi* that codes for a protein of the carotenoid oxygenase family. *Mol Genet Genomics* 2005, 274:217-228.
52. Jin JM, Lee J, Lee YW: Characterization of carotenoid biosynthetic genes in the ascomycete *Gibberella zeae*. *FEMS Microbiol Lett* 2010, 302:197-202.
- 15 53. Moline M, Flores MR, Libkind D, Dieguez Mdel C, Farias ME, van Broock M: Photoprotection by carotenoid pigments in the yeast *Rhodotorula mucilaginosa*: the role of torularhodin. *Photochemical & photobiological sciences: Official journal of the European Photochemistry Association and the European Society for Photobiology* 2010, 9:1145-1151.
- 20 54. Moline M, Libkind D, van Broock M: Production of torularhodin, torulene, and beta-carotene by *Rhodotorula* yeasts. *Methods Mol Biol* 2012, 898:275-283.
55. Buzzini P, Innocenti M, Turchetti B, Libkind D, van Broock M, Mulinacci N: Carotenoid profiles of yeasts belonging to the genera *Rhodotorula*, *Rhodospiridium*, *Sporobolomyces*, and *Sporidiobolus*. *Can J Microbiol* 2007, 53:1024-1031.
- 25 56. Liu ZJ, Sun YJ, Rose J, Chung YJ, Hsiao CD, Chang WR, Kuo I, Perozich J, Lindahl R, Hempel J, Wang BC: The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat Struct Biol* 1997, 4:317-326.
- 30 57. Lee LY, Gelvin SB: T-DNA binary vectors and systems. *Plant Physiol* 2008, 146:325-332.
- Bonfim, K. y otros: RNAi-mediated resistance to Bean golden mosaic virus in genetically engineered common bean (*Phaseolus vulgaris*). *Mol Plant Microbe Interact* 2007, 20:717-726.
- 35 Durai, S. y otros: Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucl Acids Res* 2005, 33:5978-5990.
- Fuentes, A. y otros: Intron-hairpin RNA derived from replication associated protein C1 gene confers immunity to tomato yellow leaf curl virus infection in transgenic tomato plants. *Transgenic Res* 2006, 15:291-304.
- 40 Guo, H.S. y otros: A chemical-regulated inducible RNAi system in plants. *Plant J* 2003, 34:383-392.
- Makarova, K.S. y otros: Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011, 9:467-477.
- 45 Mali, P. y otros: RNA-guided human genome engineering via Cas9. *Science* 2013, 339:823-826.
- Mysara, M. y otros: MysiRNA-designer: a workflow for efficient siRNA design. *PLOS one* 2011, 6(10):e25642.
- Vanderschuren, H. y otros: Transgenic cassava resistance to African cassava mosaic virus is enhanced by viral DNA-A bidirectional promoter-derived siRNAs. *Plant Mol Biol* 2007a, 64:549-557.
- 50 Vanderschuren, H. y otros: Engineering resistance to geminiviruses--review and perspectives. *Plant Biotechnology Journal* 2007b, 5:207-220.
- 55 Wang, M.B. y otros: A single copy of a virus-derived transgene encoding hairpin RNA gives immunity to barley yellow dwarf virus. *Mol Plant Pathol* 2000, 1:347-356.
- Wesley, S.V. y otros: Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J* 2001, 27:581-590.
- 60 Yan, P. y otros: High-throughput construction of intron-containing hairpin RNA vectors for RNAi in plants. *PLOS one* 2012, 7(5):e38186.
- Zrachya, A. y otros: Production of siRNA targeted against TYLCV coat protein transcripts leads to silencing of its expression and resistance to the virus. *Transgenic Res* 2007, 16:385-398.
- 65



ES 2 787 225 T3

aag gag gga cat gcg cac tac gag ctg tcg atg gtt cac gtc ctc cac 484  
 Lys Glu Gly His Ala His Tyr Glu Leu Ser Met Val His Val Leu His  
 140 145 150 155

5 cgc aac ttc acc tcg ctc ctc tcg atg gtc cgc cgg tct ctg atc atc 532  
 Arg Asn Phe Thr Ser Leu Leu Ser Met Val Arg Pro Ser Leu Ile Ile  
 160 165 170

10 cag ctc cgc aag ctc cat ccc ttt gtc tct gtc tat tcg cgc gcg acc 580  
 Gln Leu Arg Lys Leu His Pro Phe Val Ser Val Tyr Ser Arg Ala Thr  
 175 180 185

aag tac ttc aag acg gac cgc atg cgg aga gcg ttc acc ttt gcg tcc 628  
 Lys Tyr Phe Lys Thr Asp Arg Met Arg Arg Ala Phe Thr Phe Ala Ser  
 190 195 200

15 atg tac ctt ggc atg tct ccc ttc gac gct ctc ggc gcc tac aac ctc 676  
 Met Tyr Leu Gly Met Ser Pro Phe Asp Ala Leu Gly Ala Tyr Asn Leu  
 205 210 215

20 ctc cag tac acc gag cac tgc gaa ggc atc ctc tac cct ctc ggt ggt 724  
 Leu Gln Tyr Thr Glu His Cys Glu Gly Ile Leu Tyr Pro Leu Gly Gly  
 220 225 230 235

ttc ggt cgc atc cct caa acc ctc caa aaa ctc gcc gaa aag agc gcc 772  
 Phe Gly Arg Ile Pro Gln Thr Leu Gln Lys Leu Ala Glu Lys Ser Gly  
 240 245 250

25 gcc aag ttc cgc ttc aac agt ccc gtc aag cgc gtc acg gtg gag aat 820  
 Ala Lys Phe Arg Phe Asn Ser Pro Val Lys Arg Val Thr Val Glu Asn  
 255 260 265

30 ggc acg gcc aag ggt gtt gaa ctc gag agt ggc gag aag ttg aag gcc 868  
 Gly Thr Ala Lys Gly Val Glu Leu Glu Ser Gly Glu Lys Leu Lys Ala  
 270 275 280

gag atc gtc ctc gtc aat gcg gat ttg gtg tgg agt atg gcg cat ttg 916  
 Glu Ile Val Leu Val Asn Ala Asp Leu Val Trp Ser Met Ala His Leu  
 285 290 295

35 tac gag gag acg agc tac tcg aag agg ctc gag gag cgc ccc gtc agc 964  
 Tyr Glu Glu Thr Ser Tyr Ser Lys Arg Leu Glu Glu Arg Pro Val Ser  
 300 305 310 315

40 tgc tcg tcc atc tcg ttt tac tgg tcg atg aac cgc aag ata ccc cag 1012  
 Cys Ser Ser Ile Ser Phe Tyr Trp Ser Met Asn Arg Lys Ile Pro Gln  
 320 325 330

ctc gac tcg cat acc atc ttc ctc gca gag gag tac cga gag tcc ttc 1060  
 Leu Asp Ser His Thr Ile Phe Leu Ala Glu Glu Tyr Arg Glu Ser Phe  
 335 340 345

gac tcg atc ttc cgc gaa cac cgt atc cgg cat gag cct tcc ttc tac 1108  
 Asp Ser Ile Phe Arg Glu His Arg Ile Pro His Glu Pro Ser Phe Tyr  
 350 355 360

50 gtc aac gtc ccc agc cgt cac gac cct tct gcc gct ccc gcc gac aaa 1156  
 Val Asn Val Pro Ser Arg His Asp Pro Ser Ala Ala Pro Ala Asp Lys  
 365 370 375

gac gcc gtc atc gtc ctc gtc ccc gtc ggg cac att tcc gcc gcc ctc 1204  
 Asp Ala Val Ile Val Leu Val Pro Val Gly His Ile Ser Ala Ala Leu  
 380 385 390 395

ES 2 787 225 T3

ccc tcc tct tcc gac tgg gac aaa gtg gtc gaa gac acg cgt aac aag 1252  
 Pro Ser Ser Ser Asp Trp Asp Lys Val Val Glu Asp Thr Arg Asn Lys  
 400 405 410

5  
 att atc ggc gag atc gag cgc cga ctc gac atc aag gac ctc cga ggc 1300  
 Ile Ile Gly Glu Ile Glu Arg Arg Leu Asp Ile Lys Asp Leu Arg Gly  
 415 420 425

10  
 tgc atc gag cac gag acg atc aac acg cct atc act tgg ggc gag aag 1348  
 Cys Ile Glu His Glu Thr Ile Asn Thr Pro Ile Thr Trp Gly Glu Lys  
 430 435 440

15  
 ttc aac ttg cac cgc ggc agt atc ctt gga ctc agt cac gac ttc ttc 1396  
 Phe Asn Leu His Arg Gly Ser Ile Leu Gly Leu Ser His Asp Phe Phe  
 445 450 455

20  
 aac gtc ctc tct ttc cgc ccc aag acc cgc cac ccg agc gtc aag aac 1444  
 Asn Val Leu Ser Phe Arg Pro Lys Thr Arg His Pro Ser Val Lys Asn  
 460 465 470 475

25  
 gct tac ttc gtc ggc gcg tct gcg cac ccg gga act ggc gtc ccc atc 1492  
 Ala Tyr Phe Val Gly Ala Ser Ala His Pro Gly Thr Gly Val Pro Ile  
 480 485 490

30  
 gtc ctc gcc ggc gcc cgc ctc gtc gca acc cag atc ctc aac gac ctc 1540  
 Val Leu Ala Gly Ala Arg Leu Val Ala Thr Gln Ile Leu Asn Asp Leu  
 495 500 505

35  
 ggc atg ccc atc ccc tcg cgc tgg aac gtc tcc tcc tcc gaa ctc gcg 1588  
 Gly Met Pro Ile Pro Ser Arg Trp Asn Val Ser Ser Ser Glu Leu Ala  
 510 515 520

40  
 acg cac aag acg atc cgc gat gcg gcg gga ggg ttc acc ctc ctc tcg 1636  
 Thr His Lys Thr Ile Arg Asp Ala Ala Gly Gly Phe Thr Leu Leu Ser  
 525 530 535

45  
 gtg ttg ttt ggg ctg atc gcg ttg ttg gtc atg tat ctg cgc gga tga 1684  
 Val Leu Phe Gly Leu Ile Ala Leu Leu Val Met Tyr Leu Arg Gly  
 540 545 550

<210> 2  
 <211> 554  
 <212> PRT  
 <213> Rhodosporidium toruloides  
 <400> 2  
 Met Ala Ala Ala Asn Gly His Gly Lys Gly Lys Pro Ser Val Leu Ile  
 1 5 10 15

50  
 Val Gly Ala Gly Val Gly Gly Thr Ala Ser Ala Ala Arg Leu Ala Gln  
 20 25 30

55  
 Ser Gly Phe Asp Val Thr Val Leu Glu Lys Asn Asp Phe Ala Gly Gly  
 35 40 45

60  
 Arg Cys Ser Leu Phe Thr Asp Pro Thr Lys Ser Phe Arg Phe Asp Gln  
 50 55 60

65

ES 2 787 225 T3

Gly Pro Ser Leu Phe Leu Ile Pro Arg Leu Phe Asp Glu Thr Phe Ser  
 65 70 75 80

5 Asp Leu Gly Thr Ser Leu Glu Asn Glu Gly Ile Lys Leu Val Lys Cys  
 85 90 95

10 Glu Pro Asn Tyr Arg Ile Val Phe Pro Asp Lys Glu Val Val Glu Met  
 100 105 110

Ser Ser Asp Leu Thr Arg Met Lys Lys Gln Val Glu Arg Trp Glu Gly  
 115 120 125

15 Glu Lys Gly Phe Glu Gly Phe Leu Gly Phe Leu Lys Glu Gly His Ala  
 130 135 140

20 His Tyr Glu Leu Ser Met Val His Val Leu His Arg Asn Phe Thr Ser  
 145 150 155 160

25 Leu Leu Ser Met Val Arg Pro Ser Leu Ile Ile Gln Leu Arg Lys Leu  
 165 170 175

His Pro Phe Val Ser Val Tyr Ser Arg Ala Thr Lys Tyr Phe Lys Thr  
 180 185 190

30 Asp Arg Met Arg Arg Ala Phe Thr Phe Ala Ser Met Tyr Leu Gly Met  
 195 200 205

35 Ser Pro Phe Asp Ala Leu Gly Ala Tyr Asn Leu Leu Gln Tyr Thr Glu  
 210 215 220

40 His Cys Glu Gly Ile Leu Tyr Pro Leu Gly Gly Phe Gly Arg Ile Pro  
 225 230 235 240

Gln Thr Leu Gln Lys Leu Ala Glu Lys Ser Gly Ala Lys Phe Arg Phe  
 245 250 255

45 Asn Ser Pro Val Lys Arg Val Thr Val Glu Asn Gly Thr Ala Lys Gly  
 260 265 270

50 Val Glu Leu Glu Ser Gly Glu Lys Leu Lys Ala Glu Ile Val Leu Val  
 275 280 285

Asn Ala Asp Leu Val Trp Ser Met Ala His Leu Tyr Glu Glu Thr Ser  
 290 295 300

55 Tyr Ser Lys Arg Leu Glu Glu Arg Pro Val Ser Cys Ser Ser Ile Ser

60

65

ES 2 787 225 T3

	305		310							315					320
5	Phe Tyr Trp Ser Met Asn Arg Lys Ile Pro Gln Leu Asp Ser His Thr		325						330					335	
	Ile Phe Leu Ala Glu Glu Tyr Arg Glu Ser Phe Asp Ser Ile Phe Arg		340					345					350		
10	Glu His Arg Ile Pro His Glu Pro Ser Phe Tyr Val Asn Val Pro Ser		355				360					365			
15	Arg His Asp Pro Ser Ala Ala Pro Ala Asp Lys Asp Ala Val Ile Val		370			375					380				
20	Leu Val Pro Val Gly His Ile Ser Ala Ala Leu Pro Ser Ser Ser Asp		385			390				395					400
	Trp Asp Lys Val Val Glu Asp Thr Arg Asn Lys Ile Ile Gly Glu Ile		405					410					415		
25	Glu Arg Arg Leu Asp Ile Lys Asp Leu Arg Gly Cys Ile Glu His Glu		420				425					430			
30	Thr Ile Asn Thr Pro Ile Thr Trp Gly Glu Lys Phe Asn Leu His Arg		435				440					445			
35	Gly Ser Ile Leu Gly Leu Ser His Asp Phe Phe Asn Val Leu Ser Phe		450			455					460				
	Arg Pro Lys Thr Arg His Pro Ser Val Lys Asn Ala Tyr Phe Val Gly		465			470				475					480
40	Ala Ser Ala His Pro Gly Thr Gly Val Pro Ile Val Leu Ala Gly Ala		485						490				495		
45	Arg Leu Val Ala Thr Gln Ile Leu Asn Asp Leu Gly Met Pro Ile Pro		500					505				510			
	Ser Arg Trp Asn Val Ser Ser Ser Glu Leu Ala Thr His Lys Thr Ile		515				520					525			
50	Arg Asp Ala Ala Gly Gly Phe Thr Leu Leu Ser Val Leu Phe Gly Leu		530			535					540				
55	Ile Ala Leu Leu Val Met Tyr Leu Arg Gly		545			550									
	<210> 3														
	<211> 1916														
	<212> ADN														
60	<213> Rhodosporidium toruloides														
	<220>														
	<221> CDS														
	<222> (78) (1916)														
	<400> 3														

ES 2 787 225 T3

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

gggctgttct cgctattctc gaggggtcgt cctggggctg tctgtgactt gctatcgact 60  
 gctagactcg cgctcgc atg ggc gga ctg gac tac tgg ctc gtc cac ctc 110  
 Met Gly Gly Leu Asp Tyr Trp Leu Val His Leu  
 1 5 10  
 cgc tgg act atc ccg cca gcg cta gtc ctc tgg agc acc ttc aga aag 158  
 Arg Trp Thr Ile Pro Pro Ala Leu Val Leu Trp Ser Thr Phe Arg Lys  
 15 20 25  
 ctt agg aca cgg cgc gat gtc tac aag acg ctc ttc ctc atc acc atc 206  
 Leu Arg Thr Arg Arg Asp Val Tyr Lys Thr Leu Phe Leu Ile Thr Ile  
 30 35 40  
 gca gta acg gcg acg att ccc tgg gac tcg tac ctc atc cgg cac agg 254  
 Ala Val Thr Ala Thr Ile Pro Trp Asp Ser Tyr Leu Ile Arg His Arg  
 45 50 55  
 atc tgg tca tac ccc gag tca tcc gtc gtc ggg ccg acc ctc ttc gcg 302  
 Ile Trp Ser Tyr Pro Glu Ser Ser Val Val Gly Pro Thr Leu Phe Ala  
 60 65 70 75  
 ata ccc tac gaa gag atc ttc ttc ttc ttc gtc caa acc tac atc acc 350  
 Ile Pro Tyr Glu Glu Ile Phe Phe Phe Phe Val Gln Thr Tyr Ile Thr  
 80 85 90  
 gcg acc gtc tac gcc ctc ttc agc cgc cca gtc gtc cac gcc gtc ctc 398  
 Ala Thr Val Tyr Ala Leu Phe Ser Arg Pro Val Val His Ala Val Leu  
 95 100 105  
 ctc cct cgg aaa cct agc gac gga cga gca gcg agg tgg att gga acg 446  
 Leu Pro Arg Lys Pro Ser Asp Gly Arg Ala Ala Arg Trp Ile Gly Thr  
 110 115 120  
 gcg gcg ttc ttg gcc atc ttt gcg ctc gcg tgg gca aag ttg gag gag 494  
 Ala Ala Phe Leu Gly Ile Phe Ala Leu Ala Trp Ala Lys Leu Glu Glu  
 125 130 135  
 gga gga gag ggg acg tac ctc gcg ttg att gtt gga tgg gtg gcg ccg 542  
 Gly Gly Glu Gly Thr Tyr Leu Ala Leu Ile Val Gly Trp Val Ala Pro  
 140 145 150 155  
 ttc ctt gcg ctg ctt tgg ttc att gcc tca acc cac atc ctc gcc atg 590  
 Phe Leu Ala Leu Leu Trp Phe Ile Ala Ser Thr His Ile Leu Ala Met  
 160 165 170  
 ccc cgc tgg gct gtc ggt ctc ccc atc ctc cta ccg acg ctc tac ctg 638  
 Pro Arg Trp Ala Val Gly Leu Pro Ile Leu Leu Pro Thr Leu Tyr Leu  
 175 180 185  
 tgg gag tgc gac gcg cga gct ctg caa cgc ggg act tgg gtc atc gag 686  
 Trp Glu Cys Asp Ala Arg Ala Leu Gln Arg Gly Thr Trp Val Ile Glu

ES 2 787 225 T3

	190		195		200													
5	aag ggc acg aag ctg ggc ttg gct ttc cgc ggt ctc gag att gag gag																	734
	Lys Gly Thr Lys Leu Gly Leu Ala Phe Arg Gly Leu Glu Ile Glu Glu																	
	205																	
	gcc gtc ttc ttc ctc ttg acg aac gtc atg atc gtc ttc ggc ctg gtc																	782
	Ala Val Phe Phe Leu Leu Thr Asn Val Met Ile Val Phe Gly Leu Val																	
	220																	
10	gcc tgc gac tac tgc ctc gca gtt cac gac ctc cgc tcc tac gac aag																	830
	Ala Cys Asp Tyr Cys Leu Ala Val His Asp Leu Arg Ser Tyr Asp Lys																	
	240																	
15	cgc acc tca tcc gtc ttc cca ccc ctg cgc gac ttc ctc ccg atc ctc																	878
	Arg Thr Ser Ser Val Phe Pro Pro Leu Arg Asp Phe Leu Pro Ile Leu																	
	255																	
	ctc aac tcg ccc gac gcc gca cag cga caa cgc atc gag gac ttg cag																	926
	Leu Asn Ser Pro Asp Ala Ala Gln Arg Gln Arg Ile Glu Asp Leu Gln																	
	270																	
20	gcg gct atc gag atc ttg tcg att cac tcg aag agc ttc tcg acg gcg																	974
	Ala Ala Ile Glu Ile Leu Ser Ile His Ser Lys Ser Phe Ser Thr Ala																	
	285																	
25	agt cag gtg ttt gag ggc agg ttg agg ctg gac ctc ctc tcg ctc tac																	1022
	Ser Gln Val Phe Glu Gly Arg Leu Arg Leu Asp Leu Leu Ser Leu Tyr																	
	300																	
	gcc tgg tgc cga gtc tgc gac gac ctg atc gac aac gcc tcg aca gtc																	1070
	Ala Trp Cys Arg Val Cys Asp Asp Leu Ile Asp Asn Ala Ser Thr Val																	
	320																	
30	gca gca gcc gaa tcc aac atc gac atg att tcg ggc tgc ctc gac ctc																	1118
	Ala Ala Ala Glu Ser Asn Ile Asp Met Ile Ser Gly Cys Leu Asp Leu																	
	335																	
35	ctc tac cct ccc tcc tcc tcc acg ccc acc tct ctc ccc gtc cgc gtt																	1166
	Leu Tyr Pro Pro Ser Ser Ser Thr Pro Thr Ser Leu Pro Val Arg Val																	
	350																	
	tcg aac aag cag atc gag gcg gcc ttg ccc ggc ttg agc gag ccc gaa																	1214
	Ser Asn Lys Gln Ile Glu Ala Ala Leu Pro Gly Leu Ser Glu Pro Glu																	
	365																	
40	cga ggc gca ttc cgc ctc ctc agc ctt ctc cct att gcc cgc ccg ccg																	1262
	Arg Gly Ala Phe Arg Leu Leu Ser Leu Leu Pro Ile Ala Arg Pro Pro																	
	380																	
45	ctt aac gaa ctc ctc gac ggc ttc cgc acc gac ctc tcc ttc ctc gct																	1310
	Leu Asn Glu Leu Leu Asp Gly Phe Arg Thr Asp Leu Ser Phe Leu Ala																	
	400																	
50	ctc tcc gac tcg aag ggt gtc aag acg aac ggc agc gca aac ggt aac																	1358
	Leu Ser Asp Ser Lys Gly Val Lys Thr Asn Gly Ser Ala Asn Gly Asn																	
	415																	
	ggg aac ggc ata tcg agt atc tcc gcc gag ttg ccc atc aag acc gac																	1406
	Gly Asn Gly Ile Ser Ser Ile Ser Ala Glu Leu Pro Ile Lys Thr Asp																	
	430																	
55	tcg gat ctc ctc gtc tac gcc aat aac gtc gcc tcg tcc gtc gcc gat																	1454

ES 2 787 225 T3

Ser Asp Leu Leu Val Tyr Ala Asn Asn Val Ala Ser Ser Val Ala Asp  
445 450 455

5 ctc tgc gtc caa ctc gtc tgg gca cac tgc acg cct tac tgc cgc aca 1502  
Leu Cys Val Gln Leu Val Trp Ala His Cys Thr Pro Tyr Ser Arg Thr  
460 465 470 475

ccc gct caa tca gtc ccg cgc gac ccg acc ctc tca gaa gcg gag aac 1550  
Pro Ala Gln Ser Val Pro Arg Asp Pro Thr Leu Ser Glu Ala Glu Asn  
480 485 490

10 gca cat gtt ctc gct gcg gcg agg gag atg gga cag gct ctt cag ctc 1598  
Ala His Val Leu Ala Ala Ala Arg Glu Met Gly Gln Ala Leu Gln Leu  
495 500 505

15 gtc aat atc gcg cgg gac gta ccg gcg gat ctg aag att ggg cgg atc 1646  
Val Asn Ile Ala Arg Asp Val Pro Ala Asp Leu Lys Ile Gly Arg Ile  
510 515 520

20 tac ctc cct ggt cgc ggg ctc gac acg cct gtg ccc gag ttg acg gcg 1694  
Tyr Leu Pro Gly Arg Gly Leu Asp Thr Pro Val Pro Glu Leu Thr Ala  
525 530 535

25 gat agg cgg gcc cta ctt gct cgt gcg aac gag atg gct gca cag agt 1742  
Asp Arg Arg Ala Leu Leu Ala Arg Ala Asn Glu Met Ala Ala Gln Ser  
540 545 550 555

aag gat gcg ata gag aag ttg ccg caa gag gcg aga gga ggg atc agg 1790  
Lys Asp Ala Ile Glu Lys Leu Pro Gln Glu Ala Arg Gly Gly Ile Arg  
560 565 570

30 gcg gcg tgt ttg gtt tat ctc agc att ggg gac gcg gtc ggg agg gcc 1838  
Ala Ala Cys Leu Val Tyr Leu Ser Ile Gly Asp Ala Val Gly Arg Ala  
575 580 585

35 ttg gac gag ggg agg gtc atg gag cgc gcg agg gtg tcc aag ggg gcg 1886  
Leu Asp Glu Gly Arg Val Met Glu Arg Ala Arg Val Ser Lys Gly Ala  
590 595 600

agg gcg cgc aaa gcg tgg cag gcg ttg tga 1916  
Arg Ala Arg Lys Ala Trp Gln Ala Leu  
605 610

40 <210> 4  
<211> 612  
<212> PRT  
<213> Rhodosporidium toruloides  
<400> 4

45 Met Gly Gly Leu Asp Tyr Trp Leu Val His Leu Arg Trp Thr Ile Pro  
1 5 10 15

50 Pro Ala Leu Val Leu Trp Ser Thr Phe Arg Lys Leu Arg Thr Arg Arg  
20 25 30

Asp Val Tyr Lys Thr Leu Phe Leu Ile Thr Ile Ala Val Thr Ala Thr  
35 40 45

55  
60  
65

ES 2 787 225 T3

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

ES 2 787 225 T3

Gly Arg Leu Arg Leu Asp Leu Leu Ser Leu Tyr Ala Trp Cys Arg Val  
 305 310 315 320  
 5 Cys Asp Asp Leu Ile Asp Asn Ala Ser Thr Val Ala Ala Ala Glu Ser  
 325 330 335  
 10 Asn Ile Asp Met Ile Ser Gly Cys Leu Asp Leu Leu Tyr Pro Pro Ser  
 340 345 350  
 Ser Ser Thr Pro Thr Ser Leu Pro Val Arg Val Ser Asn Lys Gln Ile  
 355 360 365  
 15 Glu Ala Ala Leu Pro Gly Leu Ser Glu Pro Glu Arg Gly Ala Phe Arg  
 370 375 380  
 20 Leu Leu Ser Leu Leu Pro Ile Ala Arg Pro Pro Leu Asn Glu Leu Leu  
 385 390 395 400  
 Asp Gly Phe Arg Thr Asp Leu Ser Phe Leu Ala Leu Ser Asp Ser Lys  
 405 410 415  
 25 Gly Val Lys Thr Asn Gly Ser Ala Asn Gly Asn Gly Asn Gly Ile Ser  
 420 425 430  
 30 Ser Ile Ser Ala Glu Leu Pro Ile Lys Thr Asp Ser Asp Leu Leu Val  
 435 440 445  
 Tyr Ala Asn Asn Val Ala Ser Ser Val Ala Asp Leu Cys Val Gln Leu  
 450 455 460  
 35 Val Trp Ala His Cys Thr Pro Tyr Ser Arg Thr Pro Ala Gln Ser Val  
 465 470 475 480  
 40 Pro Arg Asp Pro Thr Leu Ser Glu Ala Glu Asn Ala His Val Leu Ala  
 485 490 495  
 45 Ala Ala Arg Glu Met Gly Gln Ala Leu Gln Leu Val Asn Ile Ala Arg  
 500 505 510  
 Asp Val Pro Ala Asp Leu Lys Ile Gly Arg Ile Tyr Leu Pro Gly Arg  
 515 520 525  
 50 Gly Leu Asp Thr Pro Val Pro Glu Leu Thr Ala Asp Arg Arg Ala Leu  
 530 535 540  
 55 Leu Ala Arg Ala Asn Glu Met Ala Ala Gln Ser Lys Asp Ala Ile Glu  
 545 550 555 560

ES 2 787 225 T3

Lys Leu Pro Gln Glu Ala Arg Gly Gly Ile Arg Ala Ala Cys Leu Val  
 565 570 575  
 5 Tyr Leu Ser Ile Gly Asp Ala Val Gly Arg Ala Leu Asp Glu Gly Arg  
 580 585 590  
 10 Val Met Glu Arg Ala Arg Val Ser Lys Gly Ala Arg Ala Arg Lys Ala  
 595 600 605  
 Trp Gln Ala Leu  
 610  
 15 <210> 5  
 <211> 1121  
 <212> ADN  
 <213> Rhodosporidium toruloides  
 <220>  
 20 <221> CDS  
 <222> (42) (1121)  
 <400> 5  
 actcgcctccc gctgctctcgc ctgcgctgctg cttgtctgag g atg tcg ctg gat tgg 56  
 Met Ser Leu Asp Trp  
 1 5  
 25 tac gac aac ttt atc gac aag gtg cag ggc acg ccg tcg tgg cag ccc 104  
 Tyr Asp Asn Phe Ile Asp Lys Val Gln Gly Thr Pro Ser Trp Gln Pro  
 10 15 20  
 30 gcg cag gag cag gtc ctg aca gag ccg tac acc tac ctc gcc tcg atc 152  
 Ala Gln Glu Gln Val Leu Thr Glu Pro Tyr Thr Tyr Leu Ala Ser Ile  
 25 30 35  
 35 ccg ggc aag gag gtg cgc tcg gcg ctc atc gcg gcg ttc aac cag tgg 200  
 Pro Gly Lys Glu Val Arg Ser Ala Leu Ile Ala Ala Phe Asn Gln Trp  
 40 45 50  
 40 atg ggt gtc gca gac gtc gat ctc gag att gtc aag aaa gtc gtc ggg 248  
 Met Gly Val Ala Asp Val Asp Leu Glu Ile Val Lys Lys Val Val Gly  
 55 60 65  
 45 atg ctg cac acg gcc agc ctg ctg atg gac gac gtc gag gac gac tcg 296  
 Met Leu His Thr Ala Ser Leu Leu Met Asp Asp Val Glu Asp Asp Ser  
 70 75 80 85  
 50 cac ctc cgt cga gcc atg cct gtc gca cac aag atc tac gga atc ccg 344  
 His Leu Arg Arg Gly Met Pro Val Ala His Lys Ile Tyr Gly Ile Pro  
 90 95 100  
 55 cag acg atc aac tcg gcc aac tat gtc tac ttt ctc gcg ttt caa gaa 392  
 Gln Thr Ile Asn Ser Ala Asn Tyr Val Tyr Phe Leu Ala Phe Gln Glu  
 105 110 115  
 60 ctc cag cgg ata cac ccg cgg cca gcc atc aag gtc gaa gag atg gtc 440  
 Leu Gln Arg Ile His Pro Arg Pro Gly Ile Lys Val Glu Glu Met Val

ES 2 787 225 T3

	120		125			130		
5	act gaa gag cta ttg aac ctg cat cgc gga caa ggg atg gac cta ttc							488
	Thr Glu Glu Leu Leu Asn Leu His Arg Gly Gln Gly Met Asp Leu Phe							
	135		140			145		
10	tgg cgc gag aac ctg atc tgt ccg aca gaa ccc gag tac atc gac atg							536
	Trp Arg Glu Asn Leu Ile Cys Pro Thr Glu Pro Glu Tyr Ile Asp Met							
	150		155			160		165
15	gtc aac aac aag acg gga gga ctg ttc cgc att gcg atc aag ttg atg							584
	Val Asn Asn Lys Thr Gly Gly Leu Phe Arg Ile Ala Ile Lys Leu Met							
		170				175		180
20	atg gcc gct tcg cct gct cca cca cgg gat tac gtc ccg ctc gcc aac							632
	Met Ala Ala Ser Pro Ala Pro Pro Arg Asp Tyr Val Pro Leu Ala Asn							
		185				190		195
25	ctg atc ggc atc atc ttc cag atc cgc gac gac tac gtg aac ctg caa							680
	Leu Ile Gly Ile Ile Phe Gln Ile Arg Asp Asp Tyr Val Asn Leu Gln							
		200				205		210
30	tcc gtc gag tac gca aac aat aag ggc tac tgc gaa gac ttc tcc gaa							728
	Ser Val Glu Tyr Ala Asn Asn Lys Gly Tyr Cys Glu Asp Phe Ser Glu							
		215				220		225
35	ggc aaa ttc tcc ttc ccc atc gtc cac tcg atc cgc tcc gac acc tcg							776
	Gly Lys Phe Ser Phe Pro Ile Val His Ser Ile Arg Ser Asp Thr Ser							
		230				235		240
40	aac cgc caa atc ctc aac atc ctg cgc gag cgg cct tcc tct ccc ggt							824
	Asn Arg Gln Ile Leu Asn Ile Leu Arg Glu Arg Pro Ser Ser Pro Gly							
		250				255		260
45	ccg aaa gag tac gcc gtc agc tac atg gag aca cgg acg ggc tcg ttc							872
	Pro Lys Glu Tyr Ala Val Ser Tyr Met Glu Thr Arg Thr Gly Ser Phe							
		265				270		275
50	gcg tat acg cgc gag gtg ttg cgc aag ttg acg cag cag gcg agg gac							920
	Ala Tyr Thr Arg Glu Val Leu Arg Lys Leu Thr Gln Gln Ala Arg Asp							
		280				285		290
55	gag gtt gcg cgg ctg gga ggg aac agg ggc gtc gag gcg att ctc gac							968
	Glu Val Ala Arg Leu Gly Gly Asn Arg Gly Val Glu Ala Ile Leu Asp							
		295				300		305
60	aag ctt gtg ctg gag gaa ccg cag gtg aag gtc aat ggc gtc gag ggc							1016
	Lys Leu Val Leu Glu Glu Pro Gln Val Lys Val Asn Gly Val Glu Gly							
		310				315		320
65	gag gcg atg gag agg aag ctt gag gag gtc gtc aag agc aaa ccg gtc							1064
	Glu Ala Met Glu Arg Lys Leu Glu Glu Val Val Lys Ser Lys Pro Val							
		330				335		340
70	aag gcg gtg acg aac ggc gtc aac ggc gtc cat gcg cac gcg ctc ccc							1112
	Lys Ala Val Thr Asn Gly Val Asn Gly Val His Ala His Ala Leu Pro							
		345				350		355
75	aaa gcc tga							1121
	Lys Ala							
	<210> 6							
	<211> 359							
	<212> PRT							
	<213> Rhodosporidium toruloides							
	<400> 6							

ES 2 787 225 T3

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

ES 2 787 225 T3

	Glu	Asp	Phe	Ser	Glu	Gly	Lys	Phe	Ser	Phe	Pro	Ile	Val	His	Ser	Ile	
	225					230					235					240	
5	Arg	Ser	Asp	Thr	Ser	Asn	Arg	Gln	Ile	Leu	Asn	Ile	Leu	Arg	Glu	Arg	
					245					250					255		
10	Pro	Ser	Ser	Pro	Gly	Pro	Lys	Glu	Tyr	Ala	Val	Ser	Tyr	Met	Glu	Thr	
				260					265					270			
15	Arg	Thr	Gly	Ser	Phe	Ala	Tyr	Thr	Arg	Glu	Val	Leu	Arg	Lys	Leu	Thr	
			275					280					285				
20	Gln	Gln	Ala	Arg	Asp	Glu	Val	Ala	Arg	Leu	Gly	Gly	Asn	Arg	Gly	Val	
	290						295					300					
25	Glu	Ala	Ile	Leu	Asp	Lys	Leu	Val	Leu	Glu	Glu	Pro	Gln	Val	Lys	Val	
	305					310					315					320	
30	Asn	Gly	Val	Glu	Gly	Glu	Ala	Met	Glu	Arg	Lys	Leu	Glu	Glu	Val	Val	
					325					330					335		
35	Lys	Ser	Lys	Pro	Val	Lys	Ala	Val	Thr	Asn	Gly	Val	Asn	Gly	Val	His	
				340					345					350			
40	Ala	His	Ala	Leu	Pro	Lys	Ala										
				355													
45	<210>	7															
	<211>	2430															
50	<212>	ADN															
	<213>	Rhodosporidium toruloides															
55	<400>	7															
	accactc	gct	agtgaag	caa	tggcag	ctgc	gaacgg	acac	ggcaagg	gaa	agccct	cggt				60	
60	gctcat	cgtc	ggagc	ggcg	tggagg	cac	tgcgt	ccgc	gctcgc	cctc	cccagt	ccgg				120	
	gttcgac	gtg	acaggt	gcgc	acactc	gtct	cctgcg	ccgt	cctgcgt	oga	tcgggg	gacg				180	
65	tgggag	ctat	gaggg	tgact	gtcgag	aggt	actgac	gagt	tcgcct	cgtc	gacagt	cctc				240	
	gagaaga	aacg	actttg	ccgg	cggacg	atgc	tccctc	ttca	ccgatc	ccgac	caagt	ccttc				300	
70	cgcttc	gacc	agggccc	gag	cctgtt	cctc	atccc	gcgac	tgttc	gacga	gacctt	cagc				360	
	gacctc	ggga	cgagc	cttga	gaacg	agggc	atcaag	cttg	tcaagt	gcga	gccaa	actac				420	
75	cggatc	gtct	tcccc	gaaa	ggaggt	cgtc	gagat	gagca	gcgact	tgac	gaggat	gaag				480	
	aagcag	gtcg	agcgg	tggga	gggag	agaag	ggcttt	gaag	ggtgag	tggc	ggcga	agtt				540	
80	gaagc	gagac	cacgag	cgga	gatg	ccccctt	cgctc	gtgtt	ccctc	gcgcg	tttt	tttcca				600	
	gttaact	gac	gttcat	gttc	gatac	agatt	tctc	ggcttc	ctgaag	gagg	gacat	gcgca				660	

ES 2 787 225 T3

	ctacgagctg tcgatggttc acgtcctcca cgcgaacttc acctcgtcc tctc gatggt	720
	ccgccgctct ctgatcatcc agctccgcaa gctccatccc tttgtctctg tcgtacgtca	780
5	cctcatcaag gctgatacga ctcgtcggga acgcaggatt gctgatgagc gtcgcatcga	840
	ctcagtattc gcgcgcgacc aagtacttca agacggaccg catgctggaga gcgttcacct	900
	ttgcgtccat gtaccttggt agttggagca cattacaata cgagcgcag ctatggcgcc	960
10	tcaattgctc gccaa gatcc tctcctcgtg acccgcacct cctctttcgc agggcatgtc	1020
	tcccttcgac gctctcggcg cctacaacct cctccagtac accgagcact gcgaaggcat	1080
	cctctaccct ctcggtggtt tcggtcgcct ccctcaaacc ctccaaaaac tcgccgaaaa	1140
15	gagcggcgcc aagttccgct tcaacagtcc cgtcaagcgc gtcacggtgg aga atggcac	1200
	ggccaagggt gttgaactcg agagtggcga gaagttgaag gccgagatcg tctcgtcaa	1260
20	tgccgatttg gtgtggagta tggcgcattt gtacgaggag acgagctact cgaagaggct	1320
	cgaggagcgc cccgtcagct gctcgtccat ctcgttttac tggtcgatga accggtgcgt	1380
	tgagattgtc actcttcaact tatctcagag cgggtgatcc tgacgctgtt cgagccattc	1440
25	cacagcaaga taccacgct cgactcgcct accatcttcc tcgcagagga gtaccgagag	1500
	tgagcacgag atcgcctcggg tactcgcacac aagattgact gctgacggat acgcactgtc	1560
	caggtccttc gactcgcctt tccgcgaaca ccgtatccc catgagcctt ccttctacgt	1620
30	caacgtcccc agccgtcaag acccttcgtg cgttttgctc acctcgttag ctccgccctc	1680
	tcgctcactc attgtccgct cgcagtgccg ctcccgcga caaagacgcc gtcacgtcc	1740
35	tcgtccccgt cgggcacatt tccgcgcgcc tcccctctc tccgactgg gacaaagtgg	1800
	tcgaagacac gcgtaacaag attatcggcg agatcgagcg ccgactcgac atcaaggacc	1860
	tccgaggctg catcgagcac gagacgatca acacgcctat cacttggggc gagaagttca	1920
40	acttgaccg cggcagatc cttggactca gtcacgactt cttgtgagtc tcgcgcgcta	1980
	gtctctggtc ctgctcgtg acgttagcgc tgaccttccc aatctttcgt ttctctcagc	2040
	aacgtcctct ctttccgccc caagaccgc caccgagcg tcaagaacgc ttacttcgtc	2100
45	ggcgcgtctg cgcaccggg aactgggtga gtcacggtgg tcttcgctcg tctcgttgc	2160
	cgttgctccc tgggcagcgc cgcgcgtgtg ggcggaagcc ttgtgactga ctctcgtcgt	2220
	cgttgacgcg tccccatcgt cctcgcggc gccgcctcg tcgcaacca gatcctcaac	2280
50	gacctcggca tgccccatccc ctgcgctgg aacgtctcct cctccgaact cgcgacgcac	2340
	aagacgatcc gcgatcggc gggagggttc accctcctct cgggtgtgtt tgggctgatc	2400
55	gcgttggttg tcatgtatct gcgcggatga	2430
	<210> 8	
	<211> 2334	
	<212> ADN	
	<213> Rhodosporidium toruloides	
60	<400> 8	

65

ES 2 787 225 T3

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

gggctgttct cgctattctc gaggggtcgt cctggggctg tctgtgactt gctatcgact 60  
 gctagactcg cgctcgcacg ggccgactgg actactggct cgtgtgcgtc cacctcgcctc 120  
 ccgttcactc gcatcctccg gactcaccct cgtgttgctc gtaacgcgcg acagccacct 180  
 ccgctggact atcccgccag cgctagtcct ctggagcacc ttcagaaagc ttaggacacg 240  
 gcgcgatgtc tacaagacgc tcttcctcat caccgtgcgt cctcaccgct tcctccgtcg 300  
 gctcgatgct ccaatcgcca actgacctct gctcgtcccg cagatcgag taacggttcg 360  
 tgcagctcat gtaaaccgga agacgagaga ggctgaccga aagacgcgca ggcgacgatt 420  
 ccctgggact cgtacctcat ccggcacagg gtgcgccaag tcttgctctg tatccccaaa 480  
 accgccgtct aacgctcgcg ccggacagat ctggtcatac cccgagtcac ccgtcgtcgg 540  
 gccgaccctc ttccgcgatac cctacgaaga gatcttcttc ttcttcgtcc aaacctacat 600  
 cacccgcacc gtctacgcc tcttcagccg cccagtcgtc cacgccgtcc tcctccctcg 660  
 gaaacctagc gacggacgag cagcgaggtg gattggaacg gcggcgttct tgggcatctt 720  
 tgcgctcgcg tgggcaaagt tggaggaggg aggagagggg acgtacctcg cgttgattgt 780  
 tggatgggtg gcgccgttcc ttgcctgct ttggtgcggt cagcgagttt ccggacgagg 840  
 aggatgactg acttgtgtcg cgcaggttca ttgcctcaac ccacatcctc gccatgcccc 900  
 gctgggctgt cggctcctcc atcctcctac cgacgctcta cctgtgggag tgcgacgcgc 960  
 gagctctgca acgcgggact tgggtcatcg agaagggcac gaagctgggc ttggctttcc 1020  
 gcggctctga gattgagtgc gtcctgccgt ctttcgtctt tctccgcgct tgctgactgt 1080  
 acgtgcttca cagggaggcc gtcttcttcc tcttgacgaa cgtcatgac gtcttcggcc 1140  
 tggtcgcctg cgactactgc ctccagttc acgacctccg ctccctacgac aagcgcacct 1200  
 catccgtctt cccaccctcg cgcgacttcc tcccgatcct cctcaactcg cccgacgccg 1260  
 cacagcgaca acgcatcgag gacttgcagg cggctatcga gatcttgtcg atcactcga 1320  
 agagcttctc gacggcgagt caggtgtttg agggcaggtt gaggctggac ctccctctcg 1380  
 tgtgagtgcg gctccttccg aagcaaagac tcgagctgac cttgagagta tcgcagctac 1440  
 gcctgggtgcc gagtctgcga cgacctgatc gacaacgcct cgacagtcgc agcagccgaa 1500  
 tccaacatcg acatgatttc gggctgcctc gacctcctct accctccctc ctccctccag 1560  
 cccacctctc tccccgtccg cgtttogaac aagcagatcg aggcggcctt gcccggttg 1620  
 agcgagcccg aacgaggcgc attccgcctc ctccagcttc tcctattgc ccgccgcg 1680  
 cttaacgaac tcctcgacgg ctccgcacc gacctctcct tcctcgtct ctccgactcg 1740

ES 2 787 225 T3

	aagggtgtca agacgaacgg cagcgcaaac ggtaacggga acggcatatc gagtatctcc	1800
	gccgagttgc ccatcaagac cgactcggat ctctcgtct acgccaataa cgtcgcctcg	1860
5	tccgtcgcgg atctctgCGT ccaactcgtc tgggcacact gcacgcctta ctcgcgcaca	1920
	cccgctcaat cagtcccggc cgacccgacc ctctcagaag cggagaacgc acatgttctc	1980
10	gctcggcgca gggagatggg acaggtctt cagctcgtca atatcgcggc ggacgtaccg	2040
	gcggatctga agattggcg gatctacctc cctggtcgCG ggcctcgcac gcctgtgccc	2100
	gagttgacgg cggataggcg ggcctactt gctcgtcga acgagatggc tgcacagagt	2160
15	aaggatgcga tagagaagt gccgcaagag gcgagaggag ggatcagggc ggcgtgtttg	2220
	gtttatctca gcattgggga cgcggtcggg agggccttgg acgaggggag ggtcatggag	2280
	cgcgcgaggg tgtccaaggg ggcgagggcg cgcaaagcgt ggcaggcgtt gtga	2334
20	<210> 9	
	<211> 1546	
	<212> ADN	
	<213> Rhodosporidium toruloides	
	<400> 9	
25	actcgtctcc gctgctctcg ctcgtcgtg cttgtctgag gatgtcgtg gattggtacg	60
	acaactttat cgacaaggcg cagggcacgc cgtcgtggca gcccgcgag gagcaggtgc	120
	gctcgcggcg caccgcccgc gcctcccac atcctcacga gtcgtcgaag ccagcgtga	180
30	cacgctcgag gacaggtcct gacagagccg tacacctacc tcgcctcgat cccgggcaag	240
	gaggtgCGT cggcgtcat cgcgCGTt caccagtgga tgggtgCGC agacgtcgat	300
	ctcagagattg tcaagaaagt cgtcgggatg ctgcacacgg ccagcctgct gtgcgctatc	360
35	gagacgccct cgctgtccag ctacccggg tgcctcacgat ccgctcgat tcccgcagga	420
	tggacgacgt cgaggacgac tcgcacctcc gtcgaggcat gcctggtcag tccccgatct	480
	ccgagccgcc tcccgtcga gagactgatc ccgtgtaccg tcgcacagtc gcacacaaga	540
40	tctacggaat cccgcagacg atcaactcgg ccaactatgt ctactttctc gcgtttcaag	600
	aactccagcg gatacaccgg cggccaggca tcaaggtcga agagatggc actggtgggt	660
45	ccgccttttc ttgactaaag cttcgggtgc taacgcagtc cccgcagaag agctattgaa	720
	cctgcacgCG ggacaaggga tggacctatt ctggcgcgag aacctgatct gtccgacaga	780
	acccgagtac atcgacatgg tcaacaacag tgcgtcctgc gaggctgttc tatgCGat	840
50	acgctgactc ggtcCGaatc gcagagacgg gaggactgtt ccgattgCG atcaagttga	900
	tgatggcCGc ttcgctgct ccaccacggc gcgctccctg cctcGacttg ctgggactCG	960
	cagctgatga ttgcgttGca gggattacgt cccgctcGCC aacctgatCG gcatcatctt	1020
55	ccagatccCG gacgactacg tgaacctGca atccgctgag gtgcgtttCG ctctcattat	1080
60		
65		

ES 2 787 225 T3

	gggctgctcc aacgctaacg gatcggtcgc acagtacgca aacaataagg gctactgcca	1140
	agacttctcc gaaggcaaat tctccttccc catcgtccac tcgatccgct ccgacacctc	1200
5	gaaccgcaa atcctcaaca tctcgcgca gggccctcc tctcccggc cgaagagta	1260
	cgccgtcagc tacatggaga cacggacggg ctcgttcgcg tatacgcgcg aggtgttgcg	1320
10	caagttgacg cagcagcga gggacgaggt tgcgcggctg ggagggaaaca gggcgtcga	1380
	ggcgattctc gacaagcttg tgctggagga accgcaggtg aaggtcaatg gcgtcgaggg	1440
	cgaggcgatg gagaggaagc ttgaggaggt cgtcaagagc aaaccggtca aggcggtgac	1500
15	gaacggcgtc aacggcgtcc atgcgcacgc gctccccaaa gcctga	1546
	<210> 10	
	<211> 2079	
	<212> ADN	
	<213> Rhodosporidium toruloides	
20	<400> 10	
	atggggacga agccccgca cacctacaag caggggaact acctccccgt gcgagaggag	60
	cgcacgctcg aggagtgcgc ttgcaaggt cagctcccgc aagagctcgt cggcgggtgc	120
25	gtttccgcgc cctttcctgc tcccatacag tgcttacagt gctcctacag gatgtacgtg	180
	cggaatgggg gtgagcctgc gctggcagag ttgagaggag acgactccgc agacggacct	240
	gcctaccact ggggtcgttc ggcctcgtcc gcctgcgcat tctacacgag gagctgacac	300
30	ttgacgacga cagttcgaag gtgacggcat gctgacgggc gtgtacttca agcccgtccc	360
	ctcctcctcg cccttcgacc cgaacccgcg tatcgccccg ctcttcgtca acaatacgt	420
	cctcaccgac gtcttcctcg cctcgaagc gctcggcgtc acacatccca tcctaccctc	480
35	catcgccaca ctctcggct caatctggtc cctccacctc atcctctact ccgttttctg	540
	cgccctcttc ctgccttct gtccttctt caccgaatcg cctttgcgac atctctcagt	600
40	cgccaacacg toggctcctgt ggcacgacgg acgggcgttg gcgagttgcg agagcgggcc	660
	gttgacatgg gttacgctgc cggagctcga tacggctcggg ttctggagtc tggagggcga	720
	caatggcgag aaggggctgc gagagcaggg tggcatgctc ggctggatga aggagtggac	780
45	gacggcgcac gtgagtcggc tgcttgctt cgacagtcct cagagttgac cgtctctttg	840
	cagccgaagc ggcacccgca caccggcgaa cttatgctct tccacatgct gttcctcccg	900
	ccctacctcc actactcctg catcccttcc actcacgctc ctctccctc cgcccgttcc	960
50	gagaaggccc caccactcc tcgcatctc gccgccccg ttcccatcgc cgcacctcgc	1020
	atgatgcacg atatggcgc ctctcggacg cactcgatcc tcctcgacat gccctctcgc	1080
	ctcgacccgc gcaatctcgc actcggcaag ccogtcatct cctacgacct gtccaaacct	1140
55	gctcgtctcg gcgtcttgc tcgccacaaa ccgacactcg tcaagtggta catcgccccg	1200
60		
65		

ES 2 787 225 T3

gcgtgcataa tctttcacac cgctttcgcg tacgacgagc cccctcaac ttcccacgac 1260  
 gaggtcgagg cagtcaacct cgtctgctgc cgcctcaact ctccccgctt catctactcc 1320  
 5 gccggtaacc tcgtcctccc cgagtctcaa tccctcccag caggcgccaa agagtcctgc 1380  
 gagctctact actaccgctt cccctcttcc tccctccgct cgcgcacct cgaaccgctc 1440  
 cacgccttcc cccctcgccg gattcccttc gagttcccga ccgtaccgca ggaccgcgcg 1500  
 10 gtcggcccggt cgaagtacgt ctatgggtgc tcggtgaagc acgggtcggt cgatgcgggc 1560  
 cttgggtcgg cggcgaagat cgactgcctc gtcaaggtca acgtcgactt gctcgtccgg 1620  
 15 cagggcaaga agcgcaacga ggcggcgag ggacagtcgg agcgaccggt cgacgagcga 1680  
 ggaatcctcg acgtgattgc ccagcagacg ccgagacgag gggaccgcac ctcttcggac 1740  
 gacgacgagc cgatccgat cttcgagctc cctccgctgc actacgcaca ggagagctcg 1800  
 20 ttcgtccctc gccaaaacc cggtcagaa gacgacggct acctcctcac ctacgtcttt 1860  
 gacgaacgcc aactcgaccc ctcgaccggt cgcgctgtcg aaggcgctaa gagcgagttg 1920  
 tgggtcatcg atgctggac gatgaaggac gttgtgtgca gggtaagtt gccgcagagg 1980  
 25 gtgccgtacg ggttgacgg gcattggttc tcgaggggaag agatcgagga gcagcgcagg 2040  
 gctccctcag tccgctcgcg tccctcccca aagtcctag 2079  
 <210> 11  
 <211> 1911  
 <212> ADN  
 <213> Rhodosporidium toruloides  
 <220>  
 <221> CDS  
 <222> (1) (1911)  
 35 <400> 11  
 atg ggg acg aag ccc cgg cac acc tac aag cag ggg aac tac ctc ccc 48  
 Met Gly Thr Lys Pro Arg His Thr Tyr Lys Gln Gly Asn Tyr Leu Pro  
 1 5 10 15  
 40 gtg cga gag gag cgc acg ctc gag gag tgc gct tgc gaa ggt cag ctc 96  
 Val Arg Glu Glu Arg Thr Leu Glu Glu Cys Ala Cys Glu Gly Gln Leu  
 20 25 30  
 45 ccg caa gag ctc gtc ggc ggg atg tac gtg cgg aat ggg ggt gag cct 144  
 Pro Gln Glu Leu Val Gly Gly Met Tyr Val Arg Asn Gly Gly Glu Pro  
 35 40 45  
 50 gcg ctg gca gag ttg cga gga gac gac tcc gca gac gga cct gcc tac 192  
 Ala Leu Ala Glu Leu Arg Gly Asp Asp Ser Ala Asp Gly Pro Ala Tyr  
 50 55 60  
 55 cac tgg ttc gac ggt gac ggc atg ctg acg ggc gtg tac ttc aag ccc 240  
 His Trp Phe Asp Gly Asp Gly Met Leu Thr Gly Val Tyr Phe Lys Pro  
 65 70 75 80  
 60 gtc ccc tcc tcc tcg ccc ttc gac ccg aac ccg cgt atc gcc ccg ctc 288  
 65

ES 2 787 225 T3

	Val	Pro	Ser	Ser	Ser	Pro	Phe	Asp	Pro	Asn	Pro	Arg	Ile	Ala	Pro	Leu	
				85						90					95		
5	ttc	gtc	aac	aaa	tac	gtc	ctc	acc	gac	gtc	ttc	ctc	gcc	tcg	aaa	gcg	336
	Phe	Val	Asn	Lys	Tyr	Val	Leu	Thr	Asp	Val	Phe	Leu	Ala	Ser	Lys	Ala	
			100						105					110			
10	ctc	ggc	gtc	aca	cat	ccc	atc	cta	ccc	tcc	atc	gcc	aca	ctc	ctc	ggc	384
	Leu	Gly	Val	Thr	His	Pro	Ile	Leu	Pro	Ser	Ile	Ala	Thr	Leu	Leu	Gly	
			115					120					125				
15	tca	atc	tgg	tcc	ctc	cac	ctc	atc	ctc	tac	tcc	gtc	ttt	cgc	gcc	ctc	432
	Ser	Ile	Trp	Ser	Leu	His	Leu	Ile	Leu	Tyr	Ser	Val	Phe	Arg	Ala	Leu	
		130				135						140					
20	ttc	ctc	gcc	ttc	tgc	tcc	ttc	ttc	acc	gaa	tcg	cct	ttg	cga	cat	ctc	480
	Phe	Leu	Ala	Phe	Cys	Ser	Phe	Phe	Thr	Glu	Ser	Pro	Leu	Arg	His	Leu	
		145				150					155					160	
25	tca	gtc	gcc	aac	acg	tcg	gtc	ctg	tgg	cac	gac	gga	cgg	gcg	ttg	gcg	528
	Ser	Val	Ala	Asn	Thr	Ser	Val	Leu	Trp	His	Asp	Gly	Arg	Ala	Leu	Ala	
				165						170					175		
30	agt	tgc	gag	agc	ggg	ccg	ttg	aca	tgg	ggt	acg	ctg	ccg	gag	ctc	gat	576
	Ser	Cys	Glu	Ser	Gly	Pro	Leu	Thr	Trp	Val	Thr	Leu	Pro	Glu	Leu	Asp	
				180					185					190			
35	acg	gtc	ggg	ttc	tgg	agt	ctg	gag	ggc	gac	aat	ggc	gag	aag	ggg	ctg	624
	Thr	Val	Gly	Phe	Trp	Ser	Leu	Glu	Gly	Asp	Asn	Gly	Glu	Lys	Gly	Leu	
			195				200					205					
40	cga	gag	cag	ggt	ggc	atg	ctc	ggc	tgg	atg	aag	gag	tgg	acg	acg	gcg	672
	Arg	Glu	Gln	Gly	Gly	Met	Leu	Gly	Trp	Met	Lys	Glu	Trp	Thr	Thr	Ala	
			210				215					220					
45	cat	ccg	aag	cgc	gac	ccg	cac	acc	ggc	gaa	ctt	atg	ctc	ttc	cac	atg	720
	His	Pro	Lys	Arg	Asp	Pro	His	Thr	Gly	Glu	Leu	Met	Leu	Phe	His	Met	
					230						235					240	
50	tcg	ttc	ctc	ccg	ccc	tac	ctc	cac	tac	tcc	gtc	atc	cct	tcc	act	cac	768
	Ser	Phe	Leu	Pro	Pro	Tyr	Leu	His	Tyr	Ser	Val	Ile	Pro	Ser	Thr	His	
					245					250					255		
55	gct	cct	cct	ccc	tcc	gcc	cgt	tcc	gag	aag	gcc	gca	ccc	act	cct	cgc	816
	Ala	Pro	Pro	Pro	Ser	Ala	Arg	Ser	Glu	Lys	Ala	Ala	Pro	Thr	Pro	Arg	
				260				265						270			
60	atc	ctc	gcc	ggc	ccc	gtt	ccc	atc	gcc	gca	cct	cgc	atg	atg	cac	gat	864
	Ile	Leu	Ala	Ala	Pro	Val	Pro	Ile	Ala	Ala	Pro	Arg	Met	Met	His	Asp	
			275				280					285					
65	atg	gcc	gcc	tct	cgg	acg	cac	tcg	atc	ctc	ctc	gac	atg	ccc	ctc	tcg	912
	Met	Ala	Ala	Ser	Arg	Thr	His	Ser	Ile	Leu	Leu	Asp	Met	Pro	Leu	Ser	
			290				295					300					
70	ctc	gac	ccg	cgc	aat	ctc	gca	ctc	ggc	aag	ccc	gtc	atc	tcc	tac	gac	960
	Leu	Asp	Pro	Arg	Asn	Leu	Ala	Leu	Gly	Lys	Pro	Val	Ile	Ser	Tyr	Asp	
					310						315					320	
75	ccg	tcc	aaa	ccc	gct	cgc	ttc	ggc	gtc	ttg	cct	cgc	cac	aaa	ccg	aca	1008
	Pro	Ser	Lys	Pro	Ala	Arg	Phe	Gly	Val	Leu	Pro	Arg	His	Lys	Pro	Thr	
				325						330					335		

ES 2 787 225 T3

ctc gtc aag tgg tac atc gcc ccg gcg tgc ata atc ttt cac acc gct 1056  
 Leu Val Lys Trp Tyr Ile Ala Pro Ala Cys Ile Ile Phe His Thr Ala  
                   340  345  350

5    ttc gcg tac gac gag ccc ccc tca act tcc cac gac gag gtc gag gca 1104  
 Phe Ala Tyr Asp Glu Pro Pro Ser Thr Ser His Asp Glu Val Glu Ala  
                   355  360  365

10    gtc aac ctc gtc tgc tgc cgc ctc aac tct ccc cgc ctc atc tac tcc 1152  
 Val Asn Leu Val Cys Cys Arg Leu Asn Ser Pro Arg Leu Ile Tyr Ser  
                   370  375  380

15    gcc ggt aac ctc gtc ctc ccc gag tct caa tcc ctc cca gca ggc gcc 1200  
 Ala Gly Asn Leu Val Leu Pro Glu Ser Gln Ser Leu Pro Ala Gly Ala  
 385  390  395  400

20    aaa gag tcc tgc gag ctc tac tac tac cgc ttc ccc tct tcc tcc tcc 1248  
 Lys Glu Ser Cys Glu Leu Tyr Tyr Tyr Arg Phe Pro Ser Ser Ser Ser  
   405  410  415

25    gcc tcc gtc acc ctc gaa ccg tgc cac gcc ttc ccc ctc gcc gcg att 1296  
 Ala Ser Val Thr Leu Glu Pro Ser His Ala Phe Pro Leu Ala Ala Ile  
   420  425  430

30    ccc ttc gag ttc ccg acc gta ccg cag gac cgc gcg gtc ggc ccg tgc 1344  
 Pro Phe Glu Phe Pro Thr Val Pro Gln Asp Arg Ala Val Gly Pro Ser  
   435  440  445

35    aag tac gtc tat ggg tgc tgc gtg aag cac ggg tgc ttc gat gcg gcg 1392  
 Lys Tyr Val Tyr Gly Cys Ser Val Lys His Gly Ser Phe Asp Ala Ala  
   450  455  460

40    ctt ggg tgc gcg gcg aag atc gac tgc ctc gtc aag gtc aac gtc gac 1440  
 Leu Gly Ser Ala Ala Lys Ile Asp Cys Leu Val Lys Val Asn Val Asp  
 465  470  475  480

45    ttg ctc gtc cgg cag gcc aag aag cgc aac gag gcg gcc gag gga cag 1488  
 Leu Leu Val Arg Gln Gly Lys Lys Arg Asn Glu Ala Gly Glu Gly Gln  
   485  490  495

50    tcg gag cga ccg gtc gac gag cga gga atc ctc gac gtg att gcc cag 1536  
 Ser Glu Arg Pro Val Asp Glu Arg Gly Ile Leu Asp Val Ile Ala Gln  
   500  505  510

55    cag acg ccg aga cga ggg gac cgc acc tct tgc gac gac gac gag ccg 1584  
 Gln Thr Pro Arg Arg Gly Asp Arg Thr Ser Ser Asp Asp Asp Glu Pro  
   515  520  525

60    atc ccg atc ttc gag ctc cct ccg ctg cac tac gca cag gag agc tgc 1632  
 Ile Arg Ile Phe Glu Leu Pro Pro Leu His Tyr Ala Gln Glu Ser Ser  
   530  535  540

65    ttc gtc cct cgc caa aac ccc ccg tca gaa gac gac gcc tac ctc ctc 1680  
 Phe Val Pro Arg Gln Asn Pro Arg Ser Glu Asp Asp Gly Tyr Leu Leu  
 545  550  555  560

70    acc tac gtc ttt gac gaa cgc caa ctc gac ccc tgc acc ggt cgc gct 1728  
 Thr Tyr Val Phe Asp Glu Arg Gln Leu Asp Pro Ser Thr Gly Arg Ala  
   565  570  575

75    gtc gaa gcc gct aag agc gag ttg tgg gtc atc gat gcg tgg acg atg 1776  
 Val Glu Gly Ala Lys Ser Glu Leu Trp Val Ile Asp Ala Trp Thr Met  
   580  585  590

80    aag gac gtt gtg tgc agg gtc aag ttg ccg cag agg gtg ccg tac ggg 1824  
 Lys Asp Val Val Cys Arg Val Lys Leu Pro Gln Arg Val Pro Tyr Gly  
   595  600  605

85    ttg cac ggg cat tgg ttc tgc agg gaa gag atc gag gag cag cgc agg 1872  
 Leu His Gly His Trp Phe Ser Arg Glu Glu Ile Glu Glu Gln Arg Arg  
   610  615  620

90    gct ccc tca gtc cgc tgc cgt cct ccg cca aag tcc tag 1911  
 Ala Pro Ser Val Arg Ser Arg Pro Pro Pro Lys Ser  
 625  630  635

ES 2 787 225 T3

<210> 12  
 <211> 636  
 <212> PRT  
 <213> Rhodosporidium toruloides  
 <400> 12  
 5 Met Gly Thr Lys Pro Arg His Thr Tyr Lys Gln Gly Asn Tyr Leu Pro  
 1 5 10 15  
 10 Val Arg Glu Glu Arg Thr Leu Glu Glu Cys Ala Cys Glu Gly Gln Leu  
 20 25 30  
 15 Pro Gln Glu Leu Val Gly Gly Met Tyr Val Arg Asn Gly Gly Glu Pro  
 35 40 45  
 20 Ala Leu Ala Glu Leu Arg Gly Asp Asp Ser Ala Asp Gly Pro Ala Tyr  
 50 55 60  
 25 His Trp Phe Asp Gly Asp Gly Met Leu Thr Gly Val Tyr Phe Lys Pro  
 65 70 75 80  
 30 Val Pro Ser Ser Ser Pro Phe Asp Pro Asn Pro Arg Ile Ala Pro Leu  
 85 90 95  
 35 Phe Val Asn Lys Tyr Val Leu Thr Asp Val Phe Leu Ala Ser Lys Ala  
 100 105 110  
 40 Leu Gly Val Thr His Pro Ile Leu Pro Ser Ile Ala Thr Leu Leu Gly  
 115 120 125  
 45 Ser Ile Trp Ser Leu His Leu Ile Leu Tyr Ser Val Phe Arg Ala Leu  
 130 135 140  
 50 Phe Leu Ala Phe Cys Ser Phe Phe Thr Glu Ser Pro Leu Arg His Leu  
 145 150 155 160  
 55 Ser Val Ala Asn Thr Ser Val Leu Trp His Asp Gly Arg Ala Leu Ala  
 165 170 175  
 60  
 65

ES 2 787 225 T3

Ser Cys Glu Ser Gly Pro Leu Thr Trp Val Thr Leu Pro Glu Leu Asp  
 180 185 190  
 5 Thr Val Gly Phe Trp Ser Leu Glu Gly Asp Asn Gly Glu Lys Gly Leu  
 195 200 205  
 10 Arg Glu Gln Gly Gly Met Leu Gly Trp Met Lys Glu Trp Thr Thr Ala  
 210 215 220  
 His Pro Lys Arg Asp Pro His Thr Gly Glu Leu Met Leu Phe His Met  
 225 230 235 240  
 15 Ser Phe Leu Pro Pro Tyr Leu His Tyr Ser Val Ile Pro Ser Thr His  
 245 250 255  
 20 Ala Pro Pro Pro Ser Ala Arg Ser Glu Lys Ala Ala Pro Thr Pro Arg  
 260 265 270  
 25 Ile Leu Ala Ala Pro Val Pro Ile Ala Ala Pro Arg Met Met His Asp  
 275 280 285  
 Met Ala Ala Ser Arg Thr His Ser Ile Leu Leu Asp Met Pro Leu Ser  
 290 295 300  
 30 Leu Asp Pro Arg Asn Leu Ala Leu Gly Lys Pro Val Ile Ser Tyr Asp  
 305 310 315 320  
 35 Pro Ser Lys Pro Ala Arg Phe Gly Val Leu Pro Arg His Lys Pro Thr  
 325 330 335  
 40 Leu Val Lys Trp Tyr Ile Ala Pro Ala Cys Ile Ile Phe His Thr Ala  
 340 345 350  
 Phe Ala Tyr Asp Glu Pro Pro Ser Thr Ser His Asp Glu Val Glu Ala  
 355 360 365  
 45 Val Asn Leu Val Cys Cys Arg Leu Asn Ser Pro Arg Leu Ile Tyr Ser  
 370 375 380  
 50 Ala Gly Asn Leu Val Leu Pro Glu Ser Gln Ser Leu Pro Ala Gly Ala  
 385 390 395 400  
 Lys Glu Ser Cys Glu Leu Tyr Tyr Tyr Arg Phe Pro Ser Ser Ser Ser  
 405 410 415  
 55 Ala Ser Val Thr Leu Glu Pro Ser His Ala Phe Pro Leu Ala Ala Ile  
 60  
 65

ES 2 787 225 T3

				420					425					430			
5	Pro	Phe	Glu	Phe	Pro	Thr	Val	Pro	Gln	Asp	Arg	Ala	Val	Gly	Pro	Ser	
			435					440					445				
10	Lys	Tyr	Val	Tyr	Gly	Cys	Ser	Val	Lys	His	Gly	Ser	Phe	Asp	Ala	Ala	
		450					455					460					
15	Leu	Gly	Ser	Ala	Ala	Lys	Ile	Asp	Cys	Leu	Val	Lys	Val	Asn	Val	Asp	
	465					470					475					480	
20	Ser	Glu	Arg	Pro	Val	Asp	Glu	Arg	Gly	Ile	Leu	Asp	Val	Ile	Ala	Gln	
				500					505					510			
25	Gln	Thr	Pro	Arg	Arg	Gly	Asp	Arg	Thr	Ser	Ser	Asp	Asp	Asp	Glu	Pro	
			515					520						525			
30	Ile	Arg	Ile	Phe	Glu	Leu	Pro	Pro	Leu	His	Tyr	Ala	Gln	Glu	Ser	Ser	
		530					535					540					
35	Phe	Val	Pro	Arg	Gln	Asn	Pro	Arg	Ser	Glu	Asp	Asp	Gly	Tyr	Leu	Leu	
	545					550					555					560	
40	Thr	Tyr	Val	Phe	Asp	Glu	Arg	Gln	Leu	Asp	Pro	Ser	Thr	Gly	Arg	Ala	
					565					570					575		
45	Val	Glu	Gly	Ala	Lys	Ser	Glu	Leu	Trp	Val	Ile	Asp	Ala	Trp	Thr	Met	
				580					585					590			
50	Lys	Asp	Val	Val	Cys	Arg	Val	Lys	Leu	Pro	Gln	Arg	Val	Pro	Tyr	Gly	
			595					600					605				
55	Leu	His	Gly	His	Trp	Phe	Ser	Arg	Glu	Glu	Ile	Glu	Glu	Gln	Arg	Arg	
		610					615					620					
60	Ala	Pro	Ser	Val	Arg	Ser	Arg	Pro	Pro	Pro	Lys	Ser					
	625					630					635						
65	<210>	13															
	<211>	793															
	<212>	ADN															
	<213>	Rhodosporidium toruloides															
	<400>	13															
55	atgctccacc	ctctcgcacg	cctccgcgcc	gcctcgcctcg	tcctgctcgc	aacctcaca											60

ES 2 787 225 T3

ccctcgcac tcggaaacac cgaatcctc gtctcacgct tgccgttoga gcttgcgat 120  
 cttccgaaag ctgcgcacag tgcatcagtg tacgtacacg cccgtacact ccctcttgcg 180  
 5 cggtcgaggg ctcaccaact cttcatcagt tctccccaca tcctcgaact ttattcgccg 240  
 caaacgctct ctatccgcct gactgagggg caggatgttc cgacagagct catcataccg 300  
 ctcgatgcag ctggaaaagc gccgacgggc ttgctcggcc gactggacag gcttgagcgg 360  
 10 gcgatgggac tggaaatgcg aactgtaagg ctgtcgtggc ctgcttcggg gcgtacagct 420  
 tactgtctcc tcttgtgtg aaccagctaa cctcgtagcc gcacagcacc cgaccacctt 480  
 ccacctctcg acccaccgcg ctccctcctc ctgcctgac ccttcctcc ctcacctcct 540  
 15 catctccgcc actccatcct tcgtctccc ctctgtgac caaacctct acgtcccctt 600  
 caccatcctt ctggaaccaa tccacttcgg cggcgtgccg gaatcgacgc tgccgtttgt 660  
 20 tctggtactg gttgggttgg tgggagtgat gggagtgagt ggagtggcgg gagggatggg 720  
 gaggtggttg gaggagctgg cagaggccga ttggaagctt ggtcgggccc gagaagacga 780  
 gaaggcggat tga 793  
 25 <210> 14  
 <211> 675  
 <212> ADN  
 <213> Rhodosporidium toruloides  
 <220>  
 30 <221> CDS  
 <222> (1) (675)  
 <400> 14  
 atg ctc cac oct ctc gca cgc ctc cgc gcc gcc tcg ctc gtc ctg ctc 48  
 Met Leu His Pro Leu Ala Arg Leu Arg Ala Ala Ser Leu Val Leu Leu  
 35 1 5 10 15  
 gca acc ctc aca ccc ctc gca ctc gga aac acc gaa atc ctc gtc tca 96  
 Ala Thr Leu Thr Pro Leu Ala Leu Gly Asn Thr Glu Ile Leu Val Ser  
 20 25 30  
 40 cgc ttg ccg ttc gag ctt gcg gat ctt ccg aaa gct gcg cac agt gca 144  
 Arg Leu Pro Phe Glu Leu Ala Asp Leu Pro Lys Ala Ala His Ser Ala  
 35 40 45  
 45 tca gtt tct ccc cac atc ctc gaa ctt tat tcg ccg caa acg ctc tct 192  
 Ser Val Ser Pro His Ile Leu Glu Leu Tyr Ser Pro Gln Thr Leu Ser  
 50 55 60  
 atc cgc ctg act gag ggt cag gat gtt ccg aca gag ctc atc ata ccg 240  
 Ile Arg Leu Thr Glu Gly Gln Asp Val Pro Thr Glu Leu Ile Ile Pro  
 65 70 75 80  
 50 ctc gat gca gct gga aaa gcg ccg acg ggc ttg ctc ggc cga ctg gac 288  
 Leu Asp Ala Ala Gly Lys Ala Pro Thr Gly Leu Leu Gly Arg Leu Asp  
 85 90 95  
 55 agg ctt gag ccg gcg atg gga ctg gaa atg cga act gta agg ctg tcg 336

60

65

ES 2 787 225 T3

	Arg	Leu	Glu	Arg	Ala	Met	Gly	Leu	Glu	Met	Arg	Thr	Val	Arg	Leu	Ser	
				100					105					110			
5	tgg	cct	gct	tcg	cac	ccg	acc	acc	ttc	cac	ctc	tcg	acc	cac	cgc	gct	384
	Trp	Pro	Ala	Ser	His	Pro	Thr	Thr	Phe	His	Leu	Ser	Thr	His	Arg	Ala	
			115				120					125					
10	ccc	tca	tcc	tcg	cct	gac	cct	tcc	ctc	cct	cac	ctc	ctc	atc	tcc	gcc	432
	Pro	Ser	Ser	Ser	Pro	Asp	Pro	Ser	Leu	Pro	His	Leu	Leu	Ile	Ser	Ala	
			130				135				140						
15	act	cca	tcc	ttc	gtc	tcc	ccc	tcc	tgt	gac	caa	acc	ctc	tac	gtc	ccc	480
	Thr	Pro	Ser	Phe	Val	Ser	Pro	Ser	Cys	Asp	Gln	Thr	Leu	Tyr	Val	Pro	
	145					150					155				160		
20	ttc	acc	atc	ctt	ctc	gaa	cca	atc	cac	ttc	ggc	ggc	gtg	ccg	gaa	tcg	528
	Phe	Thr	Ile	Leu	Leu	Glu	Pro	Ile	His	Phe	Gly	Gly	Val	Pro	Glu	Ser	
				165					170						175		
25	acg	ctg	ccg	ttt	gtt	ctg	gta	ctg	gtt	ggg	ttg	gtg	gga	gtg	atg	gga	576
	Thr	Leu	Pro	Phe	Val	Leu	Val	Leu	Val	Gly	Leu	Val	Gly	Val	Met	Gly	
				180					185					190			
30	gtg	agt	gga	gtg	gcg	gga	ggg	atg	ggg	agg	tgg	ttg	gag	gag	ctg	gca	624
	Val	Ser	Gly	Val	Ala	Gly	Gly	Met	Gly	Arg	Trp	Leu	Glu	Glu	Leu	Ala	
			195				200						205				
35	gag	gcc	gat	tgg	aag	ctt	ggt	cgg	gcg	gga	gaa	gac	gag	aag	gcg	gat	672
	Glu	Ala	Asp	Trp	Lys	Leu	Gly	Arg	Ala	Gly	Glu	Asp	Glu	Lys	Ala	Asp	
		210				215					220						
40	tga																675
	<210>	15															
	<211>	224															
	<212>	PRT															
	<213>	Rhodosporidium toruloides															
45	<400>	15															
	Met	Leu	His	Pro	Leu	Ala	Arg	Leu	Arg	Ala	Ala	Ser	Leu	Val	Leu	Leu	
	1				5					10					15		
50	Ala	Thr	Leu	Thr	Pro	Leu	Ala	Leu	Gly	Asn	Thr	Glu	Ile	Leu	Val	Ser	
				20					25					30			
55	Arg	Leu	Pro	Phe	Glu	Leu	Ala	Asp	Leu	Pro	Lys	Ala	Ala	His	Ser	Ala	
			35					40					45				
60	Ser	Val	Ser	Pro	His	Ile	Leu	Glu	Leu	Tyr	Ser	Pro	Gln	Thr	Leu	Ser	
		50				55					60						
65	Ile	Arg	Leu	Thr	Glu	Gly	Gln	Asp	Val	Pro	Thr	Glu	Leu	Ile	Ile	Pro	
	65					70					75					80	
70	Leu	Asp	Ala	Ala	Gly	Lys	Ala	Pro	Thr	Gly	Leu	Leu	Gly	Arg	Leu	Asp	
					85					90					95		
75																	
80																	
85																	
90																	
95																	

ES 2 787 225 T3

Arg Leu Glu Arg Ala Met Gly Leu Glu Met Arg Thr Val Arg Leu Ser  
 100 105 110  
 5 Trp Pro Ala Ser His Pro Thr Thr Phe His Leu Ser Thr His Arg Ala  
 115 120 125  
 10 Pro Ser Ser Ser Pro Asp Pro Ser Leu Pro His Leu Leu Ile Ser Ala  
 130 135 140  
 15 Thr Pro Ser Phe Val Ser Pro Ser Cys Asp Gln Thr Leu Tyr Val Pro  
 145 150 155 160  
 20 Phe Thr Ile Leu Leu Glu Pro Ile His Phe Gly Gly Val Pro Glu Ser  
 165 170 175  
 25 Val Ser Gly Val Ala Gly Gly Met Gly Arg Trp Leu Glu Glu Leu Ala  
 195 200 205  
 30 Glu Ala Asp Trp Lys Leu Gly Arg Ala Gly Glu Asp Glu Lys Ala Asp  
 210 215 220  
 <210> 16  
 <211> 2565  
 <212> ADN  
 <213> Rhodosporidium toruloides  
 <400> 16  
 cgctcgacga tggccgccat gcaggacact cccatcgaca gcatccccc ggtacatccc 60  
 35 acgctgaccc gcgcttctct cgctcgcttg cgtgctgctc tcgctttcgg agcttcgaaa 120  
 caggcgggac gaggggatcg tgggtcagcg cgcgcctggg gaggattcgc tggacatcgg 180  
 40 ggatattgct ggaggatact caggactttc tggattggtc ccgcccttcc cgcgcccgtc 240  
 gtcgctctcc agcatccatt cccactcga cactccggca aactcgccgc tgactctccc 300  
 cttcgctcat tccgcaggct tacgacaccg tcaccaaggc gttcttgtcg gaaagacta 360  
 45 ggcccatcgc ctggcgcaag gcgcagatca agaaactcgg gttccttgtc gtgagttcgc 420  
 ggttctttcg cgcggtcttt gcgacggctc gtccagtgca ggtcgcgctg gtgcgcgctg 480  
 tcaatgtagt tgcgatgaaa cctgaccag caactcccc tcccacagca agacaacgaa 540  
 50 gacgccttcg tccgcgctct cgagcaagac tttggccgtc ccgctttoga gacaatcaca 600  
 gccgagatca accccgtcaa ggctgagatt aacgaggtct acgaccacct cgagaagtgg 660  
 55 gccaaagccga ggcgcgtcaa gacttcggcg acgtggtacg ctaccaagcc gacggtctac 720  
 60  
 65

ES 2 787 225 T3

	tcggagccta aggggtgcac gtcggtatc gggacgtgga actgtgcgtg cacttgtttt	780
	gtacgaggtt gagtgtgtgt actgacggaa gatgtcgccg cagacccgat cacgctcctc	840
5	ctcgtcccg cttctcggcgc catctctgcc ggctgcaccg cgctcgtcaa ggtgcgttgc	900
	cgttcgaagt ccgtcgtact gcatgtctca ctgacactcg togctcccac agcccgtga	960
	gcaagcccct cacgtcgcg cgctcgtcgc cgacctcctg cccaagtacc tcgaccccac	1020
10	cgccctcatc tgcatacaac gcgccatccc tcaagcgacc gctctcctca aactcaagtt	1080
	cgatcacatc ttctacaccg gttcgggaac ggtcggcaag atcgttgcgc gtgcggcggc	1140
	ggagcacctt tgcccgggta cgcttgagtt ggggggaaag agtccggcgg ttgtgctgga	1200
15	tgatgcggat attgaggtt gggcgaggag gatcgttttg gccaaagtta ccaacgctgg	1260
	gcaggtgcgt cgcgagaacc gggttgtgtc gttggtctcg ccgaagcggg cgcagatgct	1320
	tagtcatccg cttgttgttg cgcacagatc tgcatactca cagactacgt cctcacgacc	1380
20	ccgcagaccg agcccgaagct cctcgaagcc ctcaagcgcg ctctcgcgc cttctccgcc	1440
	aaccccgcgc cctcctcctc ctoggaaaag tcgtcaacct cgctcgtgca caacccgaac	1500
25	tactcgcgca tcatcaacca gaaccactac aaccgcgttt cgaagtgtct tgatgcgact	1560
	aagggcgagg tggttgttgg agggcggagg gacgagaagg agcgcaagat tgaggttacg	1620
	attgtgaggg gcgtcaagcc ggatgactcg ctcatgtcgg gtgcgactg cggtctccc	1680
30	cctgaagaac gaatgtggct gacgaatgcg accgagcaga ggagattttc ggcccctcc	1740
	tcccgatcat gaccctocca acgctcgacg acatgggtcaa gttcatccag tcgcgcgaca	1800
	cccctctcgc gctctacgtc ttcacgcaga gcaagaagaa ccgcgacttc agtgcgttcc	1860
35	ccgcctctct ctcgctcgtc gaccttcgac tgactcgggtg gtggatgcag tctttgagcg	1920
	cactcgtcgc ggaggatctg tccagaacga tgtgtcgtt cagtcatga tccctgggct	1980
	gccgttcggc ggtacgggcg cggcggggtta cggaaactac cacggcaggc ggtgcgtccc	2040
40	gcttcttcgg caccgtgctt ccccgagtcc ggctgacctg ctcgcacgca gcacctcga	2100
	cacgttctcg cacgagcgcg cgtcggccaa tgtccccacc tggatggaca tgatcatggc	2160
45	gtcgcggtac cctccctaca cccgttcgtc cgggtcgtcc ctctctctct gtgctcgtca	2220
	acacactcgt cccacacgtg cagagaagaa gctgaagatg ctctgttcg cgaccaaggc	2280
	ggtgatcaag aagcccagca agtttggtc gatctcgcgc ttgctcaagg tgattgccgc	2340
50	gatggtcgtc ctcttggtg tcagggccag gctctgactg acccgtcgtc gtcccacccc	2400
	tcccctctc caccactcct ctctctctcg ggtcttggga atcgtgtgcg ctgggcagaa	2460
	gttgacggga caggcgtgag gcgggacagg gttgatccaa cccctcagat ctttcaactc	2520
55	ttccacaagt gttgtctgta cctctcctc accctctccc aatca	2565
	<210> 17	
	<211> 1628	
	<212> ADN	
60	<213> Rhodosporidium toruloides	
	<220>	
	<221> CDS	
	<222> (28) (1542)	
	<400> 17	

ES 2 787 225 T3

actcgctcac gcttgctacg ctcgacg atg gcc gcc atg cag gac act ccc atc 54  
Met Ala Ala Met Gln Asp Thr Pro Ile  
1 5

5 gac agc atc ccc cag gct tac gac acc gtc acc aag gcg ttc ttg tcg 102  
Asp Ser Ile Pro Gln Ala Tyr Asp Thr Val Thr Lys Ala Phe Leu Ser  
10 15 20 25

10 gga aag act agg ccc atc gcc tgg cgc aag gcg cag atc aag aaa ctc 150  
Gly Lys Thr Arg Pro Ile Ala Trp Arg Lys Ala Gln Ile Lys Lys Leu  
30 35 40

15 ggg ttc ctt gtc caa gac aac gaa gac gcc ttc gtc cgc gct ctc gag 198  
Gly Phe Leu Val Gln Asp Asn Glu Asp Ala Phe Val Arg Ala Leu Glu  
45 50 55

20 caa gac ttt ggc cgt ccc gct ttc gag aca atc aca gcc gag atc aac 246  
Gln Asp Phe Gly Arg Pro Ala Phe Glu Thr Ile Thr Ala Glu Ile Asn  
60 65 70

25 ccc gtc aag gct gag att aac gag gtc tac gac cac ctc gag aag tgg 294  
Pro Val Lys Ala Glu Ile Asn Glu Val Tyr Asp His Leu Glu Lys Trp  
75 80 85

30 gcc aag ccg agg cgc gtc aag act tcg gcg acg tgg tac gct acc aag 342  
Ala Lys Pro Arg Arg Val Lys Thr Ser Ala Thr Trp Tyr Ala Thr Lys  
90 95 100 105

35 ccg acg gtc tac tcg gag cct aag ggt gtc acg ctc gtt atc ggg acg 390  
Pro Thr Val Tyr Ser Glu Pro Lys Gly Val Thr Leu Val Ile Gly Thr  
110 115 120

40 tgg aac tac ccg atc acg ctc ctc ctc gtc ccg ctt ctc ggc gcc atc 438  
Trp Asn Tyr Pro Ile Thr Leu Leu Leu Val Pro Leu Leu Gly Ala Ile  
125 130 135

45 tct gcc ggc tgc acc gcg ctc gtc aag ccc gct gag caa gcc cct cac 486  
Ser Ala Gly Cys Thr Ala Leu Val Lys Pro Ala Glu Gln Ala Pro His  
140 145 150

50 gtc gcc gcg ctc gtc gcc gac ctc ctg ccc aag tac ctc gac ccc acc 534  
Val Ala Ala Leu Val Ala Asp Leu Leu Pro Lys Tyr Leu Asp Pro Thr  
155 160 165

55 gcc ttc atc tgc atc aac ggc gcc atc cct caa gcg acc gct ctc ctc 582  
Ala Phe Ile Cys Ile Asn Gly Ala Ile Pro Gln Ala Thr Ala Leu Leu  
170 175 180 185

60 aaa ctc aag ttc gat cac atc ttc tac acc ggt tcg gga acg gtc ggc 630  
Lys Leu Lys Phe Asp His Ile Phe Tyr Thr Gly Ser Gly Thr Val Gly  
190 195 200

ES 2 787 225 T3

	aag atc gtt gcg cgt gcg gcg gcg gag cac ctt tgc ccg gtt acg ctt Lys Ile Val Ala Arg Ala Ala Ala Glu His Leu Cys Pro Val Thr Leu 205 210 215	678
5	gag ttg ggg gga aag agt ccg gcg gtt gtg ctg gat gat gcg gat att Glu Leu Gly Gly Lys Ser Pro Ala Val Val Leu Asp Asp Ala Asp Ile 220 225 230	726
10	gag gtt gtg gcg agg agg atc gtt tgg gcc aag ttt acc aac gct ggg Glu Val Val Ala Arg Arg Ile Val Trp Ala Lys Phe Thr Asn Ala Gly 235 240 245	774
15	cag atc tgc atc tcg aca gac tac gtc ctc acg acc ccg cag acc gag Gln Ile Cys Ile Ser Thr Asp Tyr Val Leu Thr Thr Pro Gln Thr Glu 250 255 260 265	822
20	ccc aag ctc ctc gaa gcc ctc aag cgc gct ctc gcc gcc ttc tcc gcc Pro Lys Leu Leu Glu Ala Leu Lys Arg Ala Leu Ala Ala Phe Ser Ala 270 275 280	870
25	aac ccc gcc gcc tcc tcc tcc tcg gaa aag tcg tca acc tcg ctc gtg Asn Pro Ala Ala Ser Ser Ser Ser Glu Lys Ser Ser Thr Ser Leu Val 285 290 295	918
30	cac aac ccg aac tac tcg cgc atc atc aac cag aac cac tac aac cgc His Asn Pro Asn Tyr Ser Arg Ile Ile Asn Gln Asn His Tyr Asn Arg 300 305 310	966
35	gtt tcg aag ttg ctt gat gcg act aag ggc gag gtg gtt gtt gga gcc Val Ser Lys Leu Leu Asp Ala Thr Lys Gly Glu Val Val Val Gly Gly 315 320 325	1014
40	ggg agg gac gag aag gag cgc aag att gag gtt acg att gtg agg ggc Gly Arg Asp Glu Lys Glu Arg Lys Ile Glu Val Thr Ile Val Arg Gly 330 335 340 345	1062
45	gtc aag ccg gat gac tcg ctc atg tcg gag gag att ttc gcc ccc gtc Val Lys Pro Asp Asp Ser Leu Met Ser Glu Glu Ile Phe Gly Pro Val 350 355 360	1110
50	ctc ccg atc atg acc ctc cca acg ctc gac gac atg gtc aag ttc atc Leu Pro Ile Met Thr Leu Pro Thr Leu Asp Asp Met Val Lys Phe Ile 365 370 375	1158
55	cag tcg cgc gac acc cct ctc gcg ctc tac gtc ttc acg cag agc aag Gln Ser Arg Asp Thr Pro Leu Ala Leu Tyr Val Phe Thr Gln Ser Lys 380 385 390	1206
60	aag aac cgc gac ttc atc ttt gag cgc act cgc tcg gga gga ttc gtc Lys Asn Arg Asp Phe Ile Phe Glu Arg Thr Arg Ser Gly Gly Phe Val 395 400 405	1254
65	cag aac gat gtg ctc gtt cag ttc atg atc cct ggg ctg ccg ttc gcc Gln Asn Asp Val Leu Val Gln Phe Met Ile Pro Gly Leu Pro Phe Gly 410 415 420 425	1302
70	ggt acg ggc gcg gcg ggt tac gga aac tac cac gcc agg cgc acc ttc Gly Thr Gly Ala Ala Gly Tyr Gly Asn Tyr His Gly Arg Arg Thr Phe 430 435 440	1350
75	gac acg ttc tcg cac gag cgc gcg tcg gcc aat gtc ccc acc tgg atg Asp Thr Phe Ser His Glu Arg Ala Ser Ala Asn Val Pro Thr Trp Met	1398

ES 2 787 225 T3

	445	450	455	
5	gac atg atc atg gcg tcg cgg tac cct ccc tac acc cag aag aag ctg			1446
	Asp Met Ile Met Ala Ser Arg Tyr Pro Pro Tyr Thr Gln Lys Lys Leu			
	460	465	470	
10	aag atg ctc ctg ttc gcg acc aag gcg gtg atc aag aag ccc agc aag			1494
	Lys Met Leu Leu Phe Ala Thr Lys Ala Val Ile Lys Lys Pro Ser Lys			
	475	480	485	
15	ttt ggc tcg atc tcg cgc ttg ctc aag aag ttg acg gga cag gcg tga			1542
	Phe Gly Ser Ile Ser Arg Leu Leu Lys Lys Leu Thr Gly Gln Ala			
	490	495	500	
20	ggcgggacag ggttgatcca acccctcaga tctttcactc cttccacaag tgttgtctgt			1602
	accctctcct caccctctcc caatca			1628
	<210> 18			
	<211> 504			
	<212> PRT			
	<213> Rhodosporidium toruloides			
	<400> 18			
25	Met Ala Ala Met Gln Asp Thr Pro Ile Asp Ser Ile Pro Gln Ala Tyr			
	1	5	10	15
30	Asp Thr Val Thr Lys Ala Phe Leu Ser Gly Lys Thr Arg Pro Ile Ala			
	20	25	30	
35	Trp Arg Lys Ala Gln Ile Lys Lys Leu Gly Phe Leu Val Gln Asp Asn			
	35	40	45	
40	Glu Asp Ala Phe Val Arg Ala Leu Glu Gln Asp Phe Gly Arg Pro Ala			
	50	55	60	
45	Phe Glu Thr Ile Thr Ala Glu Ile Asn Pro Val Lys Ala Glu Ile Asn			
	65	70	75	80
50	Glu Val Tyr Asp His Leu Glu Lys Trp Ala Lys Pro Arg Arg Val Lys			
	85	90	95	
55	Thr Ser Ala Thr Trp Tyr Ala Thr Lys Pro Thr Val Tyr Ser Glu Pro			
	100	105	110	
60	Lys Gly Val Thr Leu Val Ile Gly Thr Trp Asn Tyr Pro Ile Thr Leu			
	115	120	125	
65	Leu Leu Val Pro Leu Leu Gly Ala Ile Ser Ala Gly Cys Thr Ala Leu			
	130	135	140	
70	Val Lys Pro Ala Glu Gln Ala Pro His Val Ala Ala Leu Val Ala Asp			

ES 2 787 225 T3

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

ES 2 787 225 T3

Glu Arg Thr Arg Ser Gly Gly Phe Val Gln Asn Asp Val Leu Val Gln  
 405 410 415

5 Phe Met Ile Pro Gly Leu Pro Phe Gly Gly Thr Gly Ala Ala Gly Tyr  
 420 425 430

10 Gly Asn Tyr His Gly Arg Arg Thr Phe Asp Thr Phe Ser His Glu Arg  
 435 440 445

15 Ala Ser Ala Asn Val Pro Thr Trp Met Asp Met Ile Met Ala Ser Arg  
 450 455 460

20 Tyr Pro Pro Tyr Thr Gln Lys Lys Leu Lys Met Leu Leu Phe Ala Thr  
 465 470 475 480

Lys Ala Val Ile Lys Lys Pro Ser Lys Phe Gly Ser Ile Ser Arg Leu  
 485 490 495

25 Leu Lys Lys Leu Thr Gly Gln Ala  
 500  
 <210> 19  
 <211> 3136  
 <212> ADN  
 <213> Rhodosporidium toruloides  
 <400> 19

30 cagccccggt ctacctttcg ccgacctcgt cctcagctcc ctgccccatg gacgcgctcg 60  
 ctgccacgat ctgagcaacc acgcatgtcg actctagcca tcagaccoga cgaaccaccg 120  
 aacgagcgac ttcgcgacac cgtccccgac ctgacccgt ctggcggagg agtgcgtccg 180

35 gcgatgtctc ctcgagatcg gtcgcctgcc agctctcgtg gctcgcagga cgacagctcg 240  
 tcgacattgt cggatcgtgg gcgacagcgc aggcgttcgg caagctcoga acggtcggat 300  
 cagcctgaag aagcggagaa gactgoggat ccacctcagg cttcaggtgc gtccggctcg 360

40 ctctcctgac gctcgcgcog ctcccagacg ttggtactga ctgtgtggtc atcgcgtcga 420  
 caggcctggc gcatcccget tctcccagag cacctcatcc gatcgccagg tcctcctccg 480

45 ccaccacgtc tgtcccgcg cttctgtctt ccgtgcgcgc cgatcgctc gaccaccagc 540  
 ttcagaccac gtctccagat tgcccgcgcc catctcgcat gcgcgcgctc gctcgtgtct 600  
 ctcccattc ccgcaacgag cgcccgtgc cgtcgcceat cccttctcc aacttctccg 660

50 cccaccatcg accagcccac ggtctgcgtc gacaatctc cctccctcca tcgagccgcc 720  
 gcgcgagggc gagctcgtg agcccgcctg cgacctcgtc aggtctcgtt tcttctcgt 780  
 accgtcctc ccggcacctg gatcctgacc agcattggct ttcagaggac gaggcgcacg 840

55 acgcccgtcag ggttgcggat cggcccgcgtc ggccacggcg gacgaggtcg agcaggagcg 900

60

65

ES 2 787 225 T3

acgactcggg aggtcccag cgcgacagcc ccaagctcct cgcccgctcg ggccttggcg 960  
 aagacccgac gtcggctcac cacatccaca accgaccctc caaggcgttc agaccgtcgg 1020  
 5 cgacgcccgc gatggcagag ttgacggacg ggcgatcgct gtaccgtgtc gtcccggacc 1080  
 tcgctcgcct cctcctctgc ccgtcctgcc accgcctcta caccgatccc gccacactcg 1140  
 cctgcggtca ctcgcgctgc ctctgtctgt cgggagcgtc agataccctc gccaccccgc 1200  
 10 cgatcgagac ggcgtcgatg acgaccacga ccgccttcac cttcactccg cccgcccgtc 1260  
 cgcggcatac ggccggcatt gcctcgcctc tcccttcgcc ccccgtcgtc gcacacctcg 1320  
 acaacaacca gctgtatcgc acctgtcggc cgacatcgac gacttcgaca ccttcctcgt 1380  
 15 cctcgaccgc tcagacgcac atcccctcgt atcttcccaa tctcacctgc cccgactcgt 1440  
 cgtgcgacta caacttctcg cacctcatcg tccccacct cccgcttcat gtcgactaca 1500  
 20 ccctgcgcaa ggtcagcag atgctgcaga aggctgtgcc tggactggca gcatgggcca 1560  
 cgcgtcttgc gtcgaagaac gaccgcgctc tcgcgcccgt gttctcggat gtcgattgg 1620  
 cgagtgagc gccgaccgat gtcgaggaga acgcccagca gccgatggtc gagggcgacg 1680  
 25 tccagcctgg tgcggtatcg cgcacgagca gcggcagcag tggcggcggc gacgaggaga 1740  
 cgatcgaagg gcacgacgac agcaagcggg cgcacaagtc gcgcaagtcg tggcaggcga 1800  
 gcaagaagac gcgaacggtg tggttcagcg acgtctcgat ggcttcgcct gccgtcggtc 1860  
 30 catctcgac gtcgacgccc acgcccacg ccgcacaacg agacgacaag cacctcgaca 1920  
 tcgctggctt gtcgcccgtc ttctcaccg acctgcacaa cgagtgcgag tgtcaagtgt 1980  
 gcttccagct cttccacgag ccggtcacct cgccttgcgg ccaactcgtc tgccgacaat 2040  
 35 gcctcgcgcc ctctgacgat cactccgaca agtgccctct ctgcccgcgc gacctcccgc 2100  
 ccctcgceta cttccgctgg caacgacca acatcgcgct caccaagatt atcgagaccg 2160  
 40 ctctccctca acaagccgct gagcgcgccc cgaccgtcaa ggaggaagaa cttgcgctgc 2220  
 ttgcgtcggc ccccgtgttc gtctgtacga ccgcatggcc aggcatacaag tgcttcctgc 2280  
 acatcttca gccctcggat cgcctcatgg tccgcccgtc gctcgagacg cccgaacgtt 2340  
 45 cgttcggcat ggtccttcct ctccgacgag ccggacccga cgctgtcaac gattatggca 2400  
 ccatgttgcg cgtcacgagc tgtcagatgc tcgaggacgg ccgcctcctc ctcgagacga 2460  
 tcgggacgta ccgcttcgac ctctcagagc gcagcatggt cgacgggtac aacgtcggca 2520  
 50 aggtcgaacg tgtcagcagc gtctcgcggg aacaagaagc ggagcttgag cgggtcgcgc 2580  
 tcgagaggaa cgataccctc caagacgagt acgcccaccc tgaaccagtc ggaccggaca 2640  
 gcgtgcccct ctcgagaccg cccatgaccg gcaacgtcga gttgtcgacc gaccagctca 2700  
 55 tgcagatctg cctcgacttc ctaccacgc tgcgcccgtc ctgcccgcgc tggattatcg 2760

60

65

ES 2 787 225 T3

agcggctgaa ccggacgggt gcgtctccat ctcaacttgc gcatctcttg ccggtactga 2820  
 ccttcgttct gcagtcggcg aggtgcccaa caaccgcac gacttcagct ggttcgccgc 2880  
 5 cgaggtcttc cccgtcgaag atcacgtcaa ggtcacgctt ctccagtgcg tcgtccgctc 2940  
 ttcccatcc ttcttcgctc cgctgatttc accgctgcgc aggatcacga gtgtccgcga 3000  
 acgactgcgc ttgatcgtct tctggatcga gcagttcagg tcgtcgtggt ggtactccc 3060  
 10 aggctgtgcg ttccgtttcc tcttctccg cctgacgttg ctgacgctc tccgatgcac 3120  
 aggcaacatc gcctag 3136  
 <210> 20  
 15 <211> 2889  
 <212> ADN  
 <213> Rhodosporidium toruloides  
 <220>  
 <221> CDS  
 20 <222> (85) (2889)  
 <400> 20  
 cagccccggt ctacctttcg ccgacctcgt cctcagctcc ctgccccatg gacgcgctcg 60  
 ctgccacgat ctgagcaacc acgc atg tcg act cta gcc atc aga ccc gac 111  
 25 Met Ser Thr Leu Ala Ile Arg Pro Asp  
 1 5  
 gaa cca ccg aac gag cga ctt cgc gac acc gtc ccc gac ctc gac ccg 159  
 Glu Pro Pro Asn Glu Arg Leu Arg Asp Thr Val Pro Asp Leu Asp Pro  
 10 15 20 25  
 30 tct ggc gga gga gtg cgt ccg gcg atg tct cct cga gat cgg tcg cct 207  
 Ser Gly Gly Gly Val Arg Pro Ala Met Ser Pro Arg Asp Arg Ser Pro  
 30 35 40  
 35 gcc agc tct cgt ggc tcg cag gac gac agc tcg tcg aca ttg tcg gat 255  
 Ala Ser Ser Arg Gly Ser Gln Asp Asp Ser Ser Ser Thr Leu Ser Asp  
 45 50 55  
 40 cgt ggg cga cag cgc agg cgt tcg gca agc tcg gaa cgg tcg gat cag 303  
 Arg Gly Arg Gln Arg Arg Arg Ser Ala Ser Ser Glu Arg Ser Asp Gln  
 60 65 70  
 45 cct gaa gaa gcg gag aag act gcg gat cca cct cag gct tca ggc ctg 351  
 Pro Glu Glu Ala Glu Lys Thr Ala Asp Pro Pro Gln Ala Ser Gly Leu  
 75 80 85  
 50 gcg cat ccc gct tct ccc gag gca cct cat ccg atc gcc agg tcc tcc 399  
 Ala His Pro Ala Ser Pro Glu Ala Pro His Pro Ile Ala Arg Ser Ser  
 90 95 100 105  
 55 tcc gcc acc acg tct gtc ccc gcg cct tct gct tcc gtg cgc gcc gat 447  
 Ser Ala Thr Thr Ser Val Pro Ala Pro Ser Ala Ser Val Arg Ala Asp  
 110 115 120  
 cgc ctc gac cac cag ctt cag acc acg tct cca gat tgc ccc gcc cca 495  
 Arg Leu Asp His Gln Leu Gln Thr Thr Ser Pro Asp Cys Pro Ala Pro  
 125 130 135

ES 2 787 225 T3

tct cgc atg cgc cgc gtc gct cgt gtc tct ccc aat tcc cgc aac gag 543  
 Ser Arg Met Arg Arg Val Ala Arg Val Ser Pro Asn Ser Arg Asn Glu  
 140 145 150

5 cgc ccc gtg ccg tcc gcc atc cct tcc tcc aac ttc tcc gcc cac cat 591  
 Arg Pro Val Pro Ser Ala Ile Pro Ser Ser Asn Phe Ser Ala His His  
 155 160 165

10 cga cca gcc cac ggt ctg cgt cga caa tcc tcc ctc cct cca tcg agc 639  
 Arg Pro Ala His Gly Leu Arg Arg Gln Ser Ser Leu Pro Pro Ser Ser  
 170 175 180 185

cgc cgc gcg agg gcg agc tcg ctg agc ccg ccg tcg acc tcg tca ggt 687  
 Arg Arg Ala Arg Ala Ser Ser Leu Ser Pro Pro Ser Thr Ser Ser Gly  
 190 195 200

15 ctc gct tct tcc tcg tac cgt ccg tcc cgg cac ctg gat cct gac cag 735  
 Leu Ala Ser Ser Ser Tyr Arg Pro Ser Arg His Leu Asp Pro Asp Gln  
 205 210 215

20 cat tgg ctt tca gag gac gag gcg cac gac gcc gtc agg gtt gcg gat 783  
 His Trp Leu Ser Glu Asp Glu Ala His Asp Ala Val Arg Val Ala Asp  
 220 225 230

cgg ccg cgt cgg cca cgg ccg acg agg tcg agc agg agc gac gac tcg 831  
 Arg Pro Arg Arg Pro Arg Arg Thr Arg Ser Ser Arg Ser Asp Asp Ser  
 235 240 245

25 gaa ggt ccc gag cgc gac agc ccc aag ctc ctc gcc gcg tcg ggc ctt 879  
 Glu Gly Pro Glu Arg Asp Ser Pro Lys Leu Leu Ala Ala Ser Gly Leu  
 250 255 260 265

30 ggc gaa gac ccg acg tcg gct cac cac atc cac aac cga ccc tcc aag 927  
 Gly Glu Asp Pro Thr Ser Ala His His Ile His Asn Arg Pro Ser Lys  
 270 275 280

gcg ttc aga ccg tcg gcg acg ccg ccg atg gca gag ttg acg gac ggg 975  
 Ala Phe Arg Pro Ser Ala Thr Pro Pro Met Ala Glu Leu Thr Asp Gly  
 285 290 295

cga tcg ctg tac gct gtc gtc ccg gac ctc gct cgc ctc ctc ctc tgc 1023  
 Arg Ser Leu Tyr Ala Val Val Pro Asp Leu Ala Arg Leu Leu Leu Cys  
 300 305 310

40 ccg tcc tgc cac cgc ctc tac acc gat ccc gcc aca ctc gcc tgc ggt 1071  
 Pro Ser Cys His Arg Leu Tyr Thr Asp Pro Ala Thr Leu Ala Cys Gly  
 315 320 325

cac tcg cgc tgc ctc gtc tgc tcg gga gcg tca gat acc ctc gcc acc 1119  
 His Ser Arg Cys Leu Val Cys Ser Gly Ala Ser Asp Thr Leu Ala Thr  
 330 335 340 345

ccg ccg atc gag acg gcg tcg atg acg acc acg acc gcc ttc acc ttc 1167  
 Pro Pro Ile Glu Thr Ala Ser Met Thr Thr Thr Thr Ala Phe Thr Phe  
 350 355 360

50 act ccg ccc gcc gtc gcg ccg cat acg gcc ggc atc gcc tcg cct ctc 1215  
 Thr Pro Pro Ala Val Ala Arg His Thr Ala Gly Ile Ala Ser Pro Leu  
 365 370 375

55 cct tcg ccc ccc gtc gtc gca cac ctc gac aac aac cag ctg tat cgc 1263  
 Pro Ser Pro Pro Val Val Ala His Leu Asp Asn Asn Gln Leu Tyr Arg

ES 2 787 225 T3

	380		385		390		
5	acc ttg tcg gcg aca tcg acg act tcg aca cct tcc gtc gcc tcg acc					1311	
	Thr Leu Ser Ala Thr Ser Thr Thr Ser Thr Pro Ser Val Ala Ser Thr						
	395		400		405		
	gct cag acg cac atc ccc gtc gat ctt ccc aat ctc acc tgc ccc gac					1359	
	Ala Gln Thr His Ile Pro Val Asp Leu Pro Asn Leu Thr Cys Pro Asp						
	410		415		420		425
10	tcg tcg tgc gac tac aac ttc tcg cac ctc atc gtc ccc cac ctc ccg					1407	
	Ser Ser Cys Asp Tyr Asn Phe Ser His Leu Ile Val Pro His Leu Pro						
			430		435		440
15	ctt cat gtc gac tac acc ctg cgc aag gtc agc gag atg ctg cag aag					1455	
	Leu His Val Asp Tyr Thr Leu Arg Lys Val Ser Glu Met Leu Gln Lys						
			445		450		455
	gct gtg cct gga ctg gca gca tgg gcg acg cgt ctt gcg tcg aag aac					1503	
	Ala Val Pro Gly Leu Ala Ala Trp Ala Thr Arg Leu Ala Ser Lys Asn						
			460		465		470
20	gac ccg cat ctc gcg ccg ttg ttc tcg gat gtg cga ttg gcg agt gag					1551	
	Asp Pro His Leu Ala Pro Leu Phe Ser Asp Val Arg Leu Ala Ser Glu						
			475		480		485
25	gcg ccg acc gat gtc gag gag aac ggc gag cag ccg atg gtc gag ggc					1599	
	Ala Pro Thr Asp Val Glu Glu Asn Gly Glu Gln Pro Met Val Glu Gly						
			490		495		500
	gac gtc cag cct ggt gcg gta tcg cgc acg agc agc ggc agc agt ggc					1647	
	Asp Val Gln Pro Gly Ala Val Ser Arg Thr Ser Ser Gly Ser Ser Gly						
			510		515		520
30	ggc ggc gac gag gag acg atc gaa ggg cac gac gac agc aag ccg gcg					1695	
	Gly Gly Asp Glu Glu Thr Ile Glu Gly His Asp Asp Ser Lys Arg Ala						
			525		530		535
35	cac aag tcg cgc aag tcg tgg cag gcg agc aag aag acg cga acg gtg					1743	
	His Lys Ser Arg Lys Ser Trp Gln Ala Ser Lys Lys Thr Arg Thr Val						
			540		545		550
40	tcg ttc agc gac gtc tcg atg gct tcg cct gcc gtc ggt cca tct cgc					1791	
	Ser Phe Ser Asp Val Ser Met Ala Ser Pro Ala Val Gly Pro Ser Arg						
			555		560		565
	acg tcg acg ccg acg ccg cac gcc gca caa cga gac gac aag cac ctc					1839	
	Thr Ser Thr Pro Thr Pro His Ala Ala Gln Arg Asp Asp Lys His Leu						
			570		575		580
45	gac atc gct ggc ttg tcg ccg tcg ttc ctc acc gac ctg cac aac gag					1887	
	Asp Ile Ala Gly Leu Ser Pro Ser Phe Leu Thr Asp Leu His Asn Glu						
			590		595		600
50	tgc gag tgt caa gtg tgc ttc cag ctc ttc cac gag ccg gtc acc tcg					1935	
	Cys Glu Cys Gln Val Cys Phe Gln Leu Phe His Glu Pro Val Thr Ser						
			605		610		615
	cct tgc ggc cac tcg ttc tgc cga caa tgc ctc gcc cgc tcg tac gat					1983	
	Pro Cys Gly His Ser Phe Cys Arg Gln Cys Leu Ala Arg Ser Tyr Asp						
			620		625		630
55	cac tcc gac aag tgc cct ctc tgc cgc gcc gac ctc ccg ccc ctc gcc					2031	

60

65

ES 2 787 225 T3

	His	Ser	Asp	Lys	Cys	Pro	Leu	Cys	Arg	Ala	Asp	Leu	Pro	Pro	Leu	Ala	
	635						640					645					
5	tac	ttc	cgc	tgg	caa	cga	ccc	aac	atc	gcg	ctc	acc	aag	att	atc	gag	2079
	Tyr	Phe	Arg	Trp	Gln	Arg	Pro	Asn	Ile	Ala	Leu	Thr	Lys	Ile	Ile	Glu	
	650					655					660					665	
10	acc	gct	ctc	cct	caa	caa	gcc	gct	gag	gcg	gcc	gcg	acc	gtc	aag	gag	2127
	Thr	Ala	Leu	Pro	Gln	Gln	Ala	Ala	Glu	Arg	Ala	Ala	Thr	Val	Lys	Glu	
					670						675				680		
15	gaa	gaa	ctt	gcg	ctg	ctt	gcg	tcg	gtc	ccc	gtg	ttc	gtc	tgt	acg	acc	2175
	Glu	Glu	Leu	Ala	Leu	Leu	Ala	Ser	Val	Pro	Val	Phe	Val	Cys	Thr	Thr	
				685					690					695			
20	gca	tgg	cca	ggc	atc	aag	tgc	ttc	ctg	cac	atc	ttc	gag	cct	cgg	tat	2223
	Ala	Trp	Pro	Gly	Ile	Lys	Cys	Phe	Leu	His	Ile	Phe	Glu	Pro	Arg	Tyr	
			700					705					710				
25	cgc	ctc	atg	gtc	cgc	cgc	gtg	ctc	gag	acg	ccc	gaa	cgt	tcg	ttc	ggc	2271
	Arg	Leu	Met	Val	Arg	Arg	Val	Leu	Glu	Thr	Pro	Glu	Arg	Ser	Phe	Gly	
			715				720						725				
30	atg	gtc	ctt	cct	ctc	cgc	agc	gcc	gga	ccc	gac	gct	gtc	aac	gag	tat	2319
	Met	Val	Leu	Pro	Leu	Arg	Ser	Ala	Gly	Pro	Asp	Ala	Val	Asn	Glu	Tyr	
						735					740					745	
35	ggc	acc	atg	ttg	cgc	gtc	acg	agc	tgt	cag	atg	ctc	gag	gac	ggc	cgc	2367
	Gly	Thr	Met	Leu	Arg	Val	Thr	Ser	Cys	Gln	Met	Leu	Glu	Asp	Gly	Arg	
					750					755					760		
40	ctc	atc	ctc	gag	acg	atc	ggg	acg	tac	cgc	ttc	cgc	ctc	ctc	gag	cgc	2415
	Leu	Ile	Leu	Glu	Thr	Ile	Gly	Thr	Tyr	Arg	Phe	Arg	Leu	Leu	Glu	Arg	
				765					770						775		
45	agc	atg	gtc	gac	ggg	tac	aac	gtc	ggc	aag	gtc	gaa	cgt	gtc	gac	gac	2463
	Ser	Met	Val	Asp	Gly	Tyr	Asn	Val	Gly	Lys	Val	Glu	Arg	Val	Asp	Asp	
			780					785					790				
50	gtc	tcg	ccg	gaa	caa	gaa	gcg	gag	ctt	gag	cgg	gtc	gcg	ctc	gcg	agg	2511
	Val	Ser	Pro	Glu	Gln	Glu	Ala	Glu	Leu	Glu	Arg	Val	Ala	Leu	Ala	Arg	
			795				800						805				
55	aac	gat	acc	ctc	caa	gac	gag	tac	gcc	gac	cct	gaa	cca	gtc	gga	ccg	2559
	Asn	Asp	Thr	Leu	Gln	Asp	Glu	Tyr	Ala	Asp	Pro	Glu	Pro	Val	Gly	Pro	
					815						820					825	
60	gac	agc	gtg	ccc	ctc	tcg	aga	ccg	ccc	atg	acc	ggc	aac	gtc	gag	ttg	2607
	Asp	Ser	Val	Pro	Leu	Ser	Arg	Pro	Pro	Met	Thr	Gly	Asn	Val	Glu	Leu	
					830					835					840		
65	tcg	acc	gac	cag	ctc	atg	cag	atc	tgc	ctc	gac	ttc	ctc	acc	acg	ctg	2655
	Ser	Thr	Asp	Gln	Leu	Met	Gln	Ile	Cys	Leu	Asp	Phe	Leu	Thr	Thr	Leu	
				845					850					855			
70	cgc	gct	agc	tcg	gcg	ccg	tgg	att	atc	gag	cgg	ctg	aac	cgg	acg	gtc	2703
	Arg	Ala	Ser	Ser	Ala	Pro	Trp	Ile	Ile	Glu	Arg	Leu	Asn	Arg	Thr	Val	
				860				865						870			
75	ggc	gag	gtg	ccc	aac	aac	ccg	cac	gac	ttc	agc	tgg	ttc	gcc	gcc	gag	2751
	Gly	Glu	Val	Pro	Asn	Asn	Pro	His	Asp	Phe	Ser	Trp	Phe	Ala	Ala	Glu	
				875			880					885					
80	gtc	ttc	ccc	gtc	gaa	gat	cac	gtc	aag	gtc	acg	ctt	ctc	cag	atc	acg	2799
	Val	Phe	Pro	Val	Glu	Asp	His	Val	Lys	Val	Thr	Leu	Leu	Gln	Ile	Thr	
					895					900					905		
85	agt	gtc	cgc	gaa	cga	ctg	cgc	ttg	atc	gtc	ttc	tgg	atc	gag	cag	ttc	2847
	Ser	Val	Arg	Glu	Arg	Leu	Arg	Leu	Ile	Val	Phe	Trp	Ile	Glu	Gln	Phe	
					910					915					920		
90	agg	tcg	tcg	tgg	tgg	tac	tcc	cga	ggc	tgc	aac	atc	gcc	tag			2889
	Arg	Ser	Ser	Trp	Trp	Tyr	Ser	Arg	Gly	Cys	Asn	Ile	Ala				
				925					930								

ES 2 787 225 T3

<210> 21  
 <211> 934  
 <212> PRT  
 <213> Rhodosporidium toruloides  
 <400> 21  
 5 Met Ser Thr Leu Ala Ile Arg Pro Asp Glu Pro Pro Asn Glu Arg Leu  
 1 5 10 15  
 10 Arg Asp Thr Val Pro Asp Leu Asp Pro Ser Gly Gly Gly Val Arg Pro  
 20 25 30  
 15 Ala Met Ser Pro Arg Asp Arg Ser Pro Ala Ser Ser Arg Gly Ser Gln  
 35 40 45  
 20 Asp Asp Ser Ser Ser Thr Leu Ser Asp Arg Gly Arg Gln Arg Arg Arg  
 50 55 60  
 25 Ser Ala Ser Ser Glu Arg Ser Asp Gln Pro Glu Glu Ala Glu Lys Thr  
 65 70 75 80  
 30 Ala Asp Pro Pro Gln Ala Ser Gly Leu Ala His Pro Ala Ser Pro Glu  
 85 90 95  
 35 Ala Pro His Pro Ile Ala Arg Ser Ser Ser Ala Thr Thr Ser Val Pro  
 100 105 110  
 40 Ala Pro Ser Ala Ser Val Arg Ala Asp Arg Leu Asp His Gln Leu Gln  
 115 120 125  
 45 Thr Thr Ser Pro Asp Cys Pro Ala Pro Ser Arg Met Arg Arg Val Ala  
 130 135 140  
 50 Arg Val Ser Pro Asn Ser Arg Asn Glu Arg Pro Val Pro Ser Ala Ile  
 145 150 155 160  
 55 Pro Ser Ser Asn Phe Ser Ala His His Arg Pro Ala His Gly Leu Arg  
 165 170 175  
 60  
 65

ES 2 787 225 T3

Arg Gln Ser Ser Leu Pro Pro Ser Ser Arg Arg Ala Arg Ala Ser Ser  
 180 185 190  
 5 Leu Ser Pro Pro Ser Thr Ser Ser Gly Leu Ala Ser Ser Ser Tyr Arg  
 195 200 205  
 10 Pro Ser Arg His Leu Asp Pro Asp Gln His Trp Leu Ser Glu Asp Glu  
 210 215 220  
 Ala His Asp Ala Val Arg Val Ala Asp Arg Pro Arg Arg Pro Arg Arg  
 225 230 235 240  
 15 Thr Arg Ser Ser Arg Ser Asp Asp Ser Glu Gly Pro Glu Arg Asp Ser  
 245 250 255  
 20 Pro Lys Leu Leu Ala Ala Ser Gly Leu Gly Glu Asp Pro Thr Ser Ala  
 260 265 270  
 His His Ile His Asn Arg Pro Ser Lys Ala Phe Arg Pro Ser Ala Thr  
 275 280 285  
 25 Pro Pro Met Ala Glu Leu Thr Asp Gly Arg Ser Leu Tyr Ala Val Val  
 290 295 300  
 30 Pro Asp Leu Ala Arg Leu Leu Leu Cys Pro Ser Cys His Arg Leu Tyr  
 305 310 315 320  
 35 Thr Asp Pro Ala Thr Leu Ala Cys Gly His Ser Arg Cys Leu Val Cys  
 325 330 335  
 Ser Gly Ala Ser Asp Thr Leu Ala Thr Pro Pro Ile Glu Thr Ala Ser  
 340 345 350  
 40 Met Thr Thr Thr Thr Ala Phe Thr Phe Thr Pro Pro Ala Val Ala Arg  
 355 360 365  
 45 His Thr Ala Gly Ile Ala Ser Pro Leu Pro Ser Pro Pro Val Val Ala  
 370 375 380  
 His Leu Asp Asn Asn Gln Leu Tyr Arg Thr Leu Ser Ala Thr Ser Thr  
 385 390 395 400  
 50 Thr Ser Thr Pro Ser Val Ala Ser Thr Ala Gln Thr His Ile Pro Val  
 405 410 415  
 55 Asp Leu Pro Asn Leu Thr Cys Pro Asp Ser Ser Cys Asp Tyr Asn Phe  
 420 425 430

60

65

ES 2 787 225 T3

Ser His Leu Ile Val Pro His Leu Pro Leu His Val Asp Tyr Thr Leu  
 435 440 445  
 5 Arg Lys Val Ser Glu Met Leu Gln Lys Ala Val Pro Gly Leu Ala Ala  
 450 455 460  
 10 Trp Ala Thr Arg Leu Ala Ser Lys Asn Asp Pro His Leu Ala Pro Leu  
 465 470 475 480  
 Phe Ser Asp Val Arg Leu Ala Ser Glu Ala Pro Thr Asp Val Glu Glu  
 485 490 495  
 15 Asn Gly Glu Gln Pro Met Val Glu Gly Asp Val Gln Pro Gly Ala Val  
 500 505 510  
 20 Ser Arg Thr Ser Ser Gly Ser Ser Gly Gly Gly Asp Glu Glu Thr Ile  
 515 520 525  
 Glu Gly His Asp Asp Ser Lys Arg Ala His Lys Ser Arg Lys Ser Trp  
 530 535 540  
 25 Gln Ala Ser Lys Lys Thr Arg Thr Val Ser Phe Ser Asp Val Ser Met  
 545 550 555 560  
 30 Ala Ser Pro Ala Val Gly Pro Ser Arg Thr Ser Thr Pro Thr Pro His  
 565 570 575  
 35 Ala Ala Gln Arg Asp Asp Lys His Leu Asp Ile Ala Gly Leu Ser Pro  
 580 585 590  
 Ser Phe Leu Thr Asp Leu His Asn Glu Cys Glu Cys Gln Val Cys Phe  
 595 600 605  
 40 Gln Leu Phe His Glu Pro Val Thr Ser Pro Cys Gly His Ser Phe Cys  
 610 615 620  
 45 Arg Gln Cys Leu Ala Arg Ser Tyr Asp His Ser Asp Lys Cys Pro Leu  
 625 630 635 640  
 Cys Arg Ala Asp Leu Pro Pro Leu Ala Tyr Phe Arg Trp Gln Arg Pro  
 645 650 655  
 50 Asn Ile Ala Leu Thr Lys Ile Ile Glu Thr Ala Leu Pro Gln Gln Ala  
 660 665 670  
 55 Ala Glu Arg Ala Ala Thr Val Lys Glu Glu Glu Leu Ala Leu Leu Ala  
 60  
 65

ES 2 787 225 T3

675                      680                      685

5      Ser Val Pro Val Phe Val Cys Thr Thr Ala Trp Pro Gly Ile Lys Cys  
           690                                      695                                      700

10      Phe Leu His Ile Phe Glu Pro Arg Tyr Arg Leu Met Val Arg Arg Val  
           705                                      710                                      715                                      720

15      Leu Glu Thr Pro Glu Arg Ser Phe Gly Met Val Leu Pro Leu Arg Ser  
                                     725                                      730                                      735

20      Ala Gly Pro Asp Ala Val Asn Glu Tyr Gly Thr Met Leu Arg Val Thr  
                                     740                                      745                                      750

25      Ser Cys Gln Met Leu Glu Asp Gly Arg Leu Ile Leu Glu Thr Ile Gly  
                                     755                                      760                                      765

30      Thr Tyr Arg Phe Arg Leu Leu Glu Arg Ser Met Val Asp Gly Tyr Asn  
           770                                      775                                      780

35      Val Gly Lys Val Glu Arg Val Asp Asp Val Ser Pro Glu Gln Glu Ala  
           785                                      790                                      795                                      800

40      Glu Leu Glu Arg Val Ala Leu Ala Arg Asn Asp Thr Leu Gln Asp Glu  
                                     805                                      810                                      815

45      Tyr Ala Asp Pro Glu Pro Val Gly Pro Asp Ser Val Pro Leu Ser Arg  
                                     820                                      825                                      830

50      Pro Pro Met Thr Gly Asn Val Glu Leu Ser Thr Asp Gln Leu Met Gln  
                                     835                                      840                                      845

55      Ile Cys Leu Asp Phe Leu Thr Thr Leu Arg Ala Ser Ser Ala Pro Trp  
           850                                      855                                      860

60      Ile Ile Glu Arg Leu Asn Arg Thr Val Gly Glu Val Pro Asn Asn Pro  
           865                                      870                                      875                                      880

65      His Asp Phe Ser Trp Phe Ala Ala Glu Val Phe Pro Val Glu Asp His  
                                     885                                      890                                      895

70      Val Lys Val Thr Leu Leu Gln Ile Thr Ser Val Arg Glu Arg Leu Arg  
                                     900                                      905                                      910

75      Leu Ile Val Phe Trp Ile Glu Gln Phe Arg Ser Ser Trp Trp Tyr Ser  
           915                                      920                                      925

80      Arg Gly Cys Asn Ile Ala  
           930

85      <210> 22  
           <211> 2357  
           <212> ADN  
           <213> Rhodosporidium toruloides  
           <400> 22

ES 2 787 225 T3

tcctctggcc tctctcgtct cgttcgcctg gcgcgcgcc caaacccgctt tggtgacgct 60  
 atcagcctcc cgcgagcgtc agcaggctct cttagccgcc tcacgcctac ctccggaatg 120  
 5 tcctccccac ctccgttcgc gcccggttcg acctcgcca cgcagtcgag gcccggtgcg 180  
 acgcacagca gcccaacaca aggtcccttc gcaacgcctg cgcgcccagg tgcgagctca 240  
 gcggcgcacc aaccgatcc ttcaacctcg acagcaggcg ccgcctccct cctcacgtct 300  
 10 tctcctcact acacgacttc gctgcgctct cgacactcgc tttacggcac ggaagaccgt 360  
 gttgtgctgg atctcggctc gcggatatgg aaggctgggt tcagcgggga gccgcagccg 420  
 cgcgagtgcc ggagcgtcgt gacgcagttg gcgcatgagc gggctgggcg aagagcaggg 480  
 15 ccgtccatag gcgcgagagg ggacgacgac gaagaggact gcttttgggc gctcgagaag 540  
 gccgagccga gcgaggagga gtggttgatt cgcgaggaga ggggtaagcg attactgcgc 600  
 20 aagatctggt tcgagtgagt tcaaccggtc tttcttttga taccaggact tgcgctcgcc 660  
 aacctgggcg gcgctttctg ctaagaggct cgagtacgaa ccggacaagc ggagtctttt 720  
 ccgccggctg acttcgttct cccgcagaaa cctcatgac gaccogaaga cgcgcaaagt 780  
 25 catcgtcgtg gagaaccac tgctgtcgac gcgcgtgaag gagatgatcg cgcgggtctt 840  
 gttogacaac ttgcaggtaa gccatccgcc gcttcgccag gccgttcacc cgtactgacc 900  
 gcttcgctac gcagatcccg tcgctcagct tcgcttccgc cccttgcctc gccttgatgg 960  
 30 cggccggcac agtgaccggt ctgtggctcg acgtcgaaa cctcgagacg accgttcttc 1020  
 ccgtgagttc gccttccgct ctgcctgct tactaaactt gcgctgatgt agagctaaa 1080  
 35 caggtctttc acgctcgcgc gctctttccc tccctacca caactcctcg cgcgggctct 1140  
 cgctgaacc gccgcctccg ctctctctc ctgcattcg gctcatacgc acctcctccc 1200  
 tcctccctca actccatgac gcctcccgct atcggacgga taccgaagga gctcttgacg 1260  
 40 gaggagctca tagaggagat caagacgcgg ctgtgttttg tcggtgagga ggtcccgctc 1320  
 gatgcgagcc ggggcgagag ggaggcgtcc gcgttcagcg gatcggccat gagcgtcgac 1380  
 actgcatcag cagcgcacaa ctttgacgac ccttccgacc ccgacaacgc actactaaag 1440  
 45 gagctttact cccgcttcgc cgcgacatcg accgccaaac ccgtttcctt ccggataccc 1500  
 aacttgtccc aacctgcgat cgcgaacggt acaggctcag ggtggatcca ggtcccggc 1560  
 50 tggatcaggg agcgcgcgcg cgaggtggtg tgggaggagg atggtgatgg agacgagcgt 1620

55

60

65

ES 2 787 225 T3

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

```

gggcttgccg ccgttgtcct tgactgccta ttgaaggtgc gcaagctgac cacttcgctc 1680
gagcgaagta gctgacactg cacccgttcc cccagctccc ccttgacctt cgaagaccga 1740
tggcctcgtc aatcctcctc acggggggca ccgccatgct ccccggett cttcccacgct 1800
tcaaggccgc tcttcttgcc cagctcgacc gctcgcaccc tccttcccct ccgccttccc 1860
cgcctctgcc agcagcatcc gtcgaacctc cttcttccga ccctgcccga cccatgtcga 1920
cggacacggc ggcacgcctt gcgccctctc gctcgcagca agtcaatgcg aagcggcggc 1980
gaaagcatgc cctcgcgact cgactgcaca acctgcccga ttcgcctcga tacgccccgc 2040
ttgtccctct cgctcgacac ctgcacatct tgaaccacc atctccgaac tcctcggcat 2100
cgtcgcagcc gccctcatca accctcgtc gacagcgcga aggttcggct ccgtctttct 2160
cgccccctt gcaaagctgg atcggcggca gcctcgcagg agcgcctcaag acgggggggc 2220
ctgagattgc gagggagcag tgggatgcag ggttgccggtt tgccgaggca gaagaggcgg 2280
agggagagga aggagaggag tttgaggagc gggaagtgat ccggcccggct ctgccggact 2340
ggacgaggat cgcgtag 2357
<210> 23
<211> 2046
<212> ADN
<213> Rhodosporidium toruloides
<220>
<221> CDS
<222> (118) (2046)
<400> 23
tcctctggcc tctctcgtc cgttcgcctg gcgcgcgcc caaaccgctt tgttgacgct 60
atcagcctcc cgcgagcgtt agcaggctct cttagccgcc tcacgcctac ctgcgga 117
atg tcc tcc cca cct cgc ttc gcg ccc ggt tcg acc tcg ccc acg cag 165
Met Ser Ser Pro Pro Arg Phe Ala Pro Gly Ser Thr Ser Pro Thr Gln
1 5 10 15
tcg agg ccg cgt gcg acg cac agc agc cca aca caa ggt ccc ttc gca 213
Ser Arg Pro Arg Ala Thr His Ser Ser Pro Thr Gln Gly Pro Phe Ala
20 25 30
acg cct gcg cgc cca ggt gcg agc tca gcg gcg cac caa ccc gat cct 261
Thr Pro Ala Arg Pro Gly Ala Ser Ser Ala Ala His Gln Pro Asp Pro
35 40 45
tca acc tcg aca gca ggc gcc gcc tcc ctc ctc acg tct tct cct cac 309
Ser Thr Ser Thr Ala Gly Ala Ala Ser Leu Leu Thr Ser Ser Pro His
50 55 60
tac acg act tcg ctg cgc tct cga cac tcg ctt tac gcc acg gaa gac 357
Tyr Thr Thr Ser Leu Arg Ser Arg His Ser Leu Tyr Gly Thr Glu Asp
65 70 75 80

```

ES 2 787 225 T3

5 cgt gtt gtg ctg gat ctc ggc tgc cgg ata tgg aag gtc ggg ttc agc 405  
 Arg Val Val Leu Asp Leu Gly Ser Arg Ile Trp Lys Val Gly Phe Ser  
 85 90 95

5 ggg gag ccg cag ccg cgc gag tgc cgg agc gtc gtg agc gag ttg gcg 453  
 Gly Glu Pro Gln Pro Arg Glu Cys Arg Ser Val Val Ser Glu Leu Ala  
 100 105 110

10 cat gag cgg gct ggg cga aga gca ggg ccg tcc ata ggc gcg aga ggg 501  
 His Glu Arg Ala Gly Arg Arg Ala Gly Pro Ser Ile Gly Ala Arg Gly  
 115 120 125

15 gac gac gac gaa gag gac tgc ttt tgg gcg ctc gag aag gcc gag ccg 549  
 Asp Asp Asp Glu Glu Asp Phe Trp Ala Leu Glu Lys Ala Glu Pro  
 130 135 140

15 agc gag gag gag tgg ttg att cgc gag gag agg gtg aag cga tta ctg 597  
 Ser Glu Glu Glu Trp Leu Ile Arg Glu Glu Arg Val Lys Arg Leu Leu  
 145 150 155 160

20 cgc aag atc tgg ttc gaa aac ctc atg atc gac ccg aag acg cgc aaa 645  
 Arg Lys Ile Trp Phe Glu Asn Leu Met Ile Asp Pro Lys Thr Arg Lys  
 165 170 175

25 gtc atc gtc gtg gag aac cca ctg ctg tgc acg cgc gtg aag gag atg 693  
 Val Ile Val Val Glu Asn Pro Leu Leu Ser Thr Arg Val Lys Glu Met  
 180 185 190

25 atc gcg cgg gtc ttg ttc gac aac ttg cag atc ccg tgc ctc agc ttc 741  
 Ile Ala Arg Val Leu Phe Asp Asn Leu Gln Ile Pro Ser Leu Ser Phe  
 195 200 205

30 gct tcc gcc ccc ttg ctc gcc ttg atg gcg gcc gcc aca gtg acc ggt 789  
 Ala Ser Ala Pro Leu Leu Ala Leu Met Ala Ala Gly Thr Val Thr Gly  
 210 215 220

35 ctt gtg gtc gac gtc gga aac ctc gag acg acc gtt ctt ccc gtc ttt 837  
 Leu Val Val Asp Val Gly Asn Leu Glu Thr Thr Val Leu Pro Val Phe  
 225 230 235 240

35 cac gct cgc ccg ctc ttt ccc tcc ctc acc aca act cct cgc gcg gcc 885  
 His Ala Arg Pro Leu Phe Pro Ser Leu Thr Thr Thr Pro Arg Ala Gly  
 245 250 255

40 tct cgc ctg aac cgc cgc ctc cgc tot ctc ctc ctc gca ttc gcc tca 933  
 Ser Arg Leu Asn Arg Arg Leu Arg Ser Leu Leu Leu Ala Phe Gly Ser  
 260 265 270

45 tac gca cct cct ccc tcc tcc ctc aac tcc atg acg cct ccc gct atc 981  
 Tyr Ala Pro Pro Pro Ser Ser Leu Asn Ser Met Thr Pro Pro Ala Ile  
 275 280 285

50 gga cgg ata ccg aag gag ctc ttg acg gag gag ctc ata gag gag atc 1029  
 Gly Arg Ile Pro Lys Glu Leu Leu Thr Glu Glu Leu Ile Glu Glu Ile  
 290 295 300

50 aag acg cgg ctg tgt ttt gtc ggt gag gag gtc ccg ctc gat gcg agc 1077  
 Lys Thr Arg Leu Cys Phe Val Gly Glu Glu Val Pro Leu Asp Ala Ser  
 305 310 315 320

55 cgg gcc gag agg gag gcg tcc gcg ttc agc gga tog gcc atg agc gtc 1125  
 Arg Gly Glu Arg Glu Ala Ser Ala Phe Ser Gly Ser Ala Met Ser Val  
 325 330 335

ES 2 787 225 T3

5 gac act gca tca gca cgc gac aac ttt gac gac cct tcc gac ccc gac 1173  
 Asp Thr Ala Ser Ala Arg Asp Asn Phe Asp Asp Pro Ser Asp Pro Asp  
 340 345 350

5 aac gca cta cta aag gag ctt tac tcc cgc ttc gcc gcg aca tcg acc 1221  
 Asn Ala Leu Leu Lys Glu Leu Tyr Ser Arg Phe Ala Ala Thr Ser Thr  
 355 360 365

10 gcc aaa ccc gtt tcc ttc cgg ata ccc aac ttg tcc caa cct gcg atc 1269  
 Ala Lys Pro Val Ser Phe Arg Ile Pro Asn Leu Ser Gln Pro Ala Ile  
 370 375 380

15 gcg aac ggt aca ggt cga ggg tgg atc cag gtg ccc gcc tgg atc agg 1317  
 Ala Asn Gly Thr Gly Arg Gly Trp Ile Gln Val Pro Gly Trp Ile Arg  
 385 390 395 400

15 gag cgc gcc gcc gag gtg ttg tgg gag gag gat ggt gat gga gac gag 1365  
 Glu Arg Ala Ala Glu Val Leu Trp Glu Glu Asp Gly Asp Gly Asp Glu  
 405 410 415

20 cgt ggg ctt gcc gcc gtt gtc ctt gac tgc cta ttg aag ctc ccc ctt 1413  
 Arg Gly Leu Ala Ala Val Val Leu Asp Cys Leu Leu Lys Leu Pro Leu  
 420 425 430

25 gac ctt cga aag ccg atg gcc tgg tca atc ctc ctc acg gcc gcc acc 1461  
 Asp Leu Arg Lys Pro Met Ala Ser Ser Ile Leu Leu Thr Gly Gly Thr  
 435 440 445

25 gcc atg ctc ccc gcc ttc ttc cca cgc ttc aag gcc gct ctt ctt gcc 1509  
 Ala Met Leu Pro Gly Phe Phe Pro Arg Phe Lys Ala Ala Leu Leu Ala  
 450 455 460

30 cag ctc gac cgc tcg cat cct cct tcc cct ccg cct tcc ccg cct ctg 1557  
 Gln Leu Asp Arg Ser His Pro Pro Ser Pro Pro Pro Ser Pro Pro Leu  
 465 470 475 480

35 cca gca gca tcc gtc gaa cct cct tct tcc gac cct gcc gga ccc atg 1605  
 Pro Ala Ala Ser Val Glu Pro Pro Ser Ser Asp Pro Ala Gly Pro Met  
 485 490 495

35 tcg acg gac acg gcg gca tcg cct gcg ccc tct cgc tcg agc gaa gtc 1653  
 Ser Thr Asp Thr Ala Ala Ser Pro Ala Pro Ser Arg Ser Ser Glu Val  
 500 505 510

40 aat gcg aag cgg cgg cga aag cat gcc ctc gcg act cga ctg cac aac 1701  
 Asn Ala Lys Arg Arg Arg Lys His Ala Leu Ala Thr Arg Leu His Asn  
 515 520 525

45 ctg cgg cat tcg cct cga tac gcc ccg ctt gtc cct ctc gct cga cac 1749  
 Leu Arg His Ser Pro Arg Tyr Ala Pro Leu Val Pro Leu Ala Arg His  
 530 535 540

45 ctc gcc atc ttg aac cac cca tct ccg aac tcc tcg gca tcg tcg acg 1797  
 Leu Ala Ile Leu Asn His Pro Ser Pro Asn Ser Ser Ala Ser Ser Thr  
 545 550 555 560

50 gcg ccc tca tca acc ctc gct cga cag cgc gaa ggt tcg gct ccg tct 1845  
 Ala Pro Ser Ser Thr Leu Ala Arg Gln Arg Glu Gly Ser Ala Pro Ser  
 565 570 575

55 ttc tcg ccc gcc ttg caa agc tgg atc gcc gcc agc ctc gca gga gcg 1893  
 Phe Ser Pro Ala Leu Gln Ser Trp Ile Gly Gly Ser Leu Ala Gly Ala

ES 2 787 225 T3

	580	585	590	
5	ctc aag acg ggc ggc cct gag att gcg agg gag cag tgg gat gca ggg Leu Lys Thr Gly Gly Pro Glu Ile Ala Arg Glu Gln Trp Asp Ala Gly 595 600 605			1941
10	ttg cgg ttt gcg gag gca gaa gag gcg gag gga gag gaa gga gag gag Leu Arg Phe Ala Glu Ala Glu Glu Ala Glu Gly Glu Glu Gly Glu Glu 610 615 620			1989
15	ttt gag gag cgg gaa gtg atc cgg ccg gct ctg ccg gac tgg acg agg Phe Glu Glu Arg Glu Val Ile Arg Pro Ala Leu Pro Asp Trp Thr Arg 625 630 635 640			2037
20	atc gcg tag Ile Ala <210> 24 <211> 642 <212> PRT <213> Rhodosporidium toruloides <400> 24 Met Ser Ser Pro Pro Arg Phe Ala Pro Gly Ser Thr Ser Pro Thr Gln 1 5 10 15			2046
25	Ser Arg Pro Arg Ala Thr His Ser Ser Pro Thr Gln Gly Pro Phe Ala 20 25 30			
30	Thr Pro Ala Arg Pro Gly Ala Ser Ser Ala Ala His Gln Pro Asp Pro 35 40 45			
35	Ser Thr Ser Thr Ala Gly Ala Ala Ser Leu Leu Thr Ser Ser Pro His 50 55 60			
40	Tyr Thr Thr Ser Leu Arg Ser Arg His Ser Leu Tyr Gly Thr Glu Asp 65 70 75 80			
45	Arg Val Val Leu Asp Leu Gly Ser Arg Ile Trp Lys Val Gly Phe Ser 85 90 95			
50	Gly Glu Pro Gln Pro Arg Glu Cys Arg Ser Val Val Ser Glu Leu Ala 100 105 110			
55	His Glu Arg Ala Gly Arg Arg Ala Gly Pro Ser Ile Gly Ala Arg Gly 115 120 125			
60	Asp Asp Asp Glu Glu Asp Cys Phe Trp Ala Leu Glu Lys Ala Glu Pro 130 135 140			
65	Ser Glu Glu Glu Trp Leu Ile Arg Glu Glu Arg Val Lys Arg Leu Leu			

ES 2 787 225 T3

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

ES 2 787 225 T3

Glu Arg Ala Ala Glu Val Leu Trp Glu Glu Asp Gly Asp Gly Asp Glu  
 405 410 415  
 5 Arg Gly Leu Ala Ala Val Val Leu Asp Cys Leu Leu Lys Leu Pro Leu  
 420 425 430  
 10 Asp Leu Arg Lys Pro Met Ala Ser Ser Ile Leu Leu Thr Gly Gly Thr  
 435 440 445  
 Ala Met Leu Pro Gly Phe Phe Pro Arg Phe Lys Ala Ala Leu Leu Ala  
 450 455 460  
 15 Gln Leu Asp Arg Ser His Pro Pro Ser Pro Pro Pro Ser Pro Pro Leu  
 465 470 475 480  
 20 Pro Ala Ala Ser Val Glu Pro Pro Ser Ser Asp Pro Ala Gly Pro Met  
 485 490 495  
 Ser Thr Asp Thr Ala Ala Ser Pro Ala Pro Ser Arg Ser Ser Glu Val  
 500 505 510  
 25 Asn Ala Lys Arg Arg Arg Lys His Ala Leu Ala Thr Arg Leu His Asn  
 515 520 525  
 30 Leu Arg His Ser Pro Arg Tyr Ala Pro Leu Val Pro Leu Ala Arg His  
 530 535 540  
 35 Leu Ala Ile Leu Asn His Pro Ser Pro Asn Ser Ser Ala Ser Ser Thr  
 545 550 555 560  
 Ala Pro Ser Ser Thr Leu Ala Arg Gln Arg Glu Gly Ser Ala Pro Ser  
 565 570 575  
 40 Phe Ser Pro Ala Leu Gln Ser Trp Ile Gly Gly Ser Leu Ala Gly Ala  
 580 585 590  
 45 Leu Lys Thr Gly Gly Pro Glu Ile Ala Arg Glu Gln Trp Asp Ala Gly  
 595 600 605  
 50 Leu Arg Phe Ala Glu Ala Glu Glu Ala Glu Gly Glu Glu Gly Glu Glu  
 610 615 620  
 Phe Glu Glu Arg Glu Val Ile Arg Pro Ala Leu Pro Asp Trp Thr Arg  
 625 630 635 640  
 55 Ile Ala  
 <210> 25  
 <211> 45  
 <212> PRT  
 <213> Aspergillus fumigatus Af293  
 60 <400> 25  
 65

ES 2 787 225 T3

1 Leu Ile Gln Cys Ser His Cys Leu Arg Pro Leu Arg Ser Pro Leu Arg  
 1 5 10 15  
 5 Leu Pro Cys Gly Asn Thr Phe Cys Arg Ala Cys Leu Pro Pro Leu Tyr  
 20 25 30  
 10 Glu Arg Lys Gly Ile Thr Tyr Pro Ala Asp Glu Gly Arg  
 35 40 45  
 <210> 26  
 <211> 44  
 <212> PRT  
 <213> Blakeslea trispora  
 15 <400> 26  
 Ser Thr Glu Cys Pro Ile Cys Cys Ser Arg Phe Asn Asn Pro Thr Thr  
 1 5 10 15  
 20 Thr Pro Cys Gly His Thr Phe Cys Arg Asn Cys Leu Ile Arg Ser Leu  
 20 25 30  
 25 Asp His Gln Arg Ser Cys Pro Phe Cys Arg Asp Asn  
 35  
 <210> 27  
 <211> 45  
 <212> PRT  
 <213> Fusarium fujikuroi  
 <400> 27  
 30 Leu Phe Gln Cys Gln Ile Cys Ser Leu Pro Leu Asp Glu Pro Ile Ser  
 1 5 10 15  
 35 Leu Pro Cys Gly Lys Ser Leu Cys Arg Arg Cys Leu Pro Gly Thr His  
 20 25 30  
 40 Met Arg Ala Asn Ile Thr Tyr Pro Ala Ala Pro Glu Arg  
 35 40 45  
 <210> 28  
 <211> 44  
 <212> PRT  
 <213> Mucor circinelloides  
 <400> 28  
 45 Pro Thr Glu Cys Pro Ile Cys Cys Thr Arg Phe Thr Asn Ala Thr Thr  
 1 5 10 15  
 50 Thr Pro Cys Gly His Val Phe Cys Arg Asn Cys Leu Val Arg Ser Leu  
 20 25 30  
 55 Asp His Gln Arg Ser Cys Pro Phe Cys Arg Asp Ser  
 35 40  
 <210> 29  
 <211> 47  
 <212> PRT  
 <213> Phycomyces blakesleeanus  
 <400> 29  
 60  
 65

ES 2 787 225 T3

Thr Ser His Gly Val Glu Arg Arg Ile Arg Glu Lys Asp Pro Thr Leu  
 1 5 10 15  
  
 5 Leu Thr Phe Gly Pro Asp Pro Cys Lys Thr Ile Phe Gly Arg Ile Gln  
 20 25 30  
  
 10 Ser Val Tyr His His Ile Ser Lys Phe Gly Ala Asn Ala Gln Ile  
 35 40 45  
 <210> 30  
 <211> 64  
 <212> PRT  
 <213> Rhodosporidium toruloides  
 <400> 30  
 15 Leu Leu Leu Cys Pro Ser Cys His Arg Leu Tyr Thr Asp Pro Ala Thr  
 1 5 10 15  
  
 20 Leu Ala Cys Gly His Ser Arg Cys Leu Val Cys Ser Gly Ala Ser Asp  
 20 25 30  
  
 25 Thr Leu Ala Thr Pro Pro Ile Glu Thr Ala Ser Met Thr Thr Thr Thr  
 35 40 45  
  
 30 Ala Phe Thr Phe Thr Pro Pro Ala Val Ala Arg His Thr Ala Gly Ile  
 50 55 60  
 <210> 31  
 <211> 74  
 <212> PRT  
 <213> Ustilago maydis 521  
 <400> 31  
 35 Tyr Ala Ser Ser Pro Leu Pro Arg Ser Pro His Gln Arg Pro Ser Leu  
 1 5 10 15  
  
 40 Pro Ala Ser Ser Gly Ala Ser Val Ala Ser Pro Arg Ser His His Ala  
 20 25 30  
  
 45 Arg Val Arg Arg Thr Ala Ala Ala Ser Ser Val Leu Ser Thr Ser Ile  
 35 40 45  
  
 50 Arg Pro Ser Thr Gly Ser Arg Ala Arg Ala Ala Thr Arg Arg Arg Arg  
 50 55 60  
  
 55 Arg Ser Pro Leu Asp Arg Ser Ser Val Arg  
 65 70  
 <210> 32  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 32  
 ttactagtg gacggcttgt tctctcctg 29  
 <210> 33  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 33  
 65 ttccatggt gagtgatctg gtgtgttc 29  
 <210> 34

<211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 5 <223> oligonucleótido sintético  
 <400> 34  
 ttactagtg gacggctgt tctcctg 29  
 <210> 35  
 <211> 29  
 10 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 35  
 15 ttccatggt gagtgatctg gtgtgttc 29  
 <210> 36  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 20 <220>  
 <223> oligonucleótido sintético  
 <400> 36  
 ggaactcatc cgctcgatcg 20  
 <210> 37  
 25 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 30 <400> 37  
 caggccttcg ccatcggatt 20  
 <210> 38  
 <211> 20  
 <212> ADN  
 35 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 38  
 40 tcctctccg actgggaaa 20  
 <210> 39  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 45 <223> oligonucleótido sintético  
 <400> 39  
 cccaacaac accgagagga 20  
 <210> 40  
 <211> 35  
 50 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 40  
 55 aaactgat agttttgga aggtgacgc acctc 35  
 <210> 41  
 <211> 36  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 60 <223> oligonucleótido sintético  
 <400> 41  
 tcgagctcg taccaggag gagaagaagg tgatgg 36  
 <210> 42  
 <211> 21  
 65 <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 42  
 5 tcgctggatt ggtacgaca c 21  
 <210> 43  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 10 <220>  
 <223> oligonucleótido sintético  
 <400> 43  
 ccaccagtga ccatctctc g 21  
 <210> 44  
 15 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 20 <400> 44  
 aaaggtaccg actgtccga gcgagagac 29  
 <210> 45  
 <211> 30  
 <212> ADN  
 25 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 45  
 30 aaaaagctta gactccagaa cccgaccgta 30  
 <210> 46  
 <211> 28  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 35 <223> oligonucleótido sintético  
 <400> 46  
 ttggatccc gagtctcaat ccctcca 28  
 <210> 47  
 <211> 30  
 <212> ADN  
 40 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 47  
 45 tttaggcctg gaggacgggc gatacaactc 30  
 <210> 48  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 50 <220>  
 <223> oligonucleótido sintético  
 <400> 48  
 gtcttcgcg ccctctctc c 21  
 <210> 49  
 55 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 49  
 60 cgtaggat gacgggctt c 21  
 <210> 50  
 <211> 30  
 <212> ADN  
 65 <213> Secuencia artificial  
 <220>

<223> oligonucleótido sintético  
 <400> 50  
 tttaggcctc tcgctctcct gcacacttcg 30  
 <210> 51  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 51  
 ttaagcttc gcattccag tcccatcgc 29  
 <210> 52  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 52  
 ttggatcca ccctctacgt cccctcacc 30  
 <210> 53  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 53  
 ttgagctca acgcctcgat cctgactgic 30  
 <210> 54  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 54  
 gtctgctcg caaccctcac 20  
 <210> 55  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 55  
 gagacgaagg atggagtggc g 21  
 <210> 56  
 <211> 19  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 56  
 caccgtcct ctccgcttc 19  
 <210> 57  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 57  
 cctcgctctt tcgctggttc 20  
 <210> 58  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 58

5 cagccacatt cgttctcag g 21  
 <210> 59  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 59  
 10 tggatgatgc g gatattgag g 21  
 <210> 60  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 60  
 15 ttccatgga ggactccc atcgacagc 29  
 <210> 61  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 61  
 20 ttggatccc ctgtcccgtc aactctgc 29  
 <210> 62  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 62  
 25 tttagcctc agccaagttc aagcacaacc 30  
 <210> 63  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 63  
 30 ttaagcttc gaccgatctc gaggagacat 30  
 <210> 64  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 64  
 35 aaaggatccg gaacgatacc ctccaagacg 30  
 <210> 65  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 65  
 40 aaagactct gggagttgcg aggtcataga 30  
 <210> 66  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 66  
 45 ttgtctcgg atgtcgatt gg 22  
 <210> 67

<211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 67  
 5    ataatcttgg tgagcgcgat gtt       23  
       <210> 68  
       <211> 28  
 10    <212> ADN  
       <213> Secuencia artificial  
       <220>  
       <223> oligonucleótido sintético  
       <400> 68  
 15    ttccatggc gactctagcc atcagacc       28  
       <210> 69  
       <211> 29  
       <212> ADN  
       <213> Secuencia artificial  
       <220>  
 20    <223> oligonucleótido sintético  
       <400> 69  
       ttgatatcg aggctaggcg atgttcag       29  
       <210> 70  
 25    <211> 27  
       <212> ADN  
       <213> Secuencia artificial  
       <220>  
       <223> oligonucleótido sintético  
       <400> 70  
 30    tttactagtc aagatctacg aggcgac       27  
       <210> 71  
       <211> 32  
       <212> ADN  
       <213> Secuencia artificial  
       <220>  
 35    <223> oligonucleótido sintético  
       <400> 71  
       ttgtttaa cagtgccca acgacttct ac       32  
 40    <210> 72  
       <211> 21  
       <212> ADN  
       <213> Secuencia artificial  
       <220>  
       <223> oligonucleótido sintético  
       <400> 72  
 45    cgctgacct cccaatctt c       21  
       <210> 73  
       <211> 21  
       <212> ADN  
       <213> Secuencia artificial  
       <220>  
       <223> oligonucleótido sintético  
       <400> 73  
 50    cttccgacc gacttctgc t       21  
       <210> 74  
       <211> 21  
       <212> ADN  
       <213> Secuencia artificial  
       <220>  
 55    <223> oligonucleótido sintético  
       <400> 74  
       gaaccgcagg tgaaggtaa t       21  
       <210> 75  
       <211> 24  
 60    <212> ADN  
       <213> Secuencia artificial  
       <220>  
       <223> oligonucleótido sintético  
       <400> 74  
       gaaccgcagg tgaaggtaa t       21  
       <210> 75  
       <211> 24  
 65    <212> ADN

<213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 75  
 5   tatcggaag gtacgtctct cttc       24  
     <210> 76  
     <211> 24  
     <212> ADN  
     <213> Secuencia artificial  
 10   <220>  
     <223> oligonucleótido sintético  
     <400> 76  
     caggttcat cgcaactaca ttga       24  
     <210> 77  
 15   <211> 24  
     <212> ADN  
     <213> Secuencia artificial  
     <220>  
     <223> oligonucleótido sintético  
 20   <400> 77  
     aacagagcga gttgaagagt agcc       24  
     <210> 78  
     <211> 20  
     <212> ADN  
 25   <213> Secuencia artificial  
     <220>  
     <223> oligonucleótido sintético  
     <400> 78  
     gcgacgacta cgtgaacctg       20  
 30   <210> 79  
     <211> 20  
     <212> ADN  
     <213> Secuencia artificial  
     <220>  
 35   <223> oligonucleótido sintético  
     <400> 79  
     cgatggggaa ggagaatttg       20  
     <210> 80  
     <211> 20  
 40   <212> ADN  
     <213> Secuencia artificial  
     <220>  
     <223> oligonucleótido sintético  
     <400> 80  
 45   gcacactgca cgccttactc       20  
     <210> 81  
     <211> 20  
     <212> ADN  
     <213> Secuencia artificial  
 50   <220>  
     <223> oligonucleótido sintético  
     <400> 81  
     acgagctgaa gagcctgtcc       20  
     <210> 82  
 55   <211> 20  
     <212> ADN  
     <213> Secuencia artificial  
     <220>  
     <223> oligonucleótido sintético  
 60   <400> 82  
     gcaagatacc ccagctcgac       20  
     <210> 83  
     <211> 20  
     <212> ADN  
 65   <213> Secuencia artificial  
     <220>

<223> oligonucleótido sintético  
 <400> 83  
 ggggacgttg acgtagaagg 20  
 <210> 84  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 84  
 ggctggatga aggagtggac 20  
 <210> 85  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 85  
 aggaggagcg tgagtggaag 20  
 <210> 86  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 86  
 atgggactgg aaatgcgaac 20  
 <210> 87  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 87  
 gggagacgaa ggatggagtg 20  
 <210> 88  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 88  
 tcgtgcacaa cccgaactac 20  
 <210> 89  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 89  
 atcttgcgct ccttctcgtc 20  
 <210> 90  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 90  
 accagcttca gaccagctc c 21  
 <210> 91  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 91

ES 2 787 225 T3

agaagttgga ggaagggatg g 21  
 <210> 92  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 92  
 cgacaacttt gacgaccctt c 21  
 <210> 93  
 <211> 21  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> oligonucleótido sintético  
 <400> 93  
 caggttggga caagttgggt a 21  
 <210> 94  
 <211> 597  
 <212> ADN  
 <213> Ustilago maydis  
 <400> 94  
 gattagatct tgctgatagg caggtttgct tggagaatgg ggggaaaaga ctgaccgaag 60  
 aaacagcgag atctagaagt gataagcggga aagaatctga cttgctgtga tcagcagcca 120  
 atttttttttt cgttttttttt ttttcaactcc acatcgtcgt gcgtgcacgg tctgcatgtg 180  
 taaattgtat tcatcgaaag ccacagttga atacatcagc ccgatgtgga tttcgaaaac 240  
 caattaatct tggaattcac gcgctcagat cagtcocatag agtcgacttc ggctgtttcc 300  
 aagagcttct tctctgcgag gtggttgccc gtgtttctcg ctgggaaaaa aggatcgatt 360  
 attattcget tctacctcgc tcgcaccctt ggctcgtga aggaaacagc gccgagactc 420  
 ggtcacgggt gctgggctcc gtgttgatgc tgggacggcg caaagtgggg cccgcgcact 480  
 cttcgagcca aggacctcac tcttcaagaa caagcgtgt cgccatcgtc ttcttctttc 540  
 tgctccacca tcgaatcttt ctttctcgtt tcgaaaccaa aacctcttc caccatg 597  
 <210> 95  
 <211> 1434  
 <212> ADN  
 <213> Rhodosporidium toruloides  
 <400> 95  
 ctgcagaact acgccctctc acacccaact tccgactcga ccggcgggtac gagcacgacc 60  
 tacttctact gcctgccatc gacatccggg cgggtcgtctg cctaccctgt gcgttctgcg 120  
 cctccctcgt tctcgggagg cagtgtctga cagaagcttt gcgcgcagta ccccgtaag 180  
 atgcaactct acgcaacggt cggcacagaa gtgcgcaagc tccgcgcacg gccgcctcaa 240  
 gctctcgcgc tgcccgcgag tgcgtctat tacgaggcgg agaagctcga gttgccggct 300  
 ttgccagcgg cggtaacggt tgaggtggag acggagaagg cgggagtagc gggggaggac 360  
 aatgaggcga agggtagat ggtgctggtg gagactctta cggtaggca ggaggagatt 420  
 gaattgggct cgggagtcgt gcagattgag gtagctgtgc tcgtcaagct ggaggtcagc 480  
 ggctgatcct tccgttcggt gcaaggatcg tctgcatgtt tcgcttctct caatgacaca 540  
 acctggagag cgctcccgtc agcgagaatc gaggacattc cgcagctcgt gagcaagcgg 600  
 aggtgcgagg ctccctcgaa agctgcgcct cttcagacgg cttgttctct cctgctctgg 660

ES 2 787 225 T3

	tgggctggcc tgacatgtaa tgtgctccgc cgcaagtccg tcgtcggctc caattcgacg	720
	ttgaaagggc atagcgcaag gaagaaccct ctgcgacat gcagaattac tggctcgcct	780
5	gctccttcgt ctactggaat aagtccctgtc tcggttaaagc cccaacgtcg tttttcgacg	840
	tttgtaaggc gcaagaggtg ctatgggcta cgcaggaagc tgagaggaca tagaagtcgg	900
	gggaggaacg gcgcagagcg gcagttgctg aagcatgagc aaagcgagac ggtccagcat	960
10	ctgcagcgcc aatccgcaat ctcttggttg agcctgcacc ggaagcgtcg gaacagtatg	1020
	cgagagtcg aacgcaagta agaaagacgc accctcacac tcgcttactt cgagccatac	1080
15	aacggatcaa agctgcgcgt atctcggctt gtaagggccg gaaagcaacc tcggagatgg	1140
	acacgtcaca tcaccaactt atcgatctcg gccgtcgacg tcgcagagag ggcgagagaa	1200
	gcggtgaagc agggaaacaa cccctcgaga gcatgatccg accgaatctg cagcgcagga	1260
20	agccggtaca agcccgcctc gagcgcaggt cgggtccagc cgggggacga aacgcgcgag	1320
	gctgattcgt gagcgaagga agccgcacgc acaagttcgc tcccctttgc cctctttccc	1380
	atcacccggt ctgccttac ccgctcagaa caacaccaga tcactcacia tgtc	1434
25	<210> 96	
	<211> 888	
	<212> ADN	
	<213> Aspergillus nidulans	
	<400> 96	
30	ctgtacagtg accggtgact ctttctggca tgcggagaga cggacggacg cagagagaag	60
	ggctgagtaa taagcgcac tgcgccagac agctctggcg gctctgaggt gcagtggatg	120
	attattaatc cgggaccggc cgcctcccg ccccggaagtg gaaaggctgg tgtgccctc	180
35	gttgaccaag aatctattgc atcatcggag aatatggagc ttcacgcaat caccggcagt	240
	aagcgaagga gaatgtgaag ccaggggtgt atagccgtcg gcgaaatagc atgccattaa	300
	cctaggtaca gaagtccaat tgcttccgat ctggtaaaag attcacgaga tagtaccttc	360
40	tccgaagtag gtagagcgag taccggcgc gtaagctccc taattggccc atccggcatc	420
	tgtagggcgt ccaaatatcg tgctctcct gctttgcccg gtgtatgaaa ccggaaaggc	480
45	cgctcaggag ctggccagcg gcgcagaccg ggaacacaag ctggcagtcg acccatccgg	540
	tgctctgcac tcgacctgct gagtcctc agtcctcgtt aggcagcttt gcccgtctg	600
	tcggcccggg gtgtcggcgg ggttgacaag gtogttgctt cagtccaaca tttgttgcca	660
50	tattttcctg ctctccccac cagctgctct tttctttctt cttcttttc ccatcttcag	720
	tatattcatc ttccatcca agaaccttta tttcccctaa gtaagtactt tgctacatcc	780
	atactccatc cttcccatcc cttattcctt tgaacctttc agttcgagct ttcccacttc	840
55	atcgcagctt gactaacagc taccocgctt gagcagacat caccatgg	888
	<210> 97	
	<211> 617	
	<212> ADN	
60	<213> Aspergillus niger	
	<400> 97	

65

ES 2 787 225 T3

gtcgacgaga tcgtaggagt gagtaccggc cgtgatggag ggggagcacg ctcatgggc 60  
 cgtacggcag ctgccgaggg ggagcaggag atocaaatat cgtgagtctc ctgctttgcc 120  
 5 cgggtgatga aaccggaag gactgctggg gaactgggga gcggcgcaag ccgggaatcc 180  
 cagctgacaa ttgaccatc ctcatgccgt ggcagagctt gaggtagctt ttgccccgtc 240  
 tgtctccccg gtgtgcgcat tcgactgggc ggggcatctg tgcctcctcc aggagcggag 300  
 10 gaccagtag taagtaggcc tgacctggtc gttgcgtcag tccagaggtt ccctccccta 360  
 ccctttttct acttcccctc ccccgcgct caacttttct tccctttta cttctctct 420  
 15 ctcttctct tcatccatc tctcttcatc acttccctct tcccttcatc caattcatct 480  
 tccaagtgag tcttctccc catctgtccc tccatctttc ccatcatcat ctcccctccc 540  
 agctcctccc ctctctcgt ctctcacga agcttgacta accattaccg cgccacatag 600  
 20 acacatctaa accatgg 617  
 <210> 98  
 <211> 388  
 <212> ADN  
 <213> Ashibia gossipii  
 25 <400> 98  
 cactatacgt gcctcgtccc cgcgggtca cccggccagc gacatggagg ccagaatac 60  
 cctccttgac agtcttgacg tgcgcagctc aggggcatga tgtgactgtc gcccgtaac 120  
 30 ttagcccata catcccacatg tataatcatt tgcattcata cttttgatg gccgcacggc 180  
 gcgaagcaaa aattacggct cctcgtgca gacctgcgag cagggaacg ctcccctcac 240  
 agacgcggtg aattgtcccc acgcccggcc cctgtagaga aatataaaag gttaggattt 300  
 35 gccactgagg ttcttcttcc atatacttcc ttttaaaatc ttgctaggat acagtttca 360  
 catcacatcc gaacataaac aacctagg 388  
 <210> 99  
 40 <211> 963  
 <212> ADN  
 <213> Rhodoturula grammis  
 <400> 99  
 ggcgcgcta cgtctacgtc aagggcaatg ccgacgtgac caaggccatc gcccaggacc 60  
 45 tcgccttctt ctcggtccct gtcgagctcg gcgtgcgtcc cgcgctctc tctctctct 120  
 tctctcggcc ggcctcacg tgatccacga cgtcgtactg acccttgoga atgtgcgcgc 180  
 50  
 55  
 60  
 65

ES 2 787 225 T3

	ccgcagccca acggcgtoga gaaggtgcac ccgctcggcg acctgacggc gttogagaag	240
	gagctcctcg aggcgtgcct cggcgagctg cccgggtcca tctccaaggg cgagtcgttc	300
5	atccagggct ccaagctctg actcgcggc gcacgcagcg gcgcgagcca caaggcgagg	360
	atgtgagagg aggcgtttcc tccaccttgg accccatctg ccgcctcctt ttctctctct	420
	ttctttccct tctctctct ctctctctct ctogttctcc tccttctggg cctctcggac	480
10	ctcttctctg ccgtcgactc gtgaaaatgc agtgccgctt tctgtacctt gtctcgcgag	540
	agagatctgg ttctcgcagg gtgagtcgct gccttggccg tggcacgcct cgcgcgagcg	600
	agagagaaga ggccacggtc cagcagcagc acgacgagga ggaagcgcaa aaggcgagac	660
15	accgagtgcc atcgattccc cgctcgaacc tgctcacggc tgtcgaaggc ggtcgcgccac	720
	ggtgcttgcg ggagcgaag caagctggcg tcgtcctctt gaactggttc gactcctga	780
	gggcccgcac gagaactcag gcgaggtgct ccgctcggaa caagccgggc ttgtggtcga	840
20	gggagcgaga gcgaggcagc gccctcgtcg ccgaggcaag agcggcatcg acaagttggc	900
	ccgtcgcctc tcgctccctc ttctcctcct cccaccacca cctttctcca gctcgaacca	960
	tgg	963
25	<210> 100	
	<211> 1009	
	<212> ADN	
	<213> Sporobolomyces roseus	
	<400> 100	
30	ggcgcgcga agttatacct cagaggtctc aaaaacgaaa aagtcatgca agaattctct	60
	ttgacgtgag ggttatctct ctctctctag tagtctacga gaatcgcaaa gatcggaaaa	120
	ctgatgcatc tttgtgttca cgggtagcg atttgatctt ttcgattccc aaaatcgtat	180
35	cgttcctgtc gcagggaaact acgctcaaag ccggcactct gatcatcacg ggagtgagtt	240
	ttgagctctc cctctatgag agtgcaaggc tcgctcgtga tgggtgaatc cgctcatgcc	300
	ttcccctcta cettctcctt tgtccattct ctctactacg gttgtcacat ctctctctc	360
40	cgacagacc cgcacggaat tggagcgtac togaatcctc cggaaattctt caaggacgga	420
	gacgtcttca gggtcgagat ctccggaggc atccggagtt tggtaacaa gatcgaatat	480
45	gaaaagtaga taatccgta ctcaggtcaa tggatggct tcgaagatgc tggaaatcagc	540
	cggaaagcaa agctggagag aaaaatcgag attgcgaaac gtgcgatgtc atttcgtttc	600
	gagctcgcga ccatctcgta tccctctgag ctacatacaa acgtcactac ggcctcggag	660
50	tgactccctg cgagcggatt gaaggagatc acggtcgaat cagctagacc ttcgcaacgt	720
	tttcgcgctc gcacgttctt atcgatctac tgagattgac tcgaaaaagt ctctctctac	780
	ggtogattga actttgaatg aactctcagg ttgcgcgaga gccaaatcga gccgaccaga	840
55	ggcaattcgg agcttcccgg aacgttccaa ggagagggat tttccgagag attacgattg	900
	cgagatagaa aaaaggctag ctttcgattt cgagagagat tactttcaag ttcgctgctt	960
60	ccaactcttg ctccaacccc ctccactcct tctctacaaa acaccatgg	1009
	<210> 101	
	<211> 2148	
	<212> ADN	
	<213> Rhodotorula glutinis ATCC 204091	
65	<400> 101	

ES 2 787 225 T3

	ggcgcgccac catctcctog tcgcttcttc cctctccttc ggcgcccaca ccgcttcgca	60
	gggctcacgg actgctcaca tcgtttgtgt gcgtcgctgt gcatgtccac gcaccactcc	120
	cagccccac gagcgcctca aaagacgcgg acgcagacgc ccgccgaacg acggcacgcc	180
5	cctcttctca ctagcgcgac gaaccagctg cgacgattcg tgcgcttatg ttagccggac	240
	ttctggcttg ctttgccctg ctgcgtccgt cttgtggtgc ggatcggctc gatgggggtt	300
10	tgctcgtttg ctgggagacg gtcgcctctc cctcctcctc ttcactcctc gttagctttc	360
	tacgctcatt ggttctgcga accatctaca tcacgctcgc tcgctatgct cgtactacga	420
	tcaacacccc tgctcgtcgt gctttccctc ctctccgtcc tctcggccgc gtccagcgac	480
15	ttgccagcc aacttcccc gcacgccggg gagtctccca cacttccttg cgaccccaac	540
	ccagcatctg acatccgcat cacgcagccc tcccgccttc ccaactcctcc ctcttcaccg	600
	actcctcctc ctccctccct gattcctcgt ccctcaaagc cccgcagcct cttcccttca	660
20	aatcaagcg cccccgctcg ctcgaacaag tgcagcagaa cctcgggaag aggctggcga	720
	agcgcggcga ggaggggagt aagacggaga gggtgccgtt tggtcagagg agtgcgacgg	780
25	cggcgagtgc ggggtggaca ggtggagcgg ggacggggag ggcgacgcag cgcgttacgg	840
	gcggaggaag cagaggtgca ggagggggcg gaggggagtgt cgcggctgct cagcctgtcc	900
	cttcgactac ccagacggtc gagacaggct ctaagatcgt ctcgactggt cttctgaccg	960
30	tagcgtcgcc gtcgacggca gatggaggag gcgggacggt caccaggtc gagacggcct	1020
	cctcaggggt attgatcacc agcacggcgg gagcggcgag ttcagcggcg gcgtcggacg	1080
	tcgctagcgc acaggcagcg gaggcgacgt cgagtacgag catgatcagc ggaggagcgg	1140
35	cggctggcgg gagtttaagc aggatgctgg cgggaggagt tgcgggtgca gccctgatcc	1200
	tcctcgtgcg gtgagcaggc gaagcgagga gctcatgtag atacagcata gacagtatat	1260
	atcgccagga tagcttgcaa cagccgccgg tcggtttatt ccattgtcct cgaccccatg	1320
40	cgaaggcgag ctctgctcgt cagctggcca agctggccag cagacgagcg ttgggggtggc	1380
	ggaacgcaa cggcatggag taaagcagcc gtgaggatga cggaggagct cgggcgaggt	1440
45	gatggggatt ctagcaggaa cagcagagcg gcgaggagga gaggaaccgg aagcacagtc	1500
	tcgtggccgc ttgttgacaga tcccagtgtc gctagagtgc tcgtcgtcat cagagcgagt	1560
	gaacaaagcg atgccctgaa gaacgatgag cgaatgagtc gaagcggcgt ctaccgggtga	1620
50	actcggggtg tggcaaatga gcgagacgag gagtgcccgc cagagttgcc acgtcgacct	1680
	cacgtcggaa tcgacgttga tagagtgaac gaagccattg cagaccccag aaggtggcca	1740
	tgttgaggaa gcgagggcag gagcagggg agaaggcgag gaggaggagg ggctggggaa	1800
55	gcccgctccg gaatggcgca gctgggtgcc ggggatgtgc gcgagtggcg gaggagtcca	1860
	gcgtgagagt tctggaacac ggggcgcgca caagggtcga gggccgtgac gagttcgccg	1920
	ggcgggtggtc gggctgaggg cgagcgcgcg ttggggacga cgacgccga cgcctcgtct	1980
60	cttcgtcctc accgcttccc ggagaacttt gctgtactct gcttctccct tcacactctc	2040
	acacccactc acacaccctt ccatccacac acaagctatc cgcacacctc tcacaccgca	2100
65	cccagctcg ccccatcctc ttcgcacccg gctcatcgaa aaccatgg	2148

ES 2 787 225 T3

<210> 102  
 <211> 1015  
 <212> ADN  
 5 <213> Rhodotorula glutinis ATCC 204091  
 <400> 102  
 gtgcgagaag aagcgaggca cgcgaagcgg tagaagcaat gaagcgaggg gagagcgaga 60  
 gaggcagggc ttcagccatg tccagctgat cggctgtaac gtcgcccgg gccagtctgt 120  
 10 tgaatttggt gcgtcgctg agcgtaatag aagtgcagta gtctactccg catgccgaga 180  
 acgtcgaaga gcgcgaagta gggagtcgag ggaagcgagg gtggcaaaca cagcaacgac 240  
 aagcggttcc gcttcgctca aaagctcgtt gacgttgttt tgacgttttg aagacagtac 300  
 15 aacagcagca agagggctgc gaagcgttgg tggcgagagc agcgacaagg agggaggaat 360  
 gagggagtgg tggcgagggc tcgcaaacgg gcgtacgcct cgaatggaga cgtgcgagtc 420  
 20 gttcttcgac gtccgagggg tgcagagcgc cgagacggag cacgcaacga gcgagaggag 480  
 agcagcccgcg caaggtgatt cgagtggcgc aagcggagga cgacgaggag acggacgagg 540  
 gagggaggag gatggcgagc gagcatcggg cggcggggcg cgagagacgg cgtgaggagc 600  
 25 cgggtgtgga gagtttgagg aggcgcggga tgcgaagtgg ctgggtgtgc ggagtgagcg 660  
 gtggcaaaga gcgcacttag agtctagagc gaggcagtag tagtagagct gtatgaatga 720  
 atacaaagtg tgaatacaac agtttgtaat gcgattctga gcttggacgt gtgcgcgcgga 780  
 30 gagggcgact tgcaagccag cgcgcgctcg ctcttcttcc ttctgcacct cgcgtcaacc 840  
 ctgcgatctc acacctacac tcgattcaa agtgcgtaca ctctcccacg acacacgggg 900  
 35 acggcgcaca ccaccgcgog tcgcttgaac ggcgtcgcca cttcgagccg tcaactgactt 960  
 cgtcctcgtc ctccctcctc tactctcttg tactgtactg tgtactgggg gggaa 1015  
 <210> 103  
 <211> 944  
 <212> ADN  
 40 <213> Rhodotorula glutinis ATCC 204091  
 <400> 103

45

50

55

60

65

ES 2 787 225 T3

catgctgctg ctgcccctca aaggtcctct cgtccacgtc cgacgagtct ggacagcttt 60  
 cacagtcccg agagtgcaag agcgaggcgg gctcacggtt cgcgaaagga gcgagaggtc 120  
 5 cgaccgccgg ccggtctcct tgccgcctc gcctcacctc ctcttgacagc aggttcacct 180  
 ctctgaggtc actcgatcgc tcgcagcgat gcgcaggtac aagtacgcta ggcgagagcg 240  
 tcgaaagcgg ggttctgcga gggactggac gctgcagagc gcggtcgaga gaggtcgag 300  
 10 tggcgctttg accgctcgac gcaaggcatg cgctcctccg tttgagctcg cagatactgc 360  
 cgtgcgaaga cgagcatagg ctgtggctgc ggtagcaagg agccggcgag agaaagctgt 420  
 15 gctcgagcag gacgagagac ggtccgcgcg cttgagaagg tcgaggtgag gcgtcgcaac 480  
 cgggttggat ctcgattctc ggcaactac ggcttcggcg agggccaaag cgacggcagc 540  
 ccgcgcaagc tggccaggcg agagcgcgag agtcgcgagc tgaagcgggc gcgggtaga 600  
 20 gcaagctggg gaagcgagag agggagagag agagagtgag ggggtggcga ggtggagacg 660  
 aggcgagcgg ttggcttgcg cgcgcgcgag agggatcgag gcgagaggcg agccccgaga 720  
 gtggaaggaa ggacgaggaa acctgcgtgc ggaggcgcg cgcgcgctg ccacctggct 780  
 25 gagcacgggc ccgagcttga gggagctggg ggcgcgcgag cgagacgagg gcagggcgag 840  
 cccgcgcgtg gcggccgcct cgcaacccaa ggctcgcctt ggccgcgcgt ctctgctctc 900  
 ttctccacc ttcgcgtctc accactcgaa tctcacttca tcca 944  
 30 <210> 104  
 <211> 1462  
 <212> ADN  
 <213> Rhodotorula graminis cepa WP1  
 <400> 104  
 35 caaagggaga ggagcggcgg gaggatggtg gtgccggaca gggcgagagg gaaggtcgag 60  
 ggagagatgg gagcgcgaaa ggtcgaggcg gggagaggga ggggtgggtc gggaccaagg 120  
 gggcagagag gctcacaagg acggaggagc ttactccgcc ttgaccttgc gggtagcggc 180  
 40 ggtgccctcg cggaagctgc ggcgcgagga aggtcgtcag ggccgggccc aggggaggaa 240  
 cgaggacggc gacgacgacg cacccttct tgaagcggcg cgagacgcc ttcaggtgct 300  
 45 gcatgcggcc agtgccagtg gtgtggcggc gcttggcctt ctgtccccac tcgtctgcgc 360  
 ggcgagggag aaggcgaggt gagcacgacg gcgcgcgagg gccggacgag gctgagaggg 420  
 ggacgcacac gagcggagct tggccgcggg gtagccgcac gaagcgcacg acttgtgctg 480  
 50 cttgtggaac gagcggcggc cgcagcgacg gcacagactg tgggaccacg agggtaacc 540

55

60

65

ES 2 787 225 T3

	gggtgctcgc gagacaggag cgcggcttgt ctogaagcac gggcaaagag agcgttggac	600
	gcacgtgtga ctcttggtgt ggcggagacc gaacgaggtg gtacccttcg tctgtggggc	660
5	gcaaggagga gtgggtcagc gtcgggcctc gagggcctc ggtcgtcgac ctcgcccgc	720
	cccgatcctc gcgccgtcct gtcctcctc tctccaacc tgcgacgtgt tgcggcagca	780
	gcagcttget gggacatgtg gggagggcgg caaggcgagg ggaggtcgag gtgcgaatgt	840
10	gggtggctgc gcttggcggg gcagcatgtc gtcgcgccct cgagccgggc gggcgacctg	900
	gtggccgggt cgagcgagag gcgtgggagg gagtggcgca aatggcgtgc gctcagaggc	960
15	gggttgtcga ggcgtcgagg cggacgaggt cgaggaggtc gaggtgggaa gctgctgctg	1020
	ctgctcgggc gtcgtcgccg cgtcccagat gcccctgctg cgcccctgct gccgctcctt	1080
	gggccgtcct ggtcccacct gcccgtgccc tcctccacga gagcgcgagt ggggctgtgc	1140
20	gccgggttgc gctccaactt tgcgagagag cgaggacggg ggcatggctc gctcgccggc	1200
	ctcgggtcgt tcgaggggtc gggggcgggt tgcgggaggg tggtcgagg tggcgggctt	1260
	accattgtcg cgtcggagag gggggttgg cctgcgagaa gacgaggaga cgagaggccg	1320
25	ggggaggcga ggcggcgagg cggcgagacg gctcggacca agcgcgcgcc gccaaagtct	1380
	gcctcgccgc tcgcgctcgc ctccctcttg ctctccacct cctcctagga ccacaaaggc	1440
30	acccttgtgt aggcgtaggt ca	1462
	<210> 105	
	<211> 1315	
	<212> ADN	
	<213> Rhodotorula glutinis ATCC 204091	
	<400> 105	
35	tgagcggggc agccgcgagc gaggagcgtt gaggaggaag ggagggaggg aggagggaga	60
	gggagggagg ggacccccat cttctattc attcaciaag acgacggtgc ggaggggtcc	120
	ctcgagtgtt tgggctgggc ttcggagtct cgtagcgagc aagtagtgtt tctctccggt	180
40	tcgacagctc gtattgtcat ttcttgttca ttgtcgttc cggcgactgc aggtacgctg	240
	atcttcggcg gagacgaaa gcacgtgggt tgtgagcagc gagttgagca agaaaagcg	300
45	gacgaaggcc ctcgtcgggg gcttcaagtc aagattctgc ggagattctg cgagagactg	360
	caagcgttga acctgttgag atctcgtcgg acgacagcac agtgtccgtc tcgctcaatg	420
	cgataggaag cgagagagag gaggaggata tcggaggaag gctgtttgc gttcgctcca	480
50	ggcgtcgcaa gatccggcgt agagcacaat cgtcgttggg tcgacgtttg tagttcgtca	540
	cgagtgaggg cgaagcctgg caagcaaaga aggggacgag cgactcggca gctatcgctg	600
	gaggagggcg actttgtggc ccgtttccgt cgagctcgac gcgagtgagc gcagggctcg	660
55	tccgaaccga tgccatggac gcagtgagcg aggccggatg tgcgatgctg tttcaagcga	720
60		
65		

ES 2 787 225 T3

	gcgaaggaag ggagaaagcg agcgagaggt cctcctcctg tcttcctcac gccttccgaa	780
	ggccgacaag aggcgtagac gtcgacgagt caacggtttg acgtcgctca ggctgtagc	840
5	gggtcgtcgg aagctgggaa agagaggaac caacgagtaa caagcgcgag agtctcctca	900
	aggcggacaa ttgcctcgtc tcggtcccgg tcgagctctt ccagtaccag cgagggcgaa	960
10	agtcgctgat gcgtgcgcat ccaaggccaa gcgtcgcagt cgagaagagc gagagtgaag	1020
	cgagtgaagc gggagagtga gacgggtaa tccgcgtact tacgagtggg ttgtattcct	1080
	tcttgtaatg gcagattacc tcgattggcc acgtcacggt ccgggagtgcc ccgggcgtgg	1140
15	gcaaaagggc gagcgcggcg cctctctctc ttgcttcctc agcagagcag ctctcccctc	1200
	gagtacgtcg acgggctcac tacagctagc aacagcaagg ctaccacgcc agctacacgc	1260
20	cagctcacc c aactcacacc gctcgttgtc gccgcgcgcc gcaggaaaac tttgt	1315
	<210> 106	
	<211> 1497	
	<212> ADN	
	<213> Rhodotorula glutinis ATCC 204091	
	<400> 106	
25	gctcatcadc gagcggggc agcagcgacc tgcggactgg ccgaagaact tccccggcca	60
	gcgcttcgag gacaaggaca ttcgacgcc tcgctctcag tggcgtaca tcaagctggc	120
	gaccacagac gacctctcgc cgactgagga gaacacgacg tcctgcgccg tccggtacgg	180
30	cgaggactcg cagctcggca tcttccacgt tcccggcaag gggacttctt gcacgcagca	240
	gatgtgtcct cgtgcgtcgc gcctgctctc tctttcttta tttgctggct gactcgtgct	300
	gactcggccc aaacctcaga caagcgcgcc ttcctcctcg agcatggcat cgtctcggac	360
35	gacgggagtg gccacctcta cgtctcctgc ccgctccaca agcgcgaact ccgcctcgac	420
	aacggcgact gcctcaacga cgaggagtac aagatcctcg cgtttgatgt caaggaggag	480
	aacggcgact tgctcgttca ggttcctccg cctgacgagc tcgacgcttt gattgggtgc	540
40	gtctcgctta gccctctctc aaagacctga gctgaccctt ctgattgtcc gcagctcgtc	600
	gaagtggatg gtgcgcaaag cgaccgccga agccttcggt cgcaacgcag cgacagccat	660
45	cgagtgcgtc ccctccaagc ttctgttttc cgcgcgcaca ctaggctgac gacaagtctc	720
	tgcaggatcg tcggaccgtc aggcgaggtt gacgaggaca agaaggcagc gggaacagag	780
	tgcggcgaag cggataagtc ttgcgggacg cacaagctcg agtggtgatt cttgcgggtc	840
50	cgtcacagcc aatgtateta tctctagatg tccttctcgg gtatatcagt tgttcgtgca	900
	tcgtagacgt cgttttagcag ctctcgttca gccacttgcg aaggcccgtc tcttcgacga	960
	caaggacggc ttcgcttctt ttacctcgtc gtctgagcgt tctcaaggga ccctoctacg	1020
55	cccttctctc cacaggagcg gccgacgagg cagccttgct ggcttatcgt cgcttccgcc	1080
60		
65		

ES 2 787 225 T3

	tttcatgctc gagcaagtcc tcctgogagt gtcccgacgt cggcccgcct tgcccaaggt	1140
	cgccgactgt cctatcgcga cactgcgaat gcaactgctgt ccgcgcgga gactgtgcgg	1200
5	cgcaattga gggcaaagtc gtgcatttgc gaaacggtat ccgctcgaag ggcccacgat	1260
	agacctccac cggcctcaaa cttggcgaca gggtcgcttc cgacggcgga cagcaagtta	1320
	ggctttggcg tcgtcgctgc gatccgcttt gggggacccc ttatcgcgac tgccggatc	1380
10	gattggcgat atctctcgct cgctggcctc gctggacagc tggacagtct ctgcagcgtc	1440
	gaagcgacgt cgataaagtc agcgacgtcc tcgcgaacca agaagaatca cccgccg	1497
15	<210> 107 <211> 1594 <212> ADN <213> Rhodotorula glutinis ATCC 204091 <400> 107	
	cgccagggaa cgcaagaaag gcgggacgag cgagggtgag tcgcgcaagt cgacgaagaa	60
20	gcccaggtca acgtcgaggg ttgtcaagtt ctggcagacg gacagcagcc aactaatcg	120
	ctgccgactt ctgatgcac gtgcgcgggg tgtcagctcc tagccgtgga aggaggaagc	180
	caggacgttc tcaccggggc acgcattgct tgtgccagac ctcgaggcga gcttgtcgcg	240
25	tcgaaagacc ggcgacctcg tccggctcaa attccagccc gagtgattgc cctgacaaca	300
	aaggcgaaaa gctgaagcgg gtaccaaagg tcggtccatc gaaagtcgcg ctccgaagac	360
	tggcgtcgac ggatctgacc atcgctgccc ctccctgcgc tgctttgagg caccttacag	420
30	cctgctcgtc tcgttcggag cctccgcac ctgcttgag gaccacctcg cgaccagtga	480
	cctcccttgc gatggctcgc caagtcttgc atactccggc gacgttgcgg aaggcgcagg	540
35	tggggcagga gatgcggagc gttgtgaact gtccgttgac gaggagtgtc ggcgaaagaa	600
	gggcgggtggt gagagagtag gtgaggatct tgaggaggag ttcaggagga agtgaggaaa	660
	ggtctgccgg tgactggtaa ggctgaagca tgatggcgag tgtagccaag tgatccgagc	720
40	gacgatcaag agacgaagga cgagacaacg cttcagcgcg cgaagagagc gagcggaggc	780
	cctcctggtc gagaggctat ccagtcgcca accggtagcc atccagtttg cagggttgaa	840
	acacagctga gaggatcagc gactggtagc gcaaactcct aaggcgtga acgtcaagga	900
45	cagcagcgtg gagcgtgtgg aagcgaactt cgaaggccaa actcgtgtcg cgctggccaa	960
	ccgcctgccc gctttgacgc gcttctgcgc cctccgccta ttcagagagt atgcttcgtc	1020
50	acggcgtggg cgccaacatc ggcgaggag ctggcgggac gggaaagaaag ccgcaaccgc	1080
	ggtcctcgac cttcaacgtc ccgggaggcc cgtccacgac tcccagacgt ctctgcttgt	1140
	tgttctacgt cgtcgcggcg ttgtgcagag tccagcgcgc gcccgctcgtc gacttctgac	1200
55	aagcgataaa ttccgagacc agcgggagaa ggcggaacga gaggaggagg cgagctggcg	1260
	tccttgcgac ctcgctgagc agttcaagcg agcagattga gcagcagtgc gtcgagtgag	1320
	ccaactcacg ttctcatatc ggtccctgag cgatatcgat gaggcgaagg acgacgacga	1380
60	gcgaactgat ctgcgcgtct cctcttccc cttcactctt tccactcaga aacaacacgt	1440
	gcgtcttctc tgaacgctat cagacaatcc aggaccatcg ctgaccgcgc gctcactcgt	1500
	cgctgactt catcgcccaa ccaaccgct cgtcaccgac tggatctctc cctcccctca	1560
65	caccacctgt tgcgctgcga tactcctctc acag	1594

ES 2 787 225 T3

<210> 108  
 <211> 1396  
 <212> ADN  
 5 <213> Rhodotorula glutinis ATCC 204091  
 <400> 108  
 cggcgaagag gagggtggtt aggagggaga gctgtgag ggggagaggt cagtccgagc 60  
 acgataccgc agcaggccaa gcggcttcat gttcactcca agctcgatgc ggtcgagaag 120  
 10 tacgagctcc ttgacgagtg aaagggagag aagagagact caccaacccc ggcgagacct 180  
 tcacgaaccg caacttcccc tcacacgcct cactacacgc cctctcaaac acctcgagcg 240  
 tcgcgtagtg cgagagcttg aagaataggt aggcgaggaa ggagagggcg aagagggcg 300  
 15 ccatctagcg aacaagtgg gttagctggg tgggggaaag ggaagagggg aggaagaggg 360  
 ggaggggtacc atgatccagc ctatgtcgac ttgcagttgg aggggcatcc tgtgcgcgta 420  
 ctggcgctcag cggcggacga caaggcaggt agacgagaga gggggccggg gaacgcactc 480  
 20 gcagttgcbc acgtccgaga ggaacatgta cgagtagccg gccacatga cgctcagcag 540  
 cgccgcaagc gtgtagtggg agatgagcct gccattggc agtcagcgcc gacgaggaca 600  
 25 aactctgggt gaagagaaag gagagaacgc accacttctc ctccctcagc actcccctcg 660  
 cccagaccgc cccaacaca atacagacga gatcgatagt tgccccagtc acagcgagcg 720  
 cgagttggct gtagacgagg tgtttgagcc gtctgtggc gggagaggtg cggatgagct 780  
 30 ggatctgcbc ggggagaggt tttgggacga agttggggtc gtcggccatg gtgagcccgt 840  
 cgtagtagct gtctgagcga gcctagtagt gcgctggacg agcagagccc agagtcgaga 900  
 cgagcgtgag caggagacga ggttcggagt gtccgaggag ggcgacgaga cgacgagcga 960  
 35 gcttgggaga agcgcgagca tgtccagcag cgtagtctcg aggccgagcag cagtagtaga 1020  
 gcacagcaat gaggcaggaa ggagcgcaag ggaggggaaag agcgcgacga agggtcgag 1080  
 40 tgatgaagtc caaggacagg gggaccaccc tcgcccgtt ctccctcgct ctccccacga 1140  
 agtgaccact tgtaaggctg gtaattcatt ccatacagtc tacatacact tgcagccatc 1200  
 cgcttcccct gcgatgccag tttcggtcac cgtgggactc cgatgcgatg atgcggccga 1260  
 45 gttggcttcc tcgaccgct ctcacagct cataccagcc tctcccagcc tgctaccgct 1320  
 ctctggctct gccaaacacc cactcgagca caccaccca accagcgaac tcgcccagcc 1380  
 50 tttgaaccgc aatggc 1396  
 <210> 109  
 <211> 1468  
 <212> ADN  
 <213> Rhodotorula graminis cepa WP1  
 <400> 109  
 55  
 60  
 65

ES 2 787 225 T3

	cgtgcagaag gaacccgagg aggtcagtgc gcggtcgaga gaggaaaagg agagagagac	60
	gcaccgagcg aggcatcgcg cgtgccagcg ctccagcctgt acctcgcgac tgctgatgaa	120
5	agcgtcgctc ccgcccgaca tgaccgccgc ctccgctcgac tggtcgccga cgacgtcgac	180
	gccggaagga ggtgcccc agccgcaggc ggccgcctc aggctcacgt cgacgtgccg	240
10	gacgttcgac gctcgccacg ggtcgtgctc gagggtttg acgacgagct cgtcgcgctc	300
	ggcgtgccg caccattgg cgaacgtgac gtccctggccg agcacctct gcgcgatggc	360
	gcgccacagc cggcagacgg acgcgacctg gcggtaggcg cacgtcgac aggtgacctg	420
15	gaggtgcgca aactggccgt ggacgaggag aggtcagag atgagcgagg tcgtgagggc	480
	gtggcggatg accttagga ggagctcgtg cggcacggca gagagggctc gcatgatggt	540
	gtcggcgcg gtcggcagtc tcgagagaga tgtgtagagg aagaacgatg tcgccagatc	600
20	ggtcgagcag gagccggtgc gaggcggctc gaggacctc gcggtcgagg accggtcacg	660
	gctggacgat cgaggagacg cccccctc gagcgcagcg gccagacgca agcagacacc	720
	tttgaggctg tactccaaaa cccggagcgc cggctcggga gccgtgtcct cgcaggatcc	780
25	tcggtcgaca gcgccgagtc ggagagggcc agccgacctc gggccgcccg acgccggcc	840
	gcagctcctc cggctccgacc tgcagctcat cccagcagat cgactttgag agcgaagccc	900
30	ccaggaagct gcctgagcga cctcgaggct tgggaaggtc gccgagccac ggtgggaga	960
	gcgagctccc tcacagtcga gaccggctcc aagtccaatc gcacactcgt agctgcaccg	1020
	caaaagtgtg tgcagagctg gagcgagcga ccgcgcgagg cgcgagggtc gcgagaaagc	1080
35	ggcgagcggg tgcgagtgcg cccgagacgc cgagagaggg cgcgagggcg agcgggcctc	1140
	gcgagccctc tggagcgtgc agagggcgcg gggaggagca gagtgagggg ggaagaccc	1200
	tccagagctg gcaggagcca acggagcgcg gaaatcagtg agatcgatgc ggttctcgag	1260
40	acgattcgac cgcctcgtc gtcaacgtcg cgcctcgtc cctctcctct tcccaccacc	1320
	tctccggtac ctctacacga gtgcgttctg tcccgagatc tgatctcgac gccgcacggc	1380
	actgactgac cgcctcctc gtctccctcg cccgtcccaactctcctt ccgacctccc	1440
45	acctcctcgc tcaaccctc tcgctcgt	1468
	<210> 110	
	<211> 1520	
	<212> ADN	
50	<213> Rhodotorula graminis cepa WP1	
	<400> 110	
55		
60		
65		

ES 2 787 225 T3

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

cgggaggatt gatgatcggg ttgcggtaca aggcgaggcc gaggacgagc tggagggcgc 60  
 cgaggacgag ggtcgagatg ccgagctgcg aggggacgag gccgccaatc gaggtcgtgt 120  
 cagacggccg taccgagtga tctcatagcg ggtccgccga cgcagaaaga cgaagagaaa 180  
 ctcacgacga gggggctcag cttgacaaaag ttgagctttg tcgtgcagcc agacgagctg 240  
 cgataggagc gagggtcagt gcgcccgctc tcggatgaat gcgcagtcga ggagggacag 300  
 gggcgcacca ggtgacgatg aacatgtcga ggacgccgta ctgcagcatc ttgcggaaga 360  
 ggtagaaatg ccacgccgaa gcgaaggtga cggcaaactg cgagagggac gaggtcagtg 420  
 cgaggggtccg caacagggag agggcactcg tcacggacca tcatccagcc caggtcgatc 480  
 tgtagctgat cagggaaact gtgccgggtg cgagagaggt caatgtcgaa gcttggcagc 540  
 tcgtcgagga agaagaggac ggcgagggac gcacgccatg aacttgatgt gacgctcgaa 600  
 gacgaccgag tagccgagcc acatcattgt gaggacgccg gcgcagacgt agtggagat 660  
 gaacctgaga gggcaagagg tcagtctcga aacgagggag gaagccggct cgagcaggac 720  
 gaggcgggcy caacggacgc accatctctc ggcggcgacg atattgcgag cccacacggc 780  
 cccgagcact agacaggtga tgttggcgac cgcgccgccg acggcgagag cgagaacgct 840  
 gtagacgaag tgcttgatgc gtcggtgat tggcgtcgt cggatcctcg cgatctgggc 900  
 cttggtgagc ggtggtggcg ggccaagcgg tgggcccagca gctgtgctca tcgcagcagc 960  
 ggtgccccgc aagagcact gtggagctcg agggagagga gcgcggcagg gaaagcgag 1020  
 accgaggag agcgagcgcg gacagggcag gcggaccgga cgttccggtg cggctcgact 1080  
 ggcgtgagc aggagcaggc cgtcggcggg agcagccgtg tccggcggaa agagccaggc 1140  
 gcgcgagcgg ggcggagcag acagcggcgg tccgagcgcg cggggcaggt tcgacgaaag 1200  
 tcgggctcgg gtcaggctcg cgcgagcgca tgagatgccg tcgagcagac ccatgtacag 1260  
 agtcgagcga gagagcgaag tgctggaag gagagtgtc caagagtga gcgccgtgga 1320  
 gatgagacag atgatggcga acctcggcca cagcctctcg gtcctgccac agcagctctg 1380  
 tgagtctccc tgaccgccca gcccgcgctt cagaactcac agaccaccta cacagactcg 1440  
 cgcaccagct cgaaccgcgc cagaccaccg cctcgcgcc tccccacctc gactgcttcc 1500  
 gaacctcaca agctcgacca 1520  
 <210> 111  
 <211> 1210  
 <212> ADN  
 <213> Rhodotorula graminis cepa WP1  
 <400> 111

ES 2 787 225 T3

	gcgaggggaa gggcaggaga gtcgccgaa caccgggcaa ggagacaagg acaaagagcg	60
	agcgcgcacg caccgtctgg cggtcgggct gggtagggcg ggtccagta gcgctgaccg	120
5	aagagcgagg aggagaaaa gcgagcgcgg cgggcgtggg aaggaggca aggagggcg	180
	cggcgacgcc aaaaaaggcc agattcgcgg ggcagtgatc gagtgccgtc cgcgagctca	240
10	accagcgagc gctctctcgc gcgacgggt gcgcttcttt cggcagccga tgcgcaccgt	300
	tcagaagcac gtcgcccgca ccgagagcgc ctctcgcga gcctgtgacc acctcgtcga	360
	cccgttccc gcggcttttc tcgcccgcct ggaccgccgc tatcagatcg tgcccatgag	420
15	acaagcgact cgtcgaagg acgacgatct cgtagtactg ggtccctgcg caacgctcag	480
	ccgtccgtcc ccgtcaaagt gcttcggcgg aggggaccgt gcgagagacg cccaagttgg	540
	cctcctcaag tcggtagatc cagcttaacg ctatcaaggg ttgcatggtg tagttggtca	600
20	tcacgtcagt ttaacattca gttcactgaa ggtcctcagt tcaaacctgg gtgcgatcac	660
	ctttttggct cggcggcatt gcgcccttac acccgcacgg gtctacttcc ctttgcaagc	720
	gaccaagcga agcatcctct cgcctgtaa gctgccggcg aggaggtcag acgggttggc	780
25	gggccgtcga aggtcggctc acctcaacg ctgccggctg accacgccag gcgagctatc	840
	attgctttga aagcttcgaa aacgcccagg catgcacaga aagccgcccg cgagaggctc	900
30	aagttggcgc cgagctcggg tcgagagacg acgacgacgt gggagctccc tcgcctctcc	960
	tcctttctct cccaccccat cagcccaagt gagtgcctcg ctcttccgca agggtcagcg	1020
	cacgcggttc tcgcgcacag ggcagcgcgt gcgctacca gggcccccg ttcgcccggc	1080
35	gagttggcac tgacgaggtg ccttgcccc ctccgctccc ctccccttg gcctcctctc	1140
	tcgcacgcac actctctccc tgcacccctt gcaccttccc gacactctcc cccccttcc	1200
40	caccgtccga	1210
	<210> 112	
	<211> 1495	
	<212> ADN	
	<213> Rhodotorula graminis cepa WP1	
	<400> 112	
45	actacctcgt caccgaccag ggtgcggtcc ccttctcgcg gcgtccacag ccacgacgtt	60
	gcgagtccct aacgcgtccg ctccccctg caggtggtcc tcgcacggc gcctacacgc	120
	agacatcgtc gacgccgtcg caccgacccc actttgccgt ccagaccctg tcgcgcttcg	180
50	gcctcgcgta cctcctcgcg tcgtccaaca cgctcaagga cacgtggctg agcgtgtgcg	240
	cgcccgccgg cgccaagggc cccgaaccgg acgtcgacga cctcgagctc gagaagcgcg	300
	agcaccgcga caagtggctc ctcgggcgca tcatgggcca gggcaagcag gactcggcgc	360
55		
60		
65		

ES 2 787 225 T3

	tcggggacgc agtcgctgtc gtgagtcctc tactctcggc cgttctcgag acttgggggtg	420
	cgagattgac cttgcgctcc cgctctcgc ttgcagcaat tccacaagca ctttccgcac	480
5	ctgcgctcgg cgcacctctt ccccggttt gtctttacgt gcgtccctct cgttcccccc	540
	tctctccacg tgcgccacca gcctgactcg cccctcgcct gccccgtccc gcagcaacgc	600
10	cctcgctcgg acctcgctcg tcccctcgcc gatcctgtcg ctgtacaacc tcgtcgggcc	660
	cctcgcgggc cgcatcctgc ctttggcaa cctgcccag acgtacgccg acgtgcccgt	720
	gtacgtcggc gccaacccgg cagcgcgcag ccaagggctc gagtactgca acgagcgcac	780
15	gaagccgctc gggagcccgg cgtgggccga gggcgcgacg ggcgcaaagg tgtgggacgg	840
	gctgagggcc atgatcgagg agtgagctgg tggcgggcg agcgaggagc cggagaggag	900
	ggcggaacg tgtttgagaa ggtcgcgctt tgctcgtcgg tcgcgggcgc agccgtggct	960
20	gtagccagtc tcgctttgca gtgtcactct tgtacatagc tgagcaaggc cttagcgtcgc	1020
	gagagagctg cgctgtggcg cctggctcag gcccgagagc gtcgcgctca gggcgagct	1080
	gctcgcggct caccaagggg ctcgagcggg gcgcgctcga caggggaccg agagctgcag	1140
25	gagacagacc ggaggaaaaa gctctggcga gcgaggagcg gggccacact gagtctgggg	1200
	aagcgacgga cgaggatgag cgcatccact cttgagttc gccgaggcgc gagctggcgg	1260
30	tcgacaaccg agcaagctcc tcctcttccct ccaccacact cgcccctagc acacgtgagt	1320
	ctcgcctcct cgccactgtc gaccagcaca cgctcgtcca ccgccctgtg cgccctgtgc	1380
	ggcttgccgt cgagcgaggc cgcgggtcgg gtctctgcca cccgaggaac catcgatgtc	1440
35	gctgacgctt cgctcctcgt cctcctctc cteccaccgc ccgcagctac ctaca	1495
	<210> 113	
	<211> 1634	
	<212> ADN	
	<213> Rhodotorula glutinis ATCC 204091	
40	<400> 113	
	cctgaagctg tacatcgagg tggacgacag cgcgggcaag gatgcgccag caatcgtttg	60
	tgggtcgtt ctttcctcgc agcacgcttt tgtcggctcc ctgatcagca cacaagctaa	120
45	ctaacgctct ggtttcgtcg gcagtcatgc acggccttgg ctcgtcaacc tcgttctggg	180
	aagcgccctt ctcccgtca aacctgtcct cccgcttccg cctcatccgc tacgacttcg	240
	acggccacgg tctctcggcc gtctcgtccc tcgacgcagc agatgacggc gccatgatcc	300
50	cgctcgacga cctcgtcggg gacttggcgg ctgtgatcga gtgggctggg gtggagaagg	360
	ttgcgggagt tgttggacac tcgatgagcg ggctgggtggc gagcacattt gcggccaagt	420
	accgcagaa gctcgacaag ctcggtgagt cgcattgaac cttcctccgc cgtctcttct	480
55	ccgctgacga ttcgctgact tggccctgct tctcgcgcag tcctcctcgg cgcaatgcgc	540
60		
65		

ES 2 787 225 T3

	tctctgaacc ctaccgtcca aagcaacatg ctcaagcgag ccgatacagt cctcgaatcc	600
	ggcctctcag caatcgtcgc acaagtcgtc tccgccgctt tgtccgacaa gtcaaagcag	660
5	gactcgcccc tctcggcagc gatgggtgca acgctcgtgc ttggaacgga cccgagaggg	720
	tacgcgccgc cgtgtagggc gcttgccggg gcgaaggacc cggattactc gagcatcaag	780
	gccgagacgt tgggtgctt cgcttgttct ccttcctctg cttttctccc agcaactgac	840
10	gcaagcgtct gcaacacagt cgtcgcaggc gagtttgact acctctcgaa caaggagacg	900
	accgacgcgc tgggtcaacga catcccgggc gcggagaagg tccagatgga cagtgtcggc	960
	cactggcacg ccgtcagagga ccccgcttga ctogccaaga tcctcagatgg gttcttcttg	1020
	caggggaaat gaggttggga aggggggata gactggggag aacggcaggt gcgtacgcag	1080
	cggacgtcgc tcgggaggac tttttcgggg aggatattcg ctgactgact ccgacgtcgc	1140
20	tttctcctt gcagtatctt cagaagggat gggaggaggc gaactgcaag ggtaatgaac	1200
	gagacaacgc cgagggagga agcgcgggaa ctctcggggg cgaagaagga gtggtgtctt	1260
	cgccagcgaa cagcttccgg ggtgggttg acagcgccag tagaattcca gcgtcgcaac	1320
25	agagctctag tcgaccgga tcacccacaa ggacgagagc gggtcgcgcc ttgtccgctt	1380
	ccccatcctc gtcctgctct tgctctcttc cctaccacac tctcccgtt gcgggctctc	1440
	tttctcgtt ggcgctcctg ctaccgctac tctagactct cctagtctcc ctgcacaacc	1500
30	atccctatcc cctccgcctc tctcgcacac cccccacagc ttcgttcccc aacttcaact	1560
	ccgatgccgt gcgtcgcctc cttttgcctt ggcgggcccg cgctctctc cgaggacaac	1620
35	tactgattgt ggga	1634
	<210> 114	
	<211> 1438	
	<212> ADN	
	<213> Rhodosporidium toruloides	
40	<400> 114	
	ggcgcgctag tctgcagaac tacgccctct cacacccaac ttccgactcg accggcggtg	60
	cgagcacgac ctacttctac tgccctgcat cgacatccgg gcgggtcgtt gcctaccctg	120
	tgcgttctgc gccctccctc gtctcgggag gcagtgtctg acagaagctt tgcgcgcagt	180
45	accccgctca gatgcaactc tacgcaacgt tcggcacaga agtcgccaag ctccgcgcat	240
	cgccgcctca agctctcggc ctgcccagac gtgtcgtcta ttacgaggcg gagaagctcg	300
50	agttgccggc tttgccagcg gcggtcaagg ttgaggtgga gacggagaag gcgggagtag	360
	cgggggagga caatgaggcg aagggtgaga tgggtcgtgt ggagactctt acggtggagc	420
	aggaggagat tgaattgggc tcgggagtcg tgcagattga ggagtcggtt ctcgtcaagc	480
55	tggaggtcag cggctgatcc ttccgttcgt tgcgaagatc gtctgcatgt ttcgcttctc	540
60		
65		

ES 2 787 225 T3

tcaatgacac aacctggaga gcgctcccgt cagcgagaat cgaggacatt ccgcagctcg 600

tgagcaagcg gaggtgagcg gctcccctcga aagctgcgcc tcttcagacg gcttggtctc 660

5 tcttgctctg gtgggctggc ctgacatgta atgtgctccg ccgcaagtcc gtcgtcggtc 720

tcaattcgac gttgaaaggg catagcgcaa ggaagaacc tctgctgaca tgcagaatta 780

ctggctcgcc tgctccttcg tctactggaa taagtctctg ctctgtaag ccccaacgtc 840

10 gtttttcgac gtttgtaagg cgcaagaggt gctatgggct acgcaggaag ctgagaggac 900

atagaagtgc ggggaggaac ggcgcagagc ggcagttgag gaagcatgag gaaagcgaga 960

15 cgggccagca tctgcagcgc caatccgcaa tctcctggtt gagcctgcac cggaaagcgtc 1020

ggaacagtat gcgcagagtc gaacgcaagt aagaaagacg caccctcaca ctgcgttact 1080

tcgagccata caacggatca aagctgcgag tatctcggct tgtaagggcc gaaagcaac 1140

20 ctccggagatg gacacgtcac atcaccaact tatcgatctc ggccgtcgac gtcgcagaga 1200

ggcgagagaga agcgggtaag gagggaaaca acccctcgag agcatgatcc gaccgaatct 1260

gcagcgcagc aagccgttac aagcccgcct cgagcgcagc tcgggtccag ccgggggagc 1320

25 aaacgcgcga ggctgattcg tgagcgaagg aagccgcacg gacaagtctg ctcccctttg 1380

ccctctttcc catcaccggt tctcgcctta ccogctcaga acaacaccag atcactca 1438

30 <210> 115  
<211> 660  
<212> ADN  
<213> Ustilago maydis  
<400> 115

ggcgcgctag tccttaaaac tgaaggcggg aaacgacaat ctgatccaag ctcaagctaa 60

35 gctctagtga ttagatcttg ctgataggca ggtttgcttg gagaatgggg ggaaaagact 120

gaccgaagaa acagcgagat ctagaagtga taagcggaaa gaatctgact tgctgtgatc 180

40 agcagccaat ttttttttcg tttttttttt ttcactccac atcgtcgtgc gtgcacggtc 240

tgcatgtgta aattgtattc atcgaaagcc acagttgaat acatcagccc gatgtggatt 300

tcgaaaacca attaactctg gaattcacgc gctcagatca gtccatagag tcgacttcgg 360

45 ctgtttccaa gagcttcttc tctgcgaggt ggttgcccgt gtttctcgtc gggaaaaaag 420

gatcgattat tattcgcttc tacctcgttc gcacccttgg cctgctgaag gaaacagcgc 480

cgagactcgg tcacgggttc tgggctccgt gttgatgctg ggacggcgca aagtggggcc 540

50 cgcgcactct tcgagccaag gacctcactc ttcaagaaca agcgtgctcg ccatcgtctt 600

cttctttctg ctccaccatc gaatctttct ttctcgtttc gaaacaaaa cactcttcca 660

55 <210> 116  
<211> 951  
<212> ADN  
<213> Rhodotorula graminis cepa WP1  
<400> 116

60

65

ES 2 787 225 T3

	cctacgtcta cgtcaagggc aatgccgacg tgaccaaggc catcggccag gacctcgcct	60
	tctttctcgg tccctgtcag ctccggcgtgc gtcccgccgc tctctctctc tctttctctc	120
5	ggccgcgcct cactgatcc acgactcgt actgaccctt gcgaatgtgc gcgccgcag	180
	cccaacggcg tcgagaaggt gcaccgcctc ggcgacctga cggcgttoga gaaggagctc	240
10	ctcgagggcgt gcctcggcga gctgcccggg tccatctcca agggcgagtc gttcatccag	300
	ggctccaagc tctgactcgc cggcgcacgc acggggcgcga gccacaaggc gaggatgtga	360
	gaggagggcgt ttcctccacc ttggacccca tctgccgcct ccctttctct ctctttcttt	420
15	cccttctctc ctctctctct ctctctcgtt ctctctcttc tgggcctctc ggacctcttc	480
	ctcgccgtcg actcgtgaaa atgcagtgcg cgtttctgta ccttgcctg cgagagagat	540
	ctggttctgc gagggtgagt cgttgccctt gccgtggcac gcctcgcgc agcgagagag	600
20	aagagggcac ggtccaggac gacgacgacg aggaggaagc gcaaaaggcg agacaccgag	660
	tgccatcgat tccccgctcg aacctgctca cggctgtcga aggcggtgcg ccacgggtgct	720
	tgccgggagc aaagcaagct ggctgctcc tcttgaactg gttcgagtcc gtgagggcgg	780
25	cgacgagaac tcaggcgagg tgctcgcgtc ggaacaagcc gggcttggg tcgagggagc	840
	gagagcgagg cagcgcctgc gtcgccgagg caagagcggc atcgacaagt tggcccgtcg	900
30	cctctcgtc cctcttctcc tctcccacc accaccttc tccagctoga a	951
	<210> 117	
	<211> 2136	
	<212> ADN	
	<213> Rhodotorula glutinis ATCC 204091	
	<400> 117	
35	ccaccatctc ctcgtcgtt ctccctctc cttcggcgc cacaccgctt cgcagggctc	60
	acggactgct cacatcgttt gtgtgcgtcg ctgtgcatgt ccacgcacca ctcccagccc	120
40	ccacgagcgc ctcaaaagac gcggacgcag acgcccgcg aacgacggca cgcctctctt	180
	ctcactagcg cgacgaacca gctgcgacga ttogtgcgt tatgttagcc ggacttctgg	240
	cttgccttgc gctgctcgt ccgtcttctg gtgcggatcg gctcgatggg ggtttgctcg	300
45	tttgcctggg gacggtcgc tctccctct cctcttcaact cctcgttagc tttctacgct	360
	cattggttct gcgaaccatc tacatcacgc tcgctcgtca tgctcgtact acgatcaaca	420
	cccctgctcg tcgtgcttcc cctcctctcc gtcctctcgg ccgctccag cgacttgccc	480
50	agccaacttc ccccgcacgc cggtagtct cccacacttc cttgcgacct caaccagca	540
	tctgacatcc gcatcacgca gccctcccgc cttcccactc ctccctcttc accgactcct	600
	cctcctctc cctgattec tcgtccctca aagccccgca gcctcttccc ttcaaatca	660
55		
60		
65		

ES 2 787 225 T3

agcgcccccg ctcgctcgaa caagtgcagc agaacctcgg gaagaggctg gcgaagcgcg 720  
 gcgaggaggg gagtaagacg gagaggggtgc cgtttggtca gaggagtgcg acggcggcga 780  
 5 gtgcggggtg acaagtgga gcggggacgg ggagggcgac gcagcgcgtt acgggcggag 840  
 gaagcagagg tgcagagagga ggcggagggga gtgtcgcggc tgctcagcct gtcccttcga 900  
 ctaccagac ggtcgagaca ggctctaaga tcgtctcgac tggctctctg accgtagcgt 960  
 10 cgccgtcgac ggcagatgga ggagggggga cggtcacca ggtcgagacg gcctcctcag 1020  
 gggattatgat caccagcacg gcgggagcgg cgagttcagc ggccggcgtcg gacgtcgcta 1080  
 15 gcgcacaggg agcggagggc acgtcgagta cgagcatgat cagcggagga gcggcggctg 1140  
 gcgggagttt aagcaggatg ctggcgggag gagttgcggg tcagaccctg atcctcctcg 1200  
 tgcggtgagc aggcgaagcg aggagctcat gtagatacag catagacagt atatatgcc 1260  
 20 aggatagcct gcaacagccg ccggtogggt tattccattg tcctcgacct catgogaagg 1320  
 cgagctctgc tcgtcagctg gccaaagctgg ccagcagacg agcgttgggg tggcggaacg 1380  
 ccaacggcat ggagtaaagc agccgtgagg atgacggagg agctcgggcg agtgatggg 1440  
 25 gattctagca ggaacagcag agcggcgagg aggagaggaa ccggaagcac agtctcgtgg 1500  
 ccgcttggtg cagatcccag tgtcgtaga gtgctcgtcg tcatcagagc gaggtaaaa 1560  
 agcgatgcc tgaagaacga tgagcgaatg agtcgaagcg gcgtctaccg gtgaactcgg 1620  
 30 ggtgtggcaa atgagcgaga cgaggagtgc ccgccagagt tgccacgtcg accccacgtc 1680  
 ggaatcgacg ttgatagagt gaacgaagcc attgcagacc ccagaagggt gccatggtgt 1740  
 35 ggaagcgagg gcaggagcga ggggagaagg cgaggaggag gaggggctgg ggaagcccgt 1800  
 ccgggaatgg cgcagctggg tgccggggat gtgcgcgagt ggccggaggag tcgagcgtga 1860  
 gaggttctgga acacggggcg cgcacaaggg tcgagggccg tgacgagttc gccgggcggt 1920  
 40 ggtcgggctg agggcgagcg cgcgttgggg acgacgacgc ccgacgccct cgtctctcgt 1980  
 cctcaccgct tcccggagaa ctttctgta ctctgcttct cccttcacac tctcacacc 2040  
 actcacacac ccttccatcc acacacaagc tatccgcaca cctctcacac ccgaccccag 2100  
 45 ctgccccat cctcttcgca cccggctcat cgaaaa 2136  
 <210> 118  
 <211> 997  
 <212> ADN  
 50 <213> Sporobolomyces roseus  
 <400> 118  
 ccgaagttat acctcagagg tctcaaaaac gaaaaagtca tgcaagaatc tcctttgacg 60  
 tgagggttat ttctctcct ctagtagtct acgagaatcg caaagatcgg aaaactgatg 120  
 55 catctttgtg ttcacgggtt agcgatttga tcttttcgat tcccaaaatc gtatcgttcc 180

60

65

ES 2 787 225 T3

	tgatgcaggg aactacgctc aaagccggca ctctgatcat cacgggagtg agttttgagc	240
	tctccctcta tgagagtgca aggttogtgc ctgatggtgt aatccgctca tgccttcccc	300
5	tctaccttct cctttgtcca ttctctctac tacggttgtc acatcttctt tctccgacag	360
	accccgcacg gaattggagc gtactcgaat cctccggaat tcttcaagga cggagacgctc	420
	ttcagggtcg agatctcggg aggcacgagg agtttgggtc acaagatcga atatgaaaag	480
10	tagataatcc gttactcagg tcaatggtat ggcttcgaag atgctggaat cagccggaaa	540
	gcaaagctgg agagaaaaat cgagattgcg aaacgtgcga tgtcatttcg tttcgagctc	600
15	gcaaccatct cgtatccctc tgagctacat acaaacgtca ctacggcctc ggagtgactc	660
	cctgcgagcg gattgaagga gatcacggtc gaatcagcta gaccttogca acgttttcgc	720
	gctcgcacgt tcttatcgat ctactgagat tgactcgaaa aagtcttctc tcacggtcga	780
20	ttgaactttg aatgaactct caggttgccg gagagccaat acgagccgac cagaggcaat	840
	tcggagcttc ccggaacggt ccaaggagag ggattttccg agagattacg attgagagat	900
	agaaaaaagg ctagctttcg atttcgagag agattacttt caagttcgtt gcttccaact	960
25	cttgctccaa cccctccac tccttctcta caaaaca	997
30		

## REIVINDICACIONES

1. Un método para ajustar el nivel de producción y la composición de carotenoides en un huésped fúngico que comprende:
  - (a) manipular genéticamente uno o más polinucleótidos en la biosíntesis de carotenoides en un huésped fúngico, en donde uno o más polinucleótidos se seleccionan de (i) los polinucleótidos expuestos en las SEQ ID NOs: 1, 3, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19 y 20 o una secuencia homóloga que comparte al menos el 75 % de identidad con los mismos o (ii) uno o más polinucleótidos que codifican uno o más polipéptidos expuestos en las SEQ ID NOs: 2, 4, 6, 12, 15, 18 y 21 o una secuencia homóloga que comparte al menos el 75 % de identidad con los mismos, y en donde el huésped fúngico es *Rhodospiridium* o *Rhodotorula*,y
  - (b) cultivar el huésped fúngico para producir carotenoides, en donde los carotenoides se seleccionan del grupo que consiste en licopeno, beta-caroteno, gamma-caroteno, toruleno y torularodina o derivados de estos, mediante lo cual se ajusta el nivel de producción o la composición de los carotenoides.
2. El método de la reivindicación 1, en donde la manipulación genética es la regulación negativa del uno o más polinucleótidos, en donde la regulación negativa se compara con un huésped fúngico que no tiene la manipulación genética.
3. El método de la reivindicación 2, en donde uno o más polinucleótidos se regulan negativamente por ARNi, un represor transcripcional artificial o un promotor débil.
4. El método de la reivindicación 1, en donde la manipulación genética es la sobreexpresión de uno o más polinucleótidos, en donde la sobreexpresión se compara con un huésped fúngico que no tiene la manipulación genética.
5. El método de la reivindicación 1, en donde la manipulación genética es la inactivación total de una o más funciones enzimáticas de uno o más polipéptidos codificados por uno o más polinucleótidos.
6. El método de la reivindicación 5, en donde la manipulación genética es la delección de todo o una parte de uno o más de los polinucleótidos, en donde el uno o más polinucleótidos se seleccionan de (i) los polinucleótidos expuestos en las SEQ ID NOs:10, 11, 16, 17, 19 y 20 o (ii) uno o más polinucleótidos codifican uno o más polipéptidos expuestos en las SEQ ID NOs:12, 18 y 21.
7. El método de la reivindicación 6, en donde la delección se realiza mediante recombinación homóloga, preferentemente en donde la delección se realiza mediante el uso de una nucleasa artificial seleccionada de una nucleasa de dedos de zinc o un complejo Cas9-ARNg.
8. El método de la reivindicación 2 o 3, en donde la manipulación genética incluye además la sobreexpresión de uno o más polinucleótidos diferentes al uno o más polinucleótidos regulados negativamente, en donde la sobreexpresión se compara con un huésped fúngico que no tiene la manipulación genética.
9. El método de cualquiera de las reivindicaciones 5 a 7, en donde la manipulación genética incluye además la sobreexpresión de uno o más polinucleótidos diferentes al uno o más polinucleótidos que codifican uno o más polipéptidos que tienen pérdida de función enzimática, en donde la sobreexpresión se compara con un huésped fúngico que no tiene la manipulación genética.
10. El método de cualquiera de las reivindicaciones 4, 8 o 9, en donde el uno o más polinucleótidos se sobreexpresan mediante la introducción de una o más moléculas de ADN en el huésped fúngico, en donde cada molécula de ADN comprende uno o más constructos y en donde cada constructo comprende un promotor heterólogo unido operativamente a uno de los polinucleótidos unidos operativamente a un terminador transcripcional.
11. El método de cualquiera de las reivindicaciones 5 a 7, en donde la manipulación genética incluye además la regulación negativa de uno o más polinucleótidos diferentes al uno o más polinucleótidos que codifican uno o más polipéptidos que tienen pérdida de función enzimática, en donde la regulación negativa se compara con un huésped fúngico que no tiene la manipulación genética.
12. El método de la reivindicación 9 o 10, en donde la manipulación genética incluye además la regulación negativa de uno o más polinucleótidos diferentes al uno o más polinucleótidos que codifican uno o más polipéptidos que tienen pérdida de la función enzimática o uno o más polinucleótidos que se sobreexpresan, en donde la regulación negativa se compara con un huésped fúngico que no tiene la manipulación genética.
13. El método de la reivindicación 11 o 12, en donde el uno o más polinucleótidos son regulados negativamente por ARNi, un represor transcripcional artificial o un promotor débil.
14. El método de cualquiera de las reivindicaciones 1 a 13, en donde el huésped fúngico se cultiva en un medio que comprende al menos el 5 % de una fuente de carbono a una temperatura de aproximadamente 25 °C a

aproximadamente 35 °C, preferentemente bajo iluminación, preferentemente en donde la fuente de carbono es glucosa, manosa, glicerol, sacarosa, xilosa o sus combinaciones.

- 5 15. El método de la reivindicación 18 o 19, en donde el medio comprende 30-100 g/l de glucosa, 1,5 g/l de extracto de levadura, 0,5 g/l de  $(\text{NH}_4)_2\text{SO}_4$ , 2,05 g/l de  $\text{K}_2\text{HPO}_4$ , 1,45 g/l de  $\text{KH}_2\text{PO}_4$ , 0,6 g/l de  $\text{MgSO}_4$ , 0,3 g/l de  $\text{NaCl}$ , 10 mg de  $\text{CaCl}_2$ , 1 mg/l de  $\text{FeSO}_4$ , 0,5 mg/l de  $\text{ZnSO}_4$ , 0,5 mg/l de  $\text{CuSO}_4$ , 0,5 mg/l de  $\text{H}_3\text{BO}_4$ , 0,5 mg/l de  $\text{MnSO}_4$ , 0,5 mg/l de  $\text{NaMoO}_4$ , en donde el pH del medio es de aproximadamente 5 a aproximadamente 7.

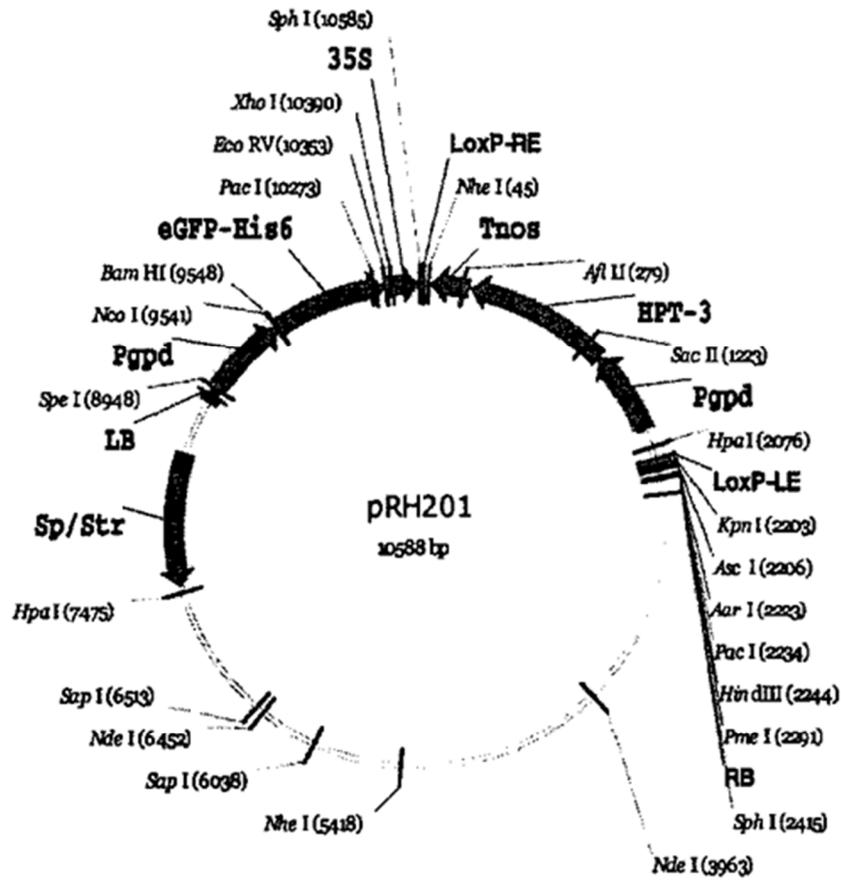
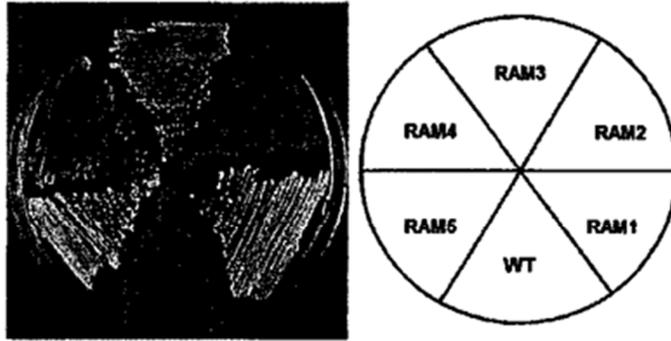
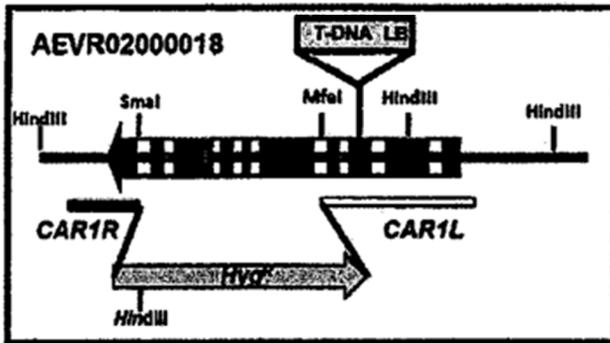


Figura 1

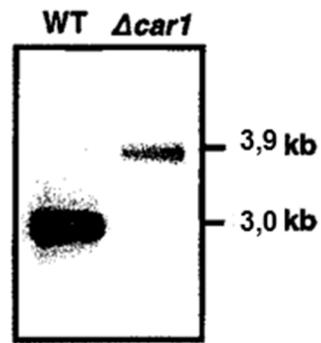
A



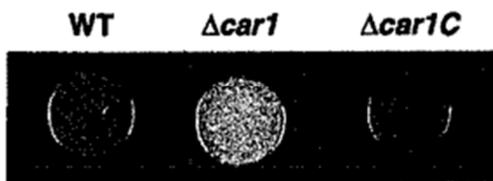
B



C



D



E

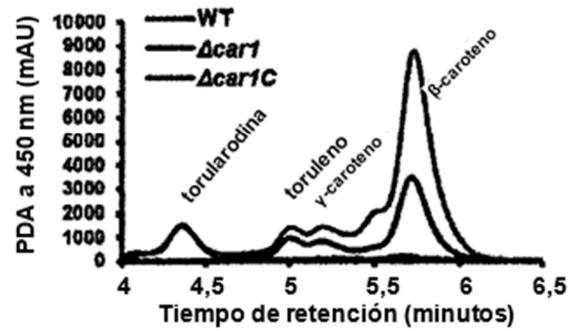
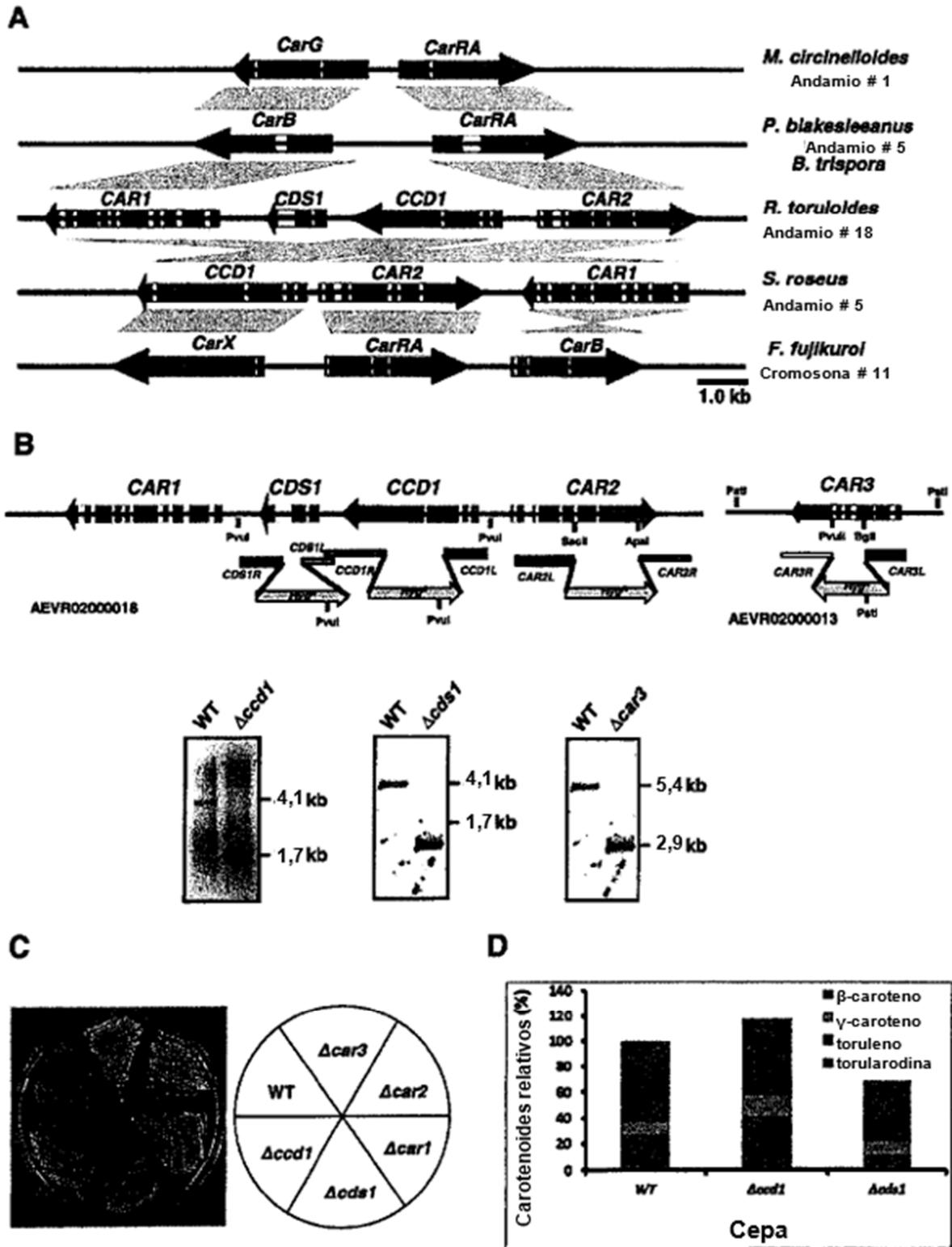


Figura 2

Figura 3



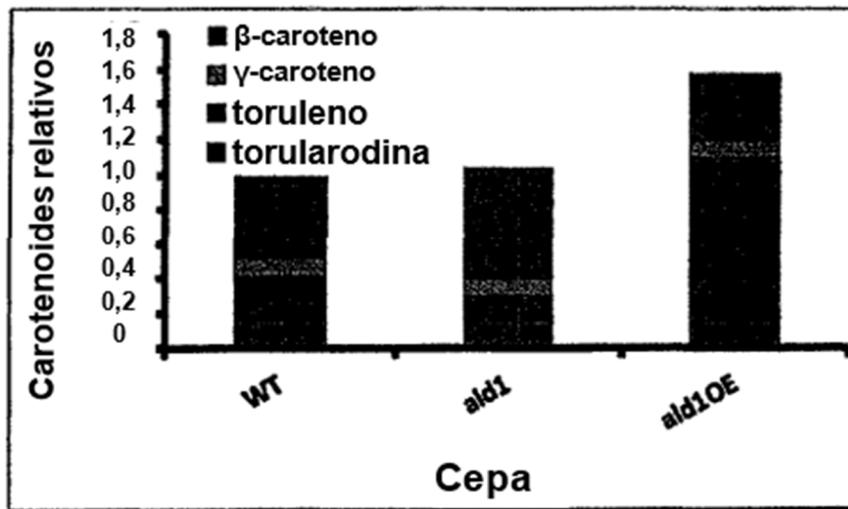


Figura 4

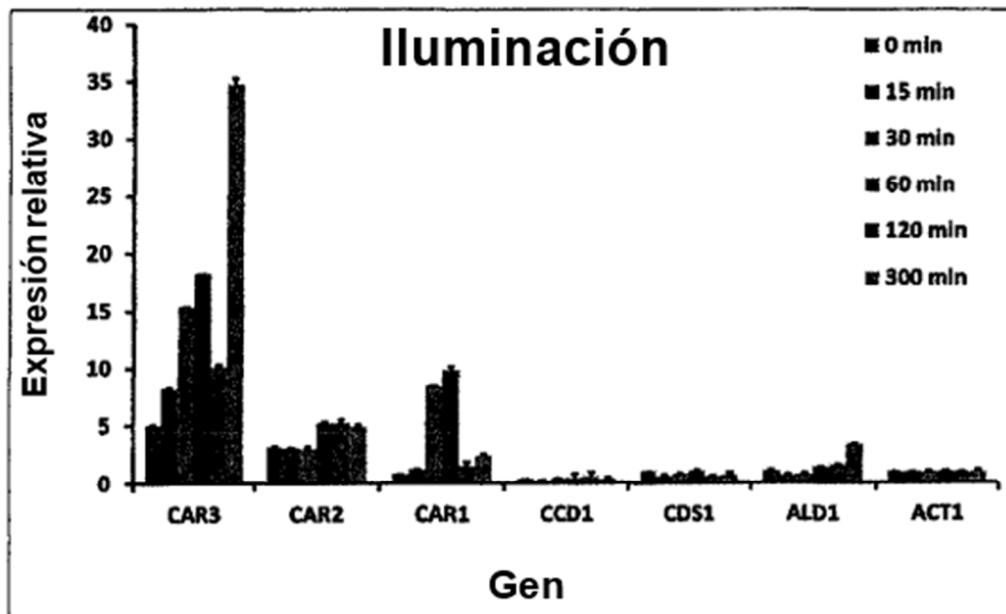


Figura 5

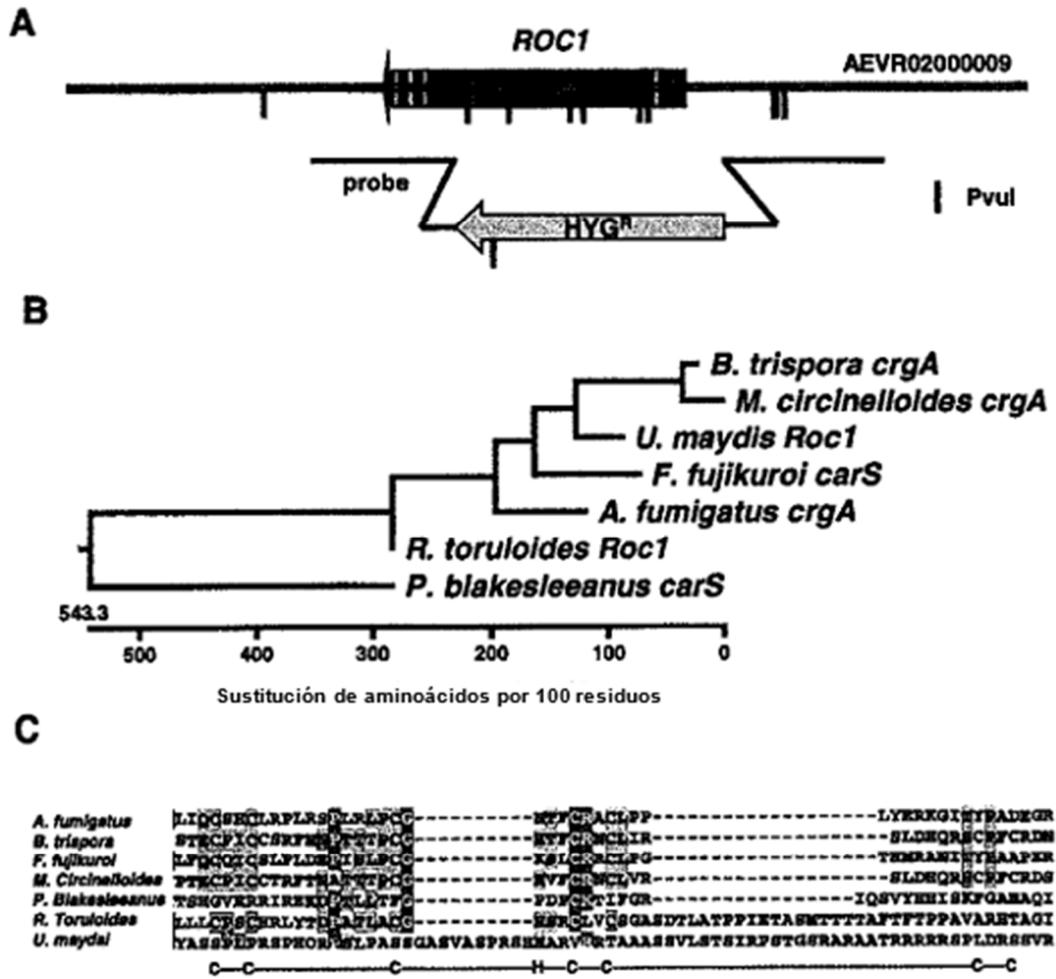


Figura 6

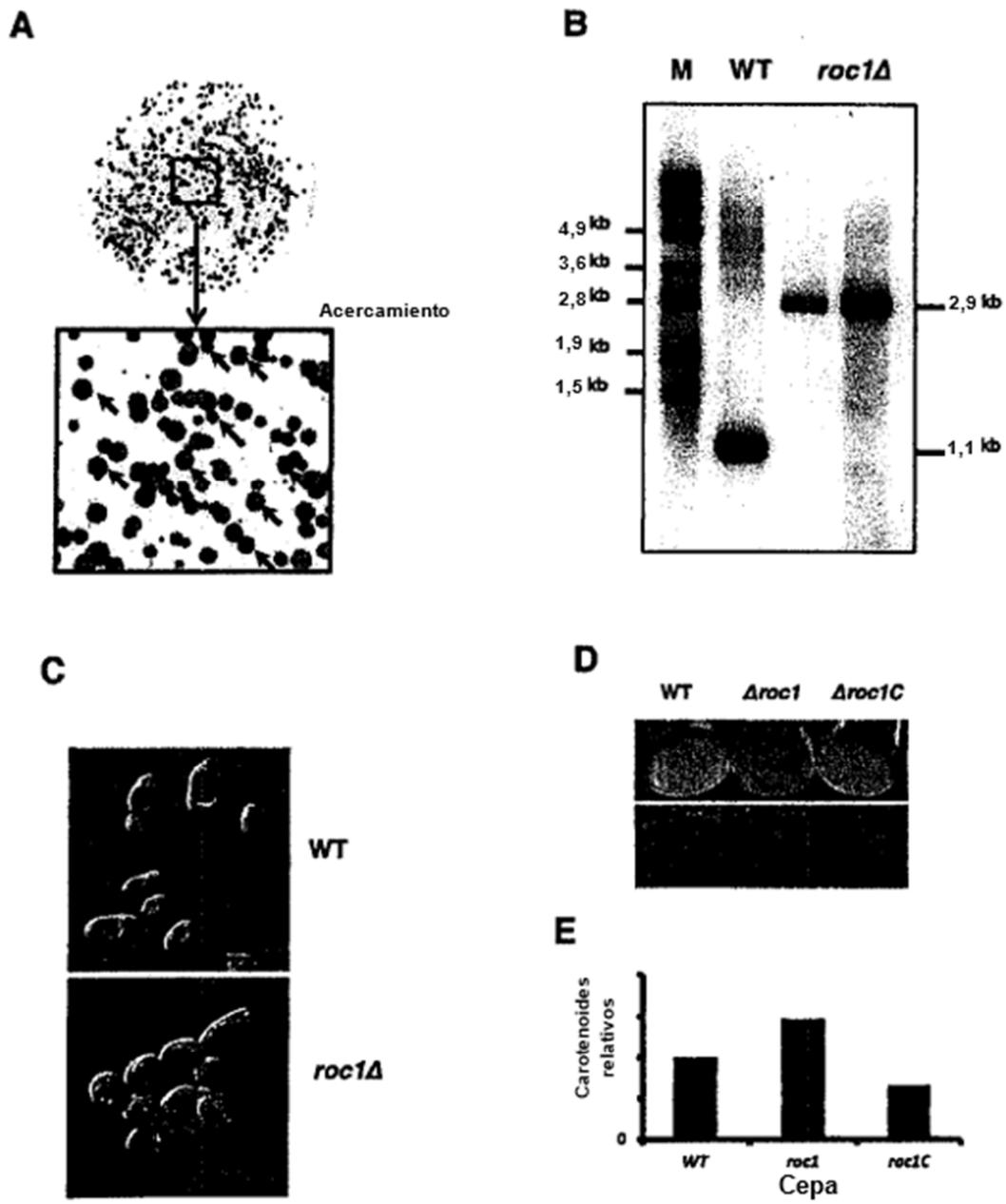


Figura 7