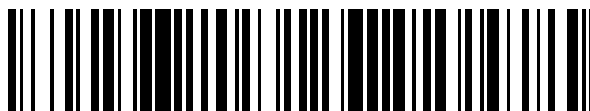


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 567 305**

51 Int. Cl.:

<b>C12N 1/20</b>	(2006.01)	<b>C12N 15/52</b>	(2006.01)
<b>C12N 5/10</b>	(2006.01)	<b>C12P 7/62</b>	(2006.01)
<b>C12N 15/00</b>	(2006.01)	<b>C12P 7/64</b>	(2006.01)
<b>C12N 15/09</b>	(2006.01)	<b>C12N 15/82</b>	(2006.01)
<b>C12N 15/63</b>	(2006.01)	<b>C12N 9/00</b>	(2006.01)
<b>C12N 15/70</b>	(2006.01)		
<b>C12N 15/74</b>	(2006.01)		
<b>C12P 19/62</b>	(2006.01)		
<b>C07H 21/02</b>	(2006.01)		
<b>C07H 21/04</b>	(2006.01)		

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.04.2002 E 10178950 (1)**

97 Fecha y número de publicación de la concesión europea: **13.01.2016 EP 2366772**

54 Título: **Sistemas de PUFA policétido sintasa y usos de los mismos**

30 Prioridad:

**16.04.2001 US 284066 P**  
**15.06.2001 US 298796 P**  
**18.09.2001 US 323269 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**21.04.2016**

73 Titular/es:

**DSM IP ASSETS B.V. (100.0%)**  
**Het Overloon 1**  
**6411 TE Heerlen, NL**

72 Inventor/es:

**METZ, JAMES G.;**  
**BARCLAY, WILLIAM R.;**  
**FLATT, JAMES H. y**  
**KUNER, JERRY M.**

74 Agente/Representante:

**ISERN JARA, Jorge**

ES 2 567 305 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

Sistemas de PUFA policétido sintasa y usos de los mismos

5 Campo de la invención

Esta invención se refiere a sistemas de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) para microorganismos, incluyendo organismos eucariotas, tales como microorganismos traustoquitridios. Más particularmente, esta invención se refiere a ácidos nucleicos que codifican sistemas PUFA PKS no bacterianos, a sistemas PUFA PKS no bacterianos, a organismos genéticamente modificados que comprenden sistemas PUKA PKS no bacterianos, y a métodos para preparar y usar los sistemas PUFA PKS no bacterianos descritos en este documento.

15 Antecedentes de la invención

Los sistemas de policétido sintasa (PKS) son generalmente conocidos en la técnica como complejos enzimáticos derivados de sistemas de ácido graso sintasa (FAS), pero que a menudo están muy modificados para producir productos especializados que típicamente muestran poco parecido a ácidos grasos. Los investigadores han intentado explotar los sistemas de policétidos sintasa (PKS) que se han descrito en la bibliografía como dentro de uno de tres tipos básicos, típicamente mencionados como: Tipo II, Tipo I y modular. El sistema Tipo II se caracteriza por proteínas separables, cada una de las cuales realiza una reacción enzimática distinta. Las enzimas trabajan en concierto para producir el producto final y cada enzima individual del sistema típicamente participa varias veces en la producción del producto final. Este tipo de sistema funciona de un modo análogo a los sistemas de ácido graso sintasa (FAS) encontrados en plantas y bacterias. Los sistemas PKS Tipo I son similares al sistema Tipo II porque las enzimas se usan de un modo iterativo para producir el producto final. El Tipo I difiere del Tipo II porque las actividades enzimáticas, en lugar de estar asociadas con proteínas separables, existen como dominios de proteínas más grandes. Este sistema es análogo a los sistemas FAS Tipo I encontrados en animales y hongos.

En contraste con los sistemas Tipo I y II, en los sistemas PKS modulares, cada dominio enzimático se usa solamente una vez en la producción del producto final. Los dominios se encuentran en proteínas muy grandes y el producto de cada reacción se pasa a otro dominio en la proteína PKS. Adicionalmente, en todos los sistemas PKS descritos anteriormente, si se incorpora un doble enlace carbono-carbono en el producto final, siempre está en configuración *trans*.

En los sistemas PKS Tipo I y Tipo II descritos anteriormente, se realiza el mismo conjunto de reacciones en cada ciclo hasta que se obtiene el producto final. No se permite la introducción de reacciones únicas durante el procedimiento biosintético. Los sistemas PKS modulares requieren proteínas grandes que no utilizan la economía de las reacciones iterativas (es decir, se requiere un dominio distinto para cada reacción). Adicionalmente, como se ha indicado anteriormente, se introducen en dobles enlaces carbono-carbono en la configuración *trans* en todos los sistemas PKS descritos previamente.

Los ácidos grasos poliinsaturados (PUFA) son componentes críticos de los lípidos de membrana en la mayoría de los eucariotas (Lauritzen *et al.*, Prog. Lipid Res. 40 1 (2001); McConn *et al.*, Plant J. 15, 521 (1998)) y son precursores de ciertas hormonas y moléculas de señalización (Heller *et al.*, Drugs 55, 487 (1998); Creelman *et al.*, Annu. Rev. Plant Physiol. Plant Mol. Biol. 48, 355 (1997)). Las rutas conocidas de la síntesis de PUFA implican el procesamiento de ácidos grasos saturados 16:0 o 18:0 (la abreviatura X:Y indica un grupo acilo que contiene X átomos de carbono e Y dobles enlaces *cis*; las posiciones de doble enlace de los PUFA se indican respecto al carbono de metilo de la cadena de ácido graso ( $\omega 3$  u  $\omega 6$ ) con interrupción de metileno sistemática de los dobles enlaces) derivados de ácido graso sintasa (FAS) por reacciones de elongación y desaturación aeróbica (Sprecher, Curr. Opin. Clin. Nutr. Metab. Care 2,135 (1999); Parker-Barnes *et al.*, Proc. Natl. Acad. Sci. USA 97, 8284 (2000); Shanklin *et al.*, Annu. Rev. Plant Physiol. Plant Mol. Biol. 49, 611 (1998)). Partiendo de acetil-CoA, la síntesis de DHA requiere aproximadamente 30 actividades enzimáticas distintas y casi 70 reacciones incluyendo las cuatro etapas repetitivas del ciclo de síntesis de ácido graso. Las policétido sintasas (PKS) realizan algunas de las mismas reacciones que FAS (Hopwood *et al.*, Annu. Rev. Genet. 24, 37 (1990); Bentley *et al.*, Annu. Rev. Microbiol. 53, 411 (1999)) y usa la misma proteína pequeña (o dominio), proteína transportadora de acilo (ACP), como sitio de unión covalente para el crecimiento de la cadena de carbono. Sin embargo, en estos sistemas enzimáticos, el ciclo completo de reducción, deshidratación y reducción observado en FAS a menudo se abrevia de modo que se produce una cadena de carbono altamente derivatizada, que típicamente contiene muchos grupos ceto e hidroxilo así como dobles enlaces carbono-carbono en la configuración *trans*. Los productos lineales de PKS a menudo se ciclan para formar agentes bioquímicos complejos que incluyen antibióticos y otros muchos productos secundarios (Hopwood *et al.*, (1990) *supra*; Bentley *et al.*, (1999), *supra*; Keating *et al.*, Curr. Opin. Chem. Biol. 3, 598 (1999)).

Se han presentado PUFA de cadena muy larga tales como ácido docosahexaenoico (DHA; 22:6 $\omega$ 3) y ácido eicosapentaenoico (EPA; 20:5 $\omega$ 3) de varias especies de bacterias marinas, incluyendo *Shewanella* sp (Nichols *et al.*, Curr. Op. Biotechnol. 10, 240 (1999); Yazawa, Lipids 31, S (1996); DeLong *et al.*, Appl. Environ. Microbiol. 51, 730 (1986)). El análisis de un fragmento genómico (clonado como plásmido pEPA) de *Shewanella* sp cepa

SCRC2738 condujo a la identificación de cinco fases de lectura abierta (Orf), que hace un total de 20 Kb, que son necesarias y suficientes para la producción de EPA en *E. coli* (Yazawa, (1996), *supra*). Varios de los dominios proteicos predichos eran homólogos de enzimas FAS, mientras que otras regiones no mostraron homología con proteínas de función conocida. En base a estas observaciones y estudios bioquímicos, se sugirió que la síntesis de PUFA en *Shewanella* implicaba la elongación de ácidos grasos de 16 o 18 carbonos producidos por FAS y la inserción de dobles enlaces por desaturasa aeróbicas indefinidas (Watanabe *et al.*, J. Biochem. 122, 467 (1997)). El reconocimiento de que esta hipótesis era incorrecta comenzó con un reexamen de las secuencias proteicas codificadas por las cinco Orf de *Shewanella*. Fueron identificables al menos 11 regiones dentro de las cinco Orf como dominios enzimáticos putativos (véase, Metz *et al.*, Science 293:290-293 (2001)). Cuando se comparan con secuencias en las bases de datos de genes, siete de estas estaban más fuertemente relacionadas con proteínas PKS que a proteínas FAS. Incluidos en este grupo estaban los dominios que putativamente codifican malonil-CoA:ACP aciltransferasa (MAT), 3-cetoacil-ACP sintasa (KS), 3-cetoacil-ACP reductasa (KR), aciltransferasa (AT), fosfopanteteína transferasa, factor de longitud de cadena (o inicio de cadena) (CLF) y un grupo muy inusual de seis dominios ACP (es decir, la presencia de más de dos dominios ACP agrupados no se ha presentado previamente en secuencias PKS o FAS). Sin embargo, tres regiones fueron mucho más homólogas a proteínas FAS bacterianas. Una de estas fue similar a la enoil reductasa (ER) resistente a Triclosán recién descrita de *Streptococcus pneumoniae* (Heath *et al.*, Nature 406, 145 (2000)); una comparación del péptido ORF8 con la enol reductasa de *S. pneumoniae* usando el programa LALIGN (matriz, BLOSUM50; penalización por abertura de hueca -10; penalización por elongación -1) indicó una similitud del 49 % sobre un solapamiento de 386aa). Dos regiones fueron homólogas de la proteínas FAS de *E. coli* codificada por *fabA*, que cataliza la síntesis de *trans*-2-decenoil-ACP y la isomerización reversible de este producto en *cis*-3-decenoil-ACP (Heath *et al.*, J. Biol. Chem., 271,27795 (1996)). En esta base, pareció probable que al menos algunos de los dobles enlaces en EPA de *Shewanella* se introducían por mecanismo de deshidrasa-isomerasa catalizado por dominios tipo FabA en Orf7.

Células de *E. coli* cultivadas de forma anaeróbica que albergaban el plásmido pEPA acumularon EPA a los mismos niveles que cultivos aeróbicos (Metz *et al.*, 2001, *supra*), lo que indica que no está implicada una desaturasa dependiente de oxígeno en la síntesis de EPA. Cuando se introdujo pEPA en un mutante *fabB'* de *E. coli*, que es incapaz de sintetizar ácidos grasos monosaturados y requiere ácidos grasos insaturados para el crecimiento, las células resultantes perdieron su auxotrofismo de ácidos grasos. También se acumularon niveles mucho mayores de EPA que otras cepas que contienen pEPA, lo que sugiere que EPA compite con ácidos grasos monoinsaturados producidos de forma endógena para la transferencia a glicerolípidos. Cuando se cultivaron células de *E. coli* que contenían pEPA en presencia de [<sup>13</sup>C]-acetato, los datos del análisis de <sup>13</sup>C-NMR de EPA purificado a partir de las células confirmó la identidad de EPA y proporcionó evidencias de que este ácido graso se sintetizaba a partir de acetil-CoA y malonil-CoA (véase Metz *et al.*, 2001, *supra*). Un homogeneizado libre de células de células *fabB'* que contenían pEPA sintetizaba tanto EPA como ácidos grasos saturados a partir de [<sup>14</sup>C]-malonil-CoA. Cuando se separó el homogeneizado en un sedimento de alta velocidad de 200.000 xg y una fracción sobrenadante sin membrana, la síntesis de ácidos grasos saturados se confinaba al sobrenadante, lo que es coherente con la naturaleza soluble de las enzimas FAS Tipo II (Magnuson *et al.*, Microbiol. Rev. 57, 522 (1993)). Se encontró síntesis de EPA solamente en la fracción sedimentada a alta velocidad, lo que indica que la síntesis de EPA puede suceder sin dependencia de las enzimas de FAS de *E. coli* o de intermedios solubles (tales como 16:0-ACP) de la fracción citoplasmática. Como las proteínas codificadas por los genes de EPA de *Shewanella* no son particularmente hidrófobas, la restricción de la actividad de síntesis de EPA a esta fracción puede reflejar una necesidad de molécula aceptora asociada a membrana. Adicionalmente, en contraste con FAS de *E. coli*, la síntesis de EPA es específicamente dependiente de NADPH y no requiere NADH. Todos estos resultados son coherentes con los genes de pEPA que codifican una PKS multifuncional que actúa independientemente de FAS, elongasa, y actividades desaturasa para sintetizar EPA directamente. Es probable que la ruta de PKS para la síntesis de PUFA que se ha identificado en *Shewanella* esté propagada en bacterias marinas. Genes con alta homología al grupo génico de *Shewanella* se han identificado en *Photobacterium profundum* (Allen *et al.*, Appli. Environ. Microbiol. 65:1710 (1999)) y en *Moritella marina* (*Vibrio marinus*) (Tanaka *et al.*, Biotechnol. Lett. 21:939 (1999)).

Los análisis bioquímicos y genéticos moleculares realizados con *Shewanella* proporcionan evidencias convincentes para policétidos sintasas que son capaces de sintetizar PUFA a partir de malonil-CoA. Se ha propuesto un esquema completo para la síntesis de EPA por la PKS de *Shewanella*. La identificación de dominios proteicos homólogos a la proteína FabA de *E. coli*, y la observación de que la síntesis de EPA bacteriana sucede de forma anaeróbica, proporciona evidencias para un mecanismo donde la inserción de dobles enlaces *cis* sucede a través de la acción de una deshidratasa/2-*trans*, 3-*cis* isomerasa bifuncional (DH/2,3I). En *E. coli*, la condensación del intermedio 3-*cis* acilo con malonil-ACP requiere una cetoacil-ACP sintasa particular y esto puede proporcionar un fundamento la presencia de dos KS en el grupo génico de *Shewanella* (en Orf 5 y Orf 7). Sin embargo, el ciclo de PKS prolonga la cadena en incrementos de dos carbonos mientras que los dobles enlaces en el producto EPA suceden cada tres carbonos. Esta disyuntiva puede resolverse si los dobles enlaces en C-14 y C-8 de EPA se generan por isomerización 2-*trans*, 2-*cis* (DH/2,2I) seguida por incorporación del doble enlace *cis* en la cadena de ácido graso en elongación. Se sabe que sucede la conversión enzimática de un doble enlace *trans* en la configuración *cis* sin migración del enlace, por ejemplo en la síntesis de 11-*cis*-retinal en el ciclo retinoide (Jang *et al.*, J. Biol. Chem. 275, 28128 (2000)). Aunque dicha función enzimática no se ha identificado aún en la PKS de *Shewanella*, puede residir en uno de los dominios proteicos no asignados.

Las rutas de PKS para la síntesis de PUFA en *Shewanella* y otras bacterias marinas, *Vibrio marinus*, se describen en detalle en la patente de Estados Unidos n.º 6.140.486 (expedida a partir de la solicitud de Estados Unidos n.º de serie 09/090.793, presentada el 4 de junio, 1998, titulada "Production of Polyunsaturated Fatty Acids by Expression of Polyketide-like Synthesis Genes in Plants", que se incorpora por referencia en este documento en su totalidad).

Los ácidos grasos poliinsaturados (PUFA) se consideran útiles para propósitos nutricionales, farmacéuticos, industriales y otros propósitos. Un suministro expansivo de PUFA a partir de fuentes naturales y de síntesis química no es suficiente para las necesidades comerciales. Como se requieren varias enzimas desaturasa y elongasa diferentes para la síntesis de ácidos grasos a partir de ácido linoleico (LA, 18:2  $\Delta$  9, 12), común en la mayoría de las especies vegetales, en los PUFA de cadena más saturada y más larga, la modificación por ingeniería de células hospedadoras vegetales para la expresión de PUFA tales como EPA y DHA puede requerir la expresión de cinco o seis actividades enzimáticas diferentes para conseguir la expresión, al menos para EPA y DHA. Adicionalmente, para la producción de cantidades utilizables de dichos PUFA, pueden requerirse esfuerzos adicionales de ingeniería, por ejemplo, la regulación negativa de enzimas que compiten por el sustrato, la modificación por ingeniería de actividades enzimáticas superiores tales como por mutagénesis o direccionamiento de enzimas a orgánulos plastidios. Por lo tanto, es de interés obtener material genético implicado en la biosíntesis de PUFA a partir de especies que producen de forma natural estos ácidos grasos y expresan el material aislado solo o en combinación en un sistema heterólogo que puede manipularse para permitir la producción de cantidades comerciales de PUFA.

El descubrimiento de un sistema PUFA PKS en bacterias marinas tales como *Shewanella* y *Vibrio marinus* (véase la patente de Estados Unidos n.º 6.140.486, *ibid.*) proporciona un recurso para nuevos métodos de producción comercial de PUFA. Sin embargo, estas bacterias marinas tienen limitaciones que finalmente restringirán su utilidad a un nivel comercial. En primer lugar, aunque la patente de Estados Unidos n.º 6.140.486 describe que los sistemas PUFA PKS de bacterias marinas pueden usarse para modificar genéticamente plantas, las bacterias marinas viven y crecen de forma natural en entornos marinos fríos y los sistemas enzimáticos de estas bacterias no funcionan bien por encima de 30 °C. En contraste, muchas plantas de cultivo, que son dianas atractivas para manipulación genética usando el sistema PUFA PKS, tienen condiciones de crecimiento normal a temperaturas por encima de 30 °C que varían hasta más de 40 °C. Por lo tanto, se predice que el sistema PUFA PKS de bacterias marinas no será fácilmente adaptable a expresión en plantas en condiciones normales de crecimiento. Además, los genes de PUFA PKS de bacterias marinas, que son de una fuente bacteriana, pueden no ser compatibles con los genomas de células hospedadoras eucariotas, o al menos pueden requerir adaptación significativa para funcionar en hospedadores eucariotas. Adicionalmente, los sistemas PUFA PKS de bacterias marinas conocidos no producen directamente triglicéridos, mientras que sería deseable la producción directa de triglicéridos porque los triglicéridos son un producto de almacenamiento de lípidos en microorganismos y como resultado pueden acumularse a niveles muy altos (por ejemplo, hasta el 80-85 % del peso celular) en células microbianas/vegetales (en oposición a un producto lipídico "estructural" (por ejemplo, fosfolípidos) que generalmente pueden acumularse solamente a bajos niveles (por ejemplo, menos del 10-15 % del peso celular como máximo)).

Por lo tanto, existe la necesidad en la técnica de otros sistemas PUFA PKS que tengan mayor flexibilidad para uso comercial.

#### Sumario de la invención

Una realización de la presente invención se refiere a una molécula aislada de ácido nucleico según la reivindicación 1.

En aspectos alternativos, la secuencia de ácido nucleico codifica una secuencia de aminoácidos que es al menos un 70 % idéntica a una secuencia de aminoácidos SEC ID N° 20. En una realización preferida, la secuencia de ácido nucleico codifica una secuencia de aminoácidos elegida entre: SEC ID N° 20, y/o fragmentos biológicamente activos de las mismas. En un aspecto, la secuencia de ácido nucleico se elige entre: SEC ID N° 19.

Otra realización de la presente invención se refiere a una molécula de ácido nucleico recombinante que comprende la molécula de ácido nucleico descrita anteriormente, unida de forma funcional a al menos una secuencia de control de la transcripción. En otra realización, la presente invención se refiere a una célula recombinante transfectada con la molécula de ácido nucleico recombinante descrita justo anteriormente.

Otra realización más de la presente invención se refiere a un microorganismo genéticamente modificado, donde el microorganismo expresa un sistema PKS que comprende al menos un dominio biológicamente activo de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS). El al menos un dominio del sistema PUFA PKS está codificado por una secuencia de ácido nucleico elegida entre las secuencias de ácido nucleico de la invención.

En esta realización, el microorganismo se modifica genéticamente para que afecte a la actividad del sistema PKS.

En un aspecto, el microorganismo genéticamente modificado es un traustochytridio, que puede incluir, aunque sin limitación, un traustochytridio de un género elegido entre *Schizochytrium* y *Thraustochytrium*. En otro aspecto, el microorganismo se ha modificado genéticamente de forma adicional para que exprese de forma recombinante al

menos una molécula de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS bacteriano, de un sistema PKS Tipo I, de un sistema PKS Tipo II, y/o de un sistema PKS modular.

5 En otro aspecto de esta realización, el microorganismo expresa de forma endógena un sistema PUFA PKS que comprende el al menos un dominio biológicamente activo de un sistema PUFA PKS, y donde la modificación genética comprende la expresión de una molécula de ácido nucleico recombinante seleccionada entre el grupo que consiste en una molécula de ácido nucleico recombinante que codifica al menos un dominio biológicamente activo de un segundo sistema PKS y una molécula de ácido nucleico recombinante que codifica una proteína que afecta a la actividad del sistema PUFA PKS. Preferiblemente, la molécula de ácido nucleico recombinante comprende una  
10 cualquiera de las secuencias de ácido nucleico descritas anteriormente.

En un aspecto de esta realización, la molécula de ácido nucleico recombinante codifica una fosfopanteteína transferasa. En otro aspecto, la molécula de ácido nucleico recombinante comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS bacteriano, de un sistema PKS Tipo I, de un sistema PKS Tipo II, y/o de un sistema PKS modular.  
15

En otro aspecto de esta realización, el microorganismo se modifica genéticamente por transfección con una molécula de ácido nucleico recombinante que codifica el al menos un dominio de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS). Dicha molécula de ácido nucleico recombinante puede incluir cualquier molécula de ácido nucleico recombinante que comprenda cualquiera de las secuencias de ácido nucleico descritas anteriormente. En un aspecto, el microorganismo se ha modificado genéticamente de forma adicional para que exprese de forma recombinante al menos una molécula de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS bacteriano, de un sistema PKS Tipo I, de un sistema PKS Tipo II, o de un sistema PKS modular.  
20

Otra realización más de la presente invención se refiere a una planta genéticamente modificada, donde la planta se ha modificado genéticamente para que exprese de forma recombinante un sistema PKS que comprende al menos un dominio biológicamente activo de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS). El dominio puede estar codificado por cualquiera de las secuencias de ácido nucleico descritas anteriormente. En un aspecto, la planta se ha modificado genéticamente de forma adicional para que exprese de forma recombinante al menos una molécula de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS bacteriano, de un sistema PKS Tipo I, de un sistema PKS Tipo II, y de un sistema PKS modular.  
25

Otra realización más de la presente invención se refiere a un método para producir una molécula bioactiva que se produce por un sistema de policétido sintasa. El método incluye la etapa de cultivar en condiciones eficaces para producir la molécula bioactiva un organismo modificado genéticamente que expresa un sistema PKS que comprende al menos un dominio biológicamente activo de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS).  
30

El dominio del sistema PUFA PKS está codificado por cualquiera de las secuencias de ácido nucleico descrita anteriormente.  
35

En un aspecto de esta realización, el organismo expresa de forma endógena un sistema PKS que comprende el al menos un dominio del sistema PUFA PKS, y la modificación genética está en una secuencia de ácido nucleico que codifica el al menos un dominio de sistema PUFA PKS. Por ejemplo, la modificación genética puede cambiar al menos un producto producido por el sistema PKS endógeno, en comparación con un organismo de tipo silvestre.  
40

En otro aspecto de esta realización, el organismo expresa de forma endógena un sistema PKS que comprende el al menos un dominio biológicamente activo del sistema PUFA PKS, y la modificación genética comprende transfección del organismo con una molécula de ácido nucleico recombinante seleccionada entre el grupo que consiste en: una molécula de ácido nucleico recombinante que codifica al menos un dominio biológicamente activo de un segundo sistema PKS y una molécula de ácido nucleico recombinante que codifica una proteína que afecta a la actividad del sistema PUFA PKS. Por ejemplo, la modificación genética puede cambiar al menos un producto producido por el sistema PKS endógeno, en comparación con un organismo de tipo silvestre.  
45

En otro aspecto más de esta realización, el organismo se modifica genéticamente por transfección con una molécula de ácido nucleico recombinante que codifica el al menos un dominio del sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS). En otro aspecto, el organismo produce un perfil de ácido graso poliinsaturado (PUFA) que difiere del organismo de origen natural sin modificación genética. En otro aspecto, el organismo expresa de forma endógena un sistema PUFA PKS no bacteriano, y donde la modificación genética comprende sustitución de un dominio de un sistema PKS diferente en el lugar de una secuencia de ácido nucleico que codifica al menos un dominio del sistema PUFA PKS no bacteriano.  
50

En otro aspecto más, el organismo expresa de forma endógena un sistema PUFA PKS no bacteriano que se ha modificado transfectando el organismo con una molécula de ácido nucleico recombinante que codifica una proteína que regula la longitud de cadena de ácidos grasos producidos por el sistema PUFA PKS. Por ejemplo, la molécula de ácido nucleico recombinante que codifica una proteína que regula la longitud de cadena de ácidos grasos puede  
55

5 remplazar una secuencia de ácido nucleico que codifica un factor de longitud de cadena en el sistema PUFA PKS no bacteriano. En otro aspecto, la proteína que regula la longitud de cadena de los ácidos grasos producidos por el sistema PUFA PKS es un factor de longitud de cadena. En otro aspecto, la proteína que regula la longitud de cadena de los ácidos grasos producidos por el sistema PUFA PKS es un factor de longitud de cadena que dirige la síntesis de unidades C20.

10 En un aspecto, el organismo expresa un sistema PUFA PKS no bacteriano que comprende una modificación genética en un dominio elegido entre: un dominio que codifica dominio  $\beta$ -hidroxi acil ACP deshidrasa (DH) tipo FabA y un dominio que codifica  $\beta$ -ceto acil ACP sintasa (KS), donde la modificación altera la relación de ácidos grasos de cadena larga producidos por el sistema PUFA PKS en comparación con la ausencia de la modificación. En un aspecto, la modificación comprende sustituir un dominio DH que no posee actividad de isomerización por una  $\beta$ -hidroxi acil ACP deshidrasa (DH) tipo FabA en el sistema PUFA PKS no bacteriano. En otro aspecto, la modificación se selecciona entre el grupo que consiste en una delección de todo o parte del dominio, una sustitución de un dominio homólogo de un organismo diferente para el dominio, y una mutación del dominio.

15 En otro aspecto, el organismo expresa un sistema PKS y la modificación genética comprende sustituir un dominio de  $\beta$ -hidroxi acil ACP deshidrasa (DH) tipo FabA de un sistema PUFA PKS para un dominio DH que no posee actividad de isomerización.

20 En otro aspecto, el organismo expresa un sistema PUFA PKS no bacteriano que comprende una modificación en un dominio de enoil ACP reductasa (ER), donde la modificación provoca la producción de un compuesto diferente en comparación con la ausencia de la modificación. Por ejemplo, la modificación puede seleccionarse entre el grupo que consiste en una delección de todo o parte del dominio ER, una sustitución de un dominio ER de un organismo diferente por el dominio ER, y una mutación del dominio ER.

25 En un aspecto, la molécula bioactiva producida por el presente método puede incluir, aunque sin limitación, una formulación antiinflamatoria, un agente quimioterapéutico, un excipiente activo, un fármaco para osteoporosis, un antidepresivo, un anticonvulsivo, un fármaco anti-*Helicobacter pylori*, un fármaco para el tratamiento de enfermedad neurodegenerativa, un fármaco para el tratamiento de enfermedad hepática degenerativa, un antibiótico, y una formulación para disminuir el colesterol. En un aspecto, la molécula bioactiva es un ácido graso poliinsaturado (PUFA). En otro aspecto, la molécula bioactiva es una molécula que incluye dobles enlaces carbono-carbono en la configuración *cis*. En otro aspecto, la molécula bioactiva es una molécula que incluye un doble enlace cada tres carbonos.

30 En un aspecto de esta realización, el organismo es un microorganismo, y en otro aspecto, el organismo es una planta.

35 Otra realización de la presente invención se refiere a un método para producir una planta que tiene un perfil de ácido graso poliinsaturado (PUFA) que difiere de la planta de origen natural, que comprende modificar genéticamente células de la planta para que expresen un sistema PKS que comprende al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS. El dominio del sistema PUFA PKS está codificado por cualquiera de las secuencias de ácido nucleico descritas anteriormente.

40 Otra realización más de la presente invención se refiere a un método para producir leche animal humanizada, que comprende modificar genéticamente células productoras de leche de un animal no humano productor de leche con al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS. El dominio del sistema PUFA PKS está codificado por cualquiera de las secuencias de ácido nucleico descritas anteriormente.

45 Otra realización más de la presente invención se refiere a un método para producir un microbio recombinante, que comprende modificar genéticamente células microbianas para expresar al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS. El dominio del sistema PUFA PKS está codificado por cualquiera de las secuencias de ácido nucleico descritas anteriormente.

50 Otra realización más de la presente invención se refiere a una célula hospedadora recombinante que se ha modificado para que exprese un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS), donde la PKS cataliza reacciones enzimáticas tanto iterativas como no iterativas. El sistema PUFA PKS comprende: (a) al menos dos dominios enoil ACP reductasa (ER); (b) al menos seis dominios de proteína de transporte de acilo (ACP); (c) al menos dos dominios  $\beta$ -ceto acil ACP sintasa (KS); (d) al menos un dominio acil transferasa (AT); (e) al menos un dominio cetorreductasa (KR); (f) al menos dos dominios  $\beta$ -hidroxi acil ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP acil transferasa (MAT). En un aspecto, el sistema PUFA PKS es un sistema PUFA PKS eucariota. En otro aspecto, el sistema PUFA PKS es un sistema PUFA PKS de algas, y preferiblemente un sistema PUFA PKS de

Trastoquitriales, que puede incluir, aunque sin limitación, un sistema PUFA PKS de *Schizochytrium* o un sistema PUFA PKS de *Thraustochytrium*.

En esta realización, el sistema PUFA PKS puede expresarse en una célula hospedadora procarionta o en una célula hospedadora eucariota. En un aspecto, la célula hospedadora es una célula vegetal. Por consiguiente, una realización de la invención es un método para producir un producto que contiene al menos un PUFA, que comprende cultivar una planta que comprende dicha célula vegetal en condiciones eficaces para producir el producto. La célula hospedadora es una célula microbiana y en este caso, una realización de la presente invención es un método para producir un producto que contiene al menos un PUFA, que comprende cultivar un cultivo que contiene dicha célula microbiana en condiciones eficaces para producir el producto. En un aspecto, el sistema PKS cataliza la producción directa de triglicéridos.

Otra realización más de la presente invención se refiere a un microorganismo modificado genéticamente que comprende un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS), donde la PKS cataliza reacciones enzimáticas tanto iterativas como no iterativas. El sistema PUFA PKS comprende: (a) al menos dos dominios enoil ACP reductasa (ER); (b) al menos seis dominios de proteína de transporte de acilo (ACP); (c) al menos dos dominios  $\beta$ -ceto acil ACP sintasa (KS); (d) al menos un dominio acil transferasa (AT); (e) al menos un dominio cetereductasa (KR); (f) al menos dos dominios  $\beta$ -hidroxi acil ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP acil transferasa (MAT). La modificación genética afecta a la actividad del sistema PUFA PKS. En un aspecto de esta realización, el microorganismo es un microorganismo eucariota.

Otra realización más de la presente invención se refiere a una célula hospedadora recombinante que se ha modificado para que exprese un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) no bacteriano, donde la PUFA PKS no bacteriana cataliza reacciones enzimáticas tanto iterativas como no iterativas. El sistema PUFA PKS no bacteriano comprende: (a) al menos un dominio enoil ACP reductasa (ER); (b) múltiples dominios de proteína de transporte de acilo (ACP); (c) al menos dos dominios  $\beta$ -ceto acil ACP sintasa (KS); (d) al menos un dominio acil transferasa (AT); (e) al menos un dominio cetereductasa (KR); (f) al menos dos dominios  $\beta$ -hidroxi acil ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP acil transferasa (MAT).

#### Breve descripción de las figuras

La Fig. 1 es una representación gráfica de la estructura de dominios del sistema PUFA PKS de *Schizochytrium*.

La Fig. 2 muestra una comparación de dominios PKS de *Schizochytrium* y *Shewanella*.

La Fig. 3 muestra una comparación de dominios PKS de *Schizochytrium* y un sistema PKS relacionado de *Nostoc* cuyo producto es un ácido graso de cadena larga que no contiene ningún doble enlace.

#### Descripción detallada de la invención

La presente invención se refiere en líneas generales a sistemas de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) derivados no bacterianos, a organismos modificados genéticamente que comprenden sistemas PUFA PKS no bacterianos, a métodos para preparar y usar dichos sistemas para la producción de productos de interés, incluyendo moléculas bioactivas, y a nuevos métodos para identificar nuevos microorganismos eucariotas que tengan dicho sistema PUFA PKS. Como se usa en este documento, un sistema PUFA PKS generalmente tiene las siguientes características de identificación: (1) produce PUFA como producto natural del sistema; y (2) comprende varias proteínas multifuncionales ensambladas en un complejo que realiza procesamiento tanto iterativo de la cadena de ácido graso así como proceso *no iterativo*, incluyendo isomerización *trans-cis* y reacciones de reducción de enoil en ciclos seleccionados (véase la Fig. 1, por ejemplo).

Más específicamente, en primer lugar, un sistema PUFA PKS que forma la base de esta invención produce ácidos grasos poliinsaturados (PUFA) como productos (es decir, un organismo que contiene de forma endógena (de forma natural) dicho sistema PKS produce PUFA usando este sistema). Los PUFA mencionados en este documento son preferiblemente ácidos grasos poliinsaturados con una longitud de cadena de carbono de al menos 16 carbonos, y más preferiblemente al menos 18 carbonos, y más preferiblemente al menos 20 carbonos, y más preferiblemente 22 o más carbonos, con al menos 3 o más dobles enlaces, y preferiblemente 4 o más, y más preferiblemente 5 o más, e incluso más preferiblemente 6 o más dobles enlaces, donde todos los dobles enlaces están en configuración *cis*. Un objeto de la presente invención es encontrar o crear mediante manipulación genética o manipulación del producto final, sistemas PKS que produzcan ácidos grasos poliinsaturados de longitud de cadena deseada y con cantidades deseadas de dobles enlaces. Ejemplos de PUFA incluyen, aunque sin limitación, DHA (ácido docosahexaenoico (C22:6,  $\omega$ -3)), DPA (ácido docosapentaenoico (C22:5,  $\omega$ -6)), y EPA (ácido eicosapentaenoico (C20:5,  $\omega$ -3)).

En segundo lugar el sistema PUFA PKS descrito en este documento incorpora reacciones tanto iterativas como no iterativas, que distinguen el sistema de los sistemas PKS descritos previamente (por ejemplo, tipo I, tipo II o modular). Más particularmente, el sistema PUFA PKS descrito en este documento contiene dominios que parecen funcionar durante cada ciclo así como aquellos que parecen funcionar durante solamente algunos de los ciclos. Un

aspecto clave de esto puede estar relacionado con los dominios que muestran homología con las enzimas FabA bacterianas. Por ejemplo, la enzima FabA de *E. coli* ha demostrado poseer dos actividades enzimáticas. Posee una actividad de deshidratación en que se extrae una molécula de agua (H<sub>2</sub>O) de una cadena de carbono que contiene un grupo hidroxilo, dejando un doble enlace *trans* en esa cadena de carbono. Además, tiene una actividad isomerasa en que el doble enlace *trans* se convierte en la configuración *cis*. Esta isomerización se consigue junto con una migración de la posición del doble enlace a carbonos adyacentes. En sistemas PKS (y FAS), la cadena principal de carbono se prolonga en incrementos de dos carbonos. Por lo tanto, se puede predecir la cantidad de reacciones de extensión necesarias para producir los productos PUFA de estos sistemas PKS. Por ejemplo, para producir DHA (C22:6, todo *cis*) se requieren 10 reacciones de extensión. Como existen solamente 6 dobles enlaces en el producto final, significa que durante algunos de los ciclos de reacción, se retiene doble enlace (como isómero *cis*), y en otros, el doble enlace se reduce antes de la siguiente extensión.

Antes del descubrimiento de un sistema PUFA PKS en bacterias marinas (véase la patente de Estados Unidos n.º 6.140.486), no se sabía que los sistemas PKS poseían esta combinación de reacciones enzimáticas iterativas y selectivas, y no se creía que fueran capaces de producir dobles enlaces carbono-carbono en configuración *cis*. Sin embargo, el sistema PUFA PKS descrito por la presente invención tiene la capacidad de introducir dobles enlaces *cis* y la capacidad de variar la secuencia de reacción en el ciclo.

Por lo tanto, los presentes inventores proponen usar estas características del sistema PUFA PKS para producir un intervalo de moléculas bioactivas que no podían producirse por los sistemas PKS descritos previamente (Tipo II y Tipo I y modular). Estas moléculas bioactivas incluyen, aunque sin limitación, ácidos grasos poliinsaturados (PUFA), antibióticos u otros compuestos bioactivos, muchos de los cuales se analizarán a continuación. Por ejemplo, usando el conocimiento de las estructuras génicas de PUFA PKS descritas en este documento, puede usarse cualquiera de varios métodos para alterar los genes de PUFA PKS, o combinar partes de estos genes con otros sistemas de síntesis, incluyendo otros sistemas PKS, de modo que se produzcan nuevos productos. La capacidad inherente de este tipo particular de sistema para hacer reacciones tanto iterativas como selectivas posibilitará que este sistema produzca productos que no se encontrarían si se aplicaran métodos similares a otros tipos de sistemas PKS.

En una realización, un sistema PUFA PKS de acuerdo con la presente invención comprende al menos los siguientes dominios biológicamente activos: (a) al menos dos dominios enoil ACP reductasa (ER); (b) al menos seis dominios de proteína de transporte de acilo (ACP); (c) al menos dos dominios β-ceto acil ACP sintasa (KS); (d) al menos un dominio acil transferasa (AT); (e) al menos un dominio cetorreductasa (KR); (f) al menos dos dominios β-hidroxi acil ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP acil transferasa (MAT). Las funciones de estos dominios generalmente son conocidas de forma individual en la técnica y se describirán en detalle a continuación con respecto al sistema PUFA PKS de la presente invención.

En otra realización, el sistema PUFA PKS comprende al menos los siguientes dominios biológicamente activos: (a) al menos un dominio enoil ACP reductasa (ER); (b) múltiples dominios de proteína de transporte de acilo (ACP) (al menos cuatro, y preferiblemente al menos cinco, y más preferiblemente al menos seis, incluso más preferiblemente siete, ocho, nueve o más de nueve); (c) al menos dos dominios β-ceto acil ACP sintasa (KS); (d) al menos un dominio acil transferasa (AT); (e) al menos un dominio cetorreductasa (KR); (f) al menos dos dominios β-hidroxi acil ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP acil transferasa (MAT). Preferiblemente, dicho sistema PUFA PKS es un sistema PUFA PKS no bacteriano.

En una realización, un sistema PUFA PKS de la presente invención es un sistema PUFA PKS no bacteriano. En otras palabras, en una realización, el sistema PUFA PKS de la presente invención se aísla de un organismo que no es una bacteria, o es un homólogo de o deriva de un sistema PUFA PKS de un organismo que no es una bacteria, tal como un eucariota o una arqueobacteria. Los eucariotas se separan de los procariotas en base al grado de diferenciación de las células. El grupo superior con más diferenciación se llama eucariótico. El grupo inferior con menos células diferenciadas se llama procariótico. En general, los procariotas no poseen una membrana nuclear como no muestran mitosis durante la división celular, tienen solamente un cromosoma, su citoplasma contiene ribosomas 70S, no poseen ninguna mitocondria, retículo endoplasmático, cloroplastos, lisosomas o aparato de golgi, sus flagelos (si están presentes) consisten en una única fibrilla. En contraste, los eucariotas tienen una membrana nuclear, muestran mitosis durante la división celular, tienen muchos cromosomas, su citoplasma contiene ribosomas 80S, poseen mitocondrias, retículo endoplasmático, cloroplastos (en algas), lisosomas y aparato de golgi, y sus flagelos (si están presentes) consisten en muchas fibrillas. En general, las bacterias son procariotas, mientras que las algas, hongos, protistas, protozoos y plantas superiores son eucariotas. Los sistemas PUFA PKS de las bacterias marinas (por ejemplo, *Shewanella* y *Vibrio marinus*) no son la base de la presente invención, aunque la presente invención contempla el uso de dominios de estos sistemas PUFA PKS bacterianos *junto con* dominios de los sistemas PUFA PKS no bacterianos de la presente invención. Por ejemplo, de acuerdo con la presente invención, pueden producirse organismos modificados genéticamente que incorporan dominios funcionales de PUFA PKS no bacterianos con dominios funcionales de PUFA PKS bacterianos, así como dominios funcionales de PKS o proteínas de otros sistemas PKS (tipo I, tipo II, modular) o sistemas FAS.



*Schizochytrium* es un microorganismo marino traustozouario que acumula grandes cantidades de triacilglicérolos ricos en DHA y ácido docosapentaenoico (DPA; 22:5 ω-6); por ejemplo, 30 % DHA + DPA en peso seco (Barclay *et al.*, J. Appl. Phycol. 6, 123 (1994)). En eucariotas que sintetizan PUFA de 20 y 22 carbonos mediante una ruta de elongación/desaturación, las combinaciones de intermedios de 18, 20 y 22 carbonos son relativamente grandes de modo que experimentos de marcaje *in vivo* usando [<sup>14</sup>C]-acetato revelan una clara cinética de precursor-producto para los intermedios predichos (Gellerman *et al.*, Biochim. Biophys. Acta 573:23 (1979)). Además, los intermedios radiomarcados proporcionados de forma exógena a dichos organismos se convierten en los productos PUFA finales. Los presentes inventores han demostrado que [<sup>14</sup>C]-acetato se captaba rápidamente por células de *Schizochytrium* y se incorporaba en ácidos grasos, pero en el tiempo de marcaje más corto (1 min), DHA contenía un 31 % del marcador recuperado en ácidos grasos, y este porcentaje permanecía esencialmente inalterado durante los 10-15 min de incorporación de [<sup>14</sup>C]-acetato y las posteriores 24 horas de crecimiento en cultivo (véase el Ejemplo 3). Asimismo, DPA representaba el 10 % del marcador en todo el experimento. No existen evidencias de una relación precursor-producto entre ácidos grasos de 16 o 18 carbonos y los ácidos grasos poliinsaturados de 22 carbonos. Estos resultados son coherentes con la rápida síntesis de DHA a partir de [<sup>14</sup>C]-acetato que implica combinaciones muy pequeñas (posiblemente unidas a enzima) de intermedios. Un homogeneizado libre de células derivado de cultivos de *Schizochytrium* incorporaba [<sup>14</sup>C]-malonil-CoA en DHA, DPA, y ácidos grasos saturados. Se recuperaron las mismas actividades biosintéticas por una fracción sobrenadante de 100.000 xg pero no estaban presentes en el sedimento de membrana. Por tanto, la síntesis de DHA y DPA en *Schizochytrium* no implica desaturasas unidas a membrana o enzimas de elongación de ácidos grasos como los descritos para otros eucariotas (Parker-Barnes *et al.*, 2000, *supra*; Shanklin *et al.*, 1998, *supra*). Estos datos de fraccionamiento contrastan con los obtenidos de las enzimas de *Shewanella* (véase Metz *et al.*, 2001, *supra*) y pueden indicar el uso de una molécula aceptora de acilo diferente (soluble), tal como CoA, por la enzima de *Schizochytrium*.

En la solicitud de Estados Unidos n.º de serie 09/231.899 (patente de Estados Unidos n.º 6.566.583) en trámite junto con la presente, se construye una biblioteca de ADNc a partir de *Schizochytrium* y se secuenciaron aproximadamente 8.000 clones aleatorios (EST). Dentro de este conjunto de datos, solamente un gen expresado moderadamente (0,3 % de todas las secuencias) se identificó como ácido graso desaturasa, aunque una segunda desaturasa putativa estaba representada por un único clon (0,01 %). Por el contrario, secuencias que mostraban homología a 8 de los 11 dominios de los genes PKS de *Shewanella* mostrados en la Fig. 2 se identificaron todos a frecuencias de 0,2-0,5 %. En la solicitud de Estados Unidos n.º de serie 09/231.899, se secuenciaron varios clones de ADNc que mostraban homología con los genes PKS de *Shewanella*, y se ensamblaron diversos clones en secuencias de ácido nucleico que representaban dos fases parciales de lectura abierta y una fase completa de lectura abierta. Los nucleótidos 390-4443 de la secuencia de ADNc que contenía la primera fase parcial de lectura abierta descrita en la solicitud de Estados Unidos n.º de serie 09/231.899 (indicada en la misma como SEC ID N° 69) coincide con los nucleótidos 4677-8730 (más el codón de parada) de la secuencia indicada en este documento como OrfA (SEC ID N° 1). Los nucleótidos 1-4876 de la secuencia de ADNc que contenía la segunda fase parcial de lectura abierta descrita en la solicitud de Estados Unidos n.º de serie 09/231.899 (indicada en la misma como SEC ID N° 71) coincide con los nucleótidos 1311-6177 (más el codón de parada) de la secuencia indicada en este documento como OrfB (SEC ID N° 3). Los nucleótidos 145-4653 de la secuencia de ADNc que contiene la fase completa de lectura abierta descrita en la solicitud de Estados Unidos n.º de serie 09/231.899 (indicada en la misma como SEC ID N° 76 y denominada incorrectamente como fase parcial de lectura abierta) coincide con la secuencia completa (más el codón de parada) de la secuencia indicada en este documento como OrfC (SEC ID N° 5).

La secuenciación adicional de clones de ADNc y genómicos por los presentes inventores permitió la identificación de la secuencia genómica de longitud completa de cada uno de OrfA, OrfB y OrfC y la identificación completa de los dominios con homología a aquellos en *Shewanella* (véase la Fig. 2). Se indica que en *Schizochytrium*, el ADN genómico y el ADNc son idénticos, debido a la ausencia de intrones en el genoma del organismo, según el conocimiento de los presentes inventores. Por lo tanto, una referencia a una secuencia de nucleótidos de *Schizochytrium* puede referirse a ADN genómico o ADNc. En base a la comparación de los dominios PKS de *Schizochytrium* con *Shewanella*, claramente, el genoma de *Schizochytrium* codifica proteínas que son muy similares a las proteínas en *Shewanella* que son capaces de catalizar la síntesis de EPA. Las proteínas en *Schizochytrium* constituyen un sistema PUFA PKS que cataliza la síntesis de DHA y DPA. Como se analiza en detalle en este documento, una modificación simple del esquema de reacción identificado para *Shewanella* permitirá la síntesis de DHA en *Schizochytrium*. La homología entre los genes procariontes de *Shewanella* y eucariotas de *Schizochytrium* sugiere que la PUFA PKS ha experimentado transferencia génica lateral.

La Fig. 1 es una representación gráfica de las tres fases de lectura abierta del sistema PUFA PKS de *Schizochytrium*, e incluye la estructura de dominios de este sistema PUFA PKS. Como se describe en el Ejemplo 1 a continuación, la estructura de dominios de cada fase de lectura abierta es la siguiente:

#### Fase de lectura abierta A (OrfA):

La secuencia completa de nucleótidos para OrfA está representada en este documento como SEC ID N° 1. Los nucleótidos 4677-8730 de la SEC ID N° 1 corresponden a los nucleótidos 390-4443 de la secuencia indicada como SEC ID N° 69 en la solicitud de Estados Unidos n.º de serie 09/231.899. Por lo tanto, los nucleótidos 1-4676 de la SEC ID N° 1 representan secuencia adicional que no se había descrito en la solicitud de Estados Unidos n.º de serie

- 09/231.899. Esta nueva región de la SEC ID N° 1 codifica los siguientes dominios en OrfA: (1) El dominio OrfA-KS; (2) el dominio OrfA-MAT; y (3) al menos una parte de la región del dominio ACP (por ejemplo, al menos los dominios ACP 1-4). Se indica que los nucleótidos 1-389 de la SEC ID N° 69 en la solicitud de Estados Unidos n.º de serie 09/231.899 no coinciden con los 389 nucleótidos que están cadena arriba de la posición 4677 en la SEC ID N° 1 descrita en este documento. Por lo tanto, las posiciones 1-389 de la SEC ID N° 69 en la solicitud de Estados Unidos n.º de serie 09/231.899 parecen estar colocados incorrectamente cerca de los nucleótidos 390-4443 de esa secuencia. La mayoría de estos primeros 389 nucleótidos (aproximadamente las posiciones 60-389) son coincidentes con una parte cadena arriba de OrfA (SEC ID N° 1) de la presente invención y por lo tanto, se cree que sucedió un error en el esfuerzo por preparar el contig de las construcciones de ADNc en la solicitud de Estados Unidos n.º de serie 09/231.899. La región en la cual sucedió el error de alineación en la solicitud de Estados Unidos n.º de serie 09/231.899 está dentro de la región de secuencia altamente repetitiva (es decir, la región ACP, analizada a continuación), que probablemente creó alguna confusión en el ensamble de esa secuencia a partir de diversos clones de ADNc.
- OrfA es una secuencia de 8730 nucleótidos (sin incluir el codón de parada) que codifica una secuencia de 2910 aminoácidos, representada en este documento como SEC ID N° 2. Dentro de la OrfA hay doce dominios: (a) un dominio  $\beta$ -ceto acil ACP sintasa (KS); (b) un dominio malonil-CoA:ACP acil transferasa (MAT); (c) nueve dominios de proteína de transporte de acilo (ACP); y (d) un dominio cetorreductasa (KR).
- La secuencia de nucleótidos para OrfA se ha depositado en GenBank como el n.º de acceso AF378327 (secuencia de aminoácidos con n.º de acceso AAK728879). Se comparó OrfA con secuencias conocidas en una búsqueda convencional BLAST (búsqueda de homología BLAST en BLAST 2.0 Basic usando blastp para búsquedas de aminoácidos, blastn para búsqueda de ácido nucleico, y BlastX para búsquedas de ácido nucleico y búsquedas de la secuencia traducida de aminoácidos en las 6 fases de lectura abierta con parámetros normales por defecto, donde la secuencia de consulta se filtra para regiones de baja complejidad por defecto (descrito en Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. y Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res.* 25:3389-3402, incorporado por referencia en este documento en su totalidad).
- A nivel de ácido nucleico OrfA no tiene homología significativa con ninguna secuencia conocida de nucleótidos. A nivel de aminoácidos, las secuencias con el mayor grado de homología a OrfA fueron: glucolípidos sintasa de heterocistos de *Nostoc* sp. 7120 (n.º de acceso NC\_003272), que era un 42 % idéntico a OrfA sobre 1001 restos de aminoácido; y Orf8 de *Moritella marinus* (*Vibrio marinus*) (n.º de acceso AB025342), que era un 40 % idéntica a OrfA sobre 993 restos de aminoácido.
- El primer dominio en OrfA es un dominio KS, también mencionado en este documento como OrfA-KS. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde el punto de inicio entre aproximadamente las posiciones 1 y 40 de la SEC ID N° 1 (OrfA) hasta un punto final entre aproximadamente las posiciones 1428 y 1500 de la SEC ID N° 1. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio OrfA-KS está representada en este documento como SEC ID N° 7 (posiciones 1-1500 de la SEC ID N° 1). La secuencia de aminoácidos que contiene el dominio KS abarca desde el punto de inicio entre aproximadamente las posiciones 1 y 14 de la SEC ID N° 2 (OrfA) hasta un punto final entre aproximadamente las posiciones 476 y 500 de la SEC ID N° 2. La secuencia de aminoácidos que contiene el dominio OrfA-KS está representada en este documento como la SEC ID N° 8 (posiciones 1-500 de la SEC ID N° 2). Se aprecia que el dominio OrfA-KS contiene un motivo de sitio activo: DXAC\* (\*C<sub>215</sub> del sitio de unión a acilo).
- De acuerdo con la presente invención, un dominio o proteína que tenga actividad biológica (función) 3-ceto acil ACP sintasa (KS) se caracteriza como la enzima que realiza la etapa inicial del ciclo de reacción de elongación de FAS (y PKS). El grupo acilo destinado para elongación se une a un resto de cisteína en el sitio activo de la enzima por un enlace tio-éster. En la reacción de múltiples etapas, la acil-enzima experimenta condensación con malonil ACP para formar ceto acil ACP, CO<sub>2</sub> y enzima libre. La KS desempeña un papel clave en el ciclo de elongación y en muchos sistemas ha demostrado poseer mayor especificidad de sustrato que otras enzimas del ciclo de reacción. Por ejemplo, *E. coli* tiene tres enzimas KS distintas-cada una con su propio papel particular en la fisiología del organismo (Magnuson *et al.*, *Microbiol. Rev.* 57, 522 (1993)). Los dos dominios KS del sistema PUFA PKS podrían tener distintos papeles en la secuencia de reacción biosintética de PUFA.
- Como una clase de enzimas, las KS se han caracterizado bien. Se conocen las secuencias de muchos genes KS verificados, se han identificado los motivos de sitio activo y se han determinado las estructuras cristalinas de varios. Las proteínas (o dominios de proteínas) pueden identificarse fácilmente como pertenecientes a la familia KS de enzimas por homología a secuencias KS conocidas.
- El segundo dominio en OrfA es un dominio MAT, también mencionado en este documento como OrfA-MAT. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1723 y 1798 de la SEC ID N° 1 (OrfA) hasta un punto final entre aproximadamente las posiciones 2805 y 3000 de la SEC ID N° 1. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio OrfA-MAT está representada en este documento como SEC ID N° 9 (posiciones 1723-3000 de la SEC ID

Nº 1). La secuencia de aminoácidos que contiene el dominio MAT abarca desde un punto de inicio entre aproximadamente las posiciones 575 y 600 de la SEC ID Nº 2 (OrfA) hasta un punto final entre aproximadamente las posiciones 935 y 1000 de la SEC ID Nº 2. La secuencia de aminoácidos que contiene el dominio OrfA-MAT está representada en este documento como SEC ID Nº 10 (posiciones 575-1000 de la SEC ID Nº 2). Se aprecia que el dominio OrfA-MAT contiene un motivo de sitio activo: GHS\*<sup>XG</sup> (\*S<sub>706</sub> del sitio de unión a acilo), representado en este documento como SEC ID Nº 11.

De acuerdo con la presente invención, un dominio o proteína que tiene actividad biológica (función) malonil-CoA:ACP acil transferasa (MAT) se caracteriza como una que transfiere el resto malonilo desde malonil-CoA a ACP. Además del motivo de sitio activo (GxSxG), estas enzimas poseen un motivo prolongado (aminoácidos R y Q en posiciones clave) que las identifica como enzimas MAT (en contraste con el dominio AT de OrfB de *Schizochytrium*). En algunos sistemas PKS (pero no el dominio PUFA PKS) los dominios MAT preferentemente cargarán metil o etil malonato en el grupo ACP (a partir del éster de CoA correspondiente), introduciendo de ese modo ramificaciones en la cadena lineal de carbono. Los dominios MAT pueden reconocerse por su homología a secuencias MAT conocidas y por su estructura de motivo prolongado.

Los dominios 3-11 de OrfA son nueve dominios ACP en tándem, también mencionados en este documento OrfA-ACP (el primer dominio en la secuencia es OrfA-ACP 1, el segundo dominio es OrfA-ACP 2, el tercer dominio es OrfA-ACP 3, etc.). El primer dominio ACP, OrfA-ACP 1, está contenido dentro de la secuencia de nucleótidos que abarca desde aproximadamente la posición 3343 a aproximadamente la posición 3600 de la SEC ID Nº 1 (OrfA). La secuencia de nucleótidos que contiene la secuencia que codifica el dominio OrfA-ACP 1 está representada en este documento como SEC ID Nº 12 (posiciones 3343-3600 de la SEC ID Nº 1). La secuencia de aminoácidos que contiene el primer dominio ACP abarca desde aproximadamente la posición 1115 a aproximadamente la posición 1200 de la SEC ID Nº 2. La secuencia de aminoácidos que contiene el dominio OrfA-ACP 1 está representada en este documento como SEC ID Nº 13 (posiciones 1115-1200 de la SEC ID Nº 2). Se aprecia que el dominio OrfA-ACP 1 contiene un motivo de sitio activo: LGIDS\* (\*S<sub>1157</sub> del motivo de unión a panteteína), representado en este documento por la SEC ID Nº 14.

Las secuencias de nucleótidos y aminoácidos de los nueve dominios ACP están altamente conservadas y por lo tanto, la secuencia para cada dominio no está representada en este documento por un identificador de secuencia individual. Sin embargo, en base a la información descrita en este documento, un experto en la materia puede determinar fácilmente la secuencia que contiene cada uno de los otros ocho dominios ACP (véase el análisis a continuación).

Los nueve dominios ACP juntos abarcan una región de OrfA desde aproximadamente la posición 3283 hasta aproximadamente la posición 6388 de la SEC ID Nº 1, que corresponde a las posiciones de aminoácido de aproximadamente 1095 a aproximadamente 2096 de la SEC ID Nº 2. La secuencia de nucleótidos para la región ACP completa que contiene los nueve dominios está representada en este documento como SEC ID Nº 16. La región representada por la SEC ID Nº 16 incluye los segmentos enlazadores entre dominios ACP individuales. El intervalo repetitivo para los nueve dominios está aproximadamente cada 330 nucleótidos de la SEC ID Nº 16 (el número real de aminoácidos medido entre serinas de sitios activos adyacentes varía de 104 a 116 aminoácidos). Cada uno de los nueve dominios ACP contiene un motivo de unión a panteteína LGIDS\* (representado en este documento por la SEC ID Nº 14), donde S\* es la serina (S) del sitio de unión a panteteína. La serina (S) del sitio de unión a panteteína está localizada cerca del centro de cada secuencia de dominio ACP. En cada extremo de la región de dominio ACP y entre cada dominio ACP hay región que está muy enriquecida en prolina (P) y alanina (A), que se cree que es una región enlazadora. Por ejemplo, entre los dominios ACP 1 y 2 está la secuencia: APAPVKAAPAAPVASAPAPA, representada en este documento como SEC ID Nº 15. Las localizaciones de los restos de serina del sitio activo (es decir, el sitio de unión a panteteína) para cada uno de los nueve dominios ACP, con respecto a la secuencia de aminoácidos de la SEC ID Nº 2, son las siguientes: ACP1 = S<sub>1157</sub>; ACP2 = S<sub>1266</sub>; ACP3 = S<sub>1377</sub>; ACP4 = S<sub>1488</sub>; ACP5 = S<sub>1604</sub>; ACP6 = S<sub>1715</sub>; ACP7 = S<sub>1819</sub>; ACP8 = S<sub>1930</sub>; y ACP9 = S<sub>2034</sub>. Dado que el tamaño promedio de un dominio ACP es de aproximadamente de 85 aminoácidos, excluyendo el enlazador, y de aproximadamente 110 aminoácidos incluyendo el enlazador, estando la serina del sitio activo aproximadamente en el centro del dominio, un experto en la materia puede determinar fácilmente las posiciones de cada uno de los nueve dominios ACP en OrfA.

De acuerdo con la presente invención, un dominio o proteína que tiene actividad biológica (función) de proteína de transporte de acilo (ACP) se caracteriza como polipéptidos pequeños (típicamente, de 80 a 100 aminoácidos de longitud), que funciona como transportadores para cadenas de acilo graso en crecimiento mediante un enlace tioéster a un co-factor unido covalentemente de la proteína. Existen como unidades separadas o como dominios dentro de proteínas más grandes. Las ACP se convierten desde apo-formas inactivas en holo-formas funcionales por transferencia del resto fosfopanteteínico de CoA a un resto de serina altamente conservado de la ACP. Los grupos acilo se unen a ACP mediante un enlace tioéster en el extremo libre del resto fosfopanteteínico. Las ACP pueden identificarse por marcaje con panteteína radiactiva y por homología de secuencia con ACP conocidas. La presencia de variaciones del motivo mencionado anterior (LGIDS\*) también es un distintivo de una ACP.

El dominio 12 en OrfA es un dominio KR, también mencionado en este documento como OrfA-KR. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio de aproximadamente la posición 6598 de la SEC ID N° 1 hasta un punto final de aproximadamente la posición 8730 de la SEC ID N° 1. La secuencia de nucleótidos que contienen la secuencia que codifica el dominio OrfA-KR está representada en este documento como SEC ID N° 17 (posiciones 6598-8730 de la SEC ID N° 1). La secuencia de aminoácidos que contiene el dominio KR abarca desde el punto de inicio de aproximadamente la posición 2200 de la SEC ID N° 2 (OrfA) hasta un punto final de aproximadamente la posición 2910 de la SEC ID N° 2. La secuencia de aminoácidos que contiene el dominio OrfA-KR está representada en este documento como SEC ID N° 18 (posiciones 2200-2910 de la SEC ID N° 2). Dentro del dominio KR hay una región central con homología aldehído de cadena corta-deshidrogenasas. (KR es un miembro de esta familia). Esta región central abarca desde aproximadamente la posición 7198 hasta aproximadamente la posición 7500 de la SEC ID N° 1, que corresponde con las posiciones de aminoácido 2400-2500 de la SEC ID N° 2.

De acuerdo con la presente invención, un dominio o proteína que tiene actividad cetorreductasa, también mencionada como actividad biológica (función) 3-ceto acil ACP reductasa (KR), se caracteriza como una que cataliza la reducción dependiente de nucleótido piridina de formas 3-ceto acilo de ACP. Es la primera etapa reductora en el ciclo de elongación de biosíntesis de ácidos grasos *de novo* y una reacción a menudo realizada en biosíntesis de policétidos. Se observa similitud significativa de secuencia con una familia de enoil ACP reductasas (ER), la otra reductasa de FAS (pero no la familia ER presente en el sistema PUFA PKS), y la familia de alcohol de cadena corta deshidrogenasa. El análisis Pfam de la región PUFA PKS indicada anteriormente revela la homología con la familia de alcohol de cadena corta deshidrogenasa en la región central. El análisis Blast de la misma región revela coincidencias en el área central con enzimas KR conocidas así como una región prolongada de homología a dominios de los otros sistemas PUFA PKS caracterizados.

#### 25 Fase de lectura abierta B (OrfB):

La secuencia completa de nucleótidos para OrfB está representada en este documento como SEC ID N° 3. Los nucleótidos 1311-6177 de la SEC ID N° 3 corresponden a los nucleótidos 1-4867 de la secuencia indicada como SEC ID N° 71 en la solicitud de Estados Unidos n.º de serie 09/231.899 (La secuencia de ADNc en la solicitud de Estados Unidos n.º de serie 09/231.899 contiene aproximadamente 345 nucleótidos adicionales más allá del codón de parada, incluyendo una cola poliA). Por lo tanto, los nucleótidos 1-1310 de la SEC ID N° 1 representan secuencia adicional que no se describía en la solicitud de Estados Unidos n.º de serie 09/231.899. Esta nueva región de la SEC ID N° 3 contiene la mayor parte del dominio KS codificado por OrfB.

35 OrfB es una secuencia de 6177 nucleótidos (sin incluir el codón de parada) que codifica una secuencia de 2059 aminoácidos, representada en este documento como SEC ID N° 4. Dentro de la OrfB hay cuatro dominios: (a) un dominio  $\beta$ -ceto acil-ACP sintasa (KS); (b) un dominio de factor de longitud de cadena (CLF); (c) un dominio acil transferasa (AT); y, (d) un dominio enoil ACP-reductasa (ER).

40 La secuencia de nucleótidos para OrfB se ha depositado en GenBank con el n.º de acceso AF378328 (secuencia de aminoácidos n.º de acceso AAK728880). Se comparó OrfB con secuencias conocidas en una búsqueda BLAST convencional como se ha descrito anteriormente. A nivel de ácido nucleico, OrfB no tiene homología significativa con ninguna secuencia conocida de nucleótidos. A nivel de aminoácidos, las secuencias con el mayor grado de homología a OrfB fueron: proteína hipotética de *Shewanella sp.* (n.º de acceso U73935), que era un 53 % idéntica a ORFB sobre 458 restos de aminoácido; ORF11 de *Moritella marinus (Vibrio marinus)* (n.º de acceso AB025342), que era un 53 % idéntica a ORFB sobre 460 restos de aminoácido; ácido graso omega-3 poliinsaturado sintasa PfaD de *Photobacterium profundum* (n.º de acceso AF409100), que era un 52 % idéntica a ORFB sobre 457 restos de aminoácido; y proteína hipotética de *Nostoc sp.* 7120 (n.º de acceso NC\_003272), que era un 53 % idéntica a ORFB sobre 430 restos de aminoácido.

50 El primer dominio en OrfB es un dominio KS, también mencionado en este documento como ORFB-KS. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde el punto de inicio entre aproximadamente las posiciones 1 y 43 de la SEC ID N° 3 (OrfB) hasta un punto final entre aproximadamente las posiciones 1332 y 1350 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFB-KS está representada en este documento como SEC ID N° 19 (posiciones 1-1350 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio KS abarca desde el punto de inicio entre aproximadamente las posiciones 1 y 15 de la SEC ID N° 4 (OrfB) hasta un punto final entre aproximadamente las posiciones 444 y 450 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-KS está representada en este documento como SEC ID N° 20 (posiciones 1-450 de la SEC ID N° 4). Se aprecia que el dominio ORFB-KS contiene un motivo de sitio activo: DXAC\* (\*C<sub>196</sub> del sitio de unión a acilo). La actividad biológica de KS y métodos para identificar proteínas o dominios que tengan dicha actividad se ha descrito anteriormente.

65 El segundo dominio en OrfB es un dominio CLF, también mencionado en este documento como ORFB-CLF. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde el punto de inicio entre aproximadamente las posiciones 1378 y 1402 de la SEC ID N° 3 (OrfB) hasta un punto final entre aproximadamente las posiciones 2682 y 2700 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica

el dominio ORFB-CLF está representada en este documento como SEC ID N° 21 (posiciones 1378-2700 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio CLF abarca desde un punto de inicio entre aproximadamente las posiciones 460 y 468 de la SEC ID N° 4 (ORFB) hasta un punto final entre aproximadamente las posiciones 894 y 900 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-CLF está representada en este documento como SEC ID N° 22 (posiciones 460-900 de la SEC ID N° 4). Se observa que el dominio ORFB-CLF contiene un motivo de sitio activo de KS sin la cisteína de unión a acilo.

De acuerdo con la presente invención, un dominio o proteína se menciona como factor de longitud de cadena (CLF) en base al siguiente fundamento. El CLF se describió originalmente como característico de sistemas PKS Tipo II (enzimas disociadas) y se planteó la hipótesis de que desempeñaba un papel en la determinación de la cantidad de ciclos de elongación, y por tanto la longitud de cadena, del producto final. Las secuencias de aminoácidos de CLF muestran homología a dominios KS (y se cree que forman heterodímeros con una proteína KS), pero carecen de la cisteína del sitio activo. El papel de CLF en los sistemas PKS es actualmente controvertido. Nuevas evidencias (C. Bisang *et al.*, *Nature* 401,502 (1999)) sugieren un papel en el cebado (que proporciona el grupo acilo inicial a elongar) de los sistemas PKS. En este papel, se cree que el dominio CLF descarboxila malonato (como malonil-ACP), formando por tanto un grupo acetato que puede transferirse al sitio activo KS. Este acetato por lo tanto actúa como la molécula de "cebado" que puede experimentar la reacción inicial de elongación (condensación). Se han identificado homólogos del CLF Tipo II como dominios "de carga" en algunos sistemas PKS modulares. Un dominio con las características de secuencia del CLF se encuentra en todos los sistemas PUFA PKS actualmente identificados y en cada caso se encuentra como parte de una proteína de múltiples dominios.

El tercer dominio en OrfB es un dominio AT, también mencionado en este documento como ORFB-AT. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde el punto de inicio entre aproximadamente las posiciones 2701 y 3598 de la SEC ID N° 3 (OrfB) hasta un punto final entre aproximadamente las posiciones 3975 y 4200 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFB-AT está representada en este documento como SEC ID N° 23 (posiciones 2701-4200 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio AT abarca desde el punto de inicio entre aproximadamente las posiciones 901 y 1200 de la SEC ID N° 4 (ORFB) hasta un punto final entre aproximadamente las posiciones 1325 y 1400 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-AT está representada en este documento como SEC ID N° 24 (posiciones 901-1400 de la SEC ID N° 4). Se aprecia que el dominio ORFB-AT contiene un motivo de sitio activo de GxS\*xG (\*S<sub>1140</sub> del sitio de unión a acilo) que es característico de proteínas aciltransferasa (AT).

Una "aciltransferasa" o "AT" se refiere a una clase general de enzimas que puede realizar varias reacciones distintas de transferencia de acilo. El dominio de *Schizochytrium* muestra buena homología aun dominio presente en todos los demás sistemas PUFA PKS actualmente examinados y homología muy débil a algunas aciltransferasas cuyas funciones específicas se han identificado (por ejemplo, a malonil-CoA:ACP aciltransferasa, MAT). A pesar de la débil homología a MAT, este dominio AT no se cree que funcione como MAT porque no posee una estructura de motivo prolongado característico de dichas enzimas (véase la descripción del dominio MAT, anteriormente). Para los propósitos de esta descripción, las funciones del dominio AT en un sistema PUFA PKS incluyen, aunque sin limitación: transferencia del grupo acilo graso desde el dominio o dominios ORFA ACP a agua (es decir, una tioesterasa - que libera el grupo acilo graso como un ácido graso libre), transferencia de un grupo acilo graso a un aceptor tal como CoA, transferencia del grupo acilo entre los diversos dominios ACP, o transferencia del grupo acilo graso a una molécula aceptora lipófila (por ejemplo, a ácido lisofosfatídico).

El cuarto dominio en OrfB es un dominio ER, también mencionado en este documento como ORFB-ER. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde el punto de inicio de aproximadamente la posición 4648 de la SEC ID N° 3 (OrfB) hasta un punto final de aproximadamente la posición 6177 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFB-ER está representada en este documento como SEC ID N° 25 (posiciones 4648-6177 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio ER abarca desde un punto de inicio de aproximadamente la posición 1550 de la SEC ID N° 4 (ORFB) hasta un punto final de aproximadamente la posición 2059 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-ER está representada en este documento como SEC ID N° 26 (posiciones 1550-2059 de la SEC ID N° 4).

Describimos que este dominio tiene actividad biológica enoil reductasa (ER). La enzima ER reduce el doble enlace *trans* (introducido por la actividad DH) en el acilo graso-ACP, provocando saturación completa de esos carbonos. El dominio ER en la PUFA PKS muestra homología a una familia recién caracterizada de enzimas ER (Heath *et al.*, *Nature* 406, 145 (2000)). Heath y Rock identificaron esta nueva clase de enzimas ER clonando un gen de interés de *Streptococcus pneumoniae*, purificando una proteína expresada a partir de ese gen, y mostrando que tenía actividad ER en un ensayo *in vitro*. La secuencia del dominio ER de *Schizochytrium* de OrfB muestra homología a la proteína ER de *S. pneumoniae*. Todos los sistemas PUFA PKS actualmente examinados contienen al menos un dominio con homología de secuencia muy elevada al dominio ER de *Schizochytrium*. El sistema PUFA PKS de *Schizochytrium* contiene dos dominios ER (uno en OrfB y uno en OrfC).

Fase de lectura abierta C (OrfC):

La secuencia completa de nucleótidos para OrfC está representada en este documento como SEC ID N° 5. Los nucleótidos 1-4509 de la SEC ID N° 5 (es decir, la secuencia completa de fase de lectura abierta, sin incluir el codón de parada) corresponden a los nucleótidos 145-4653 de la secuencia indicada como SEC ID N° 76 en la solicitud de Estados Unidos n.º de serie 09/231.899 (La secuencia de ADNc en la solicitud de Estados Unidos n.º 09/231.899 contiene aproximadamente 144 nucleótidos cadena arriba del codón de inicio para OrfC y aproximadamente 110 nucleótidos más allá del codón de parada, incluyendo una cola poliA). OrfC es una secuencia de 4509 nucleótidos (sin incluir el codón de parada) que codifica una secuencia de 1503 aminoácidos, representada en este documento como SEC ID N° 6. Dentro de la OrfC hay tres dominios: (a) dos dominios  $\beta$ -hidroxi acil-ACP deshidrasa (DH) tipo FabA; y (b) un dominio enoil ACP-reductasa (ER).

La secuencia de nucleótidos para OrfC se ha depositado en GenBank con el n.º de acceso AF378329 (secuencia de aminoácidos n.º de acceso AAK728881). Se comparó OrfC con secuencias conocidas en una búsqueda BLAST convencional como se ha descrito anteriormente. A nivel de ácido nucleico, OrfC no tiene homología significativa con ninguna secuencia conocida de nucleótidos. A nivel de aminoácidos (Blastp), las secuencias con el mayor grado de homología a ORFC fueron: ORF11 de *Moritella marinus* (*Vibrio marinus*) (n.º de acceso ABO25342), que es un 45% idéntica a ORFC sobre 514 restos de aminoácido, proteína hipotética 8 de *Shewanella sp.* (n.º de acceso U73935), que es un 49% idéntica a ORFC sobre 447 restos de aminoácido, proteína hipotética de *Nostoc sp.* (n.º de acceso NC\_003272), que es un 49% idéntica a ORFC sobre 430 restos de aminoácido, y proteína hipotética 7 de *Shewanella sp.* (n.º de acceso U73935), que es un 37% idéntica a ORFC sobre 930 restos de aminoácido.

El primer dominio en OrfC es un dominio DH, también mencionado en este documento como ORFC-DH1. Este es uno de los dos dominios DH en OrfC, y por lo tanto se denomina DH1. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1 y 778 de la SEC ID N° 5 (OrfC) hasta un punto final entre aproximadamente las posiciones 1233 y 1350 de la SEC ID N° 5. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFC-DH1 está representada en este documento como SEC ID N° 27 (posiciones 1-1350 de la SEC ID N° 5). La secuencia de aminoácidos que contiene el dominio DH1 abarca desde el punto de inicio entre aproximadamente las posiciones 1 y 260 de la SEC ID N° 6 (ORFC) hasta un punto final entre aproximadamente las posiciones 411 y 450 de la SEC ID N° 6. La secuencia de aminoácidos que contiene el dominio ORFC-DH1 está representada en este documento como SEC ID N° 28 (posiciones 1-450 de la SEC ID N° 6).

Las características de ambos dominios DH (véase a continuación para DH 2) en los sistemas PUFA PKS se han descrito en las secciones precedentes. Esta clase de enzima elimina HOH de un  $\beta$ -ceto acil-ACP y deja un doble enlace *trans* en la cadena de carbono. Los dominios DH de los sistemas PUFA PKS muestran homología a enzimas DH bacterianas asociadas con sus sistemas FAS (en lugar de a los dominios DH de otros sistemas PKS). Un subconjunto de DH bacterianas, las DH tipo FabA, poseen actividad *cis-trans* isomerasa (Heath *et al.*, J. Biol. Chem., 271, 27795 (1996)). Es la homología a las DH tipo FabA lo que indica que uno o ambos dominios DH son responsables de la inserción de los dobles enlaces *cis* en los productos de PUFA PKS.

El segundo dominio en OrfC es un dominio DH, también mencionado en este documento como ORFC-DH2. Este es el segundo de los dos dominios DH en OrfC, y por lo tanto se denomina DH2. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1351 y 2437 de la SEC ID N° 5 (OrfC) hasta un punto final entre aproximadamente las posiciones 2607 y 2850 de la SEC ID N° 5. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFC-DH2 está representada en este documento como SEC ID N° 29 (posiciones 1351-2850 de la SEC ID N° 5). La secuencia de aminoácidos que contiene el dominio DH2 abarca desde el punto de inicio entre aproximadamente las posiciones 451 y 813 de la SEC ID N° 6 (ORFC) hasta un punto final entre aproximadamente las posiciones 869 y 950 de la SEC ID N° 6. La secuencia de aminoácidos que contiene el dominio ORFC-DH2 está representada en este documento como SEC ID N° 30 (posiciones 451-950 de la SEC ID N° 6). La actividad biológica DH se ha descrito anteriormente.

El tercer dominio en OrfC es un dominio ER, también mencionado en este documento como ORFC-ER. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio de aproximadamente la posición 2998 de la SEC ID N° 5 (OrfC) hasta un punto final de aproximadamente la posición 4509 de la SEC ID N° 5. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFC-ER está representada en este documento como SEC ID N° 31 (posiciones 2998-4509 de la SEC ID N° 5). La secuencia de aminoácidos que contiene el dominio ER abarca desde el punto de inicio de aproximadamente la posición 1000 de la SEC ID N° 6 (ORFC) hasta un punto final de aproximadamente la posición 1502 de la SEC ID N° 6. La secuencia de aminoácidos que contiene el dominio ORFC-ER está representada en este documento como SEC ID N° 32 (posiciones 1000-1502 de la SEC ID N° 6). La actividad biológica ER se ha descrito anteriormente.

Una realización de la presente invención se refiere a una molécula aislada de ácido nucleico que comprende una secuencia de ácido nucleico de un sistema PUFA PKS no bacteriano, un homólogo de la misma, un fragmento de la misma, y/o una secuencia de ácido nucleico que es complementaria a cualquiera de dichas secuencias de ácido nucleico. En un aspecto, la presente invención se refiere a una molécula aislada de ácido nucleico como se expone en la reivindicación 1.

De acuerdo con la presente invención, una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS es una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio del sistema PUFA PKS descrito en detalle en este documento, como se ejemplifica por el sistema PUFA PKS *Schizochytrium*. Las actividades biológicas de los diversos dominios dentro del sistema PUFA PKS de *Schizochytrium* se han descrito en detalle anteriormente. Por lo tanto, una molécula aislada de ácido nucleico descrita en este documento puede codificar el producto de traducción de cualquier fase de lectura abierta de PUFA PKS, dominio de PUFA PKS, fragmento biológicamente activo del mismo, o cualquier homólogo de una fase de lectura abierta o dominio de PUFA PKS de origen natural que tenga actividad biológica. Un homólogo de una proteína o dominio dado es una proteína o polipéptido que tiene una secuencia de aminoácidos que difiere de la secuencia de aminoácidos de referencia de origen natural (es decir, de la proteína o dominio de referencia) porque al menos uno o unos pocos, aunque sin limitación a uno o unos pocos, aminoácidos se han delecionado (por ejemplo, una versión truncada de la proteína, tal como un péptido o fragmento), insertado, invertido, sustituido y/o derivatizado (por ejemplo, por glucosilación, fosforilación, acetilación, miristoilación, prenilación, palmitación, amidación y/o adición de glucosilfosfatidil inositol). Homólogos preferidos de una proteína o dominio de PUFA PKS se describen en detalle a continuación. Se aprecia que los homólogos pueden incluir homólogos producidos de forma sintética, variantes alélicas de origen natural de una proteína o dominio dado, o secuencias homólogas de organismos diferentes al organismo del cual se obtuvo la secuencia de referencia.

En general, la actividad biológica o acción biológica de una proteína o dominio se refiere a cualquier función mostrada o realizada por la proteína o dominio que está atribuida a la forma de origen natural de la proteína o dominio medida u observada *in vivo* (es decir, en el entorno fisiológico natural de la proteína) o *in vitro* (es decir, en condiciones de laboratorio). Las actividades biológicas de los sistemas PUFA PKS y las proteínas/dominios individuales que componen un sistema PUFA PKS se han descrito en detalle en otra parte en este documento. Las modificaciones de una proteína o dominio, tal como en un homólogo o mimético (analizado a continuación), pueden producir proteínas o dominios que tienen la misma actividad biológica que la proteína o dominio de origen natural, o proteínas o dominios que tienen actividad biológica disminuida o aumentada en comparación con la proteína o dominio de origen natural. Modificaciones que provocan una disminución en la expresión o una disminución en la actividad de la proteína o dominio, pueden mencionarse como inactivación (completa o parcial), regulación negativa, o acción disminuida de una proteína o dominio. Asimismo, modificaciones que provocan un aumento en la expresión o un aumento en la actividad de la proteína o dominio, pueden mencionarse como amplificación, sobreproducción, activación, potenciación, regulación positiva o acción aumentada de una proteína o dominio. Un dominio funcional de un sistema PUFA PKS es un dominio (es decir, un dominio puede ser una parte de una proteína) que es capaz de realizar una función biológica (es decir, tiene actividad biológica).

De acuerdo con la presente invención, una molécula aislada de ácido nucleico es una molécula de ácido nucleico que se ha retirado de su entorno natural (es decir, que se ha sometido a manipulación humana), siendo su entorno natural el genoma o cromosoma en que se encuentra la molécula de ácido nucleico en la naturaleza. Por tanto, "aislado" no refleja necesariamente el grado al cual se ha purificado la molécula de ácido nucleico, sino que indica que la molécula no incluye un genoma completo o un cromosoma completo en que se encuentra la molécula de ácido nucleico en la naturaleza. Una molécula aislada de ácido nucleico puede incluir un gen. Una molécula aislada de ácido nucleico que incluye un gen no es un fragmento de un cromosoma que incluye dicho gen, sino que en su lugar incluye la región codificante y regiones reguladoras asociadas con el gen, pero no genes adicionales encontrados de forma natural en el mismo cromosoma. Una molécula aislada de ácido nucleico también puede incluir una secuencia específica de ácido nucleico flanqueada por (es decir, en el extremo 5' y/o 3' de la secuencia) ácidos nucleicos adicionales que no flanquean de forma normal la secuencia específica de ácido nucleico en la naturaleza (es decir, secuencias heterólogas). La molécula aislada de ácido nucleico puede incluir ADN, ARN (por ejemplo, ARNm), o derivados de ADN o ARN (por ejemplo, ADNc). Aunque la expresión "molécula de ácido nucleico" se refiere principalmente a la molécula física de ácido nucleico y la expresión "secuencia de ácido nucleico" se refiere principalmente a la secuencia de nucleótidos en la molécula de ácido nucleico, las dos expresiones pueden usarse de forma intercambiable, especialmente con respecto a una molécula de ácido nucleico, o una secuencia de ácido nucleico, que es capaz de codificar una proteína o dominio de una proteína.

Preferiblemente, una molécula aislada de ácido nucleico de la presente invención se produce usando tecnología de ADN recombinante (por ejemplo, amplificación por reacción en cadena de la polimerasa (PCR), clonación) o síntesis química. Las moléculas aisladas de ácido nucleico incluyen moléculas naturales de ácido nucleico y homólogos de las mismas incluyendo, aunque sin limitación, variantes alélicas naturales y moléculas modificadas de ácido nucleico en que se han insertado delecionado, sustituido y/o invertido nucleótidos de tal modo que dichas modificaciones proporcionen el efecto deseado sobre la actividad biológica del sistema PUFA PKS como se describe en este documento. Se han analizado homólogos de proteína (por ejemplo, proteínas codificadas por homólogos de ácidos nucleico) en detalle anteriormente.

Una molécula de ácido nucleico homóloga puede producirse usando varios métodos conocidos para los expertos en la materia (véase, por ejemplo, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press, 1989). Por ejemplo, pueden modificarse moléculas de ácido nucleico usando una diversidad de técnicas incluyendo, aunque sin limitación, técnicas clásicas de mutagénesis y técnicas de ADN recombinante, tales como

mutagénesis dirigida al sitio, tratamiento químico de una molécula de ácido nucleico para inducir mutaciones, escisión con enzima de restricción de un fragmento de ácido nucleico, ligamiento de fragmento de ácido nucleico, amplificación por PCR y/o mutagénesis de regiones seleccionadas de una secuencia de ácido nucleico, síntesis de mezclas de oligonucleótidos y ligamiento de grupos mezclados para "componer" una mezcla de moléculas de ácido nucleico y combinaciones de las mismas. Pueden seleccionarse moléculas de ácido de nucleico homólogas de una mezcla de ácidos nucleicos modificados seleccionando la función de la proteína codificada por el ácido nucleico y/o por hibridación con un gen de tipo silvestre.

El tamaño mínimo de una molécula de ácido nucleico de la presente invención es un tamaño suficiente para formar una sonda o cebador oligonucleotídico que sea capaz de formar un híbrido estable (por ejemplo, en condiciones de rigurosidad moderada, alta o muy alta) con la secuencia complementaria de una molécula de ácido nucleico útil en la presente invención, o de un tamaño suficiente para codificar una secuencia de aminoácidos que tenga una actividad biológica de al menos un dominio de un sistema PUFA PKS de acuerdo con la presente invención. Por tanto, el tamaño de la molécula de ácido nucleico que codifica dicha proteína puede depender de la composición del ácido nucleico y el porcentaje de homología o identidad entre la molécula de ácido nucleico y la secuencia complementaria así como de las condiciones de hibridación *per se* (por ejemplo, temperatura, concentración salina, y concentración de formamida). El tamaño mínimo de una molécula de ácido nucleico que se usa como cebador oligonucleotídico o como sonda es típicamente de al menos aproximadamente 12 a aproximadamente 15 nucleótidos de longitud si las moléculas de ácido nucleico son ricas en GC y de al menos aproximadamente 15 a aproximadamente 18 bases de longitud si son ricas en AT. No existe límite, salvo un límite práctico, en el tamaño máximo de una molécula de ácido nucleico de la presente invención, porque la molécula de ácido nucleico puede incluir una secuencia suficiente para codificar un fragmente biológicamente activo de un dominio de un sistema PUFA PKS, un dominio completo de un sistema PUFA PKS, varios dominios dentro de una fase de lectura abierta (Orf) de un sistema PUFA PKS, una Orf completa de un sistema PUFA PKS, o más de un Orf de un sistema PUFA PKS.

En una realización de la presente invención, una molécula aislada de ácido nucleico comprende o consiste esencialmente en una secuencia de ácido nucleico seleccionada entre el grupo que consiste en: SEC ID N° 20, o fragmentos biológicamente activos de las mismas. En un aspecto, la secuencia de ácido nucleico se selecciona entre el grupo de la SEC ID N° 19.

En una realización de la presente invención, cualquiera de las secuencias de aminoácidos de PUFA PKS descritas anteriormente, así como homólogos de dichas secuencias, puede producirse con de al menos uno, y hasta aproximadamente 20, aminoácidos heterólogos adicionales flanqueando cada uno de los extremos C y/o N terminales de la secuencia dada de aminoácidos. La proteína o polipéptido resultante puede mencionarse como "que consiste esencialmente en" una secuencia dada de aminoácidos. De acuerdo con la presente invención, los aminoácidos heterólogos son una secuencia de aminoácidos que no se encuentran de forma natural (es decir, no se encuentran en la naturaleza, *in vivo*) que flanquean la secuencia dada de aminoácidos o que podrían no estar codificados por los nucleótidos que flanquean la secuencia de ácido nucleico de origen natural que codifica la secuencia dada de aminoácidos tal como existe en el gen, así dichos nucleótidos en la secuencia de origen natural se traducen usando el uso convencional de codones para el organismo del cual se obtiene la secuencia dada de aminoácidos. Asimismo, la expresión "que consiste esencialmente en", cuando se usa con referencia a una secuencia de ácido nucleico en ese documento, se refiere a una secuencia de ácido nucleico que codifica una secuencia dada de aminoácidos que puede estar flanqueada por de al menos uno, y hasta como mucho aproximadamente 60, nucleótidos heterólogos adicionales en cada uno de los extremos 5' y/o 3' de la secuencia de ácido nucleico que codifica la secuencia dada de aminoácidos. Los nucleótidos heterólogos no se encuentran de forma natural (es decir, no se encuentran en la naturaleza, *in vivo*) flanqueando la secuencia de ácido nucleico que codifica la secuencia dada de aminoácidos tal como existe en el gen natural.

La presente invención también incluye una molécula aislada de ácido nucleico que comprende una secuencia de ácido nucleico que codifica una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS. En un aspecto, dicha secuencia de ácido nucleico codifica un homólogo de cualquiera de las Orf o dominios de PUFA PKS de *Schizochytrium*, incluyendo: SEC ID N° 20, donde el homólogo tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS como se ha descrito previamente en este documento.

En un aspecto de la invención, un homólogo de una proteína o dominio de PUFA PKS de *Schizochytrium* abarcado por la presente invención comprende una secuencia de aminoácidos que es al menos un 60 % idéntica a una secuencia de aminoácidos elegida entre: SEC ID N° 20, donde dicha secuencia de aminoácidos tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS. En un aspecto adicional, la secuencia de aminoácidos del homólogo es al menos un 65 % idéntica, y más preferiblemente al menos un 70 % idéntica, y más preferiblemente al menos un 75 % idéntica, y más preferiblemente al menos un 80 % idéntica y más preferiblemente al menos un 85 % idéntica, y más preferiblemente al menos un 90 % idéntica, y más preferiblemente al menos un 95 % idéntica, y más preferiblemente al menos un 96 % idéntica, y más preferiblemente al menos un 97 % idéntica, y más preferiblemente al menos un 98 % idéntica, y más preferiblemente al menos un 99 % idéntica a una secuencia de aminoácidos elegida entre: SEC ID N° 20, donde la secuencia de aminoácidos tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS.



De acuerdo con la presente invención, el término "contiguo" o "consecutivo", con respecto a secuencias de ácido nucleico o aminoácidos descritas en este documento, significa conectados en una secuencia ininterrumpida. Por ejemplo, para que una primera secuencia comprenda 30 aminoácidos contiguos "o consecutivos" de una segunda secuencia, significa que la primera secuencia incluye una secuencia ininterrumpida de 30 restos de aminoácido que es 100 % idéntica a una secuencia ininterrumpida de 30 restos de aminoácido en la segunda secuencia. Asimismo, para que una primera secuencia tenga "identidad del 100 %" con una segunda secuencia significa que la primera secuencia coincide exactamente con la segunda secuencia sin huecos entre nucleótidos o aminoácidos.

Como se usa en este documento, salvo que se especifique de otro modo, referencias a un porcentaje (%) de identidad se refiere a un evaluación de homología que se realiza usando: (1) una búsqueda de homología en BLAST 2.0 Basic BLAST usando blastp para búsquedas de aminoácidos, blastn para búsquedas de ácidos nucleicos, y blastX para búsqueda de ácido nucleico y búsquedas de aminoácidos traducidos en las 6 fases de lectura abierta, todas con parámetros convencionales por defecto, donde la secuencia de consulta se filtra para regiones de baja complejidad por defecto (descrito en Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. y Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402, incorporado por referencia en este documento en su totalidad); (2) una alineación BLAST 2 (usando los parámetros descritos a continuación); (3) y/o PSI-BLAST con los parámetros convencionales por defecto (BLAST con iteración específica de posición). Se aprecia que debido a algunas diferencias en los parámetros convencionales entre BLAST 2.0 Basic BLAST y BLAST 2, podrían reconocerse dos secuencias específicas como teniendo homología significativa usando el programa BLAST 2, mientras que una búsqueda realizada en BLAST 2.0 Basic BLAST usando una de las secuencias como secuencia de consulta puede no identificar la segunda secuencia en las máximas coincidencias. Además, PSI-BLAST proporciona una versión automatizada, fácil de usar de una búsqueda "de perfil", que es un modo sensible de buscar homólogos de secuencia. El programa primero realiza una búsqueda en base de datos BLAST con huecos. El programa PSI-BLAST usa la información de cualquier alineación significativa devuelta para construir una matriz de valores específica de posición, que reemplaza la secuencia de consulta por la siguiente ronda de búsqueda en base de datos. Por lo tanto, debe entenderse que el porcentaje de identidad puede determinarse usando uno cualquiera de estos programas.

Pueden alinearse dos secuencias específicas entre sí usando la secuencia BLAST 2 como se describe en Tatusova and Madden, (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174:247-250, incorporado por referencia en este documento en su totalidad. La alineación de secuencia BLAST 2 se realiza en blastp o blastn usando el algoritmo BLAST 2.0 para realizar una búsqueda BLAST con huecos (BLAST 2.0) entre las dos secuencias que permita la introducción de huecos (delecciones e inserciones) en la alineación resultante. Para propósitos de claridad en este documento, se realiza una alineación de secuencia BLAST 2 usando los siguientes parámetros convencionales por defecto.

Para blastn, usando matriz 0 BLOSUM62:

Recompensa por coincidencia = 1  
 Penalización por desapareamiento = -2  
 Penalizaciones por abertura de hueco (5) y extensión de hueco (2)  
 Filtro (activo) de tamaño de palabra (11) esperado (10) con x\_disminución de hueco (50).

Para blastp, usando matriz 0 BLOSUM62:

Penalizaciones por abertura de hueco (11) y extensión de hueco (1)  
 Filtro (activo) de tamaño de palabra (3) esperado (10) con x\_disminución de hueco (50).

En otra realización de la invención, una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio de un sistema PUFA PKS de la presente invención incluye una secuencia de aminoácidos que es suficientemente similar a una proteína o polipéptido de PUFA PKS de origen natural, que es una secuencia de ácido nucleico que codifica la secuencia de aminoácidos capaz de hibridar en condiciones de rigurosidad moderada, alta o muy alta (descritas a continuación) a (es decir, con) una molécula de ácido nucleico que codifica la proteína o polipéptido de PUFA PKS de origen natural (es decir, al complemento de la hebra de ácido nucleico que codifica la proteína o polipéptido de PUFA PKS de origen natural). Una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio de un sistema PUFA PKS de la presente invención está codificada por una secuencia de ácido nucleico que hibrida en condiciones de rigurosidad moderada, alta o muy alta con el complemento de una secuencia de ácido nucleico que codifica una proteína que comprende una secuencia de amino ácidos representada por la SEC ID N° 20.

Los métodos para deducir una secuencia complementaria son conocidos para los expertos en la materia. Debe apreciarse que como las tecnologías de secuenciación de aminoácidos y secuenciación de ácidos nucleicos no están completamente libres de errores, las secuencias presentadas en este documento, en el mejor de los casos, representan secuencias evidentes de dominios y proteínas de PUFA PKS de la presente invención.

Como se usa en este documento, las condiciones de hibridación se refieren a condiciones convencionales de hibridación en las que se usan moléculas de ácido nucleico para identificar moléculas similares de ácido nucleico. Dichas condiciones convencionales se describen, por ejemplo, en Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press, 1989. Sambrook *et al.*, *ibid.*, se incorpora por referencia en este documento en su totalidad (véase específicamente, las páginas 9.31-9.62). Además, se describen fórmulas para calcular las condiciones apropiadas de hibridación y lavado para conseguir hibridación que permita grados variables de desapareamiento de nucleótidos se describen, por ejemplo, en Meinkoth *et al.*, 1984, Anal. Biochem. 138, 267-284, Meinkoth *et al.*, *ibid.*, se. incorpora por referencia en este documento en su totalidad.

Más particularmente, condiciones de hibridación y lavado de rigurosidad moderada, como se menciona en este documento, se refieren a condiciones que permiten el aislamiento de moléculas de ácido nucleico que tengan al menos aproximadamente un 70 % de identidad de secuencia de ácido nucleico con la molécula de ácido nucleico que se estaba usando para sondear en la reacción de hibridación (es decir, condiciones que permiten aproximadamente un 30 % o menos de desapareamiento de nucleótidos). Condiciones de hibridación y lavado de rigurosidad alta, como se menciona en este documento, se refieren a condiciones que permiten el aislamiento de moléculas de ácido nucleico que tienen al menos aproximadamente un 80 % de identidad de secuencia de ácido nucleico con la molécula de ácido nucleico que se está usando para sondear en la reacción de hibridación (es decir, condiciones que permiten aproximadamente un 20 % o menos de desapareamiento de nucleótidos). Condiciones de hibridación y lavado de rigurosidad muy alta, como se menciona en este documento, se refiere a condiciones que permiten el aislamiento de moléculas de ácido nucleico que tienen al menos aproximadamente un 90 % de identidad de secuencia de ácido nucleico con la molécula de ácido nucleico que se está usando para sondear en la reacción de hibridación (es decir, condiciones que permiten aproximadamente un 10 % o menos de desapareamiento de nucleótidos). Como se ha analizado anteriormente, un experto en la materia puede usar las fórmulas en Meinkoth *et al.*, *ibid* para calcular las condiciones apropiadas de hibridación y lavado para conseguir estos niveles particulares de desapareamiento de nucleótidos. Dichas condiciones variarán, dependiendo de si se están formando híbridos ADN:ARN o ADN:ADN. Las temperaturas de fusión calculadas para híbridos de ADN:ADN son 10 °C menores que para híbridos de ADN:ARN. En realizaciones particulares, las condiciones de hibridación rigurosas para híbridos de ADN:ADN incluyen hibridación a una fuerza iónica de SSC 6X (Na<sup>+</sup> 0,9 M) a una temperatura entre aproximadamente 20 °C y aproximadamente 35 °C (rigurosidad más baja), más preferiblemente, entre aproximadamente 28 °C y aproximadamente 40 °C (más riguroso), e incluso más preferiblemente, entre aproximadamente 35 °C y aproximadamente 45 °C (incluso más riguroso), con condiciones apropiadas de lavado. En realizaciones particulares, las condiciones de hibridación rigurosas para híbridos de ADN:ARN incluyen hibridación a una fuerza iónica de SSC 6X (Na<sup>+</sup> 0,9 M) a una temperatura entre aproximadamente 30 °C centígrados y aproximadamente 45 °C, más preferiblemente, entre aproximadamente 38 °C y aproximadamente 50 °C, e incluso más preferiblemente, entre aproximadamente 45 °C y aproximadamente 55 °C, con condiciones de lavado igualmente rigurosas. Estos valores se basan en cálculos de una temperatura de fusión para moléculas mayores de aproximadamente 100 nucleótidos, formamida al 0 % y un contenido de G + C de aproximadamente del 40 %. Como alternativa, la T<sub>m</sub> puede calcularse de forma empírica como se expone en Sambrook *et al.*, *supra*, páginas 9.31 to 9.62. En general, las condiciones de lavado deben ser tan rigurosas como sea posible y deben ser apropiadas para las condiciones elegidas de hibridación. Por ejemplo, las condiciones de hibridación pueden incluir una combinación de condiciones salinas y de temperatura que son aproximadamente 20-25 °C inferiores a la T<sub>m</sub> calculada de un híbrido particular, y las condiciones de lavado típicamente incluyen una combinación de condiciones salinas y de temperatura que son aproximadamente 12-20 °C inferiores a la T<sub>m</sub> calculada del híbrido particular. Un ejemplo de condiciones de hibridación adecuadas para su uso con híbridos de ADN:ADN incluyen una hibridación de 2-24 horas en SSC 6X (formamida al 50 %) a aproximadamente 42 °C, seguida por etapas de lavado que incluyen uno o más lavado a temperatura ambiente en SSC 2X, seguidos por lavados adicionales a temperaturas mayores o inferiores fuerzas iónicas (por ejemplo, al menos un lavado como aproximadamente 37 °C en SSC aproximadamente 0,1X-0,5X, seguido por al menos un lavado a aproximadamente 68 °C en SSC aproximadamente 0,1X-0,5X).

Otra realización de la presente invención incluye una molécula de ácido nucleico recombinante que comprende un vector recombinante y una molécula de ácido nucleico que comprende una secuencia de ácido nucleico que codifica una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS de acuerdo con la invención. Dichas secuencias de ácido nucleico se han descrito en detalle anteriormente. De acuerdo con la presente invención, un vector recombinante es una molécula de ácido nucleico modificada por ingeniería (es decir, producida de forma artificial) que se usa como herramienta para manipular una secuencia de ácido nucleico de elección y para introducir dicha secuencia de ácido nucleico en una célula hospedadora. El vector recombinante por lo tanto es adecuado para su uso en clonación, secuenciación, y/o manipulación de otro modo de la secuencia de ácido nucleico de elección, tal como mediante la expresión y/o suministro de la secuencia de ácido nucleico de elección en una célula hospedadora para formar una célula recombinante. Dicho vector típicamente contiene secuencias heterólogas de ácido nucleico, es decir secuencias de ácido nucleico que no se encuentran de forma natural adyacentes a la secuencia de ácido nucleico a clonar o suministrar, aunque el vector también puede contener secuencias reguladoras de ácido nucleico (por ejemplo, promotores, regiones no traducidas) que se encuentran de forma natural adyacentes a moléculas de ácido nucleico de la presente invención o que son útiles para la expresión de las moléculas de ácido nucleico de la presente invención (analizadas en detalle a continuación). El vector puede ser ARN o ADN, procariota o eucariota, y típicamente es un plásmido. El vector puede mantenerse

como un elemento extracromosómico (por ejemplo, un plásmido) o puede integrarse en el cromosoma de un organismo recombinante (por ejemplo, un microbio o una planta). El vector completo puede permanecer en su sitio dentro de una célula hospedadora, o en ciertas condiciones, el ADN plasmídico puede delecionarse, dejando detrás la molécula de ácido nucleico de la presente invención. La molécula de ácido nucleico integrada puede estar bajo el control del promotor cromosómico, bajo el control del promotor nativo o plasmídico, o bajo una combinación de varios controles de promotor. Pueden integrarse una o múltiples copias de la molécula de ácido nucleico en el cromosoma. Un vector recombinante de la presente invención puede contener al menos un marcador de selección.

En una realización, un vector recombinante usado en una molécula de ácido nucleico recombinante de la presente invención es un vector de expresión. Como se usa en este documento, la expresión "vector de expresión" se usa para hacer referencia a un vector que es adecuado para la producción de un producto codificado (por ejemplo, una proteína de interés). En esta realización, una secuencia de ácido nucleico que codifica el producto a producir (por ejemplo, un dominio de PUFA PKS) se inserta en el vector recombinante para producir una molécula de ácido nucleico recombinante. La secuencia de ácido nucleico que codifica la proteína a producir se inserta en el vector de un modo que une de forma funcional la secuencia de ácido nucleico a secuencias reguladoras en el vector, lo que posibilita la transcripción y traducción de la secuencia de ácido nucleico dentro de la célula hospedadora recombinante.

En otra realización, un vector recombinante usado en una molécula de ácido nucleico recombinante de la presente invención es un vector de direccionamiento. Como se usa en este documento, la expresión "vector de direccionamiento" se usa para hacer referencia a un vector que se usa para suministrar una molécula de ácido nucleico particular en una célula hospedadora recombinante, donde la molécula de ácido nucleico se usa para delecionar o inactivar un gen endógeno dentro de la célula o microorganismo hospedador (es decir, se usa para alteración génica dirigida o tecnología knock-out). Dicho vector también puede conocerse en la técnica como vector "knock-out". En un aspecto de esta realización, una parte del vector, pero más típicamente la molécula de ácido nucleico insertada en el vector (es decir, el inserto), tiene una secuencia de ácido nucleico que es homóloga a una secuencia de ácido nucleico de un gen diana en la célula hospedadora (es decir, un gen que es una diana para delecionarse o inactivarse). La secuencia de ácido nucleico del inserto del vector se diseña para unirse al gen diana de modo que el gen diana y su inserto experimenten recombinación homóloga, mediante lo cual el gen diana endógeno se deleciona, se inactiva o atenúa (es decir, por al menos una parte del gen diana endógeno que se está mutando o delecionando).

Típicamente, una molécula de ácido nucleico recombinante incluye al menos una molécula de ácido nucleico de la presente invención unida de forma funcional a una o más secuencias de control de la transcripción. Como se usa en este documento, la expresión "molécula recombinante" o "molécula de ácido nucleico recombinante" se refiere principalmente a una molécula de ácido nucleico o secuencia de ácido nucleico unida de forma funcional a una secuencia de control de la transcripción, pero puede usarse de forma intercambiable con la expresión "molécula de ácido nucleico", cuando dicha molécula de ácido nucleico es una molécula recombinante como se analiza en este documento. De acuerdo con la presente invención, la expresión "unido de forma funcional" se refiere a unir una molécula de ácido nucleico a una secuencia de control de la transcripción de un modo tal que la molécula sea capaz de expresarse cuando se introduce por transfección (es decir, transformación, transducción, transfección, conjugación o conducción) en una célula hospedadora. Las secuencias de control de la transcripción son secuencias que controlan el inicio, elongación, o terminación de la transcripción. Las secuencias de control de la transcripción particularmente importantes son aquellas que controlan el inicio de la transcripción, tales como secuencias promotoras, potenciadoras, operadoras y represoras. Las secuencias de control de la transcripción adecuadas incluyen cualquier secuencia de control de la transcripción que pueda funcionar en una célula hospedadora u organismo en que tiene que introducirse la molécula de ácido nucleico recombinante.

Las moléculas de ácido nucleico recombinantes de la presente invención también pueden contener secuencias reguladoras adicionales, tales como secuencias reguladoras de la traducción, orígenes de replicación, y otras secuencias reguladoras que son compatibles con la célula recombinante. En una realización, una molécula recombinante de la presente invención, incluyendo aquellas que se integran en el cromosoma de la célula hospedadora, también contiene señales de secreción (es decir, se secuencias de ácido nucleico del segmento señal) para posibilitar que una proteína expresada se secrete desde la célula que produce la proteína. Los segmentos señal adecuados incluyen un segmento señal que está asociado de forma natural con la proteína a expresar o cualquier segmento señal heterólogo capaz de dirigir la secreción de la proteína de acuerdo con la presente invención. En otra realización, una molécula recombinante de la presente invención comprende una secuencia líder para posibilitar que una proteína expresada se suministre e inserte en la membrana de una célula hospedadora. Las secuencias líder adecuadas incluyen una secuencia líder que está asociada de forma natural con la proteína, o cualquier secuencia líder heteróloga capaz de dirigir el suministro e inserción de la proteína a la membrana de una célula.

Los presentes inventores han descubierto que las Orf A y B de PUFA PKS de *Schizochytrium* están estrechamente vinculadas en el genoma y se secuenció la región entre las Orf. Las Orf están orientadas en direcciones opuestas y 4244 pares de bases separan los codones de inicio (ATG) (es decir, están dispuestas del siguiente modo: 3'OrfA5' - 4244 pb - 5'OrfB3'). El examen de la región intergénica de 4244 pb no reveló ninguna Orf obvia (no

se encontraron coincidencias significativas en una búsqueda BlastX). Ambas Orf A y B se expresan altamente en *Schizochytrium*, al menos durante el tiempo de producción de aceite, lo que implica que los elementos promotores activos están incluidos en esta región intergénica. Se cree que estos elementos genéticos tienen utilidad como secuencia promotora bidireccional para aplicaciones transgénicas. Por ejemplo, en una realización preferida, podría clonarse esta región, colocar cualquier gen de interés en cada extremo e introducir la construcción en *Schizochytrium* (o algún otro hospedador en que los promotores puedan mostrar que funcionan). Se predice que los elementos reguladores, en las condiciones apropiadas, proporcionarían un alto nivel de expresión coordinado de los dos genes introducidos. La secuencia completa de nucleótidos para la región reguladora que contiene elementos reguladores de PUFA PKS de *Schizochytrium* (por ejemplo, un promotor) están representados en este documento como SEC ID N° 36.

De un modo similar, la OrfC se expresa altamente en *Schizochytrium* durante el tiempo de producción de aceite y se espera que los elementos reguladores residan en la región cadena arriba de su codón de inicio. Una región de ADN genómico cadena arriba de OrfC se ha clonado y secuenciado y está representada en este documento como (SEC ID N° 37). Esta secuencia contiene los 3886 nt inmediatamente cadena arriba del codón de inicio de OrfC. El examen de esta región no reveló ninguna Orf obvia (es decir, no se encontraron coincidencias significativa en una búsqueda BlastX). Se cree que los elementos reguladores contenidos en esta región, en las condiciones apropiadas, proporcionarán expresión de alto nivel de un gen colocado detrás de ellos. Adicionalmente, en las condiciones apropiadas, el nivel de expresión puede coordinarse con genes bajo el control de la región intergénica A-B (SEC ID N° 36).

Por lo tanto, en una realización, una molécula de ácido nucleico recombinante útil en la presente invención, como se describe en este documento, puede incluir una región reguladora de PUFA PKS contenida en la SEC ID N° 36 y/o SEC ID N° 37. Dicha región reguladora puede incluir cualquier parte (fragmento) de la SEC ID N° 36 y/o SEC ID N° 37 que tiene al menos actividad transcripcional basal de PUFA PKS.

Una o más moléculas recombinantes de la presente invención pueden usarse para producir un producto codificado (por ejemplo, un dominio, proteína, o sistema PUFA PKS) de la presente invención. En una realización, un producto codificado se produce expresando una molécula de ácido nucleico como se describe en este documento en condiciones eficaces para producir la proteína. Un método preferido para producir una proteína codificada es transfectar una célula hospedadora con una o más moléculas recombinantes para formar una célula recombinante. Las células hospedadoras adecuadas a transfectar incluyen, aunque sin limitación, cualquier célula bacteriana, fúngica (por ejemplo, levadura), de insecto, vegetal o animal que pueda transfectarse. Las células hospedadoras pueden ser células sin transfectar o células que ya se han transfectado con al menos otra molécula de ácido nucleico recombinante.

De acuerdo con la presente invención, el término "transfección" se usa para hacer referencia a cualquier método por el cual puede insertarse una molécula exógena de ácido nucleico (es decir, una molécula de ácido nucleico recombinante) en una célula. El término "transformación" puede usarse de forma intercambiable con el término "transfección" cuando dicho término se usa para hacer referencia a la introducción de moléculas de ácido nucleico en células microbianas, tales como algas, bacterias y levaduras. En sistemas microbianos, el término "transformación" se usa para describir un cambio hereditario debido a la adquisición de ácidos nucleicos exógenos por el microorganismo y es esencialmente sinónimo al término "transfección". Sin embargo, en células animales, la transformación ha adquirido un segundo significado que puede hacer referencia a cambios en las propiedades de crecimiento de las células en cultivo después de que se conviertan en cancerosas, por ejemplo. Por lo tanto, para evitar confusiones, el término "transfección" se usa preferiblemente con respecto a la introducción de ácidos nucleicos exógenos en células animales, y el término "transfección" se usará en este documento para abarcar en líneas generales transfección de células animales, células vegetales y transformación de células microbianas, en la medida en que los términos pertenecen a la introducción de ácidos nucleicos exógenos en una célula. Por lo tanto, las técnicas de transfección incluyen, aunque sin limitación, transformación, bombardeo de partículas, electroporación, microinyección, lipofección, adsorción, infección y fusión de protoplastos.

Los expertos en la materia apreciarán que el uso de tecnologías de ADN recombinante puede mejorar el control de la expresión de moléculas de ácido nucleico transfectadas manipulando, por ejemplo, la cantidad de copias de las moléculas de ácido nucleico dentro de la célula hospedadora, la eficacia con que se transcriben esas moléculas de ácido nucleico, la eficacia con que se traducen los transcritos resultantes, y la eficacia de modificaciones post-traduccionales. Adicionalmente, la secuencia promotora podría modificarse por ingeniería genética para mejorar el nivel de expresión en comparación con el promotor nativo. Las técnicas recombinantes útiles para controlar la expresión de moléculas de ácido nucleico incluyen, aunque sin limitación, integración de las moléculas de ácido nucleico en uno o más cromosomas de la célula hospedadora, la adición de secuencias de estabilidad del vector a los plásmidos, sustituciones o modificaciones de las señales de control de la transcripción (por ejemplo, promotores, operadores, potenciadores), sustituciones o modificaciones de las señales de control de la traducción (por ejemplo, sitios de unión al ribosoma, secuencias de Shine-Dalgarno), modificación de moléculas de ácido nucleico para que correspondan al uso de codones de la célula hospedadora, y delección de secuencias que desestabilizan los transcritos.

El análisis general anterior con respecto a moléculas de ácido nucleico recombinantes y transfección de células hospedadoras pretende tener aplicación a cualquier molécula de ácido nucleico recombinante analizada en este documento, incluyendo aquellas que codifican cualquier secuencia de aminoácidos que tenga una actividad biológica de al menos un dominio de una PUFA PKS, aquellas que codifican secuencias de aminoácidos de otros sistemas PKS, y aquellas que codifican otras proteínas o dominios.

Describimos un nuevo método para identificar un microorganismo que tenga un sistema PUFA PKS que sea homólogo en estructura, organización de dominios y/o función a un sistema PUFA PKS de *Schizochytrium*. El microorganismo puede ser un microorganismo no bacteriano, y el microorganismo identificado por este método puede ser un microorganismo eucariota. Además, describimos los microorganismos identificados por dicho método y el uso de estos microorganismos y el sistema PUFA PKS de estos microorganismos en las diversas aplicaciones para un sistema PUFA PKS (por ejemplo, organismos modificados genéticamente y métodos para producir moléculas bioactivas). El método de selección único descrito y demostrado en este documento posibilita la rápida identificación de nuevas cepas microbianas que contienen un sistema PUFA PKS homólogo al sistema PUFA PKS de *Schizochytrium* de la presente invención. Los solicitantes han usado este método para descubrir y describir en este documento que un microorganismo *Thraustochytrium* contiene un sistema PUFA PKS que es homólogo al encontrado en *Schizochytrium*. Este descubrimiento se describe en detalle en el Ejemplo 2 a continuación.

Los organismos microbianos con un sistema PUFA PKS similar al encontrado en *Schizochytrium*, tal como el microorganismo *Thraustochytrium* descubierto por los presentes inventores y descrito en el Ejemplo 2, puede identificarse/aislarse/seleccionarse fácilmente por los siguientes métodos usados por separado o en cualquier combinación de estos métodos.

En general, el método para identificar un microorganismo no bacteriano que tiene un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) incluye una primera etapa de (a) seleccionar un microorganismo que produce al menos un PUFA; y una segunda etapa de (b) identificar un microorganismo de (a) que tiene una capacidad de producir PUFA aumentados en condiciones de oxígeno disuelto de menos de aproximadamente el 5 % de saturación en el medio de fermentación, en comparación con la producción de PUFA mediante dicho organismo en condiciones de oxígeno disuelto de más del 5 % de saturación, más preferiblemente 10 % de saturación, más preferiblemente más del 15 % de saturación y más preferiblemente más del 20 % de saturación en el medio de fermentación. Un microorganismo que produce al menos un PUFA y tiene una capacidad de producir PUFA aumentados en condiciones de oxígeno disuelto de menos de aproximadamente el 5 % de saturación se identifica como candidato para contener un sistema PUFA PKS. Posterior a la identificación de un microorganismo que es un fuerte candidato para que contenga un sistema PUFA PKS, el método puede incluir una etapa adicional (c) de detectar si el organismo identificado en la etapa (b) comprende un sistema PUFA PKS.

La etapa (b) puede realizarse cultivando el microorganismo seleccionado para el proceso de detección en condiciones de bajo oxígeno/anóxicas y condiciones aeróbicas y, además de medir el contenido de PUFA en el organismo, se determina el perfil de ácidos grasos, así como el contenido de grasas. Comparando los resultados en las condiciones de bajo oxígeno/anóxicas con los resultados en condiciones aeróbicas, el método proporciona un fuerte indicio de si el microorganismo de ensayo contiene un sistema PUFA PKS de la presente invención. Esta realización se describe en detalle a continuación.

Inicialmente, las cepas microbianas a examinar para la presencia de un sistema PUFA PKS se cultivan en condiciones aeróbicas, para inducir la producción de una gran cantidad de células (biomasa bacteriana). Como un elemento del proceso de identificación, estas células después se colocan en condiciones de cultivo de bajo oxígeno o anóxicas (por ejemplo, oxígeno disuelto de menos de aproximadamente el 5 % de saturación, más preferiblemente menos de aproximadamente el 2 %, incluso más preferiblemente menos de aproximadamente el 1 %, y mucho más preferiblemente oxígeno disuelto de aproximadamente el 0 % de saturación en el medio de cultivo) y se dejaron crecer durante aproximadamente otras 24-72 horas. En este proceso, los microorganismos deben cultivarse a una temperatura mayor de aproximadamente 15 °C, y más preferiblemente mayor de aproximadamente 20 °C, e incluso más preferiblemente mayor de aproximadamente 25 °C, e incluso más preferiblemente mayor de 30 °C. El entorno de cultivo bajo o anóxico puede mantenerse fácilmente en cámaras de cultivo capaces de inducir este tipo de entorno atmosférico en la cámara (y por tanto en los cultivos) o cultivando las células de un modo que induzca el entorno de bajo oxígeno directamente en el propio matraz/recipiente de cultivo.

En un método de cultivo, los microbios pueden cultivarse en matraces de agitación que, en lugar de contener normalmente una pequeña cantidad de medio de cultivo - menos de aproximadamente el 50 % de la capacidad total y habitualmente menos de aproximadamente el 25 % de la capacidad total - para mantener el medio aireado según se agita en una mesa oscilatoria, se llenan en su lugar hasta más de aproximadamente el 50 % de su capacidad, y más preferiblemente más del aproximadamente el 60 %, y mucho más preferiblemente más de aproximadamente el 75 % de su capacidad con medio de cultivo. La alta carga del matraz de agitación con medio de cultivo evita que se mezcle muy bien en el matraz cuando se coloca en una mesa oscilatoria, evitando la difusión de oxígeno en el cultivo. Por lo tanto, según crecen los microbios, usan el oxígeno existente en el medio y naturalmente crean un entorno bajo o ausente en oxígeno en el matraz de agitación.

Después del periodo de cultivo, las células se recogen y analizan para el contenido de compuestos bioactivos de interés (por ejemplo, lípidos), pero más particularmente, para compuestos que contengan dos o más enlaces insaturados, y más preferiblemente tres o más dobles enlaces, e incluso más preferiblemente cuatro o más dobles enlaces. Para lípidos, estas cepas que poseen dichos compuestos de más de aproximadamente el 5 %, y más preferiblemente más de aproximadamente el 10 %, y más preferiblemente más de aproximadamente el 15 %, e incluso más preferiblemente más de aproximadamente el 20 % del peso seco del microorganismo se identifican como que contienen de forma predecible un nuevo sistema PKS del tipo descrito anteriormente. Para otros compuestos bioactivos, tales como antibióticos o compuestos que se sintetizan en cantidades más pequeñas, esas cepas que poseen dichos compuestos a más de aproximadamente el 0,5 %, y más preferiblemente más de aproximadamente el 0,1 %, y más preferiblemente más de aproximadamente el 0,25 %, y más preferiblemente más de aproximadamente el 0,5 %, y más preferiblemente más de aproximadamente el 0,75 %, y más preferiblemente más de aproximadamente el 1 %, y más preferiblemente más de aproximadamente el 2,5 %, y más preferiblemente más de aproximadamente el 5 % del peso seco del microorganismo se identifican como que contienen de forma predecible un nuevo sistema PKS del tipo descrito anteriormente.

Como alternativa, o junto con este método, pueden identificarse cepas microbianas potenciales que contienen nuevos sistemas PUFA PKS como se describe en este documento examinando el perfil de ácidos grasos de la cepa (obtenida por cultivo del organismo o a través de fuentes publicadas u otras fuentes fácilmente disponibles). Si el microbio contiene más de aproximadamente el 30 %, y más preferiblemente más de aproximadamente el 40 %, y más preferiblemente más de aproximadamente el 45 %, e incluso y más preferiblemente más de aproximadamente el 50 %, de sus ácidos grasos totales como C14:0, C16:0 y/o C16:1, aunque produciendo también al menos un ácido graso de cadena larga con tres o más enlaces insaturados, y más preferiblemente 4 o más dobles enlaces, y más preferiblemente 5 o más dobles enlaces, e incluso más preferiblemente 6 o más dobles enlaces, entonces esta cepa microbiana se identifica como un candidato probable a poseer un nuevo sistema PUFA PKS del tipo descrito en esta invención. La selección de este organismo en las condiciones de bajo oxígeno descritas anteriormente, y la confirmación de la producción de moléculas bioactivas que contienen dos o más enlaces insaturados sugeriría la existencia de un nuevo sistema PUFA PKS en el organismo, que podría confirmarse adicionalmente por análisis del genoma de los microbios.

El éxito de este método también puede potenciarse seleccionando cepas eucariotas que se sabe que contienen ácidos grasos C17:0 y o C17:1 (junto con los grandes porcentajes de ácidos grasos C14:0, C16:0 y C16:1 descritos anteriormente) - porque los ácidos grasos C17:0 y C17:1 son marcadores potenciales para un sistema de producción de ácidos grasos basado o influenciado por bacterias (procariotas). Otro marcador para identificar cepas que contienen nuevos sistemas PUFA PKS es la producción de perfiles simples de ácidos grasos por el organismo. De acuerdo con la presente invención, un "perfil simple de ácidos grasos" se define como 8 o menos ácidos grasos que se producen por la cepa a niveles mayores del 10 % de los ácidos grasos totales.

El uso de cualquiera de estos métodos o marcadores (individualmente o preferiblemente en combinación) posibilitaría a un experto en la materia identificar fácilmente cepas microbianas con alta predicción de contener un nuevo sistema PUFA PKS del tipo descrito en esta invención.

Combinando muchos de los métodos y marcadores descritos anteriormente, se ha desarrollado un nuevo filtro biorracional (usando cultivos en matraz de agitación) para detectar microorganismos que contienen sistemas PKS productores de PUFA. Este sistema de selección se realiza del siguiente modo:

Una parte de un cultivo de la cepa/microorganismo a ensayar se coloca en matraz de agitación con separadores de 250 ml con 500 ml de medio de cultivo (tratamiento aeróbico), y otra parte de cultivo de la misma cepa se coloca en un matraz de agitación sin separadores de 250 ml con 250 ml de medio de cultivo (tratamiento anóxico/de bajo oxígeno). Pueden emplearse diversos medios de cultivo dependiente del tipo y cepa de microorganismo que se está evaluando. Ambos matraces se colocan en una mesa oscilatoria a 200 rpm. Después de 48-72 horas de tiempo de cultivo, los cultivos se recogen por centrifugación y las células se analizan para el contenido de éster metílico de ácido graso mediante cromatografía de gases para determinar los siguientes datos para cada cultivo: (1) perfil de ácidos grasos; (2) contenido de PUFA; y (3) contenido de grasas (aproximadamente como cantidad total de ácidos grasos/peso seco de la célula).

Estos datos después se analizan haciendo las siguientes cinco preguntas (Sí/No):

Comparación de los datos del matraz de bajo O<sub>2</sub>/anóxico con los datos del matraz aeróbico:

(1) ¿El DHA (u otro contenido de PUFA) (como % FAME (ésteres metílicos de ácido graso)) permanece aproximadamente igual o preferiblemente aumentado en el cultivo de bajo oxígeno en comparación con el cultivo aeróbico?

(2) ¿Es C14:0 + C16:0 + C16:1 mayor de aproximadamente el 40 % de TFA en el cultivo anóxico?

(3) ¿Existen muy pocos (<1 % como FAME) o ningún precursor (C18:3n-3 + C18:2n-6+C18:3n-6) para la ruta convencional de elongasa/desaturasa dependiente de oxígeno en el cultivo anóxico?

(4) ¿Aumentó el contenido de grasas (como cantidad total de ácidos grasos/peso seco de la célula) en el cultivo de bajo oxígeno en comparación con el cultivo aeróbico?

(5) ¿Aumentó DHA (u otro contenido de PUFA) como % de peso seco de la célula en el cultivo de bajo oxígeno en comparación con el cultivo aeróbico?

5 Si las tres primeras preguntas se responden sí, esta es un buen indicio de que la cepa contiene un sistema genético PKS para fabricar PUFA de cadena larga. Cuantas más preguntas se contesten sí (preferiblemente las tres primeras preguntas deben contestarse sí), más fuerte es el indicio de que la cepa contiene dicho sistema genético PKS. Si las cinco preguntas se contestan sí, entonces hay un indicio muy fuerte de que la cepa contiene un sistema genético PKS para fabricar PUFA de cadena larga. La ausencia de 18:3n-3/18:2n-6/18:3n-6 indicaría que las condiciones de bajo oxígeno habrían inactivado o inhibido la ruta convencional para la síntesis de PUFA. Un ácido graso alto 14:0/16:0/16:1 es un indicador preliminar de un perfil de síntesis de ácidos grasos influenciado de forma bacteriana (la presencia de C17:0 y 17:1 también es un indicador de esto) y de un perfil simple de ácidos grasos. La síntesis aumentada de PUFA y la síntesis de grasas que contienen PUFA en las condiciones de bajo oxígeno son directamente indicativas de un sistema PUFA PKS, ya que este sistema no requiere oxígeno para fabricar ácidos grasos altamente insaturados.

Finalmente, en el método de identificación, una vez identificado un fuerte candidato, el microbio se selecciona preferiblemente para detectar si el microbio contiene o no un sistema PUFA PKS. Por ejemplo, puede seleccionarse el genoma del microbio para detectar la presencia de una o más secuencias de ácido nucleico que codifiquen un dominio de un sistema PUFA PKS como se describe en este documento. Preferiblemente, esta etapa de detección incluye un método adecuado de detección de ácido nucleico, tal como hibridación, amplificación y/o secuenciación de una o más secuencias de ácido nucleico en el microbio de interés. Las sondas y/o cebadores usados en los métodos de detección pueden derivarse de cualquier sistema PUFA PKS conocido, incluyendo los sistemas PUFA PKS de bacterias marinas descritos en la patente de Estados Unidos n.º 6.140.486 o los sistemas PUFA PKS de traustoquitridio descritos en la solicitud de Estados Unidos n.º de serie 09/231.899 y en este documento. Una vez se han identificado nuevos sistemas PUFA PKS, también puede usarse el material genético de estos sistemas para detectar nuevos sistemas PUFA PKS adicionales. Los métodos de hibridación, amplificación y secuenciación de ácidos nucleicos con el fin de identificar y detectar una secuencia son bien conocidos en la técnica. Usando estos métodos de detección, puede evaluarse la homología de secuencia y estructura de dominios, (por ejemplo, la presencia, cantidad y/o disposición de diversos dominios funcional de PUFA PKS) y compararse con los sistemas PUFA PKS conocidos descritos en este documento.

Puede identificarse un sistema PUFA PKS usando ensayos biológicos. Por ejemplo, en la solicitud de Estados Unidos n.º de serie 09/231.899, Ejemplo 7, se describen los resultados de un experimento clave usando un inhibidor bien conocido de algunos tipos de sistemas de síntesis de ácidos grasos, es decir, tiolactomicina. Los inventores demostraron que la síntesis de PUFA en células completas de *Schizochytrium* podría bloquearse específicamente sin bloquear la síntesis de ácidos grasos saturados de cadena corta. El significado de ese resultado es el siguiente: los inventores conocían a partir del análisis de las secuencias de ADNc de *Schizochytrium* que está presente un sistema de ácido grado sintasa Tipo I en *Schizochytrium*. Se sabía que la tiolactomicina no inhibe sistemas FAS Tipo I, y esto es coherente con los datos de los inventores - es decir, la producción de los ácidos grasos saturados (principalmente C 14:0 y C16:0 en *Schizochytrium*) no se inhibía por el tratamiento con tiolactomicina. No hay indicaciones en la bibliografía o en los propios datos de los inventores de que la tiolactomicina tengan ningún inhibidor sobre la elongación de ácidos grasos C14:0 o C16:0 o su desaturación (es decir, la conversión de ácidos grasos saturados de cadena corta, en PUFA mediante la ruta clásica). Por lo tanto, el hecho de que la producción de PUFA en *Schizochytrium* se bloqueara por tiolactomicina indica fuertemente que la ruta clásica de síntesis de PUFA no producen los PUFA en + *Schizochytrium*, sino que en su lugar está implicada una ruta diferente de síntesis. Además, se había determinado previamente que el sistema PUFA PKS de *Shewanella* se inhibe por tiolactomicina (obsérvese que el sistema PUFA PKS de la presente invención tiene elementos de sistemas tanto Tipo I como Tipo II), y se sabía que la tiolactomicina es un inhibidor de los sistemas FAS Tipo II (tal como el encontrado en *E. coli*). Por lo tanto, este experimento indicó que *Schizochytrium* producía PUFA como resultado de una ruta que no implica FAST Tipo I. Podría usarse un fundamento y etapa de detección similares para detectar un sistema PUFA PKS en un microbio identificado usando el nuevo método de selección descrito en este documento.

Además, el Ejemplo 3 muestra datos bioquímicos adicionales que proporcionan evidencias de que no se producen PUFA en *Schizochytrium* mediante la ruta clásica (es decir, no se observa cinética de producto precursor entre C16:0 y DHA en células completas y, puede separarse la síntesis de PUFA *in vitro* de la fracción de membrana - todas las ácido graso desaturadas de la ruta clásica de síntesis de PUFA, con la excepción de la delta 9 desaturasa que inserta el primer doble enlace de la serie, están asociadas con membranas celulares). Este tipo de datos bioquímicos podría usarse para detectar actividad PUFA PKS en microbios identificados por el nuevo método de selección descrito anteriormente.

Las cepas microbianas a seleccionar usando el método de selección/identificación de la presente invención se eligen entre el grupo que consiste en: bacterias, algas, hongos, protozoos o protistas, pero muy preferiblemente entre los microbios eucariotas que consisten en algas, hongos, protozoos y protistas. Estos microbios son preferiblemente capaces de crecer y producir los compuestos bioactivos que contienen dos o más enlaces insaturados a

temperaturas mayores de aproximadamente 15 °C, más preferiblemente mayores de aproximadamente 20 °C e incluso más preferiblemente mayores de aproximadamente 25 °C y mucho más preferiblemente mayores de aproximadamente 30 °C.

5 Pueden identificarse nuevos sistemas PUFA PKS bacterianos en bacterias que producen PUFA a temperaturas que exceden aproximadamente 20 °C, preferiblemente que exceden aproximadamente 25 °C e incluso más preferiblemente que exceden aproximadamente 30 °C. Como se ha descrito previamente en este documento, las bacterias marinas *Shewanella* y *Vibrio marinus*, descritas en la patente de Estados Unidos n.º 6.140.486, no producen PUFA a temperaturas mayores, lo que limita la utilidad de sistemas PUFA PKS derivados de estas bacterias, particularmente en aplicaciones en plantas en condiciones de campo. Por lo tanto, en una realización, el método de selección de la presente invención puede usarse para identificar bacterias que tengan un sistema PUGA PKS que sea capaz de crecer y producir PUFA a temperaturas mayores (por ejemplo, por encima de aproximadamente 20, 25 o 30 °C). Pueden añadirse inhibidores de crecimiento eucariota tales como nistatina (antifúngico) o cicloheximida (inhibidor de la síntesis de proteínas eucariotas) a placas de agar usadas para cultivar/seleccionar cepas iniciales de muestras de agua/muestras de suelo recogida de los tipos de hábitats/nichos descritos a continuación. Este proceso podría ayudar a seleccionar el enriquecimiento de cepas bacterianas sin (o mínima) contaminación de cepas eucariotas. Este proceso de selección, en combinación con el cultivo de las placas a elevadas temperaturas (por ejemplo 30 °C), y después la selección de las cepas que producen al menos un PUFA identificaría inicialmente cepas bacterianas candidatas con un sistema PUFA PKS que es funcional a elevadas temperaturas (en oposición a esas cepas bacterianas de la técnica anterior que solamente muestran producción de PUFA a temperaturas menores de aproximadamente 20 °C y más preferiblemente por debajo de aproximadamente 5 °C).

25 Las localizaciones para la recogida de los tipos preferidos de microbios para la selección de un sistema PUFA PKS de acuerdo con la presente invención incluyen cualquiera de las siguientes: entornos de bajo oxígeno (o localizaciones cerca de estos tipos de entornos de bajo oxígeno incluyendo en los intestinos de animales incluyendo invertebrados que consumen microbios o alimentos que contienen microbios (incluyendo tipos de organismos de alimentación por filtración), hábitats acuáticos que contienen bajo oxígeno o nada de oxígeno (incluyendo de agua dulce, salinos y marinos), y especialmente en o cerca de entornos de bajo oxígeno (regiones) en los océanos. Las cepas microbianas preferiblemente no serían anaerobios obligados sino que estarían adaptados a vivir en entornos tanto aeróbicos como bajos o anóxicos. Los entornos de suelo que contienen entornos tanto anaeróbicos como de bajo oxígeno o anóxicos también serían entornos excelentes para encontrar estos organismos y especialmente en estos tipos de suelo en hábitats acuáticos o hábitats acuáticos temporales.

35 Una cepa microbiana particularmente preferida sería una cepa (seleccionada entre el grupo que consiste en algas, hongos (incluyendo levaduras), protozoos o protistas) que, durante una parte de su ciclo vital, es capaz de consumir células bacterianas completas (bacterívoro) por mecanismos tales como fagocitosis, capacidad fagocítica o endocítica y/o tiene una fase de su ciclo vital en que existe como fase ameboide o protoplasto desnudo. Este método de nutrición aumentaría enormemente el potencial de transferencia de un sistema PKS bacteriano a una célula eucariota si sucediera un error y la célula bacteriana (o su ADN) no se digiriera y en su lugar se incorporara de forma funcional a la célula eucariota.

45 Las cepas de microbios (diferentes de los miembros de los traustocitridios) con capacidad de bacterívoro (especialmente por fagocitosis o endocitosis) pueden encontrarse en las siguientes clases microbianas (incluyendo aunque sin limitación los géneros ejemplares):

50 En las algas y microbios tipo algas (incluyendo estramenópilos): de la clase Euglenophyceae (por ejemplo, géneros *Euglena*, y *Peranema*), la clase Chrysophyceae (por ejemplo, el género *Ochromonas*), la clase Dinobryaceae (por ejemplo, los géneros *Dinobryon*, *Platychrysis*, y *Chrysochromulina*), la Dinophyceae (incluyendo los géneros *Cryptothecodinium*, *Gymnodinium*, *Peridinium*, *Ceratium*, *Gyrodinium*, y *Oxyrrhis*), la clase Cryptophyceae (por ejemplo, los géneros *Cryptomonas*, y *Rhodomonas*), la clase Xanthophyceae (por ejemplo, el género *Olisthodiscus*) (e incluyendo formas de algas en que existe una fase ameboide como en los flagelados Rhizochloridaceae, y zoosporas/gametos de *Aphanochaete pascheri*, *Bumilleria stigeoclonium* y *Vaucheria geminata*), la clase Eustigmatophyceae, y la clase Prymnesiopyceae (incluyendo los géneros *Prymnesium* y *Diacronema*).

55 En los estramenópilos incluyendo los: Proteromonas, Opalinas, Developayella, Diplopterys, Labirintúlidos, Traustocitridios, Bicosoecidas, Oomicetes, Hipoquitridiomicetes, Commation, Reticulosphaera, Pelagomonas, Pelapococcus, Ollicola, Aureococcus, Parmales, Raphidiofitos, Synuridos, Rhizochromulinales, Pedinellales, Dictyochales, Chrysomeridales, Sarcinochrysidales, Hydrurales, Hibberdiales, y Chromulinales.

60 En los Hongos: clase Mixomicetes (forma mixameba) - moho mucilaginoso, clase Acrasieae incluyendo los órdenes Acrasiceae (por ejemplo, el género *Sappinia*), clase Guttulinaceae (por ejemplo, los géneros *Guttulinopsis*, y *Guttulina*), clase Dictysteliaceae (por ejemplo, los géneros *Acrasis*, *Dictyostelium*, *Polysphondylium*, y *Coenonia*), y clase Phycomyces incluyendo los órdenes Chytridiales, Ancylistales, Blastocladiales, Monoblepharidales, Saprolegniales, Peronosporales, Mucorales, y Entomophthorales.

65



En los Protozoos: cepas de Protozoos con fases vitales con capacidad de bacterívoro (incluyendo por fagocitosis) pueden seleccionarse entre los tipos clasificados como ciliados, flagelados o amebas. Protozoos ciliados incluyen los grupos: Chonotriches, Colpodides, Cyrtophores, Haptorides, Karyorelictos, Oligohymenophora, Polyhymenophora (espirotricos), Prostomes y Suctoria. Protozoos flagelados incluyen los Biosoecidas, Bodonidas, Cercomonadas, Crisofitos (por ejemplo, los géneros Anthophysa, Chrysamoemba, Chrysosphaerella, Dendromonas, Dinobryon, Mallomonas, Ochromonas, Paraphysomonas, Poterioochromonas, Spumella, Syncrypta, Synura, y Uroglena), flagelados Collar, Criptófitos (por ejemplo, los géneros Chilomonas, Cryptomonas, Cyanomonas, y Goniomonas), Dinoflagelados, Diplomonados, Euglenoides, Heterolobosea, Pedinellidos, Pelobiontes, Phalansteriidos, Pseudodendromonadas, Spongomonadas y Volvocales (y otros flagelados incluyendo los géneros de flagelados no asignados de Artodiscus, Clautriavia, Helkesimastix, Kathablepharis y Multicilia). Protozoos ameboides incluyen los grupos: Actinophryidos, Centrohelidos, Desmothoricos, Diplophryidos, Eumamoebae, Heterolobosea, Leptomyxidos, ameba filosa Nucleariid, Peleobiontes, amebas Testate y Vampyrellidos (e incluyendo los géneros ameboides no asignados Gymnophrys, Biomyxa, Microcometes, Reticulomyxa, Belonocystis, Elaeorhanis, Allelogromia, Gromia o Lieberkuhnia). Los órdenes de protozoos incluyen los siguientes: Percolomonadeae, Heterolobosea, Lyromonadea, Pseudociliata, Trichomonadea, Hypermastigea, Heteromiteae, Telonemea, Cyathobodonea, Ebridea, Ppytomyxea, Opalineae, Kinetomonadea, Hemimastigea, Protostelea, Myxagastrea, Dictyostelea, Choanomonadea, Apicomonadea, Eogregarinea, Neogregarinea, Coelotrophaea, Eucoccidea, Haemosporea, Piroplasmae, Spirotrichea, Prostomatea, Litostomatea, Phyllopharyngea, Nassophorea, Oligohymenophorea, Colpodea, Karyorelicta, Nucleohelea, Centrohelea, Acantharea, Sticholonchea, Polycystinea, Phaeodarea, Lobosea, Filosea, Athalamea, Monothalamea, Polythalamea, Xenophyophorea, Schizocladea, Holosea, Entamoebae, Myxosporea, Actinomyxea, Halosporea, Paramyxea, Rhombozoa y Orthonectea.

Describimos cepas de los microorganismos enumerados anteriormente que se han recogido de uno de los hábitats preferidos enumerados anteriormente.

Describimos microorganismos identificados usando el nuevo método de detección de PUFA PKS descrito anteriormente, los genes PUFA PKS y proteínas codificados por el mismo, y el uso de dichos microorganismos y/o de genes y proteínas de PUFA PKS (incluyendo homólogos y fragmentos de los mismos) en cualquiera de los métodos descritos en este documento. En particular, se identificaron organismos por el método de selección de la presente invención que después se modificaron genéticamente para regular la producción de moléculas bioactivas por dicho sistema PUFA PKS.

La presente invención se refiere a una molécula de ácido nucleico aislada que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo o fragmento biológicamente activo del mismo de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) de un microorganismo traustroquitridio. Como se ha analizado anteriormente, los presentes inventores han usado de forma satisfactoria el método para identificar un microorganismo no bacteriano que tenga un sistema PUFA PKS para identificar miembros adicionales del orden Traustroquitriales que contengan un sistema PUFA PKS. La identificación de tres de estos microorganismos se describe en el Ejemplo 2. Específicamente, los presentes inventores han usado el método de selección de la presente invención para identificar *Thraustochytrium* sp. 23B (ATCC 20892) como con alta predicción de contener un sistema PUFA PKS, seguido por detección de secuencias en el genoma de *Thraustochytrium* sp. 23B que hibrida con los genes de PUFA PKS de *Schizochytrium* descritos en este documento. También se han identificado *Schizochytrium limacium* (IFO 32693) y *Ulkenia* (BP-5601) como buenos candidatos para contener sistemas PUFA PKS. En base a estos datos y a las similitudes entre miembros del orden Traustroquitriales, se cree que ahora pueden identificarse fácilmente muchos otros sistemas PUFA PKS de Traustroquitriales usando los métodos y herramientas proporcionados por la presente invención. Por lo tanto, los sistemas PUFA PKS de Traustroquitriales y partes y/u homólogos de los mismos (por ejemplo, proteínas, dominios y fragmentos de los mismos), organismos modificados genéticamente que comprenden dichos sistemas y partes y/u homólogos de los mismos y métodos para usar dichos microorganismos y sistemas PUFA PKS, están abarcados por la presente invención.

Los avances han provocado la revisión de la taxonomía de los traustroquitridios. Los teóricos taxonómicos colocan a los traustroquitridios con las algas o protistas tipo algas. Sin embargo, a causa de la inexactitud taxonómica, sería mejor para los propósitos de la presente invención considerar las cepas descritas en la presente invención como traustroquitridios (Orden: Traustroquitriales; Familia: Thraustochytriaceae; Género: *Thraustochytrium*, *Schizochytrium*, *Labyrinthuloides*, o *Japonochytrium*). Para la presente invención, los miembros de los labirintúlidos se consideran incluidos en los traustroquitridios. Los cambios taxonómicos se resumen a continuación. Las cepas de ciertos microorganismos unicelulares descritos en este documento son miembros del orden Traustroquitriales. Los traustroquitridios son eucariotas marinos con una historia taxonómica evolutiva. Los problemas con la colocación taxonómica de los traustroquitridios se han revisado por Moss (1986), Bahnweb y Jackle (1986) y Chamberlain y Moss (1988). De acuerdo con la presente invención, las expresiones "traustroquitridio", "microorganismo de Traustroquitriales" y "microorganismo del orden Traustroquitriales" pueden usarse de forma intercambiable.

Para propósitos de conveniencia, los taxónomos colocaron en primer lugar los traustroquitridios con otros eucariotas zoospóricos incoloros en los Ficomícetes (hongos tipo alga). El nombre Ficomícetes, sin embargo, se bajó finalmente de estatus taxonómico, y los traustroquitridios se retuvieron en los Oomicetes (los hongos zoospóricos biflagelados). Inicialmente se asumió que los Oomicetes estaban relacionados con las algas Heterokonta, y finalmente un amplio

intervalo de estudios de ultra-estructura y bioquímicos, resumidos por Barr (Barr, 1981, Biosystems 14:359-370) apoyaron esta suposición. Los Oomicetes estaban de hecho aceptados por Leedale (Leedale, 1974, Taxon 23:261-270) y otros ficólogos como parte de las algas Heterokonta. Sin embargo, por cuestión de conveniencia resultante de su naturaleza heterotrófica, los Oomicetes y los traustocitridios se han estudiado en gran medida por los micólogos (científicos que estudian los hongos) en lugar de los ficólogos (científicos que estudian las algas).

Desde otra perspectiva taxonómica, los biólogos evolutivos han desarrollado dos escuelas de pensamiento generales del modo en que evolucionan los eucariotas. Una teoría propone un origen exógeno de orgánulos unidos a membrana a través de una serie de endosimbiosis (Margulis, 1970, Origin of Eukaryotic Cells. Yale University Press, New Haven); por ejemplo, las mitocondrias se obtuvieron de endosimbiosis bacteriana, los cloroplastos de cianofitos, y los flagelos de espiroquetas. La otra teoría sugiere una evolución gradual de los orgánulos unidos a membrana a partir de los sistemas no unidos a membrana del ancestro procariota mediante un proceso autógeno (Cavalier-Smith, 1975, Nature (Lond.) 256:462-468). Sin embargo, ambos grupos de biólogos evolutivos han retirado los Oomicetes y los traustocitridios de los hongos y los han colocado con las algas cromofitas en el reino Chromophyta (Cavalier-Smith, 1981, BioSystems 14:461-481) (este reino se ha expandido más recientemente para incluir otros protistas y los miembros de este reino se llaman ahora estramenópilos) o con todas las algas en el reino Protoctista (Margulis y Sagen, 1985, Biosystems 18:141-147).

Con el desarrollo de microscopía electrónica, estudios sobre la ultra-estructura de las zoosporas de dos géneros de traustocitridios, *Thraustochytrium* y *Schizochytrium*, (Perkins, 1976, pág. 279-312 en "Recent Advances in Aquatic Mycology" (ed. E.B.G. Jones), John Wiley y Sons, Nueva York; Kazama, 1980, Can. J. Bot. 58:2434-2446; Barr, 1981, Biosystems 14:359-370) han proporcionado buenas evidencias de que Thraustochytriaceae están relacionados solamente de forma distante con los Oomicetes. Adicionalmente, los datos genéticos que representan una análisis de correspondencia (una forma de estadística multivariada) de secuencias de ARN ribosómico 5 S indican que los Traustocitriales son claramente un grupo único de eucariotas, completamente separado de los hongos, y muy estrechamente relacionado con las algas rojas y pardas, y con miembros de los Oomicetes (Mannella, *et al.*, 1987, Mol. Evol. 24:228-235). La mayoría de los taxonomistas han acordado retirar los traustocitridios de los Oomicetes (Bartnicki-Garcia, 1987, pág. 389-403 en "Evolutionary Biology of the Fungi" (eds. Rayner, A.D.M., Brasier, C.M. y Moore, D.), Cambridge University Press, Cambridge).

En resumen, empleando el sistema taxonómico de Cavalier-Smith (Cavalier-Smith, 1981, BioSystems 14:461-481, 1983; Cavalier-Smith, 1993, Microbiol Rev. 57:953-994), los traustocitridios se clasifican con las algas cromofitas en el reino Chromophyta (estrmenópilos). Esta colocación taxonómica se ha reafirmado más recientemente por Cavalier-Smith *et al.* usando las características de ARNr 18s del Heterokonta para demostrar que los traustocitridios son cromistas no hongos (Cavalier-Smith *et al.*, 1994, Phil. Tran. Roy. Soc. London Series BioSciences 346:387-397). Esto los coloca en un reino completamente diferente de los hongos, que están todos colocados en el reino Eufungi. La colocación taxonómica de los traustocitridios por lo tanto se resume del siguiente modo:

Reino: Chromophyta (estrmenópilos)  
 Filo: Heterokonta  
 Orden: Traustocitriales  
 Familia: Thraustochytriaceae  
 Género: *Thraustochytrium*, *Schizochytrium*, *Labyrinthuloides*, o *Japonochytrium*

Algunos de los primeros taxónomos separaron unos pocos miembros originales del género *Thraustochytrium* (aquellos con una fase de vida ameboide) en un género diferente llamado *Ulkenia*. Sin embargo, ahora se sabe que la mayoría, sino todos, los traustocitridios (incluyendo *Thraustochytrium* y *Schizochytrium*), muestran fases ameboides y, por tanto, *Ulkenia* no se considera por nadie como un género válido. Como se usa en este documento, el género *Thraustochytrium* incluirá *Ulkenia*.

A pesar de la inexactitud de la colocación taxonómica con clasificaciones superiores de Filo y Reino, los traustocitridios siguen siendo una agrupación distintiva y característica cuyos miembros siguen siendo clasificables dentro del orden Traustocitriales.

Los ácidos grasos poliinsaturados (PUFA) son componentes esenciales de membrana en eucariotas superiores y los precursores de muchas moléculas de señalización derivadas de lípidos. El sistema PUFA PKS de la presente invención usa rutas para la síntesis de PUFA que no requieren desaturación y elongación de ácidos grasos saturados. Las rutas catalizadas por PUFA PKS son distintas de las PKS previamente reconocidas tanto en estructura como en mecanismo. Se sugiere que la generación de dobles enlaces *cis* implica isomerasas específicas de posición; se cree que estas enzimas son útiles en la producción de nuevas familias de antibióticos.

Para producir rendimientos significativamente elevados de diversas moléculas bioactivas usando el sistema PUFA PKS de la presente invención, un organismo, preferiblemente un microorganismo o una planta, puede modificarse genéticamente para afectar a la actividad de un sistema PUFA PKS. En un aspecto, dicho organismo puede contener de forma endógena y expresar un sistema PUFA PKS, y la modificación genética puede ser una modificación genética de uno o más de los dominios funcionales del sistema PUFA PKS endógeno, mediante lo cual

la modificación tiene algún efecto sobre la actividad del sistema PUFA PKS. En otro aspecto, dicho organismo puede contener de forma endógena y expresar un sistema PUFA PKS, y la modificación genética puede ser una introducción de al menos una secuencia exógena de ácido nucleico (por ejemplo, una molécula de ácido nucleico recombinante), donde la secuencia exógena de ácido nucleico codifica al menos un dominio o proteína biológicamente activa de un segundo sistema PKS y/o una proteína que afecta a la actividad de dicho sistema PUFA PKS (por ejemplo, una fosfopanteteinil transferasa (PPTasa), analizada a continuación). En otro aspecto más, el organismo no contiene necesariamente de forma endógena (de forma natural) un sistema PUFA PKS, pero se modifica genéticamente para introducir al menos una molécula de ácido nucleico recombinante que codifique una secuencia de aminoácidos que tenga la actividad biológica de al menos un dominio de un sistema PUFA PKS. En este aspecto, la actividad de PUFA PKS se ve afectada por la introducción o aumento de la actividad de PUFA PKS en el organismo. Diversas realizaciones asociadas con cada uno de estos aspectos se analizarán en mayor detalle a continuación.

Por lo tanto, de acuerdo con la presente invención, una realización se refiere a un microorganismo modificado genéticamente, donde el microorganismo expresa un sistema PKS que comprende al menos un dominio biológicamente activo de acuerdo con la presente invención de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS).

La modificación genética afecta a la actividad del sistema PKS en el organismo. El proceso de selección mencionado en la parte (b) se ha descrito en detalle anteriormente e incluye las etapas de: (a) seleccionar un microorganismo que produce al menos un PUFA; y, (b) identificar un microorganismo de (a) que tenga capacidad de producir PUFA aumentados en condiciones de oxígeno disuelto de menos de aproximadamente el 5 % de saturación en el medio de fermentación, en comparación con la producción de PUFA por el microorganismo en condiciones de oxígeno disuelto de más de aproximadamente el 5 % de saturación, y preferiblemente aproximadamente el 10 %, y más preferiblemente aproximadamente el 15 %, y más preferiblemente aproximadamente el 20 % de saturación en el medio de fermentación. El microorganismo modificado genéticamente descrito en este documento puede incluir una cualquiera o más de las secuencias de ácido nucleico identificadas anteriormente, y/o cualquiera de los otros homólogos de cualquier ORF o dominio de PUFA PKS de *Schizochytrium* descrito en detalle anteriormente.

Como se usa en este documento, un microorganismo modificado genéticamente puede incluir una bacteria, protista, microalga, hongo, u otro microbio modificado genéticamente y, particularmente, cualquiera de los géneros del orden Traustoquitriales (por ejemplo, un traustoquitridio) descrito en este documento (por ejemplo, *Schizochytrium*, *Thraustochytrium*, *Japonochytrium*, *Labyrinthuloides*). Dicho microorganismo modificado genéticamente tiene un genoma que está modificado (es decir, mutado o cambiado) a partir de su forma normal (es decir, de tipo silvestre o de origen natural) de modo que se consiga el resultado deseado (es decir, actividad PUFA PKS aumentada o modificada y/o producción de un producto deseado usando el sistema PKS). La modificación genética de un microorganismo puede conseguirse usando técnicas de desarrollo clásico de cepas y/o técnicas genéticas moleculares. Dichas técnicas son conocidas en la técnica y se describen en líneas generales para microorganismos, por ejemplo, en Sambrook *et al.*, 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press. La referencia Sambrook *et al.*, *ibid*, se incorpora por referencia en este documento en su totalidad. Un microorganismo modificado genéticamente puede incluir un microorganismo en que se han insertado, delecionado o modificado moléculas de ácido nucleico (es decir, mutado; por ejemplo, por inserción, deleción, sustitución y/o inversión de nucleótidos), de tal modo que dichas modificaciones proporcionan el efecto deseado dentro del microorganismo.

Las células hospedadoras del microorganismo preferido a modificar de acuerdo con la presente invención incluyen, aunque sin limitación, cualquier bacteria, protista, microalga, hongo o protozoo. En un aspecto, los microorganismos preferidos para modificar genéticamente incluyen, aunque sin limitación, cualquier microorganismo del orden Traustoquitriales. Células hospedadoras particularmente preferidas para su uso en la presente invención podrían incluir microorganismo de un género que incluye, aunque sin limitación: *Thraustochytrium*, *Labyrinthuloides*, *Japonochytrium*, y *Schizochytrium*. Especies preferidas dentro de estos géneros incluyen, aunque sin limitación: cualquier especie de *Schizochytrium*, incluyendo *Schizochytrium aggregatum*, *Schizochytrium limacinum*, *Schizochytrium minutum*; cualquier especie de *Thraustochytrium* (incluyendo las antiguas especies *Ulkenia* tales como *U. visurgensis*, *U. amoeboida*, *U. sarkariana*, *U. profunda*, *U. radiata*, *U. minuta* y *Ulkenia sp.* BP-5601), e incluyendo *Thraustochytrium striatum*, *Thraustochytrium aureum*, *Thraustochytrium roseum*; y cualquier especie de *Japonochytrium*. Cepas particularmente preferidas de Traustoquitriales incluyen, aunque sin limitación: *Schizochytrium sp.* (S31)(ATCC 20888); *Schizochytrium sp.* (S8)(ATCC 20889); *Schizochytrium sp.* (LC-RM)(ATCC 18915); *Schizochytrium sp.* (SR21); *Schizochytrium aggregatum* (Goldstein et Belsky)(ATCC 28209); *Schizochytrium limacinum* (Honda et Yokochi)(IFO 32693); *Thraustochytrium sp.* (23B)(ATCC 20891); *Thraustochytrium striatum* (Schneider)(ATCC 24473); *Thraustochytrium aureum* (Goldstein)(ATCC 34304); *Thraustochytrium roseum* (Goldstein)(ATCC 28210); y *Japonochytrium sp.* (L1)(ATCC 28207). Otros ejemplos de microorganismos hospedadores adecuados para modificación genética incluyen, aunque sin limitación, levaduras incluyendo *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, u otra levadura como *Candida*, *Kluyveromyces*, u otros hongos, por ejemplo, hongos filamentosos tales como *Aspergillus*, *Neurospora*, *Penicillium*, etc. También pueden usarse células bacterianas como hospedadores. Estas incluyen *Escherichia coli*, que puede ser útil en procesos de fermentación. Como alternativa, puede usarse un hospedador tal como una especie de *Lactobacillus* o una especie de *Bacillus* como hospedador.

Otra realización de la presente invención se refiere a una planta modificada genéticamente, donde la planta se ha modificado genéticamente para que exprese de forma recombinante un sistema PKS que comprende al menos un dominio biológicamente activo de acuerdo con la invención de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS).

La planta modificada genéticamente puede incluir uno cualquiera o más de las secuencias de ácido nucleico identificadas anteriormente, y/o cualquiera de los otros homólogos de cualquiera de las ORF o dominios de PUFA PKS de *Schizochytrium* descritos en detalle anteriormente.

Como se usa en este documento, una planta modificada genéticamente puede incluir cualquier planta modificada genéticamente incluyendo plantas superiores y particularmente, cualquier planta consumible u plantas útiles para producir una molécula bioactiva deseada de la presente invención. Dicha planta modificada genéticamente tiene un genoma que está modificado (es decir, mutado o cambiado) a partir de su forma normal (es decir, de tipo silvestre o de origen natural) de modo que consiga el resultado deseado (es decir, actividad PUFA PKS aumentada o modificada y/o producción de un producto deseado usando el sistema PKS). La modificación genética de una planta puede conseguirse usando técnicas clásicas de desarrollo de cepas y/o técnicas de genética molecular. Los métodos para producir una planta transgénica, donde se incorpora una molécula de ácido nucleico recombinante que codifica una secuencia deseada de aminoácidos en el genoma de la planta, son conocidos en la técnica. Una planta preferida para modificar genéticamente de acuerdo con la presente invención es preferiblemente una planta adecuada para consumo por animales, incluyendo seres humanos.

Las plantas preferidas para modificar genéticamente de acuerdo con la presente invención (es decir, células hospedadoras vegetales) incluyen, aunque sin limitación, cualquier planta superior, y particularmente plantas consumibles, incluyendo planta de cultivo y especialmente plantas usadas por sus aceites. Dichas plantas pueden incluir, por ejemplo: canola, soja, colza, linaza, maíz, cártamo, girasol y tabaco. Otras plantas preferidas incluyen aquellas plantas que son conocidas por producir compuestos usados como agentes farmacéuticos, agentes aromatizantes, agentes nutracéuticos, ingredientes alimenticios funcionales o agentes cosméticamente activos o plantas que se modifican por ingeniería genética para producir estos compuestos/agentes.

De acuerdo con la presente invención, un microorganismo o planta modificado genéticamente incluye un microorganismo o planta que se ha modificado usando tecnología recombinante. Como se usa en este documento, las modificaciones genéticas que provocan una disminución en la expresión génica, en la función del gen, o en la función del producto génico (es decir, la proteína codificada por el gen) pueden mencionarse como inactivación (completa o parcial), delección, interrupción, bloqueo o regulación negativa de un gen. Por ejemplo, una modificación genética en un gen que provoca una disminución en la función de la proteína codificada por dicho gen, puede ser el resultado de una delección completa del gen (es decir, el gen no existe, y por lo tanto la proteína no existe), una mutación en el gen que provoca traducción incompleta o ausencia de traducción de la proteína (por ejemplo, la proteína no se expresa), o una mutación en el gen que disminuye o suprime la función natural de la proteína (por ejemplo, se expresa una proteína que tiene actividad o acción enzimática disminuida o ausente). Las modificaciones genéticas que provocan un aumento en la expresión génica o función pueden mencionarse como amplificación, sobreproducción, sobreexpresión, activación, potenciación, adición, o regulación positiva de un gen.

La modificación génica de un microorganismo o planta de acuerdo con la presente invención preferiblemente afecta a la actividad del sistema PKS expresado por la planta, sea el sistema PKS endógeno y modificado genéticamente, endógeno con la introducción de moléculas de ácido nucleico recombinante en el organismo, o proporcionado completamente por tecnología recombinante. De acuerdo con la presente invención, "afectar a la actividad de un sistema PKS" incluye cualquier modificación genética que cause cualquier cambio o modificación detectable o medible en el sistema PKS expresado por el organismo en comparación con la ausencia de la modificación genética. Un cambio o modificación detectable en el sistema PKS puede incluir, aunque sin limitación: la introducción de actividad del sistema PKS en un organismo de modo que el organismo ahora tenga actividad del sistema PKS medible/detectable (es decir, el organismo no contenía un sistema PKS antes de la modificación genética), la introducción en el organismo de un dominio funcional a partir de un sistema PKS diferente que un sistema PKS expresado de forma endógena por el organismo de modo que se modifique la actividad del sistema PKS (por ejemplo, un dominio de PUFA PKS bacteriano o un dominio de PKS Tipo I se introduce en un organismo que expresa de forma endógena un sistema PUFA PKS no bacteriano), un cambio en la cantidad de una molécula bioactiva producida por el sistema PKS (por ejemplo, el sistema produce más (cantidad aumentada) o menos (cantidad disminuida) de un producto dado en comparación con la ausencia de la modificación genética), un cambio en el tipo de una molécula bioactiva producida por el sistema PKS (por ejemplo, el sistema produce un producto nuevo o diferente, o una variante de un producto que se produce de forma natural por el sistema), y/o un cambio en la proporción de múltiples moléculas bioactivas producidas por el sistema PKS (por ejemplo, el sistema produce una relación diferente de un PUFA a otro PUFA, produce un perfil completamente diferente de lípidos en comparación con la ausencia de la modificación genética, o sitúa diversos PUFA en diferentes posiciones en un triacilglicerol en comparación con la configuración natural). Dicha modificación genética incluye cualquier tipo de modificación genética e incluye específicamente modificaciones hechas por tecnología recombinante y por mutagénesis clásica.

5 Debe apreciarse que referencias a aumentar la actividad de un dominio funcional o proteína en un sistema PUFA PKS se refiere a cualquier modificación genética en el organismo que contiene el dominio o proteína (o en el que tiene que introducirse el dominio o proteína) que provoca funcionalidad aumentada del dominio o sistema proteico y puede incluir mayor actividad del dominio o proteína (por ejemplo, actividad específica o actividad enzimática *in vivo*), inhibición o degradación reducida del dominio o sistema proteico, y sobreexpresión del dominio o proteína. Por ejemplo, puede aumentarse la cantidad de copias del gen, pueden aumentarse los niveles de expresión mediante el uso de un promotor que da mayores niveles de expresión que los del promotor nativo, o puede alterarse un gen por ingeniería genética o mutagénesis clásica para aumentar la actividad del dominio o proteína codificada por el gen.

10 Asimismo, referencias a disminuir la actividad de un dominio funcional o proteína en un sistema PUFA PKS se refiere a cualquier modificación genética en el organismo que contiene dicho dominio o proteína (o en cual tiene que introducirse el dominio o proteína) que provoca funcionalidad disminuida del dominio o proteína e incluye actividad disminuida del dominio o proteína, inhibición o degradación aumentada del dominio o proteína y una reducción o eliminación de la expresión del dominio o proteína. Por ejemplo, la acción del dominio o proteína de la presente invención puede disminuirse bloqueando o reduciendo la producción del dominio o proteína, "eliminando" el gen o parte del mismo que codifica el dominio o proteína, reduciendo la actividad del dominio o proteína, o inhibiendo la actividad del dominio o proteína. El bloqueo o reducción de la producción de un dominio o proteína puede incluir colocar el gen que codifica el dominio o proteína bajo el control de un promotor que requiere la presencia de un compuesto inductor en el medio de cultivo. Estableciendo las condiciones de modo que el inductor llegue a eliminarse del medio, podría inactivarse la expresión del gen que codifica el dominio o proteína (y por lo tanto, de síntesis proteica). El bloqueo o reducción de la actividad del dominio o proteína también podría incluir el uso de un enfoque de tecnología de escisión similar al descrito en la patente de Estados Unidos n.º 4.743.546, incorporada por referencia en este documento. Para usar este enfoque, el gen que codifica la proteína de interés se clona entre secuencias genéticas específicas que permiten escisión controlada específica del gen a partir del genoma. La escisión podría impulsarse por, por ejemplo, un desplazamiento en la temperatura de cultivo del cultivo, como en la patente de Estados Unidos n.º 4.743.546, o por alguna otra señal física o nutricional.

30 En una realización de la presente invención, una modificación genética incluye una modificación de una secuencia de ácido nucleico que codifica una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS no bacteriano como se describe en este documento. Dicha modificación puede ser a una secuencia de aminoácidos dentro un sistema PUFA PKS no bacteriano expresado de forma endógena (de forma natural), mediante lo cual un microorganismo que contiene de forma natural dicho sistema se modifica genéticamente por, por ejemplo, mutagénesis clásica y técnicas de selección y/o técnicas de genética molecular, incluyendo técnicas de ingeniería genética. Las técnicas de ingeniería genética pueden incluir, por ejemplo, el uso de un vector recombinante de direccionamiento para delecionar una parte de un gen endógeno, o para remplazar una parte de un gen endógeno con una secuencia heteróloga. Ejemplos de secuencias heterólogas que podrían introducirse en un genoma hospedador incluyen secuencias que codifican al menos un dominio funcional de otro sistema PKS, tal como un sistema PUFA PKS no bacteriano diferente, un sistema PUFA PKS bacteriano, un sistema PKS tipo I, un sistema PKS tipo II, o un sistema PKS modular. Otras secuencias heterólogas a introducir en el genoma de un hospedador incluyen una secuencia que codifica una proteína o dominio funcional que no es un dominio de un sistema PKS, pero que afectará a la actividad del sistema PKS endógeno. Por ejemplo, podría introducirse en el genoma hospedador una molécula de ácido nucleico que codifica una fosfopanteteinil transferasa (analizada a continuación). Se analizan modificaciones específicas que podrían hacerse a un sistema PUFA PKS endógeno en detalle a continuación.

45 En otro aspecto de esta realización de la invención, la modificación genética puede incluir: (1) la introducción de una molécula de ácido nucleico recombinante que codifica una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema PUFA PKS no bacteriano; y/o (2) la introducción de una molécula de ácido nucleico recombinante que codifica una proteína o dominio funcional que afecta a la actividad de un sistema PUFA PKS, en un hospedador. El hospedador puede incluir: (1) una célula hospedadora que no expresa ningún sistema PKS donde todos los dominios funcionales de un sistema PKS se introducen en la célula hospedadora, y donde al menos un dominio funcional es de un sistema PUFA PKS no bacteriano; (2) una célula hospedadora que expresa un sistema PKS (endógeno o recombinante) que tiene al menos un dominio funcional de un sistema PUFA PKS no bacteriano, donde la molécula introducida de ácido nucleico recombinante puede modificar al menos una función adicional de dominio de PUFA PKS, no bacteriano u otra proteína o dominio que afecte a la actividad del sistema PUFA PKS del hospedador; y (3) Una célula hospedadora que expresa un sistema PKS (endógeno o recombinante) que no incluye necesariamente una función de dominio de un PUFA PKS no bacteriano, y donde la molécula introducida de ácido nucleico recombinante incluye una secuencia de ácido nucleico que codifica al menos un dominio funcional de un sistema PUFA PKS no bacteriano. En otras palabras, la presente invención pretende abarcar cualquier organismo modificado genéticamente (por ejemplo, el microorganismo o planta), donde el organismo comprende al menos una función de dominio de PUFA PKS no bacteriano (de forma endógena o por modificación recombinante), y donde la modificación genética tiene un efecto medible sobre la función del dominio de PUFA PKS no bacteriano o sobre el sistema PKS cuando el organismo comprende un sistema PKS funcional.

65

Por lo tanto, usando los sistemas PUFA PKS no bacterianos de la presente invención que, por ejemplo, hacen uso de genes de sistemas PUFA PKS de traustroquiritridios, puede usarse mezcla génica para ampliar el intervalo de productos PUFA para que incluyan EPA, DHA, ARA, GLA, SDA y otros, así como para producir una amplia diversidad de moléculas bioactivas incluyendo antibióticos, otros compuestos farmacéuticos, y otros productos deseables. El método para obtener estas moléculas bioactivas incluye no solo la mezcla de genes de diversos organismos sino también diversos métodos para modificar genéticamente los genes de PUFA PKS no bacteriana descritos en este documento. El conocimiento de la base genética y la estructura de dominios del sistema PUFA PKS no bacteriano de la presente invención proporcionan una base para diseñar nuevos organismos modificados genéticamente que producirán una diversidad de moléculas bioactivas. Aunque la mezcla y modificación de cualquier dominio de PKS y genes relacionados se contempla por los presentes inventores, a modo de ejemplo, a continuación se analizan diversas manipulaciones posibles del sistema PUFA PKS con respecto a modificación genética y producción de moléculas bioactivas.

Por ejemplo, se alteran los productos del sistema PUFA PKS no bacteriano, tales como los producidos por traustroquiritridios, modificando el dominio CLF (factor de longitud de cadena). Este dominio es característico de sistemas PKS Tipo II (enzimas disociadas). Su secuencia de aminoácidos muestra homología a dominios KS (pares ceto sintasa), pero carece de la cisteína del sitio activo. CLF puede funcionar determinando la cantidad de ciclos de elongación, y por tanto la longitud de cadena, del producto final. Usando el estado actual del conocimiento sobre la síntesis de FAS y PKS se proporciona una estrategia racional para la producción de ARA mediante modificación dirigida del sistema PUFA PKS no bacteriano. Hay controversia en la bibliografía respecto a la función de CLF en sistema PKS (C. Bisang *et al.*, Nature 401, 502 (1999)) y se pone de manifiesto que otros dominios pueden estar implicados en la determinación de la longitud de cadena del producto final. Sin embargo, es significativo que *Schizochytrium* produce tanto DHA (C22:6,  $\omega$ -3) y DPA (C22:5,  $\omega$ -6). En el sistema PUFA-PKS los dobles enlaces *cis* se introducen durante la síntesis de la cadena de carbono creciente. Como la colocación de los dobles enlaces  $\omega$ -3 y  $\omega$ -6 sucede de forma prematura en la síntesis de las moléculas, se esperaría que afectara a la posterior determinación de la longitud de cadena del producto final. Por tanto, sin deseo de limitarse a teoría alguna, los presentes inventores creen que la introducción de un factor (por ejemplo, CLF) que dirija la síntesis de unidades C20 (el lugar de unidades C22) en el sistema PUFA PKS de *Schizochytrium* provocará la producción de EPA (C20:5,  $\omega$ -3) y ARA (C20:4,  $\omega$ -6). Por ejemplo, en sistemas heterólogos, podría explotarse el CLF sustituyendo directamente un CLF de un sistema productor de EPA (tal como uno de *Photobacterium*) en el conjunto de genes de *Schizochytrium*. Los ácidos grasos de los transformantes resultantes pueden analizarse posteriormente para alteraciones en los perfiles para identificar los transformantes que producen EPA y/o ARA.

Además de la dependencia en el desarrollo de un sistema heterólogo (sistema recombinante, tal como podría introducirse en plantas), el concepto de CLF puede explotarse en *Schizochytrium* (es decir, por modificación de un genoma de *Schizochytrium*). Se ha demostrado la transformación y recombinación homóloga en *Schizochytrium*. Puede explotarse esto construyendo un clon con el CLF de OrfB reemplazado con un CLF de un sistema C20 PUFA PKS. Se insertará un gen marcador cadena abajo de la región codificante. Después pueden transformarse las células de tipo silvestre, seleccionar el fenotipo marcador y después detectar aquellos que hubieran incorporado el nuevo CLF. De nuevo, se analizarían estos para cualquier efecto sobre los perfiles de ácidos grasos para identificar transformantes productores de EPA y/o ARA. Si se encuentra que algún factor diferente a los asociados con el CLF influye en la longitud de cadena del producto final, podría emplearse una estrategia similar para alterar esos factores.

Describimos la alteración de los productos de PUFA PKS, que implica la modificación o sustitución de los pares  $\beta$ -hidroxi acil ACP deshidrasa/ceto sintasa. Durante la síntesis de ácido *cis*-vaccénico (C8:1,  $\Delta$ 11) en *E. coli*, se cree que la creación del doble enlace *cis* depende de una enzima DH específica  $\beta$ -hidroxi acil ACP deshidrasa, el producto del gen *FabA*. Esta enzima retira HOH de un  $\beta$ -ceto acil ACP y deja un doble enlace *trans* en la cadena de carbono. Un subconjunto de DH, tipo *FabA*, posee actividad *cis-trans* isomerasa (Heath *et al.*, 1996, *supra*). Un nuevo aspecto de sistemas PUFA PKS bacterianos y no bacterianos es la presencia de dos dominios DH tipo *FabA*. Sin el deseo de limitarse a teoría alguna, los presentes inventores creen que uno de estos dos dominios DH o ambos poseerán actividad *cis-trans* isomerasa (la manipulación de los dominios DH se analiza en mayor detalle a continuación).

Otro aspecto de las síntesis de ácidos grasos insaturados en *E. coli* es la necesidad de una enzima KS particular,  $\beta$ -ceto acil ACP sintasa, el producto del gen *FabB*. Esta es la enzima que realiza la condensación de un ácido graso, unido a un resto de cisteína en el sitio activo (mediante un enlace tio-éster), con un malonil ACP. En la reacción de múltiples etapas, se libera CO<sub>2</sub> y se prolonga la cadena lineal en dos carbonos. Se cree que solamente esta KS puede prolongar una cadena de carbono que contenga un doble enlace. Esta prolongación sucede solamente cuando el doble enlace está en configuración *cis*; si está en configuración *trans*, el doble enlace se reduce por enoil ACP reductasa (ER) antes de la elongación (Heath *et al.*, 1996, *supra*). Todos los sistemas PUFA PKS caracterizados hasta ahora tienen dos dominios KS, uno de los cuales muestra mayor homología a la KS tipo *FabB* de *E. coli* que el otro. De nuevo, sin el deseo de limitarse a teoría alguna, los presentes inventores creen que en sistema PUFA PKS, las especificidades e interacciones de los dominios enzimáticos DH (tipo *FabA*) y KS (tipo *FabB*) determinan la cantidad y colocación de dobles enlaces *cis* en los productos finales. Como la cantidad de reacciones de elongación de dos carbonos es mayor que la cantidad de dobles enlaces presentes en los productos finales de

PUFA PKS, puede determinarse que en algunos ciclos de extensión sucede reducción completa. Por tanto, pueden usarse los dominios DH y KS como dianas para alteración de la relación DHA/DPA o relaciones de otros ácidos grasos de cadena larga. Estos pueden modificarse y/o evaluarse mediante la introducción de dominios homólogos de otros sistemas o por mutagénesis de estos fragmentos génicos.

5 Los dominios ER (enoil-ACP reductasa - una enzima que reduce el doble enlace *trans* en el acilo graso-ACP produciendo carbonos completamente saturados) pueden modificarse o sustituirse para cambiar el tipo de producto fabricado por el sistema PKS. Por ejemplo, los presentes inventores saben que el sistema PUFA PKS de *Schizochytrium* difiere de los sistemas bacterianos descritos previamente en que tiene dos dominios ER (en lugar de uno). Sin el deseo de limitarse a teoría alguna, los presentes inventores creen que estos dominios ER pueden influir fuertemente en el producto resultante de producción de PKS. El producto de PKS resultante podría cambiarse eliminando por separado los dominios individuales o modificando su secuencia de nucleótidos o por sustitución de los dominios ER de otros organismos.

15 En otra realización, pueden introducirse moléculas de ácido nucleico que codifican proteínas o dominios que no son parte de un sistema PKS, pero que afectan a un sistema PKS, en un organismo. Por ejemplo, todos los sistemas PUFA PKS descritos anteriormente contienen múltiples dominios ACP, en tándem. ACP (como proteína separada o como un dominio de una proteína más grande) requiere unión de un co-factor de fosfopanteteína para producir la holo-ACP activa. La unión de fosfopanteteína a la apo-ACP se realiza por miembros de la superfamilia de enzimas - las fosfopanteteinil transferasa (PPTasa) (Lambalot R.H., *et al.*, Chemistry and Biology, 3, 923 (1996)).

20 Por analogía a otros sistemas PKS y FAS, los presentes inventores suponen que la activación de los múltiples dominios ACP presentes en la proteína OrfA de *Schizochytrium* se realiza por una PPTasa endógena específica. El gen que codifica esta supuesta PPTasa aún no se ha identificado en *Schizochytrium*. Si dicho gen está presente en *Schizochytrium*, pueden idearse varios enfoques que podrían usarse en un intento por identificarlo y clonarlo. Estos podrían incluir (aunque sin limitación): generación y secuenciación parcial de una biblioteca de ADNc preparada a partir de células de crecimiento activo de *Schizochytrium* (nota, se identificó una secuencia en la biblioteca de ADNc de *Schizochytrium* actualmente disponible que mostró homología a PPTasa; sin embargo, parece ser parte de una proteína FAS de múltiples dominios, y por tanto no puede codificar la PPTasa específica de OrfA deseada); uso de cebadores oligonucleotídicos degenerados diseñados usando motivos de aminoácidos presentes en muchas PPTasa en reacciones de PCR (para obtener una molécula de sonda de ácido nucleico para seleccionar bibliotecas genómicas o de ADNc); enfoques genéticos basados en interacciones, proteína-proteína (por ejemplo, un sistema doble híbrido de levadura) en que se usarían los OrfA-dominios ACP como "cebo" para encontrar una "diana" (es decir, la PPTasa); y purificación y secuenciación parcial de la propia enzima como medio para generar una sonda de ácido nucleico para seleccionar bibliotecas genómicas o de ADN.

35 También es concebible que una PPTasa heteróloga pueda ser capaz de activar los dominios ACP de OrfA de *Schizochytrium*. Se ha demostrado que algunas PPTasas, por ejemplo la enzima sfp de *Bacillus subtilis* (Lambalot *et al.*, *supra*) y la enzima svp de *Streptomyces verticillus* (Sanchez *et al.*, 2001, Chemistry & Biology 8:725-738) tienen una amplia tolerancia de sustrato. Estas enzimas pueden ensayarse para observar si activarán los dominios ACP de *Schizochytrium*. Además, una reciente publicación describió la expresión de una proteína PKS fúngica en tabaco (Yalpani *et al.*, 2001, The Plant Cell 13:1401-1409). Se detectaron productos del sistema PKS introducido (codificado por el gen de la ácido 6-metilsalicílico sintasa de *Penicillium patulum*) en la planta transgénica, aunque la correspondiente PPTasa fúngica no estaba presente en esas plantas. Esto sugirió que una o más PPTasas vegetales endógenas reconocían y activaban el dominio ACP de PKS fúngica. De relevancia para esta observación, los presentes inventores han identificado dos secuencias (genes) en la base de datos de genoma completo de *Arabidopsis* que probablemente codifican PPTasas. Estas secuencias (números de acceso a GenBank; AAG51443 y AAC05345) están actualmente registradas como codificantes de "Proteínas Desconocidas". Pueden identificarse como PPTasas putativas en base a la presencia en las secuencias proteicas traducidas de varios motivos característicos incluyendo; G(I/V)D y WxxKE(A/S)xxK (SEC ID N° 33), (enumerados en Lambalot *et al.*, 1996 como característicos de todas las PPTasas). Además, estas dos proteínas putativas contienen dos motivos adicionales típicamente encontrados en PPTasas típicamente asociadas con PKS y sistemas de síntesis no ribosómica de péptidos; es decir, FN(I/L/V)SHS (SEC ID N° 34) y (I/V/L)G(I/L/V)D(I/L/V) (SEC ID N° 35). Además, estos motivos existen en las posiciones relativas esperadas en las secuencias proteicas. Es probable que estén presentes homólogos de los genes de *Arabidopsis* en otras plantas, tales como tabaco. De nuevo, estos genes pueden clonarse y expresarse para observar si las enzimas que codifican pueden activar los dominios ACP de OrfA de *Schizochytrium*, o como alternativa, podría expresarse OrfA directamente en la planta transgénica (dirigido al plásmido o el citoplasma).

60 Otra PPTasa heteróloga que puede reconocer los dominios ACP de OrfA como sustratos es la proteína Het I de *Nostoc* sp. PCC 7120 (antiguamente llamado *Anabaena* sp. PCC 7120). Como se aprecia en la patente de Estados Unidos n.º 6.140.486, varios genes de PUFA PKS de *Shewanella* mostraron un alto grado de homología a los dominios proteicos presentes en un grupo PKS encontrado en *Nostoc* (Figura 2 de esa patente). Este sistema PKS de *Nostoc* está asociado con la síntesis de hidroxí ácidos grasos de cadena larga (C26 o C28) que se esterifican en restos de azúcar y forman una parte de la pared celular del heterocisto. Estos dominios de PKS de *Nostoc* también

son muy homólogos a los dominios encontrados en las Orf B y C de las proteínas de PKS de *Schizochytrium* (es decir, las mismas que corresponden a las encontradas en las proteínas de PKS de *Shewanella*).

5 Hasta muy recientemente, ninguno de los dominios de PKS de *Nostoc* presente en las bases de datos GenBank mostró alta homología a ninguna de los dominios de OrfA de *Schizochytrium* (o la proteína Orf 5 homóloga de *Shewanella*). Sin embargo, el genoma completo de *Nostoc* se ha secuenciado recientemente y como resultado ahora está disponible la secuencia de la región justo cadena arriba del grupo génico PKS. En esta región hay tres Orf que muestran homología a los dominios (KS, MAT, ACP y KR) de OrfA (véase la Fig. 3). Se incluyen en este conjunto dos dominios ACP, ambos cuales muestran alta homología a los dominios ACP de OrfA. Al final del grupo 10 PKS de *Nostoc* está el gen que codifica la PPTasa Het I. Previamente, no era obvio cual podría ser el sustrato de la enzima Het I, sin embargo, la presencia de dominios ACP en tándem en la Orf recién identificada (Hgl E) del grupo sugiere fuertemente a los presentes inventores que es esos ACP. La homología de los dominios ACP de *Schizochytrium* y *Nostoc*, así como la disposición en tándem de los dominios en ambas proteínas, hace de Het I un candidato probable para activación heteróloga de las ACP de OrfA de *Schizochytrium*. Los presentes inventores creen que son los primeros en reconocer y contemplar este uso para la PPTasa Het I de *Nostoc*. 15

Como se indica en Metz *et al.*, 2001, *supra*, una nueva característica de los sistemas PUFA PKS es la presencia de dos dominios deshidratasa, ambos cuales muestran homología a las proteínas FabA de *E. coli*. Con la disponibilidad de las nuevas secuencias génicas de PKS de *Nostoc* mencionadas anteriormente, ahora pueden compararse los dos sistemas y sus productos. La secuencia de los dominios en el grupo de *Nostoc* (de HglE a Het I) como los han definido los presentes inventores es (véase, la Fig. 3): 20

KS-MAT-2xACP, KR, KS, CLF-AT, ER (HetM, HetN) HetI

25 En las Orf A, B y C de PUFA PKS de *Schizochytrium* la secuencia (OrfA - B - C) es:

KS-MAT-9xACP-KR KS-CLF-AT-ER DH-DH-ER

30 Puede observarse la correspondencia de la secuencia de los dominios (también existe una alta homología de secuencia de aminoácidos). El producto del sistema PKS de *Nostoc* es un hidroxilado ácido graso de cadena larga (C26 o C28 con uno o dos grupos hidroxilo) que no contienen dobles enlaces (*cis* o *trans*). El producto del sistema PKS de *Schizochytrium* es un ácido graso poliinsaturado de cadena larga (C22, con 5 o 6 dobles enlaces - todos *cis*). Una diferencia obvia entre los dos conjuntos de dominios es la presencia de los dos dominios DH en las proteínas de *Schizochytrium* - justo los dominios implicados en la formación de los dobles enlaces *cis* de DHA y de DPA 35 (presumiblemente HetM y HetN en el sistema de *Nostoc* están implicadas en la inclusión de los grupos hidroxilo y también contienen un dominio DH cuyo origen difiere del de los encontrados en el PUFA). Además, el papel del dominio ER duplicado en las Orf B y C de *Schizochytrium* es desconocido (el segundo dominio ER no está presente en otros sistemas PUFA PKS caracterizados). La homología de secuencia de aminoácidos entre los dos conjuntos de dominios implica una relación evolutiva. Se puede concebir que el conjunto génico de PUFA PKS deriva de (en un sentido evolutivo) un conjunto génico de PKS tipo *Nostoc* ancestral por incorporación de los dominios DH (tipo FabA). La adición de los dominios DH provocaría la introducción de dobles enlaces *cis* en la nueva estructura del producto final de PKS. 40

45 Las comparaciones de las estructuras de dominio de PKS de *Schizochytrium* y *Nostoc* así como la comparación de la organización de dominios entre las proteínas de PUFA PKS de *Schizochytrium* y *Shewanella* demuestran la capacidad natural de alterar el orden de los dominios así como de incorporar nuevos dominios para crear nuevos productos finales. Además, los genes ahora pueden manipularse en el laboratorio para crear nuevos productos, la implicación de estas observaciones es que debe ser posible continuar manipulando los sistemas de un modo dirigido o aleatorio para influir en los productos finales. Por ejemplo, en una realización preferida, podría idearse la sustitución de uno de los dominios DH (tipo FabA) del sistema PUFA PKS en el lugar de un dominio DH que no poseía actividad de isomerización, creando potencialmente una molécula con una mezcla de dobles enlaces *cis* y *trans*. Los productos actuales del sistema PUFA PKS de *Schizochytrium* son DHA y DPA (C22:5 ω6). Si se manipula el sistema para producir ácidos grasos C20, se esperaría los productos EPA y ARA (C20:4 ω6). Esto podría proporcionar una nueva fuente para ARA. También se podría sustituir dominios de sistemas PUFA PKS relacionados que produjeran una diferente relación DHA a DPA - por ejemplo usando genes de *Thraustochytrium* 23B (el sistema PUFA PKS del cual se identifica por primera vez en este documento). 50 55

Adicionalmente, podría idearse la alteración específica de uno de los dominios ER (por ejemplo, por retirada, o inactivación) en el sistema PUFA PKS de *Schizochytrium* (otros sistemas PUFA PKS descritos hasta ahora no tienen dos dominios ER) para determinar su efecto sobre el perfil de producto final. Podrían intentarse estrategias similares de un modo dirigido para cada uno de los distintos dominios de las proteínas de PUFA PKS usando enfoques más o menos sofisticados. Por supuesto, no se establecería limitación a la manipulación de dominios individuales. Finalmente, podría ampliarse el enfoque mezclando dominios del sistema PUFA PKS y otros sistemas PKS o FAS (por ejemplo, tipo I, tipo II, modular) para crear un intervalo completo de nuevos productos finales. Por ejemplo, 60 65 podrían introducirse los dominios DH de PUFA PKS en sistemas que no incorporan normalmente dobles enlaces *cis* en sus productos finales.



Por consiguiente, la presente invención abarca métodos para modificar genéticamente células microbianas o vegetales por: modificación genética de al menos una secuencia de ácido nucleico en el organismo que codifica una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio funcional de un sistema PUFA PKS no bacteriano de acuerdo con la presente invención, y/o expresión de al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica dicha secuencia de aminoácidos. Se han descrito en detalle anteriormente diversas realizaciones de dichas secuencias, métodos para modificar genéticamente un organismo, y modificaciones específicas. Típicamente, el método se usa para producir un organismo modificado genéticamente particular que produce una molécula o moléculas bioactivas particulares.

Una realización de la presente invención se refiere a una célula hospedadora recombinante que se ha modificado para que exprese un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS), donde la PKS cataliza reacciones enzimáticas tanto iterativas como no iterativas, y donde el sistema PUFA PKS comprende: (a) al menos dos dominios enoil ACP-reductasa (ER); (b) al menos seis dominios de proteína de transporte de acilo (ACP); (c) al menos dos dominios  $\beta$ -ceto acil-ACP sintasa (KS); (d) al menos un dominio acil transferasa (AT); (e) al menos un dominio cetorreductasa (KR); (f) al menos dos dominios  $\beta$ -hidroxi acil-ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP aciltransferasa (MAT). En una realización, el sistema PUFA PKS es un sistema PUFA PKS eucariota. En una realización preferida, el sistema PUFA PKS es un sistema PUFA PKS de Traustochytriales. Dichos sistemas PUFA PKS pueden incluir, aunque sin limitación, un sistema PUFA PKS de *Schizochytrium*, y un sistema PUFA PKS de *Thraustochytrium*. En una realización, el sistema PUFA PKS puede expresarse en una célula hospedadora procariota. En otra realización, el sistema PUFA PKS puede expresarse en una célula hospedadora eucariota.

Otra realización de la presente invención se refiere a una célula hospedadora recombinante que se ha modificado para que exprese un sistema PUFA PKS no bacteriano, donde el sistema PKS cataliza reacciones enzimáticas tanto iterativas como no iterativas, y donde el sistema PUFA PKS no bacteriano comprende al menos los siguientes dominios biológicamente activos: (a) al menos un dominio enoil ACP-reductasa (ER); (b) múltiples dominios de proteína de transporte de acilo (ACP) (al menos cuatro); (c) al menos dos dominios  $\beta$ -aceto acil-ACP sintasa (KS); (d) al menos un dominio aciltransferasa (AT); (e) al menos un dominio cetorreductasa (KR); (f) al menos dos dominios  $\beta$ -hidroxi acil-ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP aciltransferasa (MAT).

Un aspecto de esta realización de la invención se refiere a un método para producir un producto que contiene al menos un PUFA, que comprende cultivar una planta que comprende cualquiera de las células hospedadoras recombinantes descritas anteriormente, donde la célula hospedadora recombinante es una célula vegetal, en condiciones eficaces para producir el producto. Otro aspecto de esta realización de la invención se refiere a un método para producir un producto que contiene al menos un PUFA, que comprende cultivar un cultivo que contiene cualquiera de las células hospedadoras recombinantes descritas anteriormente, donde la célula hospedadora es una célula microbiana, en condiciones eficaces para producir el producto. En una realización preferida, el sistema PKS en la célula hospedadora cataliza la producción directa de triglicéridos.

Otra realización de la presente invención se refiere a un microorganismo que comprende un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) no bacteriano, donde la PKS cataliza reacciones enzimáticas tanto iterativas como no iterativas, y donde el sistema PUFA PKS comprende: (a) al menos dos dominios enoil ACP-reductasa (ER); (b) al menos seis dominios de proteína de transporte de acilo (ACP); (c) al menos dos dominios  $\beta$ -ceto acil-ACP sintasa (KS); (d) al menos un dominio aciltransferasa (AT); (e) al menos un dominio cetorreductasa (KR); (f) al menos dos dominios  $\beta$ -hidroxi acil-ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP aciltransferasa (MAT). Preferiblemente, el microorganismo es un microorganismo no bacteriano y más preferiblemente, un microorganismo eucariota.

Otra realización más de la presente invención se refiere a un microorganismo que comprende un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) no bacteriano, donde la PKS cataliza reacciones enzimáticas tanto iterativas como no iterativas, y donde el sistema PUFA PKS comprende: (a) al menos un dominio enoil ACP-reductasa (ER); (b) múltiples dominios de proteína de transporte de acilo (ACP) (al menos cuatro); (c) al menos dos dominios  $\beta$ -ceto acil-ACP sintasa (KS); (d) al menos un dominio aciltransferasa (AT); (e) al menos un dominio cetorreductasa (KR); (f) al menos dos dominios  $\beta$ -hidroxi acil-ACP deshidrasa (DH) tipo FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP aciltransferasa (MAT).

En una realización de la presente invención, se contempla que podría combinarse un programa de mutagénesis con un proceso de selección selectiva para obtener moléculas bioactivas de interés. Esto incluiría métodos para buscar un intervalo de compuestos bioactivos. Esta búsqueda no se restringiría a la producción de aquellas moléculas con dobles enlaces *cis*. Los métodos de mutagénesis podrían incluir, aunque sin limitación: mutagénesis química, combinación de genes, cambio de regiones de los genes que codifican dominios enzimáticos específicos, o mutagénesis restringida a regiones específicas de esos genes, así como otros métodos.

Por ejemplo, podrían usarse métodos de mutagénesis de alto rendimiento para influir en u optimizar la producción de la molécula bioactiva deseada. Una vez se ha desarrollado un sistema modelo eficaz, podrían modificarse estos genes de un modo de alto rendimiento. La utilización de estas tecnologías puede idearse a dos niveles. Primero, si puede concebirse un filtro suficientemente selectivo para la producción de un producto de interés (por ejemplo, ARA), podría usarse para intentar alterar el sistema para producir este producto (por ejemplo, en lugar de, o en concierto con, otras estrategias tales como las analizadas anteriormente). Adicionalmente, si las estrategias resumidas anteriormente produjeran un conjunto de genes que produjeran el producto de interés, entonces podrían usarse las tecnologías de alto rendimiento para optimizar el sistema. Por ejemplo, si el dominio introducido solamente funcionara a temperaturas relativamente bajas, podrían concebirse métodos de selección para permitir la eliminación de esa limitación. En una realización de la invención, se usan métodos de selección para identificar organismos no bacterianos adicionales que tienen nuevos sistemas PKS similares al sistema PUFA PKS de *Schizochytrium*, como se describe en este documento (véase anteriormente). Pueden usarse sistemas PKS homólogos identificados en dichos organismos en métodos similares a los descritos en este documento para *Schizochytrium*, así como para una fuente adicional de material genético a partir de la cual crear, modificar adicionalmente y/o mutar un sistema PKS para la expresión en ese organismo, en otro organismo, o en una planta superior, para producir una diversidad de compuestos.

Está reconocido que muchas alteraciones genéticas, ya sean aleatorias o dirigidas, que pueden introducirse en un sistema PKS nativo (endógeno, natural), provocarán una inactivación de funciones enzimáticas. Describimos un sistema para seleccionar solamente aquellas modificaciones que no bloquean la capacidad del sistema PKS de producir un producto. Por ejemplo, la cepa FabB- de *E. coli* es incapaz de sintetizar ácidos grasos insaturados y requiere suplementación del medio con ácidos grasos que puedan sustituir sus ácidos grasos insaturados normales para crecer (véase, Metz *et al.*, 2001, *supra*). Sin embargo, esta necesidad (suplementación del medio) puede eliminarse cuando la cepa se transforma con un sistema PUFA PKS funcional (es decir, uno que produce un producto PUFA en el hospedador de *E. coli* - véase (Metz *et al.*, 2001, *supra*, Figura 2A). La cepa FabB-transformada ahora requiere un sistema PUFA PKS funcional (para producir los ácidos grasos insaturados) para crecer sin suplementación. El elemento clave en este ejemplo es que la producción de un amplio intervalo de ácidos grasos insaturados bastará (incluso sustitutos de ácidos grasos insaturados tales como ácidos grasos de cadena ramificada). Por lo tanto, podría crearse una gran cantidad de mutaciones en uno o más de los genes de PUFA PKS descritos en este documento, y después transformar la cepa FabB- apropiadamente modificada (por ejemplo, crear mutaciones en una construcción de expresión que contiene un dominio ER y transformar una cepa FabB- que tiene los otros dominios esenciales en un plásmido diferente - o integrado en el cromosoma) y seleccionar solamente aquellos transformantes que crecen sin suplementación del medio (es decir, que aún poseen una capacidad de producir una molécula que podría complementar el defecto FabB-). Podrían desarrollarse filtros adicionales para buscar compuestos particulares (por ejemplo, uso de GC para ácidos grasos) que se producen en este subconjunto selectivo de un sistema PKS activo. Podrían idearse varios filtros selectivos similares para moléculas bioactivas de interés.

Como se ha descrito anteriormente, en una realización de la presente invención, un microorganismo o planta modificado genéticamente incluye un microorganismo o planta que tiene una capacidad potenciada de sintetizar moléculas bioactivas deseadas (productos) o que tiene una capacidad recién introducida de sintetizar productos específicos (por ejemplo, de sintetizar un antibiótico específico). De acuerdo con la presente invención, "una capacidad potenciada de sintetizar" un producto se refiere a cualquier potenciación, o regulación positiva, en una ruta relacionada con la síntesis del producto de modo que el microorganismo o planta produzca una cantidad aumentada del producto (incluyendo cualquier producción de un producto donde antes no había) en comparación con el microorganismo o planta de tipo silvestre, cultivado o desarrollado, en las mismas condiciones. Los métodos para producir dichos organismos modificados genéticamente se han descrito en detalle anteriormente.

Una realización de la presente invención es un método para producir moléculas bioactivas deseadas (también mencionadas como productos o compuestos) haciendo crecer o cultivando un microorganismo o planta modificada genéticamente de la presente invención (descrita en detalle anteriormente). Dicho método incluye la etapa de cultivar en un medio de fermentación o hacer crecer en un entorno adecuado, tal como suelo, un microorganismo o planta, respectivamente, que tenga una modificación genética como se ha descrito previamente en este documento y de acuerdo con la presente invención. En una realización preferida, el método para producir moléculas bioactivas de la presente invención incluye la etapa de cultivar en condiciones eficaces para producir la molécula bioactiva un organismo modificado genéticamente que expresa un sistema PKS que comprende al menos un dominio biológicamente activo de un sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS) de acuerdo con la invención. En este aspecto preferido del método, el organismo se modifica genéticamente para afectar a la actividad del sistema PKS (descrito en detalle anteriormente). Las células hospedadoras preferidas para modificación genética respecto al sistema PUFA PKS de la invención se han descrito anteriormente.

En el método de producción de compuestos bioactivos deseados de la presente invención, se cultiva o hace crecer un microorganismo modificado genéticamente en un medio adecuado, en condiciones eficaces para producir el compuesto bioactivo. Un medio apropiado, o eficaz se refiere a cualquier medio en que un microorganismo modificado genéticamente de la presente invención, cuando se cultiva, sea capaz de producir el producto deseado. Dicho medio es típicamente un medio acuoso que comprende fuentes asimilables de carbono, nitrógeno y fosfato.

Dicho medio también puede incluir sales, minerales, metales y otros nutrientes apropiados. Los microorganismos de la presente invención pueden cultivarse en biorreactores convencionales de fermentación. Los microorganismos pueden cultivarse por cualquier proceso de fermentación que incluya, aunque sin limitación, fermentación discontinua, semicontinua, con reciclado celular, y continua. Las condiciones preferidas de cultivo para microorganismos hospedadores potenciales de acuerdo con la presente invención son bien conocidas en la técnica. Las moléculas bioactivas deseadas producidas por el microorganismo modificado genéticamente pueden recuperarse del medio de fermentación usando técnicas convencionales de separación y purificación. Por ejemplo, el medio de fermentación puede filtrarse o centrifugarse para retirar los microorganismos, desechos celulares y otra materia particulada, y el producto puede recuperarse del sobrenadante sin células por métodos convencionales, tales como, por ejemplo, intercambio iónico, cromatografía, extracción, extracción de disolvente, separación en membrana, electrodiálisis, ósmosis inversa, destilación, derivatización química y cristalización. Como alternativa, los microorganismos que producen el compuesto deseado, o extractos y diversas fracciones del mismo, pueden usarse sin eliminar los componentes del microorganismo del producto.

En el método para la producción de compuestos bioactivos deseados de la presente invención, se cultiva una planta modificada genéticamente en un medio de fermentación o se hace crecer en un medio adecuado tal como suelo. Un medio apropiado o eficaz de fermentación se ha analizado en detalle anteriormente. Un medio de cultivo adecuado para plantas superiores incluye cualquier medio de cultivo para plantas, incluyendo, aunque sin limitación, suelo, arena, cualquier otro medio particulado que de soporte al crecimiento de las raíces (por ejemplo, vermiculita, perlita, etc.) o cultivo hidropónico, así como luz adecuada, agua y suplementos nutricionales que optimizan el crecimiento de la planta superior. Las plantas modificadas genéticamente de la presente invención se modifican por ingeniería para que produzcan cantidades significativas del producto deseado a través de la actividad del sistema PKS que está modificado genéticamente de acuerdo con la presente invención. Los compuestos pueden recuperarse a través de procesos de purificación que extraen los compuestos de la planta. En una realización preferida, el compuesto se recupera por recolección de la planta. En esta realización, la planta puede consumirse en su estado natural o procesarse adicionalmente en productos consumibles.

Como se ha descrito anteriormente, un microorganismo modificado genéticamente útil en la presente invención puede, en un aspecto, contener de forma endógena y expresar un sistema PUFA PKS, y la modificación genética puede ser una modificación genética de uno o más de los dominios funcionales del sistema PUFA PKS endógeno, mediante lo cual la modificación tiene algún efecto sobre la actividad del sistema PUFA PKS. En otro aspecto, dicho organismo puede contener de forma endógena y expresar un sistema PUFA PKS, y la modificación genética puede ser una introducción de al menos una secuencia exógena de ácido nucleico (por ejemplo, una molécula de ácido nucleico recombinante), donde la secuencia exógena de ácido nucleico codifica al menos un dominio o proteína biológicamente activa de un segundo sistema PKS y/o una proteína que afecta a la actividad de dicho sistema PUFA PKS (por ejemplo, una fosfopanteteinil transferasa (PPTasa), analizada a continuación). En otro aspecto más, el organismo no contiene necesariamente de forma endógena (natural) un sistema PUFA PKS, pero se modifica genéticamente para introducir al menos una molécula de ácido nucleico recombinante que codifique una secuencia de aminoácidos que tenga la actividad biológica de al menos un dominio de un sistema PUFA PKS. En este aspecto, la actividad PUFA PKS se ve afectada por la introducción o aumento de la actividad PUFA PKS en el organismo. Se han analizado en detalle anteriormente diversas realizaciones asociadas con cada uno de estos aspectos.

En una realización del método para producir compuestos bioactivos, la modificación genética cambia al menos un producto producido por el sistema PKS endógeno, en comparación con un organismo de tipo silvestre.

En otra realización, el organismo expresa de forma endógena un sistema PKS que comprende el al menos un dominio biológicamente activo del sistema PUFA PKS, y la modificación genética comprende la transfección del organismo con una molécula de ácido nucleico recombinante seleccionada entre el grupo que consiste en: una molécula de ácido nucleico recombinante que codifica al menos un dominio biológicamente activo de un segundo sistema PKS y una molécula de ácido nucleico recombinante que codifica una proteína que afecta a la actividad del sistema PUFA PKS. En esta realización, la modificación genética preferiblemente cambia al menos un producto producido por el sistema PKS endógeno, en comparación con un organismo de tipo silvestre. Un segundo sistema PKS puede incluir otro sistema PUFA PKS (bacteriano o no bacteriano), un sistema PKS tipo I, un sistema PKS tipo II, y/o un sistema PKS modular. Se han descrito anteriormente ejemplos de proteínas que afectan a la actividad de un sistema PKS (por ejemplo, PPTasa).

En otra realización, el organismo se modifica genéticamente por transfección con una molécula de ácido nucleico recombinante que codifica el al menos un dominio del sistema de ácido graso poliinsaturado (PUFA) policétido sintasa (PKS). Dichas moléculas de ácido nucleico recombinantes se ha descrito en detalle previamente en este documento.

En otra realización, el organismo expresa de forma endógena un sistema PUFA PKS no bacteriano, y la modificación genética comprende la sustitución de un dominio de un sistema PKS diferente en el lugar de una secuencia de ácido nucleico que codifica al menos un dominio del sistema PUFA PKS no bacteriano. En otra realización, el organismo expresa de forma endógena un sistema PUFA PKS no bacteriano que se ha modificado por transfección del organismo con una molécula de ácido nucleico recombinante que codifica una proteína que

regula la longitud de cadena de los ácidos grasos producidos por el sistema PUFA PKS. En un aspecto, la molécula de ácido nucleico recombinante que codifica una proteína que regula la longitud de cadena de los ácidos grasos reemplaza una secuencia de ácido nucleico que codifica un factor de longitud de cadena en el sistema PUFA PKS no bacteriano. En otro aspecto, la proteína que regula la longitud de cadena de los ácidos grasos producidos por el sistema PUFA PKS es un factor de longitud de cadena. En otro aspecto, la proteína que regula la longitud de cadena de los ácidos grasos producidos por el sistema PUFA PKS es un factor de longitud de cadena que dirige la síntesis de unidades C20.

En otra realización, el organismo expresa un sistema PUFA PKS no bacteriano que comprende una modificación genética en un dominio que codifica  $\beta$ -cetoacil-ACP sintasa (KS), donde la modificación altera la relación de ácidos grasos de cadena larga producidos por el sistema PUFA PKS en comparación con la ausencia de la modificación. En un aspecto de esta realización, la modificación se selecciona entre el grupo que consiste en una delección de todo o parte del dominio, una sustitución de un dominio homólogo de un organismo diferente en el lugar del dominio, y una mutación del dominio.

En una realización del método para producir una molécula bioactiva, el organismo produce un perfil de ácido graso poliinsaturado (PUFA) que difiere del organismo de origen natural sin una modificación genética.

Muchas otras modificaciones genéticas útiles para producir moléculas bioactivas serán evidentes para los expertos en la materia, dada la presente descripción, y se han analizado previamente en este documento otras diversas modificaciones. La presente invención contempla cualquier modificación genética relacionada con un sistema PUFA PKS descrito en este documento que provoque la producción de una molécula bioactiva deseada.

Las moléculas bioactivas, de acuerdo con la presente invención, incluyen cualquier molécula (compuestos, productos, etc.) que tenga una actividad biológica, y que pueda producirse por un sistema PKS que comprende al menos una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio funcional de un sistema PUFA PKS no bacteriano descrito en este documento. Dichas moléculas bioactivas pueden incluir, aunque sin limitación: un ácido graso poliinsaturado (PUFA), una formulación anti-inflamatoria, un agente quimioterapéutico, un excipiente activo, un fármaco para la osteoporosis, un anti-depresivo, un anti-convulsivo, un fármaco anti-*Helicobacter pylori*, un fármaco para el tratamiento de enfermedad neurodegenerativa, un fármaco para el tratamiento de enfermedad hepática degenerativa, un antibiótico, y una formulación para disminuir el nivel de colesterol. Una ventaja del sistema PUFA PKS no bacteriano de la presente invención es la capacidad de dicho sistema de introducir dobles enlaces carbono-carbono en configuración *cis*, y moléculas que incluyen un doble enlace cada tres carbonos. Esta capacidad puede utilizarse para producir una diversidad de compuestos.

Preferiblemente, los compuestos bioactivos de interés se producen por el microorganismo modificado genéticamente en una cantidad que es mayor de aproximadamente el 0,05 %, y preferiblemente mayor de aproximadamente el 0,1 %, y más preferiblemente mayor de aproximadamente el 0,25 %, y más preferiblemente mayor de aproximadamente el 0,5 %, y más preferiblemente mayor de aproximadamente el 0,75 %, y más preferiblemente mayor de aproximadamente el 1 %, y más preferiblemente mayor de aproximadamente el 2,5 %, y más preferiblemente mayor de aproximadamente el 5 %, y más preferiblemente mayor de aproximadamente el 10 %, y más preferiblemente mayor de aproximadamente el 15 %, e incluso más preferiblemente mayor de aproximadamente el 20 % del peso seco del microorganismo. Para compuestos lipídicos, preferiblemente, dichos compuestos se producen en una cantidad que es mayor de aproximadamente el 5 % del peso seco del microorganismo. Para otros compuestos bioactivos, dichos antibióticos o compuestos que se sintetizan en cantidades más pequeñas, aquellas cepas que procesan dichos compuestos en peso seco del microorganismo, se identifican de forma predecible como que contienen un nuevo sistema PKS del tipo descrito anteriormente. En algunas realizaciones, se secretan moléculas bioactivas (compuestos) particulares por el microorganismo, en lugar de acumularse. Por lo tanto, dichas moléculas bioactivas se recuperan generalmente del medio de cultivo y la concentración de molécula producida variará dependiendo del microorganismo y el tamaño del cultivo.

Otra realización más de la presente invención se refiere a un método para producir leche animal humanizada. Este método incluye las etapas de modificar genéticamente células productoras de leche de un animal no humano productor de leche con al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS de acuerdo con la invención. El sistema PUFA PKS es un sistema PUFA PKS no bacteriano.

Los métodos para modificar genéticamente una célula hospedadora y para producir un animal no humano productor de leche, modificado genéticamente, son conocidos en la técnica. Ejemplos de animales hospedadores a modificar incluyen ganado vacuno, ovejas, cerdos, cabras, yaks, etc., que son susceptibles a manipulación genética y clonación para rápida expansión de una población que expresa el transgén. Para animales, pueden adaptarse transgenes tipo PKS para su expresión en orgánulos, tejidos y fluidos corporales diana a través de modificación de las regiones reguladoras génicas. Es de particular interés la producción de PUFA en la leche materna del animal hospedador.

Los siguientes ejemplos se proporcionan con fines ilustrativos y no pretenden limitar el alcance de la presente invención.

## Ejemplos

### Ejemplo 1

El siguiente ejemplo describe el análisis adicional de secuencias relacionadas con PKS de *Schizochytrium*.

Los presentes inventores han secuenciado el ADN genómico que incluye la longitud completa de las tres fases de lectura abierta (Orf) en el sistema PUFA PKS de *Schizochytrium* usando los métodos generales resumidos en los Ejemplos 8 y 9 de la publicación PCT n.º WO 0042195 y la solicitud de Estados Unidos n.º de serie 09/231.899. Los dominios biológicamente activos en las proteínas PKS de *Schizochytrium* se representan gráficamente en la Fig. 1. La estructura de dominios del sistema PUFA PKS de *Schizochytrium* se describe más particularmente del siguiente modo.

#### Fase de lectura abierta A (OrfA):

La secuencia completa de nucleótidos para OrfA se representa en este documento como SEC ID N° 1. OrfA es una secuencia de 8730 nucleótidos (sin incluir el codón de parada) que codifica una secuencia de 2910 aminoácidos, representada en este documento como SEC ID N° 2. Dentro de OrfA hay doce dominios:

- (a) un dominio  $\beta$ -ceto acil-ACP sintasa (KS);
- (b) un dominio malonil-CoA:ACP aciltransferasa (MAT);
- (c) nueve dominios de proteína de transporte de acilo (ACP);
- (d) un dominio cetereductasa (KR).

Los dominios contenidos dentro de OrfA se han determinado en base a:

- (1) resultados de un análisis con el programa Pfam (Pfam es una base de datos de múltiples alineaciones de dominios proteicos o regiones proteicas conservadas. Las alineaciones representan alguna estructura evolutivamente conservada que tiene implicaciones para la función de la proteína. El perfil de modelos ocultos de Markov (perfil HMM) construido a partir de las alineaciones Pfam puede ser muy útil para reconocer de forma automática que una nueva proteína pertenece a una familia existente de proteínas, incluso si la homología es débil. A diferencia de los métodos convencionales de alineación por pares (por ejemplo, BLAST, FASTA), los HMM de Pfam dan sensibilidad con proteínas de múltiples dominios. La referencia proporcionada para la versión Pfam usada es: Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) *Nucleic Acids Research* 30(1):276-280); y/o
- (2) comparación de homología a sistemas PUFA-PKS bacterianos (por ejemplo, *Shewanella*) usando una búsqueda de homología BLAST 2.0 Basic BLAST usando blastp para búsquedas de aminoácidos con parámetros convencionales por defecto, donde la secuencia de consulta se filtra para regiones de baja complejidad por defecto (descrito en Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. y Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res.* 25:3389-3402, incorporado por referencia en este documento en su totalidad).

Se cree que las secuencias proporcionadas para dominios individuales contienen la longitud completa de la secuencia que codifica un dominio funcional, y puede contener secuencia flanqueante adicional dentro de la Orf.

#### ORFA-KS

El primer dominio en OrfA es un dominio KS, también mencionado en este documento como ORFA-KS. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1 y 40 de la SEC ID N° 1 (OrfA) hasta un punto final entre aproximadamente las posiciones 1428 y 1500 de la SEC ID N° 1. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFA-KS está representada en este documento como SEC ID N° 7 (posiciones 1-1500 de la SEC ID N° 1). La secuencia de aminoácidos que contiene el dominio KS abarca desde un punto de inicio entre aproximadamente las posiciones 1 y 14 de la SEC ID N° 2 (ORFA) hasta un punto final entre aproximadamente las posiciones 476 y 500 de la SEC ID N° 2. La secuencia de aminoácidos que contiene el dominio ORFA-KS está representada en este documento como SEC ID N° 8 (posiciones 1-500 de la SEC ID N° 2). Se aprecia que el dominio ORFA-KS contiene un motivo de sitio activo: DXAC\* (\*C<sub>215</sub> del sitio de unión a acilo).

#### ORFA-MAT

El segundo dominio en OrfA es un dominio MAT, también mencionado en este documento como ORFA-MAT. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1723 y 1798 de la SEC ID N° 1 (OrfA) hasta un punto final entre aproximadamente

las posiciones 2805 y 3000 de la SEC ID N° 1. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFA-MAT está representada en este documento como SEC ID N° 9 (posiciones 1723-3000 de la SEC ID N° 1). La secuencia de aminoácidos que contiene el dominio MAT abarca desde un punto de inicio entre aproximadamente las posiciones 575 y 600 de la SEC ID N° 2 (ORFA) hasta un punto final entre aproximadamente las posiciones 935 y 1000 de la SEC ID N° 2. La secuencia de aminoácidos que contiene el dominio ORFA-MAT está representada en este documento como SEC ID N° 10 (posiciones 575-1000 de la SEC ID N° 2). Se aprecia que el dominio ORFA-MAT contiene un motivo de sitio activo: GHS\*<sub>XG</sub> (\*S<sub>706</sub> del sitio de unión a acilo), representado en este documento como SEC ID N° 11.

#### 10 ORFA-ACP n.º 1-9

Los dominios 3-11 de OrfA son nueve dominios ACP en tándem, también mencionados en este documento como ORFA-ACP (el primer dominio en la secuencia es ORFA-ACP1, el segundo dominio es ORFA-ACP2, el tercer dominio es ORFA-ACP3, etc.). El primer dominio ACP, ORFA-ACP1, está contenido dentro de la secuencia de nucleótidos que abarca desde aproximadamente la posición 3343 hasta aproximadamente la posición 3600 de la SEC ID N° 1 (OrfA). La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFA-ACP1 está representada en este documento como SEC ID N° 12 (posiciones 3343-3600 de la SEC ID N° 1). La secuencia de aminoácidos que contiene el primer dominio ACP abarca desde aproximadamente la posición 1115 hasta aproximadamente la posición 1200 de la SEC ID N° 2. La secuencia de aminoácidos que contiene el dominio ORFA-ACP1 está representada en este documento como SEC ID N° 13 (posiciones 1115-1200 de la SEC ID N° 2). Se aprecia que el dominio ORFA-ACP1 contiene un motivo de sitio activo: LGIDS\* (\*S<sub>1157</sub> del motivo de unión a panteteína), representado en este documento por la SEC ID N° 14. Las secuencias de nucleótidos y aminoácidos de los nueve dominios ACP están muy conservados y, por lo tanto, la secuencia para cada dominio no está representada en este documento por un identificador de secuencia individual. Sin embargo, en base a esta información, un experto en la materia puede determinar fácilmente la secuencia para cada uno de los otros ocho dominios ACP. El intervalo repetido para los nueve dominios es de aproximadamente 110 a aproximadamente 330 nucleótidos de la SEC ID N° 1.

Los nueve dominios ACP juntos abarcan una región de OrfA desde aproximadamente la posición 3283 hasta aproximadamente la posición 6288 de la SEC ID N° 1, que corresponde a las posiciones de aminoácido desde aproximadamente 1095 hasta aproximadamente 2096 de la SEC ID N° 2. Esta región incluye los segmentos enlazadores entre dominios ACP individuales. Cada uno de los nueve dominios ACP contiene un motivo de unión a panteteína LGIDS\* (representado en este documento por la SEC ID N° 14), donde \* es la S del sitio de unión a panteteína S. En cada extremo de la región del dominio ACP y entre cada dominio ACP hay una región que está muy enriquecida en prolina (P) y alanina (A), que se cree que es una región enlazadora. Por ejemplo, entre los dominios ACP 1 y 2 está la secuencia: APAPVKAAAPVASAPAPA, representada en este documento como SEC ID N° 15.

#### 40 ORFA-KR

El dominio 12 en OrfA es un dominio KR, también mencionado en este documento como ORFA-KR. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio desde aproximadamente la posición 6598 de la SEC ID N° 1 hasta un punto final desde aproximadamente la posición 8730 de la SEC ID N° 1. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFA-KR está representada en este documento como SEC ID N° 17 (posiciones 6598-8730 de la SEC ID N° 1). La secuencia de aminoácidos que contiene el dominio KR abarca desde un punto de inicio de aproximadamente la posición 2200 de la SEC ID N° 2 (ORFA) hasta un punto final de aproximadamente la posición 2910 de la SEC ID N° 2. La secuencia de aminoácidos que contiene el dominio ORFA-KR está representada en este documento como SEC ID N° 18 (posiciones 2200-2910 de la SEC ID N° 2). Dentro del dominio KR hay una región central con homología a las aldehído de cadena corta-deshidrogenasas (KR es un miembro de esta familia). Esta región central abarca desde aproximadamente la posición 7198 hasta aproximadamente la posición 7500 de la SEC ID N° 1, que corresponde a las posiciones de aminoácido 2400-2500 de la SEC ID N° 2.

#### 55 Fase de lectura abierta B (OrfB):

La secuencia completa de nucleótidos para OrfB está representada en este documento como SEC ID N° 3. OrfB es una secuencia de 6177 nucleótidos (sin incluir el codón de parada) que codifica una secuencia de 2059 aminoácidos, representada en este documento como SEC ID N° 4. Dentro de la OrfB hay cuatro dominios:

- 60 (a) dominio β-ceto acil-ACP sintasa (KS);
- (b) un dominio de factor de longitud de cadena (CLF);
- (c) un dominio acil transferasa (AT);
- (d) un dominio enoil ACP-reductasa (ER).

65 Los dominios contenidos dentro de ORFB se han determinado en base a: (1) resultados de un análisis con el programa Pfam, descrito anteriormente; y/o (2) comparación de homología a sistemas PUFA-PKS bacterianos (por

ejemplo, *Shewanella*) usando una búsqueda de homología con BLAST 2.0 Basic BLAST, también descrito anteriormente. Se cree que las secuencias proporcionadas para dominios individuales contienen la longitud completa de la secuencia que codifica un dominio funcional, y puede contener secuencia flanqueante adicional dentro de la Orf.

5

#### ORFB-KS

El primer dominio en OrfB es un dominio KS, también mencionado en este documento como ORFB-KS. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1 y 43 de la SEC ID N° 3 (OrfB) hasta un punto final entre aproximadamente las posiciones 1332 y 1350 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFB-KS está representada en este documento como SEC ID N° 19 (posiciones 1-1350 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio KS abarca desde un punto de inicio entre aproximadamente las posiciones 1 y 15 de la SEC ID N° 4 (ORFB) hasta un punto final entre aproximadamente las posiciones 444 y 450 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-KS está representada en este documento como SEC ID N° 20 (posiciones 1-450 de la SEC ID N° 4). Se aprecia que el dominio ORFB-KS contiene un motivo de sitio activo: DXAC\* (\*C<sub>196</sub> del sitio de unión a acilo).

10

15

#### ORFB-CLF

El segundo dominio en OrfB es un dominio CLF, también mencionado en este documento como ORFB-CLF. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1378 y 1402 de la SEC ID N° 3 (OrfB) hasta un punto final entre aproximadamente las posiciones 2682 y 2700 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFB-CLF está representada en este documento como SEC ID N° 21 (posiciones 1378-2700 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio CLF abarca desde un punto de inicio entre aproximadamente las posiciones 460 y 468 de la SEC ID N° 4 (ORFB) hasta un punto final entre aproximadamente las posiciones 894 y 900 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-CLF está representada en este documento como SEC ID N° 22 (posiciones 460-900 de la SEC ID N° 4). Se aprecia que el dominio ORFB-CLF contiene un motivo de sitio activo de KS son la cisteína de unión a acilo.

20

25

30

#### ORFB-AT

El tercer dominio en OrfB es un dominio AT, también mencionado en este documento como ORFB-AT. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 2701 y 3598 de la SEC ID N° 3 (OrfB) hasta un punto final entre aproximadamente las posiciones 3975 y 4200 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFB-AT está representada en este documento como SEC ID N° 23 (posiciones 2701-4200 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio AT abarca desde un punto de inicio entre aproximadamente las posiciones 901 y 1200 de la SEC ID N° 4 (ORFB) hasta un punto final entre aproximadamente las posiciones 1325 y 1400 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-AT está representada en este documento como SEC ID N° 24 (posiciones 901-1400 de la SEC ID N° 4). Se aprecia que el dominio ORFB-AT contiene un motivo de sitio activo de AT de GxS\*xG (\*S<sub>1140</sub> del sitio de unión a acilo).

35

40

45

#### ORFB-ER

El cuarto dominio en OrfB es un dominio ER, también mencionado en este documento como ORFB-ER. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio de aproximadamente la posición 4648 de la SEC ID N° 3 (OrfB) hasta un punto final de aproximadamente la posición 6177 de la SEC ID N° 3. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFB-ER está representada en este documento como SEC ID N° 25 (posiciones 4648-6177 de la SEC ID N° 3). La secuencia de aminoácidos que contiene el dominio ER abarca desde un punto de inicio de aproximadamente la posición 1550 de la SEC ID N° 4 (ORFB) hasta un punto final de aproximadamente la posición 2059 de la SEC ID N° 4. La secuencia de aminoácidos que contiene el dominio ORFB-ER está representada en este documento como SEC ID N° 26 (posiciones 1550-2059 de la SEC ID N° 4).

50

55

#### Fase de lectura abierta C (OrfC):

La secuencia completa de nucleótidos para OrfC está representada en este documento como SEC ID N° 5. OrfC es una secuencia de 4509 nucleótidos (sin incluir el codón de parada) que codifica una secuencia de 1503 aminoácidos, representada en este documento como SEC ID N° 6. Dentro de OrfC hay tres dominios:

60

- (a) dos dominios β-hidroxi acil-ACP deshidrasa (DH) tipo FabA;
- (b) un dominio enoil ACP-reductasa (ER).

65

Los dominios contenidos dentro de ORFC se han determinado en base a: (1) resultados de un análisis con el programa Pfam, descrito anteriormente; y/o (2) comparación de homología a sistemas PUFA-PKS bacterianos (por ejemplo, *Shewanella*) usando una búsqueda de homología con BLAST 2.0 Basic BLAST, también descrito anteriormente. Se cree que las secuencias proporcionadas para dominios individuales contienen la longitud completa de la secuencia que codifica un dominio funcional, y puede contener secuencia flanqueante adicional dentro de la Orf.

#### ORFC-DH1

El primer dominio en OrfC es un dominio DH, también mencionado en este documento como ORFC-DH1. Éste es uno de los dos dominios DH en OrfC, y por lo tanto se denomina DH1. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1 y 778 de la SEC ID N° 5 (OrfC) hasta un punto final entre aproximadamente las posiciones 1233 y 1350 de la SEC ID N° 5. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFC-DH1 está representada en este documento como SEC ID N° 27 (posiciones 1-1350 de la SEC ID N° 5). La secuencia de aminoácidos que contiene el dominio DH1 abarca desde un punto de inicio entre aproximadamente las posiciones 1 y 260 de la SEC ID N° 6 (ORFC) hasta un punto final entre aproximadamente las posiciones 411 y 450 de la SEC ID N° 6. La secuencia de aminoácidos que contiene el dominio ORFC-DH1 está representada en este documento como SEC ID N° 28 (posiciones 1-450 de la SEC ID N° 6).

#### ORFC-DH2

El segundo dominio en OrfC es un dominio DH, también mencionado en este documento como ORFC-DH2. Éste es el segundo de los dos dominios DH en OrfC, y por lo tanto se denomina DH2. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio entre aproximadamente las posiciones 1351 y 2437 de la SEC ID N° 5 (OrfC) hasta un punto final entre aproximadamente las posiciones 2607 y 2850 de la SEC ID N° 5. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFC-DH2 está representada en este documento como SEC ID N° 29 (posiciones 1351-2850 de la SEC ID N° 5). La secuencia de aminoácidos que contiene el dominio DH2 abarca desde un punto de inicio entre aproximadamente las posiciones 451 y 813 de la SEC ID N° 6 (ORFC) hasta un punto final entre aproximadamente las posiciones 869 y 950 de la SEC ID N° 6. La secuencia de aminoácidos que contiene el dominio ORFC-DH2 está representada en este documento como SEC ID N° 30 (posiciones 451-950 de la SEC ID N° 6).

#### ORFC-ER

El tercer dominio en OrfC es un dominio ER, también mencionado en este documento como ORFC-ER. Este dominio está contenido dentro de la secuencia de nucleótidos que abarca desde un punto de inicio de aproximadamente la posición 2998 de la SEC ID N° 5 (OrfC) hasta un punto final de aproximadamente la posición 4509 de la SEC ID N° 5. La secuencia de nucleótidos que contiene la secuencia que codifica el dominio ORFC-ER está representada en este documento como SEC ID N° 31 (posiciones 2998-4509 de la SEC ID N° 5). La secuencia de aminoácidos que contiene el dominio ER abarca desde un punto de inicio de aproximadamente la posición 1000 de la SEC ID N° 6 (ORFC) hasta un punto final de aproximadamente la posición 1502 de la SEC ID N° 6. La secuencia de aminoácidos que contiene el dominio ORFC-ER está representada en este documento como SEC ID N° 32 (posiciones 1000-1502 de la SEC ID N° 6).

#### Ejemplo 2

El siguiente ejemplo describe el uso del proceso de selección de la presente invención para identificar otros tres organismos no bacterianos que comprenden un sistema PUFA PKS de acuerdo con la presente invención.

Se cultivó *Thraustochytrium* sp. 23B (ATCC 20892) de acuerdo con el método de selección descrito en solicitud provisional de Estados Unidos n.º de serie 60/298.796 y como se describe en detalle en este documento.

El filtro biorracional (usando cultivos en matraz de agitación) desarrollado para detectar microorganismos que contienen sistemas PKS que producen PUFA es el siguiente:

Se colocan dos ml de un cultivo de la cepa/microorganismo a ensayar en matraz de agitación con separadores de 250 ml con 50 ml medio de cultivo (tratamiento aeróbico) y se colocan otros 2 ml de cultivo de la misma cepa en un matraz de agitación sin separadores de 250 ml con 200 ml de medio de cultivo (tratamiento anóxico). Ambos matraces se colocan en una mesa oscilatoria a 200 rpm. Después de 48-72 h de tiempo de cultivo, se recogen los cultivos por centrifugación y se analizan las células para ésteres metílicos de ácidos grasos mediante cromatografía de gases para determinar los siguientes datos para cada cultivo: (1) perfil de ácidos grasos; (2) contenido de PUFA; (3) contenido de grasa (estimada como la cantidad total de ácidos grasos (TFA)).

Estos datos se analizan después haciendo las siguientes cinco preguntas:

Criterios de selección: Matraz de bajo O<sub>2</sub>/anóxico frente al matraz aeróbico (sí/no)



- (1) ¿El DHA (u otro contenido de PUFA) (como % de FAME) permanecía aproximadamente igual o preferiblemente aumentaba en el cultivo de bajo oxígeno en comparación con el cultivo aeróbico?  
 (2) ¿Es C14:0 + C16:0 + C16:1 mayor de aproximadamente el 40 % de TFA en el cultivo anóxico?  
 (3) ¿Hay muy poco (>1 % como FAME) o ningún precursor (C18:3n-3 +C18:2n-6+C18:3n-6) para la ruta convencional de elongasa/desaturasa dependiente de oxígeno en el cultivo anóxico?  
 (4) ¿El contenido de grasa (como la cantidad total de ácidos grasos/peso seco de la célula) aumentaba en el cultivo de bajo oxígeno en comparación con el cultivo aeróbico?  
 (5) ¿El DHA (u otro contenido de PUFA) estaba aumentado como % en peso seco de la célula en el cultivo de bajo oxígeno en comparación con el cultivo aeróbico?

Si las tres primeras preguntas se contestan sí, existe un buen indicio de que la cepa contiene un sistema genético PKS para fabricar PUFA de cadena larga. Cuantas más preguntas se contesten sí (preferiblemente las tres primeras preguntas deben contestarse sí), más fuerte será el indicio de que la cepa contiene dicho sistema genético PKS. Si las cinco preguntas se contestan sí, entonces existe un indicio muy fuerte de que la cepa contiene un sistema genético PKS para fabricar PUFA de cadena larga.

Siguiente el método resumido anteriormente, se usó un vial congelado de *Thraustochytrium* sp. 23B (ATCC 20892) para inocular un matraz de agitación de 250 ml que contenía 50 ml de medio RCA. El cultivo se agitó en una mesa oscilatoria (200 rpm) durante 72 h a 25 °C. El medio RCA contiene lo siguiente:

<u>Medio RCA</u>	
Agua	1000 ml
Sales marinas Reef Crystals®	40 g/l
Glucosa	20 g/l
Glutamato monosódico (MSG)	20 g/l
Extracto de levadura	1 g/l
Metales PII*	5 ml/l
Mezcla de vitaminas*	1 ml/l
pH	7,0

\*La mezcla de metales PII y la mezcla de vitaminas son iguales a las resumidas en la patente de Estados Unidos n.º 5.130.742.

Entonces se usaron 25 ml del cultivo de 72 h de antigüedad para inocular otro matraz de agitación de 250 ml que contenía 50 ml de medio RCA bajo en nitrógeno (10 g/l de MSG en lugar de 20 g/l) y los otros 25 ml de cultivo se usaron para inocular un matraz de agitación de 250 ml que contenía 175 ml de medio RCA bajo en nitrógeno. Los dos matraces después se colocaron en una mesa oscilatoria (200 rpm) durante 72 h a 25 °C. Las células después se recogieron mediante centrifugación y se secaron por liofilización. Las células secadas se analizaron para el contenido de grasa y el perfil y contenido de ácidos grasos usando procedimientos convencionales de cromatografía de gases (tales como los resumidos en el documento US 5.130.742).

Los resultados de la selección para *Thraustochytrium* 23B fueron los siguientes:

¿Aumentaba DHA como % de FAME?	Sí (38->44 %)
¿C14:0 + C16:0 + C16:1 mayor que aproximadamente el 40 % de TFA?	Sí (44 %)
¿Nada de C18:3(n-3) o C18:3(n-6)?	Sí (0 %)
¿Aumentaba el contenido de grasa?	Sí (aumento de 2 veces)
¿Aumentaba DHA (u otro contenido de PUFA)?	Sí (aumento de 2,3 veces)

Los resultados, especialmente el aumento significativo en el contenido de DHA (como % de FAME) en condiciones de bajo oxígeno, indican fuertemente la presencia de un sistema PKS que produce PUFA en la cepa de *Thraustochytrium*.

Para proporcionar datos adicionales que confirmen la presencia de un sistema PUFA PKS, se realizó transferencia de Southern de *Thraustochytrium* 23B usando sondas de PKS de *Schizochytrium* cepa 20888, una cepa que ya se ha determinado que contiene un sistema PKS que produce PUFA (es decir, SEC ID N° 1-32 descritas anteriormente). Se detectaron fragmentos de ADN genómico de *Thraustochytrium* 23B que son homólogos a sondas de hibridación de genes de síntesis de PUFA por PKS usando la técnica de transferencia de Southern. Se digirió el ADN genómico de *Thraustochytrium* 23B con endonucleasas de restricción *ClaI* o *KpnI*, se separó por electroforesis en gel de agarosa (0,7 % de agarosa, en tampón convencional Tris-Acetato-EDTA), y se transfirió a una membrana Nytran Supercharge de Schleicher & Schuell por transferencia capilar. Se usaron dos sondas de hibridación marcadas con digoxigenina - una específica para la región enoil reductasa (ER) de Orf B de PKS de *Schizochytrium* (nucleótidos 5012-5511 de Orf B; SEC ID N° 3), y la otra específica para una región conservada en el inicio de la Orf C de PKS de *Schizochytrium* (nucleótidos 76-549 de OrfC; SEC ID N° 5).

La sonda de OrfB-ER detectó un fragmento Clal de aproximadamente 13 kb y un fragmento KpnI de aproximadamente 3,6 kb en el ADN genómico de *Thraustochytrium* 23B. La sonda de OrfC detectó un fragmento Clal de aproximadamente 7,5 kb y un fragmento KpnI de aproximadamente 4,6 kb en el ADN genómico de *Thraustochytrium* 23B.

Finalmente, se seleccionó una biblioteca genómica recombinante, que consistía en fragmentos de ADN del ADN genómico de *Thraustochytrium* 23B insertado en vector lambda FIX II (Stratagene), usando sondas marcadas con digoxigenina correspondientes a los siguientes segmentos de genes de PUFA-PKS de *Schizochytrium* 20888: nucleótidos 7385-7879 de Orf A (SEC ID N° 1), nucleótidos 5012-5511 de Orf B (SEC ID N° 3), y nucleótidos 76-549 de Orf C (SEC ID N° 5). Cada una de estas sondas detectó placas positivas de la biblioteca de *Thraustochytrium* 23B, lo que indica homología extensiva entre los genes de PUFA-PKS de *Schizochytrium* y los genes de *Thraustochytrium* 23B.

En resumen, estos resultados demuestran que el ADN genómico de *Thraustochytrium* 23B contiene secuencias que son homólogas a genes de PKS de *Schizochytrium* 20888.

Este microorganismo traustochytridiano está incluido en este documento como una fuente adicional de estos genes para uso en las realizaciones anteriores.

*Thraustochytrium* 23B (ATCC 20892) es significativamente diferente de *Schizochytrium* sp. (ATCC 20888) en su perfil de ácidos grasos. *Thraustochytrium* 23B puede tener relaciones DHA:DPA(n-6) tan elevadas como de 14:1 en comparación con solamente 2-3:1 en *Schizochytrium* (ATCC 20888). *Thraustochytrium* 23B también puede tener niveles mayores de C20:5(n-3). El análisis de los dominios en el sistema PUFA PKS de *Thraustochytrium* 23B en comparación con el sistema PUFA PKS conocido de *Schizochytrium* debe proporcionarnos información clave sobre el modo de modificar estos dominios para influir en la relación y tipos de PUFA producidos usando estos sistemas.

El método de selección descrito anteriormente se ha utilizado para identificar otras cepas candidatas potenciales que contengan un sistema PUFA PKS. Dos cepas adicionales que los presentes inventores han identificado que tienen sistemas PUFA PKS son *Schizochytrium limacium* (SR21) Honda y Yokochi (IF032693) y *Ulkenia* (BP-5601). Ambas se seleccionaron como anteriormente, pero en medio N2 (glucosa: 60 g/l; KH<sub>2</sub>PO<sub>4</sub>: 4,0 g/l; extracto de levadura: 1,0 g/l; agua de remojo de maíz: 1 ml/l; NH<sub>4</sub>NO<sub>3</sub>: 1,0 g/l; sales marinas artificiales (Reef Crystals): 20 g/l; todas las concentraciones anteriores mezcladas en agua desionizada). Para ambas cepas de *Schizochytrium* y *Ulkenia*, las respuestas a las primeras tres preguntas de filtro analizadas anteriormente para *Thraustochytrium* 23B fueron sí (*Schizochytrium* - DHA % FAME 32->41 % aeróbico frente a anóxico, 58 % 14:0/16:0/16:1, 0 % precursores) y (*Ulkenia* - DHA % FAME 28->44 % aeróbico frente a anóxico, 63 % 14:0/16:0/16:1, 0 % precursores), lo que indica que estas cepas son buenos candidatos para que contengan un sistema PUFA PKS. Se obtuvieron respuestas negativas para las dos preguntas finales para cada cepa: grasa disminuida del 61 % en peso seco al 22 % en peso seco, y DHA del 21-9 % en peso seco en *S. limacium* y grasa disminuida del 59 al 21 % en peso seco en *Ulkenia* y DHA del 16 % al 9 % en peso seco. Estos microorganismos traustochytridios también se reivindican en este documento como fuentes adicionales de los genes para uso en las realizaciones anteriores.

### Ejemplo 3

El siguiente ejemplo demuestra que la síntesis de DHA y DPA en *Schizochytrium* no implica desaturasas unidas a membrana o enzimas de elongación de ácidos grasos como las descritas para otros eucariotas (Parker-Barnes *et al.*, 2000, *supra*; Shanklin *et al.*, 1998, *supra*).

*Schizochytrium* acumula grandes cantidades de triacilgliceroles ricos en DHA y ácido docosapentaenoico (DPA; 22:5ω6); por ejemplo, 30 % de DHA + DPA en peso seco. En eucariotas que sintetizan PUFA de 20 y 22 carbonos mediante una ruta de elongación/desaturación, las combinaciones de intermedios de 18, 20 y 22 carbonos son relativamente grandes de modo que experimentos de marcaje *in vivo* usando [<sup>14</sup>C]-acetato revelan una clara cinética precursor-producto para los intermedios predichos. Además, los intermedios radiomarcados proporcionados de forma exógena a dichos organismos se convierten en los productos PUFA finales.

Se suministró [1-<sup>14</sup>C]acetato a un cultivo de 2 días de antigüedad como un único pulso a tiempo cero. Después se recogieron muestras de células por centrifugación y se extrajeron los lípidos. Además, se estimó la captación de [1-<sup>14</sup>C]acetato por las células midiendo la radiactividad de la muestra antes y después de la centrifugación. Se separaron los ésteres metílicos del ácido graso derivados de los lípidos celulares totales por AgNO<sub>3</sub>-TLC (disolvente, hexano:éter dietílico:ácido acético, 70:30:2 en volumen). La identidad de las bandas de ácido graso se verificó por cromatografía de gases, y se midió la radiactividad en ellas por recuento de centelleo. Los resultados mostraron que se captaba rápidamente [1-<sup>14</sup>C]-acetato por células de *Schizochytrium* y se incorporaba en los ácidos grasos, pero en el tiempo más corto de marcaje (1 min) DHA contenía un 31 % del marcador recuperado en los ácidos grasos y este porcentaje permaneció esencialmente inalterado durante los 10-15 min de incorporación de [1-<sup>14</sup>C]-acetato y las posteriores 24 horas de crecimiento del cultivo (datos no mostrados). Asimismo, DPA representaba el 10 % del marcador en todo el experimento. No existen evidencias de una relación precursor-producto entre ácidos grasos de 16 o 18 carbonos y los ácidos grasos poliinsaturados de 22 carbonos. Estos resultados son coherentes con la rápida

síntesis de DHA a partir de [<sup>14</sup>C]-acetato que implica combinaciones muy pequeñas (posiblemente unidas a enzimas) de intermedios.

5 A continuación, las células se alteraron en tampón fosfato 100 mM (pH 7,2), que contenía DTT 2 mM, EDTA 2 mM, y glicerol al 10 %, agitando con vórtice con perlas de vidrio. El homogeneizado sin células se centrifugó a 100.000 g durante 1 hora. Se incubaron alícuotas equivalentes de las fracciones de homogeneizado total, sedimento (sedimento H-S), y sobrenadante (sobrenadante H-S) en tampón de homogenización suplementado con acetil-CoA 20 μM, [1-<sup>14</sup>C]malonil-CoA 100 μM (0,9 Gbq/mol), NADH 2 mM, y NADPH 2 mM durante 60 min a 25 °C. Se extrajeron los ensayos y se prepararon y separaron los ésteres metílicos de ácido graso como se ha descrito  
10 anteriormente antes de la detección de la radiactividad con un Instantimager (Packard Instruments, Meriden, CT). Los resultados mostraron que un homogeneizado sin células derivado de cultivos de *Schizochytrium* incorporaba [1-<sup>14</sup>C]-malonil-CoA en DHA, DPA, y ácidos grasos saturados (datos no mostrados). Se retuvieron las mismas actividades biosintéticas por una fracción de sobrenadante a 100.000 xg pero no estuvieron presentes en el sedimento de membrana. Estos datos contrastan con los obtenidos durante los ensayos de las enzimas bacterianas  
15 (véase Metz *et al.*, 2001, *supra*) y pueden indicar el uso de una molécula aceptora de acilo diferente (soluble). Por tanto, la síntesis de DHA y DPA en *Schizochytrium* no implica desaturasas unidas a membrana o enzimas de elongación de ácidos grasos como las descritas para otros eucariotas.

Aunque se han descrito en detalle diversas realizaciones de la presente invención definida en las reivindicaciones, es evidente para los expertos en la materia que encontrarán modificaciones y adaptaciones de esas realizaciones. Debe entenderse expresamente, sin embargo, que dichas modificaciones y adaptaciones están dentro del alcance de la presente invención, como se describe en las siguientes reivindicaciones.

#### LISTADO DE SECUENCIAS

25

<110> Metz, James  
Barclay, William  
Flatt, James  
Kuner, Jerry

30

<120> SISTEMAS DE PUFA POLICÉTIDO SINTASA Y USOS DE LOS MISMOS

<130> 2997-29-PCT

35

<150> 60/284.066  
<151> 16-04-2001

<150> 60/298.796  
<151> 15-06-2001

40

<150> 60/323.269  
<151> 18-09-2001

<160> 37

45

<170> PatentIn versión 3.1

<210> 1

<211> 8730

50

<212> ADN

<213> *Schizochytrium sp.*

<220>

<221> CDS

55

<222> (1)..(8730)

<223>

<400> 1

# ES 2 567 305 T3

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

ES 2 567 305 T3

Gly	Met	Pro	Glu	Glu	Asp	Val	Lys	Val	Ala	Val	Glu	Lys	Tyr	Lys	Ala		
				165					170					175			
aac	ttc	ccc	gag	tgg	cgc	ctc	gac	tcc	ttc	cct	ggc	ttc	ctc	ggc	aac		576
Asn	Phe	Pro	Glu	Trp	Arg	Leu	Asp	Ser	Phe	Pro	Gly	Phe	Leu	Gly	Asn		
			180						185				190				
gtc	acc	gcc	ggt	cgc	tgc	acc	aac	acc	ttc	aac	ctc	gac	ggc	atg	aac		624
Val	Thr	Ala	Gly	Arg	Cys	Thr	Asn	Thr	Phe	Asn	Leu	Asp	Gly	Met	Asn		
		195					200					205					
tgc	gtt	gtc	gac	gcc	gca	tgc	gcc	tcg	tcc	ctc	atc	gcc	gtc	aag	gtc		672
Cys	Val	Val	Asp	Ala	Ala	Cys	Ala	Ser	Ser	Leu	Ile	Ala	Val	Lys	Val		
	210					215					220						
gcc	atc	gac	gag	ctg	ctc	tac	ggg	gac	tgc	gac	atg	atg	gtc	acc	ggg		720
Ala	Ile	Asp	Glu	Leu	Leu	Tyr	Gly	Asp	Cys	Asp	Met	Met	Val	Thr	Gly		
	225				230					235					240		
gcc	acc	tgc	acg	gat	aac	tcc	atc	ggc	atg	tac	atg	gcc	ttc	tcc	aag		768
Ala	Thr	Cys	Thr	Asp	Asn	Ser	Ile	Gly	Met	Tyr	Met	Ala	Phe	Ser	Lys		
			245					250						255			
acc	ccc	gtg	ttc	tcc	acg	gac	ccc	agc	gtg	cgc	gcc	tac	gac	gaa	aag		816
Thr	Pro	Val	Phe	Ser	Thr	Asp	Pro	Ser	Val	Arg	Ala	Tyr	Asp	Glu	Lys		
			260					265					270				
aca	aag	ggc	atg	ctc	atc	ggc	gag	ggc	tcc	gcc	atg	ctc	gtc	ctc	aag		864
Thr	Lys	Gly	Met	Leu	Ile	Gly	Glu	Gly	Ser	Ala	Met	Leu	Val	Leu	Lys		
	275						280					285					
cgc	tac	gcc	gac	gcc	gtc	cgc	gac	ggc	gat	gag	atc	cac	gct	ggt	att		912
Arg	Tyr	Ala	Asp	Ala	Val	Arg	Asp	Gly	Asp	Glu	Ile	His	Ala	Val	Ile		
	290					295					300						
cgc	ggc	tgc	gcc	tcc	tcc	agt	gat	ggc	aag	gcc	gcc	ggc	atc	tac	acg		960
Arg	Gly	Cys	Ala	Ser	Ser	Asp	Gly	Lys	Ala	Ala	Gly	Ile	Tyr	Thr			
	305				310				315					320			
ccc	acc	att	tcg	ggc	cag	gag	gag	gcc	ctc	cgc	cgc	gcc	tac	aac	cgc		1008
Pro	Thr	Ile	Ser	Gly	Gln	Glu	Glu	Ala	Leu	Arg	Arg	Ala	Tyr	Asn	Arg		
			325						330					335			
gcc	tgt	gtc	gac	ccg	gcc	acc	gtc	act	ctc	gtc	gag	ggg	cac	ggc	acc		1056
Ala	Cys	Val	Asp	Pro	Ala	Thr	Val	Thr	Leu	Val	Glu	Gly	His	Gly	Thr		
			340					345					350				
ggg	act	ccc	ggt	ggc	gac	cgc	atc	gag	ctc	acc	gcc	ttg	cgc	aac	ctc		1104
Gly	Thr	Pro	Val	Gly	Asp	Arg	Ile	Glu	Leu	Thr	Ala	Leu	Arg	Asn	Leu		
	355						360					365					
ttt	gac	aag	gcc	tac	ggc	gag	ggc	aac	acc	gaa	aag	gtc	gct	gtg	ggc		1152
Phe	Asp	Lys	Ala	Tyr	Gly	Glu	Gly	Asn	Thr	Glu	Lys	Val	Ala	Val	Gly		
	370				375						380						
agc	atc	aag	tcc	agc	atc	ggc	cat	ctc	aag	gcc	gtc	gcc	ggg	ctc	gcc		1200
Ser	Ile	Lys	Ser	Ser	Ile	Gly	His	Leu	Lys	Ala	Val	Ala	Gly	Leu	Ala		
	385				390				395					400			
ggg	atg	atc	aag	gtc	atc	atg	gcg	ctc	aag	cac	aag	act	ctc	ccg	ggc		1248
Gly	Met	Ile	Lys	Val	Ile	Met	Ala	Leu	Lys	His	Lys	Thr	Leu	Pro	Gly		
			405					410						415			
acc	atc	aac	gtc	gac	aac	cca	ccc	aac	ctc	tac	gac	aac	acg	ccc	atc		1296
Thr	Ile	Asn	Val	Asp	Asn	Pro	Pro	Asn	Leu	Tyr	Asp	Asn	Thr	Pro	Ile		
			420					425					430				
aac	gag	tcc	tcg	ctc	tac	att	aac	acc	atg	aac	cgc	ccc	tgg	ttc	cgg		1344
Asn	Glu	Ser	Ser	Leu	Tyr	Ile	Asn	Thr	Met	Asn	Arg	Pro	Trp	Phe	Pro		
	435						440					445					
ccc	cct	ggt	gtg	ccc	cgc	cgc	gcc	ggc	att	tcg	agc	ttt	ggc	ttt	ggg		1392
Pro	Pro	Gly	Val	Pro	Arg	Arg	Ala	Gly	Ile	Ser	Ser	Phe	Gly	Phe	Gly		
	450					455					460						
ggc	gcc	aac	tac	cac	gcc	gtc	ctc	gag	gag	gcc	gag	ccc	gag	cac	acg		1440
Gly	Ala	Asn	Tyr	His	Ala	Val	Leu	Glu	Glu	Ala	Glu	Pro	Glu	His	Thr		
	465				470					475					480		
acc	cgc	tac	cgc	ctc	aac	aag	cgc	ccg	cag	ccc	gtg	ctc	atg	atg	gcc		1488
Thr	Ala	Tyr	Arg	Leu	Asn	Lys	Arg	Pro	Gln	Pro	Val	Leu	Met	Met	Ala		
			485					490						495			
gcc	acg	ccc	gcg	gcc	ctc	cag	tcg	ctc	tgc	gag	gcc	cag	ctc	aag	gag		1536

ES 2 567 305 T3

Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu	
500	505
ttc gag gcc gcc atc aag gag aac gag acc gtc aag aac acc gcc tac	1584
Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr	
515	520
atc aag tgc gtc aag ttc ggc gag cag ttc aaa ttc cct ggc tcc atc	1632
Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile	
530	535
ccg gcc aca aac gcg cgc ctc ggc ttc ctc gtc aag gat gct gag gat	1680
Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp	
545	550
gcc tgc tcc acc ctc cgt gcc atc tgc gcc caa ttc gcc aag gat gtc	1728
Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val	
565	570
acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc cgc gcc	1776
Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala	
580	585
aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc gcc cag	1824
Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln	
595	600
ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac tgg ccc	1872
Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro	
610	615
cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc aag gtc	1920
Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val	
625	630
gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc tac ccg	1968
Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro	
645	650
cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag atc tcc	2016
Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser	
660	665
ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc ggt gcc	2064
Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala	
675	680
ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc gcc ggc	2112
Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly	
690	695
cat tcg ctc ggt gag ttc gcc ctc tac gcc gcg ggc tgc gtc gac	2160
His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp	
705	710
cgc gac gag ctc ttt gag ctt gtc tgc cgc gcg gcc cgc atc atg ggc	2208
Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly	
725	730
ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc gtc att	2256
Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile	
740	745
ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc tgg ctc	2304
Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu	
755	760
ggc aac tcc aac tcg cct tcg cag acc gtc atc acc gcc tcc gtc gaa	2352
Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu	
770	775
ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc cgc gtc	2400
Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val	
785	790
gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg gag aac	2448
Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn	
805	810
gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc cgc acc	2496
Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr	
820	825
ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag acc tac	2544

ES 2 567 305 T3

Pro	Lys	Ala	Glu	Thr	Lys	Leu	Phe	Ser	Asn	Val	Ser	Gly	Glu	Thr	Tyr		
		835					840					845					
ccc	acg	gac	gcc	cgc	gag	atg	ctt	acg	cag	cac	atg	acc	agc	agc	gtc	2592	
Pro	Thr	Asp	Ala	Arg	Glu	Met	Leu	Thr	Gln	His	Met	Thr	Ser	Ser	Val		
		850					855					860					
aag	ttc	ctc	acc	cag	gtc	cgc	aac	atg	cac	cag	gcc	ggt	gcg	cgc	atc	2640	
Lys	Phe	Leu	Thr	Gln	Val	Arg	Asn	Met	His	Gln	Ala	Gly	Ala	Arg	Ile		
		865				870					875				880		
ttt	gtc	gag	ttc	gga	ccc	aag	cag	gtg	ctc	tcc	aag	ctt	gtc	tcc	gag	2688	
Phe	Val	Glu	Phe	Gly	Pro	Lys	Gln	Val	Leu	Ser	Lys	Leu	Val	Ser	Glu		
				885					890						895		
acc	ctc	aag	gat	gac	ccc	tcg	gtt	gtc	acc	gtc	tct	gtc	aac	ccg	gcc	2736	
Thr	Leu	Lys	Asp	Asp	Pro	Ser	Val	Val	Thr	Val	Ser	Val	Asn	Pro	Ala		
			900					905					910				
tcg	ggc	acg	gat	tcg	gac	atc	cag	ctc	cgc	gac	gcg	gcc	gtc	cag	ctc	2784	
Ser	Gly	Thr	Asp	Ser	Asp	Ile	Gln	Leu	Arg	Asp	Ala	Ala	Val	Gln	Leu		
			915				920						925				
gtt	gtc	gct	ggc	gtc	aac	ctt	cag	ggc	ttt	gac	aag	tgg	gac	gcc	ccc	2832	
Val	Val	Ala	Gly	Val	Asn	Leu	Gln	Gly	Phe	Asp	Lys	Trp	Asp	Ala	Pro		
			930				935					940					
gat	gcc	acc	cgc	atg	cag	gcc	atc	aag	aag	aag	cgc	act	acc	ctc	cgc	2880	
Asp	Ala	Thr	Arg	Met	Gln	Ala	Ile	Lys	Lys	Lys	Arg	Thr	Thr	Leu	Arg		
					945		950				955				960		
ctt	tcg	gcc	gcc	acc	tac	gtc	tcg	gac	aag	acc	aag	aag	gtc	cgc	gac	2928	
Leu	Ser	Ala	Ala	Thr	Tyr	Val	Ser	Asp	Lys	Thr	Lys	Lys	Val	Arg	Asp		
				965					970					975			
gcc	gcc	atg	aac	gat	ggc	cgc	tgc	gtc	acc	tac	ctc	aag	ggc	gcc	gca	2976	
Ala	Ala	Met	Asn	Asp	Gly	Arg	Cys	Val	Thr	Tyr	Leu	Lys	Gly	Ala	Ala		
			980					985					990				
ccg	ctc	atc	aag	gcc	ccg	gag	ccc	gtt	gtc	gac	gag	gcc	gcc	aag	cgc	3024	
Pro	Leu	Ile	Lys	Ala	Pro	Glu	Pro	Val	Val	Asp	Glu	Ala	Ala	Lys	Arg		
			995				1000					1005					
gag	gcc	gag	cgt	ctc	cag	aag	gag	ctt	cag	gat	gcc	cag	cgc	cag		3069	
Glu	Ala	Glu	Arg	Leu	Gln	Lys	Glu	Leu	Gln	Asp	Ala	Gln	Arg	Gln			
			1010				1015					1020					
ctc	gac	gac	gcc	aag	cgc	gcc	gcc	gag	gcc	aac	tcc	aag	ctc			3114	
Leu	Asp	Asp	Ala	Lys	Arg	Ala	Ala	Ala	Glu	Ala	Asn	Ser	Lys	Leu			
			1025				1030					1035					
gcc	gct	gcc	aag	gag	gag	gcc	aag	acc	gcc	gct	gct	tcg	gcc	aag		3159	
Ala	Ala	Ala	Lys	Glu	Glu	Ala	Lys	Thr	Ala	Ala	Ala	Ser	Ala	Lys			
			1040				1045					1050					
ccc	gca	gtt	gac	act	gct	gtt	gtc	gaa	aag	cat	cgt	gcc	atc	ctc		3204	
Pro	Ala	Val	Asp	Thr	Ala	Val	Val	Glu	Lys	His	Arg	Ala	Ile	Leu			
			1055				1060					1065					
aag	tcc	atg	ctc	gcg	gag	ctc	gat	ggc	tac	gga	tcg	gtc	gac	gct		3249	
Lys	Ser	Met	Leu	Ala	Glu	Leu	Asp	Gly	Tyr	Gly	Ser	Val	Asp	Ala			
			1070				1075					1080					
tct	tcc	ctc	cag	cag	cag	cag	cag	cag	cag	acg	gcc	ccc	gcc	ccg		3294	
Ser	Ser	Leu	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Thr	Ala	Pro	Ala	Pro			
			1085				1090					1095					
gtc	aag	gct	gct	gcg	cct	gcc	gcc	ccc	gtt	gcc	tcg	gcc	cct	gcc		3339	
Val	Lys	Ala	Ala	Ala	Pro	Ala	Ala	Pro	Val	Ala	Ser	Ala	Pro	Ala			
			1100				1105					1110					
ccg	gct	gtc	tcg	aac	gag	ctt	ctt	gag	aag	gcc	gag	act	gtc	gtc		3384	
Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val	Val			
			1115				1120					1125					
atg	gag	gtc	ctc	gcc	gcc	aag	acc	ggc	tac	gag	acc	gac	atg	atc		3429	
Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile			
			1130				1135					1140					
gag	gct	gac	atg	gag	ctc	gag	acc	gag	ctc	ggc	att	gac	tcc	atc		3474	
Glu	Ala	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile			
			1145				1150					1155					
aag	cgt	gtc	gag	atc	ctc	tcc	gag	gtc	cag	gcc	atg	ctc	aat	gtc		3519	

ES 2 567 305 T3

Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn	Val			
	1160					1165					1170						
gag	gcc	aag	gat	gtc	gat	gcc	ctc	agc	cgc	act	cgc	act	gtt	ggt			3564
Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly			
	1175					1180					1185						
gag	gtt	gtc	aac	gcc	atg	aag	gcc	gag	atc	gct	ggc	agc	tct	gcc			3609
Glu	Val	Val	Asn	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Ser	Ser	Ala			
	1190					1195					1200						
ccg	gcg	cct	gct	gcc	gct	gct	ccg	gct	ccg	gcc	aag	gct	gcc	cct			3654
Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Lys	Ala	Ala	Pro			
	1205					1210					1215						
gcc	gcc	gct	gcg	cct	gct	gct	tcg	aac	gag	ctt	ctc	gag	aag	gcc			3699
Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala			
	1220					1225					1230						
gag	acc	gtc	gtc	atg	gag	gtc	ctc	gcc	gcc	aag	act	ggc	tac	gag			3744
Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu			
	1235					1240					1245						
act	gac	atg	atc	gag	tcc	gac	atg	gag	ctc	gag	act	gag	ctc	ggc			3789
Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly			
	1250					1255					1260						
att	gac	tcc	atc	aag	cgT	gtc	gag	atc	ctc	tcc	gag	ggt	cag	gcc			3834
Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala			
	1265					1270					1275						
atg	ctc	aac	gtc	gag	gcc	aag	gac	gtc	gac	gct	ctc	agc	cgC	act			3879
Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr			
	1280					1285					1290						
cgC	act	gtg	ggt	gag	gtc	gtc	aac	gcc	atg	aag	gct	gag	atc	gct			3924
Arg	Thr	Val	Gly	Glu	Val	Val	Asn	Ala	Met	Lys	Ala	Glu	Ile	Ala			
	1295					1300					1305						
ggt	ggc	tct	gcc	ccg	gcg	cct	gcc	gcc	gct	gcc	cca	ggt	ccg	gct			3969
Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Gly	Pro	Ala			
	1310					1315					1320						
gct	gcc	gcc	cct	gcg	cct	gcc	gcc	gcc	gcc	cct	gct	gtc	tcg	aac			4014
Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn			
	1325					1330					1335						
gag	ctt	ctt	gag	aag	gcc	gag	acc	gtc	gtc	atg	gag	gtc	ctc	gcc			4059
Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala			
	1340					1345					1350						
gcc	aag	act	ggc	tac	gag	act	gac	atg	atc	gag	tcc	gac	atg	gag			4104
Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu			
	1355					1360					1365						
ctc	gag	acc	gag	ctc	ggc	att	gac	tcc	atc	aag	cgT	gtc	gag	att			4149
Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile			
	1370					1375					1380						
ctc	tcc	gag	gtc	cag	gcc	atg	ctc	aac	gtc	gag	gcc	aag	gac	gtc			4194
Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val			
	1385					1390					1395						
gac	gct	ctc	agc	cgC	acc	cgC	act	gtt	ggc	gag	gtc	gtc	gat	gcc			4239
Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val	Asp	Ala			
	1400					1405					1410						
atg	aag	gcc	gag	atc	gct	ggt	ggc	tct	gcc	ccg	gcg	cct	gcc	gcc			4284
Met	Lys	Ala	Glu	Ile	Ala	Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala			
	1415					1420					1425						
gct	gct	cct	gct	ccg	gct	gct	gcc	gcc	cct	gcg	cct	gcc	gcc	cct			4329
Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Pro			
	1430					1435					1440						
gcg	cct	gct	gtc	tcg	agc	gag	ctt	ctc	gag	aag	gcc	gag	act	gtc			4374
Ala	Pro	Ala	Val	Ser	Ser	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val			
	1445					1450					1455						
gtc	atg	gag	gtc	ctc	gcc	gcc	aag	act	ggc	tac	gag	act	gac	atg			4419
Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met			
	1460					1465					1470						
atc	gag	tcc	gac	atg	gag	ctc	gag	acc	gag	ctc	ggc	att	gac	tcc			4464



ES 2 567 305 T3

Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser		
	1475					1480					1485					
atc	aag	cgt	gtc	gag	att	ctc	tcc	gag	gtc	cag	gcc	atg	ctc	aac	4509	
Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn		
	1490					1495					1500					
gtc	gag	gcc	aag	gac	gtc	gac	gct	ctc	agc	cgc	acc	cgc	act	ggt	4554	
Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val		
	1505					1510					1515					
ggc	gag	gtc	gtc	gat	gcc	atg	aag	gcc	gag	atc	gct	ggt	ggc	tct	4599	
Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Gly	Ser		
	1520					1525					1530					
gcc	ccg	cgc	cct	gcc	gcc	gct	gct	cct	gct	ccg	gct	gct	gcc	gcc	4644	
Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala		
	1535					1540					1545					
cct	gcg	cct	gcc	gcc	cct	gcg	cct	gcc	gcc	cct	gcg	cct	gct	gtc	4689	
Pro	Ala	Pro	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Pro	Ala	Pro	Ala	Val		
	1550					1555					1560					
tcg	agc	gag	ctt	ctc	gag	aag	gcc	gag	act	gtc	gtc	atg	gag	gtc	4734	
Ser	Ser	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val		
	1565					1570					1575					
ctc	gcc	gcc	aag	act	ggc	tac	gag	act	gac	atg	att	gag	tcc	gac	4779	
Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp		
	1580					1585					1590					
atg	gag	ctc	gag	acc	gag	ctc	ggc	att	gac	tcc	atc	aag	cggt	gtc	4824	
Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val		
	1595					1600					1605					
gag	att	ctc	tcc	gag	ggt	cag	gcc	atg	ctc	aac	gtc	gag	gcc	aag	4869	
Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys		
	1610					1615					1620					
gac	gtc	gac	gct	ctc	agc	cgc	act	cgc	act	ggt	ggt	gag	gtc	gtc	4914	
Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val		
	1625					1630					1635					
gat	gcc	atg	aag	gct	gag	atc	gct	ggc	agc	tcc	gcc	tcg	gcg	cct	4959	
Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Ser	Ser	Ala	Ser	Ala	Pro		
	1640					1645					1650					
gcc	gcc	gct	gct	cct	gct	ccg	gct	gct	gcc	gct	cct	gcg	ccc	gct	5004	
Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala		
	1655					1660					1665					
gcc	gcc	gcc	cct	gct	gtc	tcg	aac	gag	ctt	ctc	gag	aaa	gcc	gag	5049	
Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala	Glu		
	1670					1675					1680					
act	gtc	gtc	atg	gag	gtc	ctc	gcc	gcc	aag	act	ggc	tac	gag	act	5094	
Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr		
	1685					1690					1695					
gac	atg	atc	gag	tcc	gac	atg	gag	ctc	gag	act	gag	ctc	ggc	att	5139	
Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile		
	1700					1705					1710					
gac	tcc	atc	aag	cggt	gtc	gag	atc	ctc	tcc	gag	ggt	cag	gcc	atg	5184	
Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met		
	1715					1720					1725					
ctc	aac	gtc	gag	gcc	aag	gac	gtc	gat	gcc	ctc	agc	cgc	acc	cgc	5229	
Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg		
	1730					1735					1740					
act	ggt	ggc	gag	ggt	gtc	gat	gcc	atg	aag	gcc	gag	atc	gct	ggt	5274	
Thr	Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly		
	1745					1750					1755					
ggc	tct	gcc	ccg	gcg	cct	gcc	gcc	gct	gcc	cct	gct	ccg	gct	gcc	5319	
Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala		
	1760					1765					1770					
gcc	gcc	cct	gct	gtc	tcg	aac	gag	ctt	ctc	gag	aag	gcc	gag	act	5364	
Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr		
	1775					1780					1785					
gtc	gtc	atg	gag	gtc	ctc	gcc	gcc	aag	act	ggc	tac	gag	acc	gac	5409	

ES 2 567 305 T3

Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	
	1790					1795					1800				
atg	atc	gag	tcc	gac	atg	gag	ctc	gag	acc	gag	ctc	ggc	att	gac	5454
Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	
	1805					1810					1815				
tcc	atc	aag	cgt	gtc	gag	att	ctc	tcc	gag	ggt	cag	ggc	atg	ctc	5499
Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	
	1820					1825					1830				
aac	gtc	gag	gcc	aag	gac	gtc	gat	gct	ctc	agc	cgc	act	cgc	act	5544
Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	
	1835					1840					1845				
gtt	ggc	gag	gtc	gtc	gat	gac	atg	aag	gct	gag	atc	ggc	ggc	agc	5589
Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Ser	
	1850					1855					1860				
tcc	gcc	ccg	gcg	cct	gcc	gcc	gct	gct	cct	gct	ccg	gct	gct	gcc	5634
Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	
	1865					1870					1875				
gct	cct	gcg	ccc	gct	gcc	gct	gcc	cct	gct	gtc	tcg	agc	gag	ctt	5679
Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Ser	Glu	Leu	
	1880					1885					1890				
ctc	gag	aag	gcc	gag	acc	gtc	gtc	atg	gag	gtc	ctc	ggc	ggc	aag	5724
Leu	Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	
	1895					1900					1905				
act	ggc	tac	gag	act	gac	atg	att	gag	tcc	gac	atg	gag	ctc	gag	5769
Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	
	1910					1915					1920				
act	gag	ctc	ggc	att	gac	tcc	atc	aag	cgt	gtc	gag	atc	ctc	tcc	5814
Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	
	1925					1930					1935				
gag	ggt	cag	gcc	atg	ctc	aac	gtc	gag	gcc	aag	gac	gtc	gat	gcc	5859
Glu	Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	
	1940					1945					1950				
ctc	agc	cgc	acc	cgc	act	ggt	ggc	gag	ggt	gtc	gat	ggc	atg	aag	5904
Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys	
	1955					1960					1965				
gcc	gag	atc	gct	ggt	ggc	tct	gcc	ccg	gcg	cct	gcc	ggc	gct	gcc	5949
Ala	Glu	Ile	Ala	Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	
	1970					1975					1980				
cct	gct	ccg	gct	gcc	gcc	gcc	cct	gct	gtc	tcg	aac	gag	ctt	ctt	5994
Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	
	1985					1990					1995				
gag	aag	gcc	gag	acc	gtc	gtc	atg	gag	gtc	ctc	gcc	ggc	aag	act	6039
Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	
	2000					2005					2010				
ggc	tac	gag	acc	gac	atg	atc	gag	tcc	gac	atg	gag	ctc	gag	acc	6084
Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	
	2015					2020					2025				
gag	ctc	ggc	att	gac	tcc	atc	aag	cgt	gtc	gag	att	ctc	tcc	gag	6129
Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	
	2030					2035					2040				
ggt	cag	gcc	atg	ctc	aac	gtc	gag	gcc	aag	gac	gtc	gac	gct	ctc	6174
Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	
	2045					2050					2055				
agc	cgc	act	cgc	act	ggt	ggc	gag	gtc	gtc	gat	gcc	atg	aag	gct	6219
Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala	
	2060					2065					2070				
gag	atc	gct	ggt	ggc	tct	gcc	ccg	gcg	cct	gcc	ggc	gct	gct	cct	6264
Glu	Ile	Ala	Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	
	2075					2080					2085				
gcc	tcg	gct	ggc	gcc	gcg	cct	gcg	gtc	aag	att	gac	tcg	gtc	cac	6309
Ala	Ser	Ala	Gly	Ala	Ala	Pro	Ala	Val	Lys	Ile	Asp	Ser	Val	His	
	2090					2095					2100				
ggc	gct	gac	tgt	gat	gat	ctt	tcc	ctg	atg	cac	ggc	aag	gtg	ggt	6354

ES 2 567 305 T3

Gly	Ala	Asp	Cys	Asp	Asp	Leu	Ser	Leu	Met	His	Ala	Lys	Val	Val	
	2105					2110					2115				
gac	atc	cgc	cgc	ccg	gac	gag	ctc	atc	ctg	gag	cgc	ccc	gag	aac	6399
Asp	Ile	Arg	Arg	Pro	Asp	Glu	Leu	Ile	Leu	Glu	Arg	Pro	Glu	Asn	
	2120					2125					2130				
cgc	ccc	gtt	ctc	gtt	gtc	gat	gac	ggc	agc	gag	ctc	acc	ctc	gcc	6444
Arg	Pro	Val	Leu	Val	Val	Asp	Asp	Gly	Ser	Glu	Leu	Thr	Leu	Ala	
	2135					2140					2145				
ctg	gtc	cgc	gtc	ctc	ggc	gcc	tgc	gcc	gtt	gtc	ctg	acc	ttt	gag	6489
Leu	Val	Arg	Val	Leu	Gly	Ala	Cys	Ala	Val	Val	Leu	Thr	Phe	Glu	
	2150					2155					2160				
ggt	ctc	cag	ctc	gct	cag	cgc	gct	ggt	gcc	gct	gcc	atc	cgc	cac	6534
Gly	Leu	Gln	Leu	Ala	Gln	Arg	Ala	Gly	Ala	Ala	Ala	Ile	Arg	His	
	2165					2170					2175				
gtg	ctc	gcc	aag	gat	ctt	tcc	gcg	gag	agc	gcc	gag	aag	gcc	atc	6579
Val	Leu	Ala	Lys	Asp	Leu	Ser	Ala	Glu	Ser	Ala	Glu	Lys	Ala	Ile	
	2180					2185					2190				
aag	gag	gcc	gag	cag	cgc	ttt	ggc	gct	ctc	ggc	ggc	ttc	atc	tcg	6624
Lys	Glu	Ala	Glu	Gln	Arg	Phe	Gly	Ala	Leu	Gly	Gly	Phe	Ile	Ser	
	2195					2200					2205				
cag	cag	gcg	gag	cgc	ttc	gag	ccc	gcc	gaa	atc	ctc	ggc	ttc	acg	6669
Gln	Gln	Ala	Glu	Arg	Phe	Glu	Pro	Ala	Glu	Ile	Leu	Gly	Phe	Thr	
	2210					2215					2220				
ctc	atg	tgc	gcc	aag	ttc	gcc	aag	gct	tcc	ctc	tgc	acg	gct	gtg	6714
Leu	Met	Cys	Ala	Lys	Phe	Ala	Lys	Ala	Ser	Leu	Cys	Thr	Ala	Val	
	2225					2230					2235				
gct	ggc	ggc	cgc	ccg	gcc	ttt	atc	ggt	gtg	gcg	cgc	ctt	gac	ggc	6759
Ala	Gly	Gly	Arg	Pro	Ala	Phe	Ile	Gly	Val	Ala	Arg	Leu	Asp	Gly	
	2240					2245					2250				
cgc	ctc	gga	ttc	act	tcg	cag	ggc	act	tct	gac	gcg	ctc	aag	cgt	6804
Arg	Leu	Gly	Phe	Thr	Ser	Gln	Gly	Thr	Ser	Asp	Ala	Leu	Lys	Arg	
	2255					2260					2265				
gcc	cag	cgt	ggt	gcc	atc	ttt	ggc	ctc	tgc	aag	acc	atc	ggc	ctc	6849
Ala	Gln	Arg	Gly	Ala	Ile	Phe	Gly	Leu	Cys	Lys	Thr	Ile	Gly	Leu	
	2270					2275					2280				
gag	tgg	tcc	gag	tct	gac	gtc	ttt	tcc	cgc	ggc	gtg	gac	att	gct	6894
Glu	Trp	Ser	Glu	Ser	Asp	Val	Phe	Ser	Arg	Gly	Val	Asp	Ile	Ala	
	2285					2290					2295				
cag	ggc	atg	cac	ccc	gag	gat	gcc	gcc	gtg	gcg	att	gtg	cgc	gag	6939
Gln	Gly	Met	His	Pro	Glu	Asp	Ala	Ala	Val	Ala	Ile	Val	Arg	Glu	
	2300					2305					2310				
atg	gcg	tgc	gct	gac	att	cgc	att	cgc	gag	gtc	ggc	att	ggc	gca	6984
Met	Ala	Cys	Ala	Asp	Ile	Arg	Ile	Arg	Glu	Val	Gly	Ile	Gly	Ala	
	2315					2320					2325				
aac	cag	cag	cgc	tgc	acg	atc	cgt	gcc	gcc	aag	ctc	gag	acc	ggc	7029
Asn	Gln	Gln	Arg	Cys	Thr	Ile	Arg	Ala	Ala	Lys	Leu	Glu	Thr	Gly	
	2330					2335					2340				
aac	ccg	cag	cgc	cag	atc	gcc	aag	gac	gac	gtg	ctg	ctc	gtt	tct	7074
Asn	Pro	Gln	Arg	Gln	Ile	Ala	Lys	Asp	Asp	Val	Leu	Leu	Val	Ser	
	2345					2350					2355				
ggc	ggc	gct	cgc	ggc	atc	acg	cct	ctt	tgc	atc	cgg	gag	atc	acg	7119
Gly	Gly	Ala	Arg	Gly	Ile	Thr	Pro	Leu	Cys	Ile	Arg	Glu	Ile	Thr	
	2360					2365					2370				
cgc	cag	atc	gcg	ggc	ggc	aag	tac	att	ctg	ctt	ggc	cgc	agc	aag	7164
Arg	Gln	Ile	Ala	Gly	Gly	Lys	Tyr	Ile	Leu	Leu	Gly	Arg	Ser	Lys	
	2375					2380					2385				
gtc	tct	gcg	agc	gaa	ccg	gca	tgg	tgc	gct	ggc	atc	act	gac	gag	7209
Val	Ser	Ala	Ser	Glu	Pro	Ala	Trp	Cys	Ala	Gly	Ile	Thr	Asp	Glu	
	2390					2395					2400				
aag	gct	gtg	caa	aag	gct	gct	acc	cag	gag	ctc	aag	cgc	gcc	ttt	7254
Lys	Ala	Val	Gln	Lys	Ala	Ala	Thr	Gln	Glu	Leu	Lys	Arg	Ala	Phe	
	2405					2410					2415				
agc	gct	ggc	gag	ggc	ccc	aag	ccc	acg	ccc	cgc	gct	gtc	act	aag	7299

ES 2 567 305 T3

Ser	Ala	Gly	Glu	Gly	Pro	Lys	Pro	Thr	Pro	Arg	Ala	Val	Thr	Lys	
	2420					2425					2430				
ctt	gtg	ggc	tct	gtt	ctt	ggc	gct	cgc	gag	gtg	cgc	agc	tct	att	7344
Leu	Val	Gly	Ser	Val	Leu	Gly	Ala	Arg	Glu	Val	Arg	Ser	Ser	Ile	
	2435					2440					2445				
gct	gcg	att	gaa	gcg	ctc	ggc	ggc	aag	gcc	atc	tac	tcg	tcg	tcg	7389
Ala	Ala	Ile	Glu	Ala	Leu	Gly	Gly	Lys	Ala	Ile	Tyr	Ser	Ser	Cys	
	2450					2455					2460				
gac	gtg	aac	tct	gcc	gcc	gac	gtg	gcc	aag	gcc	gtg	cgc	gat	gcc	7434
Asp	Val	Asn	Ser	Ala	Ala	Asp	Val	Ala	Lys	Ala	Val	Arg	Asp	Ala	
	2465					2470					2475				
gag	tcc	cag	ctc	ggt	gcc	cgc	gtc	tcg	ggc	atc	ggt	cat	gcc	tcg	7479
Glu	Ser	Gln	Leu	Gly	Ala	Arg	Val	Ser	Gly	Ile	Val	His	Ala	Ser	
	2480					2485					2490				
ggc	gtg	ctc	cgc	gac	cgt	ctc	atc	gag	aag	aag	ctc	ccc	gac	gag	7524
Gly	Val	Leu	Arg	Asp	Arg	Leu	Ile	Glu	Lys	Lys	Leu	Pro	Asp	Glu	
	2495					2500					2505				
ttc	gac	gcc	gtc	ttt	ggc	acc	aag	gtc	acc	ggt	ctc	gag	aac	ctc	7569
Phe	Asp	Ala	Val	Phe	Gly	Thr	Lys	Val	Thr	Gly	Leu	Glu	Asn	Leu	
	2510					2515					2520				
ctc	gcc	gcc	gtc	gac	cgc	gcc	aac	ctc	aag	cac	atg	gtc	ctc	ttc	7614
Leu	Ala	Ala	Val	Asp	Arg	Ala	Asn	Leu	Lys	His	Met	Val	Leu	Phe	
	2525					2530					2535				
agc	tcg	ctc	gcc	ggc	ttc	cac	ggc	aac	gtc	ggc	cag	tct	gac	tac	7659
Ser	Ser	Leu	Ala	Gly	Phe	His	Gly	Asn	Val	Gly	Gln	Ser	Asp	Tyr	
	2540					2545					2550				
gcc	atg	gcc	aac	gag	gcc	ctt	aac	aag	atg	ggc	ctc	gag	ctc	gcc	7704
Ala	Met	Ala	Asn	Glu	Ala	Leu	Asn	Lys	Met	Gly	Leu	Glu	Leu	Ala	
	2555					2560					2565				
aag	gac	gtc	tcg	gtc	aag	tcg	atc	tgc	ttc	ggt	ccc	tgg	gac	ggt	7749
Lys	Asp	Val	Ser	Val	Lys	Ser	Ile	Cys	Phe	Gly	Pro	Trp	Asp	Gly	
	2570					2575					2580				
ggc	atg	gtg	acg	ccg	cag	ctc	aag	aag	cag	ttc	cag	gag	atg	ggc	7794
Gly	Met	Val	Thr	Pro	Gln	Leu	Lys	Lys	Gln	Phe	Gln	Glu	Met	Gly	
	2585					2590					2595				
gtg	cag	atc	atc	ccc	cgc	gag	ggc	ggc	gct	gat	acc	gtg	gcg	cgc	7839
Val	Gln	Ile	Ile	Pro	Arg	Glu	Gly	Gly	Ala	Asp	Thr	Val	Ala	Arg	
	2600					2605					2610				
atc	gtg	ctc	ggc	tcc	tcg	ccg	gct	gag	atc	ctt	gtc	ggc	aac	tgg	7884
Ile	Val	Leu	Gly	Ser	Ser	Pro	Ala	Glu	Ile	Leu	Val	Gly	Asn	Trp	
	2615					2620					2625				
cgc	acc	ccg	tcc	aag	aag	gtc	ggc	tcg	gac	acc	atc	acc	ctg	cac	7929
Arg	Thr	Pro	Ser	Lys	Lys	Val	Gly	Ser	Asp	Thr	Ile	Thr	Leu	His	
	2630					2635					2640				
cgc	aag	att	tcc	gcc	aag	tcc	aac	ccc	ttc	ctc	gag	gac	cac	gtc	7974
Arg	Lys	Ile	Ser	Ala	Lys	Ser	Asn	Pro	Phe	Leu	Glu	Asp	His	Val	
	2645					2650					2655				
atc	cag	ggc	cgc	cgc	gtg	ctg	ccc	atg	acg	ctg	gcc	att	ggc	tcg	8019
Ile	Gln	Gly	Arg	Arg	Val	Leu	Pro	Met	Thr	Leu	Ala	Ile	Gly	Ser	
	2660					2665					2670				
ctc	gcg	gag	acc	tgc	ctc	ggc	ctc	ttc	ccc	ggc	tac	tcg	ctc	tgg	8064
Leu	Ala	Glu	Thr	Cys	Leu	Gly	Leu	Phe	Pro	Gly	Tyr	Ser	Leu	Trp	
	2675					2680					2685				
gcc	att	gac	gac	gcc	cag	ctc	ttc	aag	ggt	gtc	act	gtc	gac	ggc	8109
Ala	Ile	Asp	Asp	Ala	Gln	Leu	Phe	Lys	Gly	Val	Thr	Val	Asp	Gly	
	2690					2695					2700				
gac	gtc	aac	tgc	gag	gtg	acc	ctc	acc	ccg	tcg	acg	gcg	ccc	tcg	8154
Asp	Val	Asn	Cys	Glu	Val	Thr	Leu	Thr	Pro	Ser	Thr	Ala	Pro	Ser	
	2705					2710					2715				
ggc	cgc	gtc	aac	gtc	cag	gcc	acg	ctc	aag	acc	ttt	tcc	agc	ggc	8199
Gly	Arg	Val	Asn	Val	Gln	Ala	Thr	Leu	Lys	Thr	Phe	Ser	Ser	Gly	
	2720					2725					2730				
aag	ctg	gtc	ccg	gcc	tac	cgc	gcc	gtc	atc	gtg	ctc	tcc	aac	cag	8244

ES 2 567 305 T3

Lys Leu Val Pro Ala Tyr Arg Ala Val Ile Val Leu Ser Asn Gln  
 2735 2740 2745  
 ggc gcg ccc ccg gcc aac gcc acc atg cag ccg ccc tcg ctc gat 8289  
 Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro Pro Ser Leu Asp  
 2750 2755 2760  
 gcc gat ccg gcg ctc cag gcc tcc gtc tac gac gcc aag acc ctc 8334  
 Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly Lys Thr Leu  
 2765 2770 2775  
 ttc cac ggc ccg gcc ttc cgc gcc atc gat gac gtg ctc tcg tgc 8379  
 Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu Ser Cys  
 2780 2785 2790  
 acc aag agc cag ott gtg gcc aag tgc agc gct gtc ccc ggc tcc 8424  
 Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly Ser  
 2795 2800 2805  
 gac gcc gct cgc gcc gag ttt gcc acg gac act gac gcc cat gac 8469  
 Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp  
 2810 2815 2820  
 ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg 8514  
 Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val  
 2825 2830 2835  
 cgc cgc acg ctc gcc cag gct gcg ctc ccc aac tcg atc cag cgc 8559  
 Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg  
 2840 2845 2850  
 atc gtc cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att 8604  
 Ile Val Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile  
 2855 2860 2865  
 acc ctc cgc tcc aac cag tcg gcc ggt cac tcc cag cac aag cac 8649  
 Thr Leu Arg Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His  
 2870 2875 2880  
 gcc ctt cag ttc cac aac gag cag gcc gat ctc ttc att gat gtc 8694  
 Ala Leu Gln Phe His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val  
 2885 2890 2895  
 cag gct tcg gtc atc gcc acg gac agc ctt gcc ttc 8730  
 Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala Phe  
 2900 2905 2910

<210> 2  
 <211> 2910  
 <212> PRT  
 <213> *Schizochytrium sp.*

5

<400> 2

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

10

ES 2 567 305 T3

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile  
 115 120 125  
 Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe  
 130 135 140  
 Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met  
 145 150 155 160  
 Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala  
 165 170 175  
 Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn  
 180 185 190  
 Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn  
 195 200 205  
 Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val  
 210 215 220  
 Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly  
 225 230 235 240  
 Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys  
 245 250 255  
 Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys  
 260 265 270  
 Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys  
 275 280 285  
 Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile  
 290 295 300  
 Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr  
 305 310 315 320  
 Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg  
 325 330 335  
 Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr  
 340 345 350  
 Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu  
 355 360 365  
 Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly  
 370 375 380  
 Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala  
 385 390 395 400  
 Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly  
 405 410 415  
 Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile  
 420 425 430  
 Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro  
 435 440 445

ES 2 567 305 T3

Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly  
 450 455 460  
 Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr  
 465 470 475  
 Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala  
 485 490 495  
 Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu  
 500 505 510  
 Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr  
 515 520 525  
 Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile  
 530 535 540  
 Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp  
 545 550 555 560  
 Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val  
 565 570 575  
 Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala  
 580 585 590  
 Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln  
 595 600 605  
 Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro  
 610 615 620  
 Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val  
 625 630 635 640  
 Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro  
 645 650 655  
 Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser  
 660 665 670  
 Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala  
 675 680 685  
 Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly  
 690 695 700  
 His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp  
 705 710 715 720  
 Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly  
 725 730 735  
 Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile  
 740 745 750  
 Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu  
 755 760 765  
 Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu  
 770 775 780

ES 2 567 305 T3

Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val  
 785 790 795 800  
 Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn  
 805 810 815  
 Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr  
 820 825 830  
 Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr  
 835 840 845  
 Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val  
 850 855 860  
 Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile  
 865 870 875 880  
 Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu  
 885 890 895  
 Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala  
 900 905 910  
 Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu  
 915 920 925  
 Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro  
 930 935 940  
 Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg  
 945 950 955 960  
 Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp  
 965 970 975  
 Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala  
 980 985 990  
 Pro Leu Ile Lys Ala Pro Glu Pro Val Val Asp Glu Ala Ala Lys Arg  
 995 1000 1005  
 Glu Ala Glu Arg Leu Gln Lys Glu Leu Gln Asp Ala Gln Arg Gln  
 1010 1015 1020  
 Leu Asp Asp Ala Lys Arg Ala Ala Ala Glu Ala Asn Ser Lys Leu  
 1025 1030 1035  
 Ala Ala Ala Lys Glu Glu Ala Lys Thr Ala Ala Ala Ser Ala Lys  
 1040 1045 1050  
 Pro Ala Val Asp Thr Ala Val Val Glu Lys His Arg Ala Ile Leu  
 1055 1060 1065  
 Lys Ser Met Leu Ala Glu Leu Asp Gly Tyr Gly Ser Val Asp Ala  
 1070 1075 1080  
 Ser Ser Leu Gln Gln Gln Gln Gln Gln Gln Thr Ala Pro Ala Pro  
 1085 1090 1095  
 Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser Ala Pro Ala  
 1100 1105 1110



# ES 2 567 305 T3

Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val	Val
	1115					1120					1125			
Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile
	1130					1135					1140			
Glu	Ala	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile
	1145					1150					1155			
Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn	Val
	1160					1165					1170			
Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly
	1175					1180					1185			
Glu	Val	Val	Asn	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Ser	Ser	Ala
	1190					1195					1200			
Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Lys	Ala	Ala	Pro
	1205					1210					1215			
Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala
	1220					1225					1230			
Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu
	1235					1240					1245			
Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly
	1250					1255					1260			
Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala
	1265					1270					1275			
Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr
	1280					1285					1290			
Arg	Thr	Val	Gly	Glu	Val	Val	Asn	Ala	Met	Lys	Ala	Glu	Ile	Ala
	1295					1300					1305			
Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Gly	Pro	Ala
	1310					1315					1320			
Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn
	1325					1330					1335			
Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala
	1340					1345					1350			
Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu
	1355					1360					1365			
Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile
	1370					1375					1380			
Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val
	1385					1390					1395			
Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val	Asp	Ala
	1400					1405					1410			
Met	Lys	Ala	Glu	Ile	Ala	Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala
	1415					1420					1425			

# ES 2 567 305 T3

Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Pro			
	1430					1435					1440						
Ala	Pro	Ala	Val	Ser	Ser	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val			
	1445					1450					1455						
Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met			
	1460					1465					1470						
Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser			
	1475					1480					1485						
Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn			
	1490					1495					1500						
Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val			
	1505					1510					1515						
Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Gly	Ser			
	1520					1525					1530						
Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala			
	1535					1540					1545						
Pro	Ala	Pro	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Pro	Ala	Pro	Ala	Val			
	1550					1555					1560						
Ser	Ser	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val			
	1565					1570					1575						
Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp			
	1580					1585					1590						
Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val			
	1595					1600					1605						
Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys			
	1610					1615					1620						
Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val			
	1625					1630					1635						
Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Ser	Ser	Ala	Ser	Ala	Pro			
	1640					1645					1650						
Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala			
	1655					1660					1665						
Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala	Glu			
	1670					1675					1680						
Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr			
	1685					1690					1695						
Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile			
	1700					1705					1710						
Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met			
	1715					1720					1725						
Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg			
	1730					1735					1740						

# ES 2 567 305 T3

Thr	Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly
1745						1750					1755			
Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala
1760						1765					1770			
Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu	Glu	Lys	Ala	Glu	Thr
1775						1780					1785			
Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr	Gly	Tyr	Glu	Thr	Asp
1790						1795					1800			
Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr	Glu	Leu	Gly	Ile	Asp
1805						1810					1815			
Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu	Val	Gln	Ala	Met	Leu
1820						1825					1830			
Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Arg	Thr
1835						1840					1845			
Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala	Glu	Ile	Ala	Gly	Ser
1850						1855					1860			
Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Pro	Ala	Ala	Ala
1865						1870					1875			
Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Ser	Glu	Leu
1880						1885					1890			
Leu	Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys
1895						1900					1905			
Thr	Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu
1910						1915					1920			
Thr	Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser
1925						1930					1935			
Glu	Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala
1940						1945					1950			
Leu	Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys
1955						1960					1965			
Ala	Glu	Ile	Ala	Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala
1970						1975					1980			
Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro	Ala	Val	Ser	Asn	Glu	Leu	Leu
1985						1990					1995			
Glu	Lys	Ala	Glu	Thr	Val	Val	Met	Glu	Val	Leu	Ala	Ala	Lys	Thr
2000						2005					2010			
Gly	Tyr	Glu	Thr	Asp	Met	Ile	Glu	Ser	Asp	Met	Glu	Leu	Glu	Thr
2015						2020					2025			
Glu	Leu	Gly	Ile	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Leu	Ser	Glu
2030						2035					2040			
Val	Gln	Ala	Met	Leu	Asn	Val	Glu	Ala	Lys	Asp	Val	Asp	Ala	Leu
2045						2050					2055			

# ES 2 567 305 T3

Ser	Arg	Thr	Arg	Thr	Val	Gly	Glu	Val	Val	Asp	Ala	Met	Lys	Ala
2060						2065					2070			
Glu	Ile	Ala	Gly	Gly	Ser	Ala	Pro	Ala	Pro	Ala	Ala	Ala	Ala	Pro
2075						2080					2085			
Ala	Ser	Ala	Gly	Ala	Ala	Pro	Ala	Val	Lys	Ile	Asp	Ser	Val	His
2090						2095					2100			
Gly	Ala	Asp	Cys	Asp	Asp	Leu	Ser	Leu	Met	His	Ala	Lys	Val	Val
2105						2110					2115			
Asp	Ile	Arg	Arg	Pro	Asp	Glu	Leu	Ile	Leu	Glu	Arg	Pro	Glu	Asn
2120						2125					2130			
Arg	Pro	Val	Leu	Val	Val	Asp	Asp	Gly	Ser	Glu	Leu	Thr	Leu	Ala
2135						2140					2145			
Leu	Val	Arg	Val	Leu	Gly	Ala	Cys	Ala	Val	Val	Leu	Thr	Phe	Glu
2150						2155					2160			
Gly	Leu	Gln	Leu	Ala	Gln	Arg	Ala	Gly	Ala	Ala	Ala	Ile	Arg	His
2165						2170					2175			
Val	Leu	Ala	Lys	Asp	Leu	Ser	Ala	Glu	Ser	Ala	Glu	Lys	Ala	Ile
2180						2185					2190			
Lys	Glu	Ala	Glu	Gln	Arg	Phe	Gly	Ala	Leu	Gly	Gly	Phe	Ile	Ser
2195						2200					2205			
Gln	Gln	Ala	Glu	Arg	Phe	Glu	Pro	Ala	Glu	Ile	Leu	Gly	Phe	Thr
2210						2215					2220			
Leu	Met	Cys	Ala	Lys	Phe	Ala	Lys	Ala	Ser	Leu	Cys	Thr	Ala	Val
2225						2230					2235			
Ala	Gly	Gly	Arg	Pro	Ala	Phe	Ile	Gly	Val	Ala	Arg	Leu	Asp	Gly
2240						2245					2250			
Arg	Leu	Gly	Phe	Thr	Ser	Gln	Gly	Thr	Ser	Asp	Ala	Leu	Lys	Arg
2255						2260					2265			
Ala	Gln	Arg	Gly	Ala	Ile	Phe	Gly	Leu	Cys	Lys	Thr	Ile	Gly	Leu
2270						2275					2280			
Glu	Trp	Ser	Glu	Ser	Asp	Val	Phe	Ser	Arg	Gly	Val	Asp	Ile	Ala
2285						2290					2295			
Gln	Gly	Met	His	Pro	Glu	Asp	Ala	Ala	Val	Ala	Ile	Val	Arg	Glu
2300						2305					2310			
Met	Ala	Cys	Ala	Asp	Ile	Arg	Ile	Arg	Glu	Val	Gly	Ile	Gly	Ala
2315						2320					2325			
Asn	Gln	Gln	Arg	Cys	Thr	Ile	Arg	Ala	Ala	Lys	Leu	Glu	Thr	Gly
2330						2335					2340			
Asn	Pro	Gln	Arg	Gln	Ile	Ala	Lys	Asp	Asp	Val	Leu	Leu	Val	Ser
2345						2350					2355			
Gly	Gly	Ala	Arg	Gly	Ile	Thr	Pro	Leu	Cys	Ile	Arg	Glu	Ile	Thr
2360						2365					2370			

ES 2 567 305 T3

Arg Gln Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys  
 2375 2380 2385

Val Ser Ala Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu  
 2390 2395 2400

Lys Ala Val Gln Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe  
 2405 2410 2415

Ser Ala Gly Glu Gly Pro Lys Pro Thr Pro Arg Ala Val Thr Lys  
 2420 2425 2430

Leu Val Gly Ser Val Leu Gly Ala Arg Glu Val Arg Ser Ser Ile  
 2435 2440 2445

Ala Ala Ile Glu Ala Leu Gly Gly Lys Ala Ile Tyr Ser Ser Cys  
 2450 2455 2460

Asp Val Asn Ser Ala Ala Asp Val Ala Lys Ala Val Arg Asp Ala  
 2465 2470 2475

Glu Ser Gln Leu Gly Ala Arg Val Ser Gly Ile Val His Ala Ser  
 2480 2485 2490

Gly Val Leu Arg Asp Arg Leu Ile Glu Lys Lys Leu Pro Asp Glu  
 2495 2500 2505

Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly Leu Glu Asn Leu  
 2510 2515 2520

Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met Val Leu Phe  
 2525 2530 2535

Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser Asp Tyr  
 2540 2545 2550

Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu Ala  
 2555 2560 2565

Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly  
 2570 2575 2580

Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly  
 2585 2590 2595

Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg  
 2600 2605 2610

Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp  
 2615 2620 2625

Arg Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His  
 2630 2635 2640

Arg Lys Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val  
 2645 2650 2655

Ile Gln Gly Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser  
 2660 2665 2670

Leu Ala Glu Thr Cys Leu Gly Leu Phe Pro Gly Tyr Ser Leu Trp  
 2675 2680 2685

ES 2 567 305 T3

Ala Ile Asp Asp Ala Gln Leu Phe Lys Gly Val Thr Val Asp Gly  
 2690 2695 2700  
 Asp Val Asn Cys Glu Val Thr Leu Thr Pro Ser Thr Ala Pro Ser  
 2705 2710 2715  
 Gly Arg Val Asn Val Gln Ala Thr Leu Lys Thr Phe Ser Ser Gly  
 2720 2725 2730  
 Lys Leu Val Pro Ala Tyr Arg Ala Val Ile Val Leu Ser Asn Gln  
 2735 2740 2745  
 Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro Pro Ser Leu Asp  
 2750 2755 2760  
 Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly Lys Thr Leu  
 2765 2770 2775  
 Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu Ser Cys  
 2780 2785 2790  
 Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly Ser  
 2795 2800 2805  
 Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp  
 2810 2815 2820  
 Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val  
 2825 2830 2835  
 Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg  
 2840 2845 2850  
 Ile Val Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile  
 2855 2860 2865  
 Thr Leu Arg Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His  
 2870 2875 2880  
 Ala Leu Gln Phe His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val  
 2885 2890 2895  
 Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala Phe  
 2900 2905 2910

<210> 3  
 <211> 6177  
 <212> ADN  
 <213> *Schizochytrium sp.*

5

<220>  
 <221> CDS  
 <222> (1)..(6177)  
 <223>

10

<400> 3

atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag 48  
 Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys  
 1 5 10 15  
 cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc 96  
 Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr  
 20 25 30  
 aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag 144

15

ES 2 567 305 T3

Lys	Asp	Glu	Phe	Trp	Glu	Val	Leu	Met	Asn	Gly	Lys	Val	Glu	Ser	Lys		
		35				40					45						
gtg	atc	agc	gac	aaa	cga	ctc	ggc	tcc	aac	tac	cgc	gcc	gag	cac	tac		192
Val	Ile	Ser	Asp	Lys	Arg	Leu	Gly	Ser	Asn	Tyr	Arg	Ala	Glu	His	Tyr		
	50					55				60							
aaa	gca	gag	cgc	agc	aag	tat	gcc	gac	acc	ttt	tgc	aac	gaa	acg	tac		240
Lys	Ala	Glu	Arg	Ser	Lys	Tyr	Ala	Asp	Thr	Phe	Cys	Asn	Glu	Thr	Tyr		
	65				70				75					80			
ggc	acc	ctt	gac	gag	aac	gag	atc	gac	aac	gag	cac	gaa	ctc	ctc	ctc		288
Gly	Thr	Leu	Asp	Glu	Asn	Glu	Ile	Asp	Asn	Glu	His	Glu	Leu	Leu	Leu		
			85					90						95			
aac	ctc	gcc	aag	cag	gca	ctc	gca	gag	aca	tcc	gtc	aaa	gac	tcg	aca		336
Asn	Leu	Ala	Lys	Gln	Ala	Leu	Ala	Glu	Thr	Ser	Val	Lys	Asp	Ser	Thr		
			100					105					110				
cgc	tgc	ggc	atc	gtc	agc	ggc	tgc	ctc	tcg	ttc	ccc	atg	gac	aac	ctc		384
Arg	Cys	Gly	Ile	Val	Ser	Gly	Cys	Leu	Ser	Phe	Pro	Met	Asp	Asn	Leu		
		115				120						125					
cag	ggt	gaa	ctc	ctc	aac	gtg	tac	caa	aac	cat	gtc	gag	aaa	aag	ctc		432
Gln	Gly	Glu	Leu	Leu	Asn	Val	Tyr	Gln	Asn	His	Val	Glu	Lys	Lys	Leu		
	130					135					140						
ggg	gcc	cgc	gtc	ttc	aag	gac	gcc	tcc	cat	tgg	tcc	gaa	cgc	gag	cag		480
Gly	Ala	Arg	Val	Phe	Lys	Asp	Ala	Ser	His	Trp	Ser	Glu	Arg	Glu	Gln		
	145			150					155					160			
tcc	aac	aaa	ccc	gag	gcc	ggt	gac	cgc	cgc	atc	ttc	atg	gac	ccg	gcc		528
Ser	Asn	Lys	Pro	Glu	Ala	Gly	Asp	Arg	Arg	Ile	Phe	Met	Asp	Pro	Ala		
			165					170					175				
tcc	ttc	gtc	gcc	gaa	gaa	ctc	aac	ctc	ggc	gcc	ctt	cac	tac	tcc	gtc		576
Ser	Phe	Val	Ala	Glu	Glu	Leu	Asn	Leu	Gly	Ala	Leu	His	Tyr	Ser	Val		
			180					185					190				
gac	gca	gca	tgc	gcc	acg	gcg	ctc	tac	gtg	ctc	cgc	ctc	gcg	cag	gat		624
Asp	Ala	Ala	Cys	Ala	Thr	Ala	Leu	Tyr	Val	Leu	Arg	Leu	Ala	Gln	Asp		
		195				200						205					
cat	ctc	gtc	tcc	ggc	gcc	gcc	gac	gtc	atg	ctc	tgc	ggt	gcc	acc	tgc		672
His	Leu	Val	Ser	Gly	Ala	Ala	Asp	Val	Met	Leu	Cys	Gly	Ala	Thr	Cys		
	210					215					220						
ctg	ccg	gag	ccc	ttt	ttc	atc	ctt	tcg	ggc	ttt	tcc	acc	ttc	cag	gcc		720
Leu	Pro	Glu	Pro	Phe	Phe	Ile	Leu	Ser	Gly	Phe	Ser	Thr	Phe	Gln	Ala		
	225			230						235				240			
atg	ccc	gtc	ggc	acg	ggc	cag	aac	gtg	tcc	atg	ccg	ctg	cac	aag	gac		768
Met	Pro	Val	Gly	Thr	Gly	Gln	Asn	Val	Ser	Met	Pro	Leu	His	Lys	Asp		
			245					250						255			
agc	cag	ggc	ctc	acc	ccg	ggt	gag	ggc	ggc	tcc	atc	atg	gtc	ctc	aag		816
Ser	Gln	Gly	Leu	Thr	Pro	Gly	Glu	Gly	Gly	Ser	Ile	Met	Val	Leu	Lys		
			260					265					270				
cgf	ctc	gat	gat	gcc	atc	cgc	gac	ggc	gac	cac	att	tac	ggc	acc	ctt		864
Arg	Leu	Asp	Asp	Ala	Ile	Arg	Asp	Gly	Asp	His	Ile	Tyr	Gly	Thr	Leu		
		275					280					285					
ctc	ggc	gcc	aat	gtc	agc	aac	tcc	ggc	aca	ggt	ctg	ccc	ctc	aag	ccc		912
Leu	Gly	Ala	Asn	Val	Ser	Asn	Ser	Gly	Thr	Gly	Leu	Pro	Leu	Lys	Pro		
	290					295						300					
ctt	ctc	ccc	agc	gag	aaa	aag	tgc	ctc	atg	gac	acc	tac	acg	cgc	att		960
Leu	Leu	Pro	Ser	Glu	Lys	Lys	Cys	Leu	Met	Asp	Thr	Tyr	Thr	Arg	Ile		
	305			310						315				320			
aac	gtg	cac	ccg	cac	aag	att	cag	tac	gtc	gag	tgc	cac	gcc	acc	ggc		1008
Asn	Val	His	Pro	His	Lys	Ile	Gln	Tyr	Val	Glu	Cys	His	Ala	Thr	Gly		
			325					330					335				
acg	ccc	cag	ggt	gat	cgf	gtg	gaa	atc	gac	gcc	gtc	aag	gcc	tgc	ttt		1056
Thr	Pro	Gln	Gly	Asp	Arg	Val	Glu	Ile	Asp	Ala	Val	Lys	Ala	Cys	Phe		
			340					345					350				
gaa	ggc	aag	gtc	ccc	cgf	ttc	ggt	acc	aca	aag	ggc	aac	ttt	gga	cac		1104
Glu	Gly	Lys	Val	Pro	Arg	Phe	Gly	Thr	Thr	Lys	Gly	Asn	Phe	Gly	His		
		355				360						365					
acc	cts	gyc	gca	gcc	ggc	ttt	gcc	ggt	atg	tgc	aag	gtc	ctc	ctc	tcc		1152

ES 2 567 305 T3

Thr	Xaa	Xaa	Ala	Ala	Gly	Phe	Ala	Gly	Met	Cys	Lys	Val	Leu	Leu	Ser		
	370					375					380						
atg	aag	cat	ggc	atc	atc	ccg	ccc	acc	ccg	ggt	atc	gat	gac	gag	acc		1200
Met	Lys	His	Gly	Ile	Ile	Pro	Pro	Thr	Pro	Gly	Ile	Asp	Asp	Glu	Thr		
	385				390					395					400		
aag	atg	gac	cct	ctc	gtc	gtc	tcc	ggt	gag	gcc	atc	cca	tgg	cca	gag		1248
Lys	Met	Asp	Pro	Leu	Val	Val	Ser	Gly	Glu	Ala	Ile	Pro	Trp	Pro	Glu		
				405					410					415			
acc	aac	ggc	gag	ccc	aag	cgc	gcc	ggt	ctc	tcg	gcc	ttt	ggc	ttt	ggt		1296
Thr	Asn	Gly	Glu	Pro	Lys	Arg	Ala	Gly	Leu	Ser	Ala	Phe	Gly	Phe	Gly		
			420					425					430				
ggc	acc	aac	ggc	cat	gcc	gtc	ttt	gag	gag	cat	gac	ccc	tcc	aac	gcc		1344
Gly	Thr	Asn	Ala	His	Ala	Val	Phe	Glu	Glu	His	Asp	Pro	Ser	Asn	Ala		
		435					440					445					
gcc	tgc	acg	ggc	cac	gac	tcc	att	tct	gcg	ctc	tcg	gcc	cgc	tgc	ggc		1392
Ala	Cys	Thr	Gly	His	Asp	Ser	Ile	Ser	Ala	Leu	Ser	Ala	Arg	Cys	Gly		
	450					455						460					
ggt	gaa	agc	aac	atg	cgc	atc	gcc	atc	act	ggt	atg	gac	gcc	acc	ttt		1440
Gly	Glu	Ser	Asn	Met	Arg	Ile	Ala	Ile	Thr	Gly	Met	Asp	Ala	Thr	Phe		
	465				470					475					480		
ggc	gct	ctc	aag	gga	ctc	gac	gcc	ttc	gag	cgc	gcc	att	tac	acc	ggc		1488
Gly	Ala	Leu	Lys	Gly	Leu	Asp	Ala	Phe	Glu	Arg	Ala	Ile	Tyr	Thr	Gly		
			485						490					495			
gct	cac	ggt	gcc	atc	cca	ctc	cca	gaa	aag	cgc	tgg	cgc	ttt	ctc	ggc		1536
Ala	His	Gly	Ala	Ile	Pro	Leu	Pro	Glu	Lys	Arg	Trp	Arg	Phe	Leu	Gly		
			500					505					510				
aag	gac	aag	gac	ttt	ctt	gac	ctc	tgc	ggc	gtc	aag	gcc	acc	ccg	cac		1584
Lys	Asp	Lys	Asp	Phe	Leu	Asp	Leu	Cys	Gly	Val	Lys	Ala	Thr	Pro	His		
		515					520					525					
ggc	tgc	tac	att	gaa	gat	gtt	gag	gtc	gac	ttc	cag	cgc	ctc	cgc	acg		1632
Gly	Cys	Tyr	Ile	Glu	Asp	Val	Glu	Val	Asp	Phe	Gln	Arg	Leu	Arg	Thr		
	530					535					540						
ccc	atg	acc	cct	gaa	gac	atg	ctc	ctc	cct	cag	cag	ctt	ctg	gcc	gtc		1680
Pro	Met	Thr	Pro	Glu	Asp	Met	Leu	Leu	Pro	Gln	Gln	Leu	Leu	Ala	Val		
	545				550					555					560		
acc	acc	att	gac	cgc	gcc	atc	ctc	gac	tcg	gga	atg	aaa	aag	ggt	ggc		1728
Thr	Thr	Ile	Asp	Arg	Ala	Ile	Leu	Asp	Ser	Gly	Met	Lys	Lys	Gly	Gly		
			565						570					575			
aat	gtc	gcc	gtc	ttt	gtc	ggc	ctc	ggc	acc	gac	ctc	gag	ctc	tac	cg		1776
Asn	Val	Ala	Val	Phe	Val	Gly	Leu	Gly	Thr	Asp	Leu	Glu	Leu	Tyr	Arg		
			580					585					590				
cac	cgt	gct	cgc	gtc	gct	ctc	aag	gag	cgc	gtc	cgc	cct	gaa	gcc	tcc		1824
His	Arg	Ala	Arg	Val	Ala	Leu	Lys	Glu	Arg	Val	Arg	Pro	Glu	Ala	Ser		
		595					600					605					
aag	aag	ctc	aat	gac	atg	atg	cag	tac	att	aac	gac	tgc	ggc	aca	tcc		1872
Lys	Lys	Leu	Asn	Asp	Met	Met	Gln	Tyr	Ile	Asn	Asp	Cys	Gly	Thr	Ser		
	610				615						620						
aca	tcg	tac	acc	tcg	tac	att	ggc	aac	ctc	gtc	gcc	acg	cgc	gtc	tcg		1920
Thr	Ser	Tyr	Thr	Ser	Tyr	Ile	Gly	Asn	Leu	Val	Ala	Thr	Arg	Val	Ser		
	625				630					635					640		
tcg	cag	tgg	ggc	ttc	acg	ggc	ccc	tcc	ttt	acg	atc	acc	gag	ggc	aac		1968
Ser	Gln	Trp	Gly	Phe	Thr	Gly	Pro	Ser	Phe	Thr	Ile	Thr	Glu	Gly	Asn		
			645						650					655			
aac	tcc	gtc	tac	cgc	tgc	gcc	gag	ctc	ggc	aag	tac	ctc	ctc	gag	acc		2016
Asn	Ser	Val	Tyr	Arg	Cys	Ala	Glu	Leu	Gly	Lys	Tyr	Leu	Leu	Glu	Thr		
			660					665					670				
ggc	gag	gtc	gat	ggc	gtc	gtc	ggt	gcg	ggt	gtc	gat	ctc	tgc	ggc	agt		2064
Gly	Glu	Val	Asp	Gly	Val	Val	Val	Ala	Gly	Val	Asp	Leu	Cys	Gly	Ser		
		675					680						685				
gcc	gaa	aac	ctt	tac	gtc	aag	tct	cgc	cgc	ttc	aag	gtg	tcc	acc	tcc		2112
Ala	Glu	Asn	Leu	Tyr	Val	Lys	Ser	Arg	Arg	Phe	Lys	Val	Ser	Thr	Ser		
	690					695						700					
gat	acc	ccg	cgc	gcc	agc	ttt	gac	gcc	gcc	gcc	gat	ggc	tac	ttt	gtc		2160



ES 2 567 305 T3

Asp	Thr	Pro	Arg	Ala	Ser	Phe	Asp	Ala	Ala	Ala	Asp	Gly	Tyr	Phe	Val		
705					710				715						720		
ggc	gag	ggc	tgc	ggt	gcc	ttt	gtg	ctc	aag	cgt	gag	act	agc	tgc	acc	2208	
Gly	Glu	Gly	Cys	Gly	Ala	Phe	Val	Leu	Lys	Arg	Glu	Thr	Ser	Cys	Thr		
				725					730					735			
aag	gac	gac	cgt	atc	tac	gct	tgc	atg	gat	gcc	atc	gtc	cct	ggc	aac	2256	
Lys	Asp	Asp	Arg	Ile	Tyr	Ala	Cys	Met	Asp	Ala	Ile	Val	Pro	Gly	Asn		
			740					745					750				
gtc	cct	agc	gcc	tgc	ttg	cgc	gag	gcc	ctc	gac	cag	gcg	cgc	gtc	aag	2304	
Val	Pro	Ser	Ala	Cys	Leu	Arg	Glu	Ala	Leu	Asp	Gln	Ala	Arg	Val	Lys		
			755				760					765					
ccg	ggc	gat	atc	gag	atg	ctc	gag	ctc	agc	gcc	gac	tcc	gcc	cgc	cac	2352	
Pro	Gly	Asp	Ile	Glu	Met	Leu	Glu	Leu	Ser	Ala	Asp	Ser	Ala	Arg	His		
	770				775						780						
ctc	aag	gac	ccg	tcc	gtc	ctg	ccc	aag	gag	ctc	act	gcc	gag	gag	gaa	2400	
Leu	Lys	Asp	Pro	Ser	Leu	Pro	Lys	Glu	Leu	Thr	Ala	Glu	Glu	Glu			
	785				790				795					800			
atc	ggc	ggc	ctt	cag	acg	atc	ctt	cgt	gac	gat	gac	aag	ctc	ccg	cgc	2448	
Ile	Gly	Gly	Leu	Gln	Thr	Ile	Leu	Arg	Asp	Asp	Asp	Lys	Leu	Pro	Arg		
			805						810					815			
aac	gtc	gca	acg	ggc	agt	gtc	aag	gcc	acc	gtc	ggt	gac	acc	ggt	tat	2496	
Asn	Val	Ala	Thr	Gly	Ser	Val	Lys	Ala	Thr	Val	Gly	Asp	Thr	Gly	Tyr		
			820					825					830				
gcc	tct	ggt	gct	gcc	agc	ctc	atc	aag	gct	gcg	ctt	tgc	atc	tac	aac	2544	
Ala	Ser	Gly	Ala	Ala	Ser	Leu	Ile	Lys	Ala	Ala	Leu	Cys	Ile	Tyr	Asn		
	835						840					845					
cgc	tac	ctg	ccc	agc	aac	ggc	gac	gac	tgg	gat	gaa	ccc	gcc	cct	gag	2592	
Arg	Tyr	Leu	Pro	Ser	Asn	Gly	Asp	Asp	Trp	Asp	Glu	Pro	Ala	Pro	Glu		
	850				855						860						
gcg	ccc	tgg	gac	agc	acc	ctc	ttt	gcg	tgc	cag	acc	tcg	cgc	gct	tgg	2640	
Ala	Pro	Trp	Asp	Ser	Thr	Leu	Phe	Ala	Cys	Gln	Thr	Ser	Arg	Ala	Trp		
	865				870				875					880			
ctc	aag	aac	cct	ggc	gag	cgt	cgc	tat	gcg	gcc	gtc	tcg	ggc	gtc	tcc	2688	
Leu	Lys	Asn	Pro	Gly	Glu	Arg	Arg	Tyr	Ala	Ala	Val	Ser	Gly	Val	Ser		
				885				890						895			
gag	acg	cgc	tcg	tgc	tat	tcc	gtg	ctc	ctc	tcc	gaa	gcc	gag	ggc	cac	2736	
Glu	Thr	Arg	Ser	Cys	Tyr	Ser	Val	Leu	Leu	Ser	Glu	Ala	Glu	Gly	His		
			900					905					910				
tac	gag	cgc	gag	aac	cgc	atc	tcg	ctc	gac	gag	gag	gcg	ccc	aag	ctc	2784	
Tyr	Glu	Arg	Glu	Asn	Arg	Ile	Ser	Leu	Asp	Glu	Glu	Ala	Pro	Lys	Leu		
			915				920						925				
att	gtg	ctt	cgc	gcc	gac	tcc	cac	gag	gag	atc	ctt	ggt	cgc	ctc	gac	2832	
Ile	Val	Leu	Arg	Ala	Asp	Ser	His	Glu	Glu	Ile	Leu	Gly	Arg	Leu	Asp		
	930					935						940					
aag	atc	cgc	gag	cgc	ttc	ttg	cag	ccc	acg	ggc	gcc	gcc	ccg	cgc	gag	2880	
Lys	Ile	Arg	Glu	Arg	Phe	Leu	Gln	Pro	Thr	Gly	Ala	Ala	Pro	Arg	Glu		
	945				950					955				960			
tcc	gag	ctc	aag	gcg	cag	gcc	cgc	cgc	atc	ttc	ctc	gag	ctc	ctc	ggc	2928	
Ser	Glu	Leu	Lys	Ala	Gln	Ala	Arg	Arg	Ile	Phe	Leu	Glu	Leu	Leu	Gly		
				965					970					975			
gag	acc	ctt	gcc	cag	gat	gcc	gct	tct	tca	ggc	tcg	caa	aag	ccc	ctc	2976	
Glu	Thr	Leu	Ala	Gln	Asp	Ala	Ala	Ser	Ser	Gly	Ser	Gln	Lys	Pro	Leu		
			980					985					990				
gct	ctc	agc	ctc	gtc	tcc	acg	ccc	tcc	aag	ctc	cag	cgc	gag	gtc	gag	3024	
Ala	Leu	Ser	Leu	Val	Ser	Thr	Pro	Ser	Lys	Leu	Gln	Arg	Glu	Val	Glu		
			995				1000					1005					
ctc	gcg	gcc	aag	ggt	atc	ccg	cgc	tgc	ctc	aag	atg	cgc	cgc	gat		3069	
Leu	Ala	Ala	Lys	Gly	Ile	Pro	Arg	Cys	Leu	Lys	Met	Arg	Arg	Asp			
	1010					1015					1020						
tgg	agc	tcc	cct	gct	ggc	agc	cgc	tac	gcg	cct	gag	ccg	ctc	gcc	3114		
Trp	Ser	Ser	Pro	Ala	Gly	Ser	Arg	Tyr	Ala	Pro	Glu	Pro	Leu	Ala			
	1025					1030					1035						
agc	gac	cgc	gtc	gcc	ttc	atg	tac	ggc	gaa	ggt	cgc	agc	cct	tac	3159		

ES 2 567 305 T3

Ser	Asp	Arg	Val	Ala	Phe	Met	Tyr	Gly	Glu	Gly	Arg	Ser	Pro	Tyr		
	1040					1045					1050					
tac	ggc	atc	acc	caa	gac	att	cac	cgc	att	tgg	ccc	gaa	ctc	cac		3204
Tyr	Gly	Ile	Thr	Gln	Asp	Ile	His	Arg	Ile	Trp	Pro	Glu	Leu	His		
	1055					1060					1065					
gag	gtc	atc	aac	gaa	aag	acg	aac	cgt	ctc	tgg	gcc	gaa	ggc	gac		3249
Glu	Val	Ile	Asn	Glu	Lys	Thr	Asn	Arg	Leu	Trp	Ala	Glu	Gly	Asp		
	1070					1075					1080					
cgc	tgg	gtc	atg	ccg	cgc	gcc	agc	ttc	aag	tcg	gag	ctc	gag	agc		3294
Arg	Trp	Val	Met	Pro	Arg	Ala	Ser	Phe	Lys	Ser	Glu	Leu	Glu	Ser		
	1085					1090					1095					
cag	cag	caa	gag	ttt	gat	cgc	aac	atg	att	gaa	atg	ttc	cgt	ctt		3339
Gln	Gln	Gln	Glu	Phe	Asp	Arg	Asn	Met	Ile	Glu	Met	Phe	Arg	Leu		
	1100					1105					1110					
gga	atc	ctc	acc	tca	att	gcc	ttc	acc	aat	ctg	gcg	gcg	gac	gtt		3384
Gly	Ile	Leu	Thr	Ser	Ile	Ala	Phe	Thr	Asn	Leu	Ala	Arg	Asp	Val		
	1115					1120					1125					
ctc	aac	atc	acg	ccc	aag	gcc	gcc	ttt	ggc	ctc	agt	ctt	ggc	gag		3429
Leu	Asn	Ile	Thr	Pro	Lys	Ala	Ala	Phe	Gly	Leu	Ser	Leu	Gly	Glu		
	1130					1135					1140					
att	tcc	atg	att	ttt	gcc	ttt	tcc	aag	aag	aac	ggt	ctc	atc	tcc		3474
Ile	Ser	Met	Ile	Phe	Ala	Phe	Ser	Lys	Lys	Asn	Gly	Leu	Ile	Ser		
	1145					1150					1155					
gac	cag	ctc	acc	aag	gat	ctt	cgc	gag	tcc	gac	gtg	tgg	aac	aag		3519
Asp	Gln	Leu	Thr	Lys	Asp	Leu	Arg	Glu	Ser	Asp	Val	Trp	Asn	Lys		
	1160					1165					1170					
gct	ctg	gcc	ggt	gaa	ttt	aat	gcg	ctg	cgc	gag	gcc	tgg	ggc	att		3564
Ala	Leu	Ala	Val	Glu	Phe	Asn	Ala	Leu	Arg	Glu	Ala	Trp	Gly	Ile		
	1175					1180					1185					
cca	cag	agt	gtc	ccc	aag	gac	gag	ttc	tgg	caa	ggc	tac	att	gtg		3609
Pro	Gln	Ser	Val	Pro	Lys	Asp	Glu	Phe	Trp	Gln	Gly	Tyr	Ile	Val		
	1190					1195					1200					
cgc	ggc	acc	aag	cag	gat	atc	gag	gcg	gcc	atc	gcc	ccg	gac	agc		3654
Arg	Gly	Thr	Lys	Gln	Asp	Ile	Glu	Ala	Ala	Ile	Ala	Pro	Asp	Ser		
	1205					1210					1215					
aag	tac	gtg	cgc	ctc	acc	atc	atc	aat	gat	gcc	aac	acc	gcc	ctc		3699
Lys	Tyr	Val	Arg	Leu	Thr	Ile	Ile	Asn	Asp	Ala	Asn	Thr	Ala	Leu		
	1220					1225					1230					
att	agc	ggc	aag	ccc	gac	gcc	tgc	aag	gct	gcg	atc	gcg	cgt	ctc		3744
Ile	Ser	Gly	Lys	Pro	Asp	Ala	Cys	Lys	Ala	Ala	Ile	Ala	Arg	Leu		
	1235					1240					1245					
ggt	ggc	aac	att	cct	gcg	ctt	ccc	gtg	acc	cag	ggc	atg	tgc	ggc		3789
Gly	Gly	Asn	Ile	Pro	Ala	Leu	Pro	Val	Thr	Gln	Gly	Met	Cys	Gly		
	1250					1255					1260					
cac	tgc	ccc	gag	gtg	gga	cct	tat	acc	aag	gat	atc	gcc	aag	atc		3834
His	Cys	Pro	Glu	Val	Gly	Pro	Tyr	Thr	Lys	Asp	Ile	Ala	Lys	Ile		
	1265					1270					1275					
cat	gcc	aac	ctt	gag	ttc	ccc	ggt	gtc	gac	ggc	ctt	gac	ctc	tgg		3879
His	Ala	Asn	Leu	Glu	Phe	Pro	Val	Val	Asp	Gly	Leu	Asp	Leu	Trp		
	1280					1285					1290					
acc	aca	atc	aac	cag	aag	cgc	ctc	gtg	cca	cgc	gcc	acg	ggc	gcc		3924
Thr	Thr	Ile	Asn	Gln	Lys	Arg	Leu	Val	Pro	Arg	Ala	Thr	Gly	Ala		
	1295					1300					1305					
aag	gac	gaa	tgg	gcc	cct	tct	tcc	ttt	ggc	gag	tac	gcc	ggc	cag		3969
Lys	Asp	Glu	Trp	Ala	Pro	Ser	Ser	Phe	Gly	Glu	Tyr	Ala	Gly	Gln		
	1310					1315					1320					
ctc	tac	gag	aag	cag	gct	aac	ttc	ccc	caa	atc	gtc	gag	acc	att		4014
Leu	Tyr	Glu	Lys	Gln	Ala	Asn	Phe	Pro	Gln	Ile	Val	Glu	Thr	Ile		
	1325					1330					1335					
tac	aag	caa	aac	tac	gac	gtc	ttt	gtc	gag	ggt	ggg	ccc	aac	aac		4059
Tyr	Lys	Gln	Asn	Tyr	Asp	Val	Phe	Val	Glu	Val	Gly	Pro	Asn	Asn		
	1340					1345					1350					
cac	cgt	agc	acc	gca	gtg	cgc	acc	acg	ctt	ggt	ccc	cag	cgc	aac		4104

ES 2 567 305 T3

His	Arg	Ser	Thr	Ala	Val	Arg	Thr	Thr	Leu	Gly	Pro	Gln	Arg	Asn		
	1355					1360					1365					
cac	ctt	gct	ggc	gcc	atc	gac	aag	cag	aac	gag	gat	gct	tgg	acg	4149	
His	Leu	Ala	Gly	Ala	Ile	Lys	Lys	Gln	Asn	Glu	Asp	Ala	Trp	Thr		
	1370					1375					1380					
acc	atc	gtc	aag	ctt	gtg	gct	tcg	ctc	aag	gcc	cac	ctt	gtt	cct	4194	
Thr	Ile	Val	Lys	Leu	Val	Ala	Ser	Leu	Lys	Ala	His	Leu	Val	Pro		
	1385					1390					1395					
ggc	gtc	acg	atc	tcg	ccg	ctg	tac	cac	tcc	aag	ctt	gtg	gcg	gag	4239	
Gly	Val	Thr	Ile	Ser	Pro	Leu	Tyr	His	Ser	Lys	Leu	Val	Ala	Glu		
	1400					1405					1410					
gct	cag	gct	tgc	tac	gct	gcg	ctc	tgc	aag	ggt	gaa	aag	ccc	aag	4284	
Ala	Gln	Ala	Cys	Tyr	Ala	Ala	Leu	Cys	Lys	Gly	Glu	Lys	Pro	Lys		
	1415					1420					1425					
aag	aac	aag	ttt	gtg	cgc	aag	att	cag	ctc	aac	ggt	cgc	ttc	aac	4329	
Lys	Asn	Lys	Phe	Val	Arg	Lys	Ile	Gln	Leu	Asn	Gly	Arg	Phe	Asn		
	1430					1435					1440					
agc	aag	gcg	gac	ccc	atc	tcc	tcg	gcc	gat	ctt	gcc	agc	ttt	ccg	4374	
Ser	Lys	Ala	Asp	Pro	Ile	Ser	Ser	Ala	Asp	Leu	Ala	Ser	Phe	Pro		
	1445					1450					1455					
cct	gcg	gac	cct	gcc	att	gaa	gcc	gcc	atc	tcg	agc	cgc	atc	atg	4419	
Pro	Ala	Asp	Pro	Ala	Ile	Glu	Ala	Ala	Ile	Ser	Ser	Arg	Ile	Met		
	1460					1465					1470					
aag	cct	gtc	gct	ccc	aag	ttc	tac	gcg	cg	ctc	aac	att	gac	gag	4464	
Lys	Pro	Val	Ala	Pro	Lys	Phe	Tyr	Ala	Arg	Leu	Asn	Ile	Asp	Glu		
	1475					1480					1485					
cag	gac	gag	acc	cga	gat	ccg	atc	ctc	aac	aag	gac	aac	gcg	ccg	4509	
Gln	Asp	Glu	Thr	Arg	Asp	Pro	Ile	Leu	Asn	Lys	Asp	Asn	Ala	Pro		
	1490					1495					1500					
tct	tct	tct	tct	tct	tct	tct	tct	tct	tct	tct	tct	tct	tct	tct	4554	
Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser		
	1505					1510					1515					
ccg	tcg	cct	gct	cct	tcg	gcc	ccc	gtg	caa	aag	aag	gct	gct	ccc	4599	
Pro	Ser	Pro	Ala	Pro	Ser	Ala	Pro	Val	Gln	Lys	Lys	Ala	Ala	Pro		
	1520					1525					1530					
gcc	gcg	gag	acc	aag	gct	gtt	gct	tcg	gct	gac	gca	ctt	cg	agt	4644	
Ala	Ala	Glu	Thr	Lys	Ala	Val	Ala	Ser	Ala	Asp	Ala	Leu	Arg	Ser		
	1535					1540					1545					
gcc	ctg	ctc	gat	ctc	gac	agt	atg	ctt	gcg	ctg	agc	tct	gcc	agt	4689	
Ala	Leu	Leu	Asp	Leu	Asp	Ser	Met	Leu	Ala	Leu	Ser	Ser	Ala	Ser		
	1550					1555					1560					
gcc	tcc	ggc	aac	ctt	gtt	gag	act	gcg	cct	agc	gac	gcc	tcg	gtc	4734	
Ala	Ser	Gly	Asn	Leu	Val	Glu	Thr	Ala	Pro	Ser	Asp	Ala	Ser	Val		
	1565					1570					1575					
att	gtg	ccg	ccc	tgc	aac	att	gcg	gat	ctc	ggc	agc	cg	gcc	ttc	4779	
Ile	Val	Pro	Pro	Cys	Asn	Ile	Ala	Asp	Leu	Gly	Ser	Arg	Ala	Phe		
	1580					1585					1590					
atg	aaa	acg	tac	ggt	gtt	tcg	gcg	cct	ctg	tac	acg	ggc	gcc	atg	4824	
Met	Lys	Thr	Tyr	Gly	Val	Ser	Ala	Pro	Leu	Tyr	Thr	Gly	Ala	Met		
	1595					1600					1605					
gcc	aag	ggc	att	gcc	tct	gcg	gac	ctc	gtc	att	gcc	gcc	ggc	cg	4869	
Ala	Lys	Gly	Ile	Ala	Ser	Ala	Asp	Leu	Val	Ile	Ala	Ala	Gly	Arg		
	1610					1615					1620					
cag	ggc	atc	ctt	gcg	tcc	ttt	ggc	gcc	ggc	gga	ctt	ccc	atg	cag	4914	
Gln	Gly	Ile	Leu	Ala	Ser	Phe	Gly	Ala	Gly	Gly	Leu	Pro	Met	Gln		
	1625					1630					1635					
gtt	gtg	cg	gag	tcc	atc	gaa	aag	att	cag	gcc	gcc	ctg	ccc	aat	4959	
Val	Val	Arg	Glu	Ser	Ile	Glu	Lys	Ile	Gln	Ala	Ala	Leu	Pro	Asn		
	1640					1645					1650					
ggc	ccg	tac	gct	gtc	aac	ctt	atc	cat	tct	ccc	ttt	gac	agc	aac	5004	
Gly	Pro	Tyr	Ala	Val	Asn	Leu	Ile	His	Ser	Pro	Phe	Asp	Ser	Asn		
	1655					1660					1665					
ctc	gaa	aag	ggc	aat	gtc	gat	ctc	ttc	ctc	gag	aag	ggt	gtc	acc	5049	

ES 2 567 305 T3

Leu	Glu	Lys	Gly	Asn	Val	Asp	Leu	Phe	Leu	Glu	Lys	Gly	Val	Thr		
1670						1675					1680					
ttt	gtc	gag	gcc	tcg	gcc	ttt	atg	acg	ctc	acc	ccg	cag	gtc	gtg	5094	
Phe	Val	Glu	Ala	Ser	Ala	Phe	Met	Thr	Leu	Thr	Pro	Gln	Val	Val		
1685						1690					1695					
cgg	tac	cgc	gcg	gct	ggc	ctc	acg	cgc	aac	gcc	gac	ggc	tcg	gtc	5139	
Arg	Tyr	Arg	Ala	Ala	Gly	Leu	Thr	Arg	Asn	Ala	Asp	Gly	Ser	Val		
1700						1705					1710					
aac	atc	cgc	aac	cgt	atc	att	ggc	aag	gtc	tcg	cgc	acc	gag	ctc	5184	
Asn	Ile	Arg	Asn	Arg	Ile	Ile	Gly	Lys	Val	Ser	Arg	Thr	Glu	Leu		
1715						1720					1725					
gcc	gag	atg	ttc	atg	cgt	cct	gcg	ccc	gag	cac	ctt	ctt	cag	aag	5229	
Ala	Glu	Met	Phe	Met	Arg	Pro	Ala	Pro	Glu	His	Leu	Leu	Gln	Lys		
1730						1735					1740					
ctc	att	gct	tcc	ggc	gag	atc	aac	cag	gag	cag	gcc	gag	ctc	gcc	5274	
Leu	Ile	Ala	Ser	Gly	Glu	Ile	Asn	Gln	Glu	Gln	Ala	Glu	Leu	Ala		
1745						1750					1755					
cgc	cgt	ggt	ccc	gtc	gct	gac	gac	atc	gcg	gtc	gaa	gct	gac	tcg	5319	
Arg	Arg	Val	Pro	Val	Ala	Asp	Asp	Ile	Ala	Val	Glu	Ala	Asp	Ser		
1760						1765					1770					
ggt	ggc	cac	acc	gac	aac	cgc	ccc	atc	cac	gtc	att	ctg	ccc	ctc	5364	
Gly	Gly	His	Thr	Asp	Asn	Arg	Pro	Ile	His	Val	Ile	Leu	Pro	Leu		
1775						1780					1785					
atc	atc	aac	ctt	cgc	gac	cgc	ctt	cac	cgc	gag	tgc	ggc	tac	ccg	5409	
Ile	Ile	Asn	Leu	Arg	Asp	Arg	Leu	His	Arg	Glu	Cys	Gly	Tyr	Pro		
1790						1795					1800					
gcc	aac	ctt	cgc	gtc	cgt	gtg	ggc	gcc	ggc	ggt	ggc	att	ggg	tgc	5454	
Ala	Asn	Leu	Arg	Val	Arg	Val	Gly	Ala	Gly	Gly	Gly	Ile	Gly	Cys		
1805						1810					1815					
ccc	cag	gcg	gcg	ctg	gcc	acc	ttc	aac	atg	ggt	gcc	tcc	ttt	att	5499	
Pro	Gln	Ala	Ala	Leu	Ala	Thr	Phe	Asn	Met	Gly	Ala	Ser	Phe	Ile		
1820						1825					1830					
gtc	acc	ggc	acc	gtg	aac	cag	gtc	gcc	aag	cag	tcg	ggc	acg	tgc	5544	
Val	Thr	Gly	Thr	Val	Asn	Gln	Val	Ala	Lys	Gln	Ser	Gly	Thr	Cys		
1835						1840					1845					
gac	aat	gtg	cgc	aag	cag	ctc	gcg	aag	gcc	act	tac	tcg	gac	gta	5589	
Asp	Asn	Val	Arg	Lys	Gln	Leu	Ala	Lys	Ala	Thr	Tyr	Ser	Asp	Val		
1850						1855					1860					
tgc	atg	gcc	ccg	gct	gcc	gac	atg	ttc	gag	gaa	ggc	gtc	aag	ctt	5634	
Cys	Met	Ala	Pro	Ala	Ala	Asp	Met	Phe	Glu	Glu	Gly	Val	Lys	Leu		
1865						1870					1875					
cag	gtc	ctc	aag	aag	gga	acc	atg	ttt	ccc	tcg	cgc	gcc	aac	aag	5679	
Gln	Val	Leu	Lys	Lys	Gly	Thr	Met	Phe	Pro	Ser	Arg	Ala	Asn	Lys		
1880						1885					1890					
ctc	tac	gag	ctc	ttt	tgc	aag	tac	gac	tcg	ttc	gag	tcc	atg	ccc	5724	
Leu	Tyr	Glu	Leu	Phe	Cys	Lys	Tyr	Asp	Ser	Phe	Glu	Ser	Met	Pro		
1895						1900					1905					
ccc	gca	gag	ctt	gcg	cgc	gtc	gag	aag	cgc	atc	ttc	agc	cgc	gcg	5769	
Pro	Ala	Glu	Leu	Ala	Arg	Val	Glu	Lys	Arg	Ile	Phe	Ser	Arg	Ala		
1910						1915					1920					
ctc	gaa	gag	gtc	tgg	gac	gag	acc	aaa	aac	ttt	tac	att	aac	cgt	5814	
Leu	Glu	Glu	Val	Trp	Asp	Glu	Thr	Lys	Asn	Phe	Tyr	Ile	Asn	Arg		
1925						1930					1935					
ctt	cac	aac	ccg	gag	aag	atc	cag	cgc	gcc	gag	cgc	gac	ccc	aag	5859	
Leu	His	Asn	Pro	Glu	Lys	Ile	Gln	Arg	Ala	Glu	Arg	Asp	Pro	Lys		
1940						1945					1950					
ctc	aag	atg	tcg	ctg	tgc	ttt	cgc	tgg	tac	ctg	agc	ctg	gcg	agc	5904	
Leu	Lys	Met	Ser	Leu	Cys	Phe	Arg	Trp	Tyr	Leu	Ser	Leu	Ala	Ser		
1955						1960					1965					
cgc	tgg	gcc	aac	act	gga	gct	tcc	gat	cgc	gtc	atg	gac	tac	cag	5949	
Arg	Trp	Ala	Asn	Thr	Gly	Ala	Ser	Asp	Arg	Val	Met	Asp	Tyr	Gln		
1970						1975					1980					
gtc	tgg	tgc	ggt	cct	gcc	att	ggt	tcc	ttc	aac	gat	ttc	atc	aag	5994	

# ES 2 567 305 T3

	Val	Trp	Cys	Gly	Pro	Ala	Ile	Gly	Ser	Phe	Asn	Asp	Phe	Ile	Lys	
	1985						1990					1995				
	gga	act	tac	ctt	gat	ccg	gcc	gtc	gca	aac	gag	tac	ccg	tgc	gtc	6039
	Gly	Thr	Tyr	Leu	Asp	Pro	Ala	Val	Ala	Asn	Glu	Tyr	Pro	Cys	Val	
	2000						2005					2010				
	ggt	cag	att	aac	aag	cag	atc	ctt	cgt	gga	gcg	tgc	ttc	ttg	cgc	6084
	Val	Gln	Ile	Asn	Lys	Gln	Ile	Leu	Arg	Gly	Ala	Cys	Phe	Leu	Arg	
	2015						2020					2025				
	cgt	ctc	gaa	att	ctg	cgc	aac	gca	cgc	ctt	tcc	gat	ggc	gct	gcc	6129
	Arg	Leu	Glu	Ile	Leu	Arg	Asn	Ala	Arg	Leu	Ser	Asp	Gly	Ala	Ala	
	2030						2035					2040				
	gct	ctt	gtg	gcc	agc	atc	gat	gac	aca	tac	gtc	ccg	gcc	gag	aag	6174
	Ala	Leu	Val	Ala	Ser	Ile	Asp	Asp	Thr	Tyr	Val	Pro	Ala	Glu	Lys	
	2045						2050					2055				
	ctg															6177
	Leu															

<210> 4  
 <211> 2059  
 <212> PRT  
 <213> *Schizochytrium* sp.

<220>  
 <221> misc\_feature  
 <222> (370)..(370)  
 <223> El 'Xaa' en la posición 370 representa Leu.

<220>  
 <221> misc\_feature  
 <222> (371)..(371)  
 <223> El 'Xaa' en la posición 371 representa Ala. o Val.

<400> 4

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

# ES 2 567 305 T3

165					170					175					
Ser	Phe	Val	Ala	Glu	Glu	Leu	Asn	Leu	Gly	Ala	Leu	His	Tyr	Ser	Val
			180					185					190		
Asp	Ala	Ala	Cys	Ala	Thr	Ala	Leu	Tyr	Val	Leu	Arg	Leu	Ala	Gln	Asp
		195					200					205			
His	Leu	Val	Ser	Gly	Ala	Ala	Asp	Val	Met	Leu	Cys	Gly	Ala	Thr	Cys
	210					215					220				
Leu	Pro	Glu	Pro	Phe	Phe	Ile	Leu	Ser	Gly	Phe	Ser	Thr	Phe	Gln	Ala
225					230					235					240
Met	Pro	Val	Gly	Thr	Gly	Gln	Asn	Val	Ser	Met	Pro	Leu	His	Lys	Asp
				245					250					255	
Ser	Gln	Gly	Leu	Thr	Pro	Gly	Glu	Gly	Gly	Ser	Ile	Met	Val	Leu	Lys
			260					265					270		
Arg	Leu	Asp	Asp	Ala	Ile	Arg	Asp	Gly	Asp	His	Ile	Tyr	Gly	Thr	Leu
		275					280					285			
Leu	Gly	Ala	Asn	Val	Ser	Asn	Ser	Gly	Thr	Gly	Leu	Pro	Leu	Lys	Pro
	290					295					300				
Leu	Leu	Pro	Ser	Glu	Lys	Lys	Cys	Leu	Met	Asp	Thr	Tyr	Thr	Arg	Ile
305					310					315					320
Asn	Val	His	Pro	His	Lys	Ile	Gln	Tyr	Val	Glu	Cys	His	Ala	Thr	Gly
				325					330					335	
Thr	Pro	Gln	Gly	Asp	Arg	Val	Glu	Ile	Asp	Ala	Val	Lys	Ala	Cys	Phe
			340					345					350		
Glu	Gly	Lys	Val	Pro	Arg	Phe	Gly	Thr	Thr	Lys	Gly	Asn	Phe	Gly	His
		355					360					365			
Thr	Xaa	Xaa	Ala	Ala	Gly	Phe	Ala	Gly	Met	Cys	Lys	Val	Leu	Leu	Ser
	370					375					380				
Met	Lys	His	Gly	Ile	Ile	Pro	Pro	Thr	Pro	Gly	Ile	Asp	Asp	Glu	Thr
385					390					395				400	
Lys	Met	Asp	Pro	Leu	Val	Val	Ser	Gly	Glu	Ala	Ile	Pro	Trp	Pro	Glu
				405					410					415	
Thr	Asn	Gly	Glu	Pro	Lys	Arg	Ala	Gly	Leu	Ser	Ala	Phe	Gly	Phe	Gly
			420					425					430		
Gly	Thr	Asn	Ala	His	Ala	Val	Phe	Glu	Glu	His	Asp	Pro	Ser	Asn	Ala
		435					440					445			
Ala	Cys	Thr	Gly	His	Asp	Ser	Ile	Ser	Ala	Leu	Ser	Ala	Arg	Cys	Gly
	450					455					460				
Gly	Glu	Ser	Asn	Met	Arg	Ile	Ala	Ile	Thr	Gly	Met	Asp	Ala	Thr	Phe
465					470					475					480
Gly	Ala	Leu	Lys	Gly	Leu	Asp	Ala	Phe	Glu	Arg	Ala	Ile	Tyr	Thr	Gly
				485					490					495	
Ala	His	Gly	Ala	Ile	Pro	Leu	Pro	Glu	Lys	Arg	Trp	Arg	Phe	Leu	Gly

ES 2 567 305 T3

500					505					510					
Lys	Asp	Lys	Asp	Phe	Leu	Asp	Leu	Cys	Gly	Val	Lys	Ala	Thr	Pro	His
		515					520					525			
Gly	Cys	Tyr	Ile	Glu	Asp	Val	Glu	Val	Asp	Phe	Gln	Arg	Leu	Arg	Thr
	530					535					540				
Pro	Met	Thr	Pro	Glu	Asp	Met	Leu	Leu	Pro	Gln	Gln	Leu	Leu	Ala	Val
545					550					555					560
Thr	Thr	Ile	Asp	Arg	Ala	Ile	Leu	Asp	Ser	Gly	Met	Lys	Lys	Gly	Gly
				565					570					575	
Asn	Val	Ala	Val	Phe	Val	Gly	Leu	Gly	Thr	Asp	Leu	Glu	Leu	Tyr	Arg
			580					585					590		
His	Arg	Ala	Arg	Val	Ala	Leu	Lys	Glu	Arg	Val	Arg	Pro	Glu	Ala	Ser
		595					600					605			
Lys	Lys	Leu	Asn	Asp	Met	Met	Gln	Tyr	Ile	Asn	Asp	Cys	Gly	Thr	Ser
	610					615					620				
Thr	Ser	Tyr	Thr	Ser	Tyr	Ile	Gly	Asn	Leu	Val	Ala	Thr	Arg	Val	Ser
625					630					635					640
Ser	Gln	Trp	Gly	Phe	Thr	Gly	Pro	Ser	Phe	Thr	Ile	Thr	Glu	Gly	Asn
				645					650					655	
Asn	Ser	Val	Tyr	Arg	Cys	Ala	Glu	Leu	Gly	Lys	Tyr	Leu	Leu	Glu	Thr
			660					665					670		
Gly	Glu	Val	Asp	Gly	Val	Val	Val	Ala	Gly	Val	Asp	Leu	Cys	Gly	Ser
		675					680					685			
Ala	Glu	Asn	Leu	Tyr	Val	Lys	Ser	Arg	Arg	Phe	Lys	Val	Ser	Thr	Ser
	690					695					700				
Asp	Thr	Pro	Arg	Ala	Ser	Phe	Asp	Ala	Ala	Ala	Asp	Gly	Tyr	Phe	Val
705					710					715					720
Gly	Glu	Gly	Cys	Gly	Ala	Phe	Val	Leu	Lys	Arg	Glu	Thr	Ser	Cys	Thr
				725					730					735	
Lys	Asp	Asp	Arg	Ile	Tyr	Ala	Cys	Met	Asp	Ala	Ile	Val	Pro	Gly	Asn
			740					745					750		
Val	Pro	Ser	Ala	Cys	Leu	Arg	Glu	Ala	Leu	Asp	Gln	Ala	Arg	Val	Lys
		755					760					765			
Pro	Gly	Asp	Ile	Glu	Met	Leu	Glu	Leu	Ser	Ala	Asp	Ser	Ala	Arg	His
	770					775					780				
Leu	Lys	Asp	Pro	Ser	Val	Leu	Pro	Lys	Glu	Leu	Thr	Ala	Glu	Glu	Glu
785					790					795					800
Ile	Gly	Gly	Leu	Gln	Thr	Ile	Leu	Arg	Asp	Asp	Asp	Lys	Leu	Pro	Arg
				805					810					815	
Asn	Val	Ala	Thr	Gly	Ser	Val	Lys	Ala	Thr	Val	Gly	Asp	Thr	Gly	Tyr
			820					825					830		
Ala	Ser	Gly	Ala	Ala	Ser	Leu	Ile	Lys	Ala	Ala	Leu	Cys	Ile	Tyr	Asn

# ES 2 567 305 T3

835	840	845												
Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu 850 855 860														
Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp 865 870 875 880														
Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser 885 890 895														
Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His 900 905 910														
Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu 915 920 925														
Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp 930 935 940														
Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu 945 950 955 960														
Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly 965 970 975														
Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu 980 985 990														
Ala Leu Ser Leu Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu 995 1000 1005														
Leu Ala Ala Lys Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp 1010 1015 1020														
Trp Ser Ser Pro Ala Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala 1025 1030 1035														
Ser Asp Arg Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr 1040 1045 1050														
Tyr Gly Ile Thr Gln Asp Ile His Arg Ile Trp Pro Glu Leu His 1055 1060 1065														
Glu Val Ile Asn Glu Lys Thr Asn Arg Leu Trp Ala Glu Gly Asp 1070 1075 1080														
Arg Trp Val Met Pro Arg Ala Ser Phe Lys Ser Glu Leu Glu Ser 1085 1090 1095														
Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu Met Phe Arg Leu 1100 1105 1110														
Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala Arg Asp Val 1115 1120 1125														
Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu Gly Glu 1130 1135 1140														
Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile Ser 1145 1150 1155														
Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys														



# ES 2 567 305 T3

1160	1165	1170
Ala Leu Ala Val Glu Phe Asn 1175	Ala Leu Arg Glu Ala 1180	Ala Trp Gly Ile 1185
Pro Gln Ser Val Pro Lys Asp 1190	Glu Phe Trp Gln Gly 1195	Tyr Ile Val 1200
Arg Gly Thr Lys Gln Asp 1205	Ile Glu Ala Ala Ile 1210	Ala Pro Asp Ser 1215
Lys Tyr Val Arg Leu Thr 1220	Ile Asn Asp Ala 1225	Asn Thr Ala Leu 1230
Ile Ser Gly Lys Pro Asp 1235	Ala Cys Lys Ala Ala 1240	Ile Ala Arg Leu 1245
Gly Gly Asn Ile Pro Ala 1250	Leu Pro Val Thr Gln 1255	Gly Met Cys Gly 1260
His Cys Pro Glu Val Gly 1265	Pro Tyr Thr Lys Asp 1270	Ile Ala Lys Ile 1275
His Ala Asn Leu Glu Phe 1280	Pro Val Val Asp Gly 1285	Leu Asp Leu Trp 1290
Thr Thr Ile Asn Gln Lys 1295	Arg Leu Val Pro Arg 1300	Ala Thr Gly Ala 1305
Lys Asp Glu Trp Ala Pro 1310	Ser Ser Phe Gly Glu 1315	Tyr Ala Gly Gln 1320
Leu Tyr Glu Lys Gln Ala 1325	Asn Phe Pro Gln Ile 1330	Val Glu Thr Ile 1335
Tyr Lys Gln Asn Tyr Asp 1340	Val Phe Val Glu Val 1345	Gly Pro Asn Asn 1350
His Arg Ser Thr Ala Val 1355	Arg Thr Thr Leu Gly 1360	Pro Gln Arg Asn 1365
His Leu Ala Gly Ala Ile 1370	Asp Lys Gln Asn Glu 1375	Asp Ala Trp Thr 1380
Thr Ile Val Lys Leu Val 1385	Ala Ser Leu Lys Ala 1390	His Leu Val Pro 1395
Gly Val Thr Ile Ser Pro 1400	Leu Tyr His Ser Lys 1405	Leu Val Ala Glu 1410
Ala Gln Ala Cys Tyr Ala 1415	Ala Leu Cys Lys Gly 1420	Glu Lys Pro Lys 1425
Lys Asn Lys Phe Val Arg 1430	Lys Ile Gln Leu Asn 1435	Gly Arg Phe Asn 1440
Ser Lys Ala Asp Pro Ile 1445	Ser Ser Ala Asp Leu 1450	Ala Ser Phe Pro 1455
Pro Ala Asp Pro Ala Ile 1460	Glu Ala Ala Ile Ser 1465	Ser Arg Ile Met 1470
Lys Pro Val Ala Pro Lys 1475	Phe Tyr Ala Arg Leu 1480	Asn Ile Asp Glu 1485

# ES 2 567 305 T3

1475	1480	1485
Gln Asp 1490	Glu Thr Arg Asp 1495	Pro Ile Leu Asn Lys 1495
Ser Ser 1505	Ser Ser Ser Ser 1510	Ser Ser Ser Ser 1515
Pro Ser 1520	Pro Ala Pro Ser 1525	Ala Pro Val Gln Lys 1530
Ala Ala 1535	Glu Thr Lys Ala 1540	Val Ala Ser Ala Asp 1545
Ala Leu 1550	Leu Asp Leu Asp 1555	Ser Met Leu Ala Leu 1560
Ala Ser 1565	Gly Asn Leu Val 1570	Glu Thr Ala Pro Ser 1575
Ile Val 1580	Pro Pro Cys Asn 1585	Ile Ala Asp Leu Gly 1590
Met Lys 1595	Thr Tyr Gly Val 1600	Ser Ala Pro Leu Tyr 1605
Ala Lys 1610	Gly Ile Ala Ser 1615	Ala Asp Leu Val Ile 1620
Gln Gly 1625	Ile Leu Ala Ser 1630	Phe Gly Ala Gly Gly 1635
Val Val 1640	Arg Glu Ser Ile 1645	Glu Lys Ile Gln Ala 1650
Gly Pro 1655	Tyr Ala Val Asn 1660	Leu Ile His Ser Pro 1665
Leu Glu 1670	Lys Gly Asn Val 1675	Asp Leu Phe Leu Glu 1680
Phe Val 1685	Glu Ala Ser Ala 1690	Phe Met Thr Leu Thr 1695
Arg Tyr 1700	Arg Ala Ala Gly 1705	Leu Thr Arg Asn Ala 1710
Asn Ile 1715	Arg Asn Arg Ile 1720	Ile Gly Lys Val Ser 1725
Ala Glu 1730	Met Phe Met Arg 1735	Pro Ala Pro Glu His 1740
Leu Ile 1745	Ala Ser Gly Glu 1750	Ile Asn Gln Glu Gln 1755
Arg Arg 1760	Val Pro Val Ala 1765	Asp Asp Ile Ala Val 1770
Gly Gly 1775	His Thr Asp Asn 1780	Arg Pro Ile His Val 1785
Ile Ile	Asn Leu Arg Asp 1790	Arg Leu His Arg Glu 1795
		Cys Gly Tyr Pro

ES 2 567 305 T3

1790	1795	1800
Ala Asn Leu Arg Val Arg	Val Gly Ala Gly Gly Gly	Ile Gly Cys
1805	1810	1815
Pro Gln Ala Ala Leu Ala	Thr Phe Asn Met Gly Ala	Ser Phe Ile
1820	1825	1830
Val Thr Gly Thr Val Asn	Gln Val Ala Lys Gln Ser	Gly Thr Cys
1835	1840	1845
Asp Asn Val Arg Lys Gln	Leu Ala Lys Ala Thr Tyr	Ser Asp Val
1850	1855	1860
Cys Met Ala Pro Ala Ala	Asp Met Phe Glu Glu Gly	Val Lys Leu
1865	1870	1875
Gln Val Leu Lys Lys Gly	Thr Met Phe Pro Ser Arg	Ala Asn Lys
1880	1885	1890
Leu Tyr Glu Leu Phe Cys	Lys Tyr Asp Ser Phe Glu	Ser Met Pro
1895	1900	1905
Pro Ala Glu Leu Ala Arg	Val Glu Lys Arg Ile Phe	Ser Arg Ala
1910	1915	1920
Leu Glu Glu Val Trp Asp	Glu Thr Lys Asn Phe Tyr	Ile Asn Arg
1925	1930	1935
Leu His Asn Pro Glu Lys	Ile Gln Arg Ala Glu Arg	Asp Pro Lys
1940	1945	1950
Leu Lys Met Ser Leu Cys	Phe Arg Trp Tyr Leu Ser	Leu Ala Ser
1955	1960	1965
Arg Trp Ala Asn Thr Gly	Ala Ser Asp Arg Val Met	Asp Tyr Gln
1970	1975	1980
Val Trp Cys Gly Pro Ala	Ile Gly Ser Phe Asn Asp	Phe Ile Lys
1985	1990	1995
Gly Thr Tyr Leu Asp Pro	Ala Val Ala Asn Glu Tyr	Pro Cys Val
2000	2005	2010
Val Gln Ile Asn Lys Gln	Ile Leu Arg Gly Ala Cys	Phe Leu Arg
2015	2020	2025
Arg Leu Glu Ile Leu Arg	Asn Ala Arg Leu Ser Asp	Gly Ala Ala
2030	2035	2040
Ala Leu Val Ala Ser Ile	Asp Asp Thr Tyr Val Pro	Ala Glu Lys
2045	2050	2055

Leu

5 <210> 5  
 <211> 4509  
 <212> ADN  
 <213> *Schizochytrium* sp.

10 <220>  
 <221> CDS  
 <222> (1)..(4509)  
 <223>

<400> 5



ES 2 567 305 T3

305					310					315					320	
gtc	atg	gcc	gga	tcc	ctc	gtc	tcc	gac	ggc	tgc	agc	cag	atg	ctc	aag	1008
Val	Met	Ala	Gly	Ser	Leu	Val	Ser	Asp	Gly	Cys	Ser	Gln	Met	Leu	Lys	
				325					330					335		
atg	tac	atg	atc	tgg	ctc	ggc	ctc	cac	ctc	acc	acc	gga	ccc	ttt	gac	1056
Met	Tyr	Met	Ile	Trp	Leu	Gly	Leu	His	Leu	Thr	Thr	Gly	Pro	Phe	Asp	
			340					345						350		
ttc	cgc	ccg	gtc	aac	ggc	cac	ccc	aac	aag	gtc	cgc	tgc	cgc	ggc	caa	1104
Phe	Arg	Pro	Val	Asn	Gly	His	Pro	Asn	Lys	Val	Arg	Cys	Arg	Gly	Gln	
		355					360					365				
atc	tcc	ccg	cac	aag	ggc	aag	ctc	gtc	tac	gtc	atg	gag	atc	aag	gag	1152
Ile	Ser	Pro	His	Lys	Gly	Lys	Leu	Val	Tyr	Val	Met	Glu	Ile	Lys	Glu	
	370					375						380				
atg	ggc	ttc	gac	gag	gac	aac	gac	ccg	tac	gcc	att	gcc	gac	gtc	aac	1200
Met	Gly	Phe	Asp	Glu	Asp	Asn	Asp	Pro	Tyr	Ala	Ile	Ala	Asp	Val	Asn	
	385				390					395					400	
atc	att	gat	gtc	gac	ttc	gaa	aag	ggc	cag	gac	ttt	agc	ctc	gac	cgc	1248
Ile	Ile	Asp	Val	Asp	Phe	Glu	Lys	Gly	Gln	Asp	Phe	Ser	Leu	Asp	Arg	
			405					410						415		
atc	agc	gac	tac	ggc	aag	ggc	gac	ctc	aac	aag	aag	atc	gtc	gtc	gac	1296
Ile	Ser	Asp	Tyr	Gly	Lys	Gly	Asp	Leu	Asn	Lys	Lys	Ile	Val	Val	Asp	
			420					425					430			
ttt	aag	ggc	atc	gct	ctc	aag	atg	cag	aag	cgc	tcc	acc	aac	aag	aac	1344
Phe	Lys	Gly	Ile	Ala	Leu	Lys	Met	Gln	Lys	Arg	Ser	Thr	Asn	Lys	Asn	
		435					440					445				
ccc	tcc	aag	ggt	cag	ccc	gtc	ttt	gcc	aac	ggc	gcc	gcc	act	gtc	ggc	1392
Pro	Ser	Lys	Val	Gln	Pro	Val	Phe	Ala	Asn	Gly	Ala	Ala	Thr	Val	Gly	
		450				455					460					
ccc	gag	gcc	tcc	aag	gct	tcc	tcc	ggc	gcc	agc	gcc	agc	gcc	agc	gcc	1440
Pro	Glu	Ala	Ser	Lys	Ala	Ser	Ser	Gly	Ala	Ser	Ala	Ser	Ala	Ser	Ala	
	465				470				475						480	
gcc	ccg	gcc	aag	cct	gcc	ttc	agc	gcc	gat	ggt	ctt	gcg	ccc	aag	ccc	1488
Ala	Pro	Ala	Lys	Pro	Ala	Phe	Ser	Ala	Asp	Val	Leu	Ala	Pro	Lys	Pro	
			485					490					495			
ggt	gcc	ctt	ccc	gag	cac	atc	ctc	aag	ggc	gac	gcc	ctc	gcc	ccc	aag	1536
Val	Ala	Leu	Pro	Glu	His	Ile	Leu	Lys	Gly	Asp	Ala	Leu	Ala	Pro	Lys	
			500					505					510			
gag	atg	tcc	tgg	cac	ccc	atg	gcc	cgc	atc	ccg	ggc	aac	ccg	acg	ccc	1584
Glu	Met	Ser	Trp	His	Pro	Met	Ala	Arg	Ile	Pro	Gly	Asn	Pro	Thr	Pro	
		515				520						525				
tct	ttt	gcg	ccc	tcg	gcc	tac	aag	ccg	cgc	aac	atc	gcc	ttt	acg	ccc	1632
Ser	Phe	Ala	Pro	Ser	Ala	Tyr	Lys	Pro	Arg	Asn	Ile	Ala	Phe	Thr	Pro	
			530			535					540					
ttc	ccc	ggc	aac	ccc	aac	gat	aac	gac	cac	acc	ccg	ggc	aag	atg	ccg	1680
Phe	Pro	Gly	Asn	Pro	Asn	Asp	Asn	Asp	His	Thr	Pro	Gly	Lys	Met	Pro	
			545		550				555						560	
ctc	acc	tgg	ttc	aac	atg	gcc	gag	ttc	atg	gcc	ggc	aag	gtc	agc	atg	1728
Leu	Thr	Trp	Phe	Asn	Met	Ala	Glu	Phe	Met	Ala	Gly	Lys	Val	Ser	Met	
			565					570						575		
tgc	ctc	ggc	ccc	gag	ttc	gcc	aag	ttc	gac	gac	tcg	aac	acc	agc	cgc	1776
Cys	Leu	Gly	Pro	Glu	Phe	Ala	Lys	Phe	Asp	Asp	Ser	Asn	Thr	Ser	Arg	
			580					585					590			
agc	ccc	gct	tgg	gac	ctc	gct	ctc	gtc	acc	cgc	gcc	gtg	tct	gtg	tct	1824
Ser	Pro	Ala	Trp	Asp	Leu	Ala	Leu	Val	Thr	Arg	Ala	Val	Ser	Val	Ser	
			595			600						605				
gac	ctc	aag	cac	gtc	aac	tac	cgc	aac	atc	gac	ctc	gac	ccc	tcc	aag	1872
Asp	Leu	Lys	His	Val	Asn	Tyr	Arg	Asn	Ile	Asp	Leu	Asp	Pro	Ser	Lys	
			610			615						620				
ggt	acc	atg	gtc	ggc	gag	ttc	gac	tgc	ccc	gcg	gac	gcc	tgg	ttc	tac	1920
Gly	Thr	Met	Val	Gly	Glu	Phe	Asp	Cys	Pro	Ala	Asp	Ala	Trp	Phe	Tyr	
			625		630				635						640	
aag	ggc	gcc	tgc	aac	gat	gcc	cac	atg	ccg	tac	tcg	atc	ctc	atg	gag	1968
Lys	Gly	Ala	Cys	Asn	Asp	Ala	His	Met	Pro	Tyr	Ser	Ile	Leu	Met	Glu	

ES 2 567 305 T3

				645					650				655					
atc	gcc	ctc	cag	acc	tcg	ggt	gtg	ctc	acc	tcg	gtg	ctc	aag	gcg	ccc			2016
Ile	Ala	Leu	Gln	Thr	Ser	Gly	Val	Leu	Thr	Ser	Val	Leu	Lys	Ala	Pro			
			660					665					670					
ctg	acc	atg	gag	aag	gac	gac	atc	ctc	ttc	cgc	aac	ctc	gac	gcc	aac			2064
Leu	Thr	Met	Glu	Lys	Asp	Asp	Ile	Leu	Phe	Arg	Asn	Leu	Asp	Ala	Asn			
			675					680					685					
gcc	gag	ttc	gtg	cgc	gcc	gac	ctc	gac	tac	cgc	ggc	aag	act	atc	cgc			2112
Ala	Glu	Phe	Val	Arg	Ala	Asp	Leu	Asp	Tyr	Arg	Gly	Lys	Thr	Ile	Arg			
			690					695					700					
aac	gtc	acc	aag	tgc	act	ggc	tac	agc	atg	ctc	ggc	gag	atg	ggc	gtc			2160
Asn	Val	Thr	Lys	Cys	Thr	Gly	Tyr	Ser	Met	Leu	Gly	Glu	Met	Gly	Val			
						710									720			
cac	cgc	ttc	acc	ttt	gag	ctc	tac	gtc	gat	gat	gtg	ctc	ttt	tac	aag			2208
His	Arg	Phe	Thr	Phe	Glu	Leu	Tyr	Val	Asp	Asp	Val	Leu	Phe	Tyr	Lys			
				725											735			
ggc	tcg	acc	tcg	ttc	ggc	tgg	ttc	gtg	ccc	gag	gtc	ttt	gcc	gcc	cag			2256
Gly	Ser	Thr	Ser	Phe	Gly	Trp	Phe	Val	Pro	Glu	Val	Phe	Ala	Ala	Gln			
				740											750			
gcc	ggc	ctc	gac	aac	ggc	cgc	aag	tcg	gag	ccc	tgg	ttc	att	gag	aac			2304
Ala	Gly	Leu	Asp	Asn	Gly	Arg	Lys	Ser	Glu	Pro	Trp	Phe	Ile	Glu	Asn			
				755											765			
aag	gtt	ccg	gcc	tcg	cag	gtc	tcc	tcc	ttt	gac	gtg	cgc	ccc	aac	ggc			2352
Lys	Val	Pro	Ala	Ser	Gln	Val	Ser	Ser	Phe	Asp	Val	Arg	Pro	Asn	Gly			
							775								780			
agc	ggc	cgc	acc	gcc	atc	ttc	gcc	aac	gcc	ccc	agc	ggc	gcc	cag	ctc			2400
Ser	Gly	Arg	Thr	Ala	Ile	Phe	Ala	Asn	Ala	Pro	Ser	Gly	Ala	Gln	Leu			
						790									800			
aac	cgc	cgc	acg	gac	cag	ggc	cag	tac	ctc	gac	gcc	gtc	gac	att	gtc			2448
Asn	Arg	Arg	Thr	Asp	Gln	Gly	Gln	Tyr	Leu	Asp	Ala	Val	Asp	Ile	Val			
															815			
tcc	ggc	agc	ggc	aag	agc	ctc	ggc	tac	gcc	cac	ggt	tcc	aag	acg				2496
Ser	Gly	Ser	Gly	Lys	Lys	Ser	Leu	Gly	Tyr	Ala	His	Gly	Ser	Lys	Thr			
				820											830			
gtc	aac	ccg	aac	gac	tgg	ttc	ttc	tcg	tgc	cac	ttt	tgg	ttt	gac	tcg			2544
Val	Asn	Pro	Asn	Asp	Trp	Phe	Phe	Ser	Cys	His	Phe	Trp	Phe	Asp	Ser			
															845			
gtc	atg	ccc	gga	agt	ctc	ggt	gtc	gag	tcc	atg	ttc	cag	ctc	gtc	gag			2592
Val	Met	Pro	Gly	Ser	Leu	Gly	Val	Glu	Ser	Met	Phe	Gln	Leu	Val	Glu			
															860			
gcc	atc	gcc	gcc	cac	gag	gat	ctc	gct	ggc	aaa	gca	cgg	cat	tgc	caa			2640
Ala	Ile	Ala	Ala	His	Glu	Asp	Leu	Ala	Gly	Lys	Ala	Arg	His	Cys	Gln			
															880			
ccc	cac	ctt	tgt	gca	cgc	ccc	cgg	gca	aga	tca	agc	tgg	aag	tac	cgc			2688
Pro	His	Leu	Cys	Ala	Arg	Pro	Arg	Ala	Arg	Ser	Ser	Trp	Lys	Tyr	Arg			
															895			
ggc	cag	ctc	acg	ccc	aag	agc	aag	aag	atg	gac	tcg	gag	gtc	cac	atc			2736
Gly	Gln	Leu	Thr	Pro	Lys	Ser	Lys	Lys	Met	Asp	Ser	Glu	Val	His	Ile			
															910			
gtg	tcc	gtg	gac	gcc	cac	gac	ggc	ggt	gtc	gac	ctc	gtc	gcc	gac	ggc			2784
Val	Ser	Val	Asp	Ala	His	Asp	Gly	Val	Val	Val	Asp	Leu	Val	Ala	Asp	Gly		
															925			
ttc	ctc	tgg	gcc	gac	agc	ctc	cgc	gtc	tac	tcg	gtg	agc	aac	att	cgc			2832
Phe	Leu	Trp	Ala	Asp	Ser	Leu	Arg	Val	Tyr	Ser	Val	Ser	Asn	Ile	Arg			
															940			
gtg	cgc	atc	gcc	tcc	ggt	gag	gcc	cct	gcc	gcc	gcc	tcc	tcc	gcc	gcc			2880
Val	Arg	Ile	Ala	Ser	Gly	Glu	Ala	Pro	Ala	Ala	Ala	Ser	Ser	Ala	Ala			
															960			
tct	gtg	ggc	tcc	tcg	gct	tcg	tcc	gtc	gag	cgc	acg	cgc	tcg	agc	ccc			2928
Ser	Val	Gly	Ser	Ser	Ala	Ser	Ser	Val	Glu	Arg	Thr	Arg	Ser	Ser	Pro			
															975			
gct	gtc	gcc	tcc	ggc	ccg	gcc	cag	acc	atc	gac	ctc	aag	cag	ctc	aag			2976
Ala	Val	Ala	Ser	Gly	Pro	Ala	Gln	Thr	Ile	Asp	Leu	Lys	Gln	Leu	Lys			

ES 2 567 305 T3

		980					985					990					
acc	gag	ctc	ctc	gag	ctc	gat	gcc	ccg	ctc	tac	ctc	tcg	cag	gac	ccg	3024	
Thr	Glu	Leu	Leu	Glu	Leu	Asp	Ala	Pro	Leu	Tyr	Leu	Ser	Gln	Asp	Pro		
		995					1000					1005					
acc	agc	ggc	cag	ctc	aag	aag	cac	acc	gac	gtg	gcc	tcc	ggc	cag		3069	
Thr	Ser	Gly	Gln	Leu	Lys	Lys	His	Thr	Asp	Val	Ala	Ser	Gly	Gln			
	1010					1015					1020						
gcc	acc	atc	gtg	cag	ccc	tgc	acg	ctc	ggc	gac	ctc	ggg	gac	cgc		3114	
Ala	Thr	Ile	Val	Gln	Pro	Cys	Thr	Leu	Gly	Asp	Leu	Gly	Asp	Arg			
	1025					1030					1035						
tcc	ttc	atg	gag	acc	tac	ggc	gtc	gtc	gcc	ccg	ctg	tac	acg	ggc		3159	
Ser	Phe	Met	Glu	Thr	Tyr	Gly	Val	Val	Ala	Pro	Leu	Tyr	Thr	Gly			
	1040					1045					1050						
gcc	atg	gcc	aag	ggc	att	gcc	tcg	gcg	gac	ctc	gtc	atc	gcc	gcc		3204	
Ala	Met	Ala	Lys	Gly	Ile	Ala	Ser	Ala	Asp	Leu	Val	Ile	Ala	Ala			
	1055					1060					1065						
ggc	aag	cgc	aag	atc	ctc	ggc	tcc	ttt	ggc	gcc	ggc	ggc	ctc	ccc		3249	
Gly	Lys	Arg	Lys	Ile	Leu	Gly	Ser	Phe	Gly	Ala	Gly	Gly	Leu	Pro			
	1070					1075					1080						
atg	cac	cac	gtg	cgc	gcc	ccc	ctc	gag	aag	atc	cag	gcc	gcc	ctg		3294	
Met	His	His	Val	Arg	Ala	Ala	Leu	Glu	Lys	Ile	Gln	Ala	Ala	Leu			
	1085					1090					1095						
cct	cag	ggc	ccc	tac	gcc	gtc	aac	ctc	atc	cac	tcg	cct	ttt	gac		3339	
Pro	Gln	Gly	Pro	Tyr	Ala	Val	Asn	Leu	Ile	His	Ser	Pro	Phe	Asp			
	1100					1105					1110						
agc	aac	ctc	gag	aag	ggc	aac	gtc	gat	ctc	ttc	ctc	gag	aag	ggc		3384	
Ser	Asn	Leu	Glu	Lys	Gly	Asn	Val	Asp	Leu	Phe	Leu	Glu	Lys	Gly			
	1115					1120					1125						
gtc	act	gtg	gtg	gag	gcc	tcg	gca	ttc	atg	acc	ctc	acc	ccg	cag		3429	
Val	Thr	Val	Val	Glu	Ala	Ser	Ala	Phe	Met	Thr	Leu	Thr	Pro	Gln			
	1130					1135					1140						
gtc	gtg	cgc	tac	cgc	gcc	ccc	ggc	ctc	tcg	cgc	aac	gcc	gac	ggg		3474	
Val	Val	Arg	Tyr	Arg	Ala	Ala	Gly	Leu	Ser	Arg	Asn	Ala	Asp	Gly			
	1145					1150					1155						
tcg	gtc	aac	atc	cgc	aac	cgc	atc	atc	ggc	aag	gtc	tcg	cgc	acc		3519	
Ser	Val	Asn	Ile	Arg	Asn	Arg	Ile	Ile	Gly	Lys	Val	Ser	Arg	Thr			
	1160					1165					1170						
gag	ctc	gcc	gag	atg	ttc	atc	cgc	ccg	gcc	ccg	gag	cac	ctc	ctc		3564	
Glu	Leu	Ala	Glu	Met	Phe	Ile	Arg	Pro	Ala	Pro	Glu	His	Leu	Leu			
	1175					1180					1185						
gag	aag	ctc	atc	gcc	tcg	ggc	gag	atc	acc	cag	gag	cag	gcc	gag		3609	
Glu	Lys	Leu	Ile	Ala	Ser	Gly	Glu	Ile	Thr	Gln	Glu	Gln	Ala	Glu			
	1190					1195					1200						
ctc	gcg	cgc	cgc	ggt	ccc	gtc	gcc	gac	gat	atc	gct	gtc	gag	gct		3654	
Leu	Ala	Arg	Arg	Val	Pro	Val	Ala	Asp	Asp	Ile	Ala	Val	Glu	Ala			
	1205					1210					1215						
gac	tcg	ggc	ggc	cac	acc	gac	aac	cgc	ccc	atc	cac	gtc	atc	ctc		3699	
Asp	Ser	Gly	Gly	His	Thr	Asp	Asn	Arg	Pro	Ile	His	Val	Ile	Leu			
	1220					1225					1230						
ccg	ctc	atc	atc	aac	ctc	cgc	aac	cgc	ctg	cac	cgc	gag	tgc	ggc		3744	
Pro	Leu	Ile	Ile	Asn	Leu	Arg	Asn	Arg	Leu	His	Arg	Glu	Cys	Gly			
	1235					1240					1245						
tac	ccc	gcg	cac	ctc	cgc	gtc	cgc	ggt	ggc	gcc	ggc	ggg	ggc	gtc		3789	
Tyr	Pro	Ala	His	Leu	Arg	Val	Arg	Val	Gly	Ala	Gly	Gly	Gly	Val			
	1250					1255					1260						
ggc	tgc	ccg	cag	gcc	gcc	gcc	gcc	gcg	ctc	acc	atg	ggc	gcc	gcc		3834	
Gly	Cys	Pro	Gln	Ala	Ala	Ala	Ala	Ala	Leu	Thr	Met	Gly	Ala	Ala			
	1265					1270					1275						
ttc	atc	gtc	acc	ggc	act	gtc	aac	cag	gtc	gcc	aag	cag	tcc	ggc		3879	
Phe	Ile	Val	Thr	Gly	Thr	Val	Asn	Gln	Val	Ala	Lys	Gln	Ser	Gly			
	1280					1285					1290						
acc	tgc	gac	aac	gtg	cgc	aag	cag	ctc	tcg	cag	gcc	acc	tac	tcg		3924	
Thr	Cys	Asp	Asn	Val	Arg	Lys	Gln	Leu	Ser	Gln	Ala	Thr	Tyr	Ser			

ES 2 567 305 T3

1295	gat atc tgc atg gcc ccg gcc	1300	gcc gac atg ttc gag	1305	gag ggc gtc	3969
Asp Ile Cys Met Ala Pro Ala	Ala Ala Asp Met Phe Glu	Ala Ala Asp Met Phe Glu	Glu Glu Gly Val			
1310	aag ctc cag gtc ctc aag aag	1315	gga acc atg ttc ccc	1320	tcg cgc gcc	4014
Lys Leu Gln Val Leu Lys Lys	Lys Gly Thr Met Phe Pro	Lys Gly Thr Met Phe Pro	Pro Ser Arg Ala			
1325	aac aag ctc tac gag ctc ttt	1330	tgc aag tac gac tcc	1335	ttc gac tcc	4059
Asn Lys Leu Tyr Glu Leu Phe	Phe Cys Lys Tyr Asp Ser	Phe Cys Lys Tyr Asp Ser	Phe Asp Ser			
1340	atg cct cct gcc gag ctc gag	1345	cgc atc gag aag cgt	1350	atc ttc aag	4104
Met Pro Pro Ala Glu Leu Glu	Glu Arg Ile Glu Lys Arg	Glu Arg Ile Glu Lys Arg	Ile Phe Lys			
1355	cgc gca ctc cag gag gtc tgg	1360	gag gag acc aag gac	1365	ttt tac att	4149
Arg Ala Leu Gln Glu Val Trp	Glu Glu Thr Lys Asp	Glu Glu Thr Lys Asp	Phe Tyr Ile			
1370	aac ggt ctc aag aac ccg gag	1375	aag atc cag cgc gcc	1380	gag cac gac	4194
Asn Gly Leu Lys Asn Pro Glu	Lys Ile Gln Arg Ala	Lys Ile Gln Arg Ala	Glu His Asp			
1385	ccc aag ctc aag atg tcg ctc	1390	tgc ttc cgc tgg tac	1395	ctt ggt ctt	4239
Pro Lys Leu Lys Met Ser Leu	Cys Phe Arg Trp Tyr	Cys Phe Arg Trp Tyr	Leu Gly Leu			
1400	gcc agc cgc tgg gcc aac atg	1405	ggc gcc ccg gac cgc	1410	gtc atg gac	4284
Ala Ser Arg Trp Ala Asn Met	Gly Ala Pro Asp Arg	Gly Ala Pro Asp Arg	Val Met Asp			
1415	tac cag gtc tgg tgt ggc ccg	1420	gcc att ggc gcc ttc	1425	aac gac ttc	4329
Tyr Gln Val Trp Cys Gly Pro	Ala Ile Gly Ala Phe	Ala Ile Gly Ala Phe	Asn Asp Phe			
1430	atc aag ggc acc tac ctc gac	1435	ccc gct gtc tcc aac	1440	gag tac ccc	4374
Ile Lys Gly Thr Tyr Leu Asp	Pro Ala Val Ser Asn	Pro Ala Val Ser Asn	Glu Tyr Pro			
1445	tgt gtc gtc cag atc aac ctg	1450	caa atc ctc cgt ggt	1455	gcc tgc tac	4419
Cys Val Val Gln Ile Asn Leu	Gln Ile Leu Arg Gly	Gln Ile Leu Arg Gly	Ala Cys Tyr			
1460	ctg cgc cgt ctc aac gcc ctg	1465	cgc aac gac ccg cgc	1470	att gac ctc	4464
Leu Arg Arg Leu Asn Ala Leu	Arg Asn Asp Pro Arg	Arg Asn Asp Pro Arg	Ile Asp Leu			
1475	gag acc gag gat gct gcc ttt	1480	gtc tac gag ccc acc	1485	aac gcg ctc	4509
Glu Thr Glu Asp Ala Ala Phe	Val Tyr Glu Pro Thr	Val Tyr Glu Pro Thr	Asn Ala Leu			
1490		1495		1500		

<210> 6  
 <211> 1503  
 <212> PRT  
 <213> *Schizochytrium* sp.

5

<400> 6

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

10



ES 2 567 305 T3

85					90					95					
Asn	Gly	Glu	Leu	Ser	Glu	Gly	Gly	Asp	Cys	Pro	Trp	Ala	Val	Leu	Val
			100					105					110		
Glu	Ser	Gly	Gln	Cys	Asp	Leu	Met	Leu	Ile	Ser	Tyr	Met	Gly	Ile	Asp
		115					120					125			
Phe	Gln	Asn	Gln	Gly	Asp	Arg	Val	Tyr	Arg	Leu	Leu	Asn	Thr	Thr	Leu
	130					135					140				
Thr	Phe	Tyr	Gly	Val	Ala	His	Glu	Gly	Glu	Thr	Leu	Glu	Tyr	Asp	Ile
145					150					155					160
Arg	Val	Thr	Gly	Phe	Ala	Lys	Arg	Leu	Asp	Gly	Gly	Ile	Ser	Met	Phe
				165					170					175	
Phe	Phe	Glu	Tyr	Asp	Cys	Tyr	Val	Asn	Gly	Arg	Leu	Leu	Ile	Glu	Met
			180					185					190		
Arg	Asp	Gly	Cys	Ala	Gly	Phe	Phe	Thr	Asn	Glu	Glu	Leu	Asp	Ala	Gly
		195					200					205			
Lys	Gly	Val	Val	Phe	Thr	Arg	Gly	Asp	Leu	Ala	Ala	Arg	Ala	Lys	Ile
	210					215					220				
Pro	Lys	Gln	Asp	Val	Ser	Pro	Tyr	Ala	Val	Ala	Pro	Cys	Leu	His	Lys
225					230					235					240
Thr	Lys	Leu	Asn	Glu	Lys	Glu	Met	Gln	Thr	Leu	Val	Asp	Lys	Asp	Trp
				245					250					255	
Ala	Ser	Val	Phe	Gly	Ser	Lys	Asn	Gly	Met	Pro	Glu	Ile	Asn	Tyr	Lys
			260					265					270		
Leu	Cys	Ala	Arg	Lys	Met	Leu	Met	Ile	Asp	Arg	Val	Thr	Ser	Ile	Asp
		275					280					285			
His	Lys	Gly	Gly	Val	Tyr	Gly	Leu	Gly	Gln	Leu	Val	Gly	Glu	Lys	Ile
	290					295					300				
Leu	Glu	Arg	Asp	His	Trp	Tyr	Phe	Pro	Cys	His	Phe	Val	Lys	Asp	Gln
305					310					315					320
Val	Met	Ala	Gly	Ser	Leu	Val	Ser	Asp	Gly	Cys	Ser	Gln	Met	Leu	Lys
				325					330					335	
Met	Tyr	Met	Ile	Trp	Leu	Gly	Leu	His	Leu	Thr	Thr	Gly	Pro	Phe	Asp
			340					345					350		
Phe	Arg	Pro	Val	Asn	Gly	His	Pro	Asn	Lys	Val	Arg	Cys	Arg	Gly	Gln
		355					360					365			
Ile	Ser	Pro	His	Lys	Gly	Lys	Leu	Val	Tyr	Val	Met	Glu	Ile	Lys	Glu
	370					375					380				
Met	Gly	Phe	Asp	Glu	Asp	Asn	Asp	Pro	Tyr	Ala	Ile	Ala	Asp	Val	Asn
385					390					395					400
Ile	Ile	Asp	Val	Asp	Phe	Glu	Lys	Gly	Gln	Asp	Phe	Ser	Leu	Asp	Arg
				405					410					415	
Ile	Ser	Asp	Tyr	Gly	Lys	Gly	Asp	Leu	Asn	Lys	Lys	Ile	Val	Val	Asp

# ES 2 567 305 T3

420					425					430					
Phe	Lys	Gly	Ile	Ala	Leu	Lys	Met	Gln	Lys	Arg	Ser	Thr	Asn	Lys	Asn
		435					440					445			
Pro	Ser	Lys	Val	Gln	Pro	Val	Phe	Ala	Asn	Gly	Ala	Ala	Thr	Val	Gly
	450					455					460				
Pro	Glu	Ala	Ser	Lys	Ala	Ser	Ser	Gly	Ala	Ser	Ala	Ser	Ala	Ser	Ala
465					470					475					480
Ala	Pro	Ala	Lys	Pro	Ala	Phe	Ser	Ala	Asp	Val	Leu	Ala	Pro	Lys	Pro
				485					490					495	
Val	Ala	Leu	Pro	Glu	His	Ile	Leu	Lys	Gly	Asp	Ala	Leu	Ala	Pro	Lys
			500					505					510		
Glu	Met	Ser	Trp	His	Pro	Met	Ala	Arg	Ile	Pro	Gly	Asn	Pro	Thr	Pro
		515					520					525			
Ser	Phe	Ala	Pro	Ser	Ala	Tyr	Lys	Pro	Arg	Asn	Ile	Ala	Phe	Thr	Pro
	530					535					540				
Phe	Pro	Gly	Asn	Pro	Asn	Asp	Asn	Asp	His	Thr	Pro	Gly	Lys	Met	Pro
545					550					555					560
Leu	Thr	Trp	Phe	Asn	Met	Ala	Glu	Phe	Met	Ala	Gly	Lys	Val	Ser	Met
				565					570					575	
Cys	Leu	Gly	Pro	Glu	Phe	Ala	Lys	Phe	Asp	Asp	Ser	Asn	Thr	Ser	Arg
			580					585					590		
Ser	Pro	Ala	Trp	Asp	Leu	Ala	Leu	Val	Thr	Arg	Ala	Val	Ser	Val	Ser
		595					600					605			
Asp	Leu	Lys	His	Val	Asn	Tyr	Arg	Asn	Ile	Asp	Leu	Asp	Pro	Ser	Lys
	610					615					620				
Gly	Thr	Met	Val	Gly	Glu	Phe	Asp	Cys	Pro	Ala	Asp	Ala	Trp	Phe	Tyr
625					630					635					640
Lys	Gly	Ala	Cys	Asn	Asp	Ala	His	Met	Pro	Tyr	Ser	Ile	Leu	Met	Glu
				645					650					655	
Ile	Ala	Leu	Gln	Thr	Ser	Gly	Val	Leu	Thr	Ser	Val	Leu	Lys	Ala	Pro
			660					665					670		
Leu	Thr	Met	Glu	Lys	Asp	Asp	Ile	Leu	Phe	Arg	Asn	Leu	Asp	Ala	Asn
		675					680					685			
Ala	Glu	Phe	Val	Arg	Ala	Asp	Leu	Asp	Tyr	Arg	Gly	Lys	Thr	Ile	Arg
	690					695					700				
Asn	Val	Thr	Lys	Cys	Thr	Gly	Tyr	Ser	Met	Leu	Gly	Glu	Met	Gly	Val
705					710					715					720
His	Arg	Phe	Thr	Phe	Glu	Leu	Tyr	Val	Asp	Asp	Val	Leu	Phe	Tyr	Lys
				725					730					735	
Gly	Ser	Thr	Ser	Phe	Gly	Trp	Phe	Val	Pro	Glu	Val	Phe	Ala	Ala	Gln
			740					745					750		
Ala	Gly	Leu	Asp	Asn	Gly	Arg	Lys	Ser	Glu	Pro	Trp	Phe	Ile	Glu	Asn

# ES 2 567 305 T3

755	760	765																												
Lys	Val	Pro	Ala	Ser	Gln	Val	Ser	Ser	Phe	Asp	Val	Arg	Pro	Asn	Gly	770	775	780												
Ser	Gly	Arg	Thr	Ala	Ile	Phe	Ala	Asn	Ala	Pro	Ser	Gly	Ala	Gln	Leu	785	790	795												
Asn	Arg	Arg	Thr	Asp	Gln	Gly	Gln	Tyr	Leu	Asp	Ala	Val	Asp	Ile	Val	805	810	815												
Ser	Gly	Ser	Gly	Lys	Lys	Ser	Leu	Gly	Tyr	Ala	His	Gly	Ser	Lys	Thr	820	825	830												
Val	Asn	Pro	Asn	Asp	Trp	Phe	Phe	Ser	Cys	His	Phe	Trp	Phe	Asp	Ser	835	840	845												
Val	Met	Pro	Gly	Ser	Leu	Gly	Val	Glu	Ser	Met	Phe	Gln	Leu	Val	Glu	850	855	860												
Ala	Ile	Ala	Ala	His	Glu	Asp	Leu	Ala	Gly	Lys	Ala	Arg	His	Cys	Gln	865	870	875												
Pro	His	Leu	Cys	Ala	Arg	Pro	Arg	Ala	Arg	Ser	Ser	Trp	Lys	Tyr	Arg	885	890	895												
Gly	Gln	Leu	Thr	Pro	Lys	Ser	Lys	Lys	Met	Asp	Ser	Glu	Val	His	Ile	900	905	910												
Val	Ser	Val	Asp	Ala	His	Asp	Gly	Val	Val	Asp	Leu	Val	Ala	Asp	Gly	915	920	925												
Phe	Leu	Trp	Ala	Asp	Ser	Leu	Arg	Val	Tyr	Ser	Val	Ser	Asn	Ile	Arg	930	935	940												
Val	Arg	Ile	Ala	Ser	Gly	Glu	Ala	Pro	Ala	Ala	Ala	Ser	Ser	Ala	Ala	945	950	955												
Ser	Val	Gly	Ser	Ser	Ala	Ser	Ser	Val	Glu	Arg	Thr	Arg	Ser	Ser	Pro	965	970	975												
Ala	Val	Ala	Ser	Gly	Pro	Ala	Gln	Thr	Ile	Asp	Leu	Lys	Gln	Leu	Lys	980	985	990												
Thr	Glu	Leu	Leu	Glu	Leu	Asp	Ala	Pro	Leu	Tyr	Leu	Ser	Gln	Asp	Pro	995	1000	1005												
Thr	Ser	Gly	Gln	Leu	Lys	Lys	His	Thr	Asp	Val	Ala	Ser	Gly	Gln		1010	1015	1020												
Ala	Thr	Ile	Val	Gln	Pro	Cys	Thr	Leu	Gly	Asp	Leu	Gly	Asp	Arg		1025	1030	1035												
Ser	Phe	Met	Glu	Thr	Tyr	Gly	Val	Val	Ala	Pro	Leu	Tyr	Thr	Gly		1040	1045	1050												
Ala	Met	Ala	Lys	Gly	Ile	Ala	Ser	Ala	Asp	Leu	Val	Ile	Ala	Ala		1055	1060	1065												
Gly	Lys	Arg	Lys	Ile	Leu	Gly	Ser	Phe	Gly	Ala	Gly	Gly	Leu	Pro		1070	1075	1080												
Met	His	His	Val	Arg	Ala	Ala	Leu	Glu	Lys	Ile	Gln	Ala	Ala	Leu																

ES 2 567 305 T3

1085						1090						1095			
Pro	Gln	Gly	Pro	Tyr	Ala	Val	Asn	Leu	Ile	His	Ser	Pro	Phe	Asp	
1100						1105					1110				
Ser	Asn	Leu	Glu	Lys	Gly	Asn	Val	Asp	Leu	Phe	Leu	Glu	Lys	Gly	
1115						1120					1125				
Val	Thr	Val	Val	Glu	Ala	Ser	Ala	Phe	Met	Thr	Leu	Thr	Pro	Gln	
1130						1135					1140				
Val	Val	Arg	Tyr	Arg	Ala	Ala	Gly	Leu	Ser	Arg	Asn	Ala	Asp	Gly	
1145						1150					1155				
Ser	Val	Asn	Ile	Arg	Asn	Arg	Ile	Ile	Gly	Lys	Val	Ser	Arg	Thr	
1160						1165					1170				
Glu	Leu	Ala	Glu	Met	Phe	Ile	Arg	Pro	Ala	Pro	Glu	His	Leu	Leu	
1175						1180					1185				
Glu	Lys	Leu	Ile	Ala	Ser	Gly	Glu	Ile	Thr	Gln	Glu	Gln	Ala	Glu	
1190						1195					1200				
Leu	Ala	Arg	Arg	Val	Pro	Val	Ala	Asp	Asp	Ile	Ala	Val	Glu	Ala	
1205						1210					1215				
Asp	Ser	Gly	Gly	His	Thr	Asp	Asn	Arg	Pro	Ile	His	Val	Ile	Leu	
1220						1225					1230				
Pro	Leu	Ile	Ile	Asn	Leu	Arg	Asn	Arg	Leu	His	Arg	Glu	Cys	Gly	
1235						1240					1245				
Tyr	Pro	Ala	His	Leu	Arg	Val	Arg	Val	Gly	Ala	Gly	Gly	Gly	Val	
1250						1255					1260				
Gly	Cys	Pro	Gln	Ala	Ala	Ala	Ala	Ala	Leu	Thr	Met	Gly	Ala	Ala	
1265						1270					1275				
Phe	Ile	Val	Thr	Gly	Thr	Val	Asn	Gln	Val	Ala	Lys	Gln	Ser	Gly	
1280						1285					1290				
Thr	Cys	Asp	Asn	Val	Arg	Lys	Gln	Leu	Ser	Gln	Ala	Thr	Tyr	Ser	
1295						1300					1305				
Asp	Ile	Cys	Met	Ala	Pro	Ala	Ala	Asp	Met	Phe	Glu	Glu	Gly	Val	
1310						1315					1320				
Lys	Leu	Gln	Val	Leu	Lys	Lys	Gly	Thr	Met	Phe	Pro	Ser	Arg	Ala	
1325						1330					1335				
Asn	Lys	Leu	Tyr	Glu	Leu	Phe	Cys	Lys	Tyr	Asp	Ser	Phe	Asp	Ser	
1340						1345					1350				
Met	Pro	Pro	Ala	Glu	Leu	Glu	Arg	Ile	Glu	Lys	Arg	Ile	Phe	Lys	
1355						1360					1365				
Arg	Ala	Leu	Gln	Glu	Val	Trp	Glu	Glu	Thr	Lys	Asp	Phe	Tyr	Ile	
1370						1375					1380				
Asn	Gly	Leu	Lys	Asn	Pro	Glu	Lys	Ile	Gln	Arg	Ala	Glu	His	Asp	
1385						1390					1395				
Pro	Lys	Leu	Lys	Met	Ser	Leu	Cys	Phe	Arg	Trp	Tyr	Leu	Gly	Leu	

ES 2 567 305 T3

1400 1405 1410  
 Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp  
 1415 1420 1425  
 Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe  
 1430 1435 1440  
 Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro  
 1445 1450 1455  
 Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr  
 1460 1465 1470  
 Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu  
 1475 1480 1485  
 Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu  
 1490 1495 1500

<210> 7  
 <211> 600  
 <212> ADN  
 <213> *Schizochytrium sp.*

5

<220>  
 <221> CDS  
 <222> (1)..(600)  
 <223>

10

<400> 7

atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc 48  
 Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg  
 1 5 10 15  
 att gcc atc atc ggc atg tcg gcc atc ctc ccc tgc ggc acg acc gtg 96  
 Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val  
 20 25 30  
 cgc gag tcg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tog gat 144  
 Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp  
 35 40 45  
 ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag 192  
 Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys  
 50 55 60  
 acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag 240  
 Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu  
 65 70 75 80  
 tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag 288  
 Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu  
 85 90 95  
 gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc 336  
 Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala  
 100 105 110  
 ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc 384  
 Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile  
 115 120 125  
 ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc 432  
 Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe  
 130 135 140  
 tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg 480  
 Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met  
 145 150 155 160  
 ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc 528  
 Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala  
 165 170 175  
 aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac 576  
 Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn  
 180 185 190  
 gtc acc gcc ggt cgc tgc acc aac 600  
 Val Thr Ala Gly Arg Cys Thr Asn  
 195 200

15

<210> 8

ES 2 567 305 T3

<211> 200  
 <212> PRT  
 <213> *Schizochytrium* sp.

5 <400> 8

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

10 <210> 9  
 <211> 1278  
 <212> ADN  
 <213> *Schizochytrium* sp.

15 <220>  
 <221> CDS  
 <222> (1)..(1278)  
 <223>

20 <400> 9

ES 2 567 305 T3

gat gtc acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc	48
Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe	
1 5 10 15	
cgc gcc aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc	96
Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser	
20 25 30	
ggc cag ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac	144
Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn	
35 40 45	
tgg ccc cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc	192
Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser	
50 55 60	
aag gtc gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc	240
Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu	
65 70 75 80	
tac ccg cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag	288
Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys	
85 90 95	
atc tcc ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc	336
Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu	
100 105 110	
ggt gcc ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc	384
Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala	
115 120 125	
gcc ggc cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc	432
Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys	
130 135 140	
gtc gac cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc	480
Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile	
145 150 155 160	
atg ggc ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc	528
Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala	
165 170 175	
gtc att ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc	576
Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val	
180 185 190	
tgg ctc ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc	624
Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser	
195 200 205	
gtc gaa ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc	672
Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe	
210 215 220	
cgc gtc gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg	720
Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met	
225 230 235 240	
gag aac gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc	768
Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe	
245 250 255	
cgc acc ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag	816
Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu	
260 265 270	
acc tac ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc	864
Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser	
275 280 285	
agc gtc aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg	912
Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala	
290 295 300	
cgc atc ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc	960
Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val	
305 310 315 320	
tcc gag acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac	1008
Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn	

ES 2 567 305 T3

```

                                325                                330                                335
ccg gcc tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc      1056
Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val
                                340                                345                                350
cag ctc gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac      1104
Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp
                                355                                360                                365
gcc ccc gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc      1152
Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr
                                370                                375                                380
ctc cgc ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc      1200
Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val
                                385                                390                                395                                400
cgc gac gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc      1248
Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly
                                405                                410                                415
gcc gca ccg ctc atc aag gcc ccg gag ccc
Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro
                                420                                425

```

5  
 <210> 10  
 <211> 426  
 <212> PRT  
 <213> *Schizochytrium* sp.

<400> 10

```

Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe
1                               5                               10                               15

Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser
20                              25                              30

Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn
35                               40                               45

Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser
50                               55                               60

Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu
65                               70                               75                               80

Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys
85                               90                               95

Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu
100                              105                              110

Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala
115                              120                              125

Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys
130                              135                              140

Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile
145                              150                              155                              160

Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala
165                              170                              175

Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val
180                              185                              190

Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser

```

10



# ES 2 567 305 T3

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

5 <210> 11  
 <211> 5  
 <212> PRT  
 <213> *Schizochytrium sp.*

10 <220>  
 <221> MISC\_FEATURE  
 <222> (4)..(4)  
 <223> X = cualquier aminoácido

<400> 11  
 Gly His Ser Xaa Gly  
 1 5

20 <210> 12  
 <211> 258  
 <212> ADN  
 <213> *Schizochytrium sp.*

25 <220>  
 <221> CDS  
 <222> (1)..(258)  
 <223>

<400> 12

ES 2 567 305 T3

```

gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc atg gag      48
Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu
1                               5                               10                               15
gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc gag gct gac      96
Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ala Asp
                               20                               25                               30
atg gag ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag     144
Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu
                               35                               40                               45
atc ctc tcc gag gtc cag gcc atg ctc aat gtc gag gcc aag gat gtc     192
Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val
                               50                               55                               60
gat gcc ctc agc cgc act cgc act gtt ggt gag gtt gtc aac gcc atg     240
Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asn Ala Met
65                               70                               75                               80
aag gcc gag atc gct ggc
Lys Ala Glu Ile Ala Gly
                               85

```

5  
 <210> 13  
 <211> 86  
 <212> PRT  
 <213> *Schizochytrium sp.*

<400> 13

```

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

```

10  
 <210> 14  
 <211> 5  
 <212> PRT  
 <213> *Schizochytrium sp.*

<400> 14

```

Leu Gly Ile Asp Ser
1                               5

```

20  
 <210> 15  
 <211> 21  
 <212> PRT  
 <213> *Schizochytrium sp.*

<400> 15

```

Ala Pro Ala Pro Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser
1                               5                               10                               15
Ala Pro Ala Pro Ala
                               20

```

30  
 <210> 16  
 <211> 3006  
 <212> ADN

ES 2 567 305 T3

<213> *Schizochytrium sp.*

<400> 16

gccccgccc	cggtcaaggc	tgctggcct	gcccggccc	ttgcctggc	cctcgcccc	60
gctgtctcga	acgagcttct	tgagaaggcc	gagactgtog	tcatggaggt	cctcgccgc	120
aagaccggct	acgagaccga	catgatcgag	gctgacatgg	agctcgagac	cgagctcggc	180
attgactcca	tcaagcgtgt	cgagatcctc	tccgaggtcc	aggccatgct	caatgtcgag	240
gccaaaggatg	tcgatgccct	cagccgcact	cgactgttg	gtgaggttgt	caacgccatg	300
aaggccgaga	tcgctggcag	ctctgcccc	gcccctgctg	ccgctgctcc	ggctccggcc	360
aaggctgccc	ctgcccgcgc	tgccgctgct	gtctcgaacg	agcttctcga	gaaggccgag	420
accgtcgtca	tgaggtcct	cgccgccaag	actggctacg	agactgacat	gatcgagtcc	480
gacatggagc	tcgagactga	gctcggcatt	gactccatca	agcgtgtcga	gatcctctcc	540
gaggttcagg	ccatgctcaa	cgtcgaggcc	aaggacgtcg	acgctctcag	ccgcactcgc	600
actgtgggtg	aggtcgtcaa	cgccatgaag	gctgagatcg	ctggaggctc	tgccccggcg	660
cctgcccggc	ctgccccagg	tccggctgct	gcccggcctg	cgctgcccgc	cgccggcccct	720
gctgtctcga	acgagcttct	tgagaaggcc	gagaccgtcg	tcatggaggt	cctcgccgc	780
aagctggctg	acgagactga	catgatcgag	tccgacatgg	agctcgagac	cgagctcggc	840
attgactcca	tcaagcgtgt	cgagattctc	tccgaggtcc	aggccatgct	caacgtcgag	900
gccaaaggacg	tcgacgctct	cagccgcacc	cgactgttg	gagaggtcgt	cgatgccatg	960
aaggccgaga	tcgctgggtg	ctctgcccc	gcccctgccc	ccgctgctcc	tgctccggct	1020
gctgcccggc	ctgcccggc	cgcccctg	cctgctgtct	cgagcgagct	tctcgagaag	1080
gcccagactg	tcgtcatgga	ggtcctcgcc	gccaagactg	gctacgagac	tgccatgatc	1140
gagtcgaca	tgaggtcga	gaccgagctc	ggcattgact	ccatcaagcg	tgctcgagatt	1200
ctctccgagg	tccaggccat	gctcaacgtc	gaggccaagg	acgtcgagcg	tctcagccgc	1260
accgcactg	ttggcgaggt	cgtcgatgcc	atgaaggccg	agatcgctgg	tggtctgccc	1320
ccggcgctg	ccgcccgtgc	tcctgctcgg	gctgtgccc	cccctgccc	tgcccggccc	1380
gcccctgccc	cccctgccc	tgctgtctcg	agcgagcttc	tcgagaaggc	cgagactgtc	1440
gtcatggagg	tcctcgccc	caagactggc	tacgagactg	acatgattga	gtccgacatg	1500
gagctcgaga	ccgagctcgg	catgactccc	atcaagcgtg	tcgagattct	ctccgaggtt	1560
caggccatgc	tcaacgtcga	ggccaaggac	gtcgagctc	tcagccgcac	tcgactgtt	1620
ggtgaggtcg	tcgatgccat	gaaggctgag	atcgtgtgca	gctccgctc	ggcgcctgcc	1680
gcccgtcctc	ctgctccggc	tgctgcccgt	cctgcccggc	ctgcccggc	ccctgctgtc	1740
tcgaacgagc	ttctcgagaa	agccgagact	gtcgtcatgg	aggtcctcgc	cgccaagact	1800
ggctacgaga	ctgacatgat	cgagtccgac	atggagctcg	agactgagct	cggcattgac	1860
tccatcaagc	gtgtcgagat	cctctccgag	gttcaggcca	tgctcaacgt	cgaggccaag	1920
gacgtcgatg	ccctcagccg	caccgcact	ggtggcgagg	ttgtcgatgc	catgaaggcc	1980
gagatcgctg	gtggctctgc	cccggcgct	gcccggctg	cccctgctcc	ggctgcccgc	2040
gcccctgctg	tctcgaacga	gcttctcgag	aaggccgaga	ctgtcgtcat	ggaggtcctc	2100
gcccgaaga	ctggctacga	gaccgacatg	atcgagtccg	acatggagct	cgagaccgag	2160
ctogggattg	actccatcaa	gcgtgtcgag	attctctccg	aggttcaggc	catgctcaac	2220
gtcgaggcca	aggacgtcga	tgctctcagc	cgactcgca	ctggtggcga	ggctcgtcgt	2280
gcccgaagc	ctgagatcgc	cgccagctcc	gcccggcgc	ctgcccggc	tgctcctgct	2340
ccggctgctg	ccgctcctgc	gcccgtgccc	gctgcccctg	ctgtctcgag	cgagcttctc	2400
gagaaggccg	agaccgtcgt	catggaggtc	ctgcccggca	agactggcta	cgagactgac	2460
atgattgagt	ccgacatgga	gctcgagact	gagctcggca	ttgactccat	caagcgtgtc	2520
gagatcctct	ccgaggttca	ggccatgctc	aacgtcgagg	ccaaggacgt	cgatgccctc	2580
agccgcaccc	gcaactgttg	cgaggttgtc	gatgccatga	aggccgagat	cgctgggtggc	2640
tctgccccgg	cgccctgccc	cgctgcccct	gctccggctg	ccgcccggc	tgctgtctcg	2700
aacgagcttc	ttgagaaggc	cgagaccgtc	gtcatggagg	tcctcgccgc	caagactggc	2760
tacgagaccg	acatgatcga	gtccgacatg	gagctcgaga	ccgagctcgg	cattgactcc	2820
atcaagcgtg	tcgagattct	ctccgaggtt	caggccatgc	tcaacgtcga	ggccaaggac	2880
gtcgacgctc	tcagccgcac	tcgactgtt	ggcgaggtcg	tcgatgccat	gaaggctgag	2940
atcgctggtg	gctctgcccc	ggcgctgccc	gcccgtgctc	ctgcccggc	tgccggccgc	3000
cctgccc						3006

5

<210> 17  
 <211> 2133  
 <212> ADN  
 <213> *Schizochytrium sp.*

10

<220>  
 <221> CDS  
 <222> (1)..(2133)  
 <223>

15

<400> 17

ES 2 567 305 T3

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



ES 2 567 305 T3

```

          595                600                605
tcc gac gcc gct cgc ggc gag ttt gcc acg gac act gac gcc cat gac      1872
Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
    610                615                620
ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg cgc      1920
Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg
    625                630                635                640
cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc atc gtc      1968
Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val
    645                650                655
cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att acc ctc cgc      2016
Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg
    660                665                670
tcc aac cag tcg ggc ggt cac tcc cag cac aag cac gcc ctt cag ttc      2064
Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
    675                680                685
cac aac gag cag ggc gat ctc ttc att gat gtc cag gct tcg gtc atc      2112
His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile
    690                695                700
gcc acg gac agc ctt gcc ttc
Ala Thr Asp Ser Leu Ala Phe      2133
    705                710

```

<210> 18

<211> 711

5 <212> PRT

<213> *Schizochytrium sp.*

<400> 18

```

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

```

10







ES 2 567 305 T3

65	ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc	70	gag aac gag atc gac aac gag cac gaa ctc ctc ctc	75	gag aac gag cac gaa ctc ctc ctc	80	ctc ctc	288
Gly Thr Leu Asp	85	Glu Asn Glu Ile Asp	90	Asn Glu His Glu Leu Leu Leu	95	Leu Leu		
aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tcg aca	100	Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr	105	Glu Thr Ser Val Lys Asp Ser Thr	110	Leu Leu		336
Asn Leu Ala Lys	115	Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr	120	Glu Thr Ser Phe Pro Met Asp Asn Leu	125	Leu Leu		384
cgc tgc gcc atc gtc agc gcc tgc ctc tcg ttc ccc atg gac aac ctc	130	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	135	Leu Ser Phe Pro Met Asp Asn Leu	140	Leu Leu		432
Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	145	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	150	Leu Ser Phe Pro Met Asp Asn Leu	155	Leu Leu		480
cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc	160	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	165	Leu Ser Phe Pro Met Asp Asn Leu	170	Leu Leu		528
Gln Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln	175	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	180	Leu Ser Phe Pro Met Asp Asn Leu	185	Leu Leu		576
ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag	190	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	195	Leu Ser Phe Pro Met Asp Asn Leu	200	Leu Leu		624
Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln	205	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	210	Leu Ser Phe Pro Met Asp Asn Leu	215	Leu Leu		672
tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc	220	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	225	Leu Ser Phe Pro Met Asp Asn Leu	230	Leu Leu		720
Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala	235	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	240	Leu Ser Phe Pro Met Asp Asn Leu	245	Leu Leu		768
tcc ttc gtc gcc gaa ctc aac ctc gcc ctt cac tac tcc gtc	250	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	255	Leu Ser Phe Pro Met Asp Asn Leu	260	Leu Leu		816
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val	265	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	270	Leu Ser Phe Pro Met Asp Asn Leu	275	Leu Leu		864
gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat	280	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	285	Leu Ser Phe Pro Met Asp Asn Leu	290	Leu Leu		912
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp	295	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	300	Leu Ser Phe Pro Met Asp Asn Leu	305	Leu Leu		960
cat ctc gtc tcc gcc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc	310	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	315	Leu Ser Phe Pro Met Asp Asn Leu	320	Leu Leu		1008
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys	325	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	330	Leu Ser Phe Pro Met Asp Asn Leu	335	Leu Leu		1056
ctg ccg gag ccc ttt ttc atc ott tcg gcc ttt tcc acc ttc cag gcc	340	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	345	Leu Ser Phe Pro Met Asp Asn Leu	350	Leu Leu		1104
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala	355	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	360	Leu Ser Phe Pro Met Asp Asn Leu	365	Leu Leu		1152
atg ccc gtc gcc acg gcc cag aac gtg tcc atg ccg ctg cac aag gac	370	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	375	Leu Ser Phe Pro Met Asp Asn Leu	380	Leu Leu		1200
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp	385	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	390	Leu Ser Phe Pro Met Asp Asn Leu	395	Leu Leu		1248
agc cag gcc ctc acc ccg ggt gag gcc gcc tcc atc atg gtc ctc aag	400	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	405	Leu Ser Phe Pro Met Asp Asn Leu	410	Leu Leu		1296
Ser Gln Gly Leu Thr Pro Gly Glu Gly Ser Ile Met Val Leu Lys	415	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	420	Leu Ser Phe Pro Met Asp Asn Leu	425	Leu Leu		1344
cgf ctc gat gat gcc atc cgc gac gcc gac cac att tac gcc acc ctt	425	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	430	Leu Ser Phe Pro Met Asp Asn Leu	435	Leu Leu		1350
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu	440	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	445	Leu Ser Phe Pro Met Asp Asn Leu	450	Leu Leu		
ctc gcc gcc aat gtc agc aac tcc gcc aca ggt ctg ccc ctc aag ccc	450	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	455	Leu Ser Phe Pro Met Asp Asn Leu	460	Leu Leu		
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro	465	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	470	Leu Ser Phe Pro Met Asp Asn Leu	475	Leu Leu		
ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att	480	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	485	Leu Ser Phe Pro Met Asp Asn Leu	490	Leu Leu		
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile	495	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	500	Leu Ser Phe Pro Met Asp Asn Leu	505	Leu Leu		
305	310	315	320	325	330	335	340	345
aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc gcc	345	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	350	Leu Ser Phe Pro Met Asp Asn Leu	355	Leu Leu		
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly	360	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	365	Leu Ser Phe Pro Met Asp Asn Leu	370	Leu Leu		
325	330	335	340	345	350	355	360	365
acc cts gyc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc	370	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	375	Leu Ser Phe Pro Met Asp Asn Leu	380	Leu Leu		
Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser	385	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	390	Leu Ser Phe Pro Met Asp Asn Leu	395	Leu Leu		
370	375	380	385	390	395	400	405	410
atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc	410	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	415	Leu Ser Phe Pro Met Asp Asn Leu	420	Leu Leu		
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr	425	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	430	Leu Ser Phe Pro Met Asp Asn Leu	435	Leu Leu		
385	390	395	400	405	410	415	420	425
aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag	425	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	430	Leu Ser Phe Pro Met Asp Asn Leu	435	Leu Leu		
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu	440	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	445	Leu Ser Phe Pro Met Asp Asn Leu	450	Leu Leu		
405	410	415	420	425	430	435	440	445
acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt	445	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	450	Leu Ser Phe Pro Met Asp Asn Leu	455	Leu Leu		
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly	460	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	465	Leu Ser Phe Pro Met Asp Asn Leu	470	Leu Leu		
420	425	430	435	440	445	450	455	460
ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc	465	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	470	Leu Ser Phe Pro Met Asp Asn Leu	475	Leu Leu		
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala	480	Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	485	Leu Ser Phe Pro Met Asp Asn Leu	490	Leu Leu		
435	440	445	450	455	460	465	470	475
ggc tgc Ala Cys 450								

ES 2 567 305 T3

<211> 450  
 <212> PRT  
 <213> *Schizochytrium* sp.

5 <220>  
 <221> misc\_feature  
 <222> (370)..(370)  
 <223> El 'Xaa' en la posición 370 representa Leu.

10 <220>  
 <221> misc\_feature  
 <222> (371)..(371)  
 <223> El 'Xaa' en la posición 371 representa Ala. o Val.

15 <400> 20

```

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

ES 2 567 305 T3

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

5 <210> 21  
 <211> 1323  
 <212> ADN  
 <213> *Schizochytrium sp.*

10 <220>  
 <221> CDS  
 <222> (1)..(1323)  
 <223>

<400> 21

15 tcg gcc cgc tgc ggc ggt gaa agc aac atg cgc atc gcc atc act ggt 48  
 Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly  
 1 5 10 15  
 atg gac gcc acc ttt ggc gct ctc aag gga ctc gac gcc ttc gag cgc 96

ES 2 567 305 T3

Met	Asp	Ala	Thr	Phe	Gly	Ala	Leu	Lys	Gly	Leu	Asp	Ala	Phe	Glu	Arg		
			20					25					30				
gcc	att	tac	acc	ggc	gct	cac	ggg	gcc	atc	cca	ctc	cca	gaa	aag	cgc	144	
Ala	Ile	Tyr	Thr	Gly	Ala	His	Gly	Ala	Ile	Pro	Leu	Pro	Glu	Lys	Arg		
		35					40					45					
tgg	cgc	ttt	ctc	ggc	aag	gac	aag	gac	ttt	ctt	gac	ctc	tgc	ggc	gtc	192	
Trp	Arg	Phe	Leu	Gly	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Leu	Cys	Gly	Val		
	50					55					60						
aag	gcc	acc	ccg	cac	ggc	tgc	tac	att	gaa	gat	ggt	gag	gtc	gac	ttc	240	
Lys	Ala	Thr	Pro	His	Gly	Cys	Tyr	Ile	Glu	Asp	Val	Glu	Val	Asp	Phe		
	65				70					75					80		
cag	cgc	ctc	cgc	acg	ccc	atg	acc	cct	gaa	gac	atg	ctc	ctc	cct	cag	288	
Gln	Arg	Leu	Arg	Thr	Pro	Met	Thr	Pro	Glu	Asp	Met	Leu	Leu	Pro	Gln		
				85					90					95			
cag	ctt	ctg	ggc	gtc	acc	acc	att	gac	cgc	gcc	atc	ctc	gac	tcg	gga	336	
Gln	Leu	Leu	Ala	Val	Thr	Thr	Ile	Asp	Arg	Ala	Ile	Leu	Asp	Ser	Gly		
			100					105					110				
atg	aaa	aag	ggg	ggc	aat	gtc	gcc	gtc	ttt	gtc	ggc	ctc	ggc	acc	gac	384	
Met	Lys	Lys	Gly	Gly	Asn	Val	Ala	Val	Phe	Val	Gly	Leu	Gly	Thr	Asp		
		115				120						125					
ctc	gag	ctc	tac	cgt	cac	cgt	gct	cgc	gtc	gct	ctc	aag	gag	cgc	gtc	432	
Leu	Glu	Leu	Tyr	Arg	His	Arg	Ala	Arg	Val	Ala	Leu	Lys	Glu	Arg	Val		
	130				135						140						
cgc	cct	gaa	gcc	tcc	aag	aag	ctc	aat	gac	atg	atg	cag	tac	att	aac	480	
Arg	Pro	Glu	Ala	Ser	Lys	Lys	Leu	Asn	Asp	Met	Met	Gln	Tyr	Ile	Asn		
	145				150					155					160		
gac	tgc	ggc	aca	tcc	aca	tcg	tac	acc	tcg	tac	att	ggc	aac	ctc	gtc	528	
Asp	Cys	Gly	Thr	Ser	Thr	Ser	Tyr	Thr	Ser	Tyr	Ile	Gly	Asn	Leu	Val		
				165					170					175			
gcc	acg	cgc	gtc	tcg	tcg	cag	tgg	ggc	ttc	acg	ggc	ccc	tcc	ttt	acg	576	
Ala	Thr	Arg	Val	Ser	Ser	Gln	Trp	Gly	Phe	Thr	Gly	Pro	Ser	Phe	Thr		
			180					185					190				
atc	acc	gag	ggc	aac	aac	tcc	gtc	tac	cgc	tgc	gcc	gag	ctc	ggc	aag	624	
Ile	Thr	Glu	Gly	Asn	Asn	Ser	Val	Tyr	Arg	Cys	Ala	Glu	Leu	Gly	Lys		
		195				200						205					
tac	ctc	ctc	gag	acc	ggc	gag	gtc	gat	ggc	gtc	gtc	ggt	gcg	ggg	gtc	672	
Tyr	Leu	Leu	Glu	Thr	Gly	Glu	Val	Asp	Gly	Val	Val	Val	Ala	Gly	Val		
	210				215						220						
gat	ctc	tgc	ggc	agt	gcc	gaa	aac	ctt	tac	gtc	aag	tct	cgc	cgc	ttc	720	
Asp	Leu	Cys	Gly	Ser	Ala	Glu	Asn	Leu	Tyr	Val	Lys	Ser	Arg	Arg	Phe		
	225				230					235					240		
aag	gtg	tcc	acc	tcc	gat	acc	ccg	cgc	gcc	agc	ttt	gac	gcc	gcc	gcc	768	
Lys	Val	Ser	Thr	Ser	Asp	Thr	Pro	Arg	Ala	Ser	Phe	Asp	Ala	Ala	Ala		
				245					250					255			
gat	ggc	tac	ttt	gtc	ggc	gag	ggc	tgc	ggg	gcc	ttt	gtg	ctc	aag	cgt	816	
Asp	Gly	Tyr	Phe	Val	Gly	Glu	Gly	Cys	Gly	Ala	Phe	Val	Leu	Lys	Arg		
			260					265					270				
gag	act	agc	tgc	acc	aag	gac	gac	cgt	atc	tac	gct	tgc	atg	gat	gcc	864	
Glu	Thr	Ser	Cys	Thr	Lys	Asp	Asp	Arg	Ile	Tyr	Ala	Cys	Met	Asp	Ala		
	275					280						285					
atc	gtc	cct	ggc	aac	gtc	cct	agc	gcc	tgc	ttg	cgc	gag	gcc	ctc	gac	912	
Ile	Val	Pro	Gly	Asn	Val	Pro	Ser	Ala	Cys	Leu	Arg	Glu	Ala	Leu	Asp		
	290				295							300					
cag	gcg	cgc	gtc	aag	ccg	ggc	gat	atc	gag	atg	ctc	gag	ctc	agc	gcc	960	
Gln	Ala	Arg	Val	Lys	Pro	Gly	Asp	Ile	Glu	Met	Leu	Glu	Leu	Ser	Ala		
	305				310					315					320		
gac	tcc	gcc	cgc	cac	ctc	aag	gac	ccg	tcc	gtc	ctg	ccc	aag	gag	ctc	1008	
Asp	Ser	Ala	Arg	His	Leu	Lys	Asp	Pro	Ser	Val	Leu	Pro	Lys	Glu	Leu		
				325					330					335			
act	gcc	gag	gag	gaa	atc	ggc	ggc	ctt	cag	acg	atc	ctt	cgt	gac	gat	1056	
Thr	Ala	Glu	Glu	Glu	Ile	Gly	Gly	Leu	Gln	Thr	Ile	Leu	Arg	Asp	Asp		
			340					345					350				
gac	aag	ctc	ccg	cgc	aac	gtc	gca	acg	ggc	agt	gtc	aag	gcc	acc	gtc	1104	

ES 2 567 305 T3

Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val  
 355 360 365  
 ggt gac acc ggt tat gcc tct ggt gct gcc agc ctc atc aag gct gcg 1152  
 Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala  
 370 375 380  
 ctt tgc atc tac aac cgc tac ctg ccc agc aac ggc gac gac tgg gat 1200  
 Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp  
 385 390 395  
 gaa ccc gcc cct gag gcg ccc tgg gac agc acc ctc ttt gcg lgc cag 1248  
 Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln  
 405 410 415  
 acc tcg cgc gct tgg ctc aag aac cct ggc gag cgt cgc tat gcg gcc 1296  
 Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala  
 420 425 430  
 gtc tcg ggc gtc tcc gag acg cgc tcg 1323  
 Val Ser Gly Val Ser Glu Thr Arg Ser  
 435 440

<210> 22

<211> 441

5

<212> PRT

<213> *Schizochytrium* sp.

<400> 22

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

10

ES 2 567 305 T3

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

<210> 23  
 <211> 1500  
 <212> ADN  
 <213> *Schizochytrium* sp.

5

<220>  
 <221> CDS  
 <222> (1)..(1500)  
 <223>

10

<400> 23

tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac tac gag cgc gag 48  
 Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu  
 1 5 10 15  
 aac cgc atc tcg ctc gac gag gag gcg ccc aag ctc att gtg ctt cgc 96  
 Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg  
 20 25 30  
 gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac aag atc cgc gag 144

15

ES 2 567 305 T3

Ala	Asp	Ser	His	Glu	Glu	Ile	Leu	Gly	Arg	Leu	Asp	Lys	Ile	Arg	Glu		
		35					40					45					
cgc	ttc	ttg	cag	ccc	acg	ggc	gcc	gcc	cgc	cgc	gag	tcc	gag	ctc	aag	192	
Arg	Phe	Leu	Gln	Pro	Thr	Gly	Ala	Ala	Pro	Arg	Glu	Ser	Glu	Leu	Lys		
	50					55					60						
gcg	cag	gcc	cgc	cgc	atc	ttc	ctc	gag	ctc	ctc	ggc	gag	acc	ctt	gcc	240	
Ala	Gln	Ala	Arg	Arg	Ile	Phe	Leu	Glu	Leu	Leu	Gly	Glu	Thr	Leu	Ala		
	65				70					75					80		
cag	gat	gcc	gct	tct	tca	ggc	tcg	caa	aag	ccc	ctc	gct	ctc	agc	ctc	288	
Gln	Asp	Ala	Ala	Ser	Ser	Gly	Ser	Gln	Lys	Pro	Leu	Ala	Leu	Ser	Leu		
			85						90					95			
gtc	tcc	acg	ccc	tcc	aag	ctc	cag	cgc	gag	gtc	gag	ctc	gcg	gcc	aag	336	
Val	Ser	Thr	Pro	Ser	Lys	Leu	Gln	Arg	Glu	Val	Glu	Leu	Ala	Ala	Lys		
			100					105					110				
ggt	atc	ccg	cgc	tgc	ctc	aag	atg	cgc	cgc	gat	tgg	agc	tcc	cct	gct	384	
Gly	Ile	Pro	Arg	Cys	Leu	Lys	Met	Arg	Arg	Asp	Trp	Ser	Ser	Pro	Ala		
		115				120						125					
ggc	agc	cgc	tac	gcg	cct	gag	ccg	ctc	gcc	agc	gac	cgc	gtc	gcc	ttc	432	
Gly	Ser	Arg	Tyr	Ala	Pro	Glu	Pro	Leu	Ala	Ser	Asp	Arg	Val	Ala	Phe		
	130				135						140						
atg	tac	ggc	gaa	ggt	cgc	agc	cct	tac	tac	ggc	atc	acc	caa	gac	att	480	
Met	Tyr	Gly	Glu	Gly	Arg	Ser	Pro	Tyr	Tyr	Gly	Ile	Thr	Gln	Asp	Ile		
	145				150					155					160		
cac	cgc	att	tgg	ccc	gaa	ctc	cac	gag	gtc	atc	aac	gaa	aag	acg	aac	528	
His	Arg	Ile	Trp	Pro	Glu	Leu	His	Glu	Val	Ile	Asn	Glu	Lys	Thr	Asn		
			165						170					175			
cgf	ctc	tgg	gcc	gaa	ggc	gac	cgc	tgg	gtc	atg	ccg	cgc	gcc	agc	ttc	576	
Arg	Leu	Trp	Ala	Glu	Gly	Asp	Arg	Trp	Val	Met	Pro	Arg	Ala	Ser	Phe		
			180					185					190				
aag	tcg	gag	ctc	gag	agc	cag	cag	caa	gag	ttt	gat	cgc	aac	atg	att	624	
Lys	Ser	Glu	Leu	Glu	Ser	Gln	Gln	Gln	Glu	Phe	Asp	Arg	Asn	Met	Ile		
		195				200						205					
gaa	atg	ttc	cgt	ctt	gga	atc	ctc	acc	tca	att	gcc	ttc	acc	aat	ctg	672	
Glu	Met	Phe	Arg	Leu	Gly	Ile	Leu	Thr	Ser	Ile	Ala	Phe	Thr	Asn	Leu		
	210				215						220						
gcg	cgc	gac	ggt	ctc	aac	atc	acg	ccc	aag	gcc	ttt	ggc	ctc	agt		720	
Ala	Arg	Asp	Val	Leu	Asn	Ile	Thr	Pro	Lys	Ala	Ala	Phe	Gly	Leu	Ser		
	225				230					235				240			
ctt	ggc	gag	att	tcc	atg	att	ttt	gcc	ttt	tcc	aag	aag	aac	ggt	ctc	768	
Leu	Gly	Glu	Ile	Ser	Met	Ile	Phe	Ala	Phe	Ser	Lys	Lys	Asn	Gly	Leu		
			245						250					255			
atc	tcc	gac	cag	ctc	acc	aag	gat	ctt	cgc	gag	tcc	gac	gtg	tgg	aac	816	
Ile	Ser	Asp	Gln	Leu	Thr	Lys	Asp	Leu	Arg	Glu	Ser	Asp	Val	Trp	Asn		
			260					265					270				
aag	gct	ctg	gcc	ggt	gaa	ttt	aat	gcg	ctg	cgc	gag	gcc	tgg	ggc	att	864	
Lys	Ala	Leu	Ala	Val	Glu	Phe	Asn	Ala	Leu	Arg	Glu	Ala	Trp	Gly	Ile		
	275						280					285					
cca	cag	agt	gtc	ccc	aag	gac	gag	ttc	tgg	caa	ggc	tac	att	gtg	cgc	912	
Pro	Gln	Ser	Val	Pro	Lys	Asp	Glu	Phe	Trp	Gln	Gly	Tyr	Ile	Val	Arg		
	290				295						300						
ggc	acc	aag	cag	gat	atc	gag	gcg	gcc	atc	gcc	ccg	gac	agc	aag	tac	960	
Gly	Thr	Lys	Gln	Asp	Ile	Glu	Ala	Ala	Ile	Ala	Pro	Asp	Ser	Lys	Tyr		
	305				310					315					320		
gtg	cgc	ctc	acc	atc	atc	aat	gat	gcc	aac	acc	gcc	ctc	att	agc	ggc	1008	
Val	Arg	Leu	Thr	Ile	Ile	Asn	Asp	Ala	Asn	Thr	Ala	Leu	Ile	Ser	Gly		
			325						330					335			
aag	ccc	gac	gcc	tgc	aag	gct	gcg	atc	gcg	cgt	ctc	ggt	ggc	aac	att	1056	
Lys	Pro	Asp	Ala	Cys	Lys	Ala	Ala	Ile	Ala	Arg	Leu	Gly	Gly	Asn	Ile		
			340					345					350				
cct	gcg	ctt	ccc	gtg	acc	cag	ggc	atg	tgc	ggc	cac	tgc	ccc	gag	gtg	1104	
Pro	Ala	Leu	Pro	Val	Thr	Gln	Gly	Met	Cys	Gly	His	Cys	Pro	Glu	Val		
		355					360					365					
gga	cct	tat	acc	aag	gat	atc	gcc	aag	atc	cat	gcc	aac	ctt	gag	ttc	1152	

ES 2 567 305 T3

Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe  
 370 375 380  
 ccc gtt gtc gac ggc ctt gac ctc tgg acc aca atc aac cag aag cgc 1200  
 Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg  
 385 390 395 400  
 ctc gtg cca cgc gcc acg ggc gcc aag gac gaa tgg gcc cct tct tcc 1248  
 Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser  
 405 410 415  
 ttt ggc gag tac gcc ggc cag ctc tac gag aag cag gct aac ttc ccc 1296  
 Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro  
 420 425 430  
 caa atc gtc gag acc att tac aag caa aac tac gac gtc ttt gtc gag 1344  
 Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu  
 435 440 445  
 gtt ggg ccc aac aac cac cgt agc acc gca gtg cgc acc acg ctt ggt 1392  
 Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly  
 450 455 460  
 ccc cag cgc aac cac ctt gct ggc gcc atc gac aag cag aac gag gat 1440  
 Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp  
 465 470 475 480  
 gct tgg acg acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt 1488  
 Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu  
 485 490 495  
 gtt cct ggc gtc 1500  
 Val Pro Gly Val  
 500

<210> 24

<211> 500

5

<212> PRT

<213> *Schizochytrium sp.*

<400> 24

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

10



ES 2 567 305 T3

His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn  
 165 170 175  
 Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe  
 180 185 190  
 Lys Ser Glu Leu Glu Ser Gln Gln Glu Phe Asp Arg Asn Met Ile  
 195 200 205  
 Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu  
 210 215 220  
 Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser  
 225 230 235 240  
 Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu  
 245 250 255  
 Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn  
 260 265 270  
 Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile  
 275 280 285  
 Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg  
 290 295 300  
 Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr  
 305 310 315 320  
 Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly  
 325 330 335  
 Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile  
 340 345 350  
 Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val  
 355 360 365  
 Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe  
 370 375 380  
 Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg  
 385 390 395 400  
 Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser  
 405 410 415  
 Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro  
 420 425 430  
 Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu  
 435 440 445  
 Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly  
 450 455 460  
 Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp  
 465 470 475 480  
 Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu  
 485 490 495  
 Val Pro Gly Val  
 500

5 <210> 25  
 <211> 1530  
 <212> ADN  
 <213> *Schizochytrium sp.*

10 <220>  
 <221> CDS  
 <222> (1)..(1530)

ES 2 567 305 T3

<223>

<400> 25

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

ES 2 567 305 T3

Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala  
 260 265 270  
 gcg ctg gcc acc ttc aac atg ggt gcc tcc ttt att gtc acc ggc acc 864  
 Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr  
 275 280 285  
 gtg aac cag gtc gcc aag cag tcg ggc acg tgc gac aat gtg cgc aag 912  
 Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys  
 290 295 300  
 cag ctc gcg aag gcc act tac tcg gac gta tgc atg gcc ccg gct gcc 960  
 Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala  
 305 310 315 320  
 gac atg ttc gag gaa ggc gtc aag ctt cag gtc ctc aag aag gga acc 1008  
 Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr  
 325 330 335  
 atg ttt ccc tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac 1056  
 Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr  
 340 345 350  
 gac tcg ttc gag tcc atg ccc ccc gca gag ctt gcg cgc gtc gag aag 1104  
 Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys  
 355 360 365  
 cgc atc ttc agc cgc gcg ctc gaa gag gtc tgg gac gag acc aaa aac 1152  
 Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn  
 370 375 380  
 ttt tac att aac cgt ctt cac aac ccg gag aag atc cag cgc gcc gag 1200  
 Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu  
 385 390 395 400  
 cgc gac ccc aag ctc aag atg tcg ctg tgc ttt cgc tgg tac ctg agc 1248  
 Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser  
 405 410 415  
 ctg gcg agc cgc tgg gcc aac act gga gct tcc gat cgc gtc atg gac 1296  
 Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp  
 420 425 430  
 tac cag gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc 1344  
 Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile  
 435 440 445  
 aag gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc 1392  
 Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val  
 450 455 460  
 gtt cag att aac aag cag atc ctt cgt gga gcg tgc ttc ttg cgc cgt 1440  
 Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg  
 465 470 475 480  
 ctc gaa att ctg cgc aac gca cgc ctt tcc gat ggc gct gcc gct ctt 1488  
 Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu  
 485 490 495  
 gtg gcc agc atc gat gac aca tac gtc ccg gcc gag aag ctg 1530  
 Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu  
 500 505 510

<210> 26  
 <211> 510  
 <212> PRT  
 <213> *Schizochytrium* sp.

5

<400> 26

Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser  
 1 5 10 15  
 Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro  
 20 25 30  
 Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr  
 35 40 45

10

# ES 2 567 305 T3

Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala  
 50 55 60

Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser  
 65 70 75

Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu  
 85 90 95

Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile  
 100 105 110

His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe  
 115 120 125

Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu  
 130 135 140

Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala  
 145 150 155 160

Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg  
 165 170 175

Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu  
 180 185 190

Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu  
 195 200 205

Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser  
 210 215 220

Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile  
 225 230 235 240

Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn  
 245 250 255

Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala  
 260 265 270

Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr  
 275 280 285

Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys  
 290 295 300

Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val Cys Met Ala Pro Ala Ala  
 305 310 315 320

Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr  
 325 330 335

Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr  
 340 345 350

Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys  
 355 360 365

Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn  
 370 375 380

ES 2 567 305 T3

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

<210> 27  
 <211> 4512  
 <212> ADN  
 <213> *Schizochytrium sp.*

<220>  
 <221> CDS  
 <222> (1)..(4512)  
 <223>

<400> 27

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

ES 2 567 305 T3

Thr	Phe	Tyr	Gly	Val	Ala	His	Glu	Gly	Glu	Thr	Leu	Glu	Tyr	Asp	Ile		
145					150					155					160		
cgc	gtc	acc	ggc	ttc	gcc	aag	cgt	ctc	gac	ggc	ggc	atc	tcc	atg	ttc	528	
Arg	Val	Thr	Gly	Phe	Ala	Lys	Arg	Leu	Asp	Gly	Gly	Ile	Ser	Met	Phe		
			165						170					175			
ttc	ttc	gag	tac	gac	tgc	tac	gtc	aac	ggc	cgc	ctc	ctc	atc	gag	atg	576	
Phe	Phe	Glu	Tyr	Asp	Cys	Tyr	Val	Asn	Gly	Arg	Leu	Leu	Ile	Glu	Met		
			180					185					190				
cgc	gat	ggc	tgc	gcc	ggc	ttc	ttc	acc	aac	gag	gag	ctc	gac	gcc	ggc	624	
Arg	Asp	Gly	Cys	Ala	Gly	Phe	Phe	Thr	Asn	Glu	Glu	Leu	Asp	Ala	Gly		
		195					200					205					
aag	ggc	gtc	gtc	ttc	acc	cgc	ggc	gac	ctc	gcc	gcc	cgc	gcc	aag	atc	672	
Lys	Gly	Val	Val	Phe	Thr	Arg	Gly	Asp	Leu	Ala	Ala	Arg	Ala	Lys	Ile		
	210					215					220						
cca	aag	cag	gac	gtc	tcc	ccc	tac	gcc	gtc	gcc	ccc	tgc	ctc	cac	aag	720	
Pro	Lys	Gln	Asp	Val	Ser	Pro	Tyr	Ala	Val	Ala	Pro	Cys	Leu	His	Lys		
					230					235				240			
acc	aag	ctc	aac	gaa	aag	gag	atg	cag	acc	ctc	gtc	gac	aag	gac	tgg	768	
Thr	Lys	Leu	Asn	Glu	Lys	Glu	Met	Gln	Thr	Leu	Val	Asp	Lys	Asp	Trp		
			245						250				255				
gca	tcc	gtc	ttt	ggc	tcc	aag	aac	ggc	atg	ccg	gaa	atc	aac	tac	aaa	816	
Ala	Ser	Val	Phe	Gly	Ser	Lys	Asn	Gly	Met	Pro	Glu	Ile	Asn	Tyr	Lys		
			260					265					270				
ctc	tgc	cgc	aat	aag	atg	ctc	atg	att	gac	cgc	gtc	acc	agc	att	gac	864	
Leu	Cys	Ala	Arg	Lys	Met	Leu	Met	Ile	Asp	Arg	Val	Thr	Ser	Ile	Asp		
		275				280						285					
cac	aag	ggc	ggt	gtc	tac	ggc	ctc	ggt	cag	ctc	gtc	ggt	gaa	aag	atc	912	
His	Lys	Gly	Gly	Val	Tyr	Gly	Leu	Gly	Gln	Leu	Val	Gly	Glu	Lys	Ile		
	290					295				300							
ctc	gag	cgc	gac	cac	tgg	tac	ttt	ccc	tgc	cac	ttt	gtc	aag	gat	cag	960	
Leu	Glu	Arg	Asp	His	Trp	Tyr	Phe	Pro	Cys	His	Phe	Val	Lys	Asp	Gln		
	305				310					315				320			
gtc	atg	gcc	gga	tcc	ctc	gtc	tcc	gac	ggc	tgc	agc	cag	atg	ctc	aag	1008	
Val	Met	Ala	Gly	Ser	Leu	Val	Ser	Asp	Gly	Cys	Ser	Gln	Met	Leu	Lys		
			325						330				335				
atg	tac	atg	atc	tgg	ctc	ggc	ctc	cac	ctc	acc	acc	gga	ccc	ttt	gac	1056	
Met	Tyr	Met	Ile	Trp	Leu	Gly	Leu	His	Leu	Thr	Thr	Gly	Pro	Phe	Asp		
			340					345					350				
ttc	cgc	ccg	gtc	aac	ggc	cac	ccc	aac	aag	gtc	cgc	tgc	cgc	ggc	caa	1104	
Phe	Arg	Pro	Val	Asn	Gly	His	Pro	Asn	Lys	Val	Arg	Cys	Arg	Gly	Gln		
		355					360					365					
atc	tcc	ccg	cac	aag	ggc	aag	ctc	gtc	tac	gtc	atg	gag	atc	aag	gag	1152	
Ile	Ser	Pro	His	Lys	Gly	Lys	Leu	Val	Tyr	Val	Met	Glu	Ile	Lys	Glu		
	370					375				380							
atg	ggc	ttc	gac	gag	gac	aac	gac	ccg	tac	gcc	att	gcc	gac	gtc	aac	1200	
Met	Gly	Phe	Asp	Glu	Asp	Asn	Asp	Pro	Tyr	Ala	Ile	Ala	Asp	Val	Asn		
				390						395				400			
atc	att	gat	gtc	gac	ttc	gaa	aag	ggc	cag	gac	ttt	agc	ctc	gac	cgc	1248	
Ile	Ile	Asp	Val	Asp	Phe	Glu	Lys	Gly	Gln	Asp	Phe	Ser	Leu	Asp	Arg		
			405						410				415				
atc	agc	gac	tac	ggc	aag	ggc	gac	ctc	aac	aag	aag	atc	gtc	gtc	gac	1296	
Ile	Ser	Asp	Tyr	Gly	Lys	Gly	Asp	Leu	Asn	Lys	Lys	Ile	Val	Val	Asp		
			420					425					430				
ttt	aag	ggc	atc	gct	ctc	aag	atg	cag	aag	cgc	tcc	acc	aac	aag	aac	1344	
Phe	Lys	Gly	Ile	Ala	Leu	Lys	Met	Gln	Lys	Arg	Ser	Thr	Asn	Lys	Asn		
		435					440					445					
ccc	tcc	aag	ggt	cag	ccc	gtc	ttt	gcc	aac	ggc	gcc	gcc	act	gtc	ggc	1392	
Pro	Ser	Lys	Val	Gln	Pro	Val	Phe	Ala	Asn	Gly	Ala	Ala	Thr	Val	Gly		
		450				455					460						
ccc	gag	gcc	tcc	aag	gct	tcc	ggc	gcc	agc	gcc	agc	gcc	agc	gcc	agc	1440	
Pro	Glu	Ala	Ser	Lys	Ala	Ser	Ser	Gly	Ala	Ser	Ala	Ser	Ala	Ser	Ala		
					470					475				480			
gcc	ccg	gcc	aag	cct	gcc	ttc	agc	gcc	gat	ggt	ctt	gcg	ccc	aag	ccc	1488	

ES 2 567 305 T3

Ala	Pro	Ala	Lys	Pro	Ala	Phe	Ser	Ala	Asp	Val	Leu	Ala	Pro	Lys	Pro		
				485					490					495			
ggt	gcc	ctt	ccc	gag	cac	atc	ctc	aag	ggc	gac	gcc	ctc	gcc	ccc	aag	1536	
Val	Ala	Leu	Pro	Glu	His	Ile	Leu	Lys	Gly	Asp	Ala	Leu	Ala	Pro	Lys		
			500					505					510				
gag	atg	tcc	tgg	cac	ccc	atg	gcc	cgc	atc	ccg	ggc	aac	ccg	acg	ccc	1584	
Glu	Met	Ser	Trp	His	Pro	Met	Ala	Arg	Ile	Pro	Gly	Asn	Pro	Thr	Pro		
		515				520						525					
tct	ttt	gcg	ccc	tcg	gcc	tac	aag	ccg	cgc	aac	atc	gcc	ttt	acg	ccc	1632	
Ser	Phe	Ala	Pro	Ser	Ala	Tyr	Lys	Pro	Arg	Asn	Ile	Ala	Phe	Thr	Pro		
	530					535					540						
ttc	ccc	ggc	aac	ccc	aac	gat	aac	gac	cac	acc	ccg	ggc	aag	atg	ccg	1680	
Phe	Pro	Gly	Asn	Pro	Asn	Asp	Asn	Asp	His	Thr	Pro	Gly	Lys	Met	Pro		
	545			550					555						560		
ctc	acc	tgg	ttc	aac	atg	gcc	gag	ttc	atg	gcc	ggc	aag	gtc	agc	atg	1728	
Leu	Thr	Trp	Phe	Asn	Met	Ala	Glu	Phe	Met	Ala	Gly	Lys	Val	Ser	Met		
			565						570						575		
tgc	ctc	ggc	ccc	gag	ttc	gcc	aag	ttc	gac	gac	tcg	aac	acc	agc	cgc	1776	
Cys	Leu	Gly	Pro	Glu	Phe	Ala	Lys	Phe	Asp	Asp	Ser	Asn	Thr	Ser	Arg		
			580					585					590				
agc	ccc	gct	tgg	gac	ctc	gct	ctc	gtc	acc	cgc	gcc	gtg	tct	gtg	tct	1824	
Ser	Pro	Ala	Trp	Asp	Leu	Ala	Leu	Val	Thr	Arg	Ala	Val	Ser	Val	Ser		
		595				600						605					
gac	ctc	aag	cac	gtc	aac	tac	cgc	aac	atc	gac	ctc	gac	ccc	tcc	aag	1872	
Asp	Leu	Lys	His	Val	Asn	Tyr	Arg	Asn	Ile	Asp	Leu	Asp	Pro	Ser	Lys		
	610					615						620					
ggt	acc	atg	gtc	ggc	gag	ttc	gac	tgc	ccc	gcg	gac	gcc	tgg	ttc	tac	1920	
Gly	Thr	Met	Val	Gly	Glu	Phe	Asp	Cys	Pro	Ala	Asp	Ala	Trp	Phe	Tyr		
	625					630					635				640		
aag	ggc	gcc	tgc	aac	gat	gcc	cac	atg	ccg	tac	tcg	atc	ctc	atg	gag	1968	
Lys	Gly	Ala	Cys	Asn	Asp	Ala	His	Met	Pro	Tyr	Ser	Ile	Leu	Met	Glu		
			645						650						655		
atc	gcc	ctc	cag	acc	tcg	ggt	gtg	ctc	acc	tcg	gtg	ctc	aag	gcg	ccc	2016	
Ile	Ala	Leu	Gln	Thr	Ser	Gly	Val	Leu	Thr	Ser	Val	Leu	Lys	Ala	Pro		
			660					665					670				
ctg	acc	atg	gag	aag	gac	gac	atc	ctc	ttc	cgc	aac	ctc	gac	gcc	aac	2064	
Leu	Thr	Met	Glu	Lys	Asp	Asp	Ile	Leu	Phe	Arg	Asn	Leu	Asp	Ala	Asn		
		675					680						685				
gcc	gag	ttc	gtg	cgc	gcc	gac	ctc	gac	tac	cgc	ggc	aag	act	atc	cgc	2112	
Ala	Glu	Phe	Val	Arg	Ala	Asp	Leu	Asp	Tyr	Arg	Gly	Lys	Thr	Ile	Arg		
	690					695					700						
aac	gtc	acc	aag	tgc	act	ggc	tac	agc	atg	ctc	ggc	gag	atg	ggc	gtc	2160	
Asn	Val	Thr	Lys	Cys	Thr	Gly	Tyr	Ser	Met	Leu	Gly	Glu	Met	Gly	Val		
	705				710					715					720		
cac	cgc	ttc	acc	ttt	gag	ctc	tac	gtc	gat	gat	gtg	ctc	ttt	tac	aag	2208	
His	Arg	Phe	Thr	Phe	Glu	Leu	Tyr	Val	Asp	Asp	Val	Leu	Phe	Tyr	Lys		
				725					730						735		
ggc	tcg	acc	tcg	ttc	ggc	tgg	ttc	gtg	ccc	gag	gtc	ttt	gcc	gcc	cag	2256	
Gly	Ser	Thr	Ser	Phe	Gly	Trp	Phe	Val	Pro	Glu	Val	Phe	Ala	Ala	Gln		
			740					745					750				
gcc	ggc	ctc	gac	aac	ggc	cgc	aag	tcg	gag	ccc	tgg	ttc	att	gag	aac	2304	
Ala	Gly	Leu	Asp	Asn	Gly	Arg	Lys	Ser	Glu	Pro	Trp	Phe	Ile	Glu	Asn		
		755					760					765					
aag	ggt	ccg	gcc	tcg	cag	gtc	tcc	tcc	ttt	gac	gtg	cgc	ccc	aac	ggc	2352	
Lys	Val	Pro	Ala	Ser	Gln	Val	Ser	Ser	Phe	Asp	Val	Arg	Pro	Asn	Gly		
	770					775						780					
agc	ggc	cgc	acc	gcc	atc	ttc	gcc	aac	gcc	ccc	agc	ggc	gcc	cag	ctc	2400	
Ser	Gly	Arg	Thr	Ala	Ile	Phe	Ala	Asn	Ala	Pro	Ser	Gly	Ala	Gln	Leu		
	785				790					795					800		
aac	cgc	cgc	acg	gac	cag	ggc	cag	tac	ctc	gac	gcc	gtc	gac	att	gtc	2448	
Asn	Arg	Arg	Thr	Asp	Gln	Gly	Gln	Tyr	Leu	Asp	Ala	Val	Asp	Ile	Val		
				805					810						815		
tcc	ggc	agc	ggc	aag	aag	agc	ctc	ggc	tac	gcc	cac	ggt	tcc	aag	acg	2496	

ES 2 567 305 T3

Ser	Gly	Ser	Gly	Lys	Lys	Ser	Leu	Gly	Tyr	Ala	His	Gly	Ser	Lys	Thr		
			820					825					830				
gtc	aac	ccg	aac	gac	tgg	ttc	ttc	tcg	tgc	cac	ttt	tgg	ttt	gac	tcg		2544
Val	Asn	Pro	Asn	Asp	Trp	Phe	Phe	Ser	Cys	His	Phe	Trp	Phe	Asp	Ser		
			835				840					845					
gtc	atg	ccc	gga	agt	ctc	ggt	gtc	gag	tcc	atg	ttc	cag	ctc	gtc	gag		2592
Val	Met	Pro	Gly	Ser	Leu	Gly	Val	Glu	Ser	Met	Phe	Gln	Leu	Val	Glu		
	850					855					860						
gcc	atc	gcc	gcc	cac	gag	gat	ctc	gct	ggc	aaa	gca	cgg	cat	tgc	caa		2640
Ala	Ile	Ala	Ala	His	Glu	Asp	Leu	Ala	Gly	Lys	Ala	Arg	His	Cys	Gln		
	865				870					875					880		
ccc	cac	ctt	tgt	gca	cgc	ccc	cgg	gca	aga	tca	agc	tgg	aag	tac	cgc		2688
Pro	His	Leu	Cys	Ala	Arg	Pro	Arg	Ala	Arg	Ser	Ser	Trp	Lys	Tyr	Arg		
				885					890					895			
ggc	cag	ctc	acg	ccc	aag	agc	aag	aag	atg	gac	tcg	gag	gtc	cac	atc		2736
Gly	Gln	Leu	Pro	Lys	Ser	Lys	Lys	Met	Asp	Ser	Glu	Val	His	Ile			
			900					905					910				
gtg	tcc	gtg	gac	gcc	cac	gac	ggc	ggt	gtc	gac	ctc	gtc	gcc	gac	ggc		2784
Val	Ser	Val	Asp	Ala	His	Asp	Gly	Val	Val	Asp	Leu	Val	Ala	Asp	Gly		
		915					920						925				
ttc	ctc	tgg	gcc	gac	agc	ctc	cgc	gtc	tac	tcg	gtg	agc	aac	att	cgc		2832
Phe	Leu	Trp	Ala	Asp	Ser	Leu	Arg	Val	Tyr	Ser	Val	Ser	Asn	Ile	Arg		
		930				935					940						
gtg	cgc	atc	gcc	tcc	ggt	gag	gcc	cct	gcc	gcc	tcc	tcc	gcc	gcc			2880
Val	Arg	Ile	Ala	Ser	Gly	Glu	Ala	Pro	Ala	Ala	Ala	Ser	Ser	Ala	Ala		
	945				950				955						960		
tct	gtg	ggc	tcc	tcg	gct	tcg	tcc	gtc	gag	cgc	acg	cgc	tcg	agc	ccc		2928
Ser	Val	Gly	Ser	Ser	Ala	Ser	Ser	Val	Glu	Arg	Thr	Arg	Ser	Ser	Pro		
				965				970						975			
gct	gtc	gcc	tcc	ggc	ccg	gcc	cag	acc	atc	gac	ctc	aag	cag	ctc	aag		2976
Ala	Val	Ala	Ser	Gly	Pro	Ala	Gln	Thr	Ile	Asp	Leu	Lys	Gln	Leu	Lys		
			980					985					990				
acc	gag	ctc	ctc	gag	ctc	gat	gcc	ccg	ctc	tac	ctc	tcg	cag	gac	ccg		3024
Thr	Glu	Leu	Leu	Glu	Leu	Asp	Ala	Pro	Leu	Tyr	Leu	Ser	Gln	Asp	Pro		
		995					1000					1005					
acc	agc	ggc	cag	ctc	aag	aag	cac	acc	gac	gtg	gcc	tcc	ggc	cag			3069
Thr	Ser	Gly	Gln	Leu	Lys	Lys	His	Thr	Asp	Val	Ala	Ser	Gly	Gln			
		1010				1015					1020						
gcc	acc	atc	gtg	cag	ccc	tgc	acg	ctc	ggc	gac	ctc	ggt	gac	cgc			3114
Ala	Thr	Ile	Val	Gln	Pro	Cys	Thr	Leu	Gly	Asp	Leu	Gly	Asp	Arg			
		1025				1030					1035						
tcc	ttc	atg	gag	acc	tac	ggc	gtc	gtc	gcc	ccg	ctg	tac	acg	ggc			3159
Ser	Phe	Met	Glu	Thr	Tyr	Gly	Val	Val	Ala	Pro	Leu	Tyr	Thr	Gly			
	1040					1045					1050						
gcc	atg	gcc	aag	ggc	att	gcc	tcg	gcg	gac	ctc	gtc	atc	gcc	gcc			3204
Ala	Met	Ala	Lys	Gly	Ile	Ala	Ser	Ala	Asp	Leu	Val	Ile	Ala	Ala			
	1055					1060					1065						
ggc	aag	cgc	aag	atc	ctc	ggc	tcc	ttt	ggc	gcc	ggc	ggc	ctc	ccc			3249
Gly	Lys	Arg	Lys	Ile	Leu	Gly	Ser	Phe	Gly	Ala	Gly	Gly	Leu	Pro			
	1070					1075					1080						
atg	cac	cac	gtg	cgc	gcc	gcc	ctc	gag	aag	atc	cag	gcc	gcc	ctg			3294
Met	His	His	Val	Arg	Ala	Ala	Leu	Glu	Lys	Ile	Gln	Ala	Ala	Leu			
	1085					1090					1095						
cct	cag	ggc	ccc	tac	gcc	gtc	aac	ctc	atc	cac	tcg	cct	ttt	gac			3339
Pro	Gln	Gly	Pro	Tyr	Ala	Val	Asn	Leu	Ile	His	Ser	Pro	Phe	Asp			
	1100					1105					1110						
agc	aac	ctc	gag	aag	ggc	aac	gtc	gat	ctc	ttc	ctc	gag	aag	ggc			3384
Ser	Asn	Leu	Glu	Lys	Gly	Asn	Val	Asp	Leu	Phe	Leu	Glu	Lys	Gly			
	1115					1120					1125						
gtc	act	gtg	gtg	gag	gcc	tcg	gca	ttc	atg	acc	ctc	acc	ccg	cag			3429
Val	Thr	Val	Val	Glu	Ala	Ser	Ala	Phe	Met	Thr	Leu	Thr	Pro	Gln			
	1130					1135					1140						
gtc	gtg	cgc	tac	cgc	gcc	gcc	ggc	ctc	tcg	cgc	aac	gcc	gac	ggt			3474



ES 2 567 305 T3

Val	Val	Arg	Tyr	Arg	Ala	Ala	Gly	Leu	Ser	Arg	Asn	Ala	Asp	Gly		
	1145					1150					1155					
tcg	gtc	aac	atc	cgc	aac	cgc	atc	atc	ggc	aag	gtc	tcg	cgc	acc	3519	
Ser	Val	Asn	Ile	Arg	Asn	Arg	Ile	Ile	Gly	Lys	Val	Ser	Arg	Thr		
	1160					1165					1170					
gag	ctc	gcc	gag	atg	ttc	atc	cgc	cgc	gcc	ccg	gag	cac	ctc	ctc	3564	
Glu	Leu	Ala	Glu	Met	Phe	Ile	Arg	Pro	Ala	Pro	Glu	His	Leu	Leu		
	1175					1180					1185					
gag	aag	ctc	atc	gcc	tcg	ggc	gag	atc	acc	cag	gag	cag	gcc	gag	3609	
Glu	Lys	Leu	Ile	Ala	Ser	Gly	Glu	Ile	Thr	Gln	Glu	Gln	Ala	Glu		
	1190					1195					1200					
ctc	gcg	cgc	cgc	gtt	ccc	gtc	gcc	gac	gat	atc	gct	gtc	gag	gct	3654	
Leu	Ala	Arg	Arg	Val	Pro	Val	Ala	Asp	Asp	Ile	Ala	Val	Glu	Ala		
	1205					1210					1215					
gac	tcg	ggc	ggc	cac	acc	gac	aac	cgc	ccc	atc	cac	gtc	atc	ctc	3699	
Asp	Ser	Gly	Gly	His	Thr	Asp	Asn	Arg	Pro	Ile	His	Val	Ile	Leu		
	1220					1225					1230					
ccg	ctc	atc	atc	aac	ctc	cgc	aac	cgc	ctg	cac	cgc	gag	tgc	ggc	3744	
Pro	Leu	Ile	Ile	Asn	Leu	Arg	Asn	Arg	Leu	His	Arg	Glu	Cys	Gly		
	1235					1240					1245					
tac	ccc	gcg	cac	ctc	cgc	gtc	cgc	gtt	ggc	gcc	ggc	ggg	ggc	gtc	3789	
Tyr	Pro	Ala	His	Leu	Arg	Val	Arg	Val	Gly	Ala	Gly	Gly	Gly	Val		
	1250					1255					1260					
ggc	tgc	ccg	cag	gcc	gcc	gcc	gcc	gcg	ctc	acc	atg	ggc	gcc	gcc	3834	
Gly	Cys	Pro	Gln	Ala	Ala	Ala	Ala	Ala	Leu	Thr	Met	Gly	Ala	Ala		
	1265					1270					1275					
ttc	atc	gtc	acc	ggc	act	gtc	aac	cag	gtc	gcc	aag	cag	tcc	ggc	3879	
Phe	Ile	Val	Thr	Gly	Thr	Val	Asn	Gln	Val	Ala	Lys	Gln	Ser	Gly		
	1280					1285					1290					
acc	tgc	gac	aac	gtg	cgc	aag	cag	ctc	tcg	cag	gcc	acc	tac	tcg	3924	
Thr	Cys	Asp	Asn	Val	Arg	Lys	Gln	Leu	Ser	Gln	Ala	Thr	Tyr	Ser		
	1295					1300					1305					
gat	atc	tgc	atg	gcc	ccg	gcc	gcc	gac	atg	ttc	gag	gag	ggc	gtc	3969	
Asp	Ile	Cys	Met	Ala	Pro	Ala	Ala	Asp	Met	Phe	Glu	Glu	Gly	Val		
	1310					1315					1320					
aag	ctc	cag	gtc	ctc	aag	aag	gga	acc	atg	ttc	ccc	tcg	cgc	gcc	4014	
Lys	Leu	Gln	Val	Leu	Lys	Lys	Gly	Thr	Met	Phe	Pro	Ser	Arg	Ala		
	1325					1330					1335					
aac	aag	ctc	tac	gag	ctc	ttt	tgc	aag	tac	gac	tcc	ttc	gac	tcc	4059	
Asn	Lys	Leu	Tyr	Glu	Leu	Phe	Cys	Lys	Tyr	Asp	Ser	Phe	Asp	Ser		
	1340					1345					1350					
atg	cct	cct	gcc	gag	ctc	gag	cgc	atc	gag	aag	cgt	atc	ttc	aag	4104	
Met	Pro	Pro	Ala	Glu	Leu	Glu	Arg	Ile	Glu	Lys	Arg	Ile	Phe	Lys		
	1355					1360					1365					
cgc	gca	ctc	cag	gag	gtc	tgg	gag	gag	acc	aag	gac	ttt	tac	att	4149	
Arg	Ala	Leu	Gln	Glu	Val	Trp	Glu	Glu	Thr	Lys	Asp	Phe	Tyr	Ile		
	1370					1375					1380					
aac	ggt	ctc	aag	aac	ccg	gag	aag	atc	cag	cgc	gcc	gag	cac	gac	4194	
Asn	Gly	Leu	Lys	Asn	Pro	Glu	Lys	Ile	Gln	Arg	Ala	Glu	His	Asp		
	1385					1390					1395					
ccc	aag	ctc	aag	atg	tcg	ctc	tgc	ttc	cgc	tgg	tac	ctt	ggt	ctt	4239	
Pro	Lys	Leu	Lys	Met	Ser	Leu	Cys	Phe	Arg	Trp	Tyr	Leu	Gly	Leu		
	1400					1405					1410					
gcc	agc	cgc	tgg	gcc	aac	atg	ggc	gcc	ccg	gac	cgc	gtc	atg	gac	4284	
Ala	Ser	Arg	Trp	Ala	Asn	Met	Gly	Ala	Pro	Asp	Arg	Val	Met	Asp		
	1415					1420					1425					
tac	cag	gtc	tgg	tgt	ggc	ccg	gcc	att	ggc	gcc	ttc	aac	gac	ttc	4329	
Tyr	Gln	Val	Trp	Cys	Gly	Pro	Ala	Ile	Gly	Ala	Phe	Asn	Asp	Phe		
	1430					1435					1440					
atc	aag	ggc	acc	tac	ctc	gac	ccc	gct	gtc	tcc	aac	gag	tac	ccc	4374	
Ile	Lys	Gly	Thr	Tyr	Leu	Asp	Pro	Ala	Val	Ser	Asn	Glu	Tyr	Pro		
	1445					1450					1455					
tgt	gtc	gtc	cag	atc	aac	ctg	caa	atc	ctc	cgt	ggt	gcc	tgc	tac	4419	
Cys	Val	Val	Gln	Ile	Asn	Leu	Gln	Ile	Leu	Arg	Gly	Ala	Cys	Tyr		
	1460					1465					1470					
ctg	cgc	cgt	ctc	aac	gcc	ctg	cgc	aac	gac	ccg	cgc	att	gac	ctc	4464	
Leu	Arg	Arg	Leu	Asn	Ala	Leu	Arg	Asn	Asp	Pro	Arg	Ile	Asp	Leu		
	1475					1480					1485					
gag	acc	gag	gat	gct	gcc	ttt	gtc	tac	gag	ccc	acc	aac	gcg	ctc	4509	
Glu	Thr	Glu	Asp	Ala	Ala	Phe	Val	Tyr	Glu	Pro	Thr	Asn	Ala	Leu		
	1490					1495					1500					
taa															4512	

<210> 28  
<211> 1503

ES 2 567 305 T3

<212> PRT

<213> *Schizochytrium sp.*

<400> 28

5

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



ES 2 567 305 T3

Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser  
595 600 605

Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys  
610 615 620

Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr  
625 630 635 640

Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu  
645 650 655

Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro  
660 665 670

Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn  
675 680 685

Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg  
690 695 700

Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val  
705 710 715 720

His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys  
725 730 735

Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln  
740 745 750

Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn  
755 760 765

Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly  
770 775 780

Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu  
785 790 795 800

Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val  
805 810 815

Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr  
820 825 830

Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser  
835 840 845

Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu  
850 855 860

Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln  
865 870 875 880

Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg  
885 890 895

Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile  
900 905 910

Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly  
915 920 925

# ES 2 567 305 T3

Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg  
 930 935 940  
 Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala  
 945 950 955 960  
 Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro  
 965 970 975  
 Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys  
 980 985 990  
 Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro  
 995 1000 1005  
 Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln  
 1010 1015 1020  
 Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg  
 1025 1030 1035  
 Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly  
 1040 1045 1050  
 Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala  
 1055 1060 1065  
 Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro  
 1070 1075 1080  
 Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu  
 1085 1090 1095  
 Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp  
 1100 1105 1110  
 Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly  
 1115 1120 1125  
 Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln  
 1130 1135 1140  
 Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly  
 1145 1150 1155  
 Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr  
 1160 1165 1170  
 Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu  
 1175 1180 1185  
 Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu  
 1190 1195 1200  
 Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala  
 1205 1210 1215  
 Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu  
 1220 1225 1230  
 Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly  
 1235 1240 1245

ES 2 567 305 T3

Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val  
 1250 1255 1260  
 Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala  
 1265 1270 1275  
 Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly  
 1280 1285 1290  
 Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser  
 1295 1300 1305  
 Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val  
 1310 1315 1320  
 Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala  
 1325 1330 1335  
 Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser  
 1340 1345 1350  
 Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys  
 1355 1360 1365  
 Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile  
 1370 1375 1380  
 Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp  
 1385 1390 1395  
 Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu  
 1400 1405 1410  
 Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp  
 1415 1420 1425  
 Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe  
 1430 1435 1440  
 Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro  
 1445 1450 1455  
 Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr  
 1460 1465 1470  
 Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu  
 1475 1480 1485  
 Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu  
 1490 1495 1500

<210> 29  
 <211> 1500  
 <212> ADN  
 <213> *Schizochytrium sp.*

5

<220>  
 <221> CDS  
 <222> (1)..(1500)  
 <223>

10

<400> 29

15

aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc ccc gag

48

ES 2 567 305 T3

Lys	Val	Gln	Pro	Val	Phe	Ala	Asn	Gly	Ala	Ala	Thr	Val	Gly	Pro	Glu		
1				5					10					15			
gcc	tcc	aag	gct	tcc	tcc	ggc	gcc	agc	gcc	agc	gcc	agc	gcc	gcc	ccg		96
Ala	Ser	Lys	Ala	Ser	Ser	Gly	Ala	Ser	Ala	Ser	Ala	Ser	Ala	Ala	Pro		
			20					25					30				
gcc	aag	cct	gcc	ttc	agc	gcc	gat	gtt	ctt	gcg	ccc	aag	ccc	gtt	gcc		144
Ala	Lys	Pro	Ala	Phe	Ser	Ala	Asp	Val	Leu	Ala	Pro	Lys	Pro	Val	Ala		
		35					40					45					
ctt	ccc	gag	cac	atc	ctc	aag	ggc	gac	gcc	ctc	gcc	ccc	aag	gag	atg		192
Leu	Pro	Glu	His	Ile	Leu	Lys	Gly	Asp	Ala	Leu	Ala	Pro	Lys	Glu	Met		
	50					55					60						
tcc	tgg	cac	ccc	atg	gcc	cgc	atc	ccg	ggc	aac	ccg	acg	ccc	tct	ttt		240
Ser	Trp	His	Pro	Met	Ala	Arg	Ile	Pro	Gly	Asn	Pro	Thr	Pro	Scr	Phe		
	65			70						75					80		
gcg	ccc	tcg	gcc	tac	aag	ccg	cgc	aac	atc	gcc	ttt	acg	ccc	ttc	ccc		288
Ala	Pro	Ser	Ala	Tyr	Lys	Pro	Arg	Asn	Ile	Ala	Phe	Thr	Pro	Phe	Pro		
			85					90						95			
ggc	aac	ccc	aac	gat	aac	gac	cac	acc	ccg	ggc	aag	atg	ccg	ctc	acc		336
Gly	Asn	Pro	Asn	Asp	Asn	Asp	His	Thr	Pro	Gly	Lys	Met	Pro	Leu	Thr		
			100					105					110				
tgg	ttc	aac	atg	gcc	gag	ttc	atg	gcc	ggc	aag	gtc	agc	atg	tgc	ctc		384
Trp	Phe	Asn	Met	Ala	Glu	Phe	Met	Ala	Gly	Lys	Val	Ser	Met	Cys	Leu		
		115					120					125					
ggc	ccc	gag	ttc	gcc	aag	ttc	gac	tcg	aac	acc	agc	cgc	agc	ccc			432
Gly	Pro	Glu	Phe	Ala	Lys	Phe	Asp	Asp	Ser	Asn	Thr	Ser	Arg	Ser	Pro		
		130				135					140						
gct	tgg	gac	ctc	gct	ctc	gct	acc	cgc	gcc	gtg	tct	gtg	tct	gac	ctc		480
Ala	Trp	Asp	Leu	Ala	Leu	Val	Thr	Arg	Ala	Val	Ser	Val	Ser	Asp	Leu		
				150						155					160		
aag	cac	gtc	aac	tac	cgc	aac	atc	gac	ctc	gac	ccc	tcc	aag	ggt	acc		528
Lys	His	Val	Asn	Tyr	Arg	Asn	Ile	Asp	Leu	Asp	Pro	Ser	Lys	Gly	Thr		
				165					170					175			
atg	gtc	ggc	gag	ttc	gac	tgc	ccc	gcg	gac	gcc	tgg	ttc	tac	aag	ggc		576
Met	Val	Gly	Glu	Phe	Asp	Cys	Pro	Ala	Asp	Ala	Trp	Phe	Tyr	Lys	Gly		
			180					185					190				
gcc	tgc	aac	gat	gcc	cac	atg	ccg	tac	tcg	atc	ctc	atg	gag	atc	gcc		624
Ala	Cys	Asn	Asp	Ala	His	Met	Pro	Tyr	Ser	Ile	Leu	Met	Glu	Ile	Ala		
		195					200					205					
ctc	cag	acc	tcg	ggt	gtg	ctc	acc	tcg	gtg	ctc	aag	gcg	ccc	ctg	acc		672
Leu	Gln	Thr	Ser	Gly	Val	Leu	Thr	Ser	Val	Leu	Lys	Ala	Pro	Leu	Thr		
		210				215						220					
atg	gag	aag	gac	gac	atc	ctc	ttc	cgc	aac	ctc	gac	gcc	aac	gcc	gag		720
Met	Glu	Lys	Asp	Asp	Ile	Leu	Phe	Arg	Asn	Leu	Asp	Ala	Asn	Ala	Glu		
		225			230					235					240		
ttc	gtg	cgc	gcc	gac	ctc	gac	tac	cgc	ggc	aag	act	atc	cgc	aac	gtc		768
Phe	Val	Arg	Ala	Asp	Leu	Asp	Tyr	Arg	Gly	Lys	Thr	Ile	Arg	Asn	Val		
			245						250					255			
acc	aag	tgc	act	ggc	tac	agc	atg	ctc	ggc	gag	atg	ggc	gtc	cac	cgc		816
Thr	Lys	Cys	Thr	Gly	Tyr	Ser	Met	Leu	Gly	Glu	Met	Gly	Val	His	Arg		
			260					265					270				
ttc	acc	ttt	gag	ctc	tac	gtc	gat	gat	gtg	ctc	ttt	tac	aag	ggc	tcg		864
Phe	Thr	Phe	Glu	Leu	Tyr	Val	Asp	Asp	Val	Leu	Phe	Tyr	Lys	Gly	Ser		
		275					280					285					
acc	tcg	ttc	ggc	tgg	ttc	gtg	ccc	gag	gtc	ttt	gcc	gcc	cag	gcc	ggc		912
Thr	Ser	Phe	Gly	Trp	Phe	Val	Pro	Glu	Val	Phe	Ala	Ala	Gln	Ala	Gly		
		290				295						300					
ctc	gac	aac	ggc	cgc	aag	tcg	gag	ccc	tgg	ttc	att	gag	aac	aag	gtt		960
Leu	Asp	Asn	Gly	Arg	Lys	Ser	Glu	Pro	Trp	Phe	Ile	Glu	Asn	Lys	Val		
		305			310					315					320		
ccg	gcc	tcg	cag	gtc	tcc	ttt	gac	gtg	cgc	ccc	aac	ggc	agc	ggc			1008
Pro	Ala	Ser	Gln	Val	Ser	Ser	Phe	Asp	Val	Arg	Pro	Asn	Gly	Ser	Gly		
				325				330					335				
cgc	acc	gcc	atc	ttc	gcc	aac	gcc	ccc	agc	ggc	gcc	cag	ctc	aac	cgc		1056

ES 2 567 305 T3

Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg  
 340 345 350  
 cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc tcc ggc 1104  
 Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly  
 355 360 365  
 agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg gtc aac 1152  
 Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn  
 370 375 380  
 ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg gtc atg 1200  
 Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met  
 385 390 395 400  
 ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag gcc atc 1248  
 Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile  
 405 410 415  
 gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa ccc cac 1296  
 Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln Pro His  
 420 425 430  
 ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc ggc cag 1344  
 Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg Gly Gln  
 435 440 445  
 ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc gtg tcc 1392  
 Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser  
 450 455 460  
 gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc ctc 1440  
 Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu  
 465 470 475 480  
 tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg cgc 1488  
 Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg  
 485 490 495  
 atc gcc tcc ggt 1500  
 Ile Ala Ser Gly  
 500

<210> 30  
 <211> 500  
 <212> PRT  
 <213> *Schizochytrium sp.*

5

<400> 30

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

10



ES 2 567 305 T3

Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro  
 130 135 140

Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu  
 145 150 155 160

Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr  
 165 170 175

Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly  
 180 185 190

Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala  
 195 200 205

Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr  
 210 215 220

Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu  
 225 230 235 240

Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val  
 245 250 255

Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg  
 260 265 270

Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser  
 275 280 285

Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly  
 290 295 300

Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val  
 305 310 315 320

Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly Ser Gly  
 325 330 335

Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg  
 340 345 350

Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly  
 355 360 365

Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn  
 370 375 380

Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met  
 385 390 395 400

Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile  
 405 410 415

Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln Pro His  
 420 425 430

Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg Gly Gln  
 435 440 445

Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser  
 450 455 460

Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu  
 465 470 475 480

Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg  
 485 490 495

Ile Ala Ser Gly  
 500

5 <210> 31  
 <211> 1512  
 <212> ADN

ES 2 567 305 T3

<213> *Schizochytrium* sp.

<220>

<221> CDS

5

<222> (1)..(1512)

<223>

<400> 31

```

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

```

10

ES 2 567 305 T3

Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu  
 225 230 235 240  
 cgc aac cgc ctg cac cgc gag tgc ggc tac ccc gcg cac ctc cgc gtc 768  
 Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val  
 245 250 255  
 cgc gtt ggc gcc ggc ggt ggc gtc ggc tgc ccg cag gcc gcc gcc gcc 816  
 Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala  
 260 265 270  
 gcg ctc acc atg ggc gcc gcc ttc atc gtc acc ggc act gtc aac cag 864  
 Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln  
 275 280 285  
 gtc gcc aag cag tcc ggc acc tgc gac aac gtg cgc aag cag ctc tcg 912  
 Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser  
 290 295 300  
 cag gcc acc tac tcg gat atc tgc atg gcc ccg gcc gcc gac atg ttc 960  
 Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe  
 305 310 315 320  
 gag gag ggc gtc aag ctc cag gtc ctc aag aag gga acc atg ttc ccc 1008  
 Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro  
 325 330 335  
 tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc 1056  
 Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe  
 340 345 350  
 gac tcc atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc 1104  
 Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe  
 355 360 365  
 aag cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att 1152  
 Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile  
 370 375 380  
 aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac ccc 1200  
 Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro  
 385 390 395 400  
 aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt gcc agc 1248  
 Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser  
 405 410 415  
 cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac tac cag gtc 1296  
 Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val  
 420 425 430  
 tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc atc aag ggc acc 1344  
 Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr  
 435 440 445  
 tac ctc gac ccc gct gtc tcc aac gag tac ccc tgt gtc gtc cag atc 1392  
 Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile  
 450 455 460  
 aac ctg caa atc ctc cgt ggt gcc tgc tac ctg cgc cgt ctc aac gcc 1440  
 Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala  
 465 470 475 480  
 ctg cgc aac gac ccg cgc att gac ctc gag acc gag gat gct gcc ttt 1488  
 Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe  
 485 490 495  
 gtc tac gag ccc acc aac gcg ctc 1512  
 Val Tyr Glu Pro Thr Asn Ala Leu  
 500

<210> 32  
 <211> 504  
 <212> PRT  
 <213> *Schizochytrium* sp.

<400> 32

Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys  
 1 5 10 15

ES 2 567 305 T3

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

<210> 33  
 <211> 9  
 <212> PRT  
 <213> Secuencia artificial

5

<220>  
 <223> motivo

10

<220>  
 <221> MISC\_FEATURE

ES 2 567 305 T3

<222> (2)..(3)  
 <223> x = cualquier aminoácido

5 <220>  
 <221> MISC\_FEATURE  
 <222> (6)..(6)  
 <223> x = A o S

10 <220>  
 <221> MISC\_FEATURE  
 <222> (7)..(8)  
 <223> x = cualquier aminoácido

15 <400> 33  
 Trp Xaa Xaa Lys Glu Xaa Xaa Xaa Lys  
 1 5

20 <210> 34  
 <211> 6  
 <212> PRT  
 <213> Secuencia artificial

25 <220>  
 <223> motivo

30 <220>  
 <221> MISC\_FEATURE  
 <222> (3)..(3)  
 <223> x = I o L o V

<400> 34  
 Phe Asn Xaa Ser His Ser  
 1 5

35 <210> 35  
 <211> 5  
 <212> PRT  
 <213> Secuencia artificial

40 <220>  
 <223> motivo

45 <220>  
 <221> MISC\_FEATURE  
 <222> (1)..(5)  
 <223> x = I o L o V

<400> 35  
 Xaa Gly Xaa Asp Xaa  
 1 5

55 <210> 36  
 <211> 4244  
 <212> ADN  
 <213> *Schizochytrium sp.*

<400> 36

ES 2 567 305 T3

tttctctctc	togagetggt	getgetgctg	ctgctgetgc	tgettctctg	ctggttctca	60
cgctccgttc	atcaagcgct	cgctcgctcg	accgatcggt	gcgtgcgtgc	gtgcgtgagt	120
cttggtgcc	ggcagccgca	ggctgtctgt	ctgtttctgt	agttttacc	tcgggggttc	180
gggtctgcct	gectcccgct	ccgcccgc	gcgcccgt	tcaccccgc	tcgctccgc	240
ccatcgggcc	tcgctcctc	gcgcccgc	catcgccgc	atcgatgca	tcgatgctgc	300
acgcacgggg	ggacgcgcgc	ccgctgccc	ccgcccgc	cgctgctgc	tggcgatgcc	360
gtcgccgccc	tccttctctc	cctcgctcc	tccttctccc	gagccccct	gtcttctctc	420
gccccgcag	cggcgccgag	gaagcgagga	gagcggggag	gagagaagaa	aagaaaagaa	480
aagaaaagaa	aataacagcg	ccgtctcgcg	cagacgcgcg	cggccgcgtg	cgagcggcg	540
tgatggggct	tctcgtggcg	cgctgcccgc	ctggcccgc	ctcgctttg	agggtcaggc	600
tttgggagag	aagagtggga	cgcgagaaag	ataagatggt	gccatggcgc	aggacggaga	660
ggttgcgtgaa	acttcttcga	gcggcacagg	cgatggcgag	agaccgacag	ctgcccgcgc	720
ggaggggatg	gatacctccc	gaggctggca	tgagagagct	ggccgcgcgg	atctggctgg	780
ccgcccggcg	gtgggtccgg	aggcgcgagg	ttggtttct	tcatacctga	taccatacgg	840
tattcattct	tcctctccag	gaaggaagca	agtcacatag	agtatcacta	gcctaattgat	900
ggactctatg	tttttagggca	cgctggagca	gaaggcgcga	gcgattcgaa	tgcgagcgat	960
agatacccg	cagagacctt	gcggcgacgc	cggatgcagg	cagacgcga	cgacccgac	1020
gcacggcagc	ggtgcacgcg	ctcctggca	gatgcacggg	tctgcgcgc	gcctttacat	1080
tttttgatgt	taggtggtgt	gcctgccact	ttgaacatca	tcacaagtc	aacgcagcat	1140
caagaggcaa	gcaagtacat	acatccattc	gaattcaagt	tcaagagacg	cagcaacagc	1200
cgccgctcgc	ctcaagctgc	agctagctgg	ctgacagggc	tcgctggctg	tagtggaaaa	1260
ttccattcac	ttttctgcat	ccgcccgcag	caggcccgt	cgacgcttct	ctcgtttgtt	1320
tgttcggttc	tgcgctcgctg	cgctgcgtccc	agctgcctgt	ctaactctgcc	gcgcgatcca	1380
acgaccctcg	gtcgctcgcc	caagcgaaac	ccgacgcga	cctggccaat	gccgcaagaa	1440
tgctaagcgc	gcagcaatgc	tgagagtaat	cttcagcca	ccaagtcaat	atcgctgcc	1500
aagtctccat	cgacgccaca	ttcaggcttt	ctctctctct	ccctccctct	ctttctgccc	1560
ggagagaagg	aaagaccgcg	cgccgcgcgc	tctgcgcctg	tgacgggctg	tcggttgtaa	1620
gccctcttag	acagttccta	ggtgccgggc	gccgcccgc	ctccgtcgca	ggcacacgta	1680
ggcggccacg	ggttccccc	gcacctcca	caecttctc	ccccgcagcc	ggaccgcgcg	1740
ccgtctgctt	acgcacttcg	cgccgcgcgc	gccgcgcaac	ccgagcgcgt	gctgtgggcg	1800
ccgtcttccg	gccgcgtcgg	aggctcgtccc	cgccgcgcgc	tactccgggt	cctgtgctgc	1860
acgtacttaa	tattaacagt	gggacctcgc	acaggacctg	acggcagcac	agacgtcgc	1920
gcctcgcac	gctggggagc	caggcgaggc	atcccgcgc	ggccccgcac	cggggaggct	1980
gcggggcgcg	ctcttccggc	cggcggccgc	atcaggcggg	tgacgcaaga	gccctcgcag	2040
tcgctcgcct	gcgggagcgc	agcgcggcgc	cagcgtggcc	aagctcccgc	cccttctggc	2100
tggtgcatg	cctgcctgc	tgctgcctg	cgctgcgtgc	tgctgcgtg	ccttctgctg	2160
tgctgcctt	cgtgcgtgcg	tgctgagtg	cggcggaaga	gggatcatgc	gaggatgagc	2220
caccgcgcgc	acctcgactt	ttgaagaagc	cgcgatgcga	tgcgatgcga	tgcgatgcga	2280
cgcgataccg	tgcgaggcta	cgaaagcagc	ctggcccgc	gtcatacaac	gcacgttttc	2340
gagaaggagg	gctggcggag	gcgtgcatgc	cggcgaccat	tgcaaacgcg	gcgtctcgtg	2400
gctgcgcgaag	gtgcctggag	gatctaacga	ctgctgctat	gatgctatag	ctgtgctgat	2460
ccccggtcca	ttccaccaag	tcgtgctctg	ccgctgacc	tgcttggc	tttcttcaa	2520
gttctcctcc	gccgggcctt	caggaccgag	acgagacctg	cagctgcagc	tagactcgcg	2580
ctcgctcgcg	gaggattcgc	cggccgcgcg	gccggacggg	actcgcgagg	tcacacggcc	2640
cccggcgcgc	gcgatggctg	tgctgacgta	ctcgtgcgtg	gcagccgtac	gtcagcagcg	2700
cgccctccgt	attgtggatt	cgttagttgg	ttgttggttg	atttggttgat	taattttttt	2760
gttcgtaggc	ttggttatag	ctaatagttt	agtttatact	ggtgctcttc	ggtgctgatt	2820
tagctcgact	tggttccaca	ccactgcccc	tctactgtga	atggatcaat	ggacgcacga	2880
cgggcgcgac	aaagtgcgcg	agtgaggtaa	cctaagcaac	ggcggctctc	agaggggacg	2940
cacgccctcc	gtcgcaagca	gtccagacag	gcagaaaagc	gtcttaggga	ccacgcacgc	3000
acgcacgcac	gcacgcacgc	ccgcacgcac	gctccctccc	tcgctgcctc	attttttttag	3060
gcttccctcc	gcacgggcct	acctctcgtc	ccctcgcctc	gccgcaccag	gcggcagcag	3120
cgataccctg	cgggtgcgcg	tcctgcacgc	gctcagccgc	agctcagccc	agccgcgagc	3180
taggtttgtg	tcgtcctgaa	ttgtttgatt	tgatttgatt	tgatttgatc	cgatccgatc	3240
cgatctgatc	tgatttgctt	tgctttgctt	tgcttccctc	ccggcgcgga	ccaagcgtcc	3300
gtctgcgcgc	cgcagcttcc	cttcttctcc	cagccctcct	tctgctcccg	cctctcgcgc	3360
aagcacgcag	cttgcgcgcg	gcataccggc	ggtcggctcg	tcgatcgacc	cgcttgcgcg	3420
tgctgcctgtg	gccgggcttt	tcctccatcg	cgactcttcc	ttctccatac	gtcctactac	3480
gtacatacat	actgccggct	tcctcctctt	ccagcgcgcg	gacggcggca	ggctgcgacg	3540
tcgtgcgcgc	cgcgggcgcg	gcgcgcgcgc	ccgcgcgcgc	ccgcgtcgca	gggcctcgtc	3600
gccgcgcgcg	ctccgctccg	ctccgagggc	gcgagagggc	cgcggcggcg	cgatggatgg	3660
atggatggatg	ggatggatgg	atggattttg	tgatcgatg	gcggcgcgatg	ggcgagatg	3720
agcgaggagc	agcgcgcgag	cgccgcagcc	ggattcgcag	ggcctcgtc	gcctcgcgc	3780
cgctgcgcgc	cccgccttgc	gagcctgcgc	cgcgagcgcg	cgagcgcgag	agcggggctt	3840
tctttgtctc	gcgcgcgcgt	tgccctcgtg	tgcttctgct	ttgcgtagcg	ggcgcgcgcg	3900
tggaagatgg	ctcattcaat	cgaccatttc	acgcacgcac	tcggcgcgcg	agagaaggcc	3960
gaggaggagc	agcaagcaaa	ccaaaagctc	tcgcgctcgc	ggtctcgggc	tcgagcggctc	4020
tcggagagag	agtcttgccg	cgaccaccgg	cagcagcagc	agcagcagca	gcgctgtcga	4080
gcacgcagc	gagcagcagc	acgagcagca	gcattcgcagc	aagaggacag	acacgggttg	4140
caagccctag	ctcgcctgat	acagaaagag	gcgggttggg	cgtaaaaaaa	aaggagcagc	4200
caagccgcca	gccagccagc	tagctagcca	gcctgcctgc	caaa		4244

ES 2 567 305 T3

<211> 3886  
 <212> ADN  
 <213> *Schizochytrium* sp.

5 <220>  
 <221> misc\_feature  
 <222> (2115)..(2115)  
 <223> n = a, c, g o t

10 <400> 37

gatcttgatt	gccaaagctct	ggattgtcga	ttccgatgaa	tcgagctcct	tgttgtcgag	60
ctctggcttg	ccgagctttc	agaaatagac	aaaattgccg	agttcctgat	tgcggggctc	120
tcgattgcca	aggtctggtg	gattctcgaa	ctctcgattg	tcaaaatcct	ggtcgtctcg	180
tcggattcct	tcctgatttg	ttttgtcaag	acottgagat	tgtgcaaaac	cttgatcgtt	240
gacaaaaccct	tgatcgacag	cagcctttca	tcacgctcag	ctcttgatc	tgattatatt	300
ccccctgaca	gccaacacct	tgatgcaggg	tctcaacctt	gatttttggg	ggccatcatc	360
agcatcacgc	cccggcactc	accctcaaca	ttcgacagcc	aacgcttttt	tttcttcgac	420
taggatctga	gaataaaaagc	aggtcaccac	gaccgtaggc	caacgcgaca	accatggaaa	480
taaagtgaca	acgaacgact	tgcaagttta	aatgtaaaga	gcagcaattg	cccgccaca	540
gacaaatgaa	agcaggcgcc	gagtcttatt	tgaggagggtg	ggcctgtggc	aatgggcgaa	600
agaaaatcaa	ggacaaggag	agcaggttac	gtaccggtat	actggtatac	gtacatggat	660
ggttcttggc	aagttgacgg	gatgtgtgcg	agtgaccgtg	gtagttaacg	aaagagccgc	720

# ES 2 567 305 T3

aagggcaagg	aaagcaagag	aatgcagact	tttccacag	atggatgggt	ccgcagcttg	780
ccgcatgatg	aaacgctgta	tttcacctgg	cacgtgggtg	cgcacgcgcc	cacatatgat	840
cgcgggcgcg	ggtgtattat	acatthttccc	cctcaggtct	actgccatcc	ctccatgcgt	900
cgctcgtgcg	aacgacgcaa	gocthttcgca	togtgcagcc	tctttctggg	aaggcaagag	960
ctaaaccocaa	acctaaccga	aagaaacattt	ttacctctct	ctctctocca	ttggctcgct	1020
gogctccgcc	gctcgcctct	cctcctgccca	gtgtcoggcc	ctaacttccc	ccctccctcc	1080
ctccctccct	ccctccctct	ctcctgccac	cgccccctc	tccgcgctgc	gtgcgggtgct	1140
gccctggacc	aatggcatgc	tgctgcacgc	tggcggtgat	acgcaagccg	cttcgcaatt	1200
tccggatcag	atctcggcgg	ggcgtgcgcc	gcggggcac	tgcggaacctg	ccgcgcccc	1260
tgcttctttc	acatccatca	tgtcctccaa	acctccgcct	cctccacgca	cgtagcgcag	1320
cccgctcgca	cgcgcgcact	gocgctgcga	aagcaagcgc	ccgcccgcgg	cccgcgcaag	1380
ggaggcgggc	cgcggtctcc	ctccgcgggt	gocctcgtcc	cgcgcggggc	tgggccccga	1440
gcgaagcgcg	ggtggcggcg	cgcgctcccg	tcttcgtcag	cgccctacgt	cgcgcgccgc	1500
gcgagagact	acgcatgccc	ttgcgtcatg	cgctcgcagg	tagccgcgcg	ggccttagcg	1560
tttccgctgg	cgccgcgcct	aagccccggg	cgcgcacggg	attgcgcgca	taccgtacgg	1620
ccaagaccgc	cgcacagctc	ggccctctcg	cgccacagca	gccagcagcg	cagcgaggga	1680
agagcgcgca	ggcgcgcgcg	gagggcggcc	gcagagcagc	gcagagcggg	ccggagcagc	1740
gcggagcaga	acgggcagac	tcggagcggg	cagggcgggc	agagctttgg	ggtttaagga	1800
ccgggttacc	ggcgaagtga	gocgctgcgg	ggagcgcgctg	tgggaggggt	gagtagcga	1860
gcacgatgcg	agcgagagag	agacgctgcc	gcgaatcaag	aaggtaggcg	cgctgcgag	1920
gcggcgggcg	gagcgagcgg	agggagaggg	agagggagag	agagggaggg	agacgtcgcc	1980
gcggcggggc	ctggcctggc	ctggtttggc	ttggtcagcg	cgccctgtgc	cgagcgtgca	2040
gctggagttg	ggtggattca	tttggatttt	cttttgtttt	tgtttttctc	tctttcccgg	2100
aaagtgttgg	ccggncgggtg	ttctttgttt	tgattttctc	aaaagttttg	gtggtttggt	2160
ctctctcttg	gctctctgtc	aggcggtcog	gtcccagccc	cgccctctcc	tctcctctcc	2220
tctcctctcc	tctcctctcg	tatacgtacg	tacgtttgta	tacgtacata	catcccggcc	2280
gcgctgcggg	cgagggtttg	ctcagcctgg	agcaatgcga	tgcatgcgca	tgcatgcgca	2340
cgagcgcgca	cgcgagtcac	tggttcgcgc	tgtggctgtg	gcttgcttgc	ttactttgct	2400
tcgagctctc	ccgcttctct	cttctctctc	cacgccacca	ccaacgaaag	aagatcgccc	2460
ccggcacgcc	gctgagaagg	gctggcggcg	atgacggcac	gcgcgcggcg	tgccacgttg	2520
gcgctcgtcg	ctgctgctgc	tgctgctgct	gctgctgctg	ctgctgctgc	tgctgcttct	2580
gcgagcaggg	tttgccacga	ggccggcgctg	ctggccgctg	ccgcttccag	tccgcgtgga	2640
gagatcgaat	gagagataaa	ctggatggat	tcactcaggg	atgaatgaac	gatggttgga	2700
tgcttttttc	ctttttcagg	tccacagcgg	gaagcaggag	cgctgaatc	tggcccatc	2760
cgcatatcgt	tgcatcgcac	cgcatcgcac	gcacgcacgc	ctcgcgggga	gccacagacg	2820
ggcgacaggg	cgccacgcca	gccaggcagc	cagccaggca	ggcaccagag	ggccagagag	2880
cgccctctac	gcacgcgcgg	cagtgcgcgc	atcgtcgcga	gtgcagacct	tgattccccg	2940
cgcggtctct	cgcgagcccc	aaacgaagag	cgccgtacgg	gccccatcta	gcgtgcctct	3000
gcaccgcac	gcatcgcac	gcgttcccct	gagagtagta	ctgcagcaag	gcaccatttc	3060
cgcgctcctc	ttcggcgcga	tcgaggcccc	cggcgcggcg	acgatcgcgg	cgccgcggcg	3120
gctggcggcg	gccctggcgc	tcgcgctggc	ggccgcggcg	ggcgtctggc	cctggcggcg	3180
gcgggcggcg	caggaggagc	ggcagcggct	gctgcggcgc	agagaagagc	gcgcggggcc	3240
cggggagggg	cggggagggg	aaaggagaag	cgcgcaaggc	ggccccgaaa	gagaagaccc	3300
tgactttgaa	cgcgaaagaag	aaagaagaag	agaagaagtt	gaagaagaag	aagaagaagg	3360
agaggaagtt	gaagaagacg	aggagcaggg	gcgttccaag	gcgcgttctc	ttccggaggc	3420
gcgttccagc	tgcgggcgcg	ggcggggctg	cggggcgggc	gcgggcggcg	gtgcggggcag	3480
aggggacgcg	cgcgcgaggg	cgaggggggc	cgagcgggag	cccctgctgc	tgcgggggcg	3540
ccgggcccga	ggtgtggcgc	gcgcgacgac	ggaggcgacg	acgccagcgg	ccgcgacgac	3600
aaggccggcg	gcgtcggcgg	gcggaaggcc	ccgcgcggag	caggggcggg	agcaggacaa	3660
ggcgaggag	caggagcagg	gccgggagcg	ggagcgggag	cgggcgggcg	agcccgaggc	3720
agaacccaat	cgagatccag	agcgagcaga	ggccggccgc	gagcccagag	ccgcgcccga	3780
gatcactagt	accgctcggg	aatcacagca	gcagcagcag	cagcagcagc	agcagcagca	3840
gcagcagcag	ccacgagag	gagataaaga	aaaagcgcca	gagacg		3886

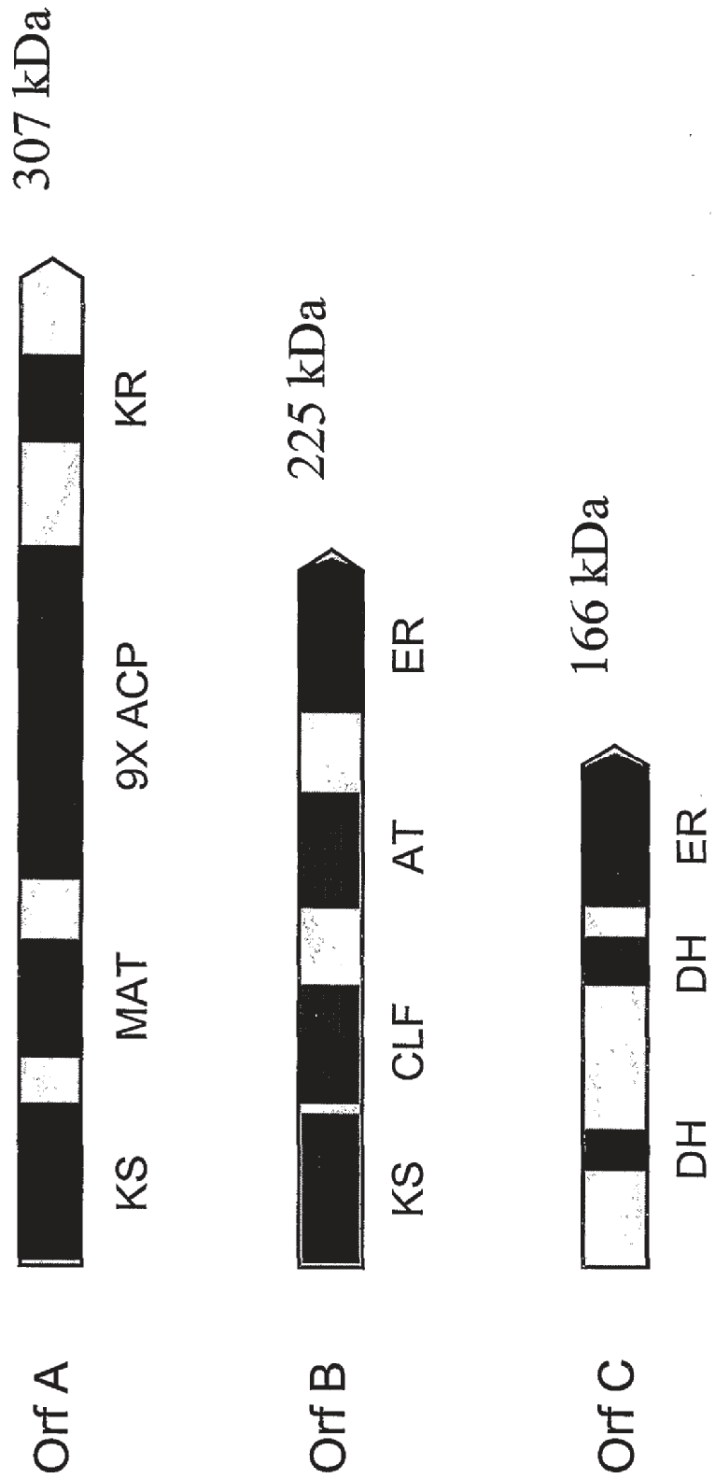


## REIVINDICACIONES

1. Una molécula aislada de ácido nucleico que comprende una secuencia de ácido nucleico seleccionada entre el grupo que consiste en:
- 5 a. una secuencia de ácido nucleico que codifica una secuencia de aminoácidos de: SEC ID N° 20, y fragmentos biológicamente activos de la misma, en la que los fragmentos biológicamente activos muestran una actividad  $\beta$ -ceto acil-ACP sintasa (KS);
- 10 b. una secuencia de ácido nucleico que codifica una secuencia de aminoácidos que es al menos un 60 % idéntica a dicha secuencia de aminoácidos de (a), en la que dicha secuencia de aminoácidos muestra una actividad  $\beta$ -ceto acil-ACP sintasa (KS); y
- c. una secuencia de ácido nucleico que es completamente complementaria a la secuencia de ácido nucleico de (a) o (b).
- 15 2. La molécula aislada de ácido nucleico de la reivindicación 1, en la que dicha molécula de ácido nucleico comprende la secuencia de ácido nucleico de: SEC ID N° 19.
3. Una molécula de ácido nucleico recombinante que comprende la molécula de ácido nucleico de la reivindicación 1 o 2, unida de forma funcional a al menos una secuencia de control de la transcripción.
- 20 4. Una célula recombinante transfectada con la molécula de ácido nucleico recombinante de la reivindicación 3.
5. Un microorganismo modificado genéticamente, en el que dicho microorganismo expresa un sistema PKS que comprende al menos un dominio biológicamente activo de un sistema PUFA PKS, en el que dicho microorganismo se ha modificado genéticamente por transfección con una molécula de ácido nucleico recombinante que codifica la secuencia de ácido nucleico de la reivindicación 1 o de la reivindicación 2, y en el que dicha molécula de ácido nucleico recombinante no es endógena a dicho microorganismo modificado genéticamente.
- 25 6. El microorganismo modificado genéticamente de la reivindicación 5, en el que dicho microorganismo se ha modificado genéticamente adicionalmente para que exprese de forma recombinante al menos una molécula de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PKS seleccionado entre el grupo que consiste en: un sistema PUFA PKS bacteriano; un sistema PKS Tipo I; un sistema PKS Tipo II; y un sistema PKS modular.
- 30 7. Una planta modificada genéticamente, en la que dicha planta se ha modificado genéticamente para que exprese de forma recombinante un sistema PKS que comprende al menos un dominio biológicamente activo de un sistema PUFA PKS, en la que dicho dominio está codificado por dicha secuencia de ácido nucleico de la reivindicación 1 o de la reivindicación 2.
- 35 8. La planta modificada genéticamente de la reivindicación 7, en la que dicha planta se ha modificado genéticamente adicionalmente para que exprese de forma recombinante al menos una molécula de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PKS seleccionado entre el grupo que consiste en: un sistema PUFA PKS bacteriano; un sistema PKS Tipo I; un sistema PKS Tipo II; y un sistema PKS modular.
- 40 9. Un método para producir una molécula bioactiva que se produce por un sistema PKS, que comprende hacer crecer o cultivar, en condiciones eficaces para producir dicha molécula bioactiva, un organismo modificado genéticamente, en el que dicho organismo modificado genéticamente es el microorganismo modificado genéticamente de acuerdo con la reivindicación 5 o 6 o la planta modificada genéticamente de la reivindicación 7 u 8 que exprese un sistema PKS que comprende al menos un dominio biológicamente activo del sistema PUFA PKS.
- 45 10. Un método para producir una planta que tiene un perfil de PUFA que difiere de la planta de origen natural, que comprende modificar genéticamente células de dicha planta para que expresen un sistema PKS que comprende al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS, en el que dicho al menos un dominio de dicho sistema PUFA PKS está codificado por una secuencia de ácido nucleico de la reivindicación 1 o de la reivindicación 2.
- 50 11. Un método para producir leche animal humanizada, que comprende modificar genéticamente células productoras de leche de un animal no humano productor de leche con al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS, en el que dicho al menos un dominio de dicho sistema PUFA PKS está codificado por una secuencia de ácido nucleico de la reivindicación 1 o de la reivindicación 2.
- 55 12. Un método para producir un microbio recombinante, que comprende modificar genéticamente células microbianas para que expresen al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica al menos un dominio biológicamente activo de un sistema PUFA PKS, en
- 60
- 65

el que dicho al menos un dominio de dicho sistema PUFA PKS está codificado por una secuencia de ácido nucleico de acuerdo con la reivindicación 1 o la reivindicación 2.

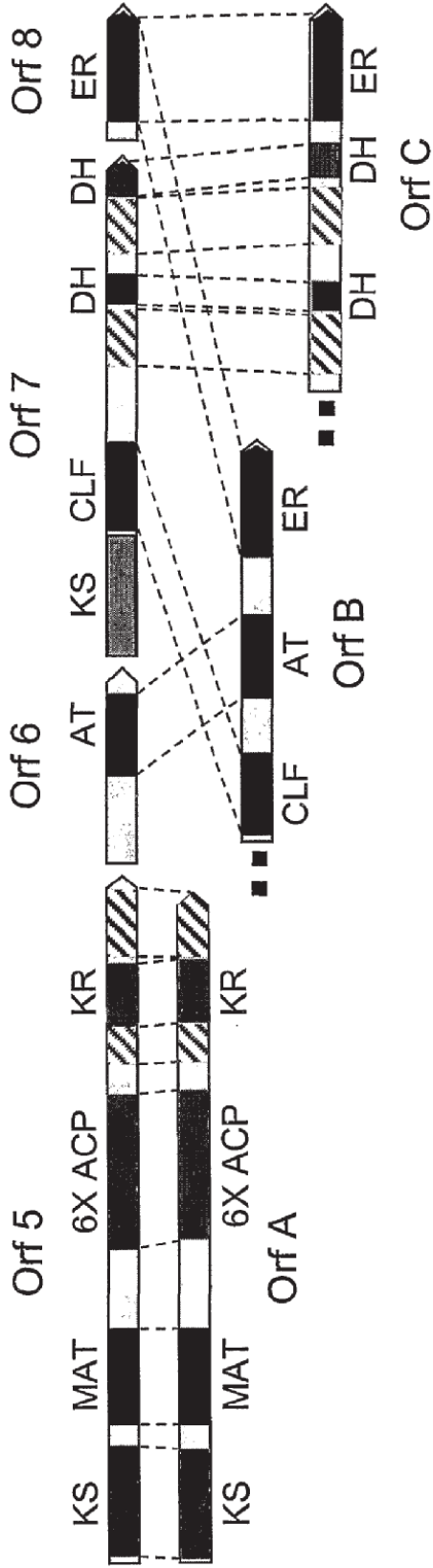
FIG. 1



# FIG. 2

## Comparación de ORF/dominios de PKS:

*Schizochytrium frente a Shewanella*



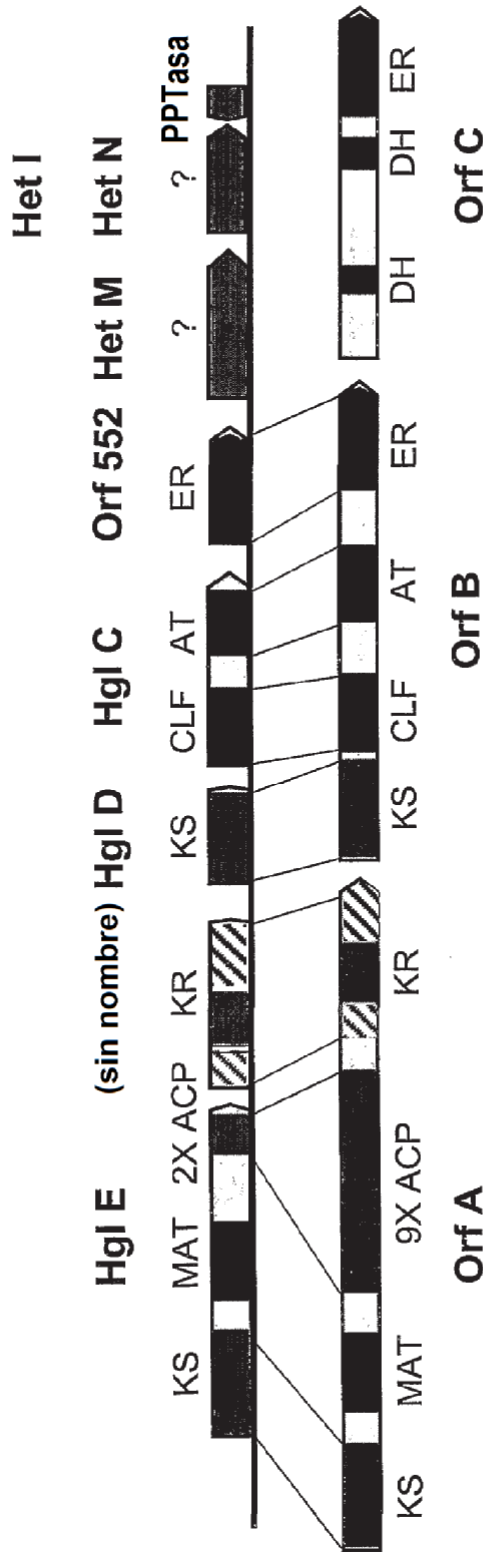
**Homología de secuencia de aminoácidos**



FIG. 3

***Nostoc (7120) frente a Schizochytrium***

ORF de Nostoc en el grupo HgI/Het  
(orden lineal en el cromosoma)



**ORF de PUFA PKS de Schizochytrium  
(Orf separadas)**

