



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 305 501**

51 Int. Cl.:

**C12N 9/10** (2006.01)

**C12N 15/81** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **03757888 .7**

86 Fecha de presentación : **23.09.2003**

87 Número de publicación de la solicitud: **1543115**

87 Fecha de publicación de la solicitud: **22.06.2005**

54 Título: **Gen escualeno sintasa (SQS).**

30 Prioridad: **27.09.2002 EP 02021619**

45 Fecha de publicación de la mención BOPI:  
**01.11.2008**

45 Fecha de la publicación del folleto de la patente:  
**01.11.2008**

73 Titular/es: **DSM IP Assets B.V.**  
**Het Overloon 1**  
**6411 TE Heerlen, NL**

72 Inventor/es: **Hoshino, Tatsuo;**  
**Ojima, Kazuyuki y**  
**Setoguchi, Yutaka**

74 Agente: **Isern Jara, Jorge**

ES 2 305 501 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

Gen escualeno sintasa (SQS).

5 La presente invención describe un gen de utilidad en un procedimiento para aumentar la producción microbiana de carotenoides.

10 El carotenoide astaxantina está distribuido en una amplia variedad de organismos tales como los animales, algas y microorganismos. Tiene una fuerte propiedad de antioxidación contra especies reactivas con el oxígeno. La astaxantina se emplea como agente colorante, especialmente en la industria del pescado de granja, tal como el salmón, debido a que la astaxantina comunica a los animales un color rojo naranja característico y contribuye a atraer al consumidor al puesto del mercado.

15 Uno de los pasos de la ruta caratogénica de p. ej. la *Phaffia rhodozyma*, a partir de un metabolito general, el acetilCoA es la isomerización del pirofosfato de isopentilo (IPP) a pirofosfato de dimetilarilo (DMAPP) por la acción de la IPPisomerasa. A continuación, el IPP y el DMAPP se convierten en una unidad de 10 átomos de carbono, el pirofosfato de geranilo (GPP) mediante la condensación de la cabeza a la cola.

20 En una similar reacción de condensación entre el GPP y el IPP, el GPP se convierte en una unidad de 15 átomos de carbono, el pirofosfato de farnesilo (FPP) el cual es un importante sustrato de colesterol en animales y ergosterol en las levaduras, y de la farnesilación de la proteína de regulación tal como la proteína RAS. En general, la biosíntesis del GPP y FPP a partir del IPP y el DMAPP está catalizada por una enzima llamada FPP sintasa. Por otro lado, en los procariotas tales como las eubacterias, el pirofosfato de isopentilo se ha sintetizado por una ruta diferente vía 1-desoxixilulosa-5-fosfato a partir del piruvato el cual está ausente en las levaduras y animales. La mayoría de los genes implicados en la ruta del mevalonato y el gen de la FPP sintasa fueron clonados a partir del *P. rhodozyma* (EP 955.363).

30 La patente WO 96/09393 describe una secuencia de ADN aislada como en ADNc a partir de especies de *Nicotiana*, p. ej. *Nicotiana benthamiana*, la cual tiene una secuencia de nucleótidos representada por SEQ ID NO:1, la cual codifica una escualeno sintetasa nativa capaz de conducir la condensación reductora de dos moléculas de difosfato de farnesilo para formar escualeno, constituyendo el primer paso efectuado en la biosíntesis del esterol en eucariotas.

35 Shimada *et al.*, Applied and Environmental Microbiology ("Microbiología aplicada y medioambiental"), Julio 1998, páginas 2676 - 2680, describe una mayor producción de carotenoides mediante la levadura de alimentos *Candida utilis* a través de la ingeniería metabólica de la ruta isoprenoide.

La presente invención se refiere al asunto del objetivo de las reivindicaciones 1-6.

40 La presente invención describe en la presente un nuevo fragmento de ADN que comprende un gen que codifica la enzima escualeno sintasa.

Más particularmente, la presente invención describe en la presente unas regiones reguladoras que contienen ADN, tales como un promotor y un terminador, así como también el marco abierto de lectura del gen del escualeno sintasa.

45 La presente invención describe en la presente un fragmento de ADN que codifica el escualeno sintasa en *Phaffia rhodozyma*. Dicho ADN significa un ADNc el cual contiene solamente el marco abierto de lectura flanqueado entre fragmentos cortos en su región 5' y 3' sin traducir, y un ADN genómico que contiene también sus secuencias reguladoras tales como su promotor y terminador los cuales son necesarios para la expresión del gen de la escualeno sintasa en el *P. rhodozyma*.

50 En consecuencia, la presente invención describe en la presente, un polinucleótido que comprende una molécula de ácido nucleico seleccionada del grupo formado por:

- 55 (a) moléculas de ácido nucleico que codifican por lo menos la forma madura del polipéptido representado en SEQ ID NO:3;
- (b) moléculas de ácido nucleico que comprenden la secuencia codificante representada en SEQ ID NO:2;
- (c) moléculas de ácido nucleico cuya secuencia de nucleótidos está degenerada como resultado del código genético de una secuencia de nucleótidos de (a) ó (b);
- 60 (d) moléculas de ácido nucleico que codifican un polipéptido derivado del polipéptido codificado por un polinucleótido de (a) a (c) por vía de sustitución, supresión y/o adición de uno o varios aminoácidos de la secuencia de aminoácidos del polipéptido codificado por un polinucleótido de (a) a (c);
- 65 (e) moléculas de ácido nucleico que codifican un polipéptido derivado del polipéptido cuya secuencia tiene una identidad del 51,3% ó más respecto a la secuencia de aminoácidos del polipéptido codificado por una molécula de ácido nucleico de (a) ó (b);

## ES 2 305 501 T3

- (f) moléculas de ácido nucleico que comprenden un fragmento o una porción que lleva un epítipo de un polipéptido codificado por una molécula de ácido nucleico de uno cualquiera de (a) a (e) y que tiene una actividad escualeno sintasa.
- 5 (g) moléculas de ácido nucleico que comprenden un polinucleótido que tiene una secuencia de una molécula de ácido nucleico amplificada a partir de una librería de ácido nucleico de *Phaffia* o *Xanthophylomyces*, empleando los cebadores representados en SEQ ID NO:4, 5 y 6;
- 10 (h) moléculas de ácido nucleico que codifican un polipéptido que tiene una actividad escualeno sintasa, en donde dicho polipéptido es un fragmento de un polipéptido codificado por uno cualquiera de (a) a (g);
- (i) moléculas de ácido nucleico que comprenden por lo menos 15 nucleótidos de un polinucleótido de uno cualquiera de (a) a (d);
- 15 (j) moléculas de ácido nucleico que codifican un polipéptido que tiene una actividad escualeno sintasa, en donde dicho polipéptido es reconocido por los anticuerpos que han sido generados contra un polipéptido codificado por una molécula de ácido nucleico de uno cualquiera de (a) a (h);
- 20 (k) moléculas de ácido nucleico que pueden obtenerse por screening de una biblioteca apropiada en condiciones restrictivas con una sonda que tiene la secuencia de la molécula de ácido nucleico de uno cualquiera de (a) a (j), y codificando un polipéptido que tiene una actividad escualeno sintasa; y
- (l) moléculas de ácido nucleico cuya cadena complementaria híbrida en condiciones restrictivas con una molécula de ácido nucleico de cualquiera de (a) a (k), que codifica un polipéptido que tiene la actividad escualeno sintasa.
- 25

Los términos “gen(es)”, “polinucleótido”, “secuencia de ácido nucleico”, “secuencia de nucleótidos”, “secuencia de ADN” ó “molécula(s) de ácido nucleico” como se emplean en la presente se refieren a una forma polimérica de nucleótidos de cualquier longitud, bien sean ribonucleótidos o bien sean desoxirribonucleótidos. Este término se refiere solamente a la estructura primaria de la molécula.

30

Así, este término incluye el ADN de cadena doble, de cadena simple y el ARN. Incluye también los tipos conocidos de modificación, por ejemplo, metilación, sustitución de los “casquetes” de uno o más de los nucleótidos que están allí de forma natural, con un análogo. De preferencia, la secuencia de ADN de la invención comprende una secuencia codificadora que codifica el polipéptido definido más arriba.

35

Una “secuencia codificadora” es una secuencia de nucleótidos que se transcribe en ARNm y/o se traduce en un polipéptido cuando se coloca bajo el control de una secuencia de regulación apropiada. Los enlaces de la secuencia de codificación se determinan mediante un codon de inicio de la traducción en el terminal 5' y un codon de paro de la traducción en el terminal 3'. Una secuencia codificadora puede incluir pero no está limitada al ARNm, ADNc, secuencias recombinantes de nucleótidos ó ADN genómico, mientras los intrones pueden estar presentes también bajo ciertas circunstancias. La SEQ ID:1 representa el ADN genómico en el cual la secuencia del intrón es insertada en la secuencia codificadora para el gen de la escualeno sintasa de *Phaffia rhodozyma*.

40

En general, el gen consta de varias partes que tienen diferentes funciones entre sí. En los eucariotas, los genes que codifican la proteína correspondiente se transcriben en ARN mensajero prematuro (pre-ARNm) que difiere de los genes del ARN ribosómico (ARNr), ARN de núcleo pequeño (ARNsn) y ARN de transferencia (ARNt). Aunque la ARN polimerasa II (PoIII) juega un papel central en esta transcripción, la PoIII no puede iniciar ella sola la transcripción sin el elemento cis que cubre una región de más arriba que contiene un promotor y una secuencia de activación de más arriba (UAS), y un factor de proteína que actúa en “trans”. En primer lugar, un complejo de iniciación de la transcripción que consta de varios componentes básicos de proteína reconoce la secuencia promotora en la región adyacente 5' del gen que se quiere expresar. En este caso, se requieren algunos participantes adicionales en el caso del gen que se expresa en alguna regulación específica, tal como una respuesta al shock por calor, o adaptación a la esteración de la nutrición, etc. En este caso, es necesario que exista una UAS en la región 5' de más arriba, sin traducir, alrededor de la secuencia promotora, y algunas proteínas de regulación positiva o negativa que reconozcan y se unan a la UAS. La fuerza de unión del complejo de iniciación de la transcripción a la secuencia del promotor está afectada por dicha unión al factor que actúa de forma “trans” alrededor del promotor, y esto permite la regulación de la actividad de la transcripción.

45

50

55

Después de la activación de un complejo de iniciación de la transcripción mediante la fosforilación, dicho complejo de iniciación de la transcripción inicia la transcripción desde el sitio de inicio de la transcripción. Algunas partes del complejo de iniciación de la transcripción se separan como un complejo de prolongación de la región promotora a la dirección 3' del gen (este paso es llamado como un caso de aclaramiento del promotor) y el complejo de prolongación continúa la transcripción hasta que alcanza una secuencia de terminación que está localizada en la región 3' adyacente de más arriba del gen. El pre-ARNm así generado se modifica en el núcleo mediante la adición de una estructura de casquete en el sitio de terminación, que casi corresponde al sitio de iniciación de la transcripción y mediante la adición de tramos de poliA en la señal de poliA, que se localiza en la región más arriba adyacente a 3'. A continuación unas estructuras de intrón se eliminan de la región codificadora y se combinan partes de exón para obtener un

60

65

## ES 2 305 501 T3

marco abierto de lectura cuya secuencia corresponde a la secuencia primaria de aminoácidos de una correspondiente proteína. Esta modificación en la cual se genera un ARNm maduro es necesaria para la expresión estable del gen. Los términos del ADNc corresponden en general a la secuencia de ADN la cual es transcrita inversamente a partir de esta secuencia madura de ARNm. Puede ser sintetizada por la transcriptasa inversa derivada de especies víricas empleando experimentalmente un ARNm maduro como molde.

Para expresar un gen el cual se deriva de un eucariota, se emplea a menudo un procedimiento en el cual el ADNc se clona en un vector de expresión para *E. coli*. Esto resulta del hecho de que una especificidad de la estructura del intrón varía entre los organismos, y de la incapacidad para reconocer la secuencia del intrón de otras secuencias. De hecho, un procarionta no tiene ninguna estructura de intrón en su propio fondo genético. Incluso en la levadura, el fondo genético es diferente entre los *Ascomycetes* a los cuales el *Saccharomyces cerevisiae* pertenece, y los *Basidiomycetes* a los cuales pertenece el *P. rhodozyma*, p. ej. la estructura del intrón del gen "actin" a partir del *P. rhodozyma* no puede ser reconocido o empalmado por la levadura ascomiceto, *Saccharomyces cerevisiae*. Las estructuras del intrón de algunas clases de genes parecen estar implicadas en la regulación de la expresión de su respectivo gen. Podría ser importante emplear un fragmento genómico que tiene sus intrones en un caso de autoclonación del gen de interés, cuya estructura del intrón implica dicha regulación de su propia expresión del gen.

Para aplicar un método de ingeniería genética para un estudio de mejora de una cepa es necesario estudiar su mecanismo genético en el caso de una transcripción y traducción. Es importante determinar una secuencia genética tal como su UAS, promotor, estructura del intrón y terminador para estudiar el mecanismo genético.

De acuerdo con esta invención, se determinó el gen que codifica la escualeno sintasa (SQS), el gen del *P. rhodozyma* incluyendo sus regiones 5' y 3' adyacentes así como también la estructura de su intrón.

La invención describe además en la presente, los polinucleótidos que difieren de una de las secuencias de nucleótidos representadas en SEQ ID NO:2 (y fragmentos de la misma) debido a la degeneración del código genético y así codifica unas escualeno sintasas como la codificada por las secuencias de nucleótidos representadas en SEQ ID NO:2. Además, el polinucleótido de la invención tiene una secuencia de nucleótidos que codifica una proteína que tiene una secuencia de aminoácidos representada en SEQ ID NO:3. Todavía en otra versión, el polinucleótido de la invención codifica una longitud completa de proteína de *Phaffia rhodozyma*, la cual es substancialmente homóloga a una secuencia de aminoácidos de SEQ ID NO:3.

Además, se apreciará por los expertos en la técnica que el polimorfismo de la secuencia de ADN, que conduce a cambios de las secuencias de aminoácidos puede existir dentro de una población (p. ej. la población de *P. rhodozyma*). Este polimorfismo genético en el gen de la escualeno sintasa puede existir entre individuos dentro de una población debido a la variación natural.

Como se emplean en la presente, los términos "gen" y "gen recombinante" se refieren a moléculas de ácido nucleico que comprenden un marco abierto de lectura que codifica una escualeno sintasa, de preferencia, una escualeno sintasa de la *P. rhodozyma*.

Estas variaciones naturales pueden dar típicamente como resultado un 1 - 5% de varianza en la secuencia de nucleótidos del gen de la escualeno sintasa. Cualquiera y todas dichas variaciones de nucleótidos y el polimorfismo resultante de los aminoácidos en la escualeno sintasa, que son el resultado de la variación natural y que no alteran la actividad funcional de la escualeno sintasa, se propone que estén dentro del ámbito de la invención.

Los polinucleótidos correspondientes a las variantes naturales y los homólogos a la *P. rhodozyma* del ADNc de la escualeno sintasa como se describen en la presente pueden ser aislados en base a su homología a los polinucleótidos de la escualeno sintasa de la *P. rhodozyma* descritos en la presente empleando el polinucleótido de la invención, o una porción del mismo, como una sonda de hibridación de acuerdo con las técnicas estándar de hibridación en condiciones restrictivas de hibridación. De acuerdo con ello, en otra versión, un polinucleótido como se ha descrito en la presente, tiene por lo menos 15 nucleótidos de longitud.

De preferencia, hibrida en condiciones restrictivas con la molécula de ácido nucleico que comprende una secuencia de nucleótidos del polinucleótido de la presente invención, p. ej. SEQ ID NO:2. En otras versiones, el ácido nucleico tiene por lo menos 20, 30, 50, 100, 250 ó más nucleótidos de longitud. La expresión "híbrida en condiciones restrictivas" se ha definido más arriba y se pretende que describe condiciones para la hibridación y lavado en las cuales las secuencias de nucleótidos por lo menos en un 60% permanecen típicamente idénticas entre sí, hibridadas entre sí. De preferencia, las condiciones son tales, que las secuencias por lo menos alrededor del 65% ó el 70%, con más preferencia por lo menos alrededor del 75% ó 80%, e incluso con mayor preferencia por lo menos alrededor del 85%, 90% ó 95% ó más idénticas entre sí permanecen típicamente hibridadas entre sí. De preferencia, el polinucleótido de la invención que hibrida en condiciones restrictivas con una secuencia de SEQ ID NO:2, corresponde a una molécula de ácido nucleico que se encuentra en la naturaleza.

En la presente invención, la secuencia de polinucleótidos incluye la SEQ ID NO:2 y fragmentos de la misma que tienen secuencias de polinucleótidos que hibridan con la SEQ ID NO:2 en condiciones restrictivas que son suficientes para identificar la unión específica con la SEQ ID NO:2. Por ejemplo, puede emplearse cualquier combinación de las siguientes condiciones de hibridación y lavado, para lograr la unión específica requerida:

## ES 2 305 501 T3

Hibridación altamente restrictiva: 6X SSC; 0,5% de SDS, 100  $\mu\text{g/ml}$  de ADN desnaturalizado de esperma de salmón, 50% de formamida, incubados durante la noche con ligero balanceo a 42°C.

5 Lavado altamente restrictivo: 1 lavado en 2X SSC, 0,5% de SDS a temperatura ambiente durante 15 minutos, seguido de otro lavado en 0,1X SSC, 0,5% de SDS a temperatura ambiente durante 15 minutos.

Hibridación de baja restrictividad: 6X SSC, 0,5% de SDS, 100  $\mu\text{g/ml}$  de ADN desnaturalizado de esperma de salmón, 50% de formamida, incubado durante la noche con un ligero balanceo a 37°C.

10 Lavado de baja restrictividad: 1 lavado en 0,1X SSC, 0,5% de SDS a temperatura ambiente durante 15 minutos.

Pueden obtenerse una condiciones restrictivas moderadas variando la temperatura a la cual la reacción de hibridación tiene lugar y/o las condiciones de lavado como se ha mencionado más arriba. En la presente invención, se prefiere emplear una hibridación y condiciones de lavado altamente restrictivas para definir la actividad antisentido contra el gen de la escualeno sintasa de la *P. rhodozyma*.

El término “homología” significa que las moléculas de ácido nucleico respectivas o las proteínas codificadas son funcionalmente y/o estructuralmente equivalentes. Las moléculas de ácido nucleico que son homólogas a las moléculas de ácido nucleico descritas más arriba y que se derivan de dichas moléculas de ácidos nucleicos son, por ejemplo, variaciones de dichas moléculas de ácidos nucleicos que representan modificaciones que tienen la misma función biológica, en particular proteínas de codificación con la misma o substancialmente la misma función biológica. Pueden ser variaciones que tienen lugar de forma natural, tales como secuencias de otras variedades de plantas o especies, o mutaciones.

25 Estas mutaciones pueden ocurrir de forma natural o pueden obtenerse mediante técnicas de mutagénesis. Las variaciones alélicas pueden ser variantes alélicas que tienen lugar de forma natural así como también variantes producidas sintéticamente o bien mediante ingeniería genética. Estructuralmente equivalentes, pueden por ejemplo identificarse mediante ensayo de la unión de dicho polipéptido a anticuerpos. Estructuralmente equivalente tienen la característica inmunológica similar, p. ej. comprenden epítomos similares.

30 Como se emplea en la presente, una molécula de ácido nucleico que “se encuentra naturalmente” se refiere a un ARN ó ADN que tiene una secuencia de nucleótidos que se encuentra en la naturaleza (p. ej. codifica una proteína natural). De preferencia, el polinucleótido codifica una escualeno sintasa de la *P. rhodozyma* natural.

35 Además de las variantes de la secuencia de escualeno sintasa que se encuentran en la naturaleza, que pueden existir en la población, los expertos especialistas apreciarán que los cambios pueden introducirse mediante la mutación en una secuencia de nucleótidos del polinucleótido que codifica la escualeno sintasa, conduciendo con ello a cambios en la secuencia de aminoácidos de la escualeno sintasa codificada, sin alteración de la capacidad funcional de la escualeno sintasa. Por ejemplo, las sustituciones de nucleótidos que conducen a sustituciones de aminoácidos en radicales aminoácidos “no esenciales” pueden hacerse en una secuencia del polinucleótido que codifica la escualeno sintasa, p. ej. SEQ ID NO:2. Un radical aminoácido “no esencial” es un radical que puede ser alterado a partir de una secuencia tipo “salvaje” de una escualeno sintasa sin alterar la actividad de dicha escualeno sintasa, mientras que para la actividad de la escualeno sintasa se requiere un radical aminoácido “esencial”. Otros radicales aminoácidos sin embargo, (p. ej. aquellos que no están conservados o están solamente semiconservados en el dominio que tiene la actividad escualeno sintasa) pueden no ser esenciales para la actividad y así están probablemente para ser aptos para la alteración sin alterar la actividad de la escualeno sintasa.

45 En consecuencia, la invención describe en la presente polinucleótidos que codifican la escualeno sintasa que contienen cambios en los radicales aminoácidos que no son esenciales para la actividad de la escualeno sintasa. Dicha escualeno sintasa difiere en la secuencia de aminoácidos de una secuencia contenida en la SEQ ID NO:3 que todavía retiene la actividad escualeno sintasa descrita en la presente. El polinucleótido puede comprender una secuencia de nucleótidos que codifica un polipéptido, en donde el polipéptido comprende una secuencia de aminoácidos por lo menos alrededor del 60% idéntica a una secuencia de aminoácidos de SEQ ID NO:3 y es capaz de participar en la síntesis del escualeno. De preferencia, la proteína codificada por la molécula de ácido nucleico es por lo menos alrededor del 60 - 65% idéntica a la secuencia en SEQ ID NO:3, con más preferencia por lo menos alrededor del 60 - 70% idéntica a una de las secuencias en SEQ ID NO:3, incluso con más preferencia por lo menos alrededor del 70 - 80%, 80 - 90%, 90 - 95% homólogo a la secuencia en SEQ ID NO:3 y con la mayor preferencia, por lo menos alrededor del 96%, 97%, 98%, ó 99% idéntica a la secuencia SEQ ID NO:3.

60 Para determinar el tanto por ciento de homología de dos secuencias de aminoácidos (p. ej. una de las secuencias de SEQ ID NO:3 y una forma mutante de la misma) ó de dos ácidos nucleicos, las secuencias se alinean con el fin de lograr una comparación óptima (p. ej. pueden introducirse huecos en la secuencia de una proteína o de un ácido nucleico para que el alineamiento sea óptimo con la otra proteína o ácido nucleico). A continuación se comparan los radicales de los aminoácidos o de los nucleótidos con las correspondientes posiciones de los aminoácidos o posiciones de los nucleótidos. Cuando una posición en una secuencia (p. ej. una de las secuencias de SEQ ID NO:2 ó 3) está ocupada por el mismo radical aminoácido o nucleótido que la correspondiente posición en la otra secuencia (p. ej. una forma mutante de la secuencia seleccionada), entonces las moléculas son homólogas en esta posición (es decir, como se emplea aquí, la “homología” de aminoácidos o ácidos nucleicos es equivalente a la “identidad” de aminoácidos o

## ES 2 305 501 T3

ácidos nucleicos). La homología en tanto por ciento entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (es decir, % de homología = números de posiciones idénticas/números totales de posiciones x 100). La homología puede determinarse mediante ciertos programas de ordenador como p. ej. el Blast 2.0 (Altschul SF, Nuc. Acid. Res. 25, 3389 - 3402, 1997). En esta invención, se emplea el programa GENETYX-SV/RC (Software Development Co., Ltd. Tokio, Japón), empleando su algoritmo por defecto como dicho programa de análisis de la homología. Este programa emplea el método de Lipman-Pearson para su algoritmo analítico.

Puede crearse una molécula de ácido nucleico que codifique un homólogo de la escualeno sintasa a una secuencia de proteína de SEQ ID NO:3, mediante la introducción de una o más substituciones, adiciones o deleciones de nucleótido en una secuencia de nucleótidos del polinucleótido de la presente invención, en particular de la SEQ ID NO:2, de forma que se introducen una o más substituciones, adiciones o deleciones de los aminoácidos en la proteína codificada. Pueden introducirse mutaciones en las secuencias de p. ej. SEQ ID NO:2 mediante técnicas estándar, tales como la mutagénesis dirigida al sitio y la mutagénesis mediada por PCR. De preferencia, se hacen substituciones conservadoras de aminoácidos en uno o más radicales de aminoácidos no esenciales predeterminados. Una “substitución conservadora de aminoácidos” es una en la cual el radical aminoácido es substituido con un radical aminoácido que tiene una cadena lateral similar. Las familias de radicales aminoácidos que tienen cadenas laterales similares han sido ya definidas en la técnica. Estas familias incluyen aminoácidos con cadenas laterales básicas (p. ej. lisina, arginina, histidina), cadenas laterales ácidas (p. ej. ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (p. ej. glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (p. ej. alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales beta-ramificadas (p. ej. treonina, valina, isoleucina) y cadenas laterales aromáticas (p. ej. tirosina, fenilalanina, triptófano, histidina). Así, un radical de aminoácido no esencial predeterminado en una escualeno sintasa se substituye de preferencia con otro radical aminoácido de la misma familia. Alternativamente, en otra versión, pueden introducirse mutaciones al azar a lo largo de todo o parte de una secuencia que codifica la escualeno sintasa tal como mediante mutagénesis de saturación, y los mutantes resultantes pueden ser sometidos a screening para detectar una actividad escualeno sintasa como se ha descrito en la presente para la identificación de mutantes que retienen la actividad escualeno sintasa. Después de la mutagénesis de una de las secuencias de SEQ ID NO:2, la proteína codificada puede expresarse recombinantemente y la actividad de la proteína puede determinarse empleando por ejemplo ensayos descritos en la presente.

En consecuencia, en una versión preferida, el polinucleótido como se ha descrito en la presente es ADN ó ARN.

Un polinucleótido como se describe en la presente, p. ej. una molécula de ácido nucleico que tiene una secuencia de nucleótidos de SEQ ID NO:2 ó una porción de la misma, puede aislarse empleando técnicas de biología molecular estándar, y la información de la secuencia proporcionada en la presente. Por ejemplo, puede aislarse el ADNc de la escualeno sintasa a partir de una biblioteca empleando toda o una porción de una de las secuencias del polinucleótido como se ha descrito en la presente como una sonda de hibridación y técnicas de hibridación estándar. Además, polinucleótidos que abarcan toda o una porción de una de las secuencias del polinucleótido como se describe en la presente pueden ser aislados mediante la reacción en cadena de la polimerasa empleando cebadores oligonucleótidos diseñados en base a esta secuencia (p. ej. una molécula de ácido nucleico que abarca todo o parte de una de las secuencias del polinucleótido como se describe en la presente puede aislarse mediante la reacción en cadena de la polimerasa empleando cebadores oligonucleótidos, p. ej. de SEQ ID NO:4, 5 ó 6, diseñados en base a esta misma secuencia de polinucleótido como se describe en la presente. Por ejemplo, puede aislarse el ARNm de células. p. ej. de células *Phaffia* (p. ej. mediante el procedimiento de extracción con tiocianato de guanidinio de Chirgwin *et al.*, y ADNc pueden ser preparados empleando la transcriptasa inversa (p. ej. transcriptasa inversa Moloney MLV ó transcriptasa inversa AMV adquirible en Promega (Madison, USA). Cebadores oligonucleótidos sintéticos para la amplificación mediante la reacción en cadena de la polimerasa, pueden diseñarse en base a una de las secuencias de nucleótidos representadas en la SEQ ID NO:2. Un polinucleótido como se describe en la presente, puede ser amplificado empleando ADNc ó alternativamente ADN genómico como un molde y cebadores oligonucleótidos apropiados de acuerdo con técnicas estándar de amplificación mediante PCR. El polinucleótido así amplificado puede ser clonado en un vector apropiado y caracterizado mediante análisis de la secuencia de ADN. Además, pueden prepararse los oligonucleótidos correspondientes a la secuencia de nucleótidos de la escualeno sintasa mediante técnicas sintéticas estándar, p. ej. empleando un sintetizador de ADN automático.

Los términos “fragmento”, “fragmento de una secuencia” ó “parte de una secuencia” significan una secuencia cortada de la secuencia original en cuestión. La secuencia cortada (secuencia de un ácido nucleico o secuencia de una proteína) puede variar ampliamente en longitud; siendo el tamaño mínimo el de una secuencia de suficiente tamaño para proporcionar una secuencia con por lo menos una función comparable y/o la actividad de la secuencia original en cuestión, mientras que el tamaño máximo no es crítico. En algunas aplicaciones, el tamaño máximo no es habitualmente substancialmente mayor que el que se requiere para proporcionar la actividad deseada y/o la(s) función(es) de la secuencia original.

Típicamente, la secuencia de aminoácidos cortada oscila desde aproximadamente 5 hasta aproximadamente 60 aminoácidos de longitud. Más típicamente, sin embargo, la secuencia tiene un máximo de aproximadamente 50 aminoácidos de longitud, de preferencia un máximo de aproximadamente 30 aminoácidos. Es habitualmente deseado seleccionar secuencias de por lo menos aproximadamente 10, 12, ó 15 aminoácidos hasta un máximo de aproximadamente 20 ó 25 aminoácidos.

## ES 2 305 501 T3

El término “epítipo” se refiere a sitios inmunorreactivos específicos sin un antígeno, también conocidos como determinantes antigénicos. Estos epítipos pueden ser una serie lineal de monómeros en una composición polimérica - tal como aminoácidos en una proteína - o constar de, o comprender una estructura secundaria o terciaria más compleja. Los expertos reconocerán que todos los inmunógenos (es decir, sustancias capaces de inducir una respuesta inmunológica) son antígenos; sin embargo, algunos antígenos, como p. ej. los haptenos, no son inmunógenos pero pueden hacerse inmunogénicos por copulación con una molécula soporte. El término “antígeno” se refiere a una sustancia contra la cual puede generarse un anticuerpo y/o a la cual el anticuerpo es específicamente inmunoreactivo.

El término “uno o varios aminoácidos” se refiere por lo menos a un aminoácido pero no más que el número de aminoácidos que resultarían en una homología inferior al 60% de identidad. De preferencia, la identidad es más del 70% u 80%, con más preferencia el 85%, 90% ó 95%, e incluso con mayor preferencia el 96%, 97%, 98% o el 99% de identidad.

El término “escualeno sintasa” o “actividad de escualeno sintasa” se refiere a la actividad enzimática de un polipéptido como se describe más adelante o el cual puede ser determinado en un método de ensayo de enzimas. Además, los polipéptidos que son inactivos en un ensayo en la presente pero son reconocidos por un anticuerpo que se une específicamente a la escualeno sintasa, es decir, que tiene uno o más epítipos de la escualeno sintasa, están también comprendidos bajo el término “escualeno sintasa”. En estos casos, la actividad se refiere a su actividad inmunológica.

Los términos “polinucleótido” y “molécula de ácido nucleico” se refieren también a polinucleótidos “aislados” o moléculas de ácido nucleico. Una molécula “aislada” de ácido nucleico es una que está separada de otras moléculas de ácido nucleico que están presentes en la fuente natural del ácido nucleico. De preferencia, un ácido nucleico “aislado” está libre de secuencias que en la naturaleza flanquean el ácido nucleico (es decir, secuencias localizadas en los extremos 5’ y 3’ del ácido nucleico) en el ADN genómico del organismo del cual se deriva el ácido nucleico.

Por ejemplo, en varias versiones, el polinucleótido PNO puede contener menos de aproximadamente 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,3 kb ó 0,1 kb de secuencias de nucleótidos que flanquean de modo natural la molécula de ácido nucleico en el ADN genómico de la célula de la cual se deriva el ácido nucleico (p. ej. una célula *Phaffia*). Además, los polinucleótidos, como se describen en la presente, en particular una molécula de ácido nucleico “aislado”, tal como una molécula de ADNc, puede estar substancialmente libre de otro material celular, o medio de cultivo cuando se obtiene mediante técnicas recombinantes, o precursores químicos u otros productos químicos cuando se sintetizan químicamente.

De preferencia, el polipéptido como se describe en la presente comprende una de las secuencias de nucleótidos mostradas en SEQ ID NO:2. La secuencia de SEQ ID NO:2 corresponde a los ADNc de la escualeno sintasa de *P. rhodozyma* de la invención.

Además, el polinucleótido como se describe en la presente, comprende una molécula de ácido nucleico la cual es un complemento de una de las secuencias de nucleótidos de los polinucleótidos más arriba mencionados, o una porción de los mismos. Una molécula de ácido nucleico la cual es complementaria a una de las secuencias de nucleótidos mostradas en la SEQ ID NO:2 es una molécula que es suficientemente complementaria a una de las secuencias de nucleótidos representadas en la SEQ ID NO:2, de manera que puede hibridar a una de las secuencias de nucleótidos mostradas en SEQ ID NO:2, formando con ello un dúplex estable.

El polinucleótido como se describe en la presente, comprende una secuencia de nucleótidos que es por lo menos aproximadamente el 60%, de preferencia por lo menos aproximadamente el 65 - 70%, con más preferencia por lo menos aproximadamente el 70 - 80%, 80 - 90%, ó 90 - 95%, e incluso con mayor preferencia, por lo menos aproximadamente el 95%, 96%, 97%, 98%, 99% ó más, homóloga a una secuencia de nucleótidos representada en la SEQ ID NO:2, ó una porción de la misma. El polinucleótido como se describe en la presente, comprende una secuencia de nucleótidos la cual hibrida, p. ej. en condiciones restrictivas como se ha definido en la presente, a una de las secuencias de nucleótidos representada en la SEQ ID NO:2 ó una porción de la misma.

Además, el polinucleótido como se define en la presente, puede comprender solamente una porción de la región que codifica una de las secuencias en la SEQ ID NO:2, por ejemplo, un fragmento que puede emplearse como una sonda o cebador o un fragmento que codifica una porción biológicamente activa de una escualeno sintasa. Las secuencias de nucleótidos determinadas a partir de la clonación del gen de la escualeno sintasa de la *P. rhodozyma* permite la generación de sondas y cebadores diseñados para emplear en la identificación y/o clonación de homólogos de la escualeno sintasa en otros tipos de células y organismos. La sonda/cebador comprende típicamente un oligonucleótido substancialmente purificado. El oligonucleótido comprende típicamente una región de una secuencia de nucleótidos que hibrida en condiciones restrictivas a nucleótidos consecutivos de por lo menos aproximadamente 12, 15, de preferencia aproximadamente 20 ó 25, con más preferencia, aproximadamente 40, 50 ó 75 de una cadena de detección de una de las secuencias mencionadas, p. ej. en SEQ ID NO:2, una secuencia anti-sentido de una de las secuencias, p. ej. mencionadas en SEQ ID NO:2, ó mutantes de las mismas presentes de forma natural. Los cebadores basados en un nucleótido como se ha descrito en la presente pueden emplearse en reacciones PCR para la clonación de homólogos de la escualeno sintasa. Pueden emplearse sondas basadas en secuencias de nucleótidos de la escualeno sintasa, para detectar secuencias transcritas o genómicas que codifican las mismas o proteínas homólogas. La sonda puede comprender además un grupo lábil unido a la misma, p. ej. el grupo lábil puede ser un radioisótopo, un compuesto fluorescente, una enzima o un co-factor de una enzima. Dichas sondas pueden emplearse como parte de un kit de

## ES 2 305 501 T3

ensayo genómico para la identificación de células que expresan una escualeno sintasa, como por ejemplo midiendo el nivel de una molécula de ácido nucleico que codifica la escualeno sintasa, en una muestra de células, p. ej. detectando los niveles del ARNm de la escualeno sintasa, o determinando si un gen genómico de la escualeno sintasa ha sido mutado o suprimido.

5 El polinucleótido como se ha descrito en la presente, codifica un polipéptido o porción del mismo, el cual incluye una secuencia de aminoácidos que es lo suficientemente homóloga a una secuencia de aminoácidos de SEQ ID NO:3 tal que la proteína o porción de la misma mantiene la capacidad de participar en la síntesis del escualeno, en particular una actividad escualeno sintasa como se describe en los ejemplos en microorganismos o plantas. Como se emplea  
10 en la presente, la expresión “suficientemente homólogo” se refiere a proteína o porciones de las mismas que tienen secuencias de aminoácidos que incluyen un mínimo número de idénticos o equivalentes (p. ej. un radical aminoácido que tiene una cadena lateral similar como un radical aminoácido en una de las secuencias del polipéptido como se ha descrito en la presente, radicales aminoácidos a una secuencia de aminoácidos de la SEQ ID NO:3, de manera que la proteína o porción de la misma es capaz de participar en la síntesis del escualeno en microorganismos o plantas.  
15 Ejemplos de una actividad escualeno sintasa están también descritos en la presente.

La proteína es por lo menos aproximadamente el 60 - 65%, de preferencia, por lo menos aproximadamente el 66 - 70%, y con más preferencia, por lo menos aproximadamente el 70 - 80%, 80 - 90%, 90 - 95% y con la mayor preferencia aproximadamente el 96%, 97%, 98%, 99% ó más homólogos a una secuencia completa de aminoácidos de SEQ ID NO:3.

Porciones de proteínas codificadas por el polinucleótido de la escualeno sintasa como se describe en la presente, son de preferencia porciones biológicamente activas de uno de los escualeno sintasa.

25 Como se menciona en la presente, la expresión “porción biológicamente activa de la escualeno sintasa” se pretende que incluya una porción, p. ej. un dominio/motivo, que participa en la biosíntesis del escualeno o que tiene una actividad inmunológica tal, que se une a un anticuerpo que se une específicamente a la escualeno sintasa. Para determinar si una escualeno sintasa o una porción biológicamente activa de la misma puede participar en el metabolismo, puede realizarse un ensayo de actividad enzimática. Dichos métodos de ensayo son ya bien conocidos por los expertos en  
30 la técnica, como se detalla en los ejemplos. Pueden prepararse fragmentos adicionales de ácido nucleico que codifican porciones biológicamente activas de una escualeno sintasa mediante el aislamiento de una porción de una de las secuencias en SEQ ID NO:2, expresión de la porción codificada de la escualeno sintasa o péptido (p. ej. mediante la expresión recombinante *in vitro* y valorando la actividad de la porción codificada de la escualeno sintasa o péptido.

35 En primer lugar, se clonó un fragmento parcial del gen que contiene una porción del gen de la SQS empleando un método de la PCR degenerada. Dicha PCR degenerada es un método para clonar un gen de interés que tiene una alta homología de la secuencia de aminoácidos con la enzima conocida de otras especies que tiene una misma o similar función. El cebador degenerado que se emplea como cebador en la PCR degenerada, fue diseñado mediante una traducción inversa de la secuencia de aminoácidos a los correspondientes nucleótidos (“degenerados”). En dicho  
40 cebador degenerado, se emplea generalmente un cebador mezcla que consta de cualquier A, C, G ó T, ó un cebador que contiene inosina en un código ambiguo. En esta invención, dichos cebadores mezcla se utilizaron para cebadores degenerados para clonar el gen de más arriba.

Un gen entero que contiene su región de codificación con su intrón así como también su región de regulación tal como un promotor o un terminador, puede ser clonado a partir de un cromosoma mediante screening de una biblioteca genómica, el cual se construye en un vector fago o un vector plásmido en un anfitrión apropiado, empleando un fragmento parcial de ADN obtenido mediante PCR degenerada como se ha descrito más arriba, como una sonda después de ser marcada. Generalmente, el *E. coli* como cepa anfitriona y el vector *E. coli*, un vector fago como el vector fago  $\lambda$ , o un vector plásmido como el vector pUC se emplean a menudo en la construcción de una biblioteca  
50 y la siguiente manipulación genética como por ejemplo un secuenciado, una digestión de restricción, una unión y similares. En esta invención, se construyó una biblioteca genómica *EcoRI* de *P. rhodozyma* en los derivados del vector  $\lambda$ ,  $\lambda$ DASHII. El tamaño del inserto, qué longitud del inserto debía ser clonada, se determinó mediante una hibridación de Southern blot para el gen antes de la construcción de la biblioteca. En esta invención, el ADN empleado para la sonda se marcó con digoxigenina (DIG), un hapteno esteroide en lugar de la marca convencional P<sup>32</sup>, siguiendo el protocolo preparado por el proveedor Boehringer-Mannheim, Mannheim, Alemania). Una biblioteca genómica  
55 construida a partir del cromosoma de *P. rhodozyma* se sometió a un screening empleando un fragmento de ADN marcado con DIG el cual tenía una porción de un gen de interés como una sonda. Las placas hibridadas fueron guardadas y empleadas para un estudio posterior. En el caso de emplear el  $\lambda$ DASHI (el tamaño del inserto fue de 9 kb a 23 kb), el  $\lambda$ ADN preparado se digirió mediante la enzima *EcoRI* seguido del clonado del inserto *EcoRI* en un vector plásmido tal como el pUC19 ó el pBluescriptII SK+. El ADN del plásmido así obtenido fue analizado para determinar su secuencia.

En esta invención, hemos empleado el secuenciador de ADN fluorescente, sistema automático ALFred (Pharmacia, Uppsala, Suecia), empleando un protocolo de secuenciado autociclo, en el cual la ADN polimerasa de Taq, se emplea  
65 en la mayor parte de los casos de secuenciado.

Después de la determinación de la secuencia genómica, se empleó la secuencia de una región de codificación para la clonación del ADNc del correspondiente gen. El método de la PCR fue también empleado para clonar el fragmento

## ES 2 305 501 T3

de ADNc. Los cebadores de la PCR cuyos secuencias fueron idénticas a la secuencia del extremo 5' y 3' del marco abierto de lectura (ORF) fueron sintetizados con la adición de un sitio de restricción apropiado, y la PCR se efectuó empleando dichos cebadores de la PCR. En la presente invención, se empleó un conjunto de ADNc como molde en esta clonación de la PCR del ADNc. Dicho conjunto de ADNc consiste en varias especies de ADNc que fueron sintetizadas *in vitro* mediante la transcriptasa inversa vírica y la Taq polimerasa (kit de CapFinder fabricado por Clontech, Palo Alto, U.S.A.) empleando el ARNm obtenido del *P. rhodozyma* como un molde. El ADNc de interés así obtenido, se confirmó determinando su secuencia.

En otra versión, la presente invención describe un método para la fabricación de un vector recombinante que comprende la inserción de un polinucleótido como se describe en la presente, en un vector.

Además, la presente invención describe un vector recombinante que contiene el polinucleótido como se describe en la presente o producido mediante dicho método de la invención.

Como se emplea en la presente, el término “vector” se refiere a una molécula de ácido nucleico capaz de transportar un polinucleótido al cual ha sido unido. Un tipo de vector es un “plásmido”, el cual se refiere a un bucle circular de ADN de doble cadena al cual pueden estar unidos segmentos adicionales de ADN. Otro tipo de vector es un vector vírico en donde pueden unirse segmentos de ADN ó ARN al genoma vírico. Ciertos vectores son capaces de una replicación autónoma en una célula anfitriona en la cual se han introducido (p. ej., vectores bacterianos que tienen un origen de replicación bacteriano, y vectores episomales de mamíferos). Otros vectores (p. ej., vectores mamíferos no episomales) son integrados en el genoma de una célula anfitriona después de la introducción en la célula hospedadora, y con ello se replican junto con el genoma del hospedador. Además, ciertos vectores son capaces de dirigir la expresión de genes a los cuales están operativamente ligados. Estos vectores reciben el nombre, en la presente, de “vectores de expresión”. En general los vectores de expresión, de utilidad en las técnicas de ADN recombinante se encuentran a menudo en forma de plásmidos. En la presente especificación “plásmido” y “vector” pueden emplearse indistintamente siendo el nombre de plásmido el que se emplea más habitualmente. Sin embargo, la invención describe en la presente otras formas de vectores de expresión, tales como vectores víricos (p. ej., retrovirus de replicación defectuosa, adenovirus y virus adenoasociados), los cuales sirven para funciones equivalentes.

La presente invención describe también en la presente, cósmidos, virus, bacteriófagos y otros vectores empleados convencionalmente en ingeniería genética que contienen una molécula de ácido nucleico de acuerdo con la invención. Los métodos que son bien conocidos por los expertos en la técnica pueden emplearse para la construcción de varios plásmidos y vectores. Alternativamente, las moléculas de ácido nucleico y los vectores como se describen en la presente pueden ser reconstituidos en liposomas para suministrar a las células diana.

La presente invención describe también en la presente un vector en el cual el polinucleótido como se describe en la presente se une operativamente a las secuencias de control de la expresión permitiendo la expresión en células anfitrionas procarióticas o eucarióticas. La naturaleza de dichas secuencias de control difiere en función del organismo anfitrión. En los procariotas, las secuencias de control incluyen generalmente el promotor, el sitio de unión ribosómica y los terminadores. En los eucariotas, las secuencias de control incluyen generalmente los promotores, los terminadores y en algunos casos, los potenciadores, transactivadores o factores de transcripción.

La expresión “secuencia de control” se pretende que incluya como mínimo, aquellos componentes cuya presencia es necesaria para la expresión, y pueden también incluir componentes adicionales ventajosos.

La expresión “unidos operablemente” se refiere a una yuxtaposición en la cual los componentes así descritos se encuentran en una relación entre sí tal, que les permite ejercer su función de la manera pretendida. Una secuencia de control “operablemente unida” a una secuencia del código está unida de tal forma que la expresión de la secuencia del código se consigue en condiciones compatibles con las secuencias de control. En el caso de que la secuencia de control sea un promotor, es obvio para un experto en la especialidad que se emplee un ácido nucleico de doble cadena.

Dichas secuencias reguladoras son ya conocidas por el experto. Las secuencias reguladoras incluyen aquellas que directamente constituyen la expresión de una secuencia de nucleótidos en muchos tipos de célula anfitriona y aquellos que dirigen la expresión de la secuencia de nucleótidos solamente en ciertas células anfitrionas o en ciertas condiciones. Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de dichos factores como la elección de la célula anfitriona que debe ser transformada, el nivel de expresión de la proteína deseada, etc. Los vectores de expresión como se han descrito en la presente pueden ser introducidos en células anfitrionas para con ello producir proteínas o péptidos, incluyendo las proteínas o péptidos de fusión, codificados por polinucleótidos como se describe en la presente.

Los vectores de expresión recombinantes como se describe en la presente pueden ser diseñados para la expresión del escualeno sintasa en células procarióticas o eucarióticas. Por ejemplo, los genes que codifican el polinucleótido como se describe en la presente pueden ser expresados en células bacterianas tales como la *E. coli*, células de insectos (empleando vectores de expresión del baculovirus), levaduras y otras células fúngicas, algas, ciliados de los tipos: *Holotrichia*, *Peritrichia*, *Spirotrichia*, *Suctorina*, *Tetra-hymena*, *Paramecium*, *Colpidium*, *Glaucoma*, *Platyophrya*, *Potomacus*, *Pseudocohnilembus*, *Euplotes*, *Engelmanniella* y *Stylonychia*, especialmente *Stylonychia lemnae* con vectores siguiendo un método de transformación como se describe en WO 98/01.572 y células vegetales multicelulares o células de mamíferos. Células anfitrionas adecuadas son ya conocidas por los expertos en la especialidad. Alternativamente,

## ES 2 305 501 T3

el vector de expresión recombinante puede ser transcrito y traducido *in vitro*, por ejemplo empleando las secuencias reguladoras del promotor T7 y la T7 polimerasa.

La expresión de las proteínas en los procariotas se efectúa la mayor parte de las veces con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas bien sea de fusión o bien sea de no-fusión. Los vectores de fusión añaden un número de aminoácidos a una proteína codificada, habitualmente al terminal amino de la proteína recombinante pero también al terminal C ó fundidos dentro de regiones adecuadas en las proteínas. Dichos vectores de fusión sirven habitualmente para tres finalidades: 1) aumentar la expresión de la proteína recombinante; 2) aumentar la solubilidad de la proteína recombinante; y 3) ayudar a la purificación de la proteína recombinante, actuando como un ligando en la purificación por afinidad. A menudo, en los vectores de expresión por fusión, se introduce un sitio de escisión proteolítica en la unión de la parte fundida y la proteína recombinante para permitir la separación de la proteína recombinante de la parte de fusión a continuación de la purificación de la proteína de fusión. Dichas enzimas, y sus secuencias de reconocimiento de los afines, incluyen el factor Xa, la trombina y la enteroquinasa.

Los vectores típicos de expresión de fusión incluyen el pGEX (Pharmacia Biotech Inc.), pMAL (New England Biolabs, Beverly, USA) y pRIT5 (Pharmacia; Piscataway, USA) los cuales funden la glutatión S-transferasa (GST), proteína de unión de la maltosa E, ó proteína A respectivamente, a la proteína recombinante diana. La secuencia que codifica el polipéptido codificado por el polinucleótido como se describe en la presente puede ser clonado en un vector de expresión pGEX para crear un vector que codifica una proteína de fusión que comprende desde el terminal N hasta el terminal C, la proteína GST-trombina del sitio X de escisión. La proteína de fusión puede purificarse mediante purificación por cromatografía de afinidad empleando resina de glutatión-agarosa. p. ej. la escualeno sintasa recombinante no fundida con la GST, puede ser recuperada mediante escisión de la proteína de fusión con la trombina.

Ejemplos de vectores de expresión de *E. coli* de no-fusión inducibles adecuados incluyen el pTrc y pET 11d. La expresión del gen diana a partir del vector pTrc se basa en la transcripción de la ARN polimerasa del anfitrión a partir de un promotor de fusión híbrido *trp-lac*. La expresión del gen diana a partir del vector pET 11d se basa sobre la transcripción del promotor de fusión T7 *gnl0-lac*, mediado por una RNA polimerasa vírica co-expresada (T7 *gnl*). Esta polimerasa vírica es suministrada por cepas anfitrionas BL21(DE3) ó HMS174(DE3) a partir de un profago  $\lambda$  residente, que contiene un gen T7 *gnl* bajo el control transcripcional del promotor *lacUV 5*.

Una estrategia para maximizar la expresión de la proteína recombinante consiste en expresar la proteína en una bacteria anfitriona con una debilitada capacidad de escindir proteolíticamente la proteína recombinante. Otra estrategia es la de alterar la secuencia de ácido nucleico del ácido nucleico que hay que insertar en un vector de expresión de forma que los codones individuales para cada aminoácido son los preferentemente utilizados en la bacteria escogida para la expresión, como p. ej. el *E. coli*. Dicha alteración de las secuencias del ácido nucleico como se describen en la presente puede ser efectuada por técnicas estándar de síntesis de ADN.

Además, el vector de escualeno sintasa puede ser un vector de expresión de una levadura. Ejemplos de vectores para la expresión en la levadura *S. cerevisiae* incluyen el pYepSecl, pMFa, pJRY88, y pYES2 (Invitrogen, San Diego, USA). Los vectores y métodos para la construcción de vectores apropiados para emplear en otros hongos, como p. ej. los hongos filamentosos, son ya conocidos por los expertos en la técnica.

Alternativamente, el polinucleótido como se describe en la presente puede ser introducido en las células de un insecto empleando vectores de expresión de baculovirus. Los vectores de expresión de baculovirus adquiribles para la expresión de proteínas en células de insectos cultivadas (p. ej. células Sf9) incluyen la serie pAc y la serie pVL.

Alternativamente, el polinucleótido como se describe en la presente se introduce en células de mamíferos empleando un vector de expresión mamífero. Ejemplos de vectores de expresión mamíferos incluyen los pCDM8 y pMT2PC. Cuando se emplean en células de mamíferos, las funciones de control de los vectores de expresión son a menudo proporcionadas por los elementos reguladores víricos. Por ejemplo, los promotores empleados corrientemente son derivados del poliovirus, Adenovirus 2, citomegalovirus y Simian Virus 40. Otros sistemas de expresión adecuados tanto para células procarióticas como eucarióticas son ya conocidos por los expertos en la especialidad.

El vector de expresión mamífero recombinante puede ser capaz de dirigir la expresión del ácido nucleico, de preferencia, en un tipo particular de célula (p. ej., se emplean elementos reguladores específicos de tejidos para expresar el ácido nucleico). Elementos reguladores específicos de tejido son ya conocidos en la técnica. Ejemplos no limitantes de promotores específicos de tejidos, adecuados, incluyen el promotor albúmina (específico del hígado), promotores específicos de linfoides, en particular promotores de células T receptores e inmunoglobulinas, promotores específicos de neuronas (p. ej. el promotor de neurofilamento), promotores específicos del páncreas, y promotores específicos de glándulas mamarias (p. ej. promotor de suero de leche; patente US 4.873.316 y EP 264.166). Están comprendidos también, promotores regulados en desarrollo, por ejemplo, los promotores del *hox* de ratón y el promotor de la fetoproteína.

Así, el gen de SQS expresado, puede ser verificado en lo que se refiere a su actividad por el método de ensayo enzimático. En la bibliografía se encuentran algunos protocolos experimentales. El que sigue es uno de los métodos que se emplean para la determinación de la actividad escualeno sintasa; las actividades escualeno sintasa se determinan monitorizando la conversión del  $[1\text{-H}^3]\text{FPP}$  en escualeno. Las mezclas de reacción (500 ml) incluyen 50 mM de Tris-

## ES 2 305 501 T3

HCl (pH 7,4), 2 mM de KF, 1 mM de MgCl<sub>2</sub>, 1 mM de la enzima NADPH, 10 mM de [1-H<sup>3</sup>]FPP (370 MBq/moles, 3,7 kBq/ml: New England Nuclear, Boston, MA). Las reacciones comienzan añadiendo [1-H<sup>3</sup>]FPP. Después de unos 10 minutos de incubación a 37°C, las reacciones se finalizan añadiendo 1 ml de etanol. Después de añadir 1 ml de H<sub>2</sub>O, las mezclas se agitan vigorosamente con 3 ml de éter de petróleo durante 30 minutos. Los lípidos extraídos se evaporan y resuspenden en 25 ml de cloroformo. Se aplican muestras a hojas plastificadas (silica gel 60, F254, Merck, Rahway, NJ) para cromatografía de capa fina (TLC), y se desarrollan en heptano durante 15 minutos. Las radiactividades incluidas en la fracción de escualeno se miden mediante el conteo de centelleo líquido. Cuando se emplea la expresión del vector para la *S. cerevisiae* puede efectuarse convenientemente un análisis complementario empleando la cepa *ERG 9* mutante condicional de la escualeno sintasa derivada de la *S. cerevisiae* como una cepa anfitriona para la confirmación de la actividad (Merkulov *et al.*, Yeast, 16, 197-206, 2000).

A continuación de la confirmación de la actividad de la enzima, se purifica la proteína expresada y se emplea para generar el anticuerpo contra la enzima purificada. El anticuerpo así preparado sería empleado para la caracterización de la expresión de la correspondiente enzima en un estudio de mejora de la cepa, un estudio de optimización de las condiciones del cultivo, y similares.

La presente invención describe también en la presente, un anticuerpo que se une específicamente al polipéptido como se describe en la presente, o partes, es decir, fragmentos específicos o epítomos de dicha proteína.

Los anticuerpos como se describen en la presente pueden emplearse para identificar y aislar otra escualeno sintasa y genes. Estos anticuerpos pueden ser anticuerpos monoclonales, anticuerpos policlonales o anticuerpos sintéticos así como fragmentos de anticuerpos, como p. ej. fragmentos Fab, Fv ó scFv, etc. Pueden prepararse anticuerpos monoclonales, por ejemplo, mediante técnicas ya conocida por los expertos, las cuales comprenden la fusión de células de mieloma de ratón a células de bazo derivadas de mamíferos inmunizados.

Además, pueden obtenerse anticuerpos o fragmentos de los mismos de los péptidos antes mencionados, empleando métodos que ya son conocidos por los expertos. Estos anticuerpos pueden emplearse, por ejemplo, para la inmunoprecipitación e inmunolocalización de proteínas como se ha descrito en la presente así como también para la monitorización de la síntesis de dichas proteínas, por ejemplo en organismos recombinantes, y para la identificación de compuestos que interactúan con la proteína como se ha descrito en la presente. Por ejemplo, la resonancia del plasmón de superficie como se emplea en el sistema BIAcore puede emplearse para aumentar la eficiencia del fago de las selecciones de anticuerpos del fago, proporcionando un alto incremento de la afinidad a partir de una simple biblioteca de anticuerpos de fagos que se unen a un epítomo de la proteína como se ha descrito en la presente. En muchos casos el fenómeno de unión de los anticuerpos a los antígenos es equivalente a otra unión ligando - antiligando.

En esta invención, el fragmento de gen para la escualeno sintasa se clonó a partir de la *P. rhodozyma* con la finalidad de aumentar su nivel de expresión en *P. rhodozyma* por el método genético empleando el fragmento de gen clonado.

La presente invención describe en la presente un procedimiento para la producción de carotenoides en donde un gen que codifica la escualeno sintasa es modificado en un anfitrión adecuado como p. ej. la *P. rhodozyma* para aumentar su expresión, y para el cultivo de dicho transformante en un medio apropiado en condiciones apropiadas de cultivo.

Para disminuir la expresión de un gen con métodos genéticos, pueden emplearse algunas estrategias. Una de las cuales es un método de disgregación genética. En este método, un fragmento parcial del gen objetivo que hay que disgregar, se liga a un cassette de resistencia a una droga en el vector de integración que no puede replicar en el organismo anfitrión. Un gen de resistencia a una droga que codifica la enzima que permite que el anfitrión sobre-viva en presencia de un antibiótico tóxico, se utiliza a menudo para el marcador de selección. El gen de resistencia G418 contenido en pGB-Ph9 es un ejemplo de un gen resistente a una droga el cual funciona en la *P. rhodozyma*. Puede también emplearse, un marcador de complemento de nutrición, en el anfitrión el cual tiene un marcador auxotrófico apropiado. La cepa ATCC24221 de *P. rhodozyma* que requiere citina para su crecimiento es un ejemplo de auxotrofo. Empleando la CTP sintetasa como ADN dador para la ATCC24221, puede establecerse un sistema de vector anfitrión empleando un complemento de nutrición.

Después de la transformación del organismo anfitrión y de la recombinación entre el fragmento de gen objetivo sobre el vector y su correspondiente fragmento de gen sobre el cromosoma de los organismos anfitriones, el vector de integración se integra en el cromosoma anfitrión por medio de una simple recombinación cruzada. Como resultado de esta recombinación, el cassette de resistencia a la droga sería insertado en el gen objetivo, cuyo producto traducido está solamente sintetizado en su forma truncada la cual no tiene su función enzimática. De una manera similar, se emplearon también dos partes del gen objetivo para el estudio de disgregación del gen, en el cual el gen resistente a la droga puede ser insertado entre dichos dos fragmentos parciales de los genes objetivo sobre el vector de integración. En el caso de este tipo de vector, se espera un caso de doble recombinación entre los fragmentos de gen alojados en el vector de integración y los correspondientes fragmentos de gen sobre el cromosoma del anfitrión. Aunque la frecuencia de esta doble recombinación cruzada es más pequeña que la simple recombinación cruzada, ningún fenotipo del gen objetivo es más estable mediante la doble recombinación cruzada que mediante la simple recombinación cruzada.

Esta estrategia se empleó para la construcción del recombinante productor de licopeno de *Candida utilis* el cual alojaba genes carotenogénicos bacterianos sobre el plásmido (Shimada *et al.*, (Applied and Environmental Microbiology ["Microbiología aplicada y del medio ambiente"], 64 (7), 2676-2680, 1998)). El gen *ERG9* que codifica la

escualeno sintasa se clonó a partir del *C. utilis* y su gen de disgregación se indujo después de una doble recombinación cruzada del gen *ERG9* sobre el cromosoma del *C. utilis* que produce el licopeno. Shimada *et al.* informó que la disgregación del gen *ERG9* daba un efecto positivo sobre la carotenogenesis mediante el *C. Utilis* recombinante especialmente derivado del anfitrión en el cual la 3-hidroxi metilglutaril-CoA reductasa se amplificó sobre el locus del ADN ribosómico multicopiado sobre el cromosoma del anfitrión.

Por otro lado, esta estrategia tiene la dificultad en el caso del gen cuya función es esencial y la disrupción es letal para el organismo anfitrión como p. ej. el gen de la escualeno sintasa. En la referencia anterior (Shimada *et al.*), la disgregación se efectuó sobre, o bien la copia del gen de la escualeno sintasa, o bien dentro de las dos copias de aquellos sobre el cromosoma del anfitrión. En dicha construcción, no se confirmó que el descenso del nivel de la actividad de escualeno sintasa era suficiente para aumentar el flujo de carbono en la ruta del carotenoide.

En este caso, pueden aplicarse otras estrategias para disminuir (no interrumpir) la expresión de un gen. Una de ellas consiste en una mutagénesis convencional para explorar el mutante cuya expresión para la escualeno sintasa ha disminuido. En este método, un recombinante apropiado en el cual se funde un gen reporter adecuado a la región promotora del gen de la escualeno sintasa a partir del organismo anfitrión, se muta y los mutantes que muestran una actividad más débil del producto del gen reporter pueden explorarse. En dichos mutantes, se espera que la expresión de la actividad de la escualeno sintasa, disminuya por encontrarse la mutación en la región promotora del gen reporter de la región que actúa como *trans*, la cual podría afectar la expresión del gen de la escualeno sintasa de distinta manera que la mutación que se encuentra en el propio gen promotor. En el caso de que la mutación tenga lugar en la región promotora de la fusión del reporter, dicha mutación puede aislarse mediante la secuencia de la región correspondiente. Así, la mutación aislada puede introducirse en una variedad de carotenoides, especialmente la astaxantina que produce mutantes derivados a partir de la *P. rhodozyma* mediante una recombinación entre el promotor original para el gen de la escualeno sintasa sobre el cromosoma y el fragmento promotor mutado. Para excluir las mutaciones que tienen lugar en la región que actúa como *trans*, puede inducirse también una mutación mediante una mutagénesis *in vitro* de un elemento *cis* en la región promotora. En este método, el cassette de un gen conteniendo un gen reporter el cual está fundido a una región promotora derivada de un gen de interés en su extremo 5', y una región terminadora de un gen de interés en su extremo 3', se mutageniza y a continuación se introduce en la *P. rhodozyma*. Mediante la detección de la diferencia de la actividad del gen reporter, puede explorarse una mutación efectiva. Dicha mutación puede introducirse en la secuencia de la región promotora nativa sobre el cromosoma mediante el mismo método que en el caso de un método de mutación *in vivo*.

Sin embargo, estos métodos tienen algunos inconvenientes como el de ser un procedimiento que consume un cierto tiempo.

Otra estrategia para disminuir la expresión de un gen es un método antisentido. Este método se aplica con frecuencia para disminuir la expresión de un gen incluso cuando se emplean organismos teleomórficos tales como el *P. rhodozyma* como organismos anfitriones, al cual es habitualmente difícil de aplicar el método de la mutación y disgregación génica. El método antisentido es un método para disminuir la expresión de un gen de interés mediante la introducción de un fragmento de un gen artificial, cuya secuencia es complementaria a un fragmento del ADNc del gen en cuestión.

Una molécula de ácido nucleico antisentido comprende una secuencia de nucleótido la cual es complementaria a una molécula de ácido nucleico "con sentido", que codifica una proteína, p. ej. complementaria a la cadena codificadora de una molécula de ADNc de doble cadena o complementaria a una secuencia de ARNm. De acuerdo con ello, una molécula de ácido nucleico antisentido puede unirse con un enlace de hidrógeno a una molécula de ácido nucleico con sentido. La molécula de ácido nucleico antisentido puede ser complementaria a una cadena completa de escualeno sintasa o solamente a una porción de la misma. En consecuencia, una molécula de ácido nucleico antisentido puede ser antisentido para una "región de codificación" de la cadena de codificación de una secuencia de nucleótidos que codifica una escualeno sintasa. La expresión "región de codificación" se refiere a la región de la secuencia de nucleótidos que comprende codones que son traducidos en radicales aminoácidos. Además, la molécula de ácido nucleico antisentido es antisentido para una "región no codificadora" de la cadena de codificación de una secuencia de nucleótidos que codifican la escualeno sintasa. La expresión "región no codificadora" se refiere a secuencias 5' y 3' que flanquean la región codificadora que no se traduce en un polipéptido (es decir, también referido a las regiones 5' y 3' no traducidas).

Dadas las secuencias codificadoras de la cadena que codifican la escualeno sintasa descrita en la presente, pueden diseñarse moléculas antisentido del ácido nucleico de acuerdo con las reglas de Watson y Crick del emparejamiento de bases. La molécula de ácido nucleico antisentido puede ser complementaria a la región de codificación completa de la escualeno sintasa ARNm, pero puede ser también un oligonucleótido que es antisentido a solamente una porción de la región codificadora o no codificadora del ARNm de la escualeno sintasa.

Por ejemplo, el oligonucleótido antisentido puede ser complementario a la región que rodea el sitio de inicio de la traducción del ARNm de la escualeno sintasa. Un oligonucleótido antisentido puede tener por ejemplo, aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50 nucleótidos de longitud. Una molécula de ácido nucleico antisentido como se describe en la presente puede construirse empleando una síntesis química y reacciones de ligación enzimáticas empleando procedimientos ya conocidos en la técnica. Por ejemplo, una molécula de ácido nucleico antisentido (p. ej. un oligonucleótido antisentido) puede sintetizarse químicamente empleando nucleótidos que se encuentran normalmente en la naturaleza o nucleótidos modificados variadamente diseñados para aumentar la estabilidad biológica

de las moléculas o para aumentar la estabilidad física del dúplex formado entre los ácidos nucleico antisentido y los ácidos nucleicos con sentido, p. ej. pueden emplearse los derivados fosforotioatos y los nucleótidos substituidos con acridina. Ejemplos de nucleótidos modificados que pueden emplearse para generar el ácido nucleico antisentido incluyen el 5-fluoruracilo, 5-bromouracilo, 5-clouracilo, 5-yodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5-(carboxihidroximetil)uracilo, 5-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluracilo, dihidrou-  
 5 racilo, beta-D-galactosilqueosina, inosina, N6-isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-adenina, 7-metilguanina, 5-inetilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxicarboximetiluracilo, 5-metoxiuracilo, 2-metiltio-N6-isopenteniladenina, ácido uracil-5-oxiacético (v), wibutoxosina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico del ácido uracil-5-oxiacético, ácido uracil-5-oxiacético (v), 5-metil-2-tio-uracilo, 3-(3-amino-3-N-2-carboxipropil)uracilo, (acp3)w, y 2,6,diaminopurina. Alternativamente, el ácido nucleico antisentido puede obtenerse biológicamente empleando un vector de expresión en el cual un polinucleótido ha sido subclonado en una orientación antisentido (es decir el ARN transcrito a partir del polinucleótido insertado será de una orientación antisentido a un polinucleótido diana de interés, descrito más adelante  
 15 en la siguiente subsección).

Las moléculas de ácido nucleico antisentido como se describen en la presente son típicamente administradas a una célula, o son generadas *in situ* de forma que hibridan con, o se unen a, ARNm celular y/o ADN genómico que codifica una escualeno sintasa para con ello inhibir la expresión de la proteína, p. ej. inhibiendo la transcripción y/o la traducción. La hibridación puede efectuarse mediante una convencional complementariedad del nucleótido para formar un dúplex estable, o por ejemplo, en el caso de una molécula de ácido nucleico antisentido que se une a los dúplex de ADN, a través de interacciones específicas en el surco principal de la doble hélice. La molécula de antisentido puede modificarse de forma que se une específicamente a un receptor o a un antígeno expresado sobre la superficie de una célula seleccionada, p. ej. uniendo la molécula de ácido nucleico antisentido a un péptido o un anticuerpo que se une a un receptor o antígeno de la superficie celular. La molécula de ácido nucleico antisentido puede también suministrarse a células empleando los vectores descritos en la presente. Para lograr unas suficientes concentraciones intracelulares de las moléculas antisentido, se prefieren aquellas construcciones vectoriales en las cuales la molécula de ácido nucleico antisentido está colocada bajo el control de un procarriota fuerte, vital, o bien un eucariota que incluye promotores vegetales.  
 20  
 25  
 30

La molécula de ácido nucleico antisentido como se describe en la presente, puede ser una molécula de ácido nucleico  $\alpha$ -anomérico. Una molécula de ácido nucleico  $\alpha$ -anomérico forma híbridos específicos de doble cadena, con ARN complementario, en el cual contrariamente a las unidades  $\beta$ , las cadenas corren paralelamente entre sí. La molécula de ácido nucleico antisentido puede también comprender un 2'-o-metilribonucleótido o un análogo ARN-ADN quimérico.  
 35

Además, la molécula de ácido nucleico antisentido como se ha descrito en la presente, puede ser un ribozima. Los ribozimas son moléculas de ARN catalíticas con actividad de ribonucleasa, las cuales son capaces de escindir un ácido nucleico de simple cadena, tal como un ARNm, con el cual tienen una región complementaria. Así, los ribozimas (p. ej. los ribozimas del "hammerhead") pueden emplearse para escindir catalíticamente transcritos de ARNm de la escualeno sintasa para con ello inhibir la traducción del ARNm. Un ribozima que tiene una especificidad para una molécula de ácido nucleico que codifica la escualeno sintasa puede diseñarse en base a la secuencia de polinucleótidos como se ha descrito en la presente. Por ejemplo, puede construirse un derivado de un ARN de Tetrahimena L-19 IVS, en el cual la secuencia de nucleótidos del sitio activo es complementaria a la secuencia de nucleótidos que hay que escindir en un ARNm codificante (patentes US 4.987.071 y US 5.116.742). Alternativamente, el ARNm de la escualeno sintasa puede emplearse para seleccionar un ARN catalítico que tiene una actividad ribonucleasa específica a partir de un conjunto de moléculas de ARN.  
 40  
 45

La aplicación de un método antisentido para la construcción de un carotenoide que sobreproduce una cepa de *P. rhodozima* ha sido ejemplificado en la patente EP 1.158.051.  
 50

La presente invención describe en la presente un método para la producción de una célula anfitriona recombinante, el cual comprende la introducción del vector o el polinucleótido como se ha descrito en la presente en una célula anfitriona.  
 55

El ADN del vector puede introducirse en células procarióticas o eucarióticas por medio de técnicas convencionales de transformación o transfección. Como se emplea en la presente, los términos "transformación" y "transfección", conjugación y transducción, se refieren a una variedad de técnicas reconocidas para la introducción de un ácido nucleico extraño (p. ej. ADN) en una célula anfitriona, incluyendo la co-precipitación con fosfato de calcio o cloruro de calcio, transfección inducida por dextrano - DEAE, lipofección, competencia natural, transferencia inducida químicamente, o electroporación. Métodos adecuados para la transformación o células anfitrionas para transfección incluyendo las células vegetales, son ya bien conocidas por los expertos.  
 60

Para la estable transfección de células mamíferas, es conocido que en función del vector de expresión y la técnica de transfección empleada, solamente una pequeña fracción de células pueden integrar el ADN extraño en su genoma. Con el fin de identificar y seleccionar estos integrantes, un gen que codifica un marcador seleccionable (p. ej. resistencia a los antibióticos) se introduce generalmente en las células anfitrionas junto con el gen de interés. Los marcadores seleccionables preferidos incluyen aquellos que confieren resistencia a ciertas drogas, tales como el G418, higromi-  
 65

## ES 2 305 501 T3

5 cina y metotrexato. El ácido nucleico que codifica un marcador seleccionable puede ser introducido en una célula anfitriona sobre el mismo vector que el que codifica el polipéptido como se describe en la presente o puede ser introducido en un vector separado. Las células establemente transfectadas con el ácido nucleico introducido pueden ser identificadas mediante, por ejemplo, selección con una droga (p. ej. las células que han incorporado el gen del marcador seleccionable sobrevivirán, mientras que las otras células, morirán).

10 Para crear un microorganismo homólogo recombinante, se prepara un vector el cual contiene por lo menos una porción del polinucleótido como se ha descrito en la presente, en el cual ha sido introducida una supresión, adición o sustitución para con ello alterar, p. ej. disgregar funcionalmente el gen de la escualeno sintasa. De preferencia, este gen de la escualeno sintasa es un gen de escualeno sintasa de *P. rhodozyma*, pero puede ser un homólogo de una fuente relacionada o diferente. Alternativamente, el vector puede ser diseñado de forma tal que después de una recombinación homóloga, el gen endógeno de la escualeno sintasa se muta, o de otra manera se altera, pero todavía codifica la proteína funcional (p. ej. la región regulatoria corriente arriba puede ser alterada para con ello alterar la expresión de la escualeno sintasa endógena). Para crear una mutación en un punto por medio de recombinación 15 homóloga pueden emplearse también híbridos de ADN - ARN conocidos como quimeraplastia.

20 El vector se introduce en una célula y las células en las cuales el gen del polinucleótido introducido se ha recombinado homologamente con el gen endógeno de la escualeno sintasa, se seleccionan, empleando técnicas ya conocidas en la especialidad.

Pueden producirse otras células anfitrionas las cuales contienen sistemas de selección que permiten la expresión regulada del gen introducido. Por ejemplo, la inclusión del polinucleótido como se ha descrito en la presente sobre un vector colocándolo bajo el control del operón lac permite la expresión del polinucleótido solamente en presencia de IPTG. Estos sistemas reguladores son ya bien conocidos en la técnica.

25 De preferencia, la molécula de ácido nucleico es extraña a la célula anfitriona.

30 Por “extraña”, se da a entender que la molécula de ácido nucleico es o bien heteróloga con respecto a la célula anfitriona, esto significa derivada de una célula o de un organismo con un fondo genómico diferente, o bien es homóloga con respecto a la célula anfitriona pero está situada en un medio ambiente genómico diferente del que se encuentra de modo natural en la parte contraria de dicha molécula de ácido nucleico. Esto significa que si la molécula de ácido nucleico es homóloga con respecto a la célula anfitriona, no está situada en su ubicación natural en el genoma de dicha molécula anfitriona, en particular está rodeada por diferentes genes. En este caso, la molécula de ácido nucleico puede estar o bien bajo el control de su propio promotor o bien bajo el control de un promotor heterólogo. El vector o molécula de ácido nucleico como se ha descrito en la presente, el cual está presente en la célula anfitriona, puede o bien ser integrado en el genoma de la célula anfitriona o bien puede ser mantenido en alguna forma extracromosómica. A este respecto, debe también comprenderse que la molécula de ácido nucleico como se ha descrito en la presente, puede emplearse para restaurar o crear un gen mutante mediante una recombinación homóloga.

40 De acuerdo con ello, en otra versión, la presente invención describe en la presente una célula anfitriona modificada mediante ingeniería genética con el polinucleótido como se describe en la presente o el vector como se describe en la presente.

45 Las expresiones “célula anfitriona” y “célula anfitriona recombinante” se emplean en la presente, intercambiabilmente. Se comprende que dichas expresiones se refieren no solamente a la célula sujeto en particular sino a la progenie o potencial progenie de dicha célula, Debido a ciertas modificaciones puede ocurrir en las generaciones venideras debido o bien a una mutación o bien a influencias del medio ambiente, que dicha progenie no puede de hecho ser idéntica a la célula original pero está todavía incluida dentro del ámbito de la expresión como se emplea en la presente.

50 Por ejemplo, un polinucleótido como se describe en la presente, puede ser introducido en células bacterianas así como también en células de insectos, células fúngicas o células mamíferas (tal como las células del ovario de hamster chino (CHO) ó células COS), algas, ciliados, células vegetales, hongos u otros microorganismos como la *E. coli*. Otras células anfitrionas adecuadas son conocidas por los expertos en la técnica. Son preferidos la *E. coli*, baculovirus, *Agrobacterium* o células fúngicas son por ejemplo las del género *Saccharomyces*, p. ej. las de las especies *S. cerevisiae* ó *P. rhodozyma* (*Xanthophylomyces dendrorhous*).

55 Además, la presente invención describe en la presente un método para la producción de transformantes fúngicos que comprenden la introducción del polinucleótido o el vector como se ha descrito en la presente, en el genoma de dicha célula fúngica.

60 Para la expresión de las moléculas de ácido nucleico como se han descrito en la presente, en la orientación con sentido o antisentido en células vegetales, las moléculas están colocadas bajo el control de elementos reguladores los cuales aseguran la expresión en células fúngicas. Estos elementos reguladores pueden ser heterólogos u homólogos con respecto a la molécula de ácido nucleico que hay que expresar, así como también con respecto a las especies fúngicas que hay que transformar.

65 En general, dichos elementos reguladores comprenden un promotor activo en células fúngicas. Para obtener una expresión constitutiva en células fúngicas, se emplean de preferencia promotores constitutivos, tales como el promotor

## ES 2 305 501 T3

gliceraldehido-3-deshidrogenasa, de la *P. rhodozyma* (patente WO 97/23.633). Pueden emplearse promotores inducibles con la finalidad de ser capaces de controlar exactamente la expresión. Un ejemplo de promotores inducibles es el promotor de genes que codifican proteínas de choque térmico. También ha sido descrito (patente EP 1.035.206), un promotor del gen de la amilasa el cual es un candidato para dichos promotores inducibles. Los elementos reguladores pueden comprender además potenciadores transcripcionales y/o translacionales que funcionan en células fúngicas. Además, los elementos reguladores pueden incluir señales de terminación de la transcripción, tales como una señal poli-A, lo cual permite la adición de una cola poli A al transcripto, lo cual puede mejorar su estabilidad.

Son ya conocidos en la técnica, métodos para la introducción de ADN extraño en células fúngicas. Estos incluyen por ejemplo, el método de transformación con LiCl, la fusión de protoplastos, electroporación, métodos biolísticos como p. ej. el bombardeo con partículas y otros métodos ya conocidos en la técnica. Métodos para la preparación de vectores apropiados son ya conocidos por los expertos. Métodos para la transformación empleando métodos biolísticos son ya bien conocidos por los expertos en la técnica.

El término “transformación” como se emplea en la presente se refiere a la transferencia de un polinucleótido exógeno en una célula anfitriona independientemente del método empleado para la transferencia. El polinucleótido puede ser temporal o establemente introducido en la célula hospedadora y puede conservarse no integrado, por ejemplo, como un plásmido o como uniones químicas o alternativamente puede ser integrado en el genoma del anfitrión.

En general, los hongos que pueden ser modificados de acuerdo con la invención y los cuales muestran o bien una sobreexpresión de una proteína como se ha descrito en la presente, o bien una reducción de la síntesis de dicha proteína, pueden ser derivados a partir de cualquier especie fúngica deseada.

Además, la presente invención describe en la presente una célula fúngica que comprende el polinucleótido, el vector o puede obtenerse mediante el método de la presente invención.

Así, la presente invención describe también en la presente, células fúngicas transgénicas las cuales contienen (de preferencia establemente integradas en el genoma) un polinucleótido como se describe en la presente, unido a elementos reguladores los cuales permiten la expresión del polinucleótido en células fúngicas y en donde el polinucleótido es extraño a la célula fúngica transformada. Para el significado del término “extraño”, véase más arriba.

La presencia y expresión del polinucleótido en las células fúngicas transformadas, modula, de preferencia disminuye, la síntesis del escualeno y conduce al aumento de la producción de carotenoides, especialmente de la producción de astaxantina en las células fúngicas transformadas así obtenidas, de preferencia, en células de *P. rhodozyma*.

Así, la invención en cuestión describe también en la presente, células fúngicas transformadas como se describe en la presente.

En consecuencia, debido a la expresión alterada de la escualeno sintasa, las rutas metabólicas de las células se modulan en la producción del rendimiento y/o la eficiencia de la producción.

Los términos “producción” o “productividad” son reconocidos en la técnica e incluyen la concentración del producto de fermentación (por ejemplo ácidos grasos, carotenoides, (poli)sacáridos, vitaminas, isoprenoides, lípidos, ésteres de ceras y/o polímeros como polihidroxialcanoatos y/o sus productos del metabolismo u otros productos químicos finos deseados como se ha mencionado en la presente), formados en un tiempo dado y un volumen de fermentación dado (p. ej. kgs de producto/hora/litro).

La expresión “eficiencia de la producción” incluye el tiempo requerido para lograr un particular nivel de producción (por ejemplo, cuanto tiempo es necesario para que la célula alcance una particular proporción de producción de dicho rendimiento alterado, en particular en carotenoides, (poli)sacáridos, lípidos, vitaminas, isoprenoides, etc).

El término “rendimiento” o “rendimiento en producto/carbono” está reconocido en la técnica e incluye la eficiencia de la conversión de la fuente de carbono en el producto (es decir, acetil CoA, ácidos grasos, carotenoides, vitaminas, isoprenoides, lípidos, etc. y/o otros compuestos como se ha definido más arriba y cuya biosíntesis está basada sobre dichos productos). Esto está generalmente expresado por ejemplo, en kgs de producto por kg de fuente de carbono. Aumentando el rendimiento o la producción del compuesto, aumenta la cantidad de moléculas recuperadas, o de moléculas recuperadas útiles de este compuesto, en una cantidad dada de cultivo, y en una cantidad dada de tiempo.

Los términos “biosíntesis” (el cual se emplea como sinónimo para “síntesis” de “producción biológica” en células, tejidos, vegetales, etc.) o “ruta biosintética” están reconocidos por la técnica e incluyen la síntesis de un compuesto, de preferencia un compuesto orgánico, mediante una célula a partir de compuestos intermedios en lo que puede ser un paso múltiple y un proceso altamente regulado.

El término “metabolismo” está reconocido en la técnica e incluye la totalidad de reacciones bioquímicas que tienen lugar en un organismo. El metabolismo de un compuesto particular (p. ej. el metabolismo de la acetil CoA, de un ácido graso, de la hexosa, lípido, isoprenoide, vitamina, carotenoide, etc.) comprende la totalidad de rutas biosintéticas, de modificación y degradación, en la célula en relación con este compuesto.

## ES 2 305 501 T3

Dicho *P. rhodozyma* formado mediante ingeniería genética debería ser cultivado en un medio apropiado y evaluado en su productividad o/y rendimiento de carotenoides, especialmente astaxantina. Un hiperproductor de astaxantina así seleccionado debería ser confirmado en cuanto a la relación entre su productividad y al nivel de expresión del gen o proteína, conseguidos mediante dicho método de ingeniería genética.

La presente invención se ilustra además con los ejemplos que se describen más adelante.

Se emplearon los siguientes materiales y métodos en el ejemplo descrito más adelante.

### Cepas

*P. rhodozyma* ATCC96594 (redepositado con el registro nº ATCC 74438 el 8 de Abril de 1998 de acuerdo con el Tratado de Budapest)

*E. coli* DH5 $\alpha$ ; F<sup>-</sup>,  $\Phi$ 80d, *lacZ* $\Delta$ M15,  $\Delta$ (*lacZYA-argF*)U169, *hsd* ( $r_K^-$ ,  $m_K^+$ ), *recA*I, *endA*I, *deoR*, *thi-1*, *supE*44, *gyrA*96, *relA*1 (Toyobo, Japón)

*E. coli* XL1 MRA (P2): $\Delta$ (*mcrA*)183,  $\Delta$ (*mcrB*)-*hsdSMR-mrr*)173, *endA*1, *supF*44, *thi-1*, *gyrA*96, *relA*1, *lac*(P2 lysogen) (Stratagene, La Jola, U.S.A.)

### Vectores

$\lambda$ DASHII (Stratagene)

pBluescriptII KS (Stratagene)

vector pMOSBlue T (Amersham Buckinghamshire, U.K.)

### Medios

La cepa *P. rhodozyma* se mantuvo rutinariamente en medio YPD (DIFCO, Detroit, U.S.A.). La cepa de *E. coli* se mantuvo en medio LB (10 g de Bacto-trypton, 5 g de extracto de levadura (DIFCO) y 5 g de NaCl por litro). El medio NZY (5 g de NaCl, 2 g de MgSO<sub>4</sub>.7H<sub>2</sub>O, 5 g de extracto de levadura (DIFCO), 10 g de NZ amina tipo A (WAKO, Osaka, Japón) por litro), se emplea para la propagación del fago  $\lambda$  en un agar blando (0,7% de agar (WAKO)). Cuando se prepara un medio con agar, se suplementa el 1,5% del agar (WAKO).

### Métodos

Las enzimas de restricción y la T4 ADN ligasa se compraron en Takara Shuzo (Ohtsu, Japón).

El aislamiento de un ADN cromosómico a partir del *P. rhodozyma* se efectuó empleando el kit QIAGEN Genomic Kit (QIAGEN, Hilden, Alemania) siguiendo el protocolo suministrado por el fabricante. Se efectuó una mini preparación de ADN del plásmido a partir del *E. coli* transformado con el sistema de aislamiento automático del ADN (PI-50, Kurabo, Co. Ltd., Osaka, Japón). Se efectuó una mini preparación del ADN del plásmido de un transformante de *E. coli* empleando una columna de QIAGEN (QIAGEN). El aislamiento del ADN  $\lambda$  se efectuó mediante el sistema Wizard de purificación de preps de ADN londa (Promega, Madison, U.S.A.) siguiendo el protocolo preparado por el fabricante. Un fragmento de ADN se aisló y purificó a partir de la agarosa empleando QIAquick ó QIAEX II (QIAGEN). La manipulación de derivados del fago  $\lambda$  se efectuó siguiendo el protocolo preparado por el fabricante (Stratagene).

El aislamiento del ARN total a partir del *P. rhodozyma* se efectuó con el método del fenol empleando Isogen (Nippon Gene, Yoyama, Japón). El ARNm se purificó del ARN total así obtenido empleando el kit de separación de ARNm (Clontech). El ADNc se sintetizó empleando el kit CapFinder de construcción de ADNc (Clontech).

El empaquetado *in vitro* se efectuó empleando el extracto de empaquetado Gigapack III gold (Stratagene).

La reacción en cadena de la polimerasa (PCR) se efectuó con el ciclador térmico de Perkin Elmer modelo 2400. Cada condición de la PCR está descrita en los ejemplos. Los cebadores de la PCR se adquirieron en un proveedor comercial. Los cebadores de ADN fluorescentes para el secuenciado del ADN se adquirieron en Pharmacia. El secuenciado del ADN se efectuó con el secuenciador de ADN fluorescente automático (ALFred, Pharmacia).

Las células competentes de DH5 $\alpha$  se compraron en Toyobo (Japón).

## ES 2 305 501 T3

### Ejemplo 1

#### *Aislamiento del ARNm del *P. rhodozyma* y construcción de una biblioteca de ADNc*

5 Para la construcción de una librería de ADNc de *P. rhodozyma*, se aisló el ARN total mediante el método de extracción con fenol inmediatamente después de la disgregación celular y el ARNm de la cepa ATCC96594 de *P. rhodozyma* se purificó empleando el kit de separación de ARNm (Clontech).

10 Se recogieron las células de la cepa ATCC96594 de 10 ml de un cultivo de dos días en medio YPD por centrifugación (1500 x g durante 10 minutos) y se lavaron una vez con tampón de extracción (10 mM de citrato de Na/HCl (pH 6,2) conteniendo 0,7 M de KCl). Después de suspender en 2,5 ml de tampón de extracción, las células se disgregaron mediante el homogeneizador a presión French (Ohtake Works Corp., Tokio, Japón) a 1500 kgf/cm<sup>2</sup> e inmediatamente se mezclaron con dos veces su volumen de isogeno (Nippon gene) de acuerdo con el método especificado por el fabricante. En este paso, se recuperaron 400 µg del ARN total.

15 A continuación, este ARN total se purificó empleando el kit de separación del ARNm (Clontech) de acuerdo con el método especificado por el fabricante. Finalmente, se obtuvieron 16 µg de ARNm de la cepa ATCC96594 de *P. rhodozyma*.

20 Para la construcción de la biblioteca de ADNc, se empleó el kit de construcción CapFinder de ADNc de PCR (Clontech) de acuerdo con el método especificado por el fabricante. Un µg de ARNm purificado se aplicó para la síntesis de una primera cadena, seguido de la amplificación mediante PCR. Después de esta amplificación mediante PCR, se obtuvo 1 mg del conjunto de ADNc.

### 25 Ejemplo 2

#### *Clonación de un gen de SGS (escualeno sintasa) parcial a partir de *P. rhodozyma**

30 Para clonar un gen parcial de SGS de *P. rhodozyma*, se empleó un método de PCR degenerado. Las especies y el número de registro para la base de datos, cuya secuencia para la escualeno sintasa se empleó para el análisis de alineación múltiple, son como sigue:

35	<i>Ustilago maydis</i>	Q92459 (SwissProt)
	<i>Schizosaccharomyces pombe</i>	P36596 (SwissProt)
	<i>Saccharomyces cerevisiae</i>	M63979 (GenBank)
40	<i>Rattus norvegicus</i>	Q02769 (SwissProt)
	<i>Mus musculus</i>	P53798 (SwissProt)
	<i>Candida albicans</i>	P78589 (SwissProt)
45	<i>Homo sapiens</i>	I38245 (Pir)
	<i>Arabidopsis thaliana</i>	U79159, AF004396
50	<i>Leishmania major</i>	U30455
	<i>Glycyrrhiza glabra</i>	D86410

55 Dos cebadores mezclados cuyas secuencias de nucleótidos fueron diseñadas y sintetizadas basadas sobre la secuencia común de los conocidos genes de escualeno sintasa de otras especies, es decir squ1 (cebador con sentido)(SEQ ID NO:4), squ4 (cebador antisentido (SEQ ID NO:5) y squ5 (cebador antisentido) (SEQ ID NO:6) (en las secuencias “n” significa los nucleótidos a, c, g ó t, “r” significa los nucleótidos a ó g, e “y” significa los nucleótidos c ó t).

60 Después de la reacción PCR de 25 ciclos de 95°C durante 30 segundos, 45°C durante 30 segundos y 72°C durante 15 segundos empleando ext. (Takara Shuzo) como una ADN polimerasa y el conjunto de ADNc obtenido en el ejemplo 1 como molde, se aplicó la mezcla de reacción sobre un gel de agarosa para electroforesis. Cada banda de la PCR que tenía una deseada longitud se recuperó de la mezcla de reacción PCR en la cual la combinación empleando los squ1 y squ4, y squ1 y squ5 respectivamente, y se purificó mediante QIAquick (QIAGEN) de acuerdo con el método del fabricante y a continuación se ligó al vector pMOSBlue-T (Amersham). Después de la transformación del competente *E. coli* DH5α, se seleccionaron 6 colonias blancas y los plásmidos se aislaron con el sistema automático de aislamiento de ADN. Como resultado del secuenciado, se descubrió que 3 clones tenían una secuencia cuya secuencia de aminoácidos deducida era similar a los genes conocidos de la escualeno sintasa. Estos clones aislados de ADNc fueron

## ES 2 305 501 T3

llamados pSQS 1007 derivados a partir de la reacción PCR empleando *squ1* y *squ5* y como pSQS 1006 derivados a partir de la reacción PCR empleando *squ1* y *squ4*, y el pSQS 1006 se empleó para otros estudios de screening.

### 5 Ejemplo 3

#### *Aislamiento del ADN genómico del *P. rhodozyma**

10 Para el aislamiento del ADN genómico del *P. rhodozyma*, se empleó el kit genómico QIAGEN, de acuerdo con el método especificado por el fabricante.

15 Las células del *P. rhodozyma* ATCC96594 a partir de 100 ml de cultivo durante la noche en medio YPD se recogieron por centrifugación (1500 x g durante 10 minutos) y se lavaron una vez con tampón TE (10 mM de Tris/HCl (pH 8,0) conteniendo 1 mM de EDTA).

20 Después de suspender en 8 ml de tampón Y1 del kit genómico QIAGEN, se añadió iticasa (SIGMA, St. Louis, U.S.A.) a la concentración de 2 mg/ml para disgregar las células mediante degradación enzimática, y la mezcla de reacción se incubó durante 90 minutos a 30°C y a continuación se efectuó el próximo paso de extracción. Finalmente se obtuvieron 20 µg de ADN genómico.

### Ejemplo 4

#### *Hibridación por Southern blot empleando el pSQS1006 como sonda*

25 Se efectuó la hibridación mediante Southern blot para clonar un fragmento genómico que contiene el gen SQS de *P. rhodozyma*. Dos µg de ADN genómico se digirieron mediante *EcoRI* y se sometieron a electroforesis con gel de agarosa seguido por un tratamiento ácido y alcalino. El ADN desnaturalizado se transfirió a una membrana de nylon (Hybond N+, Amersham) empleando transblot (Joto Rika, Tokio, Japón) durante una hora. El ADN que se transfirió a la membrana de nylon se fijó mediante un tratamiento térmico (80°C, 90 minutos). Se preparó una sonda marcando un ADN molde (pSQS 1006 digerido con *EcoRI*) con el método DIG de múltiples cebadores (Boehringer Mannheim). La hibridación se efectuó con el método especificado por el fabricante, Como resultado se visualizó una banda hibridada en el margen de 9,0 a 23,0 kilobases (kb).

### 35 Ejemplo 5

#### *Clonación de un fragmento genómico que contiene el gen SQS*

40 Cuatro µg del ADN genómico se digirió mediante *EcoRI* y se sometieron a electroforesis con gel de agarosa. A continuación, los ADN cuya longitud está dentro del margen de 9,0 a 20,0 kb se recuperaron mediante el kit de extracción con gel QIAEX II (QIAGEN) de acuerdo con el método especificado por el fabricante. El ADN purificado se ligó a 0,5 µg de λDASH II digerido con *EcoRI* y tratado con CIAP (fosfatasa alcalina de intestino de ternera), a 16°C durante la noche y empaquetado con extracto de empaquetado Gigapack III oro (Stratagene). El extracto empaquetado se infectó con la cepa MRA(P2) de *E. coli* y se recubrió con medio NZY vertido sobre medio agar LB. Se exploraron aproximadamente 5000 placas empleando pSQS1006 digerido con *EcoRI* como una sonda. Se hibridaron cinco placas a la sonda marcada.

50 Este derivado de λDASH II conteniendo gen SQS presumiblemente de *P. rhodozyma* se preparó empleando el sistema Wizard de purificación de landa preps de ADN (Promega). A continuación se efectuó la PCR empleando estos derivados λDASH II como un molde y dos cebadores *squ9* y *squ10* como cebadores. Estos cebadores *squ9* y *squ10* fueron diseñados en base a la secuencia interna de pSQS1006:*squ9* (cebador con sentido) (SEQ ID NO:7) y *squ10* (cebador antisentido) SEQ ID NO:8).

55 Como resultado de las PCR con las mismas condiciones de la PCR descritas en el ejemplo 2, se obtuvo una esperada banda de 0,5 kb. Se sugirió que todos aquellos derivados de λDASHII podían contener un gen SQS presumiblemente de *P. rhodozyma*. Se purificaron aproximadamente 20,0 kb del fragmento inserto de *EcoRI* en uno de estos derivados de λDASHII, empleando QIAEX II (QIAGEN), y se sometieron a un subclonado en el vector pBluescriptII KS (Stratagene) empleando el DH5α como cepa anfitriona y se obtuvo el pS1229.

60

### Ejemplo 6

#### *Secuenciado de un fragmento genómico que contiene el gen SQS*

65 El pSQS 1229 se secuenció con el procedimiento del “cebador andante” empleando el kit de secuenciado AutoRead (Pharmacia).

## ES 2 305 501 T3

Como resultado del secuenciado se determinó una secuencia de nucleótidos comprendiendo 4807 pares de bases de un fragmento genómico conteniendo el gen SQS de la *P. rhodozyma* conteniendo su promotor (1549 bp) y terminador (836 bp) (SEQ ID NO:1).

5 La región de codificación fue de 2422 pares de bases de largo y constaba de 9 exones y 8 intrones. Los intrones estaban dispersados a través de toda la región de codificación sin diagonales en 5' ó 3'. Se descubrió que un marco abierto de lectura (SEQ ID NO:2) consta de 512 aminoácidos (SEQ ID NO:3) cuya secuencia es notablemente similar a la secuencia conocida de aminoácidos de la escualeno sintasa de otras especies (51,3%) de identidad igual a la escualeno sintasa de la *Schizosaccharomyces pombe*) como resultado del software de búsqueda de homología GENETYX-SV/RC (Software Development Co., Ltd. Tokio, Japón).

### Ejemplo 7

#### 15 *Construcción de un plásmido antisentido para el gen SQS*

Un fragmento del gen antisentido el cual cubre el gen entero de estructura para el gen SQS, se amplifica mediante el método PCR y a continuación se clona en un vector de integración en el cual el gen SQS antisentido se transcribe mediante su propio promotor SQS en la *P. rhodozyma*.

20 Dichos cebadores incluyen una secuencia de reconocimiento de asimetría para la enzima de restricción *Sfil* (GGCCNNNNNGGCC), pero su secuencia asimétrica pendiente se diseña para que sea diferente. Esto permite una clonación direccional en el vector de expresión el cual tiene la misma secuencia asimétrica en su secuencia de ligación. El empleo de dicha construcción está ejemplificado en la patente EP 1.158.051.

25 Para el fragmento promotor y terminador que puede conducir la transcripción del gen SQS anti-sentido, el promotor y terminador SQS se clona a partir del cromosoma empleando la información de la secuencia listada en SEQ ID NO:1.

30 A continuación, el fragmento terminador se funde con la cassette resistente G418 ligando el fragmento de ADN que contiene el terminador SQS a la cassette resistente G418 del pG418Sa330 (patente EP 1.035.206) para el vector apropiado como p. ej. el p-BluescriptII KS (Stratagene).

35 A continuación, se inserta el fragmento *SacI* de 3,1 kb que contiene el locus del ADN ribosómico (ADNr) (Wery *et al.*, Gene, 184, 89-97, 1997) en el cassette G 418 corriente abajo sobre el plásmido así preparado. El fragmento de ADNr existe en multicopias sobre el cromosoma del eucariota. La integración mediante el fragmento de ADNr resultaría en una integración multicopiada sobre el cromosoma del anfitrión empleado, y esto permite la sobreexpresión de genes extraños que se alojan en el vector de expresión.

40 A continuación, se inserta el promotor SQS en el terminador SQS corriente arriba, para construir el vector de expresión que funciona en *P. rhodozyma*.

Finalmente, la construcción SQS antisentido se completa insertando el fragmento *Sfil* de 1,5 kb, que contiene el SQS antisentido en el vector de expresión así preparado que funciona en el *P. rhodozyma*.

45 Una construcción de un plásmido similar está ejemplificado en la patente EP 1.158.051.

### Ejemplo 8

#### 50 *Transformación del P. rhodozyma con el vector SQS antisentido*

El vector SQS antisentido así preparado, se transforma en la cepa tipo salvaje del *P. rhodozyma*, la ATCC96594, mediante la transformación biolística, siguiendo el protocolo descrito en la patente EP 1.158.051.

### 55 Ejemplo 9

#### *Caracterización del SQS antisentido recombinante del P. rhodozyma*

60 Se cultiva el SQS antisentido recombinante del *P. rhodozyma* ATCC96594, en 50 ml de medio YPD en un frasco Erlenmeyer de 500 ml a 20°C durante 3 días empleando su cultivo para siembra el cual crece en 10 ml de medio YPD en tubos de ensayo (21 mm de diámetro) a 20°C durante 3 días. Para el análisis del carotenoide producido, se retira un volumen apropiado de caldo de cultivo y se emplea para el análisis de su crecimiento, productividad de carotenoides, especialmente la astaxantina.

65 Para el análisis del crecimiento, se mide la densidad óptica a 660 nm empleando un fotómetro UV-1200 (Shimadzu Corp., Kioto, Japón) además de la determinación de su masa celular seca secando las células derivadas de 1 ml de caldo después de la microcentrifugación a 100°C durante un día.

## ES 2 305 501 T3

Para el análisis del contenido de astaxantina y carotenoides totales, las células se recogen a partir de 1,0 ml de caldo después de la microcentrifugación y se emplean para la extracción de los carotenoides a partir de células de *P. rhodozyma* mediante disgregación con perlas de vidrio. Después de la extracción, las células disgregadas se separan por centrifugación y el resultante se analiza para determinar el contenido en carotenoides por HPLC. Las condiciones de la HPLC empleadas son las siguientes: columna de HPLC: Chrompack Lichrosorb si-60 (4,6 mm, 250 mm); temperatura: temperatura ambiente; eluyente: acetona/hexano (18/82), añadir 1 ml/litro de agua al eluyente; volumen de inyección: 10  $\mu$ l; velocidad de flujo: 2,0 ml/minuto; detección: UV a 450 nm.

Pueden obtenerse muestras de referencia de Hoffmann La-Roche (Basilea, Suiza) (particularmente, la astaxantina) u otros proveedores comerciales (WAKO, SIGMA, etc.).

15

20

25

30

35

40

45

50

55

60

65

## ES 2 305 501 T3

### REIVINDICACIONES

1. Un organismo recombinante cuya expresión génica de la escualeno sintasa se ha reducido en comparación con el organismo anfitrión, por lo cual es capaz de producir carotenoides en un nivel potenciado en relación con el organismo anfitrión, **caracterizado** porque el organismo recombinante es una célula fúngica que contiene un polinucleótido antisentido frente a una molécula de ácido nucleico, seleccionada del grupo formado por:

(a) moléculas de ácido nucleico que codifican por lo menos la forma madura del polipéptido representado en SEQ ID NO:3;

(b) moléculas de ácido nucleico que comprenden la secuencia codificante como está representada en SEQ ID NO:2;

(c) moléculas de ácido nucleico cuya secuencia de nucleótidos está degenerada como resultado del código genético, a una secuencia de nucleótidos de (a) ó (b);

(e) moléculas de ácido nucleico que codifican un polipéptido derivado del polipéptido cuya secuencia tiene una identidad del 51,3% ó más con la secuencia de aminoácidos del polipéptido codificado por una molécula de ácido nucleico de (a) ó (b);

(g) moléculas de ácido nucleico que comprenden un polinucleótido que tiene una secuencia de una molécula de ácido nucleico amplificada a partir de una biblioteca de ácido nucleico de *Phaffia*, empleando los cebadores representados en SEQ ID NO:4, 5 y 6;

(j) moléculas de ácido nucleico que codifican un polipéptido que tiene una actividad escualeno sintasa, en donde dicho polipéptido es reconocido por anticuerpos que han sido generados contra un polipéptido codificado por una molécula de ácido nucleico de una cualquiera (a), (b), (c) y (g);

(k) moléculas de ácido nucleico que pueden obtenerse por exploración de una biblioteca apropiada en condiciones restrictivas con una sonda que tiene la secuencia de la molécula de ácido nucleico de una cualquiera de (a), (b), (c), (e), (g) y (j), y codificando un polipéptido que tiene actividad escualeno sintasa.

2. Un organismo recombinante cuya expresión génica de la escualeno sintasa se ha reducido en comparación con la del organismo anfitrión, por lo cual es capaz de producir carotenoides en un nivel potenciado en relación con el organismo anfitrión, **caracterizado** porque el organismo recombinante es una célula fúngica que contiene un polinucleótido antisentido frente a una molécula de ácido nucleico seleccionada del grupo formado por:

(m) moléculas de ácido nucleico que comprenden la secuencia de nucleótidos como está representada en SEQ ID NO:1;

(n) moléculas de ácido nucleico cuya secuencia de nucleótidos está degenerada como resultado del código genético, a una secuencia de nucleótidos de (m);

(p) moléculas de ácido nucleico que codifican un polipéptido derivado del polipéptido cuya secuencia tiene una identidad del 51,3% ó más con la secuencia de aminoácidos del polipéptido codificado por una molécula de ácido nucleico de (m);

(q) moléculas de ácido nucleico que comprenden un fragmento codificado por una molécula de ácido nucleico de una cualquiera de (m), (n) ó (p) y tienen una actividad escualeno sintasa;

(r) moléculas de ácido nucleico que comprenden un polinucleótido que tiene una secuencia de una molécula de ácido nucleico amplificada a partir de una biblioteca de ácido nucleico de *Phaffia*, empleando los cebadores representados en SEQ ID NO:4, 5 y 6;

(s) moléculas de ácido nucleico que codifican un polipéptido que tiene una actividad escualeno sintasa, en donde dicho polipéptido es un fragmento de un polipéptido codificado por una cualquiera de (m), (n), (p), (q) y (r);

(t) moléculas de ácido nucleico que comprenden por lo menos 15 nucleótidos de un polinucleótido de (m) ó (a);

(u) moléculas de ácido nucleico que codifican un polipéptido que tiene una actividad escualeno sintasa, en donde dicho polipéptido es reconocido por anticuerpos que han sido generados contra un polipéptido codificado por una molécula de ácido nucleico de uno cualquiera de (m), (n), (p), (q), (r) y (s);

(v) moléculas de ácido nucleico que pueden obtenerse explorando una biblioteca apropiada en condiciones restrictivas con una sonda que tiene la secuencia de la molécula de ácido nucleico de uno cualquiera de (m), (n), (p), (q), (r), (s), (t) y (u) y codifican un polipéptido que tiene actividad escualeno sintasa;

## ES 2 305 501 T3

(w) moléculas de ácido nucleico cuya cadena complementaria hibrida bajo condiciones restrictivas con una molécula de ácido nucleico de una cualquiera de (m), (n), (p), (q), (r), (s), (t), (u) y (v) y codifican un polipéptido que tiene actividad escualeno sintasa.

5 3. El organismo recombinante de la reivindicación 1 ó 2, que contiene un vector recombinante que comprende el polinucleótido antisentido.

10 4. El vector como se ha definido en la reivindicación 3, en el cual el polinucleótido antisentido está operativamente unido a las secuencias de control de la expresión que permiten la expresión en células fúngicas.

5 10 5. Un método para obtener un organismo recombinante, el cual comprende la introducción del vector de la reivindicación 4 en un organismo anfitrión, en donde dicho organismo anfitrión pertenece a una cepa de *Phaffia rhodozyma* ó *Xanthophylomyces dendrorhous*.

15 6. Un procedimiento para la producción de carotenoides, el cual comprende el cultivo de un organismo recombinante de una cualquiera de las reivindicaciones 1 a 3, en donde dichos carotenoides se seleccionan del grupo formado por la astaxantina,  $\beta$ -caroteno, licopeno, zeaxantina, cantaxantina.

20

25

30

35

40

45

50

55

60

65

# ES 2 305 501 T3

## LISTA DE SECUENCIAS

<110> Roche vitamins AG

5 <120> gen SOS

<130> NDR5218

10 <160> 8

<170> PatentIn version 3.1

15 <210> 1  
<211> 4807  
<212> ADN  
<213> *Phaffia rhodozyma*

20 <220>  
<221> exon  
<222> (1550)..(1577)

25 <223>

<220>

30 <221> polyA\_site  
<222> (4106)..(4107)  
<223>

<220>

35 <221> 5'UTR  
<222> (1464)..(1470)  
<223>

40 <220>  
<221> exon  
<222> (3959)..(3970)

45 <223>

<220>

50 <221> Intron  
<222> (3882)..(3958)  
<223>

<220>

55 <221> exon  
<222> (3567)..(3881)  
<223>

60 <220>  
<221> Intron  
<222> (3476)..(3564)

65 <223>

<220>

## ES 2 305 501 T3

<221> exon  
<222> (3464)..(3474)  
<223>  
5  
<220>  
<221> Intron  
<222> (3357)..(3453)  
10 <223>  
  
<220>  
<221> exon  
15 <222> (3231)..(3356)  
<223>  
  
<220>  
20 <221> Intron  
<222> (3088)..(3230)  
<223>  
  
25 <220>  
<221> exon  
<222> (2476)..(3087)  
30 <223>  
  
<220>  
<221> Intron  
35 <222> (2398)..(2474)  
<223>  
  
<220>  
40 <221> exon  
<222> (2183)..(2395)  
<223>  
  
45 <220>  
<221> Intron  
<222> (2072)..(2182)  
50 <223>  
  
<220>  
<221> exon  
55 <222> (1883)..(2071)  
<223>  
  
<220>  
60 <221> Intron  
<222> (1767)..(1882)  
<223>  
  
65 <220>  
<221> axon

# ES 2 305 501 T3

<222> (1755)..(1766)

<223>

5 <220>

<221> Intron

<222> (1578)..(1752)

<223>

10

<400> 1

15	gttcctgttc agtcaaagag tgggaaaaac atgaaagtaa aaagatgtaa tgaaagaagg	60
	ggtcagaaca tcggagatac aatggcccat agaggaagga aagctactta ccagaaacca	120
	gtgaggtttg cctaggaagt aatcccttcg tttctcaaag atatcttttt tgaaagcadc	180
20	gatgaacgac atgtcgaacc catctccatc ctcgaaatca agtttactcg atttagacct	240
	ttccagcttt tctgctctct ccagtttcgc agctttctct tcgggaagaa gctctccgcc	300
	agtcgatgtt ctgtcgacag gagaccagta gaaggcggaa ccgacaattt tggatggatc	360
25	ggaggacagg gtggctttaa caaatcggta gtacggagga tcgaacggcg cttctctcgt	420
	tcgaaggttg actcctcttg ctatgtgtat gagagcatat ccgttgatgt ctcagttaaa	480
	atctcctttt cttcttaccg ggagagtaag acacacaaaag aatcacgaag aatatgatga	540
30	ctgaccgatc cgaatatcta gcgcaggttg cttctctact ggttccattc ttcgaacgat	600
	aggttcatgt ttgaaagcat tgatcctagt tgcctctatc tgaggccagt ctgccaatgt	660
	agcaggctca atgatcactt ggggtttgtg catcttgatg ttcaaccaag tgtcgcaacg	720
35	gtcgaagattc tttttcttc ttttggctga gaaaaaaaaa cggcttcgct tcgcacgcgc	780
	gcgggggatc acccgcatat taagcggat gacgctcatc aaccggccaa gtgttcttca	840
	tcataaggtga aggttaaaac ggaatggata ggaggagcta accacgtttt tattttaatt	900
40	cgacttgggc agcctcgtcc atagtgtctg atggttatat cgtcatagaa aggcagcggc	960
	tggcgggttc gtcatggccg tgatcatctg ctttgttaga cattgtccat cagtcacctc	1020
45	aatgacagtt tcccgaagcc atcactaaga cacaaacgta tccagcacgc catgtccatc	1080
	actgaagaag gtagggtctc gtcgagccag tgcaaccaga gttacagatg aacatcaggc	1140
	cttgatcaga cccgacttat gaatatggcc gttattgtac acttcttggg gtcctctcag	1200
50	ctgctctttc gtgtttttca ctttctttcc ggatcaaacg agactgctcg tgtatctatc	1260
	tgtgcttggc atargagcat cccatgcctc tgetcaaatg atgctggagc tacgatccat	1320
	cagagacgac acaaaacggg gttgtatgaa ctctacattt cctaattgta ttggaatttt	1380
55	ctgtaatgctg ttcttcatct ttctctaatt cttttttgta gtccgctttt tcaaccttgc	1440

60

65

ES 2 305 501 T3

	cagcgtttcg cgtgtcttct tcttcctttg acgggtcatca ctttcttctc tcttctcgtt	1500
5	ccttcttccg tcttctcttc tctctcttcg tctgaacatc agcatcatc atg ggc ata Met Gly Ile 1	1558
	tca gat tac ctc gtt ctg g gtcagttctg tcttttgttt gattcttctc	1607
	Ser Asp Tyr Leu Val Leu 5	
10	tctttgcccg cngtcgcctg tcttgggtat atcatcaqca atgagaaaca tgatqttccc	1667
	cccgcgtcaa tcaactgacct tttggctctc tacttcttctc ctgtcgaatt gatcctgatt	1727
15	gatacgtgtg ccggtgctt aacagct tt cac gca tcc t gtaggtgttt	1776
	Val His Ala Ser	
	tatcgtatgc ttcattgtga tgttttagtca cgcggactga cctggccggt tgattttctg	1836
20	tatgatcgtc tgtgctaccg tctttctttg aaatccttcc catcag gc cga tct	1890
	Cys Arg Ser 15	
	gcg agc ttt aat gca gta cgc gat ctg gca tga gcc tcg aag gaa tat	1938
25	Ala Ser Phe Asn Ala Val Arg Asp Leu Ala Ala Ser Lys Glu Tyr 20 25 30	
	cac tgc aca gga gga aca tgc aac atc cgg ttg gga ccg aga aac tat	1986
	His Cys Thr Gly Gly Thr Cys Asn Ile Arg Leu Gly Pro Arg Asn Tyr 35 40 45	
30	gaa gga atg ttg gaa gta ttt gga tct gac ttc aag aag ttt cgc agc	2034
	Glu Gly Met Leu Glu Val Phe Gly Ser Asp Phe Lys Lys Phe Arg Ser 50 55 60	
	tgt cat caa aga gtt gga cgg aga tct tac ccg agt c gtacgtgttt	2081
35	Cys His Gln Arg Val Gly Arg Arg Ser Tyr Pro Ser 65 70 75	
	tcattctctc tctcctttga gatctggtcg cctccgcatt ttcttgttgc agaagggtca	2141
40	gaagctgaca acaccatctc tactgttcgg gacacggcta g at ctg ttt att cta	2196
	His Leu Phe Ile Leu 80	
	tct cgc tct tcg agg act gga tac cat tga gga tga cat gag tct atc	2244
	Ser Arg Ser Ser Arg Thr Gly Tyr His Gly His Glu Ser Ile 85 90	
45	taa tga tgt gaa gct tcc cct gct tcg gac att ctg gga aaa gct tga	2292
	Cys Glu Ala Ser Pro Ala Ser Asp Ile Leu Gly Lys Ala 95 100 105	
	ctc ccc tgg gtg gac ctt tac tgg atc cgg tcc aaa tga gaa gga tag	2340
50	Leu Pro Trp Val Asp Leu Tyr Trp Ile Arg Ser Lys Glu Gly 110 115 120	
	aga gct tct tgt tca ctt cga tgt ggc cat cgc cga gtt tgc caa ctt	2388
55	Arg Ala Ser Cys Ser Leu Arg Cys Gly His Arg Arg Val Cys Gln Leu 125 130 135	
	gga cgt c aagtgagttt ccctttatgg ttggatcatc cgctcgacag actcgaaacg	2445
	Gly Arg	
60	ctcatcactt tgggtctgctt gatgaacagc tc tcg gaa cgt cat tcg aga cat	2498

65

ES 2 305 501 T3

Leu Ser Glu Arg His Ser Arg His  
145

5	cac	tcg	caa	gat	ggg	taa	cgg	tat	ggc	cga	ctt	tgc	ttc	tct	ctc	tac	2546
	His	Ser	Gln	Asp	Gly		Arg	Tyr	Gly	Arg	Leu	Cys	Phe	Ser	Leu	Tyr	
			150						155					160			
10	gcc	ctc	caa	gcc	tgt	ggc	cga	ggt	cca	gtc	gac	cga	aga	ttt	caa	cct	2594
	Ala	Leu	Gln	Ala	Cys	Gly	Arg	Gly	Pro	Val	Asp	Arg	Arg	Phe	Gln	Pro	
			165					170					175				
15	ata	ctg	tca	tta	cgt	cgc	tgg	act	cgt	cgg	cga	ggg	act	ctc	ccg	act	2642
	Ile	Leu	Ser	Leu	Arg	Arg	Trp	Thr	Arg	Arg	Arg	Gly	Thr	Leu	Pro	Thr	
		180					185					190					
20	ctt	tgt	cgc	gac	cga	gaa	gga	acg	acc	att	ctt	ggc	caa	cca	gat	ggt	2690
	Leu	Cys	Arg	Asp	Arg	Glu	Gly	Thr	Thr	Ile	Leu	Gly	Gln	Pro	Asp	Gly	
						200					205					210	
25	act	ttc	aaa	ctc	gtt	cgg	act	cct	tct	cca	aaa	gac	aaa	cat	cct	tcg	2738
	Thr	Phe	Lys	Leu	Val	Arg	Thr	Pro	Ser	Pro	Lys	Asp	Lys	His	Pro	Ser	
					215					220					225		
30	aga	tat	tcg	gga	gga	cgc	cga	cga	agg	tcg	tgg	ctt	ctg	gcc	aag	aga	2786
	Arg	Tyr	Ser	Gly	Gly	Arg	Arg	Arg	Arg	Ser	Trp	Leu	Leu	Ala	Lys	Arg	
				230					235					240			
35	gat	ctg	ggc	caa	ccc	gat	cta	tac	tgc	gca	tgc	acc	ggg	cac	aag	gtt	2834
	Asp	Leu	Gly	Gln	Pro	Asp	Leu	Tyr	Cys	Ala	Cys	Thr	Gly	His	Lys	Val	
			245					250					255				
40	taa	ctc	gtt	gac	tga	cct	ggt	caa	gaa	aga	aaa	cat	cga	caa	agg	atc	2882
		Leu	Val	Asp		Pro	Gly	Gln	Glu	Arg	Lys	His	Arg	Gln	Arg	Ile	
			260					265						270			
45	aat	gtg	ggt	ggt	gag	tgc	gat	gac	act	cga	cgc	gat	cac	cca	tac	tac	2930
	Asn	Val	Gly	Val	Glu	Cys	Asp	Asp	Thr	Arg	Arg	Asp	His	Pro	Tyr	Tyr	
			275					280					285				
50	cga	cgc	act	gga	cta	cct	ctc	act	tct	aaa	gaa	cca	gag	tgt	ttt	caa	2978
	Arg	Arg	Thr	Gly	Leu	Pro	Leu	Thr	Ser	Lys	Glu	Pro	Glu	Cys	Phe	Gln	
			290				295					300					
55	ctt	ttg	tgc	tat	ccc	ggc	tgt	cat	gtc	gat	tgc	aac	ggt	gga	gct	atg	3026
	Leu	Leu	Cys	Tyr	Pro	Gly	Cys	His	Val	Asp	Cys	Asn	Val	Gly	Ala	Met	
						310					315					320	
60	ctt	cat	gaa	ccc	agc	ggt	ggt	cca	acg	aaa	cat	aaa	aat	cag	aaa	ggg	3074
	Leu	His	Glu	Pro	Ser	Gly	Val	Pro	Thr	Lys	His	Lys	Asn	Gln	Lys	Gly	
					325					330					335		
65	aga	agc	cgt	cga	g	gtg	cgttcgc	gcgttctggt	tctac	ctttc	ataac	attgg					3127
	Arg	Ser	Arg	Arg													
				340													
50	agg	tcttga	ctctta	agcg	tcttcca	atc	tgatgc	cctcc	aattat	catc	attttt	gtct					3187
55	gtg	caa	caa	ccc	tcg	gga	ggt	ggc	ata	cat	gtt	tag	aga	tta	tgc	tcg	3289
	Val	Gln	Gln	Pro	Ser	Gly	Gly	Gly	Ile	His	Val		Arg	Leu	Cys	Ser	
						350					355						
60	aaa	gat	tca	tgc	caa	ggc	tat	tcc	tac	aga	tcc	taa	ctt	cat	caa	ggt	3337
	Lys	Asp	Ser	Cys	Gln	Gly	Tyr	Ser	Tyr	Arg	Ser		Leu	His	Gln	Val	
						365					370						

ES 2 305 501 T3

gag cgt tgc gtg tgg tcg a gtgagttgat cgategatcc atcttttggc 3386  
 Glu Arg Cys Val Trp Ser  
 375 380

5 ttgatcatcg cgagacttga ctgatcgatt actcaaaaca tcategettc tctttcttgc 3446

tctctag at cga aca atg ggc tga gca c tgtatgttcc tccgccctc 3494  
 Asn Arg Thr Met Gly Ala  
 385

10 cttcaagttt cctctcgtt catcttttggc gagaagaggg atctgatgta tctttcttgg 3554

ttcggatcag ac ta ccc ctc att tat gat gat tgc gcc ttc gaa tga ccc 3604  
 Leu Pro Leu Ile Tyr Asp Asp Ser Ala Phe Glu Pro  
 390

15 tca aaa ccc cgc acc ctc aac ggc gct tga ccc ttt ctc agg aga cgc 3652  
 Ser Lys Pro Arg Thr Leu Asn Gly Ala Pro Phe Leu Arg Arg Arg  
 400 405 410

20 tgc ttt aag gat agc ctc taa gaa ggc tga gat cac cgc cgc tgc tct 3700  
 Ser Phe Lys Asp Ser Leu Glu Gly Asp His Arg Arg Cys Ser  
 415 420 425

25 tgt cag gaa gaa agc ccg gga tca cgc taa gtg gag aga gtc caa ggg 3748  
 Cys Gln Glu Glu Ser Pro Gly Ser Arg Val Glu Arg Val Gln Gly  
 430 435 440

att gcc tcc gag cga tcc gaa caa gcc gga caa ctc gga gga tgt taa 3796  
 Ile Ala Ser Glu Arg Ser Glu Gln Ala Gly Gln Leu Gly Gly Cys  
 445 450 455

30 ttg ggt att gat cgg cgg tat gat cgt tgg att gtt gct cgt gat ggg 3844  
 Leu Gly Ile Asp Arg Arg Tyr Asp Arg Trp Ile Val Ala Arg Asp Gly  
 460 465 470

35 cgt gct cgg ttt ggc tat cgc ttg ggt tgt tct tca g gtgcgttctt 3891  
 Arg Ala Arg Phe Gly Tyr Arg Leu Gly Cys Ser Ser  
 475 480 485

40 ccaaagagcc tttctctcat gaacacgcac ataggttgat ctaattctat cttactctgt 3951

catacag tt tga gca ata a tctcaagatt ctagtccatc ctttcgctca 4000  
 val Ala Ile

acgatctgct tcttctcctt cctctctctc gtcttctctg gtttcttttc ttactttctg 4060

45 ggatcttctt tcttgaatcc tccgatccaa tgraatctgc ataccctcgc tttagtagaa 4120

accgatcctt ccttcgatct tggcgaaaat ctaagcaaag agaatcactt ttgtctaata 4180

aaatttctt taaagagtcg gcttttctt gtggcgaagc ttcacccgt cttctcttgg 4240

50 accatctctt ctcaatattc ttgtgtctac tatatgatca agttctttga aatcaaagaa 4300

gaacatgtat ttgattttga ggttccaaga atacaaccgg cccaagtcgt tcttcgcagt 4360

ttcatcaga cagcacatat ctctctctct ctctatagaa gccgtatggg gccaatcgac 4420

55 tctcatgggt agaccgtgcc cttttgacac ggggagaaag agaacgaaag gacacttgac 4480

cgattcgtta ataaagccgt ccccacctt tctttaatgg caattcaaga agagaaaaac 4540

aacccttgcg cgcactcgag tagtcgatca gaccttccga acgacagata tcatttgcctg 4600

60 aaatcgaccg gattttaaag ctgctgccag gtcgggtgaat ccccttaggt gatctcttgg 4660

tacaagatg ttgggcacgg acttttgcac ccggatgaga acgtcgtgaa gagtttgaaa 4720

65 aagattatca acataatgtg tcttttttct ttttttctt cgtaactctc tagagaacga 4780

ggagacgtac ggtctgattt gttatcg 4807

ES 2 305 501 T3

<210> 2

<211> 1536

<212> ADN

5 <213> *Phaffia rhodozyma*

<220>

<221> CDS

10 <222> (1)..(1536)

<223>

<400> 2

15

atg ggc ata tca gat tac ctc gtt ctg gct ttc acg cat cct gcc gat 48  
Met Gly Ile Ser Asp Tyr Leu Val Leu Ala Phe Thr His Pro Ala Asp  
1 5 10 15

20

ctg cga gct tta atg cag tac gcg atc tgg cat gag cct cga agg aat 96  
Leu Arg Ala Leu Met Gln Tyr Ala Ile Trp His Glu Pro Arg Arg Asn  
20 25 30

25

atc act gca cag gag gaa cat gca aca tcc ggt tgg gac cga gaa act 144  
Ile Thr Ala Gln Glu Glu His Ala Thr Ser Gly Trp Asp Arg Glu Thr  
35 40 45

30

atg aag gaa tgt tgg aag tat ttg gat ctg act tca aga agt ttc gca 192  
Met Lys Glu Cys Trp Lys Tyr Leu Asp Leu Thr Ser Arg Ser Phe Ala  
50 55 60

35

gct gtc atc aaa gag ttg gac gga gat ctt acc cga gtc atc tgt tta 240  
Ala Val Ile Lys Glu Leu Asp Gly Asp Leu Thr Arg Val Ile Cys Leu  
65 70 75 80

40

ttc tat ctc gct ctt cga gga ctg gat acc att gag gat gac atg agt 288  
Phe Tyr Leu Ala Leu Arg Gly Leu Asp Thr Ile Glu Asp Asp Met Ser  
85 90 95

45

cta tct aat gat gtg aag ctt ccc ctg ctt cgg aca ttc tgg gaa aag 336  
Leu Ser Asn Asp Val Lys Leu Pro Leu Leu Arg Thr Phe Trp Glu Lys  
100 105 110

50

ctt gac tcc cct ggg tgg acc ttt act gga tcc ggt cca aat gag aag 384  
Leu Asp Ser Pro Gly Trp Thr Phe Thr Gly Ser Gly Pro Asn Glu Lys  
115 120 125

55

gat aga gag ctt ctt gtc cac ttc gat gtg gcc atc gcc gag ttt gcc 432  
Asp Arg Glu Leu Leu Val His Phe Asp Val Ala Ile Ala Glu Phe Ala  
130 135 140

60

aac ttg gac gtc aac tct cgg aac gtc att cga gac atc act cgc aag 480  
Asn Leu Asp Val Asn Ser Arg Asn Val Ile Arg Asp Ile Thr Arg Lys  
145 150 155 160

65

60

65

ES 2 305 501 T3

	atg ggt aac ggt atg gcc gac ttt gct tct ctc tct acg ccc tcc aag	528
	Met Gly Asn Gly Met Ala Asp Phe Ala Ser Leu Ser Thr Pro Ser Lys	
	165 170 175	
5	cct gtg gcc gag gtc cag tgc acc gaa gat ttc aac cta tac tgt cat	576
	Pro Val Ala Glu Val Gln Ser Thr Glu Asp Phe Asn Leu Tyr Cys His	
	180 185 190	
10	tac gtc gct gga ctc gtc ggc gag gga ctc tcc cga ctc ttt gtc gcg	624
	Tyr Val Ala Gly Leu Val Gly Glu Gly Leu Ser Arg Leu Phe Val Ala	
	195 200 205	
15	acc gag aag gaa cga cca ttc ttg gcc aac cag atg gta ctt tca aac	672
	Thr Glu Lys Glu Arg Pro Phe Leu Ala Asn Gln Met Val Leu Ser Asn	
	210 215 220	
20	tcg ttc gga ctc ctt ctc caa aag aca aac atc ctt cga gat att cgg	720
	Ser Phe Gly Leu Leu Leu Gln Lys Thr Asn Ile Leu Arg Asp Ile Arg	
	225 230 235 240	
25	gag gac gcc gac gaa ggt cgt ggc ttc tgg cca aga gag atc tgg gcc	768
	Glu Asp Ala Asp Glu Gly Arg Gly Phe Trp Pro Arg Glu Ile Trp Ala	
	245 250 255	
30	aac ccg atc tat act gcg cat gca ccg ggc aca agg ttt aac tcg ttg	816
	Asn Pro Ile Tyr Thr Ala His Ala Pro Gly Thr Arg Phe Asn Ser Leu	
	260 265 270	
35	act gac ctg gtc aag aaa gaa aac atc gac aaa gga tca atg tgg gtg	864
	Thr Asp Leu Val Lys Lys Glu Asn Ile Asp Lys Gly Ser Met Trp Val	
	275 280 285	
40	ttg agt gcg atg aca ctc gac gcg atc acc cat act acc gac gca ctg	912
	Leu Ser Ala Met Thr Leu Asp Ala Ile Thr His Thr Thr Asp Ala Leu	
	290 295 300	
45	gac tac ctc tca ctt cta aag aac cag agt gtt ttc aac ttt tgt gct	960
	Asp Tyr Leu Ser Leu Leu Lys Asn Gln Ser Val Phe Asn Phe Cys Ala	
	305 310 315 320	
50	atc ccg gct gtc atg tgc att gca acg ttg gag cta tgc ttc atg aac	1008
	Ile Pro Ala Val Met Ser Ile Ala Thr Leu Glu Leu Cys Phe Met Asn	
	325 330 335	
55	cca gcg gtg ttc caa cga aac ata aaa atc aga aag gga gaa gcc gtc	1056
	Pro Ala Val Phe Gln Arg Asn Ile Lys Ile Arg Lys Gly Glu Ala Val	
	340 345 350	
60	gag ctc att atg aag tgc aac aac cct cgg gag gtg gca tac atg ttt	1104
	Glu Leu Ile Met Lys Cys Asn Asn Pro Arg Glu Val Ala Tyr Met Phe	
	355 360 365	
65	aga gat tat gct cga aag att cat gcc aag gct att cct aca gat cct	1152
	Arg Asp Tyr Ala Arg Lys Ile His Ala Lys Ala Ile Pro Thr Asp Pro	
	370 375 380	
70	aac ttc atc aag ttg agc gtt gcg tgt ggt cga atc gaa caa tgg gct	1200
	Asn Phe Ile Lys Leu Ser Val Ala Cys Gly Arg Ile Glu Gln Trp Ala	
	385 390 395 400	
75	gag cac tac tac ccc tca ttt atg atg att cgg cct tcg aat gac cct	1248
	Glu His Tyr Tyr Pro Ser Phe Met Met Ile Arg Pro Ser Asn Asp Pro	
	405 410 415	
80	caa aac ccc gca ccc tca acg gcg ctt gac cct ttc tca gga gac gct	1296
	Gln Asn Pro Ala Pro Ser Thr Ala Leu Asp Pro Phe Ser Gly Asp Ala	
	420 425 430	

ES 2 305 501 T3

cgt tta agg ata gcc tct aag aag gct gag atc acc gcc gct gct ctt 1344  
 Arg Leu Arg Ile Ala Ser Lys Lys Ala Glu Ile Thr Ala Ala Ala Leu  
 435 440 445  
 5 gtc agg aag aaa gcc cgg gat cac gct aag tgg aga gag tcc aag gga 1392  
 Val Arg Lys Lys Ala Arg Asp His Ala Lys Trp Arg Glu Ser Lys Gly  
 450 455 460  
 10 ttg cct ccg agc gat ccg aac aag ccg gac aac tcg gag gat gtt aat 1440  
 Leu Pro Pro Ser Asp Pro Asn Lys Pro Asp Asn Ser Glu Asp Val Asn  
 465 470 475 480  
 15 tgg gta ttg atc ggc ggt atg atc gct gga ttg ttg ctc gtg atg ggc 1488  
 Trp Val Leu Ile Gly Gly Met Ile Val Gly Leu Leu Leu Val Met Gly  
 485 490 495  
 20 gtg ctc ggt ttg gct atc gct tgg gtt gtt ctt cag ttt gag caa taa 1536  
 Val Leu Gly Leu Ala Ile Ala Trp Val Leu Gln Phe Glu Gln  
 500 505 510  
 <210> 3  
 <211> 511  
 <212> PRT  
 25 <213> *Phaffia rhodozyma*  
 <400> 3  
 30 Met Gly Ile Ser Asp Tyr Leu Val Leu Ala Phe Thr His Pro Ala Asp  
 1 5 10 15  
 35 Leu Arg Ala Leu Met Gln Tyr Ala Ile Trp His Glu Pro Arg Arg Asn  
 20 25 30  
 40 Ile Thr Ala Gln Glu Glu His Ala Thr Ser Gly Trp Asp Arg Glu Thr  
 35 40 45  
 45 Met Lys Glu Cys Trp Lys Tyr Leu Asp Leu Thr Ser Arg Ser Phe Ala  
 50 55 60  
 50 Ala Val Ile Lys Glu Leu Asp Gly Asp Leu Thr Arg Val Ile Cys Leu  
 65 70 75 80  
 55 Phe Tyr Leu Ala Leu Arg Gly Leu Asp Thr Ile Glu Asp Asp Met Ser  
 85 90 95  
 60 Leu Ser Asn Asp Val Lys Leu Pro Leu Leu Arg Thr Phe Trp Glu Lys  
 100 105 110  
 65 Leu Asp Ser Pro Gly Trp Thr Phe Thr Gly Ser Gly Pro Asn Glu Lys  
 115 120 125  
 Asp Arg Glu Leu Leu Val His Phe Asp Val Ala Ile Ala Glu Phe Ala  
 130 135 140

ES 2 305 501 T3

5      **Asn** **Leu** **Asp** **Val** **Asn** **Ser** **Arg** **Asn** **Val** **Ile** **Arg** **Asp** **Ile** **Thr** **Arg** **Lys**  
       145                           150                           155                           160  
 10     **Met** **Gly** **Asn** **Gly** **Met** **Ala** **Asp** **Phe** **Ala** **Ser** **Leu** **Ser** **Thr** **Pro** **Ser** **Lys**  
                           165                           170                           175  
 15     **Pro** **Val** **Ala** **Glu** **Val** **Gln** **Ser** **Thr** **Glu** **Asp** **Phe** **Asn** **Leu** **Tyr** **Cys** **His**  
                           180                           185                           190  
 20     **Tyr** **Val** **Ala** **Gly** **Leu** **Val** **Gly** **Glu** **Gly** **Leu** **Ser** **Arg** **Leu** **Phe** **Val** **Ala**  
                           195                           200                           205  
 25     **Thr** **Glu** **Lys** **Glu** **Arg** **Pro** **Phe** **Leu** **Ala** **Asn** **Gln** **Met** **Val** **Leu** **Ser** **Asn**  
                           210                           215                           220  
 30     **Ser** **Phe** **Gly** **Leu** **Leu** **Leu** **Gln** **Lys** **Thr** **Asn** **Ile** **Leu** **Arg** **Asp** **Ile** **Arg**  
       225                           230                           235                           240  
 35     **Glu** **Asp** **Ala** **Asp** **Glu** **Gly** **Arg** **Gly** **Phe** **Trp** **Pro** **Arg** **Glu** **Ile** **Trp** **Ala**  
                           245                           250                           255  
 40     **Asn** **Pro** **Ile** **Tyr** **Thr** **Ala** **His** **Ala** **Pro** **Gly** **Thr** **Arg** **Phe** **Asn** **Ser** **Leu**  
                           260                           265                           270  
 45     **Thr** **Asp** **Leu** **Val** **Lys** **Lys** **Glu** **Asn** **Ile** **Asp** **Lys** **Gly** **Ser** **Met** **Trp** **Val**  
                           275                           280                           285  
 50     **Leu** **Ser** **Ala** **Met** **Thr** **Leu** **Asp** **Ala** **Ile** **Thr** **His** **Thr** **Thr** **Asp** **Ala** **Leu**  
                           290                           295                           300  
 55     **Asp** **Tyr** **Leu** **Ser** **Leu** **Leu** **Lys** **Asn** **Gln** **Ser** **Val** **Phe** **Asn** **Phe** **Cys** **Ala**  
       305                           310                           315                           320  
 60     **Ile** **Pro** **Ala** **Val** **Met** **Ser** **Ile** **Ala** **Thr** **Leu** **Glu** **Leu** **Cys** **Phe** **Met** **Asn**  
                           325                           330                           335  
 65     **Pro** **Ala** **Val** **Phe** **Gln** **Arg** **Asn** **Ile** **Lys** **Ile** **Arg** **Lys** **Gly** **Glu** **Ala** **Val**  
                           340                           345                           350  
 70     **Glu** **Leu** **Ile** **Met** **Lys** **Cys** **Asn** **Asn** **Pro** **Arg** **Glu** **Val** **Ala** **Tyr** **Met** **Phe**  
                           355                           360                           365  
 75     **Arg** **Asp** **Tyr** **Ala** **Arg** **Lys** **Ile** **His** **Ala** **Lys** **Ala** **Ile** **Pro** **Thr** **Asp** **Pro**  
                           370                           375                           380  
 80     **Asn** **Phe** **Ile** **Lys** **Leu** **Ser** **Val** **Ala** **Cys** **Gly** **Arg** **Ile** **Glu** **Gln** **Trp** **Ala**  
       385                           390                           395                           400  
 85     **Glu** **His** **Tyr** **Tyr** **Pro** **Ser** **Phe** **Met** **Met** **Ile** **Arg** **Pro** **Ser** **Asn** **Asp** **Pro**  
                           405                           410                           415



## ES 2 305 501 T3

<210> 5  
<211> 26  
<212> ADN  
5 <213> Artificial

<220>  
<221> características diversas  
10 <222> (3)..(3)  
<223> n = a, c, g o t

<220>  
15 <221> características diversas  
<222> (9)..(9)  
<223> n = a, c, g o t

<220>  
20 <221> características diversas  
<222> (15)..(15)  
25 <223> n = a, c, g o t

<220>  
<221> características diversas  
30 <222> (18)..(18)  
<223> n = a, c, g o t

<220>  
35 <221> características diversas  
<222> (21)..(21)  
<223> n = a, c, g o t

40 <400> 5

atngccatna cytgnggnat ngrca

26

45 <210> 6  
<211> 29  
<212> ADN  
50 <213> Artificial

<220>  
<221> características diversas  
55 <222> (3)..(3)  
<223> n = a, c, g o t

<220>  
60 <221> características diversas  
<222> (6)..(6)  
<223> n = a, c, g o t

<220>  
65 <221> características diversas  
<222> (9)..(9)

## ES 2 305 501 T3

	<223> n = a, c, g o t	
	<220>	
5	<221> características diversas	
	<222> (12)..(12)	
	<223> n = a, c, g o t	
10	<220>	
	<221> características diversas	
	<222> (15)..(15)	
15	<223> n =a, c, g o t	
	<400> 6	
20	ccnacngtnc engcnacrta rtgrcarta	29
	<210> 7	
	<211> 20	
25	<212> ADN	
	<213> Artificial	
	<400> 7	
30	aatgatgtgta agcttcccct	20
	<210> 8	
35	<211> 20	
	<212> ADN	
	<213> Artificial	
40	<400> 8	
	ccagatctct ctggccaga	20
45	<b>LISTADO DE SECUENCIAS</b>	
	<110> DSM IP ASSETS B.V.	
50	<120> gen SQS	
	<130> NDR5218	
55	<140> PCT/EP03/10573	
	<141> 2003-09-23	
60	<150> EP 02021619.8	
	<151> 2002-09-27	
	<160> 8	
65	<170> PatentIn version 3.2	
	<210> 1	

## ES 2 305 501 T3

<211> 4807  
<212> ADN  
<213> *Phaffia rhodozyma*

5  
<220>  
<221> 5'UTR  
<222> (1469)..(1470)

10  
<220>  
<221> exon  
<222> (1550)..(1577)

15  
<220>  
<221> Intron  
<222> (1578)..(1752)

20  
<220>  
<221> exon  
<222> (1753)..(1766)

25  
<220>  
<221> Intron  
<222> (1767)..(1882)

30  
<220>  
<221> exon  
<222> (1883)..(2071)

35  
<220>  
<221> Intron  
<222> (2072)..(2182)

40  
<220>  
<221> exon  
<222> (2183)..(2397)

45  
<220>  
<221> Intron  
<222> (2398)..(2474)

50  
<220>  
<221> exon  
<222> (2475)..(3087)

55  
<220>  
<221> Intron  
<222> (3088)..(3230)

60  
<220>  
<221> exon  
<222> (3231)..(3356)

65

## ES 2 305 501 T3

<220>

<221> Intron

<222> (3357)..(3453)

5

<220>

<221> exon

<222> (3454)..(3475)

10

<220>

<221> Intron

<222> (3476)..(3564)

15

<220>

<221> exon

<222> (3565)..(3881)

20

<220>

<221> Intron

<222> (3882)..(3958)

25

<220>

<221> exon

<222> (3959)..(3970)

30

<220>

<221> polyA\_site

<222> (4106)..(4107)

35

40

45

50

55

60

65

ES 2 305 501 T3

<400> 1

5	gttcctgttc agtcaaagag tgggaaaaac atgaaagtaa aaagatgtaa tgaagaagag	60
	ggtcagaaca tcggagatac aatggcccat agaggaagga aagctactta ccagaaacca	120
	gtgaggtttg cctaggaagt aatcccttcg tttctcaaag atatcttttt tgaagcatc	180
10	gatgaacgac atgtcgaacc catctccatc ctcgaaatca agtttactcg atttagacct	240
	ttccagcttt tctgctctct ccagtttcgc agctttctct tcggaagaa gctctccgcc	300
	agtcgatgtt ctgtcgacag gagaccagta gaaggcggaa ccgacaattt tggatggatc	360
15	ggaggacagg gtggctttta caaatcggta gtacggagga tcgaacggcg cttctctcgt	420
	tcgaaggttg actcctcttg ctatgtgtat gagagcatat ccgttgatgt ctcagttaaa	480
	atctcctttt cttcttacct ggagagtaag acacacaaag aatcacgaag aatatgatga	540
20	ctgaccgatc cgaatatcta gcgcaggttg cttctctact ggttccattc ttcgaacgat	600
	aggttcatgt ttgaaagcat tgatcctagt tgcctctatc tgaggccagt ctgccaatgt	660
	agcaggctca atgatactt ggggtttgtg catcttgatg ttcaaccaag tgtcgcaacg	720
25	gtcgagattc tttttcttc ttttggtcga gaaaaaaaaa cggcttcgct tcgcacgcgc	780
	gcgggggatc acccgcatat taagcggat gacgctcatc aaccggccaa gtgttcttca	840
	tcataaggtga aggttaaaac ggaatggata ggaggagcta accacgtttt tattttaatt	900
30	cgacttgggc agcctcgtcc atagtgtctg atggttatat cgtcatagaa aggcagcgc	960
	tggcgggttc gtcataggccg tgatcatctg ctttgttaga cattgtccat cagtcacctc	1020
	aatgacagtt tcccgcgc atcactaaga cacaaacgta tccagcgc catgtccatc	1080
35	actgaagaag gtagggctc gtcgagccag tgcaaccaga gttacagatg aacatcaggc	1140
	cttgatcaga cccgacttat gaatatggcc gttattgtac acttcttggg gctcctcgag	1200
40	ctgctctttc gtgtttttca ctttctttcc ggatcaaacg agactgctcg tgtatctatc	1260
	tgtgcttgcc atatgagcat cccatgcctc tgetcaaatg atgctggagc tacgatccat	1320
	cagagacgac acaaaacggg gttgtatgaa ctctacattt cctaattgta ttggaatttt	1380
45	ctgtaatgcg ttcttcatct ttctctaattg cttttttgta gtccgtcttt tcaaccttgc	1440
	cagcgtttcg cgtgtctctt ttctcctttg acggtcatca ctttcttctc tcttctcgtt	1500
	ctttcttccg tcttctctc tctctcttcg tctgaacatc agcatcatc atg ggc ata	1558
50		Met Gly Ile 1
	tca gat tac ctc gtt ctg g gtcagttctg tcttttgttt gattcttatc	1607
	Ser Asp Tyr Leu Val Leu	
55	5	
60		
65		



ES 2 305 501 T3

5 cga gac atc act cgc aag atg ggt aac ggt atg gcc gac ttt gct tct 2538  
 Arg Asp Ile Thr Arg Lys Met Gly Asn Gly Met Ala Asp Phe Ala Ser  
 155 160 165 170  
 5 ctc tct acg ccc tcc aag cct gtg gcc gag gtc cag tcg acc gaa gat 2586  
 Leu Ser Thr Pro Ser Lys Pro Val Ala Glu Val Gln Ser Thr Glu Asp  
 175 180 185  
 10 ttc aac cta tac tgt cat tac gtc gct gga ctc gtc gcc gag gga ctc 2634  
 Phe Asn Leu Tyr Cys His Tyr Val Ala Gly Leu Val Gly Glu Gly Leu  
 190 195 200  
 15 tcc cga ctc ttt gtc gcg acc gag aag gaa cga cca ttc ttg gcc aac 2682  
 Ser Arg Leu Phe Val Ala Thr Glu Lys Glu Arg Pro Phe Leu Ala Asn  
 205 210 215  
 20 cag atg gta ctt tca aac tcg ttc gga ctc ctt ctc caa aag aca aac 2730  
 Gln Met Val Leu Ser Asn Ser Phe Gly Leu Leu Leu Gln Lys Thr Asn  
 220 225 230  
 25 atc ctt cga gat att cgg gag gac gcc gac gaa ggt cgt gcc ttc tgg 2778  
 Ile Leu Arg Asp Ile Arg Glu Asp Ala Asp Glu Gly Arg Gly Phe Trp  
 235 240 245 250  
 30 cca aga gag atc tgg gcc aac ccg atc tat act gcg cat gca ccg gcc 2826  
 Pro Arg Glu Ile Trp Ala Asn Pro Ile Tyr Thr Ala His Ala Pro Gly  
 255 260 265  
 35 aca agg ttt aac tcg ttg act gac ctg gtc aag aaa gaa aac atc gac 2874  
 Thr Arg Phe Asn Ser Leu Thr Asp Leu Val Lys Lys Glu Asn Ile Asp  
 270 275 280  
 40 aaa gga tca atg tgg gtg ttg agt gcg atg aca ctc gac gcg atc acc 2922  
 Lys Gly Ser Met Trp Val Leu Ser Ala Met Thr Leu Asp Ala Ile Thr  
 285 290 295  
 45 cat act acc gac gca ctg gac tac ctc tca ctt cta aag aac cag agt 2970  
 His Thr Thr Asp Ala Leu Asp Tyr Leu Ser Leu Leu Lys Asn Gln Ser  
 300 305 310  
 50 gtt ttc aac ttt tgt gct atc ccg gct gtc atg tcg att gca acg ttg 3018  
 Val Phe Asn Phe Cys Ala Ile Pro Ala Val Met Ser Ile Ala Thr Leu  
 315 320 325 330  
 55 gag cta tgc ttc atg aac cca gcg gtg ttc caa cga aac ata aaa atc 3066  
 Glu Leu Cys Phe Met Asn Pro Ala Val Phe Gln Arg Asn Ile Lys Ile  
 335 340 345  
 60 aga aag gga gaa gcc gtc gag gtgcgttcgc gcgttctgtt tctaccttc 3117  
 Arg Lys Gly Glu Ala Val Glu  
 350  
 65 ataacattgg aggttcttga ctcttaagcg tcttccaatc tgatgcctcc aattatcatc 3177  
 atttttgtct tttttgcttt cctcttgttt ctttcggcg tgattcaatc cag ctc 3233  
 Leu  
 70 att atg aag tgc aac aac cct cgg gag gtg gca tac atg ttt aga gat 3281  
 Ile Met Lys Cys Asn Asn Pro Arg Glu Val Ala Tyr Met Phe Arg Asp  
 355 360 365 370  
 75 tat gct cga aag att cat gcc aag gct att cct aca gat cct aac ttc 3329  
 Tyr Ala Arg Lys Ile His Ala Lys Ala Ile Pro Thr Asp Pro Asn Phe  
 375 380 385  
 80 atc aag ttg agc gtt gcg tgt ggt cga gtgagttgat cgatcgatcc 3376  
 Ile Lys Leu Ser Val Ala Cys Gly Arg  
 390 395  
 85 atcttttgtt ttgatcatcg cgagacttga ctgatcgatt actcaaaaca tcatcgcttc 3436

ES 2 305 501 T3

tccttcttgc tctctag atc gaa caa tgg gct gag cac t gtatgttctt 3485  
 Ile Glu Gln Trp Ala Glu His  
 400

5 ccgccccctcc ttcaagtttc ctctcgettc atctttgttg agaagagggga tctgatgtat 3545

ctttctttgt tggatcag ac tac ccc tca ttt atg atg att cgg cct tcg 3596  
 Tyr Tyr Pro Ser Phe Met Met Ile Arg Pro Ser  
 405 410

10 aat gac cct caa aac ccc gca ccc tca acg gcg ctt gac cct ttc tca 3644  
 Asn Asp Pro Gln Asn Pro Ala Pro Ser Thr Ala Leu Asp Pro Phe Ser  
 415 420 425

15 gga gac gct cgt tta agg ata gcc tct aag aag gct gag atc acc gcc 3692  
 Gly Asp Ala Arg Leu Arg Ile Ala Ser Lys Lys Ala Glu Ile Thr Ala  
 430 435 440 445

gct gct ctt gtc agg aag aaa gcc cgg gat cac gct aag tgg aga gag 3740  
 Ala Ala Leu Val Arg Lys Lys Ala Arg Asp His Ala Lys Trp Arg Glu  
 450 455 460

20 tcc aag gga ttg cct ccg agc gat ccg aac aag ccg gac aac tcg gag 3788  
 Ser Lys Gly Leu Pro Pro Ser Asp Pro Asn Lys Pro Asp Asn Ser Glu  
 465 470 475

25 gat gtt aat tgg gta ttg atc ggc ggt atg atc gtt gga ttg ttg ctc 3836  
 Asp Val Asn Trp Val Leu Ile Gly Gly Met Ile Val Gly Leu Leu Leu  
 480 485 490

30 gtg atg ggc gtg ctc ggt ttg gct atc gct tgg gtt gtt ctt cag 3881  
 Val Met Gly Val Leu Gly Leu Ala Ile Ala Trp Val Val Leu Gln  
 495 500 505

gtgcgttctt ccaaagagcc tttctctcat gaacacgcac ataggttgat ctaattctat 3941

35 cttactctgt catacag ttt gag caa taa tctcaagatt ctagtccatc 3990  
 Phe Glu Gln  
 510

ctttcgtcga acgatctgct tcttctcctt ctcttctcc gtcttctctg gtttcttttc 4050

40 ttactttctg ggatcttctt tcttgaatcc tccgatccaa tgtaatctgc ataccctcgc 4110

tttagtagaa accgatcctt cattcgatct tggcgaaaat ctaagcaaag agaatcactt 4170

ttgtctaata aaatttctt taaagagtcg gctttttctt gtggcgaagc ttcacccctg 4230

45 cttctctggt accatctctt ctcaatattc tttgtgctac tatatgatca agttctttga 4290

aatcaaagaa gaacatgtat ttgattttga ggttccaaga atacaaccgg cccaagtcgt 4350

49 tcttcgcagt tttcatcaga cagcacatat ctctctcctt ctctatagaa gccgtagggg 4410

50 gccaatcgac tctcatgggt agaccgtgcc cttttgacac ggggagaaaag agaacgaaaag 4470

gacacttgac cgattcgta ataaagccgt ccccacctt tctttaatgg caattcaaga 4530

55 agagaaaaac aaccctgcg cgcactcgag tagtcgatca gaccttccga acgacagata 4590

tcatttgctg aatcgaccg gattttaaag ctgctgccag gtcggtgaat ccccctaggt 4650

gatctccttg tacaagatg ttgggcacgg acttttcgac ccggatgaga acgtcgtgaa 4710

60 gagtttgaaa aagattatca acataatgtg tcttttttct ttttttctt cgtaactctc 4770

tagagaacga ggagacgtac ggtctgattt gttatcg 4807

65 <210> 2  
 <211> 1536

ES 2 305 501 T3

<212> ADN

<213> *Phaffia rhodozyma*

5 <220>

<221> CDS

<222> (1)..(1536)

10 <400> 2

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

ES 2 305 501 T3

	tcg ttc gga ctc ctt ctc caa aag aca aac atc ctt cga gat att cgg Ser Phe Gly Leu Leu Leu Gln Lys Thr Asn Ile Leu Arg Asp Ile Arg 225 230 235 240	720
5	gag gac gcc gac gaa ggt cgt ggc ttc tgg cca aga gag atc tgg gcc Glu Asp Ala Asp Glu Gly Arg Gly Phe Trp Pro Arg Glu Ile Trp Ala 245 250 255	768
10	aac ccg atc tat act gcg cat gca ccg ggc aca agg ttt aac tcg ttg Asn Pro Ile Tyr Thr Ala His Ala Pro Gly Thr Arg Phe Asn Ser Leu 260 265 270	816
15	act gac ctg gtc aag aaa gaa aac atc gac aaa gga tca atg tgg gtg Thr Asp Leu Val Lys Lys Glu Asn Ile Asp Lys Gly Ser Met Trp Val 275 280 285	864
	ttg agt gcg atg aca ctc gac gcg atc acc cat act acc gac gca ctg Leu Ser Ala Met Thr Leu Asp Ala Ile Thr His Thr Thr Asp Ala Leu 290 295 300	912
20	gac tac ctc tca ctt cta aag aac cag agt gtt ttc aac ttt tgt gct Asp Tyr Leu Ser Leu Leu Lys Asn Gln Ser Val Phe Asn Phe Cys Ala 305 310 315 320	960
25	atc ccg gct gtc atg tcg att gca acg ttg gag cta tgc ttc atg aac Ile Pro Ala Val Met Ser Ile Ala Thr Leu Glu Leu Cys Phe Met Asn 325 330 335	1008
	cca gcg gtg ttc caa cga aac ata aaa atc aga aag gga gaa gcc gtc Pro Ala Val Phe Gln Arg Asn Ile Lys Ile Arg Lys Gly Glu Ala Val 340 345 350	1056
30	gag ctc att atg aag tgc aac aac cct cgg gag gtg gca tac atg ttt Glu Leu Ile Met Lys Cys Asn Asn Pro Arg Glu Val Ala Tyr Met Phe 355 360 365	1104
35	aga gat tat gct cga aag att cat gcc aag gct att cct aca gat cct Arg Asp Tyr Ala Arg Lys Ile His Ala Lys Ala Ile Pro Thr Asp Pro 370 375 380	1152
40	aac ttc atc aag ttg agc gtt gcg tgt ggt cga atc gaa caa tgg gct Asn Phe Ile Lys Leu Ser Val Ala Cys Gly Arg Ile Glu Gln Trp Ala 385 390 395 400	1200
	gag cac tac tac ccc tca ttt atg atg att cgg cct tcg aat gac cct Glu His Tyr Tyr Pro Ser Phe Met Met Ile Arg Pro Ser Asn Asp Pro 405 410 415	1248
45	caa aac ccc gca ccc tca acg gcg ctt gac cct ttc tca gga gac gct Gln Asn Pro Ala Pro Ser Thr Ala Leu Asp Pro Phe Ser Gly Asp Ala 420 425 430	1296
50	cgt tta agg ata gcc tct aag aag gct gag atc acc gcc gct gct ctt Arg Leu Arg Ile Ala Ser Lys Lys Ala Glu Ile Thr Ala Ala Leu 435 440 445	1344
	gtc agg aag aaa gcc ccg gat cac gct aag tgg aga gag tcc aag gga Val Arg Lys Lys Ala Arg Asp His Ala Lys Trp Arg Glu Ser Lys Gly 450 455 460	1392
55	ttg cct ccg agc gat ccg aac aag ccg gac aac tcg gag gat gtt aat Leu Pro Pro Ser Asp Pro Asn Lys Pro Asp Asn Ser Glu Asp Val Asn 465 470 475 480	1440
60	tgg gta ttg atc ggc ggt atg atc gtt gga ttg ttg ctc gtg atg ggc Trp Val Leu Ile Gly Gly Met Ile Val Gly Leu Leu Leu Val Met Gly 485 490 495	1488
65	gtg ctc ggt ttg gct atc gct tgg gtt gtt cct cag ttt gag caa taa Val Leu Gly Leu Ala Ile Ala Trp Val Val Leu Gln Phe Glu Gln 500 505 510	1536

ES 2 305 501 T3

<210> 3

<211> 511

<212> PRT

5 <213> *Phaffia rhodozyma*

<400> 3

10 Met Gly Ile Ser Asp Tyr Leu Val Leu Ala Phe Thr His Pro Ala Asp  
1 5 10 15

15 Leu Arg Ala Leu Met Gln Tyr Ala Ile Trp His Glu Pro Arg Arg Asn  
20 25 30

20 Ile Thr Ala Gln Glu Glu His Ala Thr Ser Gly Trp Asp Arg Glu Thr  
35 40 45

25 Met Lys Glu Cys Trp Lys Tyr Leu Asp Leu Thr Ser Arg Ser Phe Ala  
50 55 60

30 Ala Val Ile Lys Glu Leu Asp Gly Asp Leu Thr Arg Val Ile Cys Leu  
65 70 75 80

35 Phe Tyr Leu Ala Leu Arg Gly Leu Asp Thr Ile Glu Asp Asp Met Ser  
85 90 95

40 Leu Ser Asn Asp Val Lys Leu Pro Leu Leu Arg Thr Phe Trp Glu Lys  
100 105 110

45 Leu Asp Ser Pro Gly Trp Thr Phe Thr Gly Ser Gly Pro Asn Glu Lys  
115 120 125

50 Asp Arg Glu Leu Leu Val His Phe Asp Val Ala Ile Ala Glu Phe Ala  
130 135 140

55 Asn Leu Asp Val Asn Ser Arg Asn Val Ile Arg Asp Ile Thr Arg Lys  
145 150 155 160

60 Met Gly Asn Gly Met Ala Asp Phe Ala Ser Leu Ser Thr Pro Ser Lys  
165 170 175

65 Pro Val Ala Glu Val Gln Ser Thr Glu Asp Phe Asn Leu Tyr Cys His  
180 185 190

Tyr Val Ala Gly Leu Val Gly Glu Gly Leu Ser Arg Leu Phe Val Ala  
195 200 205

Thr Glu Lys Glu Arg Pro Phe Leu Ala Asn Gln Met Val Leu Ser Asn  
210 215 220

ES 2 305 501 T3

Ser Phe Gly Leu Leu Leu Gln Lys Thr Asn Ile Leu Arg Asp Ile Arg  
 225 230 235 240  
 5  
 Glu Asp Ala Asp Glu Gly Arg Gly Phe Trp Pro Arg Glu Ile Trp Ala  
 245 250 255  
 10  
 Asn Pro Ile Tyr Thr Ala His Ala Pro Gly Thr Arg Phe Asn Ser Leu  
 260 265 270  
 15  
 Thr Asp Leu Val Lys Lys Glu Asn Ile Asp Lys Gly Ser Met Trp Val  
 275 280 285  
 20  
 Leu Ser Ala Met Thr Leu Asp Ala Ile Thr His Thr Thr Asp Ala Leu  
 290 295 300  
 25  
 Asp Tyr Leu Ser Leu Leu Lys Asn Gln Ser Val Phe Asn Phe Cys Ala  
 305 310 315 320  
 30  
 Ile Pro Ala Val Met Ser Ile Ala Thr Leu Glu Leu Cys Phe Met Asn  
 325 330 335  
 35  
 Pro Ala Val Phe Gln Arg Asn Ile Lys Ile Arg Lys Gly Glu Ala Val  
 340 345 350  
 40  
 Glu Leu Ile Met Lys Cys Asn Asn Pro Arg Glu Val Ala Tyr Met Phe  
 355 360 365  
 45  
 Arg Asp Tyr Ala Arg Lys Ile His Ala Lys Ala Ile Pro Thr Asp Pro  
 370 375 380  
 50  
 Asn Phe Ile Lys Leu Ser Val Ala Cys Gly Arg Ile Glu Gln Trp Ala  
 385 390 395 400  
 55  
 Glu His Tyr Tyr Pro Ser Phe Met Met Ile Arg Pro Ser Asn Asp Pro  
 405 410 415  
 60  
 Gln Asn Pro Ala Pro Ser Thr Ala Leu Asp Pro Phe Ser Gly Asp Ala  
 420 425 430  
 65  
 Arg Leu Arg Ile Ala Ser Lys Lys Ala Glu Ile Thr Ala Ala Ala Leu  
 435 440 445  
 Val Arg Lys Lys Ala Arg Asp His Ala Lys Trp Arg Glu Ser Lys Gly  
 450 455 460  
 Leu Pro Pro Ser Asp Pro Asn Lys Pro Asp Asn Ser Glu Asp Val Asn  
 465 470 475 480  
 Trp Val Leu Ile Gly Gly Met Ile Val Gly Leu Leu Leu Val Met Gly  
 485 490 495  
 Val Leu Gly Leu Ala Ile Ala Trp Val Val Leu Gln Phe Glu Gln  
 500 505 510

## ES 2 305 501 T3

<210> 4  
<211> 27  
<212> ADN  
5 <213> Artificial

<220>  
<223> Cebador  
10

<220>  
<221> características diversas  
<222> (3)..(3)  
15 <223> n = a, c, g o t

<220>  
<221> características diversas  
20 <222> (4)..(4)  
<223> y = c o t

<220>  
25 <221> características diversas  
<222> (6)..(6)  
<223> n = a, c, g o t

30 <220>  
<221> características diversas  
<222> (12)..(12)  
35 <223> n = a, c, g o t

<220>  
40 <221> características diversas  
<222> (15)..(15)  
<223> n = a, c, g o t

45 <400> 4  
gcnytngaya cngtngarga ygayatg

<210> 5  
50 <211> 26  
<212> ADN  
<213> Artificial

55 <220>  
<223> Cebador

<220>  
60 <221> características diversas  
<222> (3)..(3)  
<223> n = a, c, g o t

65 <220>  
<221> características diversas

27

## ES 2 305 501 T3

<222> (9)..(9)

<223> n = a, c, g o t

5 <220>

<221> características diversas

<222> (15)..(15)

<223> n = a, c, g o t

10

<220>

<221> características diversas

<222> (18)..(18)

15

<223> n = a, c, g o t

<220>

<221> características diversas

20

<222> (21)..(21)

<223> n = a, c, g o t

25

<400> 5

atngccatna cytgnggnat ngrca

26

30

<210> 6

<211> 29

<212> ADN

<213> Artificial

35

<220>

<223> Cebador

40

<220>

<221> características diversas

<222> (3)..(3)

<223> n = a, c, g o t

45

<220>

<221> características diversas

<222> (6)..(6)

50

<223> n = a, c, g o t

<220>

55

<221> características diversas

<222> (9)..(9)

<223> n = a, c, g o t

60

<220>

<221> características diversas

<222> (12)..(12)

<223> n = a, c, g o t

65

<220>

<221> características diversas

## ES 2 305 501 T3

	<222> (15)..(15)	
	<223> n =a, c, g o t	
5	<400> 6	
	ccnacngtnc engcnacrta rtgrcarta	29
10	<210> 7	
	<211> 20	
	<212> ADN	
	<213> Artificial	
15	<220>	
	<223> Cebador	
20	<400> 7	
	aatgatgtgta agcttcccct	20
25	<210> 8	
	<211> 20	
	<212> ADN	
	<213> Artificial	
30	<220>	
	<223> Cebador	
35	<400> 8	
	ccagatctct cttggccaga	20
40		
45		
50		
55		
60		
65		