

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 467 918**

51 Int. Cl.:

C12P 7/64 (2006.01)

C12N 9/00 (2006.01)

C12N 15/52 (2006.01)

C12N 15/82 (2006.01)

C12P 7/62 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.05.2008 E 08755645 (2)**

97 Fecha y número de publicación de la concesión europea: **05.03.2014 EP 2160470**

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

30 Prioridad:

16.05.2007 US 749686

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

13.06.2014

73 Titular/es:

DSM IP ASSETS B.V. (100.0%)

Het Overloon 1

6411 TE Heerlen , NL

72 Inventor/es:

WEAVER, CRAIG, A.;

ZIRKLE, ROSS;

DOHERTY, DANIEL, H. y

METZ, JAMES, G.

74 Agente/Representante:

PÉREZ BARQUÍN, Eliana

ES 2 467 918 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 policétido sintasa de AGPI quiméricos y usos de los mismos

5 Campo de la invención

Esta invención se refiere a sistemas de policétido sintasa (PKS) de ácidos grasos poliinsaturados (AGPI) quiméricos, y particularmente, a sistemas de PKS de AGPI quiméricos de *Schizochytrium* y *Thraustochytrium*. Más particularmente, se describen en el presente documento ácidos nucleicos que codifican para tales sistemas de PKS de AGPI, estos sistemas de PKS de AGPI, organismos modificados genéticamente que comprenden tales sistemas de PKS de AGPI, y métodos de preparación y uso de tales sistemas de PKS de AGPI.

Antecedentes de la invención

Los sistemas de policétido sintasa (PKS) se conocen generalmente en la técnica como complejos enzimáticos relacionados con los sistemas de ácido graso sintasa (FAS), pero que a menudo se modifican enormemente para producir productos especializados que muestran normalmente poca semejanza con los ácidos grasos. Sin embargo, se ha mostrado ahora que los sistemas similares a PKS, también denominados en el presente documento de manera intercambiable sistemas de PKS de AGPI, sistemas de AGPI sintasa o sistemas de PKS para la producción de AGPI, existen en bacterias marinas y determinados organismos eucariotas que pueden sintetizar ácidos grasos poliinsaturados (AGPI) a partir de acetil-CoA y malonil-CoA. Las rutas de PKS de AGPI para la síntesis de AGPI en *Shewanella* y otras bacterias marinas, *Vibrio marinus*, se describen en detalle en la patente estadounidense n.º 6.140.486. Las rutas de PKS de AGPI para la síntesis de AGPI en el traustoquitridio eucariota, *Schizochytrium*, se describe en detalle en la patente estadounidense n.º 6.566.583. Las rutas de PKS de AGPI para la síntesis de AGPI en eucariotas tales como miembros de *Thraustochytriales*, incluyendo la descripción adicional de un sistema de PKS de AGPI en *Schizochytrium* y la identificación de un sistema de PKS de AGPI en *Thraustochytrium*, incluyendo detalles referentes a usos de estos sistemas, se describen en detalle en la publicación de solicitud de patente estadounidense n.º 20020194641, publicada el 19 de diciembre de 2002, y la publicación de solicitud de patente estadounidense n.º 20070089199, publicada el 19 de abril de 2007. La publicación de solicitud de patente estadounidense n.º 20040235127, publicada el 25 de noviembre de 2004, da a conocer la descripción estructural detallada de un sistema de PKS de AGPI en *Thraustochytrium*, y detalle adicional referente a la producción de ácido eicosapentaenoico (C20:5, ω -3) (EPA) y otros AGPI usando tales sistemas. La publicación de solicitud de patente estadounidense n.º 20050100995, publicada el 12 de mayo de 2005, da a conocer la descripción estructural y funcional de sistemas de PKS de AGPI en *Shewanella olleyana* y *Shewanella japonica*, y usos de tales sistemas. Estas solicitudes también dan a conocer la modificación genética de organismos, incluyendo microorganismos y plantas, con los genes que comprenden la ruta de PKS de AGPI y la producción de AGPI por tales organismos. Además, la publicación de patente PCT n.º WO 05/097982 describe un sistema de PKS de AGPI en *Ulkenia*, y la publicación de solicitud de patente estadounidense n.º 20050014231 describe genes y proteínas PKS de AGPI de *Thraustochytrium aureum*.

Se describen sistemas de PKS de AGPI adicionales y usos de los mismos en los documentos WO 02/083870, WO 2004/087879, WO 2006/044646 y WO 2006/135866..

Los investigadores han intentado aprovechar los sistemas de policétido sintasa (PKS) que se han descrito tradicionalmente en la bibliografía que se encuentran en uno de tres tipos básicos, denominados normalmente: tipo I (modular o iterativo), tipo II y tipo III. Con fines de mayor claridad, se observa que el sistema de PKS de tipo I modular anteriormente también se ha denominado simplemente sistema de PKS "modular", y el sistema de PKS de tipo I iterativo anteriormente también se ha denominado simplemente sistema de PKS "tipo I". El sistema de tipo II se caracteriza por proteínas separables, cada una de las cuales lleva a cabo una reacción enzimática distinta. Las enzimas actúan conjuntamente para producir el producto final y cada enzima individual del sistema participa normalmente varias veces en la producción del producto final. Este tipo de sistema funciona de manera análoga a los sistemas de ácido graso sintasa (FAS) hallados en plantas y bacterias. Los sistemas de PKS de tipo I iterativo son similares al sistema tipo II en que las enzimas se usan de modo iterativo para producir el producto final. El tipo I iterativo difiere del tipo II en que las actividades enzimáticas, en vez de asociarse con proteínas separables, se producen como dominios de proteínas más grandes. Este sistema es análogo a los sistemas de FAS de tipo I hallados en animales y hongos.

A diferencia de los sistemas tipo II, en los sistemas de PKS de tipo I modular, cada dominio enzimático se usa sólo 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 los sistemas de PKS descritos anteriormente, si se incorpora un doble enlace carbono-carbono en el producto final, está habitualmente en la configuración *trans*.

Se han descubierto más recientemente sistemas de tipo III y pertenecen a la familia de enzimas de condensación de la chalcona sintasa de plantas. Los PKS de tipo III son distintos de los sistemas de PKS de tipo I y tipo II y utilizan sustratos de acil-CoA libre en reacciones de condensación iterativas para producir habitualmente un producto final

heterocíclico.

Se considera que los ácidos grasos poliinsaturados (AGPI) son útiles para fines nutricionales, farmacéuticos, industriales, y otros. El suministro actual de AGPI a partir de fuentes naturales y a partir de síntesis química no es suficiente para las necesidades comerciales. Una fuente actual importante para AGPI es a partir de peces marinos; sin embargo, las reservas pesqueras están disminuyendo, y esto puede no ser un recurso sostenible. Adicionalmente, la contaminación, tanto por metales pesados como por moléculas orgánicas tóxicas, es un grave problema del aceite derivado de peces marinos. Los aceites vegetales derivados de cultivos de semillas oleaginosas son relativamente baratos y no tienen los problemas de contaminación asociados con los aceites de pescado. Sin embargo, los AGPI hallados en aceites de plantas desarrolladas comercialmente se limitan normalmente a ácido linoleico (dieciocho carbonos con 2 dobles enlaces, en las posiciones delta 9 y 12 - 18: 2 delta 9,12) y ácido linolénico (18:3 delta 9,12,15). En la ruta convencional (es decir, la ruta "habitual" o ruta "clásica") para la síntesis de AGPI, ácidos grasos saturados de longitud de cadena media (productos de un sistema de ácido graso sintasa (FAS)) se modifican mediante una serie de reacciones de elongación y desaturación. Los sustratos para la reacción de elongación son acil-CoA graso (la cadena de ácido graso que va a elongarse) y malonil-CoA (la fuente de los 2 carbonos añadidos durante cada reacción de elongación). El producto de la reacción con elongasa es un acil-CoA graso que tiene dos carbonos adicionales en la cadena lineal. Las desaturasas crean dobles enlaces *cis* en la cadena de ácido graso preexistente mediante la extracción de 2 hidrógenos en una reacción dependiente de oxígeno. Los sustratos para las desaturasas son o bien acil-CoA (en algunos animales) o bien el ácido graso que se esterifica para dar la estructura principal de glicerol de un fosfolípido (por ejemplo, fosfatidilcolina).

Por tanto, debido a que se requieren varias enzimas desaturasas y elongasas para la síntesis de ácidos grasos a partir de ácidos linoleico y linolénico para producir AGPI más insaturados y de cadena más larga, células huésped de plantas modificadas mediante ingeniería genética para la expresión de AGPI tales como EPA y ácido docosahexaenoico (DHA) pueden requerir la expresión de varias enzimas independientes para lograr la síntesis. Adicionalmente, para la producción de cantidades utilizables de tales AGPI, pueden requerirse esfuerzos de modificación mediante ingeniería genética adicionales. Por tanto, resulta de interés obtener material genético implicado en la biosíntesis de AGPI a partir de especies que producen de manera natural estos ácidos grasos (por ejemplo, a partir de un sistema de PKS de AGPI) y expresar el material aislado solo o en combinación en un sistema heterólogo que puede manipularse para permitir la producción de cantidades comerciales de AGPI.

Ha habido muchos esfuerzos para producir AGPI en plantas de cultivo de semillas oleaginosas mediante modificación de los ácidos grasos producidos de manera endógenas. La modificación genética de estas plantas con diversos genes individuales para elongasas y desaturasas de ácidos grasos ha producido hojas o semillas que contienen niveles medibles de AGPI tales como EPA, pero que contienen también un nivel significativo de AGPI de cadena más corta y menos saturados mezclados (Qi *et al.*, Nature Biotech. 22:739 (2004); publicación PCT n.º WO 04/071467; Abadi *et al.*, Plant Cell 16:1 (2004)); Napier y Sayanova, Proceedings of the Nutrition Society (2005), 64:387-393; Robert *et al.*, Functional Plant Biology (2005) 32:473-479; o publicación de solicitud de patente estadounidense 2004/0172682.

La mejora tanto en la producción microbiana como de plantas de AGPI es un objetivo comercial altamente deseable. Por tanto, sigue habiendo la necesidad en la técnica de un método para producir eficaz y eficientemente cantidades de lípidos (por ejemplo, triacilglicerol (TAG) y fosfolípido (PL)) enriquecidos en AGPI deseados, particularmente en organismos comercialmente útiles tales como microorganismos y plantas de semillas oleaginosas.

Sumario de la invención

Una realización de la invención se refiere a un sistema de PKS de AGPI quimérico, en el que un dominio deshidrosas-2 (DH2) de β 3-hidroxiacil-ACP similar a FabA de un primer sistema de PKS de AGPI se sustituye por un dominio DH2 de un segundo sistema de PKS de AGPI diferente, para producir un sistema de PKS de AGPI quimérico que produce una razón diferente de AGPI omega-3 con respecto a omega-6 en comparación con el primer sistema de PKS de AGPI. En un aspecto, una proteína que comprende el dominio DH2 del primer sistema de PKS de AGPI se sustituye por una proteína homóloga que comprende el dominio DH2 del segundo sistema de PKS de AGPI. En un aspecto, el dominio DH2 del sistema de PKS de AGPI primero o segundo corresponde a un dominio DH2 de *Schizochytrium* o *Thraustochytrium*. En un aspecto, el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y en el que el segundo sistema de PKS de AGPI es un sistema de PKS de AGPI de *Thraustochytrium*. En un aspecto, el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y en el que el OrfC del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el OrfC de un traustoitridio diferente.

En un aspecto de esta realización, el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y en el que el OrfC del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el OrfC de *Thraustochytrium* 23B. En un aspecto, un OrfC de este tipo de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*. Una secuencia de ácido nucleico a modo de ejemplo comprende SEQ ID NO: 70. En un aspecto adicional, el OrfA del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el OrfA de *Thraustochytrium* 23B. En un aspecto, un OrfA de este tipo de

Thraustochytrium 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*. Una secuencia de ácido nucleico a modo de ejemplo comprende SEQ ID NO: 71. En otro aspecto adicional, el OrfB del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el OrfB de *Thraustochytrium* 23B. En un aspecto, un OrfB de este tipo de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*. Una secuencia de ácido nucleico a modo de ejemplo comprende SEQ ID NO: 72. Otras combinaciones de OrfA, B y C resultarán evidentes basándose en esta descripción para los expertos en la técnica.

Aún en otro aspecto de esta realización, el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y el dominio DH2 del OrfC del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el dominio DH2 de *Thraustochytrium* 23B. En un aspecto, una secuencia de ácido nucleico a modo de ejemplo que comprende el dominio DH2 de *Thraustochytrium* 23B comprende SEQ ID NO: 73. En un aspecto, el dominio DH2 de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*. Una secuencia de ácido nucleico de este tipo que comprende el dominio DH2 de *Thraustochytrium* 23B se ejemplifica mediante la secuencia de ácido nucleico que comprende SEQ ID NO: 75.

Aún en otro aspecto de esta realización, el sistema de PKS de AGPI quimérico comprende una proteína que comprende una secuencia de aminoácidos que es idéntica en al menos el 95% a SEQ ID NO: 74. En un aspecto, el sistema de PKS de AGPI quimérico comprende una proteína que comprende una secuencia de aminoácidos de SEQ ID NO: 74. En un aspecto, el sistema de PKS de AGPI quimérico comprende SEQ ID NO: 2, SEQ ID NO: 4 y SEQ ID NO: 74. En otro aspecto, el sistema de PKS de AGPI quimérico comprende SEQ ID NO: 39, SEQ ID NO: 4 y SEQ ID NO: 62. En otro aspecto, el sistema de PKS de AGPI quimérico comprende SEQ ID NO: 39, SEQ ID NO: 4 y SEQ ID NO: 74. En otro aspecto, el sistema de PKS de AGPI quimérico se codifica por moléculas de ácido nucleico que comprenden: SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 70. Aún en otro aspecto, el sistema de PKS de AGPI quimérico se codifica por moléculas de ácido nucleico que comprenden: SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 73. En otro aspecto, el sistema de PKS de AGPI quimérico se codifica por moléculas de ácido nucleico que comprenden: SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 75. En otro aspecto, el sistema de PKS de AGPI quimérico se codifica por moléculas de ácido nucleico que comprenden: SEQ ID NO: 71, SEQ ID NO: 3 y SEQ ID NO: 70.

Otra realización de la invención se refiere a un método de alteración de la razón de omega-3 con respecto a omega-6 de ácidos grasos poliinsaturados (AGPI) producidos por un primer sistema de PKS de AGPI, que comprende expresar cualquiera de los sistemas de PKS de AGPI quiméricos descritos anteriormente en un organismo. En un aspecto, el sistema de PKS de AGPI quimérico se expresa por un microorganismo. En un aspecto, el microorganismo es un *Schizochytrium*. Aún en otro aspecto, el microorganismo es una levadura. En un aspecto, el sistema de PKS de AGPI quimérico se expresa por una planta.

Aún otra realización de la invención se refiere a un microorganismo o una planta o una parte de la planta modificado genéticamente, que comprende cualquiera de los sistemas de PKS de AGPI quiméricos descritos anteriormente.

Otra realización de la invención se refiere a un método de aumento de la producción de AGPI y de alteración de la razón de omega-3 con respecto a omega-6 de ácidos grasos poliinsaturados (AGPI) producidos por un primer sistema de PKS de AGPI. El método comprende expresar un sistema de PKS de AGPI quimérico en un organismo, en el que el dominio deshidrasa-2 (DH2) de β -hidroxilacil-ACP similar a FabA de un primer sistema de PKS de AGPI se sustituye por un dominio DH2 de un segundo sistema de PKS de AGPI diferente, para producir un sistema de PKS de AGPI quimérico que produce una razón diferente de AGPI omega-3 con respecto a omega-6 en comparación con el primer sistema de PKS de AGPI. El dominio DH2 del segundo sistema de PKS de AGPI está optimizado para el uso de codones del organismo del que se deriva el primer sistema de PKS de AGPI.

También se describe una molécula de ácido nucleico aislada que codifica para una proteína OrfC quimérica que es idéntica en al menos el 95% a SEQ ID NO: 74. En un aspecto, la molécula de ácido nucleico aislada comprende una secuencia de ácido nucleico que es idéntica en al menos el 95% a SEQ ID NO: 73. En un aspecto, la secuencia de ácido nucleico está optimizada para el uso de codones de un organismo en el que va a expresarse la molécula de ácido nucleico. Como ejemplo, la secuencia de ácido nucleico puede optimizarse para el uso de codones de un organismo del que se deriva una porción de la proteína quimérica. La secuencia de ácido nucleico puede ser idéntica en al menos el 95% a SEQ ID NO: 75.

También se describe una molécula de ácido nucleico recombinante que comprende cualquiera de las moléculas de ácido nucleico descritas anteriormente.

Se describe además una célula huésped recombinante que se ha transfectado con cualquiera de las moléculas de ácido nucleico descritas anteriormente. En un aspecto, la célula es un microorganismo. En un aspecto, el microorganismo es un *Schizochytrium*. En un aspecto, el microorganismo es una bacteria. En un aspecto, el microorganismo es una levadura. En un aspecto, la célula es una célula de planta.

También se describe una planta modificada genéticamente o parte de la misma, que comprende cualquiera de las

células huésped recombinantes descritas anteriormente.

También se describe un sistema de PKS de AGPI quimérico, que comprende: (a) al menos un dominio enoil-ACP reductasa (ER); (b) al menos cuatro dominios ACP; (c) al menos dos dominios β -cetoacil-ACP sintasa (KS); (d) al menos un dominio aciltransferasa (AT); (e) al menos un dominio β -cetoacil-ACP reductasa (KR); (f) al menos dos dominios β -hidroxiacil-ACP deshidrasa (DH) similar a FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); y (h) al menos un dominio malonil-CoA:ACP aciltransferasa (MAT). Al menos uno de los dominios DH es de un primer sistema de PKS de AGPI, y el resto de los dominios (a)-(h) son de un segundo sistema de PKS de AGPI diferente.

También se describe un método de aumento de la producción de AGPI por un organismo que expresa un sistema de PKS de AGPI. El método incluye modificar una molécula de ácido nucleico que codifica para al menos una proteína en el sistema de PKS de AGPI para el uso de codones optimizado del organismo o de un organismo relacionado. En un aspecto, el organismo expresa un sistema de PKS de AGPI recombinante, heterólogo. En un aspecto, el organismo es un *Schizochytrium* y una molécula de ácido nucleico que codifica para al menos una proteína en el sistema de PKS de AGPI endógeno está optimizada para el uso de codones de *Schizochytrium*.

Breve descripción de las figuras

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

La figura 2A es un dibujo esquemático que muestra la etapa 1 de la construcción de un plásmido que contiene una secuencia de ácido nucleico optimizada para codones de *Schizochytrium*, sintética que codifica para OrfC de *Thraustochytrium* 23B (pThOrfC_synPS), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 2B es un dibujo esquemático que muestra la etapa 2 de la construcción de un plásmido que contiene una secuencia de ácido nucleico optimizada para codones de *Schizochytrium*, sintética que codifica para OrfC de *Thraustochytrium* 23B (pThOrfC_synPS), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 3A es un dibujo esquemático que muestra las etapas 1-6 de la construcción de un plásmido que codifica para OrfC de *Schizochytrium* que comprende un dominio DH2 nativo de *Thraustochytrium* 23B (pDS49), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 3B es un dibujo esquemático que muestra la etapa 7 de la construcción de un plásmido que codifica para OrfC de *Schizochytrium* que comprende un dominio DH2 nativo de *Thraustochytrium* 23B (pDS49), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 3C es un dibujo esquemático que muestra las etapas 8-9 de la construcción de un plásmido que codifica para OrfC de *Schizochytrium* que comprende un dominio DH2 nativo de *Thraustochytrium* 23B (pDS49), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 4A es un dibujo esquemático que muestra la construcción de plásmido DD21 como la primera etapa en la construcción de un plásmido que codifica para OrfC de *Schizochytrium* que comprende un dominio DH2 optimizado para codones de *Schizochytrium*, sintético de *Thraustochytrium* 23B (pDD24), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 4B es un dibujo esquemático que muestra la construcción de plásmido DD22 como la segunda etapa en la construcción de un plásmido que codifica para OrfC de *Schizochytrium* que comprende un dominio DH2 optimizado para codones de *Schizochytrium*, sintético de *Thraustochytrium* 23B (pDD24), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 4C es un dibujo esquemático que muestra la construcción de plásmido pDD24 como la etapa final en la construcción de un plásmido que codifica para OrfC de *Schizochytrium* que comprende un dominio DH2 optimizado para codones de *Schizochytrium*, sintético de *Thraustochytrium* 23B (pDD24), así como plásmidos intermedios producidos mediante el procedimiento.

La figura 5 es un perfil de EMAG de levadura de control y levadura que expresa los OrfA, los OrfB, OrfC y Het I de *Schizochytrium*.

La figura 6 es el perfil de EMAG para la levadura de la figura 5, expandido para ilustrar la producción de AGPI objetivo.

Descripción detallada de la invención

La presente invención se refiere en general a sistemas de policétido sintasa (PKS) de ácidos grasos poliinsaturados

(AGPI), también conocidos como sistemas de AGPI sintasa, incluyendo sistemas de PKS de AGPI de traustochytridios (por ejemplo, *Schizochytrium* y *Thraustochytrium*), laberintúlidos, bacterias marinas, y otros organismos que contienen PKS de AGPI, y proteínas PKS de AGPI quiméricas y sistemas producidos a partir de las mismas. La presente invención se refiere a organismos modificados genéticamente que comprenden tales sistemas de PKS de AGPI, y a métodos de preparación y uso de tales sistemas para la producción de productos de interés, incluyendo moléculas bioactivas. En una realización preferida, la presente invención se refiere a un método para producir AGPI en un microorganismo o en una planta de semilla oleaginosa o parte de la planta que se ha modificado genéticamente para expresar un sistema de PKS de AGPI de la presente invención. Los aceites producidos por el microorganismo o la planta contienen al menos un AGPI producido por el sistema de PKS de AGPI, y en el caso de la planta, están sustancialmente libres de AGPI de cadena más corta y menos insaturados mezclados que son productos de ácidos grasos producidos mediante la modificación de productos de los sistemas de FAS. La presente invención incluye específicamente métodos para modificar la cantidad de AGPI y la razón de AGPI producidos por un sistema de PKS de AGPI, y en un aspecto de la invención, la razón de AGPI omega-3 con respecto a omega-6 o la razón de un AGPI con respecto a otro(s) AGPI (por ejemplo, la razón de DHA con respecto a EPA), que puede aplicarse a la creación y el uso de cualquier constructo de PKS de AGPI y/u organismo modificado genéticamente, tal como se ejemplifica y describe en detalle en el presente documento.

En primer lugar, los presentes inventores describen en el presente documento un dominio de un sistema de PKS de AGPI que es tanto necesario como suficiente para modificar la razón de AGPI que se producen por un sistema de PKS de AGPI cuando se producen más de un AGPI, y proporcionan constructos quiméricos novedosos, sistemas de PKS de AGPI quiméricos novedosos, organismos novedosos, y métodos novedosos para producir cantidades modificadas de AGPI usando este descubrimiento. En segundo lugar, los presentes inventores describen en el presente documento métodos, modificaciones, y una variedad de sistemas de PKS de AGPI quiméricos y constructos para optimizar la expresión de PKS de AGPI en huéspedes heterólogos (o en un huésped endógeno) para aumentar la producción de AGPI por el organismo. La invención incluye una descripción detallada del uso de estos dos descubrimientos, solos o en conjunto para potenciar y dirigir la producción de AGPI en un organismo.

Más particularmente, con respecto a determinadas realizaciones de la invención, trabajo previo de los presentes inventores y colaboradores (véase el ejemplo 8 en la publicación de solicitud de patente estadounidense n.º 20050100995) demostró que la región codificante de orfC de *Thraustochytrium* 23B (representada en el presente documento por SEQ ID NO: 62) podría sustituir funcionalmente la región codificante de orfC de *Schizochytrium* en el locus de OrfC en el genoma. Esto se determinó creando en primer lugar una delección exacta de la región codificante de orfC de *Schizochytrium* orfC que contenía un casete de resistencia a antibióticos en su lugar (indicado como Δ orfC::ZEO) dando como resultado una cepa (indicada como B32-Z1) con un requisito de crecimiento obligado para DHA y resistencia a Zeocin™. Entonces se construyó un plásmido en el que se clonó exactamente la región codificante de orfC de *Th.23B* entre regiones no codificantes aguas arriba y aguas abajo del orfC de *Schizochytrium*. La transformación de la cepa de Δ orfC::ZEO de *Schizochytrium* con este constructo de orfC de *Th.23B* dio como resultado la complementación de los transformantes de delección y prototróficos (que no requieren DHA), sensibles a Zeocin. Se determinó que estos transformantes derivados de acontecimientos de recombinación de doble cruzamiento en el locus de OrfC de manera que la región codificante de orfC de *Th.23B* se había sustituido exactamente por la de *Schizochytrium*; es decir, sustitución génica. El análisis del contenido en ácidos grasos de estos transformantes mostró que la razón DHA/DPA se había cambiado de aproximadamente 2,3 (en *Schizochytrium* ATCC20888 de tipo natural) a aproximadamente 8,3 (aproximadamente la de *Th.23B*). Este resultado indicó que el gen orfC (que contiene tres dominios, DH1, DH2 y ER, en *Schizochytrium* y *Thraustochytrium*) desempeña un papel importante en la determinación de la razón n-3/n-6 (omega-3/omega-6) de productos de AGPI. Sin embargo, la producción de AGPI totales en la cepa que contiene de orfC *Th.23B*, aunque significativa, fue menor que la del huésped de *Schizochytrium* de tipo natural (aproximadamente el 60%).

El examen de estas dos regiones codificantes de orfC condujo a los inventores a considerar que el gen de *Th.23B* se expresa escasamente en *Schizochytrium* debido a patrones de uso de codones notablemente diferentes entre *Schizochytrium* y *Thraustochytrium*. Los inventores han descubierto ahora que usando una región codificante de orfC de *Th.23B* "sintética" (es decir, una región codificante producida de manera sintética) con uso de codones optimizado para el patrón de *Schizochytrium*, se potenció la producción de DHA, mientras que se mantuvo la razón n-3/n-6 aumentada observada con el orfC de *Th.23B* no sintético (véanse los ejemplos 1 y 4).

Los inventores también han descrito previamente la existencia de dominios identificables dentro de la proteína OrfC para *Schizochytrium* y *Thraustochytrium*: deshidratasa 1 (DH1), deshidratasa 2 (DH2), y enoil reductasa (ER) (por ejemplo, véase la publicación de solicitud de patente estadounidense n.º 20020194641, citada anteriormente; la publicación de solicitud de patente estadounidense n.º 20040235127, citada anteriormente), y han enseñado que se cree que uno o más de los dominios en OrfC está implicado en el control del tipo y/o la razón de ácidos grasos producidos por el sistema de PKS de AGPI. En este documento, los inventores demuestran en sistemas de *Schizochytrium*, *E. coli* y levadura que el dominio DH2 solo es responsable de la mayor parte o la totalidad del efecto del sistema de PKS de AGPI sobre la razón de ácidos grasos omega-3 con respecto a omega-6 (n-3/n-6). En particular, los inventores realizaron en primer lugar experimentos en los que se usaron diversos dominios de OrfC de *Thraustochytrium* 23B para sustituir los dominios correspondientes en el OrfC de *Schizochytrium* (datos no mostrados). Los inventores encontraron que la sustitución del dominio ER de OrfC de *Schizochytrium* por el de

Thraustochytrium no cambió significativamente la razón DHA/DPA en comparación con *Schizochytrium* de tipo natural (históricamente de aproximadamente 2,3). Sin embargo, la sustitución de ambos dominios DH de *Schizochytrium* por los dominios correspondientes de *Thraustochytrium* aumentó significativamente la razón DHA/DPA hacia la de *Thraustochytrium* 23B de tipo natural (históricamente de aproximadamente 8,3-10), y la sustitución de sólo el dominio DH2 de *Schizochytrium* por el de *Thraustochytrium* 23B, fue suficiente para lograr eficazmente el mismo resultado. Los ejemplos 2, 3, 4, 5, y 6 proporcionan una variedad de resultados experimentales que demuestran el efecto del dominio DH2 sobre la razón de ácidos grasos omega-3 con respecto a omega-6 (n-3/n-6) en sistemas de PKS de AGPI.

Los presentes inventores también describen el uso de una variedad de sistemas de PKS de AGPI quiméricos para aumentar la producción de AGPI por el organismo huésped, y han realizado el descubrimiento inesperado de que determinadas combinaciones de PKS de AGPI quiméricos (por ejemplo, sistemas de PKS de AGPI quiméricos compuestos por combinaciones particulares de Orf de *Schizochytrium* y *Thraustochytrium*) tienen una producción de AGPI significativamente mayor, y en un ejemplo, de producción de DHA, que los organismos nativos o que otros sistemas de PKS de AGPI quiméricos. Por ejemplo, los inventores demuestran que un sistema de PKS de AGPI quimérico compuesto por un OrfA y OrfC de *Thraustochytrium* 23B y un OrfB de *Schizochytrium*, cuando se expresan en un organismo huésped de *Schizochytrium*, produce significativamente más ácidos grasos y significativamente más DHA específicamente, que *Schizochytrium* nativo o que otros sistemas de PKS de AGPI quiméricos derivados de estos dos organismos (ejemplo 8). Por consiguiente, la invención proporciona orientación sustancial sobre la producción de varios sistemas de PKS de AGPI diferentes que tienen una producción aumentada de AGPI y razones n-3/n-6 mejoradas, en comparación con algunas AGPI sintetas de tipo natural (no quiméricas).

Tal como se usa en el presente documento, un sistema de PKS de AGPI (que también puede denominarse un sistema de AGPI sintasa, una AGPI sintasa, o un sistema similar a PKS para la producción de AGPI) generalmente tiene las siguientes características de identificación: (1) produce AGPI, y particularmente, AGPI de cadena larga, como producto natural del sistema; y (2) comprende varias proteínas multifuncionales ensambladas en un complejo que realiza tanto procesamiento iterativo de la cadena de ácido graso así como procesamiento no iterativo, incluyendo isomerización *trans-cis* y reacciones de reducción de enoílo en ciclos seleccionados. Además, los dominios ACP presentes en las enzimas AGPI sintetas requieren activación mediante unión de un cofactor (4-fosfopanteteína). La unión de este cofactor se lleva a cabo por las fosfopanteteinil transferasas (PPTasa). Si las PPTasas endógenas del organismo huésped no pueden activar los dominios ACP de AGPI sintetas, entonces es necesario proporcionar una PPTasa que pueda llevar a cabo esa función. Los inventores han identificado la enzima Het I de *Nostoc* sp. como PPTasa a modo de ejemplo y adecuada para activar dominios ACP de AGPI sintetas. La referencia a un sistema de PKS de AGPI o una AGPI sintasa se refiere colectivamente a todos los genes y sus productos codificados que actúan en un complejo para producir AGPI en un organismo. Por tanto, el sistema de PKS de AGPI se refiere específicamente a un sistema de PKS para el que los productos naturales son AGPI.

Más específicamente, un sistema de PKS de AGPI tal como se hace referencia en el presente documento produce ácidos grasos poliinsaturados (AGPI) y particularmente, AGPI de cadena larga, como productos. Por ejemplo, un organismo que contiene de manera endógena (de manera natural) un sistema de PKS de AGPI produce AGPI usando este sistema. Según la presente invención, AGPI son ácidos grasos con una longitud de la cadena carbonada 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, en los que todos los dobles enlaces están en la configuración *cis*. La referencia a ácidos grasos poliinsaturados de cadena larga (AGPICL) en el presente documento se refiere más particularmente a ácidos grasos de longitud de la cadena carbonada de 18 y más, y preferiblemente longitud de la cadena carbonada de 20 y más, que contienen 3 o más dobles enlaces. Los AGPICL de la serie de omega-6 incluyen: ácido gamma-linolénico (C18:3), ácido di-homo-gamma-linolénico (C20:3n-6), ácido araquidónico (C20:4n-6), ácido adrenico (también denominado ácido docosatetraenoico o DTA) (C22:4n-6), y ácido docosapentaenoico (C22:5n-6). Los AGPICL de la serie de omega-3 incluyen: ácido alfa-linolénico (C18:3), ácido eicosatrienoico (C20:3n-3), ácido eicosatetraenoico (C20:4n-3), ácido eicosapentaenoico (C20:5n-3), ácido docosapentaenoico (C22:5n-3) y ácido docosahexaenoico (C22:6n-3). Los AGPICL también incluyen ácidos grasos con más de 22 carbonos y 4 o más dobles enlaces incluyendo pero sin limitarse a C28:8(n-3).

En segundo lugar, un sistema de PKS de AGPI según la presente invención comprende varias proteínas multifuncionales (y puede incluir proteínas de una única función, particularmente para sistemas de PKS de AGPI de bacterias marinas) que se ensamblan en un complejo que realiza tanto procesamiento iterativo de la cadena de ácido graso así como procesamiento no iterativo, incluyendo isomerización *trans-cis* y reacciones de reducción de enoílo en ciclos seleccionados. Estas proteínas también pueden denominarse en el presente documento el complejo enzimático de PKS de AGPI central o el sistema de PKS de AGPI central. Las funciones generales de los dominios y motivos contenidos dentro de estas proteínas se conocen individualmente en la técnica y se han descrito en detalle con respecto a diversos sistemas de PKS de AGPI de bacterias marinas y organismos eucariotas (véase, por ejemplo, la patente estadounidense n.º 6.140.486; patente estadounidense n.º 6.566.583; Metz *et al.*, Science 293:290-293 (2001); la publicación de solicitud de patente estadounidense n.º 20020194641; la publicación de solicitud de patente estadounidense n.º 20040235127; la publicación de solicitud de patente estadounidense n.º

20050100995 y la publicación PCT n.º WO 2006/135866). Los dominios pueden encontrarse como una única proteína (es decir, el dominio y la proteína son sinónimos) o como uno de dos o más (múltiples) dominios en una única proteína, tal como se mencionó anteriormente.

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

Los presentes inventores proponen usar estas características del sistema de PKS de AGPI para producir una gama de moléculas bioactivas que no podían producirse por los sistemas de PKS descritos previamente (de tipo I iterativo o modular, de tipo II o de tipo III). Estas moléculas bioactivas incluyen, pero no se limitan a, ácidos grasos poliinsaturados (AGPI), antibióticos u otros compuestos bioactivos, muchos de los cuales se comentarán más adelante. Por ejemplo, usando el conocimiento de las estructuras génicas de PKS de AGPI descritas en el presente documento, puede usarse cualquiera de varios métodos para alterar los genes de PKS de AGPI, o combinar porciones de estos genes con otros sistemas de síntesis, incluyendo otros sistemas de PKS, de manera que se producen nuevos productos. La capacidad inherente de este tipo de sistema particular para realizar reacciones tanto iterativas como selectivas permitirá que este sistema proporcione productos que no se encontrarían si se aplicaran métodos similares a otros tipos de sistemas de PKS.

Preferiblemente, un sistema de PKS de AGPI de la presente invención comprende al menos los siguientes dominios biológicamente activos que están contenidos normalmente en tres o más proteínas: (a) al menos un dominio enoil-ACP reductasa (ER); (b) múltiples dominio(s) de proteína transportadora de acilo (ACP) (por ejemplo, al menos de uno a cuatro, y preferiblemente al menos cinco dominios ACP, y en algunas realizaciones hasta seis, siete, ocho, nueve, diez, o más de diez dominios ACP); (c) al menos dos dominios β -cetoacil-ACP sintasa (KS); (d) al menos un dominio aciltransferasa (AT); (e) al menos un dominio β -cetoacil-ACP reductasa (KR); (f) al menos dos dominios β -hidroxiacil-ACP deshidrasa (DH) similares a FabA; (g) al menos un dominio de factor de longitud de cadena (CLF); (h) al menos un dominio malonil-CoA:ACP aciltransferasa (MAT). Un sistema de PKS de AGPI según la presente invención también puede comprender al menos una región que contiene un motivo de sitio activo conservado de deshidratasa (DH).

Se describe un sistema de PKS de AGPI de *Schizochytrium* que comprende al menos los siguientes dominios biológicamente activos: (a) dos dominios enoil-ACP reductasa (ER); (b) entre cuatro o cinco y diez o más dominios de proteína transportadora de acilo (ACP), y en un aspecto, nueve dominios ACP; (c) dos dominios β -cetoacil-ACP sintasa (KS); (d) un dominio aciltransferasa (AT); (e) un dominio β -cetoacil-ACP reductasa (KR); (f) dos dominios β -hidroxiacil-ACP deshidrasa (DH) similares a FabA; (g) un dominio de factor de longitud de cadena (CLF); y (h) un dominio malonil-CoA:ACP aciltransferasa (MAT). También se describe un sistema de PKS de AGPI de *Schizochytrium* que también comprende al menos una región o dominio que contiene un motivo de sitio activo conservado de deshidratasa (DH) que no es parte de un dominio DH similar a FabA. Las características estructurales y funcionales de estos dominios generalmente se conocen individualmente en la técnica y se describirán en detalle a continuación con respecto a los sistemas de PKS de AGPI de la presente invención.

Adicionalmente, se describe un sistema de PKS de AGPI de *Thraustochytrium* que comprende al menos los siguientes dominios biológicamente activos: (a) dos dominios enoil-ACP reductasa (ER); (b) entre cuatro o cinco y diez o más dominios de proteína transportadora de acilo (ACP), y en un aspecto, ocho dominios ACP; (c) dos dominios β -cetoacil-ACP sintasa (KS); (d) un dominio aciltransferasa (AT); (e) un dominio β -cetoacil-ACP reductasa (KR); (f) dos dominios β -hidroxiacil-ACP deshidrasa (DH) similares a FabA; (g) un dominio de factor de longitud de cadena (CLF); y (h) un dominio malonil-CoA:ACP aciltransferasa (MAT). También se describe un sistema de PKS de AGPI de *Thraustochytrium* que también comprende al menos una región o dominio que contiene un motivo de sitio activo conservado de deshidratasa (DH) que no es parte de un dominio DH similar a FabA. Las características estructurales y funcionales de estos dominios generalmente se conocen individualmente en la técnica y se describirán en detalle a continuación con respecto a los sistemas de PKS de AGPI de la presente invención.

Un sistema de PKS de AGPI puede incluir adicionalmente una o más proteínas auxiliares, que se definen en el presente documento como proteínas que no se considera que son parte del sistema de PKS de AGPI central tal como se describió anteriormente (es decir, no es parte del propio complejo de la enzima AGPI sintasa), pero que pueden ser, o son, necesarias para la producción de AGPI o al menos para la producción eficiente de AGPI usando el complejo de la enzima AGPI sintasa central de la presente invención, particularmente en determinados organismos huésped (por ejemplo, plantas). Por ejemplo, para producir AGPI, un sistema de PKS de AGPI debe actuar con una proteína auxiliar que transfiere un resto 4'-fosfopanteteinilo de coenzima A al/a los dominio(s) de proteína transportadora de acilo (ACP). Por tanto, un sistema de PKS de AGPI puede considerarse que incluye al menos un dominio 4'-fosfopanteteinil transferasa (PPTasa), o un dominio de este tipo puede considerarse que es un dominio o una proteína auxiliar para el sistema de PKS de AGPI. Cuando se modifican genéticamente organismos (por ejemplo, microorganismos o plantas) para expresar un sistema de PKS de AGPI según la presente invención,

algunos organismos huésped pueden expresar de manera endógena proteínas auxiliares, que son necesarias para actuar con el PKS de AGPI para producir AGPI (por ejemplo, PPTasas). Sin embargo, algunos organismos pueden transformarse con moléculas de ácido nucleico que codifican para una o más proteínas auxiliares descritas en el presente documento para permitir y/o potenciar la producción de AGPI por el organismo, incluso si el organismo produce de manera endógena una proteína auxiliar homóloga (es decir, algunas proteínas auxiliares heterólogas pueden funcionar más eficaz o eficientemente con las proteínas AGPI sintetas transformadas que con la proteína auxiliar endógena de las células huésped). La presente invención y solicitudes anteriores proporcionan ejemplos de bacterias y levaduras que se han modificado genéticamente con el sistema de PKS de AGPI de la presente invención que incluye una PPTasa auxiliar. Se han descrito plantas que se han modificado genéticamente con el sistema de PKS de AGPI que incluye una PPTasa auxiliar (véase, por ejemplo, la publicación de solicitud de patente estadounidense n.º 20070089199). Las características estructurales y funcionales de las PPTasas se describirán en más detalle a continuación.

La ruta "habitual" o "clásica" para la síntesis de AGPI de cadena larga (AGPICL) en organismos eucariotas implica la modificación de ácidos grasos monoinsaturados o saturados de longitud de cadena media (por ejemplo, los productos de los sistemas de FAS descritos anteriormente). Estas modificaciones consisten en etapas de elongación y etapas de desaturación. Los sustratos para la reacción de elongación son acil-CoA graso (la cadena de ácido graso que va a elongarse) y malonil-CoA (la fuente de los dos carbonos añadidos durante cada reacción de elongación). El producto de la reacción con elongasa es un acil-CoA graso que tiene dos carbonos adicionales en la cadena lineal. Normalmente no se producen ácidos grasos libres (AGL) en este ciclo de reacción. Las desaturasas crean dobles enlaces *cis* en la cadena de ácido graso preexistente mediante la extracción de dos hidrógenos en una reacción dependiente de oxígeno. Los sustratos para las desaturasas son o bien acil-CoA (en algunos animales) o bien ácidos grasos que se esterifican en la cadena principal de glicerol de un PL (por ejemplo, fosfatidilcolina). De nuevo, no se producen AGL en este mecanismo de reacción. Por tanto, la única vez que se producen AGL en rutas de síntesis de AGPICL "habituales" o "clásicas" es durante la liberación de los ácidos grasos de algunos sistemas de FAS. Tal como se comentó anteriormente, éstos son normalmente ácidos grasos de 16 ó 18 carbonos y habitualmente son ácidos grasos o bien saturados o bien monoinsaturados, ni AGPI de cadena más larga tales como EPA o DHA. Una consecuencia de este esquema para la producción de AGPI de cadena larga es que a menudo se acumulan productos intermedios en la ruta, que representan a menudo la mayoría de los ácidos grasos novedosos producidos por el sistema.

Por tanto, según la presente invención, la referencia a una ruta "habitual" o "clásica" para la producción de AGPI se refiere a la ruta de síntesis de ácidos grasos en la que ácidos grasos saturados de longitud de cadena media (por ejemplo, productos de un sistema de ácido graso sintasa (FAS)) se modifican mediante una serie de reacciones de elongación y desaturación. Los sustratos para la reacción de elongación son acil-CoA graso (la cadena de ácido graso que va a elongarse) y malonil-CoA (la fuente de los 2 carbonos añadidos durante cada reacción de elongación). El producto de la reacción con elongasa es un acil-CoA graso que tiene dos carbonos adicionales en la cadena lineal. Las desaturasas crean dobles enlaces *cis* en la cadena de ácido graso preexistente mediante la extracción de 2 hidrógenos en una reacción dependiente de oxígeno. Tales rutas y los genes implicados en tales rutas se conocen bien en la bibliografía.

Tal como se usa en el presente documento, el término "lípidos" incluye fosfolípidos (PL); ácidos grasos libres; ésteres de ácidos grasos; triacilglicerol (TAG); diacilglicéridos; monoacilglicéridos; fosfátido; ceras (ésteres de alcoholes y ácidos grasos); esterol y ésteres de esterol; carotenoides; xantófilas (por ejemplo, oxicarotenoides); hidrocarburos; y otros lípidos conocidos por un experto habitual en la técnica. Los términos "ácidos grasos poliinsaturados" y "AGPI" incluyen no sólo la forma de ácido graso libre, sino también otras formas, tales como la forma de TAG y la forma de PL.

La referencia a un organismo "heterólogo" o huésped "heterólogo", con respecto a la expresión de una proteína, un dominio o un sistema de PKS de AGPI por el organismo/huésped, significa que al menos una proteína, un dominio o una porción del sistema de PKS de AGPI no es una proteína, un dominio o una porción que se expresa de manera natural (de manera endógena) por el organismo, aunque el sistema de PKS de AGPI puede incluir proteínas, dominios o porciones de los mismos que se expresan de manera natural por el organismo huésped (por ejemplo, una proteína quimérica tal como se describe en el presente documento que contiene secuencias derivadas del organismo huésped y de un organismo diferente o una proteína diferente).

Determinadas moléculas de ácido nucleico (constructos) a modo de ejemplo que codifican para diversas proteínas quiméricas se describen en el presente documento (véanse los ejemplos). Según la presente invención, una "proteína quimérica" es una proteína modificada mediante ingeniería genética codificada por una secuencia de ácido nucleico que se produce mediante corte y empalme o unión (ligamiento) entre sí de dos o más genes o secuencias de ácido nucleico completos o parciales. Un "sistema de PKS de AGPI quimérico" es un sistema de PKS de AGPI que contiene proteínas y/o dominios, incluyendo proteínas y/o dominios quiméricos, de dos o más sistemas de PKS diferentes. Por ejemplo, los ejemplos describen un sistema de PKS de AGPI quimérico compuesto por los OrfA y OrfB de PKS de AGPI de *Schizochytrium* y el OrfC de PKS de AGPI de *Thraustochytrium*. Los ejemplos también describen un sistema de PKS de AGPI quimérico compuesto por los OrfA, OrfB y todo el OrfC excepto por el dominio DH2 de PKS de AGPI de *Schizochytrium*, que es el dominio DH2 de PKS de AGPI de un PKS de AGPI de

Thraustochytrium. Este último sistema de PKS de AGPI quimérico comprende por consiguiente de una proteína quimérica (una proteína OrfC quimérica). Las mismas quimeras también se describen usando secuencias de ácido nucleico de *Thraustochytrium* que se han optimizado para el uso de codones de *Schizochytrium*, que ilustran una combinación de manipulaciones genéticas que puede usarse para alterar el producto producido por un sistema de PKS de AGPI (véanse los ejemplos). Los ejemplos también describen una variedad de otros sistemas de PKS de AGPI quiméricos.

Tal como se usa en el presente documento, “optimización de codones” o expresiones derivadas de la misma se refieren al procedimiento de modificar (alterar, cambiar, mutar) una secuencia de ácido nucleico que codifica para una proteína dada para sustituir uno o más codones en la secuencia con codones que se usan de la manera más frecuente en secuencias de ácido nucleico de un organismo particular en el que va a expresarse una molécula de ácido nucleico que comprende la secuencia de ácido nucleico. La preferencia codónica y la idea general de optimización de codones se entienden por el experto en la técnica. Más particularmente, el grado en el que aparece un codón dado en el código genético puede variar significativamente entre organismos (por ejemplo, incluyendo una especie a otra dentro de un género). Cualquier codón que un organismo usa un pequeño porcentaje del tiempo, o menos que otro codón para el mismo aminoácido, puede producir problemas en la expresión de proteínas. Por consiguiente, la expresión de proteínas puede mejorar drásticamente cuando la frecuencia de codones de la secuencia de ácido nucleico que está usándose se hace coincidir con la del organismo/sistema de expresión del huésped (por ejemplo, sustituyendo codones raros o poco frecuentes o usados menos frecuentemente por otros que reflejan más estrechamente la preferencia codónica natural del sistema del huésped, sin modificar la secuencia de aminoácidos).

Los presentes inventores describen en el presente documento métodos para optimizar el uso de codones de una secuencia de ácido nucleico para la de *Schizochytrium*, aunque esto es sólo un ejemplo del uso de optimización de codones en la presente invención. Según la presente invención, la secuencia de nucleótidos de una molécula de ácido nucleico que codifica para una proteína dada (por ejemplo, una proteína PKS de AGPI) puede modificarse (por ejemplo, mediante síntesis, mutación, tecnología recombinante, etc.) para el uso óptimo (optimizado) de codones de una célula o un organismo huésped en el que va a expresarse la molécula de ácido nucleico, o en efecto, para el uso de codones optimizado de un organismo diferente (por ejemplo, una molécula de ácido nucleico que codifica para una proteína PKS de AGPI de *Thraustochytrium* para la expresión en una planta puede optimizarse para el uso de codones de *Schizochytrium*). La tabla 1 de los ejemplos ilustra el uso de codones optimizado para *Schizochytrium*.

Además, los inventores proponen en el presente documento la optimización de la secuencia de ácido nucleico de una molécula de ácido nucleico que codifica para una proteína dada para el mismo huésped del que se derivó, aprendió u obtuvo la secuencia de ácido nucleico, para la expresión en ese huésped (o en otro huésped). Esta última realización de la invención representa una evolución “dirigida” o “acelerada” de clases, en las que, por ejemplo, una molécula de ácido nucleico que codifica para una proteína de un organismo (por ejemplo, una proteína PKS de AGPI de *Schizochytrium*) se modifica (por ejemplo, volviendo a sintetizar la secuencia de ácido nucleico y sustituyendo determinados nucleótidos) para potenciar el uso de codones (optimizar el uso de codones) que se prefiere por el mismo organismo (*Schizochytrium*, en este ejemplo). Esta molécula de ácido nucleico puede expresarse entonces en *Schizochytrium* (como molécula de ácido nucleico recombinante) o en otro organismo o célula huésped (por ejemplo, en una planta). En esta realización, se propone que una secuencia de ácido nucleico dada de un organismo puede no usar los codones óptimos (preferencia codónica) que pueden determinarse para ese organismo. Por consiguiente, puede volver a sintetizarse la secuencia de ácido nucleico para mejorar la expresión de proteínas en ese organismo.

Los sistemas de PKS de AGPI y proteínas o dominios de los mismos que son útiles en la presente invención incluyen sistemas de PKS de AGPI tanto bacterianos como no bacterianos. Un sistema de PKS de AGPI no bacteriano es un sistema de PKS de AGPI que es de o se deriva de un organismo que no es una bacteria, tal como un eucariota o una arqueobacteria. Los eucariotas se separan de los procariotas basándose en el grado de diferenciación de las células, estando los eucariotas más diferenciados que los procariotas. En general, los procariotas no presentan una membrana nuclear, no muestran mitosis durante la división celular, sólo tienen un cromosoma, contienen ribosomas 70S en su citoplasma, no presentan mitocondrias, retículo endoplasmático, cloroplastos, lisosomas ni aparato de Golgi, y pueden tener flagelos, que si están presentes, contienen una única fibrilla. En cambio, los eucariotas tienen una membrana nuclear, presentan mitosis durante la división celular, tienen muchos cromosomas, contienen ribosomas 80S en su citoplasma, presentan mitocondrias, retículo endoplasmático, cloroplastos (en algas), lisosomas y aparato de Golgi, y pueden tener flagelos, que si están presentes, contienen muchas fibrillas. En general, las bacterias son los procariotas, mientras que las algas, hongos, protistas, protozoos y plantas superiores son eucariotas. Según la presente invención, pueden producirse organismos modificados genéticamente que incorporan dominios funcionales PKS de AGPI no bacterianos con dominios funcionales PKS de AGPI bacterianos, así como proteínas o dominios funcionales PKS de otros sistemas de PKS (de tipo I iterativo o modular, de tipo II o de tipo III) o sistemas de FAS.

Según la presente invención, un dominio o una proteína que tiene actividad biológica (función) 3-cetoacil-ACP sintasa (KS) se caracteriza como la enzima que lleva a cabo la etapa inicial del ciclo de reacción de elongación de

FAS (y PKS). El término “ β -cetoacil-ACP sintasa” puede usarse de manera intercambiable con los términos “3-cetoacil-ACP sintasa”, “ β -ceto acil-ACP sintasa” y “ceto-acil ACP sintasa”, y derivados similares. El grupo acilo destinado para la elongación se une a un residuo de cisteína en el sitio activo de la enzima mediante un enlace tioéster. En la reacción de múltiples etapas, la acil-enzima experimenta condensación con malonil-ACP para formar -cetoacil-ACP, CO₂ y enzima libre. La KS desempeña un papel clave en el ciclo de elongación y en muchos sistemas se ha mostrado que presenta mayor especificidad de sustrato que otras enzimas del ciclo de reacción. Por ejemplo, *E. coli* tiene tres enzimas KS distintas, cada una con su papel particular en la fisiología del organismo (Magnuson *et al.*, Microbiol. Rev. 57, 522 (1993)). Los dos dominios KS de los sistemas de PKS de AGPI descritos en bacterias marinas y los traustochytridios descritos en el presente documento pueden tener distintos papeles en la secuencia de reacción biosintética de AGPI. Como una clase de enzimas, las KS se han caracterizado bien. Las secuencias de muchos genes de KS verificados se conocen, se han identificado los motivos de sitios activos y se han determinado las estructuras cristalinas de varias. Pueden identificarse fácilmente proteínas (o dominios de proteínas) como pertenecientes a la familia de enzimas de KS mediante homología con secuencias de KS conocidas.

Según la presente invención, un dominio o una proteína que tiene actividad biológica (función) malonil-CoA:ACP aciltransferasa (MAT) se caracteriza como uno que transfiere el resto malonilo de malonil-CoA a ACP. El término “malonil-CoA:ACP aciltransferasa” puede usarse de manera intercambiable con “malonil aciltransferasa” y derivados similares. Además del motivo de sitio activo (GxSxG, SEQ ID NO: 96), estas enzimas presentan un motivo extendido de aminoácidos R y Q en posiciones clave que las identifica como enzimas MAT (por ejemplo, en contraposición a un dominio AT descrito a continuación). En algunos sistemas de PKS (pero no el dominio PKS de AGPI), los dominios MAT cargan preferentemente metil- o etil-malonato en el grupo ACP (del éster de CoA correspondiente), introduciendo de ese modo ramificaciones en la cadena de carbono lineal. Pueden reconocerse dominios MAT por su homología con secuencias MAT conocidas y por su estructura de motivos extendida.

Según la presente invención, un dominio o una proteína que tiene actividad biológica (función) de proteína transportadora de acilo (ACP) se caracteriza como polipéptidos pequeños (normalmente, de 80 a 100 aminoácidos de longitud), que funcionan como portadores para cadenas de acilo grasas en crecimiento por medio de una unión tioéster a un cofactor de la proteína unido covalentemente. Se producen como unidades separadas o como dominios dentro de proteínas más grandes. Las ACP se convierten de apoformas inactivas a holoformas funcionales mediante transferencia del resto fosfopanteteinilo de CoA a un residuo de serina altamente conservado de la ACP. Los grupos acilo se unen a la ACP por medio de una unión tioéster en el extremo terminal libre del resto fosfopanteteinilo. Las ACP pueden identificarse mediante marcaje con panteteína radiactiva y mediante homología de secuencia con ACP conocidas. La presencia de variaciones del motivo mencionado anteriormente (LGIDS*) es también un rasgo distintivo de una ACP.

Según la presente invención, un dominio o una proteína que tiene actividad cetorreductasa, también denominada actividad biológica (función) 3-cetoacil-ACP reductasa (KR), se caracteriza como uno que cataliza la reducción dependiente de nucleótidos de piridina de formas de 3-cetoacilo de ACP. Es la primera etapa reductora en el ciclo de elongación de la biosíntesis de ácidos grasos *de novo* y una reacción realizada a menudo en biosíntesis de policétidos. El término “ β -cetoacil-ACP reductasa” puede usarse de manera intercambiable con los términos “cetorreductasa”, “3-cetoacil-ACP reductasa”, “ceto-acil ACP reductasa” y derivados similares del término. Se observa similitud de secuencia significativa con una familia de enoil ACP reductasas (ER), la otra reductasa de FAS (pero no la familia de ER presente en los sistemas de PKS de AGPI), y la familia de alcohol deshidrogenasa de cadena corta. El análisis Pfam de la región de PKS de AGPI indicada anteriormente revela la homología con la familia de alcohol deshidrogenasa de cadena corta 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 de homología extendida con dominios de los otros sistemas caracterizados de PKS de AGPI.

Según la presente invención, un dominio o una proteína se denomina factor de longitud de cadena (CLF) basándose en las siguientes razones. El CLF se describió originalmente como característico de sistemas de PKS de tipo II (enzimas disociadas) y se planteó la hipótesis de que desempeña un papel en la determinación del número de ciclos de elongación, y por tanto la longitud de cadena, del producto final. Las secuencias de aminoácidos de CLF muestran homología con dominios KS (y se piensa que forman heterodímeros con una proteína KS), pero carecen de la cisteína del sitio activo. El papel del CLF en sistemas de PKS ha sido controvertido. Nuevas pruebas (C. Bisang *et al.*, Nature 401, 502 (1999)) sugieren un papel en el cebado (proporcionando el grupo acilo inicial que va a elongarse) de los sistemas de PKS. En este papel, se piensa que el dominio CLF descarboxila malonato (como malonil-ACP), formando así un grupo acetato que puede transferirse al sitio activo de KS. Por tanto, este acetato actúa como molécula de “cebado” que puede experimentar la reacción de elongación inicial (condensación). Se han identificado homólogos del CLF de tipo II como dominios de “carga” en algunos sistemas de PKS modulares. Se encuentra un dominio con las características de secuencia del CLF en todos los sistemas de PKS de AGPI identificados actualmente y en cada caso se encuentra como parte de una proteína de múltiples dominios.

Una “aciltransferasa” o “AT” se refiere a una clase general de enzimas que pueden llevar a cabo varias reacciones de transferencia de acilo distintas. El término “aciltransferasa” puede usarse de manera intercambiable con el término “acil transferasa”. Los dominios AT identificados en los sistemas de PKS de AGPI descritos en el presente documento muestran buena homología entre sí y con dominios presentes en todos los otros sistemas de PKS de

AGPI examinados actualmente y homología muy débil con algunas aciltransferasas cuyas funciones específicas se han identificado (por ejemplo con malonil-CoA:ACP aciltransferasa, MAT). A pesar de la débil homología con MAT, no se cree que este dominio AT funcione como una MAT porque no presenta una estructura de motivos extendida característica de tales enzimas (véase la descripción de dominios MAT, anteriormente). Para los fines de esta descripción, las posibles funciones del dominio AT en un sistema de PKS de AGPI incluyen, pero no se limitan a: transferencia del grupo acilo graso del/de los dominio(s) ORFA ACP a agua (es decir, una tioesterasa – liberando 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 lisofosfádico).

Según la presente invención, este dominio tiene actividad biológica enoil reductasa (ER). La enzima ER reduce el doble enlace trans (introducido por la actividad DH) en el acil-ACP graso, dando como resultado la saturación completa de esos carbonos. El dominio ER dominio en la AGPI-PKS muestra homología con 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*. Todos los sistemas de PKS de AGPI examinados actualmente contienen al menos un dominio con homología de secuencia muy alta con el dominio ER de *Schizochytrium*, que muestra homología con la proteína ER de *S. pneumoniae*.

Según la presente invención, una proteína o dominio que tiene actividad deshidrasa o deshidratasa (DH) cataliza una reacción de deshidratación. Tal como se usa generalmente en el presente documento, la referencia a actividad DH se refiere normalmente a actividad biológica β -hidroxiacil-ACP deshidrasa (DH) similar a FabA. La actividad β -hidroxiacil-ACP deshidrasa (DH) similar a FabA elimina HOH de una β -cetoacil-ACP y produce inicialmente un doble enlace trans en la cadena de carbono. El término " β -hidroxiacil-ACP deshidrasa similar a FabA" puede usarse de manera intercambiable con los términos " β -hidroxil acil-ACP deshidrasa similar a FabA", " β -hidroxiacil-ACP deshidrasa", "deshidrasa" y derivados similares. Los dominios DH de los sistemas de PKS de AGPI muestran homología con enzimas DH bacterianas asociadas con sus sistemas de FAS (en vez de con los dominios DH de otros sistemas de PKS). Un subconjunto de DH bacterianas, las DH similares a FabA, presentan actividad isomerasa cis-trans (Heath *et al.*, J. Biol. Chem., 271, 27795 (1996)). Es la homología con las proteínas DH similares a FabA la que sugiere que uno o todos los dominios DH descritos en el presente documento son responsables de la inserción de los dobles enlaces cis en los productos de PKS de AGPI.

Una proteína PKS de AGPI útil de la invención puede tener también actividad deshidratasa que no se caracteriza como similar a FabA (por ejemplo, la actividad cis-trans descrita anteriormente está asociada con actividad similar a FabA), denominada generalmente en el presente documento actividad no similar a FabA, o actividad biológica β -hidroxiacil-ACP deshidrasa (DH) no similar a FabA. Más específicamente, se encuentra un motivo de sitio activo conservado (~13 aminoácidos de longitud: L*xxHxxxGxxxxP, SEQ ID NO: 97; *en el motivo, L también puede ser I) en dominios deshidratasa en sistemas de PKS (Donadio S, Katz L. Gene. 1 de febrero de 1992; 11 1(1):51-60). Este motivo conservado, denominado también en el presente documento motivo de sitio activo conservado de deshidratasa (DH) o motivo de DH, se encuentra en una región similar de todas las secuencias de AGPI-PKS conocidas descritas hasta la fecha y en las secuencias de PKS de AGPI descritas en el presente documento, pero se cree que este motivo se ha detectado sólo recientemente. Este motivo conservado está dentro de una región no caracterizada de alta homología en la secuencia de AGPI-PKS. La biosíntesis propuesta de AGPI por medio de AGPI-PKS requiere una deshidratación no similar a FabA, y este motivo puede estar asociado con esa reacción.

Para fines de ilustración, se describe en detalle a continuación la estructura de determinados sistemas de PKS de AGPI. Sin embargo, debe entenderse que esta invención no se limita al uso de estos sistemas de PKS de AGPI. Por ejemplo, puede encontrarse una descripción detallada de sistemas bacterianos de PKS de AGPI en la patente estadounidense n.º 6.140.486 y la publicación de solicitud de patente estadounidense n.º 20050100995, y se encuentra una descripción de otros sistemas o genes de PKS de AGPI en la publicación de patente PCT n.º WO 05/097982 y la publicación de solicitud de patente estadounidense n.º 20050014231.

Sistema de PKS de AGPI de *Schizochytrium*

Schizochytrium es un microorganismo marino traustequitridio que acumula grandes cantidades de triacilglicerol ricos en DHA y ácido docosapentaenoico (DPA; 22:5 ω -6); por ejemplo, el 30% de DHA + DPA en peso seco (Barclay *et al.*, J. Appl. Phycol. 6, 123 (1994)). En eucariotas que sintetizan AGPI de 20 y 22 carbonos mediante una ruta de elongación/desaturación, las reservas de productos intermedios de 18, 20 y 22 carbonos son relativamente grandes de modo que experimentos de marcaje *in vivo* usando [^{14} C]-acetato revelan una cinética de producto precursor clara para los productos intermedios predichos (Gellerman *et al.*, Biochim. Biophys. Acta 573:23 (1979)). Además, productos intermedios radiomarcados proporcionados de manera exógena a tales organismos se convierten en los productos de AGPI finales. Los presentes inventores han mostrado que se captaba rápidamente [^{14} C]-acetato por células de *Schizochytrium* y se incorporaban en ácidos grasos, pero al tiempo de marcaje más corto (1 min), DHA contenía el 31% del marcador recuperado en ácidos grasos, y este porcentaje permanecía esencialmente sin cambios durante los 10-15 min de incorporación de [^{14} C]-acetato y las 24 horas posteriores de

crecimiento del cultivo (véase la publicación de solicitud de patente estadounidense n.º 20020194641, citada anteriormente). De manera similar, DPA representaba el 10% del marcador a lo largo de todo el experimento. No hay ninguna prueba de una relación precursor-producto entre ácidos grasos de 16 ó 18 carbonos y los ácidos grasos poliinsaturados de 22 carbonos. Estos resultados concuerdan con una síntesis rápida de DHA a partir de [¹⁴C]-acetato que implica reservas muy pequeñas (posiblemente unidas a enzima) de productos intermedios.

La figura 1 es una representación gráfica de los tres marcos de lectura abiertos del sistema de PKS de AGPI de *Schizochytrium*, e incluye la estructura de dominios de este sistema de PKS de AGPI. Hay tres marcos de lectura abiertos que forman el sistema de PKS de AGPI de *Schizochytrium* central. La estructura de dominios de cada marco de lectura abierto es tal como sigue.

Marco de lectura abierto A (OrfA) de Schizochytrium:

La secuencia de nucleótidos completa para OrfA se representa en el presente documento como SEQ ID NO: 1. OrfA es una secuencia de 8730 nucleótidos (sin incluir el codón de terminación) que codifica para una secuencia de 2910 aminoácidos, representada en el presente documento como SEQ ID NO: 2. Dentro de OrfA hay doce dominios: (a) un dominio β-ceto acil-ACP sintasa (KS); (b) un dominio malonil-CoA:ACP aciltransferasa (MAT); (c) nueve dominios de proteína transportadora de acilo (ACP); y (d) un dominio cetereductasa (KR). Se han aislado y secuenciado clones de ADN genómico (plásmidos) que codifican para OrfA de tanto *Schizochytrium* sp. ATCC 20888 como una cepa hija de ATCC 20888, indicada como *Schizochytrium* sp., cepa N230D. N230D era uno de más de 1.000 supervivientes elegidos al azar de *Schizochytrium* ATCC 20888 mutagenizada químicamente (NTG; 1-metil-3-nitro-1-nitrosoguanidina) examinados para detectar variaciones en el contenido en ácidos grasos. Se valoró esta cepa particular por su productividad de DHA mejorada.

Un clon genómico descrito en el presente documento como JK1126, aislado de *Schizochytrium* sp. ATCC 20888, comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos que abarca desde la posición 1 hasta 8730 de SEQ ID NO: 1, y codifica para la secuencia de aminoácidos correspondiente de SEQ ID NO: 2. El clon genómico pJK1126 (indicado como clon genómico pJK1126 OrfA, en forma de un vector de plásmido de *E. coli* que contiene el gen "OrfA" de *Schizochytrium* ATCC 20888) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de junio de 2006 y se le asignó el n.º de registro de la ATCC PTA-7648. La secuencia de nucleótidos del clon genómico pJK1126 OrfA, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

Dos clones genómicos descritos en el presente documento como clon genómico pJK306 OrfA y clon genómico pJK320 OrfA, aislados de *Schizochytrium* sp. N230D, juntos (clones solapantes) comprenden, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 1, y codifican para la secuencia de aminoácidos de SEQ ID NO: 2. El clon genómico pJK306 (indicado como clon genómico pJK306 OrfA, en forma de un plásmido de *E. coli* que contiene la porción en 5' del gen OrfA de *Schizochytrium* sp. N230D (2,2 kB se solapan con pJK320)) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de junio de 2006, y se le asignó el n.º de registro de la ATCC PTA-7641. La secuencia de nucleótidos del clon genómico pJK306 OrfA, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención. El clon genómico pJK320 (indicado como clon genómico pJK320 OrfA, en forma de un plásmido de *E. coli* que contiene la porción en 3' del gen OrfA de *Schizochytrium* sp. N230D (2,2 kB se solapan con pJK306)) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de junio de 2006, y se le asignó el n.º de registro de la ATCC PTA-7644. La secuencia de nucleótidos del clon genómico pJK320 OrfA, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

El primer dominio en OrfA es un dominio KS, también denominado en el presente documento ORFA-KS, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFA-KS se representa en el presente documento como SEQ ID NO: 7 (posiciones 1-1500 de SEQ ID NO: 1). La secuencia de aminoácidos que contiene el dominio ORFA-KS se representa en el presente documento como SEQ ID NO: 8 (posiciones 1-500 de SEQ ID NO: 2). Se observa que el dominio ORFA-KS contiene un motivo de sitio activo: DXAC* (SEQ ID NO: 98; *sitio de unión a acilo C₂₁₅). Además, un motivo característico en el extremo de la región KS de *Schizochytrium*, GFGG (SEQ ID NO: 99), está presente en este dominio en SEQ ID NO: 2 y, por consiguiente, en SEQ ID NO: 8.

El segundo dominio en OrfA es un dominio MAT, también denominado en el presente documento ORFA-MAT, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFA-MAT se representa en el presente documento como SEQ ID NO: 9 (posiciones 1723-3000 de SEQ ID NO: 1). La secuencia de aminoácidos que contiene el dominio ORFA-MAT se representa en el presente documento como SEQ ID NO: 10 (posiciones 575-1000 de SEQ ID NO: 2). El dominio MAT comprende un aspartato en la posición 93 y una histidina en la posición 94 (correspondientes a las posiciones 667 y 668, respectivamente, de SEQ ID NO: 2). Se observa que el dominio ORFA-MAT contiene un motivo de sitio activo: GHS*⁷⁰⁶XG (*sitio de unión a acilo S₇₀₆), representado en el presente documento como SEQ ID NO: 11.

Los dominios 3-11 de OrfA son nueve dominios ACP en tándem, también denominados en el presente documento

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 de desde aproximadamente la posición 3343 hasta aproximadamente la posición 3600 de SEQ ID NO: 1 (OrfA). La secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFA-ACP1 se representa en el presente documento como SEQ ID NO: 12 (posiciones 3343-3600 de SEQ ID NO: 1). La secuencia de aminoácidos que contiene el primer dominio ACP abarca de desde aproximadamente la posición 1115 hasta aproximadamente la posición 1200 de SEQ ID NO: 2. La secuencia de aminoácidos que contiene el dominio ORFA-ACP1 se representa en el presente documento como SEQ ID NO: 13 (posiciones 1115-1200 de SEQ ID NO: 2). Se observa que el dominio ORFA-ACP1 contiene un motivo de sitio activo: LGIDS* (*motivo de unión a panteteína S1157), representada en el presente documento por SEQ ID NO: 14.

Las secuencias de nucleótidos y aminoácidos de los nueve dominios ACP están altamente conservadas y, por tanto, la secuencia para cada dominio no se representa en el presente documento por un identificador de secuencia individual. Sin embargo, basándose en la información dada a conocer en el presente documento, un experto en la técnica puede determinar fácilmente la secuencia que contiene cada uno de los otros ocho dominios ACP. Los nueve dominios ACP juntos abarcan una región de OrfA de desde aproximadamente la posición 3283 hasta aproximadamente posición 6288 de SEQ ID NO: 1, que corresponde a las posiciones de aminoácido de desde aproximadamente 1095 hasta aproximadamente 2096 de SEQ ID NO: 2. La secuencia de nucleótidos para toda la región ACP que contiene los nueve dominios se representa en el presente documento como SEQ ID NO: 16. La región representada por SEQ ID NO: 16 incluye los segmentos de ligador entre dominios ACP individuales. El intervalo de repetición para los nueve dominios es aproximadamente cada 330 nucleótidos de SEQ ID NO: 16 (el número real de aminoácidos medidos entre serinas de sitios activos adyacentes oscila entre 104 y 116 aminoácidos). Cada uno de los nueve dominios ACP contiene un motivo de unión a panteteína LGIDS* (representado en el presente documento por SEQ ID NO: 14), en el que S* es la serina del sitio de unión a panteteína (S). La serina del sitio de unión a panteteína (S) está ubicada cerca del centro de cada secuencia de dominio ACP. En cada extremo de la región de dominio ACP y entre cada dominio ACP está una región que está altamente enriquecida en prolina (P) y alanina (A), que es que cree que es una región de ligador. Por ejemplo, entre los dominios ACP 1 y 2 está la secuencia: APAPVKAAAPAAPVASAPAPA, representada en el presente documento como SEQ ID NO: 15. Las ubicaciones de los residuos de serina de sitios activos (es decir, el sitio de unión a panteteína) para cada uno de los nueve dominios ACP, con respecto a la secuencia de aminoácidos de SEQ ID NO: 2, son las siguientes: ACP1 = S₁₁₅₇; ACP2 = S₁₂₆₆; ACP3 = S₁₃₇₇; ACP4 = S₁₄₈₈; ACP5 = S₁₆₀₄; ACP6 = S₁₇₁₅; ACP7 = S₁₈₁₉; ACP8 = S₁₉₃₀; y ACP9 = S₂₀₃₄. Dado que el tamaño promedio de un dominio ACP es de aproximadamente 85 aminoácidos, excluyendo el ligador, y de aproximadamente 110 aminoácidos incluyendo el ligador, estando la serina del sitio activo aproximadamente en el centro del dominio, un experto en la técnica puede determinar fácilmente las posiciones de cada uno de los nueve dominios ACP en OrfA.

El dominio 12 en OrfA es un dominio KR, también denominado en el presente documento ORFA-KR, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFA-KR se representa en el presente documento como SEQ ID NO: 17 (posiciones 6598-8730 de SEQ ID NO: 1). La secuencia de aminoácidos que contiene el dominio ORFA-KR se representa en el presente documento como SEQ ID NO: 18 (posiciones 2200-2910 de SEQ ID NO: 2). Dentro del dominio KR hay una región central con homología con aldehído-deshidrogenasas de cadena corta (KR es a miembro de esta familia). Esta región central abarca de desde aproximadamente la posición 7198 hasta aproximadamente la posición 7500 de SEQ ID NO: 1, que corresponde a las posiciones de aminoácido 2400-2500 de SEQ ID NO: 2.

Marco de lectura abierto B (OrfB) de *Schizochytrium*:

La secuencia de nucleótidos completa para OrfB se representa en el presente documento como SEQ ID NO: 3. OrfB es una secuencia de 6177 nucleótidos (sin incluir el codón de terminación) que codifica para una secuencia de 2059 aminoácidos, representada en el presente documento como SEQ ID NO: 4. Dentro de OrfB hay cuatro dominios: (a) un dominio -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).

Se han aislado y secuenciado clones de ADN genómico (plásmidos) que codifican para OrfB de tanto *Schizochytrium* sp. ATCC 20888 y una cepa hija de ATCC 20888, indicada como *Schizochytrium* sp., cepa N230D.

Un clon genómico descrito en el presente documento como pJK1129, aislado de *Schizochytrium* sp. ATCC 20888, comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 3, y codifica para la secuencia de aminoácidos de SEQ ID NO: 4. El clon genómico pJK1129 (indicado como clon genómico pJK1129 OrfB, en forma de un vector de plásmido de *E. coli* que contiene el gen "OrfB" de *Schizochytrium* ATCC 20888) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de junio de 2006, y se le asignó el n.º de registro de la ATCC PTA-7649. La secuencia de nucleótidos del clon genómico pJK1126 OrfB, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

Un clon genómico descrito en el presente documento como clon genómico pJK324 OrfB, aislado de *Schizochytrium*

sp. N230D, comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 3, y codifica para la secuencia de aminoácidos de SEQ ID NO: 4. El clon genómico pJK324 (indicado como clon genómico pJK324 OrfB, en forma de un plásmido de *E. coli* que contiene la secuencia del gen OrfB de *Schizochytrium* sp. N230D) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de junio de 2006, y se le asignó el n.º de registro de la ATCC PTA-7643. La secuencia de nucleótidos del clon genómico pJK324 OrfB, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

El primer dominio en OrfB es un dominio KS, también denominado en el presente documento ORFB-KS, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFB-KS se representa en el presente documento como SEQ ID NO: 19 (posiciones 1-1350 de SEQ ID NO: 3). La secuencia de aminoácidos que contiene el dominio ORFB-KS se representa en el presente documento como SEQ ID NO: 20 (posiciones 1-450 de SEQ ID NO: 4). Este dominio KS comprende una valina en la posición 371 de SEQ ID NO: 20 (también posición 371 de SEQ ID NO: 20). Se observa que el dominio ORFB-KS contiene un motivo de sitio activo: DXAC* (*sitio de unión a acilo C₁₉₆). Además, un motivo característico en el extremo de esta región KS, GFGG, está presente en este dominio en SEQ ID NO: 4 y por consiguiente, en SEQ ID NO: 20.

El segundo dominio en OrfB es un dominio CLF, también denominado en el presente documento ORFB-CLF, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFB-CLF se representa en el presente documento como SEQ ID NO: 21 (posiciones 1378-2700 de SEQ ID NO: 3). La secuencia de aminoácidos que contiene el dominio ORFB-CLF se representa en el presente documento como SEQ ID NO: 22 (posiciones 460-900 de SEQ ID NO: 4). Se observa que el dominio ORFB-CLF contiene un motivo de sitio activo de KS sin la proteína de unión a acilo.

El tercer dominio en OrfB es un dominio AT, también denominado en el presente documento ORFB-AT, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFB-AT se representa en el presente documento como SEQ ID NO: 23 (posiciones 2701-4200 de SEQ ID NO: 3). La secuencia de aminoácidos que contiene el dominio ORFB-AT se representa en el presente documento como SEQ ID NO: 24 (posiciones 901-1400 de SEQ ID NO: 4). Se observa que el dominio ORFB-AT contiene un motivo de sitio activo de GxS*xG (*sitio de unión a acilo S₁₁₄₀) que es característico de proteínas aciltransferasa (AT).

El cuarto dominio en OrfB es un dominio ER, también denominado en el presente documento ORFB-ER, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio ORFB-ER se representa en el presente documento como SEQ ID NO: 25 (posiciones 4648-6177 de SEQ ID NO: 3). La secuencia de aminoácidos que contiene el dominio ORFB-ER se representa en el presente documento como SEQ ID NO: 26 (posiciones 1550-2059 de SEQ ID NO: 4).

Marco de lectura abierto C (OrfC) de *Schizochytrium*:

La secuencia de nucleótidos completa para OrfC se representa en el presente documento como SEQ ID NO: 5. OrfC es una secuencia de 4506 nucleótidos (sin incluir el codón de terminación) que codifica para una secuencia de 1502 aminoácidos, representada en el presente documento como SEQ ID NO: 6. Dentro de OrfC hay tres dominios: (a) dos dominios -hidroxil acil-ACP deshidrasa (DH) similares a FabA; y (b) un dominio enoil ACP-reductasa (ER).

Se han aislado y secuenciado clones de ADN genómico (plásmidos) que codifican para OrfC de tanto *Schizochytrium* sp. ATCC 20888 como una cepa hija de ATCC 20888, indicada como *Schizochytrium* sp., cepa N230D.

Un clon genómico descrito en el presente documento como pJK1131, aislado de *Schizochytrium* sp. ATCC 20888, comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 5, y codifica para la secuencia de aminoácidos de SEQ ID NO: 6. El clon genómico pJK1131 (indicado como clon genómico pJK1131 OrfC, en forma de un vector de plásmido de *E. coli* que contiene el gen "OrfC" de *Schizochytrium* ATCC 20888) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de junio de 2006, y se le asignó el n.º de registro de la ATCC PTA-7650. La secuencia de nucleótidos del clon genómico pJK1131 OrfC, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

Un clon genómico descrito en el presente documento como clon genómico pBR002 OrfC, aislado de *Schizochytrium* sp. N230D, comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 5, y codifica para la secuencia de aminoácidos de SEQ ID NO: 6. El clon genómico pBR002 (indicado como clon genómico pBR002 OrfC, en forma de un vector de plásmido de *E. coli* que contiene la secuencia del gen OrfC de *Schizochytrium* sp. N230D) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de junio de 2006, y se le asignó el n.º de registro de la ATCC PTA-7642. La secuencia de nucleótidos del clon genómico pBR002 OrfC, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

El primer dominio en OrfC es un dominio DH, también denominado en el presente documento ORFC-DH1. Éste es uno de dos dominios DH en OrfC, y se designa por tanto DH1. La secuencia de nucleótidos que contiene la secuencia que codifica para el dominio OrfC-DH1 se representa en el presente documento como SEQ ID NO: 27 (posiciones 1-1350 de SEQ ID NO: 5). La secuencia de aminoácidos que contiene el dominio OrfC-DH1 se representa en el presente documento como SEQ ID NO: 28 (posiciones 1-450 de SEQ ID NO: 6).

El segundo dominio en OrfC es un dominio DH, también denominado en el presente documento ORFC-DH2. Éste es el segundo de dos dominios DH en OrfC, y por tanto se designa DH2. La secuencia de nucleótidos que contiene la secuencia que codifica para el dominio OrfC-DH2 se representa en el presente documento como SEQ ID NO: 29 (posiciones 1351-2847 de SEQ ID NO: 5). La secuencia de aminoácidos que contiene el dominio OrfC-DH2 se representa en el presente documento como SEQ ID NO: 30 (posiciones 451-949 de SEQ ID NO: 6). Este dominio DH comprende los aminoácidos H-G-I-A-N-P-T-F-V-H-A-P-G-K-I (SEQ ID NO: 100; posiciones 876-890 de SEQ ID NO: 6) en las posiciones 426-440 de SEQ ID NO: 30.

El tercer dominio en OrfC es un dominio ER, también denominado en el presente documento ORFC-ER, y la secuencia de nucleótidos que contiene la secuencia que codifica para el dominio OrfC-ER se representa en el presente documento como SEQ ID NO: 31 (posiciones 2995-4506 de SEQ ID NO: 5). La secuencia de aminoácidos que contiene el dominio OrfC-ER se representa en el presente documento como SEQ ID NO: 32 (posiciones 999-1502 de SEQ ID NO: 6).

Sistema de PKS de AGPI de *Thraustochytrium*

Hay tres marcos de lectura abiertos que forman el sistema de PKS de AGPI de *Thraustochytrium* 23B central. La organización de dominios es la misma que la que *Schizochytrium* con la excepción de que el Orf A de *Th.* 23B tiene 8 dominios ACP adyacentes, mientras que el Orf de *Schizochytrium* A tiene 9 dominios ACP adyacentes. La estructura de dominios de cada marco de lectura abierto es tal como sigue.

Marco de lectura abierto A (OrfA) de Thraustochytrium 23B:

La secuencia de nucleótidos completa para OrfA de *Th.* 23B se representa en el presente documento como SEQ ID NO: 38. OrfA de *Th.* 23B es una secuencia de 8433 nucleótidos (sin incluir el codón de terminación) que codifica para una secuencia de 2811 aminoácidos, representada en el presente documento como SEQ ID NO: 39. SEQ ID NO: 38 codifica para los siguientes dominios en OrfA de *Th.* 23B: (a) un dominio β -cetoacil-ACP sintasa (KS); (b) un dominio malonil-CoA:ACP aciltransferasa (MAT); (c) ocho dominios de proteína transportadora de acilo (ACP); y (d) un dominio β -cetoacil-ACP reductasa (KR).

Dos clones genómicos descritos en el presente documento como Th23BORfA_pBR812.1 y Th23BORfA_pBR811 (clones genómicos de OrfA), aislados de *Thraustochytrium* 23B, juntos (clones solapantes) comprenden, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 38, y codifica para la secuencia de aminoácidos de SEQ ID NO: 39. El clon genómico Th23BORfA_pBR812.1 (indicado como clon genómico Th23BORfA_pBR812.1, en forma de un vector de plásmido de *E. coli* que contiene la secuencia del gen OrfA de *Thraustochytrium* 23B) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8232. La secuencia de nucleótidos de Th23BORfA_pBR812.1, un clon genómico de OrfA, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención. El clon genómico Th23BORfA_pBR811 (indicado como clon genómico Th23BORfA_pBR811, en forma de un vector de plásmido de *E. coli* que contiene la secuencia del gen OrfA de *Thraustochytrium* 23B) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8231. La secuencia de nucleótidos de Th23BORfA_pBR811, un clon genómico de OrfA, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

El primer dominio en OrfA de *Th.* 23B es un dominio KS, también denominado en el presente documento OrfA-KS de *Th.* 23B, y está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 1 hasta aproximadamente la posición 1500 de SEQ ID NO: 38, representada en el presente documento como SEQ ID NO: 40. La secuencia de aminoácidos que contiene el dominio KS de *Th.* 23B es una región de SEQ ID NO: 39 que abarca de desde aproximadamente la posición 1 hasta aproximadamente la posición 500 de SEQ ID NO: 39, representada en el presente documento como SEQ ID NO: 41. Esta región de SEQ ID NO: 39 tiene una coincidencia de Pfam con FabB (β -cetoacil-ACP sintasa) que abarca de desde la posición 1 hasta aproximadamente la posición 450 de SEQ ID NO: 39 (también posiciones 1 a aproximadamente 450 de SEQ ID NO: 41). Se observa que el dominio *Th.* 23B OrfA-KS contiene un motivo de sitio activo: DXAC* (*sitio de unión a acilo C₂₀₇). Además, un motivo característico en el extremo de la región KS de *Th.* 23B, GFGG, está presente en las posiciones 453-456 de SEQ ID NO: 39 (también posiciones 453-456 de SEQ ID NO: 41).

El segundo dominio en OrfA de *Th.* 23B es un dominio MAT, también denominado en el presente documento OrfA-MAT de *Th.* 23B, y está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la

posición 1503 hasta aproximadamente la posición 3000 de SEQ ID NO: 38, representada en el presente documento como SEQ ID NO: 42. La secuencia de aminoácidos que contiene el dominio MAT de *Th. 23B* es una región de SEQ ID NO: 39 que abarca de desde aproximadamente la posición 501 hasta aproximadamente la posición 1000, representada en el presente documento por SEQ ID NO: 43. Esta región de SEQ ID NO: 39 tiene una coincidencia de Pfam con FabD (malonil-CoA:ACP aciltransferasa) que abarca de desde aproximadamente la posición 580 hasta aproximadamente la posición 900 de SEQ ID NO: 39 (posiciones 80-400 de SEQ ID NO: 43). Se observa que el dominio OrfA-MAT de *Th. 23B* contiene un motivo de sitio activo: GHS*⁺XG (*sitio de unión a acilo S₆₉₇), representado por las posiciones 695-699 de SEQ ID NO: 39.

Los dominios 3-10 de OrfA de *Th. 23B* son ocho dominios ACP en tándem, también denominados en el presente documento OrfA-ACP de *Th. 23B* (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 de *Th. 23B*, OrfA-ACP1 de *Th. 23B*, está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 3205 hasta aproximadamente la posición 3555 de SEQ ID NO: 38 (OrfA), representada en el presente documento como SEQ ID NO: 39, son 1128 (ACP1), 1244 (ACP2), 1360 (ACP3), 1476 (ACP4), 1592 (ACP5), 1708 (ACP6), 1824 (ACP7) y 1940 (ACP8). Las secuencias de nucleótidos y aminoácidos de los ocho dominios ACP de *Th. 23B* están altamente conservadas y, por tanto, la secuencia para cada dominio no se representa en el presente documento por un identificador de secuencia individual. Sin embargo, basándose en la información dada a conocer en el presente documento, un experto en la técnica puede determinar fácilmente la secuencia que contiene cada uno de los otros siete dominios ACP en SEQ ID NO: 38 y SEQ ID NO: 39.

Los ocho dominios ACP en OrfA de *Th. 23B* son adyacentes entre sí y pueden identificarse por la presencia del motivo de sitio de unión a fosfopanteteína, LGXDS* (representado por SEQ ID NO: 46), en el que la S* es el sitio de acoplamiento a fosfopanteteína. La posición de aminoácido de cada uno de los ocho sitios S*, con referencia a SEQ ID NO: 39, son 1128 (ACP1), 1244 (ACP2), 1360 (ACP3), 1476 (ACP4), 1592 (ACP5), 1708 (ACP6), 1824 (ACP7) y 1940 (ACP8). Las secuencias de nucleótidos y aminoácidos de los ocho dominios ACP de *Th. 23B* están altamente conservadas y, por tanto, la secuencia para cada dominio no se representa en el presente documento por un identificador de secuencia individual. Sin embargo, basándose en la información dada a conocer en el presente documento, un experto en la técnica puede determinar fácilmente la secuencia que contiene cada uno de los otros siete dominios ACP en SEQ ID NO: 38 y SEQ ID NO: 39.

Todos los ocho dominios ACP de *Th. 23B* juntos abarcan una región de OrfA de *Th. 23B* de desde aproximadamente la posición 3205 hasta aproximadamente la posición 5994 de SEQ ID NO: 38, que corresponde a las posiciones de aminoácido de desde aproximadamente 1069 hasta aproximadamente 1998 de SEQ ID NO: 39. La secuencia de nucleótidos para toda la región ACP que contiene los ocho dominios se representa en el presente documento como SEQ ID NO: 47. SEQ ID NO: 47 codifica para una secuencia de aminoácidos representada en el presente documento por SEQ ID NO: 48. SEQ ID NO: 48 incluye los segmentos de ligador entre dominios ACP individuales. El intervalo de repetición para los ocho dominios es aproximadamente cada 116 aminoácidos de SEQ ID NO: 48, y puede considerarse que cada dominio consiste en aproximadamente 116 aminoácidos centrados en el motivo de sitio activo (descrito anteriormente).

El último dominio en OrfA de *Th. 23B* es un dominio KR, también denominado en el presente documento OrfA-KR de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 6001 hasta aproximadamente la posición 8433 de SEQ ID NO: 38, representada en el presente documento por SEQ ID NO: 49. La secuencia de aminoácidos que contiene el dominio KR de *Th. 23B* es una región de SEQ ID NO: 39 que abarca de desde aproximadamente la posición 2001 hasta aproximadamente la posición 2811 de SEQ ID NO: 39, representada en el presente documento por SEQ ID NO: 50. Esta región de SEQ ID NO: 39 tiene una coincidencia de Pfam con FabG (β -cetoacil-ACP reductasa) que abarca de desde aproximadamente la posición 2300 hasta aproximadamente 2550 de SEQ ID NO: 39 (posiciones 300-550 de SEQ ID NO: 50).

Marco de lectura abierto B (OrfB) de Thraustochytrium 23B:

La secuencia de nucleótidos completa para OrfB de *Th. 23B* se representa en el presente documento como SEQ ID NO: 51, que es una secuencia de 5805 nucleótidos (sin incluir el codón de terminación) que codifica para una secuencia de 1935 aminoácidos, representada en el presente documento como SEQ ID NO: 52. SEQ ID NO: 51 codifica para los siguientes dominios en OrfB de *Th. 23B*: (a) un dominio β -cetoacil-ACP sintasa (KS); (b) un dominio de factor de longitud de cadena (CLF); (c) un dominio aciltransferasa (AT); y (d) un dominio enoil-ACP reductasa (ER).

Un clon genómico descrito en el presente documento como Th23BOrfB_pBR800 (clon genómico de OrfB), aislado de *Thraustochytrium 23B*, comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 51, y codifica para la secuencia de aminoácidos de SEQ ID NO: 52. El clon genómico Th23BOrfB_pBR800 (indicado como clon genómico Th23BOrfB_pBR800, en forma de un vector de plásmido de *E. coli* que contiene la secuencia del gen OrfB de *Thraustochytrium 23B*) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8227. La secuencia de nucleótidos de Th23BOrfB_pBR800, un clon genómico de OrfB, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

El primer dominio en el OrfB de *Th. 23B* es un dominio KS, también denominado en el presente documento OrfB-KS de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 1 hasta aproximadamente la posición 1500 de SEQ ID NO: 51 (OrfB de *Th. 23B*), representada en el presente documento como SEQ ID NO: 53. La secuencia de aminoácidos que contiene el dominio KS de *Th. 23B* es una región de SEQ ID NO: 52 que abarca de desde aproximadamente la posición 1 hasta aproximadamente la posición 500 de SEQ ID NO: 52, representada en el presente documento como SEQ ID NO: 54. Esta región de SEQ ID NO: 52 tiene una coincidencia de Pfam con FabB (β -cetoacil-ACP sintasa) que abarca de desde aproximadamente la posición 1 hasta aproximadamente la posición 450 (posiciones 1-450 de SEQ ID NO: 54). Se observa que el dominio OrfB-KS de *Th. 23B* contiene un motivo de sitio activo: DXAC*, en el que C* es el sitio de acoplamiento de grupo acilo y en el que la C* está en la posición 201 de SEQ ID NO: 52. Además, un motivo característico en el extremo de la región KS, GFGG, está presente en las posiciones de aminoácido 434-437 de SEQ ID NO: 52.

El segundo dominio en OrfB de *Th. 23B* es un dominio CLF, también denominado en el presente documento OrfB-CLF de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 1501 hasta aproximadamente la posición 3000 de SEQ ID NO: 51 (OrfB), representada en el presente documento como SEQ ID NO: 55. La secuencia de aminoácidos que contiene el dominio CLF es una región de SEQ ID NO: 52 que abarca de desde aproximadamente la posición 501 hasta aproximadamente la posición 1000 de SEQ ID NO: 52, representada en el presente documento como SEQ ID NO: 56. Esta región de SEQ ID NO: 52 tiene una coincidencia de Pfam con FabB (β -cetoacil-ACP sintasa) que abarca de desde aproximadamente la posición 550 hasta aproximadamente la posición 910 (posiciones 50-410 de SEQ ID NO: 56). Aunque CLF tiene homología con proteínas KS, carece de una cisteína de sitio activo a la que se acopla el grupo acilo en proteínas KS.

El tercer dominio en OrfB de *Th. 23B* es un dominio AT, también denominado en el presente documento OrfB-AT de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 3001 hasta aproximadamente la posición 4500 de SEQ ID NO: 51 (OrfB de *Th. 23B*), representada en el presente documento como SEQ ID NO: 58. La secuencia de aminoácidos que contiene el dominio AT de *Th. 23B* es una región de SEQ ID NO: 52 que abarca de desde aproximadamente la posición 1001 hasta aproximadamente la posición 1500 de SEQ ID NO: 52, representada en el presente documento como SEQ ID NO: 58. Esta región de SEQ ID NO: 52 tiene una coincidencia de Pfam con FabD (malonil-CoA:ACP aciltransferasa) que abarca de desde aproximadamente la posición 1100 hasta aproximadamente la posición 1375 (posiciones 100-375 de SEQ ID NO: 58). Aunque este dominio AT de las AGPI sintetas tiene homología con proteínas MAT, carece del motivo extendido del MAT (residuos arginina y glutamina clave) y no se piensa que esté implicado en transferencias de malonil-CoA. Está presente el motivo GXS*XG de aciltransferasas, siendo la S* el sitio de acoplamiento de acilo y estando ubicada en la posición 1123 con respecto a SEQ ID NO: 52.

El cuarto dominio en OrfB de *Th. 23B* es un dominio ER, también denominado en el presente documento OrfB-ER de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 4501 hasta aproximadamente la posición 5805 de SEQ ID NO: 51 (OrfB), representada en el presente documento como SEQ ID NO: 59. La secuencia de aminoácidos que contiene el dominio ER de *Th. 23B* es una región de SEQ ID NO: 52 que abarca de desde aproximadamente la posición 1501 hasta aproximadamente la posición 1935 de SEQ ID NO: 52, representada en el presente documento como SEQ ID NO: 60. Esta región de SEQ ID NO: 52 tiene una coincidencia de Pfam con una familia de dioxigenasas relacionadas con 2-nitropropano dioxigenasas que abarca de desde aproximadamente la posición 1501 hasta aproximadamente la posición 1810 (posiciones 1-310 de SEQ ID NO: 60). Puede predecirse adicionalmente que este dominio funciona como una ER debido a la homología con una enzima ER recién caracterizada de *Streptococcus pneumoniae*.

Marco de lectura abierto C (OrfC) de *Thraustochytrium 23B*:

La secuencia de nucleótidos completa para OrfC de *Th. 23B* se representa en el presente documento como SEQ ID NO: 61, que es una secuencia de 4410 nucleótidos (sin incluir el codón de terminación) que codifica para una secuencia de 1470 aminoácidos, representada en el presente documento como SEQ ID NO: 62. SEQ ID NO: 61 codifica para los siguientes dominios en OrfC de *Th. 23B*: (a) dos dominios β -hidroxiacil-ACP deshidrasa (DH) similares a FabA, ambos con homología con la proteína FabA (una enzima que cataliza la síntesis de trans-2-decenil-ACP y la isomerización reversible de este producto a cis-3-decenil-ACP); y (b) un dominio enoil-ACP reductasa (ER) con alta homología con el dominio ER de OrfB de *Schizochytrium*.

Un clon genómico descrito en el presente documento como Th23BOrfC_pBR709A (clon genómico de OrfC), aislado de *Thraustochytrium 23B*, comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 61, y codifica para la secuencia de aminoácidos de SEQ ID NO: 62. El clon genómico Th23BOrfC_pBR709A (indicado como clon genómico Th23BOrfC_pBR709A, en forma de un vector de plásmido de *E. coli* que contiene la secuencia del gen OrfC de *Thraustochytrium 23B*) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8228. La secuencia de nucleótidos de Th23BOrfC_pBR709A, un clon genómico de OrfC, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

El primer dominio en OrfC de *Th. 23B* es un dominio DH, también denominado en el presente documento OrfC-DH1 de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 1 hasta aproximadamente la posición 1500 de SEQ ID NO: 61 (OrfC), representada en el presente documento como SEQ ID NO: 63. La secuencia de aminoácidos que contiene el dominio DH1 de *Th. 23B* es una región de SEQ ID NO: 62 que abarca de desde aproximadamente la posición 1 hasta aproximadamente la posición 500 de SEQ ID NO: 62, representada en el presente documento como SEQ ID NO: 64. Esta región de SEQ ID NO: 62 tiene una coincidencia de Pfam con FabA, tal como se mencionó anteriormente, que abarca de desde aproximadamente la posición 275 hasta aproximadamente la posición 400 (posiciones 275-400 de SEQ ID NO: 64).

El segundo dominio en OrfC de *Th. 23B* es también un dominio DH, también denominado en el presente documento OrfC-DH2 de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 1501 hasta aproximadamente 3000 de SEQ ID NO: 61 (OrfC), representada en el presente documento como SEQ ID NO: 65. La secuencia de aminoácidos que contiene el dominio DH2 de *Th. 23B* es una región de SEQ ID NO: 62 que abarca de desde aproximadamente la posición 501 hasta aproximadamente la posición 1000 de SEQ ID NO: 62, representada en el presente documento como SEQ ID NO: 66. Esta región de SEQ ID NO: 62 tiene una coincidencia de Pfam con FabA, tal como se mencionó anteriormente, que abarca de desde aproximadamente la posición 800 hasta aproximadamente la posición 925 (posiciones 300-425 de SEQ ID NO: 66).

El tercer dominio en OrfC de *Th. 23B* es un dominio ER, también denominado en el presente documento OrfC-ER de *Th. 23B*, que está contenido dentro de la secuencia de nucleótidos que abarca de desde aproximadamente la posición 3001 hasta aproximadamente la posición 4410 de SEQ ID NO: 61 (OrfC), representada en el presente documento como SEQ ID NO: 67. La secuencia de aminoácidos que contiene el dominio ER de *Th. 23B* es una región de SEQ ID NO: 62 que abarca de desde aproximadamente la posición 1001 hasta aproximadamente la posición 1470 de SEQ ID NO: 62, representada en el presente documento como SEQ ID NO: 68. Esta región de SEQ ID NO: 62 tiene una coincidencia de Pfam con las dioxigenasas relacionadas con 2-nitropropano dioxigenasas, tal como se mencionó anteriormente, que abarca de desde aproximadamente la posición 1025 hasta aproximadamente la posición 1320 (posiciones 25-320 de SEQ ID NO: 68). También puede predecirse que este dominio funciona como una ER debido a la homología con una enzima ER recién caracterizada de *Streptococcus pneumoniae*.

Constructos sintéticos optimizados para codones

La invención también abarca versiones resintetizadas de cualquiera de las secuencias de ácido nucleico descritas en el presente documento, que tienen principalmente un uso de codones optimizado para un organismo heterólogo (huésped heterólogo), en el que la secuencia de aminoácidos codificada no está cambiada con referencia a la secuencia de aminoácidos natural, de tipo natural o fuente. Los presentes inventores han descubierto que la resintetización de secuencias de ácido nucleico para un uso de codones óptimo es un modo eficaz para mejorar la producción de AGPI en un huésped heterólogo que se transforma con moléculas de ácido nucleico de un sistema de PKS de AGPI. La resíntesis de todas las moléculas de ácido nucleico en un sistema de PKS de AGPI no se requiere necesariamente para una expresión y producción de AGPI óptimas en un huésped heterólogo. En efecto, los inventores han encontrado que la resíntesis de sólo algunas de las moléculas de ácido nucleico es suficiente para mejorar la producción de AGPI. Por ejemplo, mientras que la resíntesis de los Orf A y B de *Schizochytrium* mejoraba la expresión de AGPI sintasa y la producción de AGPI en levadura, el uso del OrfC de *Schizochytrium* nativo y HetI PPTasa de *Nostoc* nativa eran suficientes. Además, la optimización de codones de un constructo para su uso en un huésped heterólogo también puede ser útil para mejorar la producción de AGPI en un huésped heterólogo diferente (por ejemplo, la optimización del uso de codones de una secuencia que codifica para OrfC de *Thraustochytrium* para su uso en *Schizochytrium* también puede ser eficaz para estimular la producción de AGPI en otro organismo huésped heterólogo, tal como plantas).

Además, el uso de constructos sintéticos optimizados para codones puede ser útil en la producción de constructos PKS de AGPI quiméricos y/o sistemas de PKS de AGPI quiméricos, en los que se introduce un dominio o una proteína de un sistema de PKS de AGPI (por ejemplo, de un primer organismo) en un segundo sistema de PKS de AGPI (por ejemplo, de un segundo organismo). En tales sistemas, no sólo puede manipularse el perfil de AGPI (por ejemplo, mediante el uso de los constructos quiméricos y/o sistemas de PKS de AGPI quiméricos), sino que la producción de AGPI también puede mejorarse mediante el uso de constructos quiméricos sintéticos optimizados para codones. En efecto, la combinación de los dos conceptos (quimeras y optimización de codones) puede producir un resultado sinérgico con respecto a los perfiles de AGPI y/o la producción de AGPI. Se incluyen en la invención sistemas que contienen algunas secuencias que están optimizadas para codones para el huésped y algunas que no están optimizadas para codones para el huésped.

A continuación se describen determinadas secuencias optimizadas para codones a modo de ejemplo. Otras secuencias optimizadas para codones resultarán evidentes para los expertos en la técnica tras esta descripción:

sOrfA

SEQ ID NO: 35, indicada como sOrfA, representa la secuencia de ácido nucleico que codifica para OrfA de *Schizochytrium* (SEQ ID NO: 1) que se ha resintetizado para el uso de codones optimizado en levadura. SEQ ID NO: 1 y SEQ ID NO: 35 codifican cada una para SEQ ID NO: 2.

sOrfB

SEQ ID NO: 36, indicada como sOrfB, representa la secuencia de ácido nucleico que codifica para OrfB de *Schizochytrium* (SEQ ID NO: 3) que se ha resintetizado para el uso de codones optimizado en levadura. SEQ ID NO: 3 y SEQ ID NO: 36 codifican cada una para SEQ ID NO: 4.

OrfB*

SEQ ID NO: 37, indicada como OrfB* (pJK962), representa una secuencia de ácido nucleico que codifica para OrfB de *Schizochytrium* (SEQ ID NO: 4) que se ha resintetizado dentro de una porción de SEQ ID NO: 3 (secuencia de nucleótidos que codifica para SEQ ID NO: 4) para su uso en células de plantas, y que se derivó de una secuencia muy similar desarrollada inicialmente para el uso de codones optimizado en *E. coli*, también denominada OrfB* (pJK780), que se describe a continuación. OrfB* en ambas formas (para *E. coli* y para plantas) es idéntica a SEQ ID NO: 3 con la excepción de un BspHI resintetizado (nucleótido 4415 de SEQ ID NO: 3) para dar un fragmento SacII (sitio único en SEQ ID NO: 3). Ambas versiones (*E. coli* y planta) tienen otras dos modificaciones de codones cerca del inicio del gen en comparación con la secuencia genómica original de orfB (SEQ ID NO: 3). En primer lugar, el cuarto codón, arginina (R), se cambió de CGG en la secuencia genómica a CGC en orfB*. En segundo lugar, el quinto codón, asparagina (N), se cambió de AAT en la secuencia genómica a AAC en orfB*. Con el fin de facilitar la clonación de este gen en los vectores de plantas para crear SEQ ID NO: 37, también se modificó por ingeniería genética un sitio de PstI (CTGCAG) en la secuencia de orfB* de *E. coli* 20 bases a partir del inicio del gen. Este cambio no alteró la secuencia de aminoácidos de la proteína codificada. Tanto SEQ ID NO: 37 como SEQ ID NO: 3 (así como la forma de OrfB* para *E. coli*, descrita en SEQ ID NO: 69 a continuación) codifican para SEQ ID NO: 4.

SEQ ID NO: 69, indicada como OrfB* (pJK780), representa una secuencia de ácido nucleico que codifica para OrfB de *Schizochytrium* (SEQ ID NO: 4) que se ha resintetizado dentro de una porción de SEQ ID NO: 3 (secuencia de nucleótidos que codifica para SEQ ID NO: 4) para su uso en *E. coli*. La secuencia del constructo de OrfB* en ambas formas (para *E. coli* y para plantas) se ha descrito anteriormente. SEQ ID NO: 69 y SEQ ID NO: 3 codifican para SEQ ID NO: 4.

El plásmido descrito en el presente documento como OrfB*_pJK780 comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 69, y codifica para la secuencia de aminoácidos de SEQ ID NO: 4. El plásmido OrfB*_pJK780 (indicado como clon OrfB*_pJK780, en forma de un vector de plásmido de *E. coli*) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8225. La secuencia de nucleótidos de OrfB*_pJK780 y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

pThOrfC-synPS

SEQ ID NO: 70 representa una secuencia de ácido nucleico que codifica para un OrfC de *Thraustochytrium* 23B (SEQ ID NO: 61, que codifica para SEQ ID NO: 62) que se ha resintetizado para el uso de codones optimizado en *Schizochytrium*. Las posiciones 2000-6412 de SEQ ID NO: 70 representan la región codificante para la proteína OrfC de *Thraustochytrium* 23B (incluyendo el codón de terminación). Las posiciones 1-1999 y 6413-8394 de SEQ ID NO: 70 representan secuencias de OrfC de *Schizochytrium* en el sentido de 5' y en sentido de 3' (regiones no codificantes), respectivamente. La construcción del plásmido que contiene SEQ ID NO: 70, indicado como pThOrfC-synPS, se describe en detalle en el ejemplo 1. SEQ ID NO: 70 y SEQ ID NO: 61 codifican cada una para SEQ ID NO: 62. pThOrfC-syn PS se diseña para sustituir exactamente la región codificante (CDS) de orfC de *Schizochytrium* (SEQ ID NO: 5) por la región codificante para el orfC de *Thraustochytrium* 23B, resintetizada tal como se comentó anteriormente (SEQ ID NO: 70). La producción y el uso de organismos que se han transformado con este constructo se describen en detalle a continuación y en los ejemplos.

El plásmido descrito anteriormente como pThOrfC-synPS comprende, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 70, y codifica para la secuencia de aminoácidos correspondiente de SEQ ID NO: 62. El plásmido pThOrfC-synPS (indicado como pThOrfC-synPS, en forma de un vector de plásmido de *E. coli* que contiene un OrfC de PKS de AGPI de *Thraustochytrium* 23B con "costura perfecta" ("perfect stitch") optimizado para codones para la expresión en *Schizochytrium* o otros huéspedes heterólogos) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8229. La secuencia de nucleótidos de pThOrfC-synPS, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

pDD26

SEQ ID NO: 71 representa una secuencia de ácido nucleico que codifica para un OrfA de *Thraustochytrium* 23B (SEQ ID NO: 38, que codifica para SEQ ID NO: 39) que se ha resintetizado para el uso de codones optimizado en *Schizochytrium*. Las posiciones 2044-10479 de SEQ ID NO: 71 representan la región codificante para la proteína OrfA de *Thraustochytrium* 23B (incluyendo el codón de terminación). Las posiciones 1-2043 y 10480-12495 de SEQ ID NO: 71 representan secuencias de OrfA de *Schizochytrium* en el sentido de 5' y en el sentido de 3' (regiones no codificantes), respectivamente. La construcción del plásmido que contiene SEQ ID NO: 71, indicado como pDD26, se describe en detalle en el ejemplo 8. SEQ ID NO: 71 y SEQ ID NO: 38 codifican cada una para SEQ ID NO: 39. pDD26 se diseña para sustituir exactamente la región codificante (CDS) de orfA de *Schizochytrium* (SEQ ID NO: 1) por la región codificante para el orfC de *Thraustochytrium* 23B, resintetizada tal como se comentó anteriormente (SEQ ID NO: 71). La producción y uso de organismos que se han transformado con este constructo se describen en detalle a continuación y en los ejemplos.

Los plásmidos descritos anteriormente como pDD26 comprenden, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 71, y codifican para la secuencia de aminoácidos correspondiente de SEQ ID NO: 39. El plásmido pDD26 (indicado como pDD26, en forma de un vector de plásmido de *E. coli*) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 USA el 8 de mayo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8411. La secuencia de nucleótidos de pDD26, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

pDD32

SEQ ID NO: 72 representa una secuencia de ácido nucleico que codifica para un OrfB de *Thraustochytrium* 23B (SEQ ID NO: 51, que codifica para SEQ ID NO: 52) que se ha resintetizado para el uso de codones optimizado en *Schizochytrium*. Las posiciones 1452-7259 de SEQ ID NO: 72 representan la región codificante para la proteína OrfB de *Thraustochytrium* 23B (incluyendo el codón de terminación). Las posiciones 1-1451 y 7260-8647 de SEQ ID NO: 72 representan secuencias de OrfB de *Schizochytrium* en el sentido de 5' y en el sentido de 3' (regiones no codificantes), respectivamente. La construcción del plásmido que contiene SEQ ID NO: 72, indicada como pDD32, se describe en detalle en el ejemplo 8. SEQ ID NO: 72 y SEQ ID NO: 51 codifican cada una para SEQ ID NO: 52. pDD32 se diseña para sustituir exactamente la región codificante (CDS) de orfB de *Schizochytrium* (SEQ ID NO: 3) por la región codificante para el orfC de *Thraustochytrium* 23B, resintetizada tal como se comentó anteriormente (SEQ ID NO: 72). La producción y uso de organismos que se han transformado con este constructo se describen en detalle a continuación y en los ejemplos.

Los plásmidos descritos anteriormente como pDD32 comprenden, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 72, y codifican para la secuencia de aminoácidos correspondiente de SEQ ID NO: 52. El plásmido pDD32 (indicado como pDD32, en forma de un vector de plásmido de *E. coli*) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 8 de mayo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8412. La secuencia de nucleótidos de pDD32, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

45 *Constructos de PKS de AGPI quiméricos*

La invención también abarca constructos quiméricos que usan porciones de dos o más secuencias de ácido nucleico de PKS de AGPI diferentes, tales como las descritas en el presente documento, para producir proteínas PKS de AGPI quiméricas. Los presentes inventores demuestran en el presente documento en varios ejemplos diferentes que "mezclando y haciendo coincidir" dominios o porciones de proteínas PKS de AGPI de diferentes organismos (es decir, creando proteínas PKS de AGPI quiméricas compuestas por dominios o polipéptidos de dos o más organismos diferentes), puede modificarse el perfil de AGPI producidos por un organismo que expresa un sistema de PKS de AGPI que contiene tales proteínas quiméricas, en comparación con un sistema de PKS de AGPI nativo (que se produce de manera natural). Por ejemplo, los presentes inventores describen en el presente documento el uso del dominio DH2 de un sistema de PKS de AGPI de *Thraustochytrium* en la proteína OrfC de una proteína de *Schizochytrium*, de modo que la proteína OrfC quimérica resultante contiene los dominios DH1 y ER de *Schizochytrium*, y el dominio DH2 de *Thraustochytrium*. El constructo quimérico se modifica adicionalmente mediante el uso de un dominio DH2 de *Thraustochytrium* optimizado para codones (para *Schizochytrium*) en un constructo, y un dominio DH2 de *Thraustochytrium* nativo en otro constructo, lo que demuestra la flexibilidad y los efectos de las diversas modificaciones descritas en el presente documento.

Se describen a continuación determinados constructos quiméricos a modo de ejemplo. Otros constructos quiméricos resultarán evidentes para los expertos en la técnica tras esta descripción.

65 *pDS49*

SEQ ID NO: 73 representa una secuencia de ácido nucleico que codifica para una proteína quimérica que comprende una proteína OrfC de *Schizochytrium* (SEQ ID NO: 6) en la que el dominio DH2 (SEQ ID NO: 30) se ha sustituido por el dominio DH2 (secuencia que incluye SEQ ID NO: 66) de OrfC de *Thraustochytrium* 23B (SEQ ID NO: 62). En este constructo quimérico, la secuencia que codifica para DH2 de *Thraustochytrium* es la secuencia nativa (no optimizada para codones). La construcción del plásmido que contiene SEQ ID NO: 73, indicado como pDS49, se describe en detalle en el ejemplo 2. Las secuencias no codificantes en el sentido de 5' y en el sentido de 3' de OrfC de *Schizochytrium* que flanquean SEQ ID NO: 73 en pDS49 son las mismas que las descritas anteriormente con respecto a SEQ ID NO: 70 (no representadas en SEQ ID NO: 73). SEQ ID NO: 73 codifica para una secuencia de aminoácidos de SEQ ID NO: 74. En referencia a SEQ ID NO: 74, el polipéptido OrfC quimérico tiene 1493 residuos de aminoácido de longitud. La región DH2, definida como los aminoácidos 516-1041 de SEQ ID NO: 74, consiste en la secuencia de aminoácidos de la región DH2 de la proteína OrfC de *Th.23B*, es decir, los aminoácidos 491-1016 de SEQ ID NO: 62, que incluye toda la SEQ ID NO: 66 y alguna secuencia de aminoácidos flanqueante de SEQ ID NO: 62. Con respecto al resto de la secuencia de aminoácidos de OrfC quimérica, los residuos 1-515 y 1042-1493 de SEQ ID NO: 74 son idénticos a los residuos de OrfC de *Schizochytrium* 1-515 y 1051-1502 de SEQ ID NO: 6, respectivamente. La producción y el uso de organismos que se han transformado con este constructo se describen en detalle a continuación y en los ejemplos.

Los plásmidos descritos anteriormente como pDS49 comprenden, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 73, y codifican para la secuencia de aminoácidos correspondiente de SEQ ID NO: 74. El plásmido pDS49 (indicado como pDS49, en forma de un vector de plásmido de *E. coli*) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8230. La secuencia de nucleótidos de pDS49, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

oDD24

SEQ ID NO: 75 representa otra secuencia de ácido nucleico que codifica para una proteína quimérica que comprende una proteína OrfC de *Schizochytrium* (SEQ ID NO: 6) en la que el dominio DH2 (SEQ ID NO: 30) se ha sustituido por el dominio DH2 (secuencia que incluye SEQ ID NO: 66) de OrfC de *Thraustochytrium* 23B (SEQ ID NO: 62). En este constructo quimérico, la secuencia que codifica para DH2 de *Thraustochytrium* es una secuencia optimizada para codones para su uso en *Schizochytrium*. La construcción del plásmido que contiene SEQ ID NO: 75, indicado como pDD24, se describe en detalle en el ejemplo 3. Las secuencias no codificantes en el sentido de 5' y en el sentido de 3' de OrfC de *Schizochytrium* que flanquean SEQ ID NO: 75 en pDD24 son las mismas que las descritas anteriormente con respecto a SEQ ID NO: 70 (no representadas en SEQ ID NO: 75). SEQ ID NO: 75 codifica para una secuencia de aminoácidos de SEQ ID NO: 74. SEQ ID NO: 74 se ha descrito en detalle anteriormente con respecto a SEQ ID NO: 73, que también codifica para SEQ ID NO: 74. Sin embargo, en este constructo, tal como se comentó anteriormente la secuencia de nucleótidos que codifica para los aminoácidos 516-1041 de SEQ ID NO: 74 se derivó de la "secuencia génica sintética" para OrfC de *Thraustochytrium* 23B que está contenida en el plásmido pThOrfC-synPS (véase el ejemplo 1 y SEQ ID NO: 70) y que emplea codones que se prefieren para la expresión génica en *Schizochytrium*. La producción y el uso de organismos que se han transformado con este constructo se describen en detalle a continuación y en los ejemplos.

Los plásmidos descritos anteriormente como pDD24 comprenden, según el mejor entender de los presentes inventores, la secuencia de nucleótidos de SEQ ID NO: 75, y codifican para la secuencia de aminoácidos correspondiente de SEQ ID NO: 74. El plásmido pDD24 (indicado como pDD24, en forma de un vector de plásmido de *E. coli*) se depositó en la Colección Americana de Cultivos Tipo (ATCC), 10801 University Boulevard, Manassas, Va. 20110-2209 EE.UU. el 1 de marzo de 2007, y se le asignó el n.º de registro de la ATCC PTA-8226. La secuencia de nucleótidos de pDD24, y la secuencia de aminoácidos codificada por este plásmido se abarcan por la presente invención.

Sistemas de PKS de AGPI quiméricos

Además del uso de la optimización de codones y los constructos quiméricos descritos anteriormente, la invención incluye la producción y el uso de sistemas de PKS de AGPI quiméricos. Los sistemas de PKS de AGPI quiméricos incluyen el uso de los constructos quiméricos descritos anteriormente, en los que se crea una proteína PKS de AGPI quimérica y se usa en un sistema de PKS de AGPI, pero tales sistemas también abarcan sistemas de PKS de AGPI en los que una o más proteínas completas o proteínas de uno o más sistema(s) de PKS de AGPI se intercambian por o se añaden a la correspondiente proteína completa o proteínas de otro sistema de PKS de AGPI, de manera que el sistema de PKS de AGPI resultante comprende proteínas de dos o más sistemas de PKS de AGPI diferentes. Tales sistemas también pueden incluir el uso de proteínas quiméricas, tal como se describió anteriormente (por ejemplo, proteínas quiméricas y sustituciones de proteínas completas). Por ejemplo, el constructo descrito anteriormente como pTh23B_synPS (que comprende una secuencia que codifica para OrfC de *Thraustochytrium* 23B, optimizada para el uso de codones de *Schizochytrium*) puede sustituirse en un sistema de PKS de AGPI de *Schizochytrium* para reemplazar perfectamente la secuencia que codifica para OrfC de *Schizochytrium* nativa, creando de ese modo un sistema de PKS de AGPI quimérico. Como otro ejemplo, la secuencia que codifica para

OrfC de *Thraustochytrium* 23B nativa (no optimizada para codones) puede sustituirse en un sistema de PKS de AGPI de *Schizochytrium* para reemplazar perfectamente la secuencia que codifica para OrfC de *Schizochytrium* nativa, creando de ese modo otro sistema de PKS de AGPI quimérico. Aún como otro ejemplo, las secuencias que codifican para OrfA y OrfC de *Thraustochytrium* 23B nativas (optimizadas para codones o no) pueden sustituirse en un sistema de PKS de AGPI de *Schizochytrium* para sustituir perfectamente las secuencias que codifican para OrfA y OrfC de *Schizochytrium* nativas, respectivamente, creando de ese modo aún otro sistema de PKS de AGPI quimérico. Estos y otros sistemas de PKS de AGPI quiméricos se describen en los ejemplos a continuación. Se incluyen en los ejemplos huéspedes de *Schizochytrium* que expresan sistemas de PKS de AGPI quiméricos compuestos por: (1) OrfA, SOrfB de *Schizochytrium* (S), y OrfC de *Thraustochytrium* (Th); (2) SOrfA, ThOrfB y SOrfC; (3) ThOrfA, SOrfB y SOrfC; (4) SOrfA, ThOrfB y ThOrfC; (5) ThOrfA, SOrfB y ThOrfC; (6) ThOrfA, ThOrfB y SOrfC; y (7) ThOrfA, ThOrfB y ThOrfC.

Basándose en la descripción y los experimentos a modo de ejemplo proporcionados en el presente documento, ahora es posible mejorar y/o modificar la producción de AGPI mediante resíntesis seleccionada de moléculas de ácido nucleico de PKS de AGPI para el uso de codones del huésped, y/o el uso de constructos de PKS de AGPI quiméricos y/o sistemas de PKS de AGPI quiméricos en diversos organismos huésped, incluyendo en organismos huésped que no tienen de manera endógena un sistema de PKS de AGPI para la producción de AGPI.

Fosfopanteteinil transferasa (PPTasa)

Según la presente invención, un sistema de PKS de AGPI para la producción y/o acumulación de AGPI en un huésped heterólogo o la producción y/o acumulación mejoradas de AGPI en un huésped endógeno puede hacer uso de diversas proteínas auxiliares, que se definen en el presente documento como proteínas que no se considera que son parte del sistema de PKS de AGPI central tal como se describió anteriormente (es decir, no son parte del propio complejo de la enzima AGPI sintasa), pero que pueden ser, o son, necesarias para la producción de AGPI o al menos para la producción eficiente de AGPI usando el complejo de la enzima AGPI sintasa central de la presente invención.

Con el fin de producir AGPI, un sistema de PKS de AGPI debe actuar con una proteína auxiliar que transfiere un resto 4'-fosfopanteteinilo de coenzima A al/a los dominio(s) de proteína transportadora de acilo (ACP). Por tanto, un sistema de PKS de AGPI puede considerarse que incluye al menos un dominio 4'-fosfopanteteinil transferasa (PPTasa), o un dominio de este tipo puede considerarse que es un dominio o una proteína auxiliar para el sistema de PKS de AGPI. Las características estructurales y funcionales de PPTasas se han descrito en detalle, por ejemplo, en la publicación de solicitud de patente estadounidense n.º 20020194641; publicación de solicitud de patente estadounidense n.º 20040235127; y la publicación de solicitud de patente estadounidense n.º 20050100995.

Según la presente invención, un dominio o una proteína que tiene actividad biológica (función) 4'-fosfopanteteinil transferasa (PPTasa) se caracteriza como la enzima que transfiere un resto 4'-fosfopanteteinilo de la coenzima A a la proteína transportadora de acilo (ACP). Esta transferencia a un residuo de serina invariante de la ACP activa la apoforma inactiva a la holoforma. En la síntesis de tanto policétidos como ácidos grasos, el grupo fosfopanteteína forma tioésteres con las cadenas de acilo en crecimiento. Las PPTasas son una familia de enzimas que se han caracterizado bien en la síntesis de ácidos grasos, la síntesis de policétidos y la síntesis de péptidos no ribosómica. Se conocen las secuencias de muchas PPTasas, y se han determinado las estructuras cristalinas (por ejemplo, Reuter K, Mofid MR, Marahiel MA, Ficner R. "Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily" EMBO J. 1 de diciembre de 1999; 18(23):6823-31) así como el análisis mutacional de residuos de aminoácido importantes para la actividad (Mofid MR, Finking R, Essen LO, Marahiel MA. "Structure-based mutational analysis of the 4'-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: carrier protein recognition and reaction mechanism" Biochemistry. 13 de abril de 2004; 43(14): 4128-36). Estos aminoácidos invariantes y altamente conservados en PPTasas están contenidos dentro de los ORF de pfaE de ambas cepas de *Shewanella* descritas anteriormente.

Una PPTasa heteróloga que se ha demostrado anteriormente que reconoce los dominios ACP de OrfA descritos en el presente documento como sustratos es la proteína Het I de *Nostoc* sp. PCC 7120 (anteriormente denominado *Anabaena* sp. PCC 7120). Het I está presente en una agrupación de genes en *Nostoc* que se sabe que son responsables de la síntesis de hidroxi-ácidos grasos de cadena larga que son un componente de la capa de glicolípidos presente en heterocistos de ese organismo (Black y Wolk, 1994, J. Bacteriol. 176, 2282-2292; Campbell *et al.*, 1997, Arch. Microbiol. 167, 251-258). Es probable que Het I active los dominios ACP de una proteína, Hgl E, presente en esa agrupación. Los dos dominios ACP de Hgl E tienen un alto grado de homología de secuencia con los dominios ACP encontrados en Orf A de *Schizochytrium*. SEQ ID NO: 34 representa la secuencia de aminoácidos de la proteína Het I de *Nostoc*, y es una PPTasa funcional que puede usarse con un sistema de PKS de AGPI descrito en el presente documento, incluyendo los sistemas de PKS de AGPI de *Schizochytrium* y *Thraustochytrium*. SEQ ID NO: 34 se codifica por SEQ ID NO: 33. El codón de iniciación endógeno de Het I no se ha identificado (no hay ninguna metionina presente en la proteína supuesta). Hay varios posibles codones de iniciación alternativos (por ejemplo, TTG y ATT) cerca del extremo 5' del marco de lectura abierto. Ningún codón de metionina (ATG) está presente en la secuencia. Sin embargo, se completó la construcción de un constructo de expresión de Het I usando PCR para sustituir el posible codón de iniciación alternativo en 5' más lejano (TTG) por un codón de metionina (ATG,

como parte de un sitio de reconocimiento de enzima de restricción NdeI), e introduciendo un sitio de XhoI en el extremo 3' de la secuencia codificante, y se ha mostrado que la PPTasa codificada (SEQ ID NO: 34) es funcional.

Otra PPTasa heteróloga que se ha demostrado anteriormente que reconoce los dominios ACP de OrfA descritos en el presente documento como sustratos es sfp, derivada de *Bacillus subtilis*. Sfp se ha caracterizado bien, y se usa ampliamente debido a su capacidad para reconocer una amplia gama de sustratos. Basándose en la información de secuencia publicada (Nakana, *et al.*, 1992, Molecular and General Genetics 232: 313-321), se produjo previamente un vector de expresión para sfp clonando la región codificante, junto con secuencias de ADN flanqueantes en el sentido de 5' y en el sentido de 3' definidas, en un vector de clonación pACYC-184. Este constructo codifica para una PPTasa funcional tal como se demostró por su capacidad para coexpresarse con Orf A, B* y C de *Schizochytrium* en *E. coli* lo que, en condiciones apropiadas, dio como resultado la acumulación de DHA en esas células (véase la publicación de solicitud de patente estadounidense n.º 20040235127).

Cuando se modifican genéticamente organismos (por ejemplo, microorganismos o plantas) para expresar un sistema de PKS de AGPI según la presente invención, algunos organismos huésped pueden expresar de manera endógena proteínas auxiliares que son necesarias para actuar con la PKS de AGPI para producir AGPI (por ejemplo, PPTasas). Sin embargo, algunos organismos pueden transformarse con moléculas de ácido nucleico que codifican para una o más proteínas auxiliares descritas en el presente documento para permitir y/o potenciar la producción de AGPI por el organismo, aunque el organismo produzca de manera endógena una proteína auxiliar homóloga (es decir, algunas proteínas auxiliares heterólogas pueden funcionar más eficaz o eficientemente con las proteínas AGPI sintasas transformadas que la proteína auxiliar endógena de las células huésped). En una realización, una proteína auxiliar de este tipo incluye una PPTasa auxiliar.

Se describe una molécula de ácido nucleico aislada que comprende una secuencia de ácido nucleico de un sistema de PKS de AGPI, un homólogo de la misma, un fragmento de la misma y/o una secuencia de ácido nucleico que es complementaria a cualquiera de tales secuencias de ácido nucleico. Se describe una molécula de ácido nucleico aislada que comprende una secuencia de ácido nucleico seleccionada del grupo que consiste en: (a) una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos seleccionada del grupo que consiste en: SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 39, SEQ ID NO: 52, SEQ ID NO: 62 y fragmentos biológicamente activos de las mismas; (b) una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos seleccionada del grupo que consiste en: SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68 y fragmentos biológicamente activos de las mismas; (c) una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos que es idéntica en al menos aproximadamente el 60% a al menos 500 aminoácidos consecutivos de cualquiera de las secuencias de aminoácidos de (a), en el que la secuencia de aminoácidos tiene una actividad biológica de al menos uno, dos, tres o más dominios de un sistema de poliketido sintasa (PKS) de ácidos grasos poliinsaturados (AGPI); (d) una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos que es idéntica en al menos aproximadamente el 60% a cualquiera de las secuencias de aminoácidos de (b), en la que dicha secuencia de aminoácidos tiene una actividad biológica de al menos un dominio de un sistema de poliketido sintasa (PKS) de ácidos grasos poliinsaturados (AGPI); o (e) una secuencia de ácido nucleico que es completamente complementaria a la secuencia de ácido nucleico de (a), (b), (c) o (d). También se describen secuencias de ácido nucleico que incluyen una secuencia que codifica para los dominios de sitio activo u otros motivos funcionales descritos anteriormente para varios de los dominios PKS de AGPI.

Se describen además moléculas de ácido nucleico aisladas que codifican para proteínas quiméricas útiles en un sistema de PKS de AGPI tal como se describe en el presente documento. La presente invención incluye el uso de cualquier dominio o proteína de o derivada de un sistema de PKS de AGPI en un dominio y/o con proteínas de o derivadas de otro sistema de PKS de AGPI con el fin de crear sistemas de PKS de AGPI novedosos con cualidades únicas.

Por ejemplo, una realización de la presente invención se refiere a el uso de un dominio DH2 de un sistema de PKS de AGPI para modificar un sistema de PKS de AGPI compuesto por proteínas/dominios de un organismo u organismos diferentes, en el que la introducción del dominio DH2 (por ejemplo, en una realización, mediante sustitución del dominio DH2 endógeno o dominio similar en el huésped) modifica la razón de AGPI producidos por el sistema, y particularmente la razón de AGPI omega-3 con respecto a omega-6 producidos por el sistema. Esta realización se describe en detalle a continuación.

Se describen moléculas de ácido nucleico que incluyen una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos de SEQ ID NO: 74, y fragmentos biológicamente activos de la misma, una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos que es idéntica en al menos aproximadamente el 60% a SEQ ID NO: 74 que tiene actividad biológica de al menos uno, dos, tres o más dominios de un sistema de poliketido sintasa (PKS) de ácidos grasos poliinsaturados (AGPI), o una secuencia de ácido nucleico que es completamente complementaria a las secuencias de ácido nucleico anteriores. La molécula de ácido nucleico puede incluir una secuencia de ácido nucleico seleccionada de SEQ ID NO: 73 y SEQ ID NO: 75. La molécula de ácido

nucleico puede incluir una secuencia de ácido nucleico que codifica para la secuencia de aminoácidos codificada por un plásmido seleccionado del grupo de pDS49 y pDD24. La molécula de ácido nucleico puede incluir también la secuencia de ácido nucleico de un plásmido seleccionado del grupo de pDS49 y pDD24 que codifica para una proteína OrfC quimérica.

También se describen moléculas de ácido nucleico que comprenden una secuencia de ácido nucleico que codifica para una proteína PKS de AGPI o dominio u homólogo de la misma de un sistema de PKS de AGPI, en el que la secuencia de ácido nucleico está optimizada para el uso de codones de un organismo diferente, tal como un huésped en el que va a expresarse la secuencia de ácido nucleico. En el presente documento se describen ejemplos de tales secuencias de ácido nucleico, e incluyen, pero no se limitan a, las secuencias de ácido nucleico representadas por SEQ ID NO: 70, SEQ ID NO: 71 y SEQ ID NO: 72, así como SEQ ID NO: 75. Se describen secuencias de ácido nucleico optimizadas para codones que codifican para cualquier proteína o dominio PKS de AGPI, y particularmente, cualquiera de las secuencias de aminoácidos descritas en el presente documento. Una molécula de ácido nucleico de este tipo incluye una secuencia de ácido nucleico que codifica para la secuencia de aminoácidos codificada por un plásmido seleccionado del grupo de pThOrfC-synPS, pDD26, pDD32 o pDD24. La molécula de ácido nucleico puede incluir la secuencia de ácido nucleico de un plásmido seleccionado de pThOrfC-synPS, pDD26, pDD32 o pDD24 que codifica para una proteína o proteína quimérica útil en un sistema de PKS de AGPI.

Según la presente invención, una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema de PKS de AGPI es una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio del sistema de PKS de AGPI descrito en detalle en el presente documento, tal como se muestra a modo de ejemplo por los sistemas de PKS de AGPI de *Schizochytrium* y *Thraustochytrium*, y tal como se muestra a modo de ejemplo adicionalmente por las actividades biológicas descritas de cualquiera de las proteínas y los dominios en cualquiera de los sistemas de PKS de AGPI descritos en la patente estadounidense n.º 6.140.486, la patente estadounidense n.º 6.566.583, la publicación de solicitud de patente estadounidense n.º 20020194641, la publicación de solicitud de patente estadounidense n.º 20070089199, la publicación de solicitud de patente estadounidense n.º 20040235127, la publicación de solicitud de patente estadounidense n.º 20050100995, la publicación de patente PCT n.º WO 05/097982 o la publicación de solicitud de patente estadounidense n.º 20050014231, citados anteriormente.

Por consiguiente, una molécula de ácido nucleico aislada descrita en el presente documento puede codificar para el producto de traducción de cualquier marco de lectura abierto de PKS de AGPI, dominio PKS de AGPI, fragmento biológicamente activo del mismo o cualquier homólogo de un dominio o marco de lectura abierto de PKS de AGPI que se produce de manera natural que tiene actividad biológica. Un homólogo de un dominio o una proteína dado es una proteína o un polipéptido que tiene una secuencia de aminoácidos que difiere de la secuencia de aminoácidos de referencia que se produce de manera natural (es decir, del dominio o la proteína de referencia) en que al menos uno o unos cuantos, pero sin limitarse a uno o unos cuantos, 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, mediante glicosilación, fosforilación, acetilación, miristoilación, prenilación, palmitación, amidación y/o adición de glicosilfosfatidilinositol). A continuación se describen en detalle homólogos preferidos de un dominio o una proteína PKS de AGPI. Se observa que los homólogos pueden incluir homólogos producidos de manera sintética, variantes alélicas que se producen de manera natural de un dominio o una proteína dado o secuencias homólogas de organismos distintos del organismo del que se derivó la secuencia de referencia.

En general, la actividad biológica o acción biológica de una proteína o un dominio se refiere a cualquier función presentada o realizada por la proteína o el dominio que se atribuye a la forma que se produce de manera natural de la proteína o el dominio tal como se mide o se observa *in vivo* (es decir, en el entorno fisiológico natural de la proteína) o *in vitro* (es decir, en condiciones de laboratorio). En otra parte en el presente documento se han descrito en detalle actividades biológicas de sistemas de PKS de AGPI y las proteínas/dominios individuales que constituyen un sistema de PKS de AGPI. Modificaciones de una proteína o un dominio, tal como en un homólogo o mimético (comentado a continuación), pueden dar como resultado proteínas o dominios que tienen la misma actividad biológica que la proteína o el dominio que se produce de manera natural, o proteínas o dominios que tienen actividad biológica disminuida o aumentada en comparación con la de la proteína o el dominio que se produce de manera natural. Las modificaciones que dan como resultado una disminución en la expresión o una disminución en la actividad de la proteína o el dominio pueden denominarse inactivación (completa o parcial), regulación por disminución o disminución de la acción de una proteína o un dominio. De manera similar, las modificaciones que dan como resultado un aumento en la expresión o un aumento en la actividad de la proteína o el dominio pueden denominarse amplificación, sobreproducción, activación, potenciación, regulación por incremento o aumento de la acción de una proteína o un dominio. Un dominio funcional de un sistema de PKS de AGPI es un dominio (es decir, un dominio puede ser una porción de una proteína) que puede realizar una función biológica (es decir, tiene actividad biológica).

Según la presente invención, una molécula de ácido nucleico aislada es una molécula de ácido nucleico que se ha retirado de su medio natural (es decir, que se ha sometido a manipulación humana), siendo su medio natural el genoma o cromosoma en el que se encuentra la molécula de ácido nucleico en la naturaleza. Como tal, "aislado" no

refleja necesariamente el grado en el que se ha purificado la molécula de ácido nucleico, sino que indica que la molécula no incluye un genoma completo o un cromosoma completo en el que se encuentra la molécula de ácido nucleico en la naturaleza. Una molécula de ácido nucleico aislada puede incluir un gen. Una molécula de ácido nucleico aislada que incluye un gen no es un fragmento de un cromosoma que incluye tal gen, sino que más bien incluye la región codificante y regiones reguladoras asociadas con el gen, pero normalmente ningún gen adicional encontrado de manera natural en el mismo cromosoma, aunque algunas moléculas de ácido nucleico pueden incluir genes cercanos/unidos que no son necesariamente una parte del sistema o gen de PKS de AGPI. Una molécula de ácido nucleico aislada también puede incluir una secuencia de ácido nucleico especificada flanqueada por (es decir, en el extremo 5' y/o el 3' de la secuencia) ácidos nucleicos adicionales que no flanquean normalmente la secuencia de ácido nucleico especificada en la naturaleza (es decir, secuencias heterólogas). Una molécula de ácido nucleico aislada puede incluir ADN, ARN (por ejemplo, ARNm) o derivados de o bien ADN o bien ARN (por ejemplo, ADNc). Aunque la expresión "molécula de ácido nucleico" se refiere principalmente a la molécula de ácido nucleico física 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 manera intercambiable, especialmente con respecto a una molécula de ácido nucleico, o una secuencia de ácido nucleico, que puede codificar para una proteína o un dominio de una proteína.

Preferiblemente, se produce una molécula de ácido nucleico aislada descrita en el presente documento usando tecnología de ADN recombinante (por ejemplo, amplificación por reacción en cadena de la polimerasa (PCR), clonación) o síntesis química. Las moléculas de ácido nucleico aisladas incluyen moléculas de ácido nucleico naturales y homólogos de las mismas, incluyendo, pero sin limitarse a, variantes alélicas naturales y moléculas de ácido nucleico modificadas en las que se han insertado, deletado, sustituido y/o invertido nucleótidos de tal manera que tales modificaciones proporcionan el efecto deseado sobre la actividad biológica del sistema de PKS de AGPI tal como se describe en el presente documento. Se han comentado en detalle anteriormente homólogos de proteínas (por ejemplo, proteínas codificadas por homólogos de ácido nucleico).

Puede producirse un homólogo de molécula de ácido nucleico usando varios métodos conocidos por los expertos en la técnica (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 variedad de técnicas incluyendo, pero sin limitarse a, técnicas de mutagénesis clásicas y técnicas de ADN recombinante, tales como mutagénesis dirigida al sitio, tratamiento químico de una molécula de ácido nucleico para inducir mutaciones, escisión con enzimas de restricción de un fragmento de ácido nucleico, ligamiento de fragmentos de ácido nucleico, amplificación por PCR y/o mutagénesis de regiones seleccionadas de una secuencia de ácido nucleico, síntesis de mezclas de oligonucleótidos y ligamientos de grupos de mezclas para "construir" una mezcla de moléculas de ácido nucleico y combinaciones de las mismas. Pueden seleccionarse homólogos de moléculas de ácido nucleico a partir de una mezcla de ácidos nucleicos modificados examinando para seleccionar la función de la proteína codificada por el ácido nucleico y/o mediante hibridación con un gen de tipo natural.

El tamaño mínimo de una molécula de ácido nucleico descrita en el presente documento es un tamaño suficiente para formar una sonda o un cebador oligonucleotídico que puede 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 para una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema de PKS de AGPI según la presente invención. Como tal, el tamaño de la molécula de ácido nucleico que codifica para una proteína de este tipo puede depender de la composición de 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 de sal y concentración de formamida). El tamaño mínimo de una molécula de ácido nucleico que se usa como cebador oligonucleotídico o como sonda es normalmente de al menos aproximadamente 12 a aproximadamente 15 nucleótidos de longitud si la 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 hay ningún límite, aparte de un límite práctico, en el tamaño máximo de una molécula de ácido nucleico descrita en el presente documento, ya que la molécula de ácido nucleico puede incluir una secuencia suficiente para codificar para un fragmento biológicamente activo de un dominio de un sistema de PKS de AGPI, un dominio completo de un sistema de PKS de AGPI, varios dominios dentro de un marco de lectura abierto (Orf) de un sistema de PKS de AGPI, un Orf completo de un sistema de PKS de AGPI o más de un Orf de un sistema de PKS de AGPI.

Se describe una molécula de ácido nucleico aislada que comprende, consiste esencialmente en o consiste en una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos seleccionada del grupo de: SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68 o SEQ ID NO: 74, o fragmentos biológicamente activos de las mismas. En un aspecto, la secuencia de ácido nucleico se selecciona de: SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 12, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID

NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73 o SEQ ID NO: 75.

- 5 Cualquiera de las secuencias de aminoácidos de PKS de AGPI descritas anteriormente, así como homólogos de tales secuencias, pueden producirse con desde al menos uno, y hasta aproximadamente 20, aminoácidos heterólogos adicionales que flanquean cada uno del extremo C y/o N-terminal de la secuencia de aminoácidos dada. Puede hacerse referencia a la proteína o al polipéptido resultante como “que consiste esencialmente en” una
- 10 secuencia de aminoácidos dada. Según la presente invención, los aminoácidos heterólogos son una secuencia de aminoácidos que no se encuentran de manera natural (es decir, no se encuentran en la naturaleza, *in vivo*) flanqueando la secuencia de aminoácidos dada o que no se codificaría por los nucleótidos que flanquean la secuencia de ácido nucleico que se produce de manera natural que codifica para la secuencia de aminoácidos dada tal como se produce en el gen, si tales nucleótidos en la secuencia que se produce de manera natural se tradujesen usando el uso de codones convencional para el organismo del que se deriva la secuencia de aminoácidos dada. De
- 15 manera similar, la expresión “que consiste esencialmente en”, cuando se usa con referencia a una secuencia de ácido nucleico en el presente documento, se refiere a una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos dada que puede estar flanqueada por al menos uno, y hasta aproximadamente 60, nucleótidos heterólogos adicionales en cada uno del extremo 5' y/o el 3' de la secuencia de ácido nucleico que codifica para la secuencia de aminoácidos dada. Los nucleótidos heterólogos no se encuentran de manera natural (es decir, no se encuentran en la naturaleza, *in vivo*) flanqueando la secuencia de ácido nucleico que codifica para la
- 20 secuencia de aminoácidos dada tal como se produce en el gen natural.

También se describe una molécula de ácido nucleico aislada que comprende una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema

25 de PKS de AGPI. En un aspecto, una secuencia de ácido nucleico de este tipo codifica para un homólogo de cualquiera de los dominios o las proteínas PKS de AGPI descritos anteriormente, en la que el homólogo tiene una actividad biológica de al menos un (o dos, tres, cuatro o más) dominio de un sistema de PKS de AGPI tal como se describió anteriormente en el presente documento.

- 30 En un aspecto de la invención, un homólogo de un dominio o una proteína PKS de AGPI abarcado por la presente invención comprende una secuencia de aminoácidos que es idéntica en al menos aproximadamente el 60% a al menos 500 aminoácidos consecutivos de una secuencia de aminoácidos elegida de: SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 39, SEQ ID NO: 52, SEQ ID NO: 62 o SEQ ID NO: 74; en el que dicha secuencia de aminoácidos tiene una actividad biológica de al menos un dominio de un sistema de PKS de AGPI. En un aspecto
- 35 adicional, la secuencia de aminoácidos del homólogo es idéntica en al menos aproximadamente el 60% a al menos aproximadamente 600 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 700 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 800 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 900 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1000 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1100 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1200 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1300 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1400 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1500 aminoácidos consecutivos de cualquiera de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 39, SEQ ID NO: 52, SEQ ID NO: 62 o SEQ ID NO: 74, o a la longitud completa de SEQ ID NO: 6; SEQ
- 40 ID NO: 62 o SEQ ID NO: 74. En un aspecto adicional, la secuencia de aminoácidos del homólogo es idéntica en al menos aproximadamente el 60% a al menos aproximadamente 1600 aminoácidos consecutivos; y más preferiblemente a al menos aproximadamente 1700 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1800 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 1900 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2000 aminoácidos consecutivos de cualquiera de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 39 o SEQ ID NO: 52 o a la longitud completa de SEQ ID NO: 4 o SEQ ID NO: 52. En un aspecto adicional, la secuencia de aminoácidos del homólogo es idéntica en al menos aproximadamente el 60% a al menos aproximadamente 2100 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2200 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2300 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2400
- 45 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2500 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2600 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2700 aminoácidos consecutivos, y más preferiblemente a al menos aproximadamente 2800 aminoácidos consecutivos, e incluso más preferiblemente, a la longitud completa de SEQ ID NO: 2 o SEQ ID NO: 39.

- 60 En otro aspecto, un homólogo de un dominio o una proteína PKS de AGPI abarcado por la presente invención comprende una secuencia de aminoácidos que es idéntica en al menos aproximadamente el 65%, y más preferiblemente idéntica en al menos aproximadamente el 70%, y más preferiblemente idéntica en al menos aproximadamente el 75%, y más preferiblemente idéntica en al menos aproximadamente el 80%, y más preferiblemente idéntica en al menos aproximadamente el 85%, y más preferiblemente idéntica en al menos aproximadamente el 90%, y más preferiblemente idéntica en al menos aproximadamente el 95%, y más
- 65

preferiblemente idéntica en al menos aproximadamente el 96%, y más preferiblemente idéntica en al menos aproximadamente el 97%, y más preferiblemente idéntica en al menos aproximadamente el 98%, y más preferiblemente idéntica en al menos aproximadamente el 99% a cualquiera de las secuencias de aminoácidos descritas anteriormente, a lo largo de cualquiera de las longitudes de aminoácido consecutivos descritas en los párrafos anteriores, en el que la secuencia de aminoácidos tiene una actividad biológica de al menos un dominio de un sistema de PKS de AGPI.

En un aspecto de la invención, un homólogo de un dominio o una proteína PKS de AGPI abarcado por la presente invención comprende una secuencia de aminoácidos que es idéntica en al menos aproximadamente el 60% a una secuencia de aminoácidos elegida de: SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, o secuencias de aminoácidos que comprenden combinaciones de cualquiera de tales secuencias de aminoácidos, en el que dicha secuencia de aminoácidos tiene una actividad biológica de al menos un dominio de un sistema de PKS de AGPI o proteína auxiliar del mismo. En un aspecto adicional, la secuencia de aminoácidos del homólogo es idéntica en al menos aproximadamente el 65%, y más preferiblemente idéntica en al menos aproximadamente el 70%, y más preferiblemente idéntica en al menos aproximadamente el 75%, y más preferiblemente idéntica en al menos aproximadamente el 80%, y más preferiblemente idéntica en al menos aproximadamente el 85%, y más preferiblemente idéntica en al menos aproximadamente el 90%, y más preferiblemente idéntica en al menos aproximadamente el 95%, y más preferiblemente idéntica en al menos aproximadamente el 96%, y más preferiblemente idéntica en al menos aproximadamente el 97%, y más preferiblemente idéntica en al menos aproximadamente el 98%, y más preferiblemente idéntica en al menos aproximadamente el 99% a cualquiera de las secuencias de aminoácidos descritas anteriormente, en el que la secuencia de aminoácidos tiene una actividad biológica de al menos un dominio de un sistema de PKS de AGPI o proteína auxiliar del mismo.

Según la presente invención, el término “contiguo” o “consecutivo”, con respecto a secuencias de aminoácidos o de ácido nucleico descritas en el presente documento, significa que están conectadas en una secuencia ininterrumpida. Por ejemplo, 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 residuos de aminoácido que es idéntica al 100% a una secuencia ininterrumpida de 30 residuos de aminoácido en la segunda secuencia. De manera similar, que una primera secuencia tenga una “identidad del 100%” con una segunda secuencia significa que la primera secuencia coincide exactamente con la segunda secuencia sin huecos entre nucleótidos o aminoácidos.

Tal como se usa en el presente documento, a menos que se especifique lo contrario, la referencia a un porcentaje (%) de identidad se refiere a una evaluación de la homología que se realiza usando: (1) una búsqueda de homología BLAST 2.0 Basic BLAST usando blastp para búsquedas de aminoácidos, blastn para búsquedas de ácidos nucleicos y blastX para búsquedas de ácidos nucleicos y búsquedas de aminoácidos traducidos en los 6 marcos de lectura abiertos, todos con parámetros por defecto convencionales, en la que la secuencia de consulta se filtra para detectar regiones de baja complejidad por defecto (descrito en Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”, *Nucleic Acids Res.* 25:3389-3402); (2) una alineación BLAST 2 (usando los parámetros descritos a continuación); (3) y/o PSI-BLAST con los parámetros por defecto convencionales (BLAST con iteraciones específicas de posición). Se observa que debido a algunas diferencias en los parámetros convencionales entre BLAST 2.0 Basic BLAST y BLAST 2, podría reconocerse que dos secuencias específicas tienen 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 coincidencias superiores. 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 realiza en primer lugar una búsqueda en la base de datos BLAST con huecos. El programa PSI-BLAST usa la información de cualquier alineación significativa devuelta para construir una matriz de puntuación específica de posición, que reemplaza a la secuencia de consulta para la siguiente ronda de búsqueda en la base de datos. Por tanto, debe entenderse que el porcentaje de identidad puede determinarse usando uno cualquiera de estos programas.

Pueden alinearse dos secuencias específicas entre sí usando BLAST 2 Sequence tal como se describe en Tatusova y Madden, (1999), “Blast 2 sequence - a new tool for comparing protein and nucleotide sequence”, *FEMS Microbiol Lett.* 174:247-250. La alineación de secuencias BLAST 2 se realiza en blastp o blastn usando el algoritmo BLAST 2.0 para realizar una búsqueda en BLAST con huecos (BLAST 2.0) entre las dos secuencias permitiendo la introducción de huecos (deleciones e inserciones) en la alineación resultante. Con fines de mayor claridad en el presente documento, se realiza una alineación de secuencias BLAST 2 usando los parámetros por defecto convencionales tal como sigue.

Para blastn, usando la matriz 0 BLOSUM62:

Recompensa por coincidencia = 1

Penalización por coincidencia errónea = -2

Penalizaciones por apertura de hueco (5) y extensión de extensión (2)

5 hueco x_dropoff (50) expectativa (10) tamaño de palabra (11) filtro (activado)

Para blastp, usando la matriz 0 BLOSUM62:

Penalizaciones por apertura de hueco (11) y extensión de hueco (1)

10 hueco x_dropoff (50) expectativa (10) tamaño de palabra (3) filtro (activado).

En otra realización de la invención, una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio de un sistema de PKS de AGPI de la presente invención incluye una secuencia de aminoácidos que es lo suficientemente similar a un polipéptido o una proteína PKS de AGPI que se produce de manera natural como para que una secuencia de ácido nucleico que codifica para la secuencia de aminoácidos pueda hibridarse en condiciones de rigurosidad moderada, alta o muy alta (descritas a continuación) a (es decir, con) una molécula de ácido nucleico que codifica para el polipéptido o la proteína PKS de AGPI que se produce de manera natural (es decir, al complemento de la hebra de ácido nucleico que codifica para el polipéptido o la proteína PKS de AGPI que se produce de manera natural). Preferiblemente, una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio de un sistema de PKS de AGPI de la presente invención se codifica por una secuencia de ácido nucleico que se hibrida en condiciones de rigurosidad moderada, alta o muy alta al complemento de una secuencia de ácido nucleico que codifica para una proteína que comprende una secuencia de aminoácidos representada por cualquiera de las secuencias de aminoácidos descritas en el presente documento.

En otra realización de la invención, una secuencia de nucleótidos de la presente invención es una secuencia de nucleótidos aislada de (que puede obtenerse de), idéntica a, o un homólogo de, la secuencia de nucleótidos de un *Schizochytrium*, en la que la secuencia de nucleótidos de un *Schizochytrium* (incluyendo cualquier hebra de una molécula de ADN de *Schizochytrium*) se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos que codifica para una secuencia de aminoácidos representada por cualquiera de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30 o SEQ ID NO: 32. En una realización, el *Schizochytrium* es *Schizochytrium* ATCC 20888. En otra realización, el *Schizochytrium* es una cepa hija de *Schizochytrium* 20888, incluyendo cepas mutadas de la misma (por ejemplo, N230D). En una realización, la secuencia de ácido nucleico se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos seleccionada de: SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 12, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29 o SEQ ID NO: 31.

En otra realización de la invención, una secuencia de nucleótidos de la presente invención es una secuencia de nucleótidos aislada de (que puede obtenerse de), idéntica a, o un homólogo de, la secuencia de nucleótidos de un *Thraustochytrium*, en la que la secuencia de nucleótidos de un *Thraustochytrium* (incluyendo cualquier hebra de una molécula de ADN de *Thraustochytrium*) se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos que codifica para una secuencia de aminoácidos representada por cualquiera de SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68. En una realización, el *Thraustochytrium* es *Thraustochytrium* 23B (ATCC 20892). En una realización, la secuencia de ácido nucleico se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos seleccionada de: SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 65 o SEQ ID NO: 67.

Aún en otra realización, una secuencia de nucleótidos de la presente invención es una secuencia de nucleótidos aislada de (que puede obtenerse de), idéntica a, o un homólogo de, la secuencia de nucleótidos de un organismo eucariota (por ejemplo, un traustochytridio o un laberintúlido) o una bacteria marina, en la que la secuencia de nucleótidos se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos que codifica para cualquiera de las secuencias de aminoácidos representadas en el presente documento.

En otra realización, una secuencia de nucleótidos de la presente invención es una secuencia de nucleótidos aislada de (que puede obtenerse de), idéntica a, o un homólogo de, cualquier secuencia de nucleótidos que codifica para una proteína auxiliar descrita en el presente documento (incluyendo cualquier hebra de una molécula de ADN), en la que, en una realización, la secuencia de nucleótidos se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos que codifica para una secuencia de aminoácidos representada por SEQ ID NO: 34. En una realización, la secuencia de ácido nucleico se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos representada por SEQ ID NO: 33.

En otra realización, una secuencia de nucleótidos de la presente invención es una secuencia de nucleótidos aislada de (que puede obtenerse de), idéntica a, o un homólogo de, cualquier secuencia de nucleótidos química u optimizada para codones descrita en el presente documento (incluyendo cualquier hebra de una molécula de ADN), en la que, en una realización, la secuencia de nucleótidos se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos que codifica para una secuencia de aminoácidos representada por SEQ ID NO: 74. En una realización, la secuencia de ácido nucleico se hibrida en condiciones de rigurosidad moderada, alta o muy alta a una secuencia de nucleótidos seleccionada de SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 69, SEQ ID NO: 70, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73 o SEQ ID NO: 75.

Los expertos en la técnica conocen métodos para deducir una secuencia complementaria. Debe indicarse que puesto que las tecnologías de secuenciación de aminoácidos y secuenciación de ácido nucleico no están completamente libres de errores, las secuencias presentadas en el presente documento, en el mejor de los casos, representan secuencias aparentes de dominios y proteínas PKS de AGPI de la presente invención, o de las secuencias de nucleótidos que codifican para tales secuencias de aminoácidos.

Tal como se usa en el presente documento, las condiciones de hibridación se refieren a condiciones de hibridación convencionales en las que se usan moléculas de ácido nucleico para identificar moléculas de ácido nucleico similares. Tales condiciones convencionales se dan a conocer, por ejemplo, en Sambrook *et al.*, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Labs Press, 1989 (véanse específicamente las páginas 9.31-9.62). Además, se dan a conocer fórmulas para calcular las condiciones de hibridación y lavado apropiadas para lograr una hibridación que permite grados variables de apareamiento erróneo de nucleótidos, por ejemplo, en Meinkoth *et al.*, 1984, Anal. Biochem. 138, 267-284.

Más particularmente, las condiciones de hibridación y lavado de rigurosidad moderada, tal como se denominan en el presente documento, se refieren a condiciones que permiten el aislamiento de moléculas de ácido nucleico que tienen al menos aproximadamente el 70% de identidad de secuencia de ácido nucleico con la molécula de ácido nucleico que está usándose para estudiar con sonda en la reacción de hibridación (es decir, condiciones que permiten aproximadamente el 30% o menos de apareamiento erróneo de nucleótidos). Las condiciones de hibridación y lavado de alta rigurosidad, tal como se denominan en el presente documento, se refieren a condiciones que permiten el aislamiento de moléculas de ácido nucleico que tienen al menos aproximadamente el 80% de identidad de secuencia de ácido nucleico con la molécula de ácido nucleico que está usándose para estudiar con sonda en la reacción de hibridación (es decir, condiciones que permiten aproximadamente el 20% o menos de apareamiento erróneo de nucleótidos). Las condiciones de hibridación y lavado de rigurosidad muy alta, tal como se denominan en el presente documento, se refieren a condiciones que permiten el aislamiento de moléculas de ácido nucleico que tienen al menos aproximadamente el 90% de identidad de secuencia de ácido nucleico con la molécula de ácido nucleico que está usándose para estudiar con sonda en la reacción de hibridación (es decir, condiciones que permiten aproximadamente el 10% o menos de apareamiento erróneo de nucleótidos). Tal como se comentó anteriormente, un experto en la técnica puede usar las fórmulas en Meinkoth *et al.*, *ibidem.* para calcular las condiciones de hibridación y lavado apropiadas para lograr estos niveles particulares de apareamiento erróneo de nucleótidos. Tales condiciones variarán, dependiendo de si están formándose híbridos de ADN:ARN o ADN:ADN. Las temperaturas de fusión calculadas para híbridos de ADN:ADN son de 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 6X SSC (Na⁺ 0,9 M) a una temperatura de entre aproximadamente 20°C y aproximadamente 35°C (rigurosidad inferior), más preferiblemente, entre aproximadamente 28°C y aproximadamente 40°C (más rigurosa) e incluso más preferiblemente, entre aproximadamente 35°C y aproximadamente 45°C (incluso más rigurosa), con condiciones de lavado apropiadas. En realizaciones particulares, las condiciones de hibridación rigurosas para híbridos de ADN:ARN incluyen hibridación a una fuerza iónica de 6X SSC (Na⁺ 0,9 M) a una temperatura de entre aproximadamente 30°C 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 rigurosas de manera similar. Estos valores se basan en cálculos de una temperatura de fusión para moléculas de más de aproximadamente 100 nucleótidos, formamida al 0% y un contenido en G + C de aproximadamente el 40%. Alternativamente, la T_m puede calcularse empíricamente tal como se expone en Sambrook *et al.*, citado anteriormente, páginas 9.31 a 9.62. En general, las condiciones de lavado deben ser tan rigurosas como sea posible, y deben ser apropiadas para las condiciones de hibridación elegidas. Por ejemplo, las condiciones de hibridación pueden incluir una combinación de condiciones de sal y temperatura que están aproximadamente 20-25°C por debajo de la T_m calculada de un híbrido particular, y las condiciones de lavado incluyen normalmente una combinación de condiciones de sal y temperatura que están aproximadamente 12-20°C por debajo de la T_m calculada del híbrido particular. Un ejemplo de condiciones de hibridación adecuadas para su uso con híbridos de ADN:ADN incluye una hibridación de 2-24 horas en 6X SSC (formamida al 50%) a aproximadamente 42°C, seguido por etapas de lavado que incluyen uno o más lavados a temperatura ambiente en aproximadamente 2X SSC, seguido por lavados adicionales a temperaturas superiores y fuerza iónica inferior (por ejemplo, al menos un lavado a aproximadamente 37°C en aproximadamente 0,1X-0,5X SSC, seguido por al menos un lavado a aproximadamente 68°C en aproximadamente 0,1X-0,5X SSC).

Se describe además una molécula de ácido nucleico que comprende, que consiste esencialmente en o que consiste en, una secuencia de ácido nucleico que es idéntica a, o que es un homólogo de (tal como se definió anteriormente),

la secuencia de ácido nucleico de un plásmido seleccionado de: pJK1126 (n.º de registro de la ATCC PTA-7648), pJK1129 (n.º de registro de la ATCC PTA-7649), pJK1131 (n.º de registro de la ATCC PTA-7650), pJK306 (n.º de registro de la ATCC PTA-7641), pJK320 (n.º de registro de la ATCC PTA-7644), pJK324 (n.º de registro de la ATCC PTA-7643), pBR002 (n.º de registro de la ATCC PTA-7642), Th23BOrfA_pBR812.1 (n.º de registro de la ATCC PTA-8232) Th23BOrfA_pBR811 (n.º de registro de la ATCC PTA-8231), Th23BOrfB_pBR800 (n.º de registro de la ATCC PTA-8227) o Th23BOrfC_pBR709A (n.º de registro de la ATCC PTA-8228).

También se describe una molécula de ácido nucleico que comprende, que consiste esencialmente en o que consiste en, una secuencia de ácido nucleico que es idéntica a, o que es a homólogo de (tal como se definió anteriormente), la secuencia de ácido nucleico de un plásmido seleccionado de: pThOrfC-synPS (n.º de registro de la ATCC PTA-8229), pDS49 (n.º de registro de la ATCC PTA-8230), pDD24 (n.º de registro de la ATCC PTA-8226), pDD26 (n.º de registro de la ATCC PTA-8411), pDD32 (n.º de registro de la ATCC PTA-8412) o OrfB*_pJK780 (n.º de registro de la ATCC PTA-8225).

También se describe una molécula de ácido nucleico que comprende, que consiste esencialmente en o que consiste en, una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos que es idéntica a, o que es un homólogo de (tal como se definió anteriormente), la secuencia de aminoácidos codificada por un plásmido seleccionado de: pJK1126 (n.º de registro de la ATCC PTA-7648), pJK1129 (n.º de registro de la ATCC PTA-7649), pJK1131 (n.º de registro de la ATCC PTA-7650), pJK306 (n.º de registro de la ATCC PTA-7641), pJK320 (n.º de registro de la ATCC PTA-7644), pJK324 (n.º de registro de la ATCC PTA-7643), pBR002 (n.º de registro de la ATCC PTA-7642), Th23BOrfA_pBR812.1 (n.º de registro de la ATCC PTA-8232) Th23BOrfA_pBR811 (n.º de registro de la ATCC PTA-8231), Th23BOrfB_pBR800 (n.º de registro de la ATCC PTA-8227) o Th23BOrfC_pBR709A (n.º de registro de la ATCC PTA-8228).

También se describe una molécula de ácido nucleico que comprende, que consiste esencialmente en o que consiste en, una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos que es idéntica a, o que es un homólogo de (tal como se definió anteriormente), la secuencia de aminoácidos codificada por un plásmido seleccionado de: pThOrfC-synPS (n.º de registro de la ATCC PTA-8229), pDS49 (n.º de registro de la ATCC PTA-8230), pDD24 (n.º de registro de la ATCC PTA-8226), pDD26 (n.º de registro de la ATCC PTA-8411), pDD32 (n.º de registro de la ATCC PTA-8412) o OrfB*_pJK780 (n.º de registro de la ATCC PTA-8225).

Se describe además 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 para una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio o una proteína de un sistema de PKS de AGPI tal como se describe en el presente documento. Tales secuencias de ácido nucleico y dominios o proteínas se describieron en detalle anteriormente. Según la presente invención, un vector recombinante es una molécula de ácido nucleico diseñada por ingeniería genética (es decir, producida artificialmente) que se usa como herramienta para manipular una secuencia de ácido nucleico de elección y para introducir una secuencia de ácido nucleico de este tipo en una célula huésped. El vector recombinante es por tanto adecuado para su uso en clonación, secuenciación y/o manipulación de otra forma de la secuencia de ácido nucleico de elección, tal como expresando y/o suministrando la secuencia de ácido nucleico de elección a una célula huésped para formar una célula recombinante. Un vector de este tipo contiene normalmente secuencias de ácido nucleico heterólogas, es decir secuencias de ácido nucleico que no se encuentran de manera natural adyacentes a una secuencia de ácido nucleico que va a clonarse o suministrarse, aunque el vector también puede contener secuencias de ácido nucleico reguladoras (por ejemplo, promotores, regiones no traducidas) que se encuentran de manera 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 (comentado en detalle a continuación). El vector puede ser o bien ARN o bien ADN, o bien procariota o bien eucariota, y normalmente 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). Todo el vector puede permanecer en su lugar dentro de una célula huésped, o en determinadas condiciones, el ADN de plásmido puede deleccionarse, dejando atrás la molécula de ácido nucleico de la presente invención. La molécula de ácido nucleico integrada puede estar bajo el control de un promotor cromosómico, bajo el control de un promotor de plásmido o nativo o bajo una combinación de varios controles de promotor. Pueden integrarse copias individuales o múltiples de la molécula de ácido nucleico en el cromosoma. Un vector recombinante de la presente invención puede contener al menos un marcador seleccionable.

Un vector recombinante usado en una molécula de ácido nucleico recombinante puede ser un vector de expresión. Tal como se usa en el presente documento, la expresión "vector de expresión" se usa para referirse 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 para el producto que va a producirse (por ejemplo, un dominio PKS de AGPI) se inserta en el vector recombinante para producir una molécula de ácido nucleico recombinante. La secuencia de ácido nucleico que codifica para la proteína que va a producirse se inserta en el vector de una manera que une operativamente la secuencia de ácido nucleico a secuencias reguladoras en el vector que permiten la transcripción y traducción de la secuencia de ácido nucleico dentro de la célula huésped recombinante.

El vector recombinante usado en una molécula de ácido nucleico recombinante también puede ser un vector de direccionamiento. Tal como se usa en el presente documento, la expresión "vector de direccionamiento" se usa para referirse a un vector que se usa para suministrar una molécula de ácido nucleico particular a una célula huésped recombinante, en el que la molécula de ácido nucleico se usa para deleccionar o inactivar un gen endógeno dentro del microorganismo o la células huésped (es decir, se usa para la alteración genética dirigida o tecnología de inactivación). Un vector de este tipo también puede conocerse en la técnica como vector de "inactivación". En un aspecto de esta realización, una porción del vector, pero más normalmente, 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 huésped (es decir, un gen que se selecciona como diana para deleccionarse o inactivarse). La secuencia de ácido nucleico del inserto del vector está diseñada para unirse al gen diana de manera que el gen diana y el inserto experimentan recombinación homóloga, mediante lo cual el gen diana endógeno se delecciona, se inactiva o se atenúa (es decir, mutando o deleccionando al menos una porción del gen diana endógeno).

Normalmente, una molécula de ácido nucleico recombinante incluye al menos una molécula de ácido nucleico descrita en el presente documento operativamente unida a una o más secuencias de control de la transcripción. Tal como se usa en el presente 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 operativamente unida a una secuencia de control de la transcripción, pero puede usarse de manera intercambiable con la expresión "molécula de ácido nucleico", cuando tal molécula de ácido nucleico es una molécula recombinante tal como se comenta en el presente documento. Según la presente invención, la expresión "operativamente unida" se refiere a la unión de una molécula de ácido nucleico a una secuencia de control de la transcripción de una manera tal que la molécula puede expresarse cuando se transfecta (es decir, se transforma, se transduce, se transfecta, se conjuga o se conduce) al interior de una célula huésped. Las secuencias de control de la transcripción son secuencias que controlan la iniciación, la elongación o la terminación de la transcripción. Secuencias de control de la transcripción particularmente importantes son las que controlan la iniciación 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 o un organismo huésped en el que va a introducirse la molécula de ácido nucleico recombinante.

Las moléculas de ácido nucleico recombinantes descritas en el presente documento 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. Una molécula recombinante descrita en el presente documento, incluyendo las que se integran en el cromosoma de la célula huésped, también puede contener señales secretoras (es decir, secuencias de ácido nucleico de segmento señal) para permitir que una proteína expresada secrete de la célula que produce la proteína. Los segmentos señal adecuados incluyen un segmento señal que está asociado de manera natural con la proteína que va a expresarse o cualquier segmento señal heterólogo que pueda dirigir la secreción de la proteína según la presente invención. Una molécula recombinante descrita en el presente documento también puede comprender una secuencia líder para permitir que una proteína expresada se suministre a, y se inserte en, la membrana de una célula huésped. Las secuencias líder adecuadas incluyen una secuencia líder que está asociada de manera natural con la proteína, o cualquier secuencia líder heteróloga que pueda dirigir el suministro y la inserción de la proteína en la membrana de una célula.

Los presentes inventores han encontrado que los Orf A y B de PKS de AGPI de *Schizochytrium* y *Thraustochytrium* están estrechamente unidos en el genoma y se ha secuenciado la región entre los Orf. En *Schizochytrium*, los Orf están orientados en direcciones opuestas y 4244 pares de bases separan los codones de iniciación (ATG) (es decir, están dispuestos tal como sigue: 3'OrfA5'-4244 pb-5'OrfB3'). El examen de la región intergénica de 4244 pb no reveló ningún Orf obvio (no se encontraron coincidencias significativas en una búsqueda de BlastX). Tanto el Orf A como el B se expresan altamente en *Schizochytrium*, al menos durante el tiempo de producción de aceite, lo que implica que están incrustados elementos promotores activos en esta región intergénica. Se cree que estos elementos genéticos tienen utilidad como secuencia promotora bidireccional para aplicaciones transgénicas. Por ejemplo, podría clonarse esta región, colocar cualquier gen de interés en cada extremo e introducir el constructo en *Schizochytrium* (o algún otro huésped en el que pueda mostrarse que los promotores funcionan). Se predice que los elementos reguladores, en las condiciones apropiadas, proporcionarían expresión de alto nivel, coordinada, de los dos genes introducidos. La secuencia de nucleótidos completa para la región reguladora que contiene elementos reguladores de PKS de AGPI de *Schizochytrium* (por ejemplo, un promotor) se representa en el presente documento como SEQ ID NO: 76.

De una manera similar, OrfC se expresa altamente en *Schizochytrium* durante el tiempo de producción de aceite y se espera que residan elementos reguladores en la región en el sentido de 5' de su codón de iniciación. Se ha clonado y secuenciado una región de ADN genómico en el sentido de 5' del OrfC y se representa en el presente documento como (SEQ ID NO: 77). Esta secuencia contiene los 3886 nt inmediatamente en el sentido de 5' del codón de iniciación de OrfC. El examen de esta región no reveló ningún Orf obvio (es decir, no se encontraron coincidencias significativas en una búsqueda de 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 los mismos. Adicionalmente, en las condiciones apropiadas, puede coordinarse el nivel de expresión con genes bajo el

control de la región intergénica A-B (SEQ ID NO: 76).

Por tanto, una molécula de ácido nucleico recombinante, tal como se da a conocer en el presente documento, puede incluir una región reguladora de PKS de AGPI contenida dentro de SEQ ID NO: 76 y/o SEQ ID NO: 77. Una región reguladora de este tipo puede incluir cualquier porción (fragmento) de SEQ ID NO: 76 y/o SEQ ID NO: 77 que tiene al menos actividad transcripcional de PKS de AGPI basal (al menos actividad promotora basal).

Pueden usarse una o más moléculas recombinantes descritas en el presente documento para producir un producto codificado (por ejemplo, un dominio, una proteína o un sistema de PKS de AGPI) de la presente invención. Puede producirse un producto codificado expresando una molécula de ácido nucleico tal como se describe en el presente documento en condiciones eficaces para producir la proteína. Un método preferido para producir una proteína codificada es transfectando una célula huésped con una o más moléculas recombinantes para formar una célula recombinante. Las células huésped adecuadas para transfectar incluyen, pero no se limitan a, cualquier célula bacteriana, fúngica (por ejemplo, levadura), de insecto, vegetal o animal que pueda transfectarse. Las células huésped pueden ser o bien células no transfectadas o bien células que ya están transfectadas con al menos otra molécula de ácido nucleico recombinante.

Según la presente invención, el término “transfección” se usa para referirse a cualquier método mediante el cual puede insertarse una molécula de ácido nucleico exógeno (es decir, una molécula de ácido nucleico recombinante) en una célula. El término “transformación” puede usarse de manera intercambiable con el término “transfección” cuando tal término se usa para referirse a la introducción de moléculas de ácido nucleico en células microbianas, tales como algas, bacterias y levadura. En sistemas microbianos, el término “transformación” se usa para describir un cambio heredado debido a la adquisición de ácidos nucleicos exógenos por el microorganismo y es esencialmente sinónimo al término “transfección.” Sin embargo, en células animales, transformación ha adquirido un segundo significado que puede referirse a cambios en las propiedades de crecimiento de células en cultivo tras volverse cancerosas, por ejemplo. Por tanto, para evitar la confusión, 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 el presente documento para abarcar generalmente la transfección de células animales, células de plantas y la transformación de células microbianas, en la medida en que los términos se refieran a la introducción de ácidos nucleicos exógenos en una célula. Por tanto, las técnicas de transfección incluyen, pero no se limitan a, transformación, bombardeo con partículas, electroporación, microinyección, lipofección, adsorción, infección y fusión de protoplastos.

Un experto en la técnica apreciará que el uso de tecnologías de ADN recombinante puede mejorar el control de la expresión de moléculas de ácido nucleico transfectadas manipulando, por ejemplo, el número de copias de las moléculas de ácido nucleico dentro de la célula huésped, la eficacia con la que se transcriben esas moléculas de ácido nucleico, la eficacia con la que se traducen los transcritos resultantes y la eficacia de las modificaciones postraduccionales. Adicionalmente, la secuencia promotora puede 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, pero no se limitan a, integración de la moléculas de ácido nucleico en uno o más cromosomas de la célula huésped, adición de secuencias de estabilidad de vector a plásmidos, sustituciones o modificaciones de señales de control de la transcripción (por ejemplo, promotores, operadores, potenciadores), sustituciones o modificaciones de 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 se correspondan con el uso de codones de la célula huésped y delección de secuencias que desestabilizan los transcritos.

Se pretende que la discusión general anterior con respecto a moléculas de ácido nucleico recombinantes y la transfección de células huésped se aplique a cualquier molécula de ácido nucleico recombinante comentada en el presente documento, incluyendo las que codifican para cualquier secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un PKS de AGPI, las que codifican para secuencias de aminoácidos de otros sistemas de PKS y las que codifican para otras proteínas o dominios.

Esta invención también se refiere a sistemas de PKS de AGPI (y proteínas o dominios de los mismos) de microorganismos distintos de los descritos específicamente en el presente documento que son homólogos en estructura, organización de dominios y/o función a cualquiera del sistema de PKS de AGPI (y proteínas o dominios del mismo) tal como se describe en el presente documento. Además, esta invención se refiere al uso de estos microorganismos y los sistemas de PKS de AGPI o componentes de los mismos (por ejemplo, dominios DH2) de estos microorganismos en las diversas aplicaciones para un sistema de PKS de AGPI (por ejemplo, organismos modificados genéticamente y métodos de producción de moléculas bioactivas) según la presente invención. En la publicación de solicitud de patente estadounidense n.º 20020194641, citada anteriormente, se describe en detalle un procedimiento de examen para la identificación de microorganismos que comprenden un sistema de PKS de AGPI. El conocimiento de la estructura y función de las proteínas y los dominios PKS de AGPI descritos en el presente documento, y la secuencia de nucleótidos que codifica para el mismo, son herramientas útiles para la identificación, confirmación y/o aislamiento de homólogos de tales proteínas o polinucleótidos.

Según la presente invención, el término “traustoquitridio” se refiere a cualquier miembro del orden Thraustochytriales, que incluye la familia Thraustochytriaceae, y el término “laberintúlido” se refiere a cualquier miembro del orden Labyrinthulales, que incluye la familia Labyrinthulaceae. En algún momento se consideró que los miembros de la familia Labyrinthulaceae eran miembros del orden Thraustochytriales, pero en revisiones más recientes de la taxonomía de tales organismos, se considera ahora que la familia es un miembro del orden Labyrinthulales, y se considera que tanto Labyrinthulales como Thraustochytriales son miembros del filo Labyrinthulomycota. Los desarrollos han dado como resultado una revisión frecuente de la taxonomía de los traustoquitridios y laberintúlidos. Sin embargo, los teóricos de la taxonomía colocan ahora generalmente a ambos de estos grupos de microorganismos con las algas o protistas similares a las algas dentro del linaje de Stramenopila. La colocación taxonómica actual de los traustoquitridios y laberintúlidos puede resumirse tal como sigue:

Reino: Stramenopila (Chromista)

Filo: Labyrinthulomycota

Clase: Labyrinthulomycetes

Orden: Labyrinthulales

Familia: Labyrinthulaceae

Orden: Thraustochytriales

Familia: Thraustochytriaceae

Sin embargo, debido a las incertidumbres taxonómicas que quedan, lo mejor para los fines de la presente invención sería considerar las cepas descritas en la presente invención como traustoquitridios incluyendo los siguientes organismos: Orden: Thraustochytriales; Familia: Thraustochytriaceae; Géneros: *Thraustochytrium* (especies: sp., *arudimentale*, *aureum*, *benthicola*, *globosum*, *kinnei*, *motivum*, *multirudimentale*, *pachydermum*, *proliferum*, *roseum*, *striatum*), *Ulkenia* (especies: sp., *amoeboidea*, *kerguelensis*, *minuta*, *profunda*, *radiata*, *sailens*, *sarkariana*, *schizochytrids visurgensis*, *yorkensis*), *Schizochytrium* (especies: sp., *aggregatum*, *limnaceum*, *mangrovei*, *minutum*, *octosporum*), *Japonochytrium* (especie: sp., *marinum*), *Aplanochytrium* (especies: sp., *haliotidis*, *kerguelensis*, *profunda*, *stocchinoi*), *Althornia* (especie: sp., *crouchii*) o *Elina* (especies: sp., *marisalba*, *sinorifica*). Debe indicarse que la descripción original del género *Ulkenia* no se publicó en una revista revisada por expertos de modo que siguen existiendo algunas cuestiones en cuanto a la validez de este género y las especies colocadas dentro del mismo. Para los fines de esta invención, se considerará que las especies descritas dentro de *Ulkenia* son miembros del género *Thraustochytrium*.

Las cepas descritas en la presente invención como laberintúlidos incluyen los siguientes organismos: Orden: Labyrinthulales, Familia: Labyrinthulaceae, Géneros: *Labyrinthula* (especies: sp., *algeriensis*, *coenocystis*, *chattonii*, *macrocystis*, *macrocystis atlantica*, *macrocystis macrocystis*, *marina*, *minuta*, *roscoffensis*, *valkanovii*, *vitellina*, *vitellina pacifica*, *vitellina vitellina*, *zopfii*), *Labyrinthuloides* (especies: sp., *haliotidis*, *yorkensis*), *Labyrinthomyxa* (especie: sp., *maria*), *Diplophys* (especie: sp., *archeri*), *Pyrrhosorus* (especie: sp., *marinus*), *Sorodiplophrys* (especie: sp., *stercorea*) o *Chlamydomyxa* (especies: sp., *labyrinthuloides*, *montana*) (aunque no hay actualmente un consenso sobre la colocación taxonómica exacta de *Pyrrhosorus*, *Sorodiplophrys* o *Chlamydomyxa*).

Para producir rendimientos significativamente altos de diversas moléculas bioactivas usando el sistema de PKS de AGPI de la presente invención, un organismo, preferiblemente un microorganismo o una planta o parte de la planta (por ejemplo, una célula de planta), puede modificarse genéticamente para afectar a la actividad de un sistema de PKS de AGPI. En un aspecto, un organismo de este tipo puede contener de manera endógena y expresar un sistema de PKS de AGPI, y la modificación genética puede ser una modificación genética de uno o más de los dominios funcionales del sistema de PKS de AGPI endógeno, mediante lo cual la modificación tiene algún efecto sobre la actividad del sistema de PKS de AGPI. En otro aspecto, un organismo de este tipo puede contener de manera endógena y expresar un sistema de PKS de AGPI, y la modificación genética puede ser una introducción de al menos una secuencia de ácido nucleico exógeno (por ejemplo, una molécula de ácido nucleico recombinante), en el que la secuencia de ácido nucleico exógeno codifica para al menos un dominio biológicamente activo o proteína del mismo o un segundo sistema de PKS y/o una proteína que afecta a la actividad de dicho sistema de PKS de AGPI (por ejemplo, una fosfopanteteinil transferasa (PPTasa), comentado a continuación). Aún en otro aspecto, el organismo no contiene necesariamente de manera endógena (de manera natural) un sistema de PKS de AGPI, sino que se modifica genéticamente para introducir al menos una molécula de ácido nucleico recombinante que codifica para una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio de un sistema de PKS de AGPI. En este aspecto, la actividad de PKS de AGPI se ve afectada introduciendo o aumentando la actividad de PKS de AGPI en el organismo. A continuación se comentarán en mayor detalle diversas realizaciones asociadas con cada uno de estos aspectos.

Por tanto, según la presente invención, una realización se refiere a un microorganismo modificado genéticamente,

en el que el microorganismo expresa un sistema de PKS que comprende al menos un dominio biológicamente activo de un sistema de policétido sintasa (PKS) de ácidos grasos poliinsaturados (AGPI). El al menos un dominio del sistema de PKS de AGPI se codifica por una secuencia de ácido nucleico descrita en el presente documento. La modificación genética afecta a la actividad del sistema de PKS en el organismo. El microorganismo modificado genéticamente puede incluir una cualquiera o más de las secuencias de ácido nucleico identificadas anteriormente, y/o cualquiera de los otros homólogos de cualquiera de los ORF o dominios PKS de AGPI tal como se describió en detalle anteriormente.

Tal como se usa en el presente documento, un microorganismo modificado genéticamente puede incluir una bacteria, un protista, una microalga, un hongo u otros microbios modificados genéticamente, y particularmente, cualquiera de los géneros del orden Thraustochytriales (por ejemplo, un traustochytridio) descritos en el presente documento. Un microorganismo modificado genéticamente de este tipo tiene un genoma que está modificado (es decir, está mutado o cambiado) de su forma normal (es decir, de tipo natural o que se produce de manera natural) de manera que se logra el resultado deseado (es decir, actividad de PKS de AGPI aumentada o modificada y/o producción de un producto deseado usando el sistema de PKS de AGPI o componente del mismo). La modificación genética de un microorganismo puede lograrse usando desarrollo de cepas clásico y/o técnicas de genética molecular. Tales técnicas se conocen en la técnica y se dan a conocer generalmente para microorganismos, por ejemplo en Sambrook *et al.*, 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press. Un microorganismo modificado genéticamente puede incluir un microorganismo en el que se han insertado, delecionado o modificado moléculas de ácido nucleico (es decir, mutado; por ejemplo, mediante inserción, deleción, sustitución y/o inversión de nucleótidos), de una manera tal que tales modificaciones proporcionan el efecto deseado dentro del microorganismo.

Las células huésped de microorganismos preferidos para modificar según la presente invención incluyen, pero no se limitan a, cualquier bacteria, protista, microalga, hongo o protozoo. En un aspecto, los microorganismos preferidos para modificar genéticamente incluyen, pero no se limitan a, cualquier microorganismo del orden Thraustochytriales o cualquier microorganismo del orden Labyrinthulales. Las células huésped particularmente preferidas para su uso en la presente invención pueden incluir microorganismos de un género incluyendo, pero sin limitarse a: *Thraustochytrium*, *Ulkenia*, *Schizochytrium*, *Japonochytrium*, *Aplanochytriune*, *Althornia*, *Elina*, *Labyrinthula*, *Labyrinthuloides*, *Labyrinthosyxa*, *Diplophrys*, *Pyrrhosorus*, *Sorodiplophrys* o *Chlamydomyxa*. Otros ejemplos de microorganismos huésped adecuados para la modificación genética incluyen, pero no se limitan a, levadura incluyendo *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis* u otra levadura tal como *Candida*, *Kluyveromyces* u otros hongos, por ejemplo, hongos filamentosos tales como *Aspergillus*, *Neurospora*, *Penicillium*, etc. También pueden usarse células bacterianas como huéspedes. Esto incluye *Escherichia coli*, que puede ser útil en procesos de fermentación. Alternativamente, puede usarse un huésped un huésped tal como una especie de *Lactobacillus* o una especie de *Bacillus*.

Otra realización de la presente invención se refiere a una planta modificada genéticamente o parte de una planta (por ejemplo, en la que la planta se ha modificado genéticamente para expresar un sistema de PKS de AGPI descrito en el presente documento), que incluye al menos el complejo enzimático de PKS de AGPI central y, en una realización, al menos una proteína auxiliar de PKS de AGPI (por ejemplo, una PPTasa), de modo que la planta produce AGPI. Preferiblemente, la planta es una planta de semillas oleaginosas, en la que las semillas oleaginosas o el aceite en las semillas oleaginosas contienen AGPI producidos por el sistema de PKS de AGPI. Tales aceites contienen una cantidad detectable de al menos un AGPI primario o objetivo que es el producto del sistema de PKS de AGPI. No se conoce que las plantas contengan de manera endógena un sistema de PKS de AGPI y, por tanto, los sistemas de PKS de AGPI de la presente invención representan una oportunidad para producir plantas con capacidades de producción de ácidos grasos únicos. Una realización particularmente preferida de la presente invención es modificar por ingeniería genética plantas para producir uno o más AGPI en la misma planta, incluyendo EPA, DHA, DPA (n-3 y/o n-6); ARA, GLA, SDA y otros. La presente invención ofrece la capacidad para crear uno cualquiera de varios "aceites de diseño" en diversas razones y formas.

Se conocen bien en la técnica métodos para la modificación por ingeniería genética de plantas. Por ejemplo, se han desarrollado numerosos métodos para la transformación de plantas, incluyendo protocolos de transformación biológica y física. Véase, por ejemplo, Miki *et al.*, "Procedures for Introducing Foreign DNA into Plants" en Methods in Plant Molecular Biology and Biotechnology, Glick, B.R. y Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) págs. 67-88. Además, están disponibles vectores y métodos de cultivo *in vitro* para la transformación de tejidos o células de plantas y la regeneración de plantas. Véase, por ejemplo, Gruber *et al.*, "Vectors for Plant Transformation" en Methods in Plant Molecular Biology and Biotechnology, Glick, B.R. y Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) págs. 89-119.

El método más ampliamente utilizado para introducir un vector de expresión en plantas se basa en el sistema de transformación natural de *Agrobacterium*. Véase, por ejemplo, Horsch *et al.*, Science 227:1229 (1985). *A. tumefaciens* y *A. rhizogenes* son bacterias del suelo patógenas de plantas que transforman genéticamente células de plantas. Los plásmidos Ti y Ri de *A. tumefaciens* y *A. rhizogenes*, respectivamente, portan genes responsables de la transformación genética de la planta. Véase, por ejemplo, Kado, C.I., Crit. Rev. Plant. Sci. 10:1 (1991). Numerosas referencias proporcionan descripciones de sistemas de vector de *Agrobacterium* y métodos para la

transferencia génica mediada por *Agrobacterium*, incluyendo Gruber *et al.*, citado anteriormente, Miki *et al.*, citado anteriormente, Moloney *et al.*, Plant Cell Reports 8:238 (1989) y las patentes estadounidenses n.^{os} 4.940.838 y 5.464.763.

- 5 Otro método generalmente aplicable de transformación de plantas es transformación mediada por microproyectiles en el que se porta ADN sobre la superficie de microproyectiles. El vector de expresión se introduce en tejidos de plantas con un dispositivo biolístico que acelera los microproyectiles hasta velocidades suficientes para penetrar en las membranas y paredes de células de plantas. Sanford *et al.*, Part. Sci. Technol. 5:27 (1987), Sanford, J.C., Trends Biotech. 6:299 (1988), Sanford, J.C., Physiol. Plant 79:206 (1990), Klein *et al.*, Biotechnology 10:268 (1992).

- 10 Otro método para el suministro físico de ADN a plantas es la sonicación de células diana. Zhang *et al.*, Bio/Technology 9:996 (1991). Alternativamente, se ha usado la fusión de esferoplastos o liposomas para introducir vectores de expresión en plantas. Deshayes *et al.*, EMBO J., 4:2731 (1985), Christou *et al.*, Proc Natl. Acad. Sci. USA 84:3962 (1987). También se ha notificado la captación directa de ADN en protoplastos usando precipitación con CaCl₂, poli(alcohol vinílico) o poli-L-ornitina. Hain *et al.*, Mol. Gen. Genet.- 199:161 (1985) y Draper *et al.*, Plant Cell Physiol. 23:451 (1982). También se ha descrito la electroporación de protoplastos y tejidos y células completas. Donn *et al.*, en Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC, A2-38, pág. 53 (1990); D'Halluin *et al.*, Plant Cell 4:1495-1505 (1992) y Spencer *et al.*, Plant Mol. Biol. 24:51-61 (1994).

- 20 Tras la introducción del constructo genético en células de plantas, se hacen crecer células de plantas y tras la aparición de tejido de diferenciación tal como brotes y raíces, se generan plantas maduras. Normalmente se genera una pluralidad de plantas. Los expertos en la técnica conocerán generalmente metodologías para regenerar plantas y pueden encontrarse por ejemplo en: Plant Cell and Tissue Culture, 1994, Vasil y Thorpe Eds. Kluwer Academic Publishers y en: Plant Cell Culture Protocols (Methods in Molecular Biology 111, 1999 Hall Eds Humana Press).

- 25 Tal como se usa en el presente documento, una planta modificada genéticamente puede incluir cualquier planta modificada genéticamente incluyendo plantas superiores y, particularmente, cualquier planta consumible o planta útil para producir una molécula bioactiva deseada de la presente invención. "Partes de plantas", tal como se usa en el presente documento, incluyen cualquier parte de una planta, incluyendo, pero sin limitarse a, semillas (inmaduras o maduras), aceites, polen, embriones, flores, frutos, brotes, hojas, raíces, tallos, explantes, etc. Una planta modificada genéticamente tiene un genoma que está modificado (es decir, mutado o cambiado) con respecto a su forma normal (es decir, de tipo natural o que se produce de manera natural) de manera que se alcanza el resultado deseado (por ejemplo, actividad PKS de AGPI y producción de AGPI). La modificación genética de una planta puede lograrse usando técnicas clásicas de desarrollo de cepas y/o genética molecular. En la técnica se conocen métodos para producir una planta transgénica, en los que se incorpora en el genoma de la planta una molécula de ácido nucleico recombinante que codifica para una secuencia de aminoácidos deseada. Una planta preferida para modificarse genéticamente según la presente invención es preferiblemente una planta adecuada para su consumo por animales, incluyendo seres humanos.

- 40 Las plantas preferidas para modificarse genéticamente según la presente invención (es decir, células huésped de plantas) incluyen, pero no se limitan a, cualquier planta superior, incluyendo plantas tanto dicotiledóneas como monocotiledóneas, y particularmente plantas consumibles, incluyendo plantas cultivadas y especialmente plantas usadas por sus aceites. Tales plantas pueden incluir, pero no se limitan a, por ejemplo: canola, soja, colza, linaza, maíz, cártamos, girasoles y tabaco. Por tanto, puede seleccionarse cualquier especie de planta o célula de planta.
- 45 Las células particulares usadas en el presente documento, y las plantas que se hacen crecer o se derivan a partir de las mismas, incluyen, pero no se limitan a, células que pueden obtenerse de canola (*Brassica rapa* L.); soja (*Glicina max*); colza (*Brassica spp.*); linaza/lino (*Linum usitatissimum*); maíz (*Zea mays*); cártamo (*Carthamus tinctorius*); girasol (*Helianthus annuus*); tabaco (*Nicotiana tabacum*); *Arabidopsis thaliana*, nuez de Brasil (*Betholettia excelsa*); ricino (*Ricinus communis*); coco (*Cocos nucifera*); cilantro (*Coriandrum sativum*); algodón (*Gossypium spp.*); cacahuete (*Arachis hypogaea*); jojoba (*Simmondsia chinensis*); mostaza (*Brassica spp.* y *Sinapis alba*); palma de aceite (*Elaeis guineensis*); olivo (*Olea europaea*); arroz (*Oryza sativa*); calabacín (*Cucurbita maxima*); cebada (*Hordeum vulgare*); trigo (*Triticum aestivum*) y lenteja de agua (*Lemnaceae sp.*). Debe observarse que según el presente documento el contexto genético dentro de una especie de planta puede variar.

- 55 Otras plantas preferidas incluyen aquellas plantas que se sabe que producen compuestos usados como agentes farmacéuticos, agentes aromatizantes, agentes nutracéuticos, componentes alimenticios funcionales o agentes cosméticamente activos o plantas que se modifican por ingeniería genética para producir estos compuestos/agentes.

- 60 En una realización adicional, pueden usarse cultivos de células de plantas de acuerdo con el presente documento. En tales realizaciones no se hacen crecer células de plantas para dar plantas diferenciadas ni se cultivan usando prácticas agrícolas habituales, sino que en vez de eso se hacen crecer y se mantienen en un medio líquido.

- Según la presente invención, una planta o un microorganismo modificado genéticamente incluye una planta o un microorganismo que se ha modificado usando tecnología recombinante. Tal como se usa en el presente documento, las modificaciones genéticas que dan como resultado una disminución de la expresión génica, de la función del gen o de la función del producto génico (es decir, la proteína codificada por el gen) pueden denominarse inactivación

(completa o parcial), delección, interrupción, bloqueo o regulación por disminución de un gen. Por ejemplo, una modificación genética en un gen que da como resultado una disminución de la función de la proteína codificada por tal gen, puede ser el resultado de una delección completa del gen (es decir, el gen no existe, y por tanto la proteína no existe), una mutación en el gen que da como resultado una 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 una disminución o ausencia de acción o actividad enzimática). Las modificaciones genéticas que dan como resultado un aumento de la expresión o función génica pueden denominarse amplificación, sobreproducción, sobreexpresión, activación, potenciación, adición o regulación por incremento de un gen.

La modificación genética de una planta o un microorganismo según la presente invención afecta preferiblemente a la actividad del sistema de PKS expresado por la planta, ya sea el sistema de PKS endógeno y modificado genéticamente, endógeno con la introducción de moléculas de ácido nucleico recombinantes en el organismo o se proporcione completamente mediante tecnología recombinante. Según la presente invención, "afectar a la actividad de un sistema de PKS" incluye cualquier modificación genética que provoca cualquier cambio o modificación detectable o medible en el sistema de PKS expresado por el organismo en comparación con en ausencia de la modificación genética. Un cambio o modificación detectable en el sistema de PKS puede incluir, pero no se limita a: la introducción de actividad de sistema de PKS en un organismo de manera que el organismo ahora tiene actividad de sistema de PKS medible/detectable (es decir, el organismo no contenía un sistema de PKS antes de la modificación genética), la introducción en el organismo de un dominio funcional de un sistema de PKS diferente de un sistema de PKS expresado de manera endógena por el organismo de manera que se modifica la actividad de sistema de PKS (por ejemplo, se introduce el dominio DH2 de un sistema de PKS de AGPI en el sistema de PKS de AGPI de un organismo diferente), un cambio en la cantidad de una molécula bioactiva producida por el sistema de PKS (por ejemplo, el sistema produce más (cantidad aumentada) o menos (cantidad disminuida) de un producto dado en comparación con en ausencia de la modificación genética), un cambio en el tipo de una molécula bioactiva producida por el sistema de PKS (por ejemplo, el sistema produce un producto nuevo o diferente, o una variante de un producto que se produce de manera natural por el sistema) y/o un cambio en la razón de múltiples moléculas bioactivas producidas por el sistema de PKS (por ejemplo, el sistema produce una razón diferente de un AGPI con respecto a otro AGPI, produce un perfil lipídico completamente diferente en comparación con en ausencia de la modificación genética o coloca diversos AGPI en posiciones diferentes en un triacilglicerol en comparación con la configuración natural). Una modificación genética de este tipo incluye cualquier tipo de modificación genética e incluye específicamente modificaciones realizadas mediante tecnología recombinante y mediante mutagénesis clásica.

Debe observarse que la referencia a aumentar la actividad de un dominio o una proteína funcional en un sistema de PKS de AGPI se refiere a cualquier modificación genética en el organismo que contiene el dominio o la proteína (o en el que debe introducirse el dominio o la proteína) que da como resultado un aumento de la funcionalidad del sistema de dominio o proteína y puede incluir actividad superior del dominio o de la proteína (por ejemplo, actividad específica o actividad enzimática *in vivo*), inhibición o degradación reducida del sistema de dominio o proteína y sobreexpresión del dominio o de la proteína. Por ejemplo, puede aumentarse el número de copias del gen, pueden aumentarse los niveles de expresión mediante el uso de un promotor que proporciona niveles de expresión superiores a los del promotor nativo o puede alterarse un gen mediante ingeniería genética o mutagénesis clásica para aumentar la actividad del dominio o la proteína codificado por el gen.

De manera similar, la referencia a disminuir la actividad de un dominio o una proteína funcional en un sistema de PKS de AGPI se refiere a cualquier modificación genética en el organismo que contiene tal dominio o proteína (o en el que va a introducirse el dominio o la proteína) que da como resultado una disminución de la funcionalidad del dominio o de la proteína e incluye actividad disminuida del dominio o de la proteína, inhibición o degradación aumentada del dominio o de la proteína y una reducción o eliminación de la expresión del dominio o de la proteína. Por ejemplo, la acción de un dominio o una proteína de la presente invención puede disminuirse bloqueando o reduciendo la producción del dominio o de la proteína, "desactivación" del gen o la porción del mismo que codifica para el dominio o la proteína, reducción de la actividad del dominio o de la proteína o inhibición de la actividad del dominio o de la proteína. Bloquear o reducir la producción de un dominio o una proteína puede incluir colocar el gen que codifica para el dominio o la proteína bajo el control de un promotor que requiere la presencia de un compuesto inductor en el medio de crecimiento. Estableciendo las condiciones de manera que se agote el inductor del medio, puede inactivarse la expresión del gen que codifica para el dominio o la proteína (y por tanto, de la síntesis de proteína). Bloquear o reducir la actividad del dominio o de la proteína también puede incluir usar un enfoque de tecnología de escisión similar al descrito en la patente estadounidense n.º 4.743.546. Para usar este enfoque, se clona el gen que codifica para la proteína de interés entre secuencias genéticas específicas que permiten una escisión específica, controlada, del gen a partir del genoma. La escisión puede provocarse, por ejemplo, por un desplazamiento en la temperatura de cultivo del cultivo, como en la patente estadounidense n.º 4.743.546, o por alguna otra señal física o nutricional.

En una realización de la presente invención, una modificación genética incluye una modificación de una secuencia de ácido nucleico que codifica para una proteína o un dominio de un sistema de PKS de AGPI expresado de manera endógena (de manera natural), mediante lo cual se modifica genéticamente un microorganismo que contiene de

manera natural un sistema de este tipo, por ejemplo, mediante 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, usar un vector recombinante de direccionamiento para delecionar una porción de un gen endógeno o para sustituir una porción de un gen endógeno por una secuencia heteróloga. Los ejemplos de

5 secuencias heterólogas que pueden introducirse en un genoma huésped incluyen secuencias que codifican para al menos un dominio funcional de otro sistema de PKS, tal como un sistema de PKS de AGPI diferente (bacteriano o no bacteriano), un sistema de PKS de tipo I (iterativo o modular), un sistema de PKS de tipo II o un sistema de PKS de tipo III. Otras secuencias heterólogas para introducir en el genoma de un huésped incluyen una secuencia que codifica para una proteína o un dominio funcional que no es un dominio de un sistema de PKS central, pero que

10 afectará a la actividad del sistema de PKS endógeno. Por ejemplo, puede introducirse en el genoma huésped una molécula de ácido nucleico que codifica para una fosfopanteteinil transferasa (comentada a continuación). A continuación se comentan en detalle modificaciones específicas que pueden realizarse a un sistema de PKS de AGPI endógeno.

15 En otro aspecto de esta realización de la invención, la modificación genética incluye: (1) la introducción en una célula o un organismo huésped homólogo o heterólogo de una molécula de ácido nucleico recombinante que codifica para una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio de un sistema de PKS de AGPI; y/o (2) la introducción en una célula o un organismo huésped de una molécula de ácido nucleico recombinante que codifica para una proteína o un dominio funcional que afecta a la actividad de un sistema de PKS de AGPI. El huésped puede incluir: (1) una célula o un organismo huésped que no expresa ningún sistema de PKS para la producción de AGPI, en el que se introducen todos los dominios funcionales de un sistema de PKS de AGPI en la célula huésped; (2) una célula huésped que expresa un sistema de PKS para la producción de AGPI (endógeno o recombinante), en la que se introduce al menos un dominio o una proteína PKS de AGPI adicional en la célula o el organismo. En otras palabras, la presente invención pretende abarcar cualquier célula u organismo

20 modificado genéticamente (por ejemplo, microorganismo o planta), en el que el organismo comprende al menos un dominio o una proteína PKS de AGPI descrito en el presente documento o se ha modificado para producir un dominio o una proteína PKS de AGPI resintetizado y/o quimérico tal como se describe en el presente documento.

Por tanto, usando la orientación proporcionada en el presente documento, así como la descripción de los sistemas de PKS de AGPI descritos en el presente documento y conocidos antes de la invención, puede usarse el mezclado de genes (o mezclado de moléculas de ácido nucleico), por ejemplo, mediante la producción de proteínas quiméricas y/o sistemas de PKS de AGPI quiméricos tal como se describe en detalle en el presente documento, para ampliar la gama de productos de AGPI, las razones de los mismos y los niveles de producción de los mismos, por un organismo que expresa el sistema de PKS de AGPI. Por ejemplo, pueden usarse las enseñanzas

30 proporcionadas en el presente documento para mejorar las cantidades de AGPI producidos, para cambiar la razón de un AGPI con respecto a otro, incluyendo la razón de AGPI omega-3 con respecto a omega-6, y para ampliar la gama de productos de PKS de AGPI para incluir EPA, DPA (n-3 o n-6), DHA, ARA, GLA, SDA y otros, así como para producir una amplia variedad de moléculas bioactivas, incluyendo antibióticos, otros compuestos farmacéuticos y otros productos deseables. El método para obtener estas mejoras no sólo incluye el mezclado de genes de diversos organismos sino también diversos métodos de modificación genética de los genes de PKS de AGPI y las moléculas de ácido nucleico dadas a conocer en el presente documento. El conocimiento de la base genética y la estructura de dominios de los sistemas de PKS de AGPI tal como se describe en el presente documento proporciona una base para diseñar organismos modificados genéticamente novedosos. A modo de ejemplo, en la publicación de solicitud de patente estadounidense n.º 20020194641, la publicación de solicitud de patente estadounidense n.º 20040235127 y la publicación de solicitud de patente estadounidense n.º 20050100995, citadas anteriormente con respecto a la modificación genética y producción de moléculas bioactivas, se comentan diversas manipulaciones posibles del sistema de PKS de AGPI. Sin embargo, esta invención proporciona realizaciones novedosas referentes a la manipulación de los niveles de producción de AGPI por un organismo huésped y la manipulación de la razón de AGPI producidos por un organismo huésped.

40

50 Por consiguiente, se describen métodos para modificar genéticamente células microbianas o de plantas mediante: modificar genéticamente al menos una secuencia de ácido nucleico en el organismo que codifica para una secuencia de aminoácidos que tiene la actividad biológica de al menos un dominio funcional de un sistema de PKS de AGPI según la presente invención, y/o expresar al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica para tal secuencia de aminoácidos. Anteriormente se han descrito en detalle diversas realizaciones de tales secuencias, métodos para modificar genéticamente un organismo y modificaciones específicas. Normalmente, el método se usa para producir un organismo modificado genéticamente particular que produce una molécula o moléculas bioactivas particulares.

55

60 Se contempla que puede combinarse un programa de mutagénesis con un procedimiento de examen selectivo para obtener moléculas bioactivas de interés. Esto incluirá métodos para buscar una gama de compuestos bioactivos. Esta búsqueda no se limitará a la producción de las moléculas con dobles enlaces cis. Los métodos de mutagénesis pueden incluir, pero no se limitan a: mutagénesis química, intercambio génico, cambiar regiones de los genes que codifican para dominios enzimáticos específicos o mutagénesis limitada a regiones específicas de esos genes, así como otros métodos.

65

Por ejemplo, pueden usarse métodos de mutagénesis de alto rendimiento para influir en, u optimizar, la producción de la molécula bioactiva deseada. Una vez que se haya desarrollado un sistema modelo eficaz, pueden modificarse estos genes de una manera con alto rendimiento. Puede considerarse la utilización de estas tecnologías en dos niveles. En primer lugar, si puede concebirse un examen lo suficientemente selectivo para la producción de un producto de interés (por ejemplo, ARA), puede usarse para intentar alterar el sistema para producir este producto (por ejemplo, en lugar de, o junto con, otras estrategias tales como las comentadas anteriormente). Adicionalmente, si las estrategias expuestas anteriormente dieron como resultado un conjunto de genes que sí producían el producto de interés, entonces pueden usarse las tecnologías de alto rendimiento para optimizar el sistema. Por ejemplo, si el dominio introducido sólo funcionaba a temperaturas relativamente bajas, pueden concebirse métodos de selección para permitir eliminar esa limitación.

Se reconoce que muchas alteraciones genéticas, o bien al azar o bien dirigidas, que pueden introducirse en un sistema de PKS de AGPI nativo (endógeno, natural), darán como resultado una inactivación de funciones enzimáticas. Por tanto, se describe un sistema para seleccionar sólo aquellas modificaciones que no bloquean la capacidad del sistema de PKS de AGPI para producir un producto. Por ejemplo, la cepa FabB de *E. coli* no puede sintetizar ácidos grasos insaturados y requiere la complementación del medio con ácidos grasos que pueden sustituir a sus ácidos grasos insaturados normales con el fin de crecer (véase Metz *et al.*, 2001, citado anteriormente). Sin embargo, este requisito (de complementación del medio) puede eliminarse cuando se transforma la cepa con un sistema de PKS de AGPI funcional (es decir, uno que produce un producto de AGPI en el huésped de *E. coli*, véase (Metz *et al.*, 2001, citado anteriormente, figura 2A)). Ahora la cepa FabB transformada requiere un sistema de PKS de AGPI funcional (para producir los ácidos grasos insaturados) para crecer sin complementación. El elemento clave en este ejemplo es que bastará con la producción de una amplia gama de ácidos grasos insaturados (incluso sustitutos de ácidos grasos insaturados, tales como ácidos grasos de cadena ramificada). Por tanto, se contempla que puede crearse un gran número de mutaciones en uno o más de los genes de PKS de AGPI dados a conocer en el presente documento, y entonces transformar la cepa FabB apropiadamente modificada (por ejemplo, crear mutaciones en un constructo de expresión que contiene un dominio ER y transformar una cepa FabB que tiene los otros dominios esenciales en un plásmido separado, o integrados en el cromosoma) y seleccionar sólo aquellos transformantes que crecen sin complementación del medio (es decir, que todavía presentaban una capacidad para producir una molécula que podía complementar el defecto de FabB). Pueden desarrollarse exámenes adicionales para buscar compuestos particulares (por ejemplo el uso de CG para ácidos grasos) que están produciéndose en este subconjunto selectivo de un sistema de PKS activo. Pueden considerarse diversos exámenes selectivos similares para moléculas bioactivas de interés.

En una realización de la invención, un organismo modificado genéticamente tiene una modificación que cambia al menos un producto producido por el sistema de PKS endógeno, en comparación con un organismo de tipo natural. Los constructos novedosos usados para producir tales organismos modificados, así como las proteínas y los organismos producidos usando tales constructos, y los métodos asociados con tales modificaciones, están todos abarcados por la invención.

En una realización preferida, un organismo modificado genéticamente expresa un sistema de PKS de AGPI que comprende una modificación genética en un dominio β -hidroxil acil-ACP deshidrasa (DH) correspondiente al dominio DH2 de *Schizochytrium* o *Thraustochytrium*, en el que la modificación altera la razón de ácidos grasos de cadena larga, y particularmente, la razón de ácidos grasos de cadena larga omega-3 con respecto a omega-6, producidos por el sistema de PKS de AGPI, en comparación con en ausencia de la modificación. En un aspecto de esta realización, la modificación se selecciona del grupo que consiste en una delección de la totalidad o de una parte del dominio, una sustitución de la totalidad o de parte del dominio por un dominio homólogo o una parte del mismo de un organismo diferente (por ejemplo, un organismo diferente que produce de manera natural diferentes razones y/o cantidades de los AGPI) y una mutación del dominio.

Más específicamente, tal como se ilustra en el presente documento, la comparación de la arquitectura (organización de dominios) de PKS de AGPI de *Schizochytrium* y *Thraustochytrium* con otra arquitectura de sistema de PKS de AGPI ilustra la capacidad de la naturaleza para alterar el orden de dominios así como incorporar nuevos dominios para crear productos finales novedosos, o alterar las razones de productos finales, por ejemplo. Además, ahora pueden manipularse los genes en el laboratorio para crear nuevos productos, tal como se describe en los ejemplos. Los inventores han demostrado ahora la capacidad de emplear esta capacidad y usarla para crear organismos novedosos con cantidades de producción y perfiles de AGPI novedosos. En el presente documento se describe la manipulación de sistemas de PKS de AGPI de una manera o bien dirigida o bien al azar para influir en los productos finales. Por ejemplo, en una realización preferida, se usa la sustitución de un dominio DH (similar a FabA) o una porción biológicamente activa del mismo de un primer sistema de PKS de AGPI, y específicamente, el dominio DH2 descrito en el presente documento, por el dominio DH homólogo o una porción biológicamente activa del mismo en un segundo sistema de PKS de AGPI diferente, para alterar la razón de AGPI producidos por el segundo sistema de PKS de AGPI, y particularmente, para manipular la razón de ácidos grasos omega-3 con respecto a omega-6 producidos por el segundo sistema de PKS de AGPI. Puede alcanzarse un resultado similar sustituyendo una proteína entera o cualquier porción biológicamente activa de la misma que contiene tal dominio DH2 (por ejemplo, el OrfC de *Thraustochytrium* 23B) de un primer sistema de PKS de AGPI por la proteína homóloga o porción de la misma en un segundo sistema de PKS de AGPI. Aunque los ejemplos descritos en el presente documento utilizan

los sistemas de PKS de AGPI de *Schizochytrium* y *Thraustochytrium*, la invención abarca la manipulación similar de cualquier sistema de PKS o similar a PKS para la producción de AGPI mediante la modificación de la proteína DH2 o del dominio similar a DH2. Tal modificación puede realizarse sola o junto con otras modificaciones de un sistema de PKS de AGPI.

Por consiguiente, una realización de la presente invención comprende un sistema de PKS de AGPI quimérico y un organismo que expresa tal sistema de PKS de AGPI quimérico. En un aspecto, el sistema de PKS de AGPI quimérico comprende un primer sistema de PKS de AGPI, en el que el dominio o la proteína del primer sistema de PKS de AGPI que corresponde al dominio DH2 o porción biológicamente activa del mismo (por ejemplo, de *Schizochytrium* o *Thraustochytrium* descritos en el presente documento) se ha modificado o sustituido por un dominio DH2 o una proteína o porción biológicamente activa del mismo de un segundo sistema de PKS de AGPI diferente. Por "sistema de PKS de AGPI diferente" quiere decirse un sistema de PKS de AGPI de una cepa, especie, género u organismo diferente, o incluso un homólogo de un sistema de PKS de AGPI natural o de tipo natural. El objetivo de producir esta proteína quimérica es alterar la razón de los AGPI, y particularmente la razón de AGPI omega-3 con respecto a omega-6, producidos por el sistema de PKS de AGPI. Por tanto, la selección del sistema de PKS de AGPI diferente debe basarse en la selección de un segundo sistema que produce una razón de los AGPI diferente, o deseada, que el primer sistema de PKS de AGPI.

En un aspecto de la invención, tal sistema de PKS de AGPI quimérico comprende una proteína OrfA (SEQ ID NO: 2) y OrfB (SEQ ID NO: 4) de *Schizochytrium* tal como se describe en el presente documento, y una proteína OrfC (SEQ ID NO: 62) de *Thraustochytrium* tal como se describe en el presente documento. En los ejemplos se describen organismos de *Schizochytrium*, *E. coli* y levadura que expresan tales sistemas de PKS de AGPI quiméricos y están abarcados por la presente invención, además de plantas y partes de plantas que expresan tales sistemas de PKS de AGPI quiméricos. En otras realizaciones, ejemplificadas en los ejemplos, se producen sistemas de PKS de AGPI quiméricos que comprenden todas las combinaciones de los OrfA, B y C de *Schizochytrium* y *Thraustochytrium*.

En otro aspecto de la invención, un sistema de PKS de AGPI quimérico comprende una proteína OrfA (SEQ ID NO: 2) y OrfB (SEQ ID NO: 4) de *Schizochytrium* tal como se describe en el presente documento, y una proteína OrfC quimérica (codificada por una secuencia de ácido nucleico representada en el presente documento por SEQ ID NO: 74, codificada por SEQ ID NO: 73). El polipéptido OrfC quimérico tiene 1493 residuos de aminoácido de longitud. La región DH2, definida como los aminoácidos 516-1041 de SEQ ID NO: 74, consiste en la secuencia de aminoácidos de la región DH2 de la proteína OrfC de *Th.23B*, es decir, los aminoácidos 491-1016 de SEQ ID NO: 62, que incluye la totalidad de SEQ ID NO: 66 y alguna secuencia de aminoácidos flanqueante de SEQ ID NO: 62. Con respecto al resto de la secuencia de aminoácidos de OrfC quimérica, los residuos 1-515 y 1042-1493 de SEQ ID NO: 74 son idénticos a los residuos 1-515 y 1051-1502 del OrfC de *Schizochytrium* de SEQ ID NO: 6, respectivamente.

En otra realización de la invención, se ha modificado una célula o un organismo modificado genéticamente para expresar un sistema de PKS de AGPI o una porción del mismo, incluyendo un sistema de PKS de AGPI quimérico, en el que la(s) secuencia(s) de ácido nucleico que codifica(n) para el sistema de PKS de AGPI o porción del mismo está(n) optimizada(s) totalmente o en parte para utilizar el uso de codones preferido de la célula o el organismo huésped. Esta realización se ejemplifica a continuación e ilustra cómo puede aumentarse la producción de una molécula bioactiva (por ejemplo, un AGPI) realizando tales modificaciones. Esta realización puede usarse junto con las demás modificaciones genéticas descritas en el presente documento (por ejemplo, realizaciones de proteínas y PKS de AGPI quiméricos), para mejorar la producción de una molécula bioactiva en un organismo huésped.

En un aspecto de esta realización, un sistema de PKS de AGPI quimérico comprende una proteína OrfA (SEQ ID NO: 2) y OrfB (SEQ ID NO: 4) de *Schizochytrium* tal como se describe en el presente documento, y una proteína OrfC (SEQ ID NO: 62) de *Thraustochytrium* tal como se describe en el presente documento, en el que la secuencia de ácido nucleico que codifica para SEQ ID NO: 62 está optimizada para el uso de codones del huésped. En los ejemplos se describe un ejemplo de tal molécula optimizada para la expresión en *Schizochytrium*, representándose tal secuencia de ácido nucleico que codifica para OrfC de *Thraustochytrium* (OrfC sintético u optimizado para codones) en el presente documento por SEQ ID NO: 70. En otra realización, puede combinarse OrfA de *Thraustochytrium* (SEQ ID NO: 39) y/o OrfB de *Thraustochytrium* (SEQ ID NO: 52) con uno cualquiera o más de los Orf A, B y/o C de *Schizochytrium*, y/o con el OrfC de *Thraustochytrium*, para la expresión en *Schizochytrium*. De nuevo, en este ejemplo, la molécula de ácido nucleico que codifica para el OrfA de *Thraustochytrium* y/o el OrfB de *Thraustochytrium* puede estar optimizada para el uso de codones del huésped. En los ejemplos se describen ejemplos de tales moléculas optimizadas para la expresión en *Schizochytrium*, representándose la secuencia de ácido nucleico que codifica para OrfA de *Thraustochytrium* (OrfA sintético u optimizado para codones) en el presente documento por SEQ ID NO: 71, y representándose la secuencia de ácido nucleico que codifica para OrfB de *Thraustochytrium* (OrfB sintético u optimizado para codones) en el presente documento por SEQ ID NO: 72.

En otro aspecto de esta realización, un sistema de PKS de AGPI quimérico comprende una proteína OrfA (SEQ ID NO: 2) y OrfB (SEQ ID NO: 4) de *Schizochytrium* tal como se describe en el presente documento, y una proteína OrfC quimérica y parcialmente optimizada para codones (codificada por una secuencia de ácido nucleico representada en el presente documento por SEQ ID NO: 75). La proteína codificada por SEQ ID NO: 75 también se representa por SEQ ID NO: 74, que se describió anteriormente con respecto a SEQ ID NO: 73. Sin embargo, en este

caso la porción de la secuencia de ácido nucleico que codifica para SEQ ID NO: 66 (dominio DH2), que se deriva de *Thraustochytrium*, está optimizada para la expresión en *Schizochytrium* tal como se describe en los ejemplos.

Anteriormente y a continuación en los ejemplos se describen otras secuencias de ácido nucleico optimizadas para codones para su uso en *E. coli*, levadura y plantas.

En otra realización, se ha modificado un organismo modificado genéticamente mediante transfección del organismo con una molécula de ácido nucleico recombinante que codifica para una proteína que regula la longitud de cadena de ácidos grasos producidos por el sistema de PKS de AGPI. Por ejemplo, la proteína que regula la longitud de cadena de ácidos grasos producidos por el sistema de PKS de AGPI puede ser un factor de longitud de cadena que dirige la síntesis de unidades C20 y/o unidades C22.

En otra realización, un organismo modificado genéticamente expresa un sistema de PKS de AGPI que comprende una modificación en un dominio enoil-ACP reductasa (ER), en el que la modificación da como resultado la producción de un compuesto diferente en comparación con en ausencia de la modificación. En un aspecto de esta realización, la modificación se selecciona del grupo que consiste en una delección de la totalidad o una parte de un ER dominio, una sustitución del dominio ER por un dominio ER de un organismo diferente y una mutación de un dominio ER.

En una realización de la invención, el organismo modificado genéticamente produce un perfil de ácidos grasos poliinsaturados (AGPI) que es diferente del del organismo que se produce de manera natural sin una modificación genética.

Muchas otras modificaciones genéticas útiles para producir moléculas bioactivas resultarán evidentes para los expertos en la técnica, dada la presente divulgación, y diversas otras modificaciones se han comentado anteriormente en el presente documento. La presente invención contempla cualquier modificación genética relacionada con un sistema de PKS de AGPI tal como se describe en el presente documento que da como resultado la producción de una molécula bioactiva deseada.

Tal como se describió anteriormente, en una realización de la presente invención, un organismo modificado genéticamente, tal como una planta o un microorganismo modificado genéticamente, incluye un organismo que tiene una capacidad potenciada de sintetizar moléculas bioactivas deseadas (productos) o que tiene una capacidad recién introducida de sintetizar productos específicos (por ejemplo, sintetizar AGPI, sintetizar un perfil de AGPI diferente o sintetizar un antibiótico específico). Según la presente invención, “una capacidad potenciada de sintetizar” un producto se refiere a cualquier potenciación, o regulación por incremento, en una ruta relacionada con la síntesis del producto de manera que el microorganismo o la planta produce una cantidad aumentada del producto (incluyendo cualquier producción de un producto cuando antes no había ninguna) en comparación con el microorganismo o la planta de tipo natural, que se cultiva o se hace crecer en las mismas condiciones. Anteriormente se describieron en detalle métodos para producir tales organismos modificados genéticamente. En una realización preferida, la presente invención se refiere a una planta o parte de una planta modificada genéticamente (por ejemplo, en la que la planta se ha modificado genéticamente para expresar un sistema de PKS de AGPI, incluyendo un sistema de PKS de AGPI quimérico, descrito en el presente documento) que incluye al menos el complejo enzimático de PKS de AGPI central y, en una realización, al menos una proteína auxiliar de PKS de AGPI (por ejemplo, una PPTasa), de modo que la planta produce AGPI. Preferiblemente, la planta es una planta de semilla oleaginosa, en la que las semillas oleaginosas o el aceite en las semillas oleaginosas contienen AGPI producidos por el sistema de PKS de AGPI. Tales aceites contienen una cantidad detectable de al menos un AGPI objetivo o principal que es el producto del sistema de PKS de AGPI.

Los presentes inventores han demostrado la producción de AGPI en una planta que se ha modificado genéticamente para expresar los genes que codifican para un sistema de PKS de AGPI de *Schizochytrium* y una enzima auxiliar de PKS de AGPI, 4'-fosfopanteteinil transferasa (PPTasa) (por ejemplo, véase la publicación de solicitud de patente estadounidense n.º 20070089199, citada anteriormente). Los aceites producidos por estas plantas contienen cantidades significativas tanto de DHA (ácido docosahexaenoico (C22:6, n-3)) como de DPA (ácido docosapentaenoico (C22:5, n-6), que son los AGPI predominantes (los AGPI primarios) producidos por el *Schizochytrium* del que se derivaron los genes de PKS de AGPI. Significativamente, los aceites de plantas que producen AGPI usando la ruta de PKS de AGPI tienen un perfil de ácidos grasos diferente del de plantas que se modifican por ingeniería genética para producir los mismos AGPI mediante la ruta “habitual” descrita anteriormente. En particular, los aceites de plantas que se han modificado por ingeniería genética para producir AGPI específicos mediante la ruta de PKS de AGPI están sustancialmente libres de los diversos productos intermedios y secundarios que se acumulan en aceites que se producen como resultado del uso de la ruta de síntesis de AGPI habitual. Esta característica se comenta en detalle a continuación.

Más particularmente, esfuerzos para producir AGPI de cadena larga en plantas mediante la ruta “habitual” (descritos anteriormente) han adoptado el mismo enfoque básico, que viene dictado por esta ruta de síntesis. Estos esfuerzos se basaron en la modificación de los ácidos grasos endógenos de las plantas mediante la introducción de genes que codificaban para diversas elongasas y desaturasas. Las plantas producen normalmente ácidos grasos de 18

carbonos (por ejemplo, ácido oleico, ácido linolénico, ácido linolénico) mediante la ácido graso sintasa de tipo II (FAS) en sus plástidos. A menudo, se forma un único doble enlace mientras el ácido graso está unido a ACP, y entonces se escinde el ácido oleico (18:1) de ACP mediante la acción de una acil-ACP tioesterasa. El ácido graso libre se exporta desde el plástido y se convierte en un acil-CoA. El ácido 18:1 puede esterificarse con fosfatidilcolina (PC) y pueden añadirse hasta dos dobles enlaces cis más. Las elongasas recién introducidas pueden utilizar sustratos en la reserva de acil-CoA para añadir carbonos en incrementos de dos carbonos. Las desaturasas recién introducidas pueden utilizar o bien ácidos grasos esterificados con PC o bien aquéllos en la reserva de acil-CoA, dependiendo de la fuente de la enzima. Sin embargo, una consecuencia de este esquema para la producción de AGPI de cadena larga es que se acumulan productos intermedios o secundarios en la ruta, que a menudo representan la mayoría de los ácidos grasos novedosos en el aceite vegetal, en vez del AGPI de cadena larga objetivo.

Por ejemplo, usando la ruta habitual o clásica tal como se describió anteriormente, cuando el producto de AGPI objetivo (es decir, el producto de AGPI que es el objetivo de la producción, que se pretende producir, que se intenta producir, usando la ruta habitual) es DHA o EPA, por ejemplo (por ejemplo, se produce usando elongasas y desaturasas que producirán el DHA o EPA a partir de los productos de los sistemas de FAS), se producirá una variedad de productos intermedios y productos secundarios además de DHA o EPA, y estos productos intermedios o secundarios representan con frecuencia la mayoría de los productos producidos por la ruta, o al menos están presentes en cantidades significativas en los lípidos del organismo de producción. Tales productos intermedios y secundarios incluyen, pero no se limitan a, ácidos grasos que tienen menos carbonos y/o menos dobles enlaces que el AGPI objetivo, o primario, y pueden incluir productos secundarios de ácido graso no habituales que pueden tener el mismo número de carbonos que el AGPI objetivo o primario, pero que pueden tener dobles enlaces en posiciones no habituales. A modo de ejemplo, en la producción de EPA usando la ruta habitual (por ejemplo, véase la publicación de solicitud de patente estadounidense 2004/0172682), aunque el AGPI objetivo de la ruta es EPA (es decir, debido al uso de elongasas y desaturasas que actúan específicamente sobre los productos de los sistemas de FAS para producir EPA), los aceites producidos por el sistema incluyen una variedad de productos intermedios y secundarios incluyendo: ácido gamma-linolénico (GLA; 18:3, n-6); ácido estearidónico (STA o SDA; 18:4, n-3); ácido dihomo-gamma-linolénico (DGLA o HGLA; 20:3, n-6), ácido araquidónico (ARA, C20:4, n-6); ácido eicosatrienoico (ETA; 20:3, n-9) y diversos otros productos intermedios o secundarios, tales como 20:0; 20:1 ($\Delta 5$); 20:1 ($\Delta 11$); 20:2 ($\Delta 8,11$); 20:2 ($\Delta 11,14$); 20:3 ($\Delta 5,11,14$); 20:3 ($\Delta 11,14,17$); ácido de Mead (20:3; $\Delta 5,8,11$); o 20:4 ($\Delta 5,1,14,17$). Los productos intermedios del sistema también pueden incluir AGPI de cadena larga que no son el objetivo de la modificación genética (por ejemplo, un sistema enzimático de ruta habitual para producir DHA puede producir realmente más EPA como producto intermedio que DHA).

En cambio, la sintasa PKS de AGPI de la presente invención no utiliza los productos de ácidos grasos de sistemas de FAS. En vez de eso, produce el producto de AGPI final (el producto de AGPI primario) a partir de la misma molécula precursora pequeña que se utiliza por FAS y elongasas (malonil-CoA). Por tanto, no se liberan productos intermedios en el ciclo de síntesis en ninguna cantidad significativa, y el producto de AGPI (también denominado en el presente documento producto de AGPI primario) se transfiere eficientemente a fracciones de fosfolípidos (PL) y triacilglicerol (TAG) de los lípidos. En efecto, un sistema de PKS de AGPI puede producir dos productos de AGPI objetivo o primarios (por ejemplo, el sistema de PKS de AGPI de *Schizochytrium* produce tanto DHA como DPAn-6 como productos primarios), pero DPA no es un producto intermedio en la ruta para producir DHA. En vez de eso, cada uno es un producto separado del mismo sistema de PKS de AGPI. Por tanto, los genes de PKS de AGPI de la presente invención son un medio excelente de producción de aceites que contienen AGPI, y particularmente, AGPICL en un huésped heterólogo, tal como una planta, en el que los aceites están sustancialmente libres (definido a continuación) de los productos intermedios y productos secundarios que contaminan los aceites producidos por la ruta de AGPI "habitual".

Por tanto, un objeto de la presente invención es producir, mediante manipulación genética de plantas tal como se describe en el presente documento, ácidos grasos poliinsaturados y, por extensión, aceites obtenidos a partir de tales plantas (por ejemplo, obtenidos a partir de las semillas oleaginosas de tales plantas) que comprenden estos AGPI. Los ejemplos de los AGPI que puede producirse por la presente invención incluyen, pero no se limitan a, DHA (ácido docosahexaenoico (C22:6, n-3)), ARA (ácido eicosatetraenoico o ácido araquidónico (C20:4, n-6)), DPA (ácido docosapentaenoico (C22:5, n-6 o n-3)) y EPA (ácido eicosapentaenoico (C20:5, n-3)). La presente invención permite la producción de lípidos comercialmente valiosos enriquecidos en uno o más AGPI deseados (objetivos o primarios) mediante el desarrollo de los presentes inventores de plantas modificadas genéticamente mediante el uso del sistema de policétido sintasa de la presente invención, así como componentes del mismo, que produce AGPI.

Según la presente invención, la referencia a un "AGPI primario", "AGPI objetivo", "AGPI pretendido" o "AGPI deseado" se refiere al AGPI o a los AGPI particulares que son el producto pretendido u objetivo de la ruta enzimática que se usa para producir el/los AGPI. Por ejemplo, cuando se usan elongasas y desaturasas para modificar productos de los sistemas de FAS, pueden seleccionarse combinaciones particulares de elongasas y desaturasas que, cuando se usan juntas, producirán un AGPI objetivo o deseado (por ejemplo, DHA o EPA). Tal como se comentó anteriormente, tal AGPI objetivo o deseado producido por la ruta habitual puede no ser realmente un AGPI "primario" en cuanto a la cantidad de AGPI como porcentaje de ácidos grasos totales producidos por el sistema, debido a la formación de productos intermedios y secundarios que pueden representar realmente la mayoría de los

productos producidos por el sistema. Sin embargo, puede usarse el término “AGPI primario” incluso en ese caso para hacer referencia al producto de AGPI objetivo o pretendido producido por las elongasas o desaturasas usadas en el sistema.

5 Cuando se usa un sistema de PKS de AGPI tal como se prefiere en la presente invención, un sistema de PKS de AGPI dado derivado de un organismo particular producirá AGPI particular(es), de manera que la selección de un sistema de PKS de AGPI de un organismo particular dará como resultado la producción de AGPI objetivo o primarios especificados. Por ejemplo, el uso de un sistema de PKS de AGPI de *Schizochytrium* dará como resultado la producción de DHA y DPAn-6 como AGPI objetivo o primarios. Por otro lado, el uso de un sistema de PKS de AGPI de diversas especies de *Shewanella* dará como resultado la producción de EPA como AGPI objetivo o primario. Se observa que la razón de los AGPI objetivo o primarios puede diferir dependiendo de la selección del sistema de PKS de AGPI particular y de cómo responde ese sistema a las condiciones específicas en las que se expresa. Por ejemplo, el uso de un sistema de PKS de AGPI de *Thraustochytrium* 23B (ATCC n.º 20892) también dará como resultado la producción de DHA y DPAn-6 como AGPI objetivo o primarios; sin embargo, en el caso de *Thraustochytrium* 23B, la razón de DHA con respecto a DPAn-6 es de aproximadamente 10:1 (y puede oscilar entre aproximadamente 8:1 y aproximadamente 40:1), mientras que en *Schizochytrium*, normalmente la razón es de aproximadamente 2,5:1. Por tanto, el uso de un sistema o proteínas o dominios PKS de AGPI de *Thraustochytrium* puede alterar la razón de AGPI producidos por un organismo en comparación con *Schizochytrium* incluso aunque los AGPI objetivo sean los mismos. Sin embargo, tal como se detalló anteriormente, el uso de diversas proteínas y dominios con proteínas y dominios de otros sistemas de PKS de AGPI u otros sistemas de PKS (que producen moléculas bioactivas distintas de AGPI) puede combinarse (“mezclarse y hacerse corresponder”) para producir proteínas quiméricas y/o sistemas de PKS de AGPI quiméricos (descritos anteriormente), dando como resultado la producción de perfiles de AGPI diferentes, incluyendo tipos de AGPI, cantidades y/o razones de un AGPI con respecto a otro diferentes.

25 Cuando se usa un sistema de PKS de AGPI de la presente invención, los aceites producidos por el organismo, tal como una planta, están sustancialmente libres de productos intermedios o secundarios que no son los productos de AGPI objetivo o primarios y que no se producen de manera natural por el sistema de FAS endógeno en el organismo de tipo natural (por ejemplo, las plantas de tipo natural producen algunos AGPI de cadena más corta o media, tales como AGPI de 18 carbonos, mediante los sistemas de FAS, pero se producirán ácidos grasos nuevos, o adicionales, en la planta como resultado de la modificación genética con un sistema de PKS de AGPI). En otras palabras, en comparación con el perfil de ácidos grasos totales de la planta de tipo natural (no modificada genéticamente) o la planta original usada como receptor de la modificación genética indicada, la mayoría de los ácidos grasos adicionales en el perfil de ácidos grasos totales producidos por plantas que se han modificado genéticamente con el sistema de PKS de AGPI de la presente invención (o un componente del mismo) comprenden los productos de AGPI objetivo o pretendidos del sistema de PKS de AGPI (es decir, la mayoría de los ácidos grasos adicionales en los ácidos grasos totales que se producen por la planta modificada genéticamente son el/los AGPI objetivo).

40 Según la presente invención, la referencia a “productos intermedios” o “productos secundarios” de un sistema enzimático que produce AGPI se refiere a cualquier producto, y particularmente, producto de ácidos grasos, que se produce por el sistema enzimático como resultado de la producción del/de los AGPI objetivo o primario(s) del sistema, pero que no son el/los AGPI primario(s) u objetivo. En una realización, los productos intermedios y secundarios pueden incluir ácidos grasos no objetivo que se producen de manera natural por la planta de tipo natural o por la planta original usada como receptor para la modificación genética indicada, pero ahora se clasifican como productos intermedios o secundarios porque se producen en niveles superiores como resultado de la modificación genética, en comparación con los niveles producidos por la planta de tipo natural o por la planta original usada como receptor para la modificación genética indicada. Los productos intermedios y secundarios son particularmente significativos en la ruta habitual para la síntesis de AGPI y son sustancialmente menos significativos en la ruta de PKS de AGPI, tal como se comentó anteriormente. Se observa que un AGPI primario u objetivo de un sistema enzimático puede ser un producto intermedio de un sistema enzimático diferente en el que el producto primario u objetivo es un AGPI diferente, y esto es particularmente cierto para productos de la ruta habitual de producción de AGPI, ya que el sistema de PKS de AGPI evita sustancialmente la producción de productos intermedios. Por ejemplo, cuando se usa la ruta habitual para producir EPA, se producen ácidos grasos tales como GLA, DGLA y SDA como productos intermedios en cantidades significativas (por ejemplo, la publicación de solicitud de patente estadounidense 2004/0172682 ilustra este punto). De manera similar, y también ilustrado por la publicación de solicitud de patente estadounidense 2004/0172682, cuando se usa la ruta habitual para producir DHA, además de los ácidos grasos mencionados anteriormente, se producen ETA y EPA (de manera notable, el AGPI objetivo en el primer ejemplo anterior) en cantidades significativas y de hecho, pueden estar presentes en cantidades significativamente superiores con respecto al producto de ácidos grasos totales al propio AGPI objetivo. Este último punto también se muestra en la publicación de solicitud de patente estadounidense 2004/0172682, en la que una planta que se modificó por ingeniería genética para producir DHA por la ruta habitual produce más EPA como porcentaje de ácidos grasos totales que el DHA objetivo.

65 Además, estar “sustancialmente libre” de productos intermedios o secundarios del sistema para sintetizar AGPI, o no tener productos intermedios o secundarios presentes en cantidades sustanciales, significa que cualquier ácido graso como producto intermedio o secundario (AGPI no objetivo) que se produce en la planta (y/o partes de plantas y/o

fracción de aceite de semilla) modificada genéticamente como resultado de la introducción o presencia del sistema enzimático para producir AGPI (es decir, que no se producen por la planta de tipo natural o la planta original usada como receptor para la modificación genética indicada), están presentes en una cantidad que es de menos de aproximadamente el 10% en peso de los ácidos grasos totales producidos por la planta, y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% en peso de los ácidos grasos totales producidos por la planta, y más preferiblemente menos de aproximadamente el 0,5% en peso de los ácidos grasos totales producidos por la planta.

En una realización preferida, estar "sustancialmente libre" de productos intermedios o secundarios del sistema para sintetizar AGPI, o no tener productos intermedios o secundarios presentes en cantidades sustanciales, significa que cualquier ácido graso como producto intermedio o secundario que se produce en la planta (y/o partes de plantas y/o en fracción de aceite de semilla) modificada genéticamente como resultado del sistema enzimático para producir los AGPI (es decir, que no se producen por la planta de tipo natural o por la planta original usada como receptor para la modificación genética indicada para la producción de AGPI objetivo), están presentes en una cantidad que es de menos de aproximadamente el 10% en peso de los ácidos grasos adicionales totales producidos por la planta (definiéndose los ácidos grasos adicionales como aquellos ácidos grasos o niveles de ácidos grasos que no se producen de manera natural por la planta de tipo natural o por la planta original que se usa como receptor para la modificación genética indicada para la producción de AGPI objetivo), y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% de los ácidos grasos adicionales totales producidos por la planta. Por tanto, al contrario que el perfil de ácidos grasos de plantas que se han modificado genéticamente para producir AGPI mediante la ruta habitual, la mayoría de los productos de ácidos grasos resultantes de la modificación genética con un sistema de PKS de AGPI serán los productos de ácidos grasos objetivo o pretendidos.

Cuando el producto objetivo de un sistema de PKS de AGPI es un AGPI de cadena larga, tal como DHA o DPA (n-6 o n-3) producido por el sistema de PKS de AGPI de la invención descrito en el presente documento, los productos intermedios y productos secundarios que no están presentes en cantidades sustanciales en los lípidos totales de plantas modificadas genéticamente con tal PKS de AGPI pueden incluir, pero no se limitan a: ácido gamma-linolénico (GLA; 18:3, n-6); ácido estearidónico (STA o SDA; 18:4, n-3); ácido dihomogamma-linolénico (DGLA o HGLA; 20:3, n-6), ácido araquidónico (ARA, C20:4, n-6); ácido eicosatrienoico (ETA; 20:3, n-9) y diversos otros productos intermedios o secundarios, tales como 20:0; 20:1 ($\Delta 5$); 20:1 ($\Delta 11$); 20:2 ($\Delta 8,11$); 20:2 ($\Delta 11,14$); 20:3 ($\Delta 5,11,14$); 20:3 ($\Delta 11,14,17$); ácido de Mead (20:3; $\Delta 5,8,11$); o 20:4 ($\Delta 5,1,14,17$). Además, cuando el producto objetivo es un AGPI particular, tal como DHA, los productos intermedios y productos secundarios que no están presentes en cantidades sustanciales en los lípidos totales de las plantas modificadas genéticamente también incluyen otros AGPI, incluyendo otros AGPI que son un producto natural de un sistema de PKS de AGPI diferente, tal como EPA en este ejemplo. En algunos sistemas, un sistema de PKS de AGPI puede producir más de un AGPI, tal como tanto un AGPI C22 como uno C20, y tales combinaciones de AGPI pueden representar el producto objetivo, mientras que otros AGPI pueden representar productos intermedios o secundarios. Debe observarse que el sistema de PKS de AGPI de la presente invención también puede usarse, si se desea, para producir como AGPI objetivo un AGPI que puede incluir GLA, SDA o DGLA (haciendo referencia a realizaciones en las que se producen aceites usando componentes de un sistema de PKS de AGPI descrito en el presente documento).

Usando el conocimiento de la base genética y la estructura de dominios del sistema de PKS de AGPI descrito en el presente documento, los presentes inventores han diseñado y producido constructos que codifican para un sistema de PKS de AGPI de este tipo y han producido satisfactoriamente plantas transgénicas que expresan el sistema de PKS de AGPI. Las plantas transgénicas producen aceites que contienen AGPI, y los aceites están sustancialmente libres de productos intermedios que se acumulan en una ruta de AGPI habitual (véase la publicación de solicitud de patente estadounidense n.º 20070089199, citada anteriormente). Los presentes inventores también han demostrado el uso de los constructos para producir AGPI en *E. coli*, y también en otro eucariota, la levadura, como experimento de prueba de concepto antes de la producción de las plantas transgénicas (publicación de solicitud de patente estadounidense n.º 20070089199, citada anteriormente). Los ejemplos demuestran que la transformación tanto de levadura como de plantas con un sistema de PKS de AGPI que produce DHA y DPAn-6 como los AGPI objetivo produce ambos de estos AGPI como ácidos grasos adicionales primarios en los ácidos grasos totales de la planta (es decir, restando los ácidos grasos que se producen en la planta de tipo natural), y en la levadura y, además, que cualquier otro ácido graso que no está presente en los ácidos grasos de la planta de tipo natural son prácticamente indetectables. En otra parte en el presente documento se describen en detalle características específicas de plantas modificadas genéticamente y partes y aceites de las mismas de la presente invención.

Por consiguiente, se describe un método para producir moléculas bioactivas deseadas (también denominadas

productos o compuestos) haciendo crecer o cultivando un microorganismo modificado genéticamente o una planta modificada genéticamente de la presente invención (descrito en detalle anteriormente). Tal método incluye la etapa de cultivar en un medio de crecimiento o fermentación o hacer crecer en un entorno adecuado, tal como tierra, un microorganismo o una planta, respectivamente, que tiene una modificación genética tal como se describió anteriormente en el presente documento y según la presente invención. En una realización preferida, el método para producir moléculas bioactivas incluye la etapa de cultivar, en condiciones eficaces para producir la molécula bioactiva, un organismo modificado genéticamente que expresa un sistema de PKS que comprende al menos un dominio biológicamente activo de un sistema de policétido sintasa (PKS) de ácidos grasos poliinsaturados (AGPI) tal como se describe en el presente documento.

En el método de producción de compuestos bioactivos deseados, se cultiva o se 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 el que un microorganismo modificado genéticamente de la presente invención, cuando se cultiva, puede producir el producto deseado. Tal medio es normalmente un medio acuoso que comprende fuentes de carbono, nitrógeno y fosfato asimilables. Tal medio también puede incluir sales, minerales, metales y otros nutrientes apropiados. Los microorganismos de la presente invención pueden cultivarse en biorreactores de fermentación convencionales. Los microorganismos pueden cultivarse mediante cualquier procedimiento de fermentación incluyendo, pero sin limitarse a, fermentación discontinua, semicontinua, con recirculación de células y continua. Las condiciones de crecimiento preferidas para posibles microorganismos huéspedes según la presente invención se conocen bien en la técnica. Las moléculas bioactivas deseadas producidas por el microorganismo modificado genéticamente pueden recuperarse del medio de fermentación usando técnicas de separación y purificación convencionales. Por ejemplo, puede filtrarse o centrifugarse el medio de fermentación para eliminar microorganismos, residuos celulares y otra materia particulada, y puede recuperarse el producto a partir del sobrenadante libre de células mediante métodos convencionales, tales como, por ejemplo, intercambio iónico, cromatografía, extracción, extracción con disolventes, separación por membrana, electrodiálisis, ósmosis inversa, destilación, derivatización química y cristalización. Alternativamente, pueden usarse microorganismos que producen el compuesto deseado, o extractos y diversas fracciones del mismo, sin eliminar los componentes del microorganismo del producto.

En el método para la producción de compuestos bioactivos deseados, se cultiva una planta modificada genéticamente o parte de la planta (incluyendo una célula de planta) en un medio de crecimiento o se hace crecer en un medio adecuado tal como tierra, según sea apropiado. Anteriormente se comentó en detalle un medio de crecimiento o de cultivo apropiado, o eficaz. Un medio de crecimiento adecuado para plantas superiores incluye cualquier medio de crecimiento para plantas, incluyendo, pero sin limitarse a, tierra, arena, cualquier otro medio particulado que soporte el crecimiento de raíces (por ejemplo vermiculita, perlita, etc.) o cultivo hidropónico, así como luz, agua y complementos nutricionales adecuados que optimizan el crecimiento de la planta superior. Las plantas modificadas genéticamente de la presente invención se modifican por ingeniería genética para producir cantidades significativas del producto deseado mediante la actividad del sistema de PKS de AGPI que se modifica genéticamente según la presente invención. Los compuestos pueden recuperarse mediante procedimientos de purificación que extraen los compuestos de la planta. En una realización preferida, el compuesto se recupera cosechando la planta. En una realización particularmente preferida, se recuperan AGPI de la planta o parte de la planta cosechando el aceite de la planta o parte de la planta (por ejemplo, de las semillas oleaginosas). En esta realización, la planta puede consumirse en su estado natural o procesarse adicionalmente para obtener productos consumibles.

Las moléculas bioactivas, según la presente invención, incluyen cualquier molécula (compuesto, producto, etc.) que tiene una actividad biológica, y que puede producirse por un sistema de PKS que comprende al menos una secuencia de aminoácidos que tiene una actividad biológica de al menos un dominio funcional de un sistema de PKS de AGPI no bacteriano tal como se describe en el presente documento. Tales moléculas bioactivas pueden incluir, pero no se limitan a: un ácido graso poliinsaturado (AGPI), una formulación antiinflamatoria, un agente quimioterápico, un excipiente activo, un fármaco para la osteoporosis, un antidepresivo, un anticonvulsivo, un fármaco anti-*Helicobacter pylori*, un fármaco para el tratamiento de una enfermedad neurodegenerativa, un fármaco para el tratamiento de una enfermedad hepática degenerativa, un antibiótico y una formulación hipocolesterolemia. Una ventaja del sistema de PKS de AGPI no bacteriano de la presente invención es la capacidad de tal sistema de introducir dobles enlaces carbono-carbono en la configuración cis, y moléculas que incluyen un doble enlace cada tres carbonos. Esta capacidad puede utilizarse para producir una variedad de compuestos.

Con respecto a microorganismos, preferiblemente, se producen compuestos bioactivos de interés por el microorganismo modificado genéticamente en una cantidad que es de más de aproximadamente el 0,05%, y 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%, 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. Para compuestos lipídicos, preferiblemente, tales compuestos se producen

en una cantidad que es de más de aproximadamente el 5% del peso seco del microorganismo. Otros compuestos bioactivos, tales como antibióticos o compuestos que se sintetizan en cantidades más pequeñas, pueden producirse en cantidades conocidas por los expertos en la técnica, y las cepas que presentan tales compuestos se identifican como que contienen de manera predecible un sistema de PKS novedoso del tipo descrito en el presente documento.

En algunas realizaciones, el microorganismo secreta moléculas (compuestos) bioactivas particulares, en vez de acumularlas en las células. Por tanto, tales moléculas bioactivas se recuperan generalmente del medio de cultivo y la concentración de la molécula producida variará dependiendo del microorganismo y del tamaño del cultivo, y puede medirse en g/l, en vez de en peso celular seco.

Preferiblemente, un organismo modificado genéticamente (por ejemplo, microorganismo o planta) de la invención produce uno o más ácidos grasos poliinsaturados incluyendo, pero sin limitarse a, EPA (C20:5, n-3), DHA (C22:6, n-3), DPA (C22:5, n-6 o n-3), ARA (C20:4, n-6), GLA (C18:3, n-6), ALA (C18:3, n-3) y/o SDA (C18:4, n-3)), y más preferiblemente, uno o más ácidos grasos de cadena larga (AGPICL), incluyendo, pero sin limitarse a, EPA (C20:5, n-3), DHA (C22:6, n-3), DPA (C22:5, n-6 o n-3) o DTA (C22:4, n-6). En una realización particularmente preferida, un organismo modificado genéticamente de la invención produce uno o más ácidos grasos poliinsaturados incluyendo, pero sin limitarse a, EPA (C20:5, n-3), DHA (C22:6, n-3) y/o DPA (C22:5, n-6 o n-3).

Preferiblemente, un organismo modificado genéticamente de la invención produce al menos un AGPI (el AGPI objetivo), en el que el perfil de ácidos grasos totales en el organismo (o una parte del organismo que acumula AGPI, tal como semillas maduras o aceite de tales semillas, si el organismo es una planta de semillas oleaginosas), comprende una cantidad detectable de este AGPI o estos AGPI. Preferiblemente, el AGPI es un AGPI de al menos 20 carbonos y comprende al menos 3 dobles enlaces, y más preferiblemente al menos 4 dobles enlaces, e incluso más preferiblemente, al menos 5 dobles enlaces. En una realización, el AGPI es un AGPI que no se produce de manera natural por el organismo en cantidades detectables o significativas (por ejemplo, el organismo de tipo natural en ausencia de la modificación genética o el organismo original usado como receptor para la modificación genética indicada).

Preferiblemente, el perfil de ácidos grasos totales en el organismo (o parte del organismo que acumula AGPI) comprende al menos el 0,1% del/de los AGPI objetivo en peso de los ácidos grasos totales, y más preferiblemente al menos aproximadamente el 0,2%, y más preferiblemente al menos aproximadamente el 0,3%, y más preferiblemente al menos aproximadamente el 0,4%, y más preferiblemente al menos aproximadamente el 0,5%, y más preferiblemente al menos aproximadamente el 1%, y más preferiblemente al menos aproximadamente el 2%, y más preferiblemente al menos aproximadamente el 3%, y más preferiblemente al menos aproximadamente el 4%, y más preferiblemente al menos aproximadamente el 5%, y más preferiblemente al menos aproximadamente el 10%, y más preferiblemente al menos aproximadamente el 15%, y más preferiblemente al menos aproximadamente el 20%, y más preferiblemente al menos aproximadamente el 25%, y más preferiblemente al menos aproximadamente el 30%, y más preferiblemente al menos aproximadamente el 35%, y más preferiblemente al menos aproximadamente el 40%, y más preferiblemente al menos aproximadamente el 45%, y más preferiblemente al menos aproximadamente el 50%, y más preferiblemente al menos aproximadamente el 55%, y más preferiblemente al menos aproximadamente el 60%, y más preferiblemente al menos aproximadamente el 65%, y más preferiblemente al menos aproximadamente el 70%, y más preferiblemente al menos aproximadamente el 75%, y más preferiblemente más del 75% de al menos un ácido graso poliinsaturado (el AGPI objetivo) en peso de los ácidos grasos totales, o cualquier porcentaje desde el 0,1% hasta el 75%, o superior al 75% (hasta el 100% o aproximadamente el 100%), en incrementos del 0,1%, del/de los AGPI objetivo. Tal como se usa generalmente en el presente documento, la referencia a una cantidad en porcentaje de producción de AGPI es en peso de los ácidos grasos totales producidos por el organismo, a menos que se mencione lo contrario (por ejemplo, en algunos casos, el porcentaje en peso es con respecto a los ácidos grasos totales producidos por un complejo enzimático, tal como un sistema de PKS de AGPI). En una realización, los ácidos grasos totales producidos por una planta se presentan como porcentaje en peso tal como se determina mediante análisis por cromatografía de gases (CG) de una preparación de ésteres metílicos de ácidos grasos (EMAG).

Tal como se describió anteriormente, una característica adicional de los ácidos grasos totales producidos por la planta descrita anteriormente (y/o partes de plantas o fracción de aceite de semilla) es que estos ácidos grasos totales producidos por la planta comprenden menos de (o no contienen más de) aproximadamente el 10% en peso de cualquier ácido graso distinto del/de los AGPI objetivo que se produce(n) por el complejo enzimático que produce el/los AGPI objetivo. Preferiblemente, cualquier ácido graso que se produce por el complejo enzimático que produce el/los AGPI objetivo (por ejemplo, como resultado de la modificación genética de la planta con la enzima o el complejo enzimático que produce el/los AGPI objetivo), distinto del/de los AGPI objetivo, está presente a menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% en peso de los ácidos grasos totales producidos por la planta.

En otra realización, cualquier ácido graso que se produce por el complejo enzimático que produce el/los AGPI objetivo distinto del/de los AGPI objetivo está presente a menos de (o no contiene más de) aproximadamente el 10% en peso de los ácidos grasos totales que se producen por el complejo enzimático que produce el/los AGPI objetivo en la planta (es decir, esta medida se limita a aquellos ácidos grasos totales que se producen por el complejo enzimático que produce los AGPI objetivo), y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% en peso de los ácidos grasos totales, y más preferiblemente menos de aproximadamente el 0,5% en peso de los ácidos grasos totales que se producen por el complejo enzimático que produce el/los AGPI objetivo en la planta.

En otro aspecto de esta realización de la invención, los ácidos grasos totales producidos por la planta (y/o partes de plantas o fracción de aceite de semilla) contienen menos del (o no contienen más del) 10% de AGPI que tienen 18 o más carbonos en peso de los ácidos grasos totales producidos por la planta, distintos del/de los AGPI objetivo o de los AGPI que están presentes en la planta de tipo natural (no modificada genéticamente) o en la planta original usada como receptor para la modificación genética indicada (inicial o secuencial). En aspectos adicionales, los ácidos grasos totales producidos por la planta (y/o partes de plantas o fracción de aceite de semilla) contienen menos del 9% de AGPI que tienen 18 o más carbonos, o menos del 8% de AGPI que tienen 18 o más carbonos, o menos del 7% de AGPI que tienen 18 o más carbonos, o menos del 6% de AGPI que tienen 18 o más carbonos, o menos del 5% de AGPI que tienen 18 o más carbonos, o menos del 4% de AGPI que tienen 18 o más carbonos, o menos del 3% de AGPI que tienen 18 o más carbonos, o menos del 2% de AGPI que tienen 18 o más carbonos, o menos del 1% de AGPI que tienen 18 o más carbonos en peso de los ácidos grasos totales producidos por la planta, distintos del/de los AGPI objetivo o de los AGPI que están presentes en la planta de tipo natural (no modificada genéticamente) o la planta original usada como receptor para la modificación genética indicada.

En otro aspecto de esta realización de la invención, los ácidos grasos totales producidos por la planta (y/o partes de plantas o fracción de aceite de semilla) contienen menos del (o no contienen más del) 10% de AGPI que tienen 20 o más carbonos en peso de los ácidos grasos totales producidos por la planta, distintos del/de los AGPI objetivo o de los AGPI que están presentes en la planta de tipo natural (no modificada genéticamente) o la planta original usada como receptor para la modificación genética indicada (inicial o secuencial). En aspectos adicionales, los ácidos grasos totales producidos por la planta (y/o partes de plantas o fracción de aceite de semilla) contienen menos del 9% de AGPI que tienen 20 o más carbonos, o menos del 8% de AGPI que tienen 20 o más carbonos, o menos del 7% de AGPI que tienen 20 o más carbonos, o menos del 6% de AGPI que tienen 20 o más carbonos, o menos del 5% de AGPI que tienen 20 o más carbonos, o menos del 4% de AGPI que tienen 20 o más carbonos, o menos del 3% de AGPI que tienen 20 o más carbonos, o menos del 2% de AGPI que tienen 20 o más carbonos, o menos del 1% de AGPI que tienen 20 o más carbonos en peso de los ácidos grasos totales producidos por la planta, distintos del/de los AGPI objetivo o de los AGPI que están presentes en la planta de tipo natural (no modificada genéticamente) o la planta original usada como receptor para la modificación genética indicada.

En una realización, los ácidos grasos totales en la planta (y/o partes de plantas o fracción de aceite de semilla) contienen menos de aproximadamente el 10% en peso de los ácidos grasos totales producidos por la planta, y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% de un ácido graso seleccionado de uno cualquiera o más de: ácido gamma-linolénico (GLA; 18:3, n-6); ácido estearidónico (STA o SDA; 18:4, n-3); ácido dihomogamma-linolénico (DGLA o HGLA; 20:3, n-6), ácido araquidónico (ARA, C20:4, n-6); ácido eicosatrienoico (ETA; 20:3, n-9) y diversos otros ácidos grasos, tales como 20:0; 20:1 (Δ 5); 20:1 (Δ 11); 20:2 (Δ 8,11); 20:2 (Δ 11,14); 20:3 (Δ 5,11,14); 20:3 (Δ 11,14,17); ácido de Mead (20:3; Δ 5,8,11); o 20:4 (Δ 5,1,14,17).

En otra realización, los ácidos grasos que se producen por el sistema enzimático que produce los AGPI de cadena larga en la planta contienen menos de aproximadamente el 10% en peso de un ácido graso seleccionado de: ácido gamma-linolénico (GLA; 18:3, n-6); ácido estearidónico (STA o SDA; 18:4, n-3); ácido dihomogamma-linolénico (DGLA o HGLA; 20:3, n-6), ácido araquidónico (ARA, C20:4, n-6); ácido eicosatrienoico (ETA; 20:3, n-9) y diversos otros ácidos grasos, tales como 20:0; 20:1 (Δ 5); 20:1 (Δ 11); 20:2 (Δ 8,11); 20:2 (Δ 11,14); 20:3 (Δ 5,11,14); 20:3 (Δ 11,14,17); ácido de Mead (20:3; Δ 5,8,11); o 20:4 (Δ 5,1,14,17), como porcentaje de los ácidos grasos totales producidos por la planta, y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% de un ácido graso seleccionado de: ácido gamma-linolénico (GLA; 18:3, n-6); ácido estearidónico (STA o SDA; 18:4, n-3); ácido dihomogamma-linolénico (DGLA o HGLA; 20:3, n-6), ácido araquidónico (ARA, C20:4, n-6); ácido

eicosatrienoico (ETA; 20:3, n-9) y diversos otros ácidos grasos, tales como 20:0; 20:1 ($\Delta 5$); 20:1 ($\Delta 11$); 20:2 ($\Delta 8,11$); 20:2 ($\Delta 11,14$); 20:3 ($\Delta 5,11,14$); 20:3 ($\Delta 11,14,17$); ácido de Mead (20:3; $\Delta 5,8,11$); o 20:4 ($\Delta 5,1,14,17$).

En otra realización, los ácidos grasos que se producen por el sistema enzimático que produce los AGPI de cadena larga en la planta contienen menos de aproximadamente el 10% en peso de todos de los siguientes AGPI: ácido gamma-linolénico (GLA; 18:3, n-6), AGPI que tienen 18 carbonos y cuatro dobles enlaces carbono-carbono, AGPI que tienen 20 carbonos y tres dobles enlaces carbono-carbono, y AGPI que tienen 22 carbonos y dos o tres dobles enlaces carbono-carbono, como porcentaje de los ácidos grasos totales producidos por la planta, y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% de todos de los siguientes AGPI: ácido gamma-linolénico (GLA; 18:3, n-6), AGPI que tienen 18 carbonos y cuatro dobles enlaces carbono-carbono, AGPI que tienen 20 carbonos y tres dobles enlaces carbono-carbono, y AGPI que tienen 22 carbonos y dos o tres dobles enlaces carbono-carbono.

En otra realización, los ácidos grasos que se producen por el sistema enzimático que produce los AGPI de cadena larga en la planta contienen menos de aproximadamente el 10% en peso de cada uno de los siguientes AGPI: ácido gamma-linolénico (GLA; 18:3, n-6), AGPI que tienen 18 carbonos y cuatro dobles enlaces carbono-carbono, AGPI que tienen 20 carbonos y tres dobles enlaces carbono-carbono, y AGPI que tienen 22 carbonos y dos o tres dobles enlaces carbono-carbono, como porcentaje de los ácidos grasos totales producidos por la planta, y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% de cada uno de los siguientes AGPI: ácido gamma-linolénico (GLA; 18:3, n-6), AGPI que tienen 18 carbonos y cuatro dobles enlaces carbono-carbono, AGPI que tienen 20 carbonos y tres dobles enlaces carbono-carbono, y AGPI que tienen 22 carbonos y dos o tres dobles enlaces carbono-carbono.

En otra realización, los ácidos grasos que se producen por el sistema enzimático que produce los AGPI de cadena larga en la planta contienen menos de aproximadamente el 10% en peso de uno cualquiera o más de los siguientes AGPI: ácido gamma-linolénico (GLA; 18:3, n-6), AGPI que tienen 18 carbonos y cuatro dobles enlaces carbono-carbono, AGPI que tienen 20 carbonos y tres dobles enlaces carbono-carbono, y AGPI que tienen 22 carbonos y dos o tres dobles enlaces carbono-carbono, como porcentaje de los ácidos grasos totales producidos por la planta, y más preferiblemente menos de aproximadamente el 9%, y más preferiblemente menos de aproximadamente el 8%, y más preferiblemente menos de aproximadamente el 7%, y más preferiblemente menos de aproximadamente el 6%, y más preferiblemente menos de aproximadamente el 5%, y más preferiblemente menos de aproximadamente el 4%, y más preferiblemente menos de aproximadamente el 3%, y más preferiblemente menos de aproximadamente el 2%, y más preferiblemente menos de aproximadamente el 1% de uno cualquiera o más de los siguientes AGPI: ácido gamma-linolénico (GLA; 18:3, n-6), AGPI que tienen 18 carbonos y cuatro dobles enlaces carbono-carbono, AGPI que tienen 20 carbonos y tres dobles enlaces carbono-carbono, y AGPI que tienen 22 carbonos y dos o tres dobles enlaces carbono-carbono.

En un aspecto de esta realización de la invención, la planta produce al menos dos AGPI objetivo, y el perfil de ácidos grasos totales en la planta, o la parte de la planta que acumula AGPI (incluyendo aceites de las semillas oleaginosas), comprende una cantidad detectable de estos AGPI. En esta realización, los AGPI son preferiblemente cada uno al menos un AGPI de 20 carbonos y comprenden al menos 3 dobles enlaces, y más preferiblemente al menos 4 dobles enlaces, e incluso más preferiblemente, al menos 5 dobles enlaces. Tales AGPI se eligen lo más preferiblemente de DHA, DPAn-6 y EPA. En un aspecto, la planta produce DHA y DPAn-6, y la razón de DHA con respecto a DPAn-6 es de aproximadamente 1:10 a aproximadamente 10:1 o superior, incluyendo cualquier razón entre las mismas. En una realización, la razón de DHA con respecto a DPA es de aproximadamente 1:1 a aproximadamente 3:1, y en otra realización, de aproximadamente 2,5:1. En una realización, la planta produce DHA y EPA.

La invención incluye además cualquier semilla producida por las plantas descritas anteriormente, así como cualquier parte de plantas, aceites producidos por las plantas o semillas producidas por las plantas. La invención también incluye cualquier producto producido usando las plantas, partes de plantas, semilla o aceites descritos en el presente documento.

También se describe un método para modificar un producto final que contiene al menos un ácido graso, que comprende añadir a dicho producto final un aceite producido por una célula huésped recombinante que expresa al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica para al menos un dominio biológicamente activo de un sistema de PKS de AGPI tal como se describe en el presente documento.

Preferiblemente, el producto final se selecciona del grupo que consiste en un alimento, un suplemento dietético, una formulación farmacéutica, una leche de animal humanizada y un preparado para lactantes. Las formulaciones farmacéuticas adecuadas incluyen, pero no se limitan a, una formulación antiinflamatoria, un agente quimioterápico, un excipiente activo, un fármaco para la osteoporosis, un antidepresivo, un anticonvulsivo, un fármaco anti-*Helicobacter pylori*, un fármaco para el tratamiento de una enfermedad neurodegenerativa, un fármaco para el tratamiento de una enfermedad hepática degenerativa, un antibiótico y una formulación hipocolesterolemizante. En una realización, el producto final se usa para tratar un estado seleccionado del grupo que consiste en: inflamación crónica, inflamación aguda, trastorno gastrointestinal, cáncer, caquexia, reestenosis cardíaca, trastorno neurodegenerativo, trastorno degenerativo del hígado, trastorno de lípidos sanguíneos, osteoporosis, osteoartritis, enfermedad autoinmunitaria, preeclampsia, parto prematuro, maculopatía relacionada con la edad, trastorno pulmonar y trastorno peroxisomal.

Los productos alimenticios adecuados incluyen, pero no se limitan a, productos de panadería fina, pan y bollos, cereales para el desayuno, queso procesado y sin procesar, condimentos (*ketchup*, mayonesa, etc.), productos lácteos (leche, yogur), pudines y postres de gelatina, bebidas carbonatadas, té, mezclas de bebida en polvo, productos de pescado procesados, bebidas a base de frutas, chicle, dulces duros, productos lácteos congelados, productos cárnicos procesados, frutos secos y cremas para untar a base de frutos secos, pasta, productos avícolas procesados, jugos y salsas, *snacks* de patatas fritas y otros *snacks*, chocolate y otros dulces, sopas y mezclas de sopa, productos a base de soja (leches, bebidas, natas, sustitutos lácteos para el café), cremas para untar a base de aceite vegetal y bebidas a base de verduras.

Se describe además un método para producir una leche de animal humanizada. Este método incluye las etapas de modificar genéticamente células productoras de leche de un animal productor de leche con al menos una molécula de ácido nucleico recombinante que comprende una secuencia de ácido nucleico que codifica para al menos un dominio biológicamente activo de un sistema de PKS de AGPI tal como se describe en el presente documento.

En la técnica se conocen métodos para modificar genéticamente una célula huésped y para producir un animal productor de leche, no humano, modificado genéticamente. Ejemplos de animales huéspedes para modificar incluyen ganado, ovejas, cerdos, cabras, yaks, etc., que son susceptibles de manipulación genética y clonación para una rápida expansión de una población que expresa un transgén. Para animales, pueden adaptarse transgenes similares a PKS para la expresión en orgánulos, tejidos y líquidos corporales diana mediante modificación de las regiones reguladoras del gen. Resulta de particular interés la producción de AGPI en la leche materna del animal huésped.

Los siguientes ejemplos se proporcionan con fines de ilustración y no se pretende que limiten el alcance de la presente invención.

Ejemplos

Ejemplo 1

El siguiente ejemplo describe la construcción de un vector de clonación de orfC de *Th.23B* sintético para su uso en *Schizochytrium*.

Se combinaron datos de uso de codones para cuatro genes grandes de *Schizochytrium* (por ejemplo, ATCC 20888 o *Schizochytrium* N230D) (orfA, orfB, orfC y FAS; descritos en la publicación de solicitud de patente estadounidense n.º 20020194641, la publicación de solicitud de patente estadounidense n.º 20070089199 o la publicación de solicitud de patente estadounidense n.º 20050191679). Dado que *Schizochytrium* ATCC 20888 produce altos niveles de ácidos grasos, se espera que estos genes se expresen altamente. Se eliminaron los codones con una representación inferior a aproximadamente el 3% (dentro de aquéllos para un aminoácido dado) y se ajustó el uso relativo de los codones restantes. La tabla 1 muestra el uso de codones de *Schizochytrium*, el uso ajustado y el uso de codones para el orfC de *Th.23B* no sintético. Se usó DNA2.0 (Menlo Park, CA) para analizar estos datos de uso de codones para diseñar y sintetizar una región codificante para el orfC de 23B de *Thraustochytrium*. Se añadieron nucleótidos a ambos extremos de la región codificante para codificar para sitios de reconocimiento de enzimas de restricción que facilitarán la posterior manipulación del gen sintético. Se ajustó un pequeño número de codones (sin cambiar el aminoácido codificado de SEQ ID NO: 62) para eliminar o añadir determinadas secuencias de reconocimiento de enzimas de restricción (véase a continuación un ejemplo). Se desarrolló la secuencia sintética resultante mediante DNA2.0 dentro de un vector de plásmido y se muestra en la figura 2B como "pThOrfC synth.". La tabla 1 muestra el uso de codones de la región codificante sintética.

Tabla 1.

aminoácido	codón	A, B y C de Schizo	FAS	Uso ajustado / diana	orfC de <i>Th.23B</i>	orfC de <i>Th.23B</i> sintético
		número	fracción	fracción	número	fracción
Arg	CGG	7	0,013	0	13	0,18
						0

ES 2 467 918 T3

Arg	CGA	6	0,011	0	13	0,18	0	0
Arg	CGT	94	0,173	0,21	17	0,24	11	0,15
Arg	CGC	436	0,803	0,79	17	0,24	61	0,85
Arg	AGG	0	0,000	0,00	9	0,13	0	0
Arg	AGA	0	0,000	0,00	3	0,04	0	0
Ser	TCG	244	0,327	0,34	19	0,19	32	0,33
Ser	TCA	10	0,013	0,00	16	0,16	0	0
Ser	TCT	64	0,086	0,10	12	0,12	10	0,10
Ser	TCC	230	0,308	0,29	19	0,19	32	0,33
Ser	AGT	19	0,025	0,00	12	0,12	0	0
Ser	AGC	179	0,240	0,27	20	0,20	24	0,24
Leu	CTG	111	0,123	0,13	36	0,28	13	0,10
Leu	CTA	2	0,002	0,00	7	0,05	0	0
Leu	CTT	148	0,164	0,18	33	0,26	33	0,26
Leu	CTC	623	0,690	0,69	27	0,21	82	0,64
Leu	TTG	18	0,020	0,00	21	0,16	0	0
Leu	TTA	1	0,001	0,00	4	0,03	0	0
Gly	GGG	7	0,009	0,00	21	0,18	0	0
Gly	GGA	38	0,047	0,04	33	0,29	5	0,04
Gly	GGT	174	0,216	0,25	17	0,15	35	0,30
Gly	GGC	585	0,728	0,71	44	0,38	75	0,65
Val	GTG	198	0,242	0,29	44	0,38	29	0,25
Val	GTA	4	0,005	0,00	14	0,12	0	0
Val	GTT	103	0,126	0,13	34	0,29	18	0,16
Val	GTC	512	0,627	0,58	24	0,21	69	0,59
Ala	GCG	214	0,159	0,17	21	0,18	20	0,17
Ala	GCA	41	0,031	0,00	36	0,31	0	0
Ala	GCT	236	0,176	0,21	33	0,28	25	0,22
Ala	GCC	853	0,635	0,62	26	0,22	71	0,61
Thr	ACG	156	0,297	0,28	19	0,30	21	0,33
Thr	ACA	13	0,025	0,00	8	0,13	0	0
Thr	ACT	71	0,135	0,22	16	0,25	10	0,16
Thr	ACC	285	0,543	0,50	20	0,32	32	0,51
Pro	CCG	195	0,340	0,32	19	0,24	27	0,35
Pro	CCA	12	0,021	0,00	17	0,22	0	0

ES 2 467 918 T3

Pro	CCT	116	0,202	0,27	29	0,37	19	0,24
Pro	CCC	250	0,436	0,41	13	0,17	32	0,41
Ile	ATA	0	0,000	0,00	2	0,03	0	0
Ile	ATT	136	0,298	0,28	40	0,57	16	0,23
Ile	ATC	320	0,702	0,72	28	0,40	54	0,77
Glu	GAG	683	0,912	0,90	47	0,56	77	0,92
Glu	GAA	66	0,088	0,10	37	0,44	7	0,08
Asp	GAT	143	0,237	0,26	33	0,37	22	0,24
Asp	GAC	460	0,763	0,74	57	0,63	68	0,76
Lys	AAG	551	0,960	0,90	40	0,48	73	0,88
Lys	AAA	23	0,040	0,10	43	0,52	10	0,12
Asn	AAT	22	0,062	0,11	12	0,21	6	0,10
Asn	AAC	331	0,938	0,89	46	0,79	52	0,90
Cys	TGT	7	0,050	0,06	12	0,36	4	0,12
Cys	TGC	134	0,950	0,94	21	0,64	29	0,88
Tyr	TAT	13	0,057	0,39	15	0,34	14	0,32
Tyr	TAC	214	0,943	0,61	29	0,66	30	0,68
Phe	TTT	160	0,451	0,47	44	0,62	28	0,39
Phe	TTC	195	0,549	0,43	27	0,38	43	0,61
Gln	CAG	306	0,924	0,90	26	0,47	50	0,91
Gln	CAA	25	0,076	0,10	29	0,53	5	0,09
His	CAT	29	0,173	0,15	10	0,32	7	0,23
His	CAC	139	0,827	0,85	21	0,68	24	0,77
Met	ATG	291	1,00	1	46	1	46	1
Trp	TGG	104	1,00	1	19	1	19	1

5 Tal como se describió anteriormente, un trabajo anterior de los presentes inventores y colaboradores (véase el ejemplo 8 en la publicación de solicitud de patente estadounidense n.º 20050100995) dio como resultado la creación de un plásmido en el que se clonó la región codificante (no sintética) del orfC de *Th.23B* entre las regiones no codificantes en el sentido de 5' y en el sentido de 3' del orfC de *Schizochytrium* de manera que se generó una "costura perfecta" con la región codificante de *Th.23B*. Pueden usarse plásmidos intermedios en este procedimiento para clonar la región codificante de orfC de *Th.23B* sintética (véanse las figuras 2A y 2B). Con el fin de utilizar más

fácilmente uno de estos constructos intermedios, los inventores diseñaron una secuencia de nucleótidos de 283 pb y se sintetizó mediante DNA2.0 para crear las uniones de “costura perfecta” y utilizar sitios de restricción dentro de las regiones en el sentido de 5’/3’ del orfC de *Schizochytrium* y se diseñó en el gen orfC de *Th.23B* sintético para posteriores reacciones de clonación. Esta secuencia de ADN corta se denominó “Th23B synth orfC INT” y estaba contenida dentro del plásmido “pThOrfC stitch INT”.

El “Th23B synth orfC INT” de 283 pb consiste en cinco segmentos. El primer segmento consiste en los 102 pb finales de la región (no codificante) en el sentido de 5’ del orfC de *Schizochytrium* desde un sitio de *SpeI* hasta, pero sin incluir, el codón de iniciación ATG del orfC de *Schizochytrium* (véase SEQ ID NO: 77). El segundo segmento consiste en los 9 pb iniciales de la región codificante de orfC de *Th.23B* sintética (SEQ ID NO: 61) y contiene el ATG de iniciación que se solapa con un sitio de *SalDI* designado (GGGTCCC). Estos segmentos crean la unión de “costura perfecta” en el sentido de 5’. El tercer segmento es un sitio de restricción de *BamHI* de 6 pb (GGATCC) que funciona como espaciador. El cuarto segmento consiste en los 45 pb finales de la región codificante de orfC de *Th.23B* (SEQ ID NO: 61) desde un sitio de *Clal* designado hasta el codón de terminación TAA. El quinto segmento consiste en los 121 pb iniciales de la región en el sentido de 3’ (no codificante) de orfC de *Schizochytrium* (sin incluir el codón de terminación) hasta un sitio de *BsmI* “inverso”. Los seis nucleótidos finales del fragmento “Th23B synth orfC INT” en la orientación “directa” son 5’>GCATTC>3’. El complemento inverso, 5’>GAATGC>3’, es la secuencia de reconocimiento para *BsmI*. Los segmentos cuarto y quinto crean la unión de “costura perfecta” en el sentido de 3’.

A continuación se proporcionan detalles de construcción de la versión de “costura perfecta” de la secuencia codificante de orfC de *Th.23B* sintética (véanse también las figuras 2A y 2B).

Etapas 1 (figura 2A). Se eliminó el fragmento “Th23B synth orfC INT” de pThOrfC stitch INT mediante digestión con enzimas de restricción *SpeI* y *BsmI* y se purificó el fragmento mediante electroforesis en gel de agarosa (kit GeneClean Turbo, QBioGene). De manera similar, se obtuvo el fragmento de vector *SpeI/BsmI* grande a partir de pREZ22 (véase la publicación de solicitud de patente estadounidense n.º 20050100995), que contiene aproximadamente 2000 pb de cada una de las regiones en el sentido de 5’ y en el sentido de 3’ del orfC de *Schizochytrium* separadas por un espaciador de sitio de reconocimiento de *BamHI* clonado en pBlueScriptII SK(+)). Se ligaron estos dos fragmentos y se transformaron en *E. coli* XL-1 Blue (Stratagene, La Jolla, CA). Se identificaron los clones que contenían el plásmido deseado, “pREZ22 orfC INT”, mediante digestiones por restricción y secuenciación de ADN parcial. Este plásmido contiene las regiones en el sentido de 5’ y en el sentido de 3’ del orfC de *Schizochytrium* perfectamente cosidas a las regiones 5’ y 3’, respectivamente, de la región codificante de orfC sintética, pero carece de la mayor parte de la región codificante.

Etapas 2 (figura 2B). Se obtuvo la mayor parte de la región codificante de orfC de *Th.23B* sintética a partir de “pThOrfC synth” mediante digestión con enzimas de restricción *SalDI* y *Clal* y purificación del fragmento de ADN deseado (como anteriormente). Se ligó este fragmento en un fragmento de vector obtenido de manera similar a partir de pREZ22 orfC INT y se clonó en *E. coli* (como anteriormente). El plásmido resultante, “pThOrfC-synPS”, contiene la región codificante de orfC de *Th.23B* sintética de longitud completa perfectamente cosida a las regiones en el sentido de 5’ y en el sentido de 3’ del gen orfC de *Schizochytrium*. La secuencia de nucleótidos de la región codificante de pThOrfC-synPS se representa en el presente documento por SEQ ID NO: 70. SEQ ID NO: 70 codifica para SEQ ID NO: 62. Se depositó pThOrfC-synPS con el n.º de registro de la ATCC PTA-8229, tal como se describió anteriormente en el presente documento.

Ejemplo 2

El siguiente ejemplo describe la creación de un constructo que codifica para el OrfC de *Schizochytrium* que comprende un dominio DH2 de *Thraustochytrium* 23B.

Se sustituyó la región DH2 de orfC de *Schizochytrium* ATCC20888 (SEQ ID NO: 30) por la de *Thraustochytrium* 23B ATCC 20892 (SEQ ID NO: 66) en puntos de cruzamiento en 5’ y 3’ específicos mediante una combinación de extensión por solapamiento basada en PCR (“corte y empalme mediante extensión por solapamiento” o “SOEing” (Horton, R.M., (1993) *In vitro* Recombination and Mutagenesis of DNA. SOEing together tailor-made genes. Methods in molecular Biology, vol. 15: PCR Protocols: Current Methods and Applications, capítulo 25, págs. 251-266 (B.A. White, Ed.) Humana Press, Totawa, NJ)) y clonación por restricción.

Más específicamente, en este ejemplo, los inventores construyeron una molécula de ácido nucleico que codificaba para un polipéptido OrfC híbrido (quimérico) (secuencia de aminoácidos representada en el presente documento por SEQ ID NO: 74), de 1493 residuos de aminoácido de longitud, en el que la región DH2, definida como los aminoácidos 516-1041 de este híbrido, consiste en la secuencia de aminoácidos de la región DH2 de la proteína orfC de *Th.23B*; es decir, los aminoácidos 491-1016 de SEQ ID NO: 62, lo que incluye la totalidad de SEQ ID NO: 66 (descrita como el dominio DH2 de *Thraustochytrium* 23B en el presente documento). El resto de la secuencia de aminoácidos de OrfC híbrida, los residuos 1-515 y 1042-1493 de SEQ ID NO: 74, son idénticos a los residuos 1-515 y 1051-1502 de OrfC de *Schizochytrium* de SEQ ID NO: 6, respectivamente.

La construcción del plásmido que codifica para esta proteína quimérica se ilustra en las figuras 3A-3C.

Etapla 1. Se usaron los cebadores prREZ197 (SEQ ID NO: 78) y prREZ198 (SEQ ID NO: 79) para amplificar aproximadamente 1,5 kb del marco de lectura orfC de *Schizochytrium* en el sentido de 5' de la región DH2 usando el gen orfC de *Schizochytrium* sin modificar como molde:

prREZ197 CATATGGCGCTCCGTGTCAA

prREZ198 GCCAGGAAGCTTTGACATGGGGTGCCAGGACATCT.

El cebador prREZ197 creó un sitio de *NdeI* (subrayado) en el codón ATG de iniciación. El cebador inverso prREZ198 (35 meros) contenía el punto de cruzamiento en 5' generado mediante 20 pb de homología con la secuencia de OrfC de *Schizochytrium* (en negrita) y 15 pb de homología con la secuencia de OrfC de *Th.23B*. Condiciones de PCR: reacción de 50 μ l, 1 μ l de polimerasa *PfuUltra* (Stratagene) y tampón *PfuUltra* 1X, DMSO al 2%, 0,5 μ M de cada dNTP, 0,4 μ M de cada uno de prRZ197 y prRZ198, 10 ng de molde (región codificante de orfC de *Schizochytrium* clonada), 1 min de desnaturalización inicial a 94°C, 20 ciclos de 1 min de desnaturalización a 94°C, 1 min de hibridación a 52°C, 90 s de extensión a 72°C y 10 min de extensión final. Se purificó el producto de PCR siguiendo electroforesis en gel de agarosa usando el kit de extracción en gel QIAquick® (Qiagen, Valencia, CA).

Etapla 2. Se usaron los cebadores prREZ 199 (SEQ ID NO: 80) y prREZ200 (SEQ ID NO: 81) para amplificar la región DH2 de *Th.23B* (aproximadamente 1,5 KR) usando el gen orfC de *Th.23B* como molde.

prREZ199 TCCTGGCACCCCATGTCAAAGCTTCCTGGCAACCCTA

prREZ200 AGTATACAGAGGTGCTGACA

El cebador prREZ199 (37 meros) contenía el punto de cruzamiento en 5' generado mediante 22 pb de homología con la secuencia de orfC de *Th.23B* (DH2) y 15 pb de homología con la secuencia de orfC de *Schizochytrium* (en negrita). Estos últimos 15 pb también proporcionaron solapamiento con prREZ198 y por tanto el producto de PCR de la etapa 1. El cebador inverso prREZ200 incorporó un sitio de *BstZ171* natural en el orfC de *Th.23B* en el punto de cruzamiento en 3' (subrayado). Las condiciones de PCR y purificación de fragmentos fueron como anteriormente excepto porque se usaron los cebadores prREZ199 y prREZ200 con 10 ng en la región codificante de orfC de *Th.23B* clonada como molde.

Etapla 3. Se usó la extensión por solapamiento para crear la fusión de longitud completa entre el extremo 5' de la región codificante de orfC de *Schizochytrium* y la región DH2 de *Th.23B*. Se realizó PCR con el producto de la etapa 1 (prREZ197 x prREZ198) y la etapa 2 (prREZ199 x prREZ200) como moldes y los cebadores externos prREZ197 y prREZ200. Condiciones de PCR: reacción de 50 μ l, 1 μ l de polimerasa *PfuUltra* (Stratagene) y tampón *PfuUltra* 1X, DMSO al 2%, 0,5 μ M de cada dNTP, 0,4 μ M de cada uno de prRZ197 y prRZ200, 50 ng de cada producto de PCR de la etapas 1 y 2, 1 min de desnaturalización inicial a 94°C, 20 ciclos de 1 min de desnaturalización a 94°C, 1 min de hibridación a 52°C, 3,5 min de extensión a 72°C y 10 min de extensión final. Se purificó el producto de PCR como en la etapa 1.

Etapla 4. Se clonó el producto de la reacción de PCR en la etapa 3 en pCR-BluntII-TOPO (Invitrogen) y se transformó en TOP10 *E. coli* (Invitrogen) usando las condiciones recomendadas por el fabricante para crear pREZ171. Se confirmó que la secuencia del ADN insertado era tal como se había diseñado.

Etapla 5. Usando sitios de restricción en las secuencias de vector respectivas, se transfirió el ADN clonado en pREZ171 al vector pBC KS(+) (Stratagene) como un fragmento de *XbaI/SpeI* para crear pREZ175.

Etapla 6. Se digirió el plásmido pREZ175 (linealizado) con *BstZ171*, entonces se digirió parcialmente con *NdeI*. Se clonó un fragmento de aproximadamente 6 Kb que representa la región en 5' de orfC de *Schizochytrium* y la región DH2 de *Th.23B* fusionados en el fragmento de vector pREZ172 *NdeI/BstZ171* creando pREZ177. El plásmido pREZ172 contiene toda la región codificante de orfC de *Schizochytrium* clonada en el vector de expresión de *E. coli* pColADuet-1 (Novagen) de manera que el codón ATG de iniciación incorpora un sitio de *NdeI*. Se deriva de pREZ101 (véase el ejemplo 5) y se había modificado mediante mutagénesis dirigida al sitio (kit Quik Change, Stratagene) para insertar un sitio de *BstZ171* neutro para aminoácidos en el sitio de cruzamiento en 3'. Específicamente, se modificó a TAT el codón de tirosina TAC en la posición de aminoácido 1051.

Etapla 7. Tras el análisis de pREZ177 mediante secuenciación de ADN, se descubrió que se había delecionado un único par de bases en el sitio de *BstZ171*. Específicamente, el <GTATAC> esperado era, en vez de eso, <GTAAC>. Para corregir este error, se usó un fragmento de restricción de *PciI* que contenía el punto de cruzamiento de *BstZ171* correcto de pDS26 para sustituir al fragmento de *PciI* defectuoso en pREZ177. El plásmido pDS26 contiene una región codificante de orfC híbrida que se había creado anteriormente para otros fines. El plásmido resultante, pREZ179, contiene por tanto una región codificante de orfC entera que es predominantemente de *Schizochytrium* pero contiene una sustitución precisa de la región DH2 por la de *Th.23B* (la secuencia de aminoácidos representada en el presente documento por SEQ ID NO: 74). El plásmido pREZ179 representa además una herramienta única

para estudiar la función del gene híbrido en *E. coli* y proporciona un punto de partida para el desarrollo de vectores de expresión para otros organismos.

Las siguientes etapas adicionales (véase la figura 3C) describen la transferencia del gene híbrido de pREZ179 a un vector para la sustitución génica en *Schizochytrium*.

Etapla 8. Se aisló la región codificante de orfC de *Schizochytrium* (sin modificar) más porciones cortas de secuencias flanqueantes en el sentido de 5' y en el sentido de 3' a partir de pBR002 (un clon de la región genómica de OrfC) como fragmento de *NheI/BspEI*. Entonces se clonó este fragmento en la parte de vector de pREZ31 digerido con *NheI/BspEI* (funcionalmente equivalente a pREZ33 descrito en la publicación de solicitud de patente estadounidense n.º 20050100995, ejemplo 8). El plásmido resultante, pDS48, contiene la región codificante de orfC de *Schizochytrium* (sin modificar) más las mismas secuencias en el sentido de 5' y en el sentido de 3' que se han usado para dirigir la sustitución génica en el locus de OrfC.

Etapla 9. Una porción del marco de lectura orfC híbrido que contenía la totalidad de la región DH2 de *Th.23B* intercambiada se aisló a partir de pREZ179 como fragmento de *PstI/PfMI*. Se clonó este fragmento en la porción de vector de pDS48 digerido con *PstI/PfMI* para proporcionar pDS49. Como resultado, el plásmido pDS49 contiene el orfC híbrido dentro del mismo contexto que pREZ33 (región codificante de orfC de *Th.23B* de longitud completa como una sustitución génica de "costura perfecta"; véase la publicación de solicitud de patente estadounidense n.º 20050100995, ejemplo 8). La secuencia de nucleótidos de la región codificante de pDS49 se representa en el presente documento por SEQ ID NO: 73. SEQ ID NO: 73 codifica para SEQ ID NO: 74. Se depositó el plásmido pDS49 con el n.º de registro de la ATCC PTA-8230, tal como se describió en detalle anteriormente en el presente documento.

Ejemplo 3

El siguiente ejemplo describe la construcción de un constructo que codifica para el OrfC de *Schizochytrium* que comprende un dominio DH2 de *Thraustochytrium* 23B, en el que el dominio DH2 se ha resintetizado para estar optimizado para el uso de codones de *Schizochytrium*.

En este ejemplo, los inventores construyeron una molécula de ácido nucleico que codificaba para un polipéptido OrfC híbrido (SEQ ID NO: 74), de 1493 residuos de aminoácido de longitud, en la que la región DH2, definida como los aminoácidos 516-1041 de este híbrido, consiste en la secuencia de aminoácidos de la región DH2 de la proteína orfC de *Th.23B*; es decir, los aminoácidos 491-1016 de SEQ ID NO: 62, lo que incluye la totalidad de SEQ ID NO: 66 (descrita como el dominio DH2 de *Thraustochytrium* 23B en el presente documento). El resto de la secuencia de aminoácidos de OrfC híbrida, los residuos 1-515 y 1042-1493 de SEQ ID NO: 74, son idénticos a los residuos 1-515 y 1051-1502 de OrfC de *Schizochytrium* de SEQ ID NO: 6, respectivamente. Además, en este constructo, se derivó la secuencia de ADN que codifica para los aminoácidos 516-1041 de la "secuencia génica sintética" para OrfC de *Th.23B* que está contenida en el plásmido pThOrfC synth y pThOrfC_synPS (véase el ejemplo 1 y SEQ ID NO: 70) y que emplea codones que se prefieren para la expresión génica en *Schizochytrium*. Los detalles de construcción se ilustran en las figuras 4A-4C y se describen a continuación.

Se amplificaron las secuencias de ADN que codifican para la región DH2 del polipéptido OrfC de T23B mediante PCR (reacción 59/60) a partir de pThOrfC synth usando los cebadores oligonucleotídicos dhd59 (5' > G CAC CCC ATG AGC AAG CTC CCC GGC AAC >3; SEQ ID NO: 82) y dhd60 (5' > GT ATA CAG AGG CGC AGA CAC GTT GTA AG >3; SEQ ID NO: 83). El cebador "directo" o de cadena sentido dhd59 se solapa con la secuencia de ADN que codifica para los residuos de aminoácido 491-501 (WHPMSKLPGNP; posiciones 491-501 de SEQ ID NO: 62) de la proteína orfC de *Th.23B*. El cebador "inverso" o de cadena antisentido dhd60 se solapa con la secuencia de ADN que codifica para los residuos de aminoácido 1008-1017 (TYNVSAPLYT; posiciones 1008-1017 de SEQ ID NO: 62) de la proteína orfC de *Th.23B*. El cebador dhd60 contiene dos apareamientos erróneos con la secuencia de pThOrfC synth que se indican mediante los residuos encerrados en un recuadro en la secuencia de dhd60 anterior. Estos cambios crearon un sitio de endonucleasa de restricción *BstZ17 I*, indicado mediante la porción con subrayado doble de la secuencia de dhd60 anterior, con el fin de facilitar etapas de clonación posteriores y también introdujeron dos "mutaciones silenciosas" en la secuencia codificante de la proteína híbrida: de CTT(L) a CTG(L) y de TAC(Y) a TAT(Y). Se llevó a cabo esta amplificación en un volumen de reacción de 40 µl de tampón de reacción HF *PfuUltra*TM 1X (Stratagene, La Jolla, CA) que contenía dhd59 y dhd60 a 0,5 µM cada uno, dNTP 200 µM, 2 unidades de ADN polimerasa de alta fidelidad *PfuUltra*TM (Stratagene, La Jolla, CA) y 1 ng de ADN de pThOrfC synth. Los parámetros de ciclos fueron: 1 vez [1 min a 94°C], 28 veces [(1 min a 94°C), (0,5 min a 60°C), (1,5 min a 72°C)], 1 vez [8,5 min a 72°C] y mantener a 4°C. Se realizó la reacción en un termociclador de sistema de PCR 2400 GeneAmp® de Perkin Elmer (Applied Biosystems, Foster City, CA).

Se amplificó la secuencia de ADN que codifica para los residuos de aminoácido 331-522 de la proteína OrfC híbrida codificada por pREZ179 mediante PCR (reacción 57/58) de pREZ179 usando los cebadores oligonucleotídicos dhd57 (5' > C TGC AGC CAG ATG CTC AAG ATG TAC ATG >3; SEQ ID NO: 84) y dhd58 (5' > G GAG CTT GCT CAT GGG GTG CCA GGA CAT CTC >3; SEQ ID NO: 85). El cebador "directo" o de cadena sentido dhd57 se solapa con la secuencia de ADN que codifica para los residuos de aminoácido 330-339 (GCSQMLKMYM, SEQ ID NO: 101;

posiciones 330-339 de SEQ ID NO: 74) de la proteína OrfC híbrida codificada por pPREZ179. El cebador “inverso” o de cadena antisentido dhd58 se solapa con la secuencia de ADN que codifica para los residuos de aminoácido 513-523 (EMSWHPMSKLP, SEQ ID NO: 102; posiciones 513-523 de SEQ ID NO: 74) de la proteína OrfC híbrida. El extremo 5' del cebador directo, dhd57, se solapa con el sitio de *Pst* I presente en la secuencia codificante de OrfC híbrida contenida en pPREZ179. Se llevó a cabo esta amplificación en un volumen de reacción de 40 μ l de tampón de reacción HF *PfuUltra*TM 1X (Stratagene, La Jolla, CA) que contenía dhd57 y dhd58 a 0,5 μ M cada uno, dNTP 200 μ M, 2 unidades de ADN polimerasa de alta fidelidad *PfuUltra*TM (Stratagene, La Jolla, CA) y 1 ng de ADN de pPREZ179. Los parámetros de ciclos fueron: 1 vez [1 min a 94°C], 28 veces [(1 min a 94°C), (0,5 min a 60°C), (1,5 min a 72°C)], 1 vez [8,5 min a 72°C] y mantener a 4°C. Se realizó la reacción en un termociclador de sistema 2400 GeneAmp de Perkin Elmer.

Se hicieron pasar cuatro microlitros de cada una de las reacciones 57/58 y 59/60 sobre gel de agarosa al 1,2%. Se observaron bandas de ADN en cada caso que concordaban con los tamaños de producto esperados: 578 pb para el producto de 57/58 y 1578 pb para el producto de 59/60. Se escindieron estas bandas del gel y se recuperó el ADN de los cortes de agarosa usando un kit de extracción en gel QIAquick® (QIAGEN, Inc. Valencia, CA) según el protocolo del proveedor. Se recuperaron los productos de PCR en 40 μ l de tampón de elución.

Los 20 nucleótidos en 5' del cebador inverso dhd58 (subrayados anteriormente) comprenden el complemento inverso de los 20 nucleótidos en 5' de dhd59, también subrayados anteriormente. Como resultado, hay un solapamiento idéntico de 20 pb entre el extremo 3' del producto de la reacción 57/58 y el extremo 5' del producto de la reacción 59/60 y este solapamiento permite el posterior corte y empalme mediante PCR de estos dos productos mediante la técnica de “corte y empalme mediante extensión por solapamiento” o “SOEing” mediante PCR [Horton, R.M., (1993) *In vitro* Recombination and Mutagenesis of DNA. SOEing together tailor-made genes. Methods in molecular Biology, vol. 15: PCR Protocols: Current Methods and Applications, capítulo 25, págs. 251-266 (B.A. White, Ed.) Humana Press, Totawa, NJ]. Este fragmento sometido a corte y empalme contiene entonces sitios de restricción útiles en (*Bst*Z17 I & *Pst* I) o cerca (*Bs*W I) de sus extremos.

Se realizó la reacción de corte y empalme mediante PCR (reacción 57/60) de la siguiente manera. Un volumen de reacción de 40 μ l de tampón de reacción HF *PfuUltra*TM 1X contenía los cebadores dhd57 y dhd60 cada uno a 0,5 μ M, dNTP 200 μ M, 2 unidades de ADN polimerasa de alta fidelidad *PfuUltra*TM (Stratagene, La Jolla, CA) y 0,8 μ l de una dilución de 50 veces de cada uno de los productos de PCR purificados en gel 57/58 y 59/60. Se realizó una serie de reacciones de corte y empalme mediante PCR en las que se hizo variar la temperatura de hibridación en incrementos de 1°C entre 66-70°C. Otros parámetros de ciclos fueron constantes: 1 vez [1 min a 98°C], 33 veces [(1 min a 98°C), (1 min a 66-70°C), (2,5 min a 72°C)], 1 vez [7,5 min a 72°C] y mantener a 6°C. Se realizó la reacción en un ciclador de temperatura RoboCycler® (Stratagene, La Jolla, CA). Se hicieron pasar alícuotas de estas reacciones sobre gel de agarosa al 1% y se observó que todas las reacciones contenían un producto que concordaba en cuanto al tamaño con el producto esperado (2136 pb) pero también se observaron otras bandas a todas las temperaturas de hibridación. Por tanto, se combinaron las 3 reacciones con hibridaciones a 67, 68 y 69°C, se hicieron pasar sobre un gel de agarosa al 1% y se escindió la banda de interés de aproximadamente 2,1 kb y se recuperó el fragmento de ADN usando un kit de extracción en gel QIAquick® (QIAGEN, Inc. Valencia, CA) según el protocolo del proveedor. Se recuperó el ADN eluido en 30 μ l de tampón de elución y se clonó en el vector de clonación de fragmento de PCR pCR®-Blunt II TOPO® (Invitrogen Corp., Carlsbad, CA) usando el kit de clonación mediante PCR Zero Blunt® TOPO® (Invitrogen Corp., Carlsbad, CA) según los protocolos del proveedor. Se usaron los productos de la reacción de clonación de TOPO para transformar *E. coli* químicamente competente TOP10 One Shot® (Invitrogen) según el protocolo del proveedor. Se hicieron crecer durante la noche ocho de los transformantes resultantes y se prepararon ADN de plásmido y se analizaron mediante digestión con endonucleasa de restricción y electroforesis en gel de agarosa. Se encontró que siete de los ocho contenían el producto de PCR 57/60 de 2,1 kb clonado. Se secuenció el producto de PCR 57/60 clonado de un aislado y se mostró que correspondía exactamente a la secuencia esperada. Se realizó una secuenciación de ADN por el Biotechnology Resource Center de la Cornell University (Ithaca, Nueva York) pagando una tarifa por sus servicios usando el analizador de ADN automatizado 3730 de Applied Biosystems, con química de terminador Big Dye y ADN polimerasa AmpliTaq-FS (Applied Biosystems, Foster City, CA). El plásmido que contenía el inserto verificado por la secuencia se denominó pDD21 y se usó en etapas de construcción adicionales descritas a continuación.

El segmento de ADN que codificaba para el dominio DH2 de *Th.23B* optimizado para el uso de codones de *Schizochytrium* se escindió de pDD21 y se clonó en pPREZ179 (véase el ejemplo 2) de modo que sustituyó a la secuencia codificante de dominio DH2 de *Th.23B* nativa presente en ese constructo. Se construyó el plásmido resultante, pDD22, de la siguiente manera. Se dirigió ADN de pDD21 purificado con *Bs*WI y *Bst*Z17I (New England BioLabs, Beverly MA) según los protocolos del proveedor. Posteriormente se sometió la reacción a tratamiento usando el procedimiento de purificación por centrifugación QIAquick® y el kit de purificación mediante PCR QIAquick® (QIAGEN Inc., Valencia, CA) según el protocolo del proveedor. Se hicieron pasar los productos de digestión purificados sobre un gel de agarosa al 1% y se escindió el fragmento de *Bs*WI - *Bst*Z17I de 1940 pb y se eluyó de la agarosa usando un kit de extracción en gel QIAEX II (QIAGEN Inc., Valencia, CA) según el protocolo del proveedor. También se dirigió ADN de pPREZ179 purificado, con *Bs*WI y *Bst*Z17I y posteriormente se trató con fosfatasa antártica (New England BioLabs, Beverly, MA) según el protocolo del proveedor. También se sometieron a tratamiento los productos de digestión tratados con fosfatasa usando el procedimiento QIAquick® tal como se

describió anteriormente y se hicieron pasar sobre un gel de agarosa al 0,7%. Se escindió el fragmento de vector de *Bsi*WI - *Bst*Z171 de ~6,1 Kb del gel y se eluyó de la agarosa usando el kit de extracción en gel QIAEX II descrito anteriormente. Se ligaron estos dos fragmentos en tampón de reacción de ligasa de T4 1X usando ligasa de T4, ambos de New England BioLabs (Beverly, MA). Se usaron los productos de ligamiento para transformar *E. coli* químicamente competente TOP10 One Shot® (Invitrogen) según el protocolo del proveedor. Se analizaron ADN de plásmido de tres de los transformantes resultantes mediante digestión con endonucleasa de restricción y electroforesis en gel de agarosa y se encontró que los tres tenían la estructura del recombinante esperado. Un plásmido se denominó pDD22 y se empleó en construcciones adicionales.

Con el fin de facilitar la introducción del ADN que codifica para el OrfC híbrido que contiene la región DH2 de *Th.23B* codificada por codones preferidos por *Schizochytrium* en el genoma de *Schizochytrium*, se escindió de pDD22 un segmento de ADN de *Pst*I-*Pf*MI que abarcaba la secuencia que codifica para la región DH2 y se clonó en pDS48 (véase el ejemplo 2), un vector diseñado para la sustitución génica en secuencias en el locus del gen OrfC en *Schizochytrium*. Se construyó el plásmido resultante, pDD24, que se usó en sustituciones génicas posteriores, de la siguiente manera. Se escindió de pDD22 el segmento de ADN que codifica para el dominio DH2 de T23B y con el uso de codones optimizado y se clonó en pDS48 de modo que sustituyó a la secuencia codificante de dominio DH2 de *Schizochytrium* nativa presente en ese constructo. Se digirió ADN de pDD22 purificado con *Pst*I, *Pf*MI y *Cl*I (New England BioLabs, Beverly MA) según los protocolos del proveedor. La digestión con *Cl*I escindió un fragmento de *Pf*MI-*Pf*MI que de lo contrario migraría cerca de la posición del fragmento de interés de ~3,2 Kb de *Pst*I-*Pf*MI. Posteriormente se sometió la reacción a tratamiento usando el procedimiento de purificación por centrifugación QIAquick® y el kit de purificación mediante PCR QIAquick® (QIAGEN Inc., Valencia, CA) según el protocolo del proveedor. Se hicieron pasar los productos de digestión purificados sobre un gel de agarosa al 0,7% y se escindió el fragmento de interés de *Pst*I - *Pf*MI de ~3,2 Kb y se eluyó de la agarosa usando un kit de extracción en gel QIAEX II (QIAGEN Inc., Valencia, CA) según el protocolo del proveedor. Se digirió de manera similar ADN de pDS48 purificado, con *Pf*MI y *Pst*I, se sometió al tratamiento QIAquick® tal como se describió anteriormente y se hizo pasar sobre un gel de agarosa al 0,7%. Se escindió el fragmento de vector de *Pst*I - *Pf*MI de ~8,0 Kb del gel y se eluyó de la agarosa usando el kit de extracción en gel QIAEX II descrito anteriormente. Se ligaron estos dos fragmentos en tampón de reacción de ligasa de T4 1X usando ligasa de T4, ambos de New England BioLabs (Beverly, MA). Se usaron los productos de ligamiento para transformar *E. coli* químicamente competente TOP10 One Shot® (Invitrogen) según el protocolo del proveedor. Se hicieron crecer durante la noche los transformantes resultantes, en cultivo líquido de medios LB que contenían 100 µg/ml de ampicilina a 30°C. Se encontró que la propagación de estos transformantes a 37°C en cultivos líquidos daba como resultado inestabilidad del plásmido en algunas circunstancias. Se analizaron ADN de plásmido de tres de los transformantes resultantes mediante digestión con endonucleasa de restricción y electroforesis en gel de agarosa y se encontró que los tres tenían la estructura del recombinante esperado. Un plásmido se denominó pDD24 y se sometió a análisis con endonucleasa de restricción adicional y se empleó en experimentos de sustitución génica en *Schizochytrium* (véase el ejemplo 4). La secuencia de nucleótidos de la región codificante de pDD24 se representa en el presente documento por SEQ ID NO: 75. SEQ ID NO: 75 codifica para SEQ ID NO: 74. Se depositó el plásmido pDD24 con el n.º de registro de la ATCC PTA-8226, tal como se describió anteriormente en el presente documento.

Ejemplo 4

El siguiente ejemplo describe la expresión de diversos constructos de orfC de *Th.23B* descritos en los ejemplos 1-3 anteriores en *Schizochytrium*, y el análisis de los AGPI producidos por tales organismos.

Expresión de genes orfC de Th.23B variantes en Schizochytrium

Se transformó la cepa B32-Z1 de *Schizochytrium* (véase anteriormente y el ejemplo 8 en la publicación de solicitud de patente estadounidense n.º 20050100995), que es un *Schizochytrium* con una delección exacta de la región codificante de orfC de *Schizochytrium*, con el plásmido pThOrfC-synPS (orfC de *Th.23B* sintético de longitud completa; véase el ejemplo 1), pDS49 (región DH2 de *Th.23B* no sintética; véase el ejemplo 2) y pDD24 (región DH2 de *Th.23B* sintética; véase el ejemplo 3) mediante bombardeo de partículas usando técnicas anteriormente descritas (véase la publicación de solicitud de patente estadounidense n.º 2003/0166207). Se obtuvieron transformantes sensibles a Zeocin™ prototróficos. Tales transformantes surgieron de acontecimientos de sustitución génica de doble cruzamiento tal como se confirmó mediante transferencia de tipo Southern y/o PCR para cepas seleccionadas.

En resumen, el bombardeo de partículas utilizó el sistema de suministro de partículas Biolistic® PDS-1000/He de BioRad (Hercules, CA). Se hicieron crecer cepas de *Schizochytrium* para la transformación a 29-30°C en medio M2B (más DHA cuando era apropiado) sobre una plataforma giratoria (200 rpm) hasta DO600 = de 1 a 2,5 (BioPhotometer, Eppendorf). Se recogieron células mediante centrifugación (3000 rpm, 5 min) y se resuspendieron en Na₂SO₄ 7,5 g/l estéril hasta DO600 = 30. Se extendió un volumen de 150 µl de células suspendidas en una pequeña zona circular (6 cm de diámetro) sobre una placa Petri que contenía agar M2B (sin DHA). Para el crecimiento de auxótrofos de AGPI, se complementó M2B con DHA hasta 0,25 mM a partir de una disolución madre de DHA 25 mM en β-ciclodextrina metilada al azar al 40% (p/v) (CTD Inc, High Springs, FL.). Cuando se realizaron bombardeos para complementar la auxotrofia de DHA, se omitió DHA del medio de agar. Se llevaron a cabo bombardeos en una campana extractora de flujo laminar usando discos de ruptura de 1100 psi, un hueco de 0,25

pulgadas entre el tapón de retención del disco y la tapa de cubierta del macroportador, y el soporte de pantalla de detención en la posición central. La repisa diana está en la posición L2 (6 cm). Se incubaron a 29-30°C placas Petri que contenían cepas de *Schizochytrium* auxótrofas para DHA sometidas a bombardeo hasta que se desarrollaron colonias (prototróficas prospectivas) (3-5 días). Se cultivaron en línea colonias elegidas al azar en placas de agar M2B. Tras el crecimiento, se transfirieron diversas colonias bien aisladas a placas de M2B con y sin Zeocin® (50 µg/ml). Se seleccionaron protótrofos de DHA sensibles a Zeocin (lo que sugiere un acontecimiento de sustitución génica) para su estudio adicional.

Crecimiento de Schizochytrium para el análisis de ácidos grasos

Se inocularon matraces Erlenmeyer (250 ml) que contenían 50 ml de medio M50-20 con el contenido (1 ml) de un criovial de la cepa indicada. Se incubaron los matraces a 29-30°C en un agitador rotativo a 200 rpm durante 72 horas. Se inocularon matraces similares que contenían medio SSFM con 0,5 ml del cultivo en M50-20 y se incubaron como anteriormente durante 5 días. Se recogieron las células mediante centrifugación (4000 g, 5 min) tras dilución del caldo con un volumen igual de isopropanol al 70%. Se suspendieron los sedimentos celulares resultantes en un volumen original del 35% de isopropanol-agua y volvieron a centrifugarse. Se congelaron inmediatamente a -70°C los sedimentos celulares lavados seguido por liofilización. Se determinó el contenido en ácidos grasos de la biomasa secada preparando ésteres metílicos de ácidos grasos (EMAG) usando metanol ácido, extrayéndolos en hexano y analizándolos mediante cromatografía gas-líquido.

Medio M50-20

Los componentes por litro del medio M50-20 son los siguientes: 12,5 g de NaCl, 2,5 g de MgSO₄·7H₂O, 0,5 g de KCl, 0,05 g de CaCl₂, 20,0 g de glucosa, 20,0 g de Na-glutamato, 0,4 g de KH₂PO₄, 1,0 g de extracto de levadura, 0,4 g de NaHCO₃, 5 ml de metales traza PII (la disolución de metales traza PII 200X contiene por litro: 6,0 g de Na₂EDTA, 0,29 g de FeCl₃·6H₂O, 6,84 g de H₃BO₃, 0,86 g de MnCl₂·4H₂O, 60 mg de ZnCl₂, 26 mg de CoCl₂·6H₂O, 52 mg de NiSO₄·6H₂O, 2 mg de CuSO₄·5H₂O y 5 mg de NaMoO₄·2H₂O, pH 8,0), 1 ml de mezcla de vitaminas PII (la mezcla de vitaminas PII 1000X contiene por litro: 100 mg de tiamina, 0,5 mg de biotina y 0,5 mg de vitamina B₁₂), pH 7,0.

Medio SSFM

Los componentes por litro del medio SSFM son los siguientes: 13,62 g de Na₂SO₄, 0,72 g de K₂SO₄, 0,56 g de KCl, 2,27 g de MgSO₄·7H₂O, 0,19 g de CaCl₂, 0,0565 g de KH₂PO₄, 0,57 g de (NH₄)₂SO₄, 0,13 g de Na-glutamato, MES 100 mM (ácido 4-morfolin-etanosulfónico) pH 6,0, 50,0 g de glucosa, 0,16 mg de vitamina B₁₂, 9,75 mg de tiamina, 3,33 mg de pantotenato de calcio, 10,3 mg de FeSO₄·7H₂O, 3,1 mg de MnCl₂·4H₂O, 1,93 mg de ZnSO₄·7H₂O, 0,04 mg de CoCl₂·6H₂O, 0,04 mg de NaMoO₄·2H₂O, 2,07 mg de CuSO₄·5H₂O, 2,07 mg de NiSO₄·6H₂O, 2,0 mg de ácido cítrico.

Medio M2B

Los componentes del medio M2B son los siguientes (por litro): 10 g de glucosa, 0,8 g de (NH₄)₂SO₄, 5,0 g de Na₂SO₄, 2,0 g de MgSO₄·7H₂O, 0,5 g de KH₂PO₄, 0,5 g de KCl, 0,1 g de CaCl₂·2H₂O, 0,05 mg de vitamina B₁₂, 0,2 mg de tiamina·HCl, 0,2 mg de pantotenato de calcio, 3,0 mg de FeSO₄·7H₂O, 1,0 mg de MnCl₂·4H₂O, 0,8 mg de ZnSO₄·7H₂O, 0,02 mg de CoCl₂·6H₂O, 0,01 mg de Na₂MoO₄·2H₂O, 0,6 mg de CuSO₄·5H₂O, 0,8 mg de NiSO₄·6H₂O, tampón MES 0,1 M, pH 6,0 (ajustado con NaOH).

Análisis de AGPI de cepas de Schizochytrium recombinantes

La tabla 2 muestra el contenido en ácidos grasos totales, DHA y DPAn-6 (expresado como EMAG (éster metílico de ácidos grasos)) de *Schizochytrium* ATCC 20888 y cepas derivadas en las que la región codificante de orfC nativa está sustituida por la totalidad o parte de la región codificante de OrfC de *Thraustochytrium* 23B (descrita en los ejemplos 1-3). La sustitución de la totalidad de la región codificante de orfC de *Schizochytrium* ATCC 20888 por la de *Th.23B* (cepa B34-1) da como resultado una razón DHA/DPAn-6 superior (más próxima a la de *Th.23B*) pero un contenido en AGPI totales inferior. El hecho de que la expresión de proteínas es la causa probable del contenido en AGPI totales inferior se demuestra mediante el uso de la región codificante de orfC de *Th.23B* optimizada para codones (sintética) (por ejemplo, en la cepa B67-5; transformada con pThOrfC_syn-PS) en la que la producción de AGPI aumenta con respecto a los niveles de tipo natural mientras que se mantiene la razón DHA/DPAn-6 potenciada. Las sustituciones sólo de la región DH2 de *Schizochytrium* por la de *Thraustochytrium* muestran un patrón similar. La cepa con la región DH2 de *Th.23B* optimizada para codones (B69-2; transformada con pDD24) proporciona un contenido en AGPI superior al de la cepa con la región DH2 no optimizada (B105-1A1; transformada con pDS49). Sin embargo, la razón DHA/DPA en la cepa B105-1A1 (región DH2 no optimizada) era notablemente alta.

De manera interesante, la cepa B69-6 produce altos niveles de DHA y una razón DHA/DPA relativamente alta. Esta cepa se obtuvo como resultado de la misma transformación que la cepa B32-Z1 con el plásmido pDD24 que produjo

la cepa B69-2. Sin embargo, la cepa B69-6 no tiene una integración/sustitución génica correcta de la región codificante de orfC modificada (tal como se determinó mediante análisis por PCR), aunque se desconoce la naturaleza exacta de la discrepancia.

- 5 Dados estos datos, pueden desarrollarse fermentaciones a escala de producción con la cepa B69-2 para lograr una producción de DHA máxima, o las cepas B69-6 o B105-1A1 si se desea la mayor razón DHA/DPA.

Tabla 2. Resumen de variantes de orfC

Cepa	EMAG (% de dcw)	DHA (% de dcw)	DPA _n -6 (% de dcw)	DHA (% de EMAG)	DHA/DPA	descripción de la cepa
ATCC20888	71,4	16,5	3,64	22,9	4,5	<i>Schizochytrium</i> de tipo natural
B34-1	78,4	13,4	1,24	17,0	10,8	orfC de <i>Th.</i> 23B (no sint.)
B67-5	73,0	21,3	1,85	28,9	11,5	orfC de <i>Th.</i> 23B sint.
B105-1A1	73,5	19,4	1,31	26,4	14,8	DH2 de <i>Th.</i> 23B (no sint.)
B69-2	73,0	23,0	2,31	31,6	10,0	DH2 de <i>Th.</i> 23B sint.
B69-6	73,8	22,4	1,76	30,3	12,7	DH2 de <i>Th.</i> 23B sint.
Dcw peso celular seco						
EMAG éster metílico de ácidos grasos						

10

Ejemplos 5

El siguiente ejemplo describe la producción de DHA y DPA en *E. coli* mediante un sistema de múltiples plásmidos y además ilustra que el dominio DH2 del sistema de PKS de AGPI controla la razón de producción de ácidos grasos por el sistema.

15

Los inventores demostraron anteriormente la producción de DHA y DPA en *E. coli* mediante el uso de un sistema inducible por T7 para expresar OrfA, OrfB*, OrfC de *Schizochytrium* y HetI de *Nostoc* (ejemplo 3, página 41, publicación de solicitud de patente estadounidense n.º 20050100995). En este ejemplo anterior, OrfA, OrfB* y OrfC estaban contenidos en un único plásmido. Con el fin de crear un sistema más susceptible de manipulación genética, se clonaron las regiones codificantes individuales de *Schizochytrium* en un conjunto de plásmidos de expresión compatibles diseñados para la coexpresión de múltiples genes diana. La expresión de los genes diana se dirige de manera similar por el promotor de T7 inducible en esta serie Duet de plásmidos (Novagen). Se clonó orfA de *Schizochytrium* como un fragmento de *NdeI* - *XbaI* de pBR115L1 en el vector de expresión pETDuet-1 para crear pREZ91 (se hace referencia a pBR115L1 en la generación del plásmido de expresión final en el ejemplo 3, página 41, publicación de solicitud de patente estadounidense n.º 20050100995). Se clonó orfB* de *Schizochytrium* como un fragmento de *NdeI* - *XbaI* de pJK780 en el vector de expresión pCDFDuet-1 para crear pREZ96 (se hace referencia a pJK780 en la generación del plásmido de expresión final en el ejemplo 3, página 41, publicación de solicitud de patente estadounidense n.º 20050100995). Se clonó orfC de *Schizochytrium* como un fragmento de *NdeI* - *XbaI* de pJKS 10 en pColADuet-1 para crear pREZ101 (se hace referencia a pJK510 en la generación del plásmido de expresión final en el ejemplo 3, página 41, publicación de solicitud de patente estadounidense n.º 20050100995). El gen auxiliar requerido *hetI*, que codifica para una fosfopanteteína transferasa (PPTasa), se suministró en un plásmido basado en pACYC184, pJK737, anteriormente descrito (ejemplo 3, página 41, publicación de solicitud de patente estadounidense n.º 20050100995). Se transformaron OrfA, OrfB*, OrfC y *hetI*, contenidos por separado en los plásmidos pREZ91, pREZ96, pREZ101 y pJK737 respectivamente, en la cepa BLR (DE3) de *E. coli* (Novagen) que contiene una ARN polimerasa de T7 inducible.

30

35

Se detectó la producción de DHA y DPA en células de *E. coli* que se hicieron crecer en caldo Luria (LB) tanto a 25°C como a 30°C (véase la tabla 3 a continuación) usando estas cepas de múltiples plásmidos. Se inocularon colonias individuales en caldo LB complementado con antibióticos para mantener cada plásmido en la cepa dada y se hicieron crecer durante la noche a la temperatura deseada (25°C o 30°C). Entonces se usaron volúmenes de 300 μ l de estos cultivos para inocular cultivos principales de 30 ml de LB con antibióticos apropiados. Se hicieron crecer los cultivos principales a la temperatura indicada hasta que la DO600 (BioPhotometer, Eppendorf) estaba entre 0,45 y 0,55, punto en el cual se indujeron los cultivos con IPTG hasta una concentración final de 1 mM. Entonces se mantuvieron los cultivos en estas condiciones de expresión durante 24 horas tras lo cual se recogieron las células mediante centrifugación y se prepararon para el análisis de EMAG. El nivel típico de AGPI producidos (como porcentajes de EMAG totales) a 30°C fue del 10% de DHA y el 6% de DPA (el 16% de AGPI totales) para la cepa que portaba orfC de *Schizochytrium*. La razón DHA/DPA de 1,7 se aproxima a lo que se observa en *Schizochytrium* (véase la tabla 2 a continuación).

La expresión de los genes de *Schizochytrium* requeridos para la producción de DHA y DPA en *E. coli* en plásmidos separados proporcionó a los inventores la capacidad de estudiar más fácilmente y manipular genes de biosíntesis de AGPI. Tal como se describe en la publicación de solicitud de patente estadounidense n.º 2005/0100995, ejemplo 8, se demostró que en *Schizochytrium*, la sustitución del OrfC por el gen homólogo de *Thraustochytrium* 23B alteró el perfil de AGPI con un desplazamiento en la razón de DHA con respecto a DPA. Se llevó a cabo el experimento similar con el sistema de expresión de múltiples plásmidos de *E. coli* descrito anteriormente, en el que se sustituyó el plásmido de expresión de orfC de *Schizochytrium* (pREZ101) por un plásmido de expresión de orfC de *Thraustochytrium* 23B similar (pREZ142).

Para crear pREZ142, se clonó la región codificante de orfC de *Th.23B* de pREZ31 como un fragmento de *NcoI/SalI* en el vector Duet pColADuet-1. El plásmido pREZ31 es una variante de pREZ33, el vector de sustitución génica con "costura perfecta" (descrito en el ejemplo 1 anterior y en el ejemplo 8 de la publicación de solicitud de patente estadounidense n.º 2005/0100995), en el que se modificó por ingeniería genética un sitio de restricción de *BamHI* (subrayado a continuación) justo en el sentido de 5' del ATG de iniciación (en minúsculas a continuación). Esta modificación por ingeniería genética creó de manera casual en pREZ31 un sitio de restricción de *NcoI* (en cursiva a continuación) que contenía el ATG de iniciación que estaba compuesto por las dos últimas bases del sitio de *BamHI* y las cuatro primeras bases de la región codificante de orfC de *Th.23B*:

GGATCCatgG (SEQ ID NO: 86)

El sitio de restricción de *SalI* usado en esta clonación es nativo para la región en el sentido de 3' de orfC de *Schizochytrium* y está aproximadamente 250 pb en el sentido de 3' del codón de terminación TAA. Esta sustitución en el sistema de expresión de *E. coli* o del orfC de *Schizochytrium* por el orfC de *Th.23B* dio como resultado un perfil de AGPI alterado con un desplazamiento de la razón de DHA con respecto a DPA de 1,5 a 6,8 y la cantidad total de DHA + DPA se redujo del 10% al 4% cuando se hicieron crecer las cepas y se indujeron a 25°C (véase la tabla a continuación).

Se generaron regiones codificantes de orfC híbridas con el fin de determinar la región o el dominio del gen responsable de controlar la razón de DHA con respecto a DPA. El orfC híbrido en el plásmido de expresión pREZ179 contiene una región DH2 central derivada del orfC de *Thraustochytrium* 23B y está flanqueada en el sentido de 5' y en el sentido de 3' por secuencias de orfC de *Schizochytrium* (véase el ejemplo 2). Cuando se expresó pREZ179 en el sistema anterior en lugar de pREZ101, se observó una razón de DHA con respecto a DPA de 6,5, mientras que la cantidad de AGPI totales fue del 9% cuando se expresó y se indujo a 25°C (véase la tabla a continuación). Este desplazamiento en la razón de DHA con respecto a DPA en la expresión en el modelo de *E. coli* y el mantenimiento del rendimiento indicaron que la región DH2 central de orfG controla la mayor parte o la totalidad de la razón de DHA con respecto a DPA en la biosíntesis de AGPI. Cuando se modificó entonces este constructo con ADN flanqueante adicional y se transformó en *Schizochytrium* para sustituir al orfC nativo, se observó un desplazamiento similar en la razón de DHA con respecto a DPA así como ausencia de disminución en la producción (véase el ejemplo 4). De manera similar, cuando se expresó el orfC híbrido en un sistema de levadura, se observó de nuevo un desplazamiento en la razón de DHA con respecto a DPA (véase el ejemplo 6).

Tabla 3

forma de orfC (temperatura)	plásmido de orfC	DHA + DPA	DHA/DPA
<i>Schizochytrium</i> (30°)	pREZ101	16%	1,7
<i>Schizochytrium</i> (25°)	pREZ101	10%	1,5
<i>Th.23B</i> (25°)	pREZ142	4%	6,8
<i>Th.23B</i> DH2 (25°)	pREZ179	9%	6,5

Uso del sistema de múltiples plásmidos de expresión

Los ejemplos anteriores, en los que se usaron los sistemas de modelo de expresión de múltiples plásmidos en *E. coli* y levadura para esclarecer el papel del OrfC y, en particular, la región DH2, en el control de la razón de DHA con respecto a DPA en la biosíntesis de AGPI, demuestran la utilidad de estos sistemas heterólogos. Los resultados observados en *E. coli* y en levadura son análogos a los observados en *Schizochytrium* en cuanto al efecto relativo de la fuente de orfC sobre la razón DHA/DPA. De manera similar, se describen en el presente documento los sistemas de modelo de expresión de múltiples plásmidos en *E. coli* y levadura para investigar y modificar por ingeniería genética otros aspectos de la biosíntesis de AGPI incluyendo la longitud de cadena de AGPI, el grado de saturación de ácido graso y la posición de dobles enlaces. Estos sistemas también permitirán una fácil expresión de genes implicados en otros tipos de modificaciones de ácidos grasos tales como hidroxilación y glicosilación. De manera similar, pueden clonarse otros genes de biosíntesis de AGPI de un único organismo (tal como se ha hecho para la agrupación de *Shewanella japonica* descrita en el ejemplo 2, publicación de solicitud de patente estadounidense n.º 2005/0100995) o de más de un organismo en este sistema de *E. coli* para facilitar su estudio.

Ejemplo 6

El siguiente ejemplo describe el método mediante el cual se expresaron las subunidades A, B y C de AGPI sintasa de *Schizochytrium* y *hetI* de *Nostoc* en levadura y además ilustra que el dominio DH2 del sistema de PKS de AGPI controla la razón de la producción de ácidos grasos por el sistema.

Parte A

Experimentos de expresión preliminares indicaron que Het I y OrfC de *Schizochytrium* podían producirse como proteínas de longitud completa en levadura usando las regiones codificantes nativas. En cambio, la expresión de las regiones codificantes nativas para OrfA y B de *Schizochytrium* no dio como resultado la producción de cantidades detectables de las proteínas esperadas. El problema parecía estar asociado con la traducción del ARNm. (Transferencias de tipo Northern mostraron la presencia de ARNm del tamaño correcto). Por consiguiente, se prepararon versiones sintéticas de esas dos regiones codificantes con el objetivo de mejorar su expresión en levadura. Las secuencias de aminoácidos de las proteínas codificadas por los genes sintéticos son idénticas a las codificadas por los genes nativos (es decir, SEQ ID NO: 2 y SEQ ID NO: 4). Se realizaron un diseño génico inicial y una síntesis génica completa de orfA y orfB por Blue Heron Biotechnology, Inc. (Bothell, WA). La optimización de codones tuvo en cuenta preferencias codónicas de *S. cerevisiae*. Las secuencias completas de las regiones codificantes sintéticas (denominadas; sOrfA y sOrfB) se indican como SEQ ID NO: 35 (sOrfA) y SEQ ID NO: 36 (sOrfB). A cada región codificante sintética se le agregó ADN de la siguiente manera para facilitar la clonación en los vectores de transformación de levadura:

secuencia en el sentido de 5' (SEQ ID NO: 87) AAGCTTGTGCAGTCAAGTGCGCAAAACCATG

secuencia en el sentido de 3' (SEQ ID NO: 88) TAACCCGGGTCTAGA.

Las posiciones de los codones de iniciación y de terminación están subrayadas y los sitios de reconocimiento de enzimas de restricción para *HindIII* (en el sentido de 5') y *XbaI* (en el sentido de 3') se muestran en negrita.

Se usó la cepa InvSC1 de *S. cerevisiae* (MATa *his3-Δ1*, *leu2*, *trp1-289*, *ura3-52*) (Invitrogen, Carlsbad, CA) para estos experimentos. Se mantuvo la cepa y se transformó según las recomendaciones del proveedor. Se hicieron crecer transformantes sobre medio sólido de glucosa, caldo de rafinosa y medio de inducción de galactosa según las instrucciones del fabricante (Invitrogen). Todos los componentes de medios de levadura se adquirieron de Q-BIOgene (Carlsbad, CA).

Se clonaron *hetI* y los genes de AGPI sintasa de *Schizochytrium* en los siguientes vectores de transformación: pYES-Leu* (sOrfA; SEQ ID NO: 35), pYES3-Tryp (sOrfB; SEQ ID NO: 36), pYES2/CT (OrfC; SEQ ID NO: 5) y pYESHis* (*hetI*; SEQ ID NO: 33). A continuación se describe en detalle la creación de estos vectores. Se modificaron algunos de los vectores y genes para adaptarse a requisitos de clonación y expresión específicos (descritos en detalle a continuación). Se usaron medios de selección apropiados, dependiendo del experimento particular. Se clonaron los genes en cada caso detrás del promotor GAL1 y se indujo la expresión mediante resuspensión de células lavadas en medios que contenían galactosa según directrices proporcionadas por Invitrogen. Se hicieron crecer las células a 30°C y se recogieron (mediante centrifugación) en los momentos indicados tras haberse transferido al medio de inducción. Se secaron por congelación los sedimentos celulares y se prepararon EMAG usando metanol ácido, se extrajeron en hexano y se analizaron mediante CG.

Constructo de expresión de sOrfA: Se clonó el sOrfA en un vector personalizado, pYES-Leu/CT, construido de la siguiente manera. Se modificó un vector pYES6/CT (Invitrogen) sustituyendo una región de su ADN que contenía un gen de resistencia a blasticidina por un segmento de ADN que contenía un gen *leu2* (para la selección en medios que carecen de leucina). Se eliminó el gen de blasticidina mediante digestión de pYES6/CT con *BglII* y *NheI* y purificación en gel del fragmento de vector de ~4913 pb resultante. Se obtuvo el gen *leu2* a partir del vector de levadura pRS425 (ATCC 77106, n.º de GenBank U03452). Se usaron los cebadores PO-Leu5' (SEQ ID NO: 89) y

PO-Leu3' (SEQ ID NO: 90) en una reacción de PCR con pRS425 como molde para generar un fragmento de ADN de 1812 pb (desde el pb 664 hasta el 2475 de pRS425) que contenía el gen *leu2*.

PO-Leu5' GACTGCTAGCTTAAGCAAGGATTTTCTTAAC

PO-Leu3' GACTGGATCCTCCTGATGCGGTATTTTCTCC

Se incorporaron sitios de reconocimiento de enzimas de restricción en los cebadores para facilitar la clonación (5' *NheI* y 3' *BamHI* subrayados). Se digirió el fragmento de PCR con *BamHI* y *NheI* y se ligó al fragmento de vector de 4913 pb obtenido a partir del digesto de pYES6/CT *BglII/NheI* para formar pYES6-Leu. Se digirió este vector con *HindIII* y *XbaI* en la preparación para la inserción de sOrfA. Se digirió el plásmido de Blue Heron que contenía el sOrfA y el ADN flanqueante apropiado con *HindIII* y *XbaI*. Se purificó en gel el fragmento de 8,8 kb con el OrfA completo y se ligó al vector pYES6-Leu preparado para formar pBR882 (pYES6-Leu:sOrfA).

Constructo de expresión de sOrf B: Los inventores deseaban clonar el sOrfB en el vector de expresión de levadura pYES3 que tiene un marcador de selección de triptófano. Dado que el vector pYES3 contiene un segundo sitio de restricción de *XbaI* (el segundo sitio en el gen *trpI*), esa enzima de restricción no podía usarse convenientemente para la introducción del fragmento de ADN de sOrfB. Se modificó la región que contenía el sitio de *XbaI* en el sentido de 3' del sOrf B para introducir un sitio de *NotI* único (también disponible como sitio de clonación de inserción génica en pYES3) de la siguiente manera. Se digirió el plásmido que contenía el fragmento de sOrfB de Blue Heron con *HindIII* y *XbaI* y se purificó en gel el fragmento de interés de 6,2 kb resultante. Se ligó ese fragmento en pYES2/CT (Invitrogen) que se había cortado con las mismas enzimas, proporcionando el plásmido pBR879. Se abrió este plásmido cortando en el sitio de *XbaI* único. Se usó el ligador oligomérico autocomplementario 5'-CTAGGCGGCCGC-3' (SEQ ID NO: 91) para crear un sitio de *NotI* único (subrayado; también eliminó el sitio de *XbaI*). Esto proporcionó el plásmido pJK894. Se digirió este constructo con *HindIII* y *NotI* y se purificó en gel el fragmento de interés de 6,2 kb resultante. Se ligó ese fragmento en pYES3/CT (Invitrogen) que se había cortado con las mismas enzimas para formar pJK908 (pYES3 aOrfB).

Constructo de expresión de OrfC: El orfC nativo se había clonado previamente en un vector de expresión bacteriano y esto sirvió como fuente para el gen para la expresión en levadura. El vector bacteriano era pBluescript II KS (Stratagene) y se clonó la región codificante más ADN flanqueante en los sitios de *EcoRI* (5') y *XbaI* (3') del vector. El ADN de inserto incluía un sitio de restricción de *NdeI* como parte del codón de iniciación ATG y el codón de terminación TAA justo antes del sitio de *XbaI*. Se incluyó una secuencia de sitio de unión a ribosoma bacteriano en la región entre el sitio de *EcoRI* y el sitio de *NdeI* que contenía el codón de iniciación. Antes de la clonación en el vector de levadura, se eliminó el ADN del sitio de unión a ribosoma y se substituyó por ADN apropiado para la expresión en el sistema de levadura. Se digirió con *EcoRI* y *NdeI* el plásmido pBluescript que albergaba el orfC y se ligó a los ligadores oligonucleotídicos FL5' (AATTCAA) y FL3' (TATTG). Se digirió el plásmido resultante (denominado pKCFL) con *HindIII* (justo en el sentido de 5' del sitio de *EcoRI* en el polilígar pBluescript KS) y *XbaI* para liberar un fragmento de 4526 pb. Se ligó este fragmento a pYES2/CT digerido con *HindIII/XbaI* para generar: pYES2/ORFCwt (pYES2:OrfC).

Constructo de HetI: Se clonó el gene *hetI* de *Nostoc*, que codifica para una PPTasa, en un vector personalizado, pYES6-His/CT, que se construyó de la siguiente manera. Se modificó un vector pYES6/CT (Invitrogen) substituyendo una región de su ADN que contenía un gen de resistencia a blasticidina por un segmento de ADN que contenía un gen *his3* (para la selección en medios que carecen de histidina). Se eliminó el gen de blasticidina mediante digestión de pYES6/CT con *BglII* y *NheI* y purificación en gel del fragmento de vector de ~4913 pb resultante. Se amplificó este gen *his3* a partir del vector de levadura pRS423 (ATCC 77104, n.º de GenBank U03454) usando los cebadores PO-His5' (SEQ ID NO: 92) y PO-His3' (SEQ ID NO: 93).

PO-His5' GACTACTAGTCTAAGAAACCATTATTATCAT

PO-His3' GACTGGATCCAGCTTTAAATAATCGGTGTCA

Esto generó una región de ~1251 pb del plásmido pRS423 que contenía este gen *his3*. Se incorporaron sitios de reconocimiento de enzimas de restricción en los cebadores para facilitar la clonación (5' *SpeI* y 3' *BamHI*, subrayados). Se digirió el fragmento de PCR con *SpeI* y *BamHI* y se ligó al fragmento de vector de ~4913 pb obtenido a partir de pYES6/CT para formar pYES6-His. Se digirió este vector con *BamHI* y *XbaI* en la preparación para la inserción del gen *hetI*.

El gen *hetI* se había clonado anteriormente y usado con los genes de AGPI sintasa de *Schizochytrium* para la producción de AGPI en *E. coli* (publicación de solicitud de patente estadounidense n.º 20040235127, ejemplo 2). Tal como se indica en esa solicitud, no hay codones de metionina presentes en el marco de lectura abierto, pero hay varios posibles codones de iniciación alternativos (TTG y ATT) cerca del extremo 5' (Black y Wolk, 1994, JBC 116, 2282 - 2292). Se usó PCR para amplificar el Orf del ADN genómico de *Nostoc*. Se diseñó el cebador en 5' de modo que se substituyó la primera T del codón TTG en 5' más alejado por una A para crear un codón de metionina (ATG). El cebador en 3' incluía el codón de terminación TGA. La región amplificada se extendió desde el pb 3994 hasta el

3282 de la secuencia de nucleótidos de *Nostoc* depositada con el n.º de GenBank L22883 (siendo el nucleótido 3994 la segunda T del codón TTG alterado para formar el codón ATG). Se clonó este Orf de *hetI* amplificado en un vector pACYC 184 junto con elementos reguladores flanqueantes para la expresión en *E. coli*. Se usó este clon del Orf de *hetI* como ADN molde para amplificar el gen en la preparación para la clonación en pYES6-His. Se usaron los cebadores HetI 5' (SEQ ID NO: 94) y HetI 3' (SEQ ID NO: 95) para crear un fragmento de 740 pb que contenía el Orf de *hetI*.

HetI 5' GACTGGATCCGCCACCATGTTGCAGCATACTTGGCTACCAAACCC

10 HetI 3' GACTTCTAGATCAATAATGCCAGAATTTTGGCTGC

Se incorporaron sitios de reconocimiento de enzimas de restricción en los cebadores para facilitar la clonación (5' *Bam*HI y 3' *Xba*I, subrayados). El codón de iniciación de metionina ATG (cebador en 5') y el codón de terminación TGA (mostrado como el triplete TCA inverso en el cebador en 3') se muestran en negrita. Se digirió el producto de PCR con *Bam*HI y *Xba*I y se ligó en el vector pYES6-His anteriormente preparado para formar pYES-His/Het/CT (pYES6-His:HetI).

Resultados de expresar pYES6-Leu:sOrfA, pYES3:sOrfB, pYES2:OrfC y pYES6-His:HetI en levadura.

20 La figura 7 muestra una comparación de perfiles de CG de EMAG derivados de células de levadura que expresaban el sistema de AGPI sintasa de *Schizochytrium* (sOrfA, sOrfB, OrfC y *hetI*) y uno obtenido a partir de células control (que carecían del gen sOrfA), denominadas tales cepas de levadura en el presente documento cepas BRY 4.5 y BRY 3.3, respectivamente. Se recogieron células ~20 h tras la inducción. Puede observarse que aparecieron dos picos de EMAG novedosos en el perfil de la cepa que expresaba el sistema de AGPI sintasa completo. Estos dos
25 picos se identificaron como DPAn-6 y DHA mediante comparación del tiempo de elución con patrones auténticos y posteriormente mediante análisis de EM. Tal como se predijo a partir de la presente caracterización de la AGPI sintasa de *Schizochytrium*, aparte de DHA y DPAn-6, no resulta evidente ningún otro pico novedoso en el perfil. La figura 8 muestra la región del cromatograma de CG de la figura 8 que contiene los EMAG de AGPI. Tanto las células control como la célula que expresa la AGPI sintasa contienen un pico que eluye cerca del EMAG de DHA. Esto se
30 ha identificado como EMAG C26:0 (mediante análisis de espectro de masas) y probablemente se deriva de esfingolípidos. Aunque eluye cerca del pico de DHA, la resolución es suficiente de modo que no interfiere con la cuantificación de DHA. El pico de DPAn-6 está bien separado de otros lípidos de levadura endógenos en el perfil de EMAG. En este ejemplo particular de la cepa BRY 4.5, las células que expresaban el sistema de AGPI sintasa de *Schizochytrium* acumularon el 2,4% de DHA y el 2,0% de DPAn-6 (como porcentaje de los EMAG totales; véase la
35 tabla 4 a continuación). La suma de DHA y DPAn-6 es del 4,4% de los ácidos grasos medidos en las células. La razón de DHA con respecto a DPAn-6 observada en las células fue de ~1,2:1.

Los resultados presentados anteriormente que muestran la expresión de la AGPI sintasa *Schizochytrium* en levadura proporcionan una confirmación de la ruta propuesta en las anteriores solicitudes así como de las predicciones en
40 cuanto a las alteraciones de los perfiles de ácidos grasos que pueden esperarse en levadura y también en plantas.

Parte B

45 *Expresión de Orf A, B de AGPI sintasa de Schizochytrium y Het 1 de Nostoc en levadura en combinación con un gen híbrido que codifica para un OrfC que contiene una región DH2 derivada del homólogo de OrfC de Thraustochytrium 23B, y los efectos sobre los AGPI producidos en esas células.*

Expresión de genes orfC de Schizochytrium/Th.23B híbridos en levadura: Tal como se describe en otras secciones de esta solicitud, los inventores han descubierto que los principales determinantes de la razón de productos de AGPI n-3 con respecto a n-6 de AGPI sintasas residen en la proteína OrfC y más específicamente en la región DH2 de esa proteína. Experimentos de sustitución génica tanto en *E. coli* como en *Schizochytrium* usando el homólogo de OrfC derivado de *Th.23B* en combinación con los genes de AGPI sintasa derivados de *Schizochytrium* dieron como
50 resultado la alteración de la razón de DHA con respecto a DPAn-6 producida por esos sistemas mixtos. En *E. coli*, los productos de la AGPI sintasa se acumulan como ácidos grasos libres supuestamente sin influencia sobre la acumulación de los productos primarios de la enzima por parte de enzimas de síntesis de lípidos del organismo huésped. En *Schizochytrium*, los productos de AGPI se acumulan en los lípidos esterificados, pero es probable que las enzimas de síntesis de lípidos endógenas puedan admitir fácilmente tanto DHA como DPAn-6 ya que son
55 componentes principales de la fracción lipídica del huésped sin modificar. La expresión del sistema de AGPI sintasa mixto en levadura proporcionará un modelo para huéspedes eucariotas heterólogos (por ejemplo, plantas).

60 Los intentos para expresar los genes orfC de *Th.23B* no sintéticos o completamente sintéticos en levadura fueron insatisfactorios, ya que no pudieron detectarse las proteínas esperadas. En cambio, la expresión de los constructos de orfC híbridos (descritos a continuación) dio como resultado la producción de proteínas activas.

65 *OrfC de Schizochytrium / Th.23Bs híbridos en pYES2:* Se digirió el plásmido que contenía el orfC de *Schizochytrium* nativo, pYES2:OrfC (descrito anteriormente), con *Bs*WI y *Pml*I para eliminar la sección de ADN que codifica para la

región DH2 y algo del ADN flanqueante. La región eliminada fue de desde ~1179 pb (el sitio de *BsWI*) hasta ~3256 pb (el sitio de *Pml*) de la secuencia de orfC de *Schizochytrium* (SEQ ID NO: 5). Se purificó en gel el fragmento de 8,4 kb resultante (que contenía el vector así como las porciones en 5' y en 3' del OrfC). Se digirió un plásmido anteriormente descrito (véase el ejemplo 2) que contenía un orfC de *Schizochytrium* / *Th.23B* híbrido (pREZ179 = híbrido de pColA DUET-orfC de *Schizo.*- DH2 de *Th.23B*) con *BsWI* y *Pml* y se purificó en gel un fragmento de 2 kb que contenía la región DH2 de *Th.23B* y el ADN de *Schizochytrium* flanqueante. Se ligaron entre sí los dos fragmentos purificados para formar pYES2: OrfC-23BDH2.

Se usó una estrategia similar para crear pYES2: OrfC-s23BDH2. En este caso el plásmido usado como fuente para la región DH2 de *Th.23B* sintética (pDD22; véase el ejemplo 3) fue un orfC híbrido en el que el ADN que codifica para el dominio DH2 de *Th.23B* se derivó de una región codificante sintética cuyos codones se habían modificado para corresponder más estrechamente a las preferencias de *Schizochytrium* (véase el ejemplo 3).

Resultados de expresar pYES6-*Leu::sOrf A*, pYES3:*sOrf' B*, pYES6-*His:HetI* y pYES2:OrfC-23BDH2 orpYES2: OrfC-s23BDH2 en levadura: La tabla 4 muestra los AGPI producidos en levadura que expresaba constructos de Orf C híbrido junto con las subunidades A y B de *Schizochytrium* y *HetI* de *Nostoc*. Tal como se observó anteriormente en la parte A, los únicos picos novedosos detectados en estas muestras de levadura fueron DHA y DPAn-6. Las condiciones de crecimiento y la preparación de muestras fueron tal como se describió anteriormente. Sólo se muestran los datos de AGPI relevantes (como EMAG facilitados como % de área). Las muestras marcadas como BRY 4.21 contienen el orfC híbrido con la región DH2 de *Th.23B* nativa, mientras que la muestra marcada como BRY 4.23 contiene el orfC híbrido con la región DH2 de *Th.23B* derivada del gen sintético. Se sometieron a prueba dos muestras (a y b, de aislados independientes) para la cepa BRY 4.21 mientras que se sometió a prueba un aislado de la cepa BRY 4.23. Con respecto a las células que expresaban el orfC de *Schizochytrium*, aquellas células que expresaban cualquier forma del orfC híbrido tienen una razón DHA/DPAn-6 superior (un promedio de ~2,6 para aquellas con la DH2 de *Th.23B* nativa y un valor de ~2,9 para la muestra con la DH2 de *Th.23B* sintética). La expresión del gen orfC híbrido en levadura dio claramente como resultado un aumento en la razón de DHA con respecto a DPAn-6 con relación a la levadura que expresaba el gen orfC de *Schizochytrium* nativo. El hecho de que la razón DHA/DPAn-6 en las células de *Th.23B* o en *Schizochytrium* que expresan el orfC híbrido sea muy superior (□8-10) indica que otros factores están contribuyendo al sesgo hacia la acumulación de DHA con respecto a DPAn-6. La observación de que la razón sí que aumentó en levadura indica que este constructo es un modelo útil para expresar un sistema de AGPI sintasa en huéspedes eucariotas heterólogos (por ejemplo, levadura o plantas).

Tabla 4

Cepa	forma de orfC	DHA	DPAn-6	DHA + DPA	DHA/DPA
BRY 4.5	<i>Schizo.</i>	2,4	2,0	4,4	1,2
BRY 4.21a	DH2 de <i>Th.23B</i>	4,30	1,51	5,81	2,85
BRY 4.21b	DH2 de <i>Th.23B</i>	4,36	1,67	6,03	2,61
BRY 4.23	DH2 de <i>Th.23B</i> sint.	2,71	0,92	3,63	2,95

Ejemplo 7

El siguiente ejemplo demuestra la producción de AGPI en experimentos a escala de fermentación usando diversas cepas de *Schizochytrium* modificadas genéticamente descritas en el ejemplo 4.

Experimento 1

Usando fermentadores de 2 litros en condiciones de fermentación típicas, se cultivaron dos cultivos de un *Schizochytrium* de tipo natural (ATCC 20888) y dos cultivos de un *Schizochytrium* transgénico (B67-5, que tiene una región codificante de orfC de *Th.23B* optimizada para codones (sintética) en lugar de la región codificante de orfC de *Schizochytrium* nativa; véase el ejemplo 4) para comparar los perfiles de ácidos grasos. Se fermentó cada cepa en un medio que contenía carbono, nitrógeno, fósforo, sales, metales traza y vitaminas. Se inoculó cada fermentador con un cultivo de siembra típico, entonces se cultivó durante 80 horas y se le alimentó tanto una fuente de carbono como una fuente de nitrógeno durante el cultivo. La fuente de nitrógeno sólo se alimentó y se consumió durante la fase de crecimiento, mientras que la fuente de carbono se alimentó y se consumió durante la totalidad de la fermentación. Tras 80 horas, se centrifugaron muestras de cada fermentador, se liofilizaron y se analizaron mediante cromatografía de gases para determinar el contenido en ácidos grasos.

Condiciones de fermentación típicas:

Temperatura:

28 - 30°C

pH:

5,0 - 7,5

agitación:	100 - 300 cps
flujo de aire:	0,25 - 2,0 vvm
glucosa:	5 - 35 g/l (concentración)
inóculo:	7,5% - 15%

Los resultados fueron tal como se muestran en la tabla 5 a continuación:

5

Tabla 5

Registro de cepa	20888 de tipo natural	20888 de tipo natural	B67-5 transgénica	B67-5 transgénica
horas	80	80	80	80
fermentador	BN25	BN28	BN26	BN27
% de 10:0	0,02	0,01	0,01	0,01
% de 12:0	0,20	0,18	0,20	0,20
% de 13:0	0,00	0,00	0,07	0,00
% de 14:0	9,57	8,89	9,76	9,80
% de 16:0	33,68	32,58	34,62	34,51
% de 16:1	0,13	0,12	0,18	0,17
% de 17:0	0,08	0,09	0,07	0,07
% de 18:0	0,78	0,76	0,77	0,76
% de 18:1 n-9	0,00	0,00	0,08	0,08
% de 18:1 n-7	0,14	0,12	0,11	0,11
% de 18:3 n-6	0,14	0,15	0,08	0,08
% de 18:3 n-3	0,03	0,04	0,08	0,08
% de 20:0	0,09	0,08	0,08	0,08
% de 20:3 n-6	0,32	0,33	0,09	0,09
% de 20:4 ARA	0,25	0,30	0,10	0,11
% de 20:5 EPA	0,36	0,38	0,60	0,60
% de 22:5 n-6	14,98	15,37	6,52	6,52
% de 22:5 n-3	0,00	0,00	0,21	0,21
% de 22:6 DHA	37,32	38,64	44,47	44,58
DHA/DPA	2,49	2,51	6,82	6,84

Tal como se muestra en la tabla 5, la cepa B67-5 que contiene la región codificante de orfC de *Thraustochytrium* 23B sintética en lugar de la región codificante de *Schizochytrium* nativa produjo más DHA y tuvo una mayor razón de DHA con respecto a DPA n-6 que la cepa de *Schizochytrium* de tipo natural.

10

Experimento 2

15

Usando fermentadores de 10 litros en condiciones de fermentación típicas, se cultivaron un cultivo de un *Schizochytrium* de tipo natural (ATCC 20888) y un cultivo de *Schizochytrium* transgénico (B105-1A1; que contiene una región codificante de DH2 de *Th.23B* no optimizada para codones (nativa de *Thraustochytrium*) en lugar de la región DH2 de *Schizochytrium* nativa; véase el ejemplo 4) para comparar los perfiles de ácidos grasos. Se hizo crecer cada cepa en un medio que contenía carbono, nitrógeno, fósforo, sales, metales traza y vitaminas. Se inoculó cada fermentador con un cultivo de siembra típico, entonces se cultivó durante 72 horas y se le alimentó tanto una

fuelle de carbono como una fuente de nitr3geno durante el cultivo. La fuente de nitr3geno s3lo se aliment3 y se consumi3 durante la fase de crecimiento, mientras que la fuente de carbono se aliment3 y se consumi3 durante la totalidad de la fermentaci3n. Tras 72 horas, se centrifugaron muestras de cada fermentador, se liofilizaron y se analizaron mediante cromatograf3a de gases para determinar el contenido en 3cidos grasos.

5

Condiciones de fermentaci3n t3picas:

Temperatura:	28 - 30°C
pH:	5,0 - 7,5
agitaci3n:	100 - 300 cps
flujo de aire:	0,25 - 2,0 vvm
glucosa:	5 - 35 g/l (concentraci3n)
in3culo:	7,5%-15%

Los resultados se muestran en la tabla 6.

Registro de cepa	20888 de tipo natural	B105-1A-1
horas	72	72
recipiente	BN23	BN24
% de 10:0	0,00	0,00
% de 12:0	0,26	0,28
% de 13:0	0,09	0,10
% de 14:0	11,36	12,39
% de 16:0	37,10	40,02
% de 16:1	0,13	0,15
% de 17:0	0,07	0,06
% de 18:0	0,83	0,86
% de 18:1 n-9	0,00	0,11
% de 18:1 n-7	0,08	0,08
% de 18:3 n-6	0,13	0,05
% de 18:3 n-3	0,00	0,00
% de 20:0	0,08	0,10
% de 20:3 n-6	0,28	0,00
% de 20:4 ARA	0,26	0,00
% de 20:5 EPA	0,34	0,35
% de 22:5 n-6	14,48	4,40
% de 22:5 n-3	0,00	0,00
% de 22:6 DHA	34,07	39,56
DHA/DPA	2,53	8,98

10

La tabla 6 muestra que la cepa que comprende una regi3n DH2 de *Thraustochytrium* 23B en lugar de la regi3n DH2 de *Schizochytrium* tiene una raz3n DHA/DPA n-6 muy superior, ilustrando de nuevo la raz3n de DHA mejorada que se alcanza mediante el uso de sistemas de PKS de AGPI quim3ricos descritos en el presente documento.

15

Ejemplo 8

Este ejemplo describe la construcci3n y evaluaci3n de todas las combinaciones de regiones codificantes de orfA,

orfB y orfC de *Th. 23B* optimizadas para codones sintéticas expresadas en *Schizochytrium*.

Anteriormente (ejemplos 1 y 4) se han facilitado descripciones detalladas de métodos para la sustitución exacta de la región codificante de orfC de *Schizochytrium* por la región codificante de orfC de *Th.23B* optimizada para codones sintética. Los expertos en la técnica reconocen que estas técnicas pueden aplicarse generalmente a la mayoría de los genes de interés. Los expertos en la técnica reconocen además que tales diseños de genes y sustituciones pueden alcanzarse mediante variaciones de estos métodos u otros métodos totalmente diferentes. Por ejemplo, pueden delecionarse simultáneamente múltiples genes/regiones codificantes y sustituirse simultáneamente. En *Schizochytrium*, los genes orfA y orfB se encuentran próximos uno a otro ("ligados") en el genoma separados por una región intergénica (que comprende SEQ ID NO: 76). Estas dos regiones codificantes (junto con la región intergénica) pueden delecionarse simultáneamente mediante métodos análogos a los descritos anteriormente para orfC (publicación de solicitud de patente estadounidense n.º 20050100995). Entonces pueden usarse métodos similares a los descritos en los ejemplos 1 y 4 anteriores para crear simultáneamente sustituciones de "costura perfecta" de regiones codificantes de orfA y orfB de *Th.23B* optimizadas para codones sintéticas (incluyendo la región intergénica de *Schizochytrium* completa) en el locus de orfA/orfB de *Schizochytrium*. De esta manera se crearon cepas tales como B80-1 y B80-20 (tabla 7).

En otro ejemplo, pueden crearse delecciones de región codificante mediante un método "en dos etapas" en el que un plásmido que porta la estructura de delección marcada más un segundo marcador seleccionable se recombina inicialmente en su totalidad mediante un único acontecimiento de cruzamiento en el locus diana. Entonces, la estructura integrante "se resuelve" mediante un único acontecimiento de cruzamiento en un sitio en el lado opuesto de la estructura de delección de manera que se pierde el segundo marcador seleccionable y la estructura de delección permanece en lugar de la estructura génica original (Rothstein R., "Targeting, Disruption, Replacement, and Allele Rescue: Integrative DNA Transformation in Yeast", págs. 281-301 en *Methods in Enzymology*, vol. 194 (1991), Elsevier/Academic Press, Ámsterdam). De esta manera se creó el precursor de la cepa B71-1 (tabla 7).

Mediante los métodos explicados en este documento, se creó un conjunto de cepas de *Schizochytrium* en las que se han sustituido todas las combinaciones de las regiones codificantes de orfA, orfB y orfC de *Th.23B* sintéticas (optimizadas para codones) por las regiones codificantes de *Schizochytrium* relacionadas. El miembro del conjunto que no contiene genes de *Th.23B* es *Schizochytrium* ATCC20888 de tipo natural, y el miembro del conjunto que sólo contiene la región codificante de orfC de *Th.23B* optimizada para codones sintética (de longitud completa), B67-5, se describió en el ejemplo 4 y en la tabla 1 anteriores. Se evaluó este conjunto de ocho cepas para determinar la producción de ácidos grasos durante el crecimiento en medio SSFM tal como se describió en el ejemplo 4 anterior y se facilitan los datos en la tabla 7.

El plásmido pDD26 contiene la región codificante de orfA de *Th.23B* sintética de longitud completa unida mediante costura perfecta a las regiones en el sentido de 5' y en el sentido de 3' del gen orfA de *Schizochytrium*. La secuencia de nucleótidos de la región codificante de pDD26 se representa en el presente documento por SEQ ID NO: 71. SEQ ID NO: 71 codifica para SEQ ID NO: 39. Se depositó pDD26 con el n.º de registro de la ATCC PTA-8411, tal como se describió anteriormente en el presente documento.

El plásmido pDD32 contiene la región codificante de orfB de *Th.23B* sintética de longitud completa unida mediante costura perfecta a las regiones en el sentido de 5' y en el sentido de 3' del gen orfB de *Schizochytrium*. La secuencia de nucleótidos de la región codificante de pDD32 se representa en el presente documento por SEQ ID NO: 72. SEQ ID NO: 72 codifica para SEQ ID NO: 52. Se depositó pDD32 con el n.º de registro de la ATCC PTA-8412, tal como se describió anteriormente en el presente documento.

Los productos proteicos de las tres regiones codificantes de orf de *Th.23B* optimizadas para codones sintéticas funcionan en *Schizochytrium* e interaccionan satisfactoriamente con otros componentes de AGPI sintasa independientemente de la fuente. La expresión de la proteína OrfC de *Th.23B* (cepa B67-5) provoca un aumento en la razón DHA/DPA hasta un valor que se aproxima al de la cepa de *Th.23B* nativa, un resultado anteriormente demostrado en el ejemplo 4. Este fenómeno se observa para todas las combinaciones que expresan la proteína OrfC de *Th.23B* (B67-5, B79-11, B79-1 y B80-20). Sorprendentemente, la combinación de regiones codificantes de orfC de *Th.23B* optimizada para codones sintética más de orfA de *Th.23B* optimizadas para codones sintéticas (cepa B79-1) conduce al mayor nivel de producción de DHA, mientras que se mantiene la alta razón DHA/DPA. El aumento de la producción de DHA en esta cepa de *Schizochytrium* parece deberse tanto al aumento de la razón n-3/n-6 provocado por OrfC de *Th.23B* como al aumento de la producción de AGPI totales provocado por la interacción de OrfA de *Th.23B* con OrfC de *Th.23B*.

Estos datos demuestran que componentes del complejo de AGPI sintasa de diferentes organismos pueden funcionar conjuntamente de manera satisfactoria y pueden conferir características específicas del organismo fuente a un nuevo huésped. Además, la manipulación de la fuente y los niveles de expresión de componentes de AGPI sintasa pueden conducir a perfiles novedosos, productividades superiores y menores costes de ácidos grasos objetivo.

Tabla 7

cepa	Gen(es) orf de <i>Th. 23B</i>	EMAG	DHA	DPA (% de dcw)	DHA (% de EMAG)	DHA/DPA
		(% de dcw)	(% de dcw)			
ATCC20888	(ninguno)	73,9	16,4	5,4	22,1	3,04
871-1	A	74,2	17,2	5,15	23,2	3,34
882-3	B	67,9	15,4	4,93	22,7	3,12
B67-5	C	76,2	22,2	2,88	29,2	7,71
B80-1	AB	77,9	12,8	3,20	16,4	4,00
B79-11	BC	79,1	23,4	2,72	29,6	8,60
B79-1	AC	79,0	31,1	2,90	39,4	10,72
B80-20	ABC	77,4	20,9	2,32	27,0	9,01

Aunque se han descrito en detalle diversas realizaciones de la presente invención, resulta evidente que a los expertos en la técnica se les ocurrirán modificaciones y adaptaciones de esas realizaciones. Sin embargo, debe entenderse expresamente que tales modificaciones y adaptaciones se encuentran dentro del alcance de las reivindicaciones.

Lista de secuencias

10 <110> Martek Biosciences Corporation

<120> Sistemas de policétido sintasa de AGPI quiméricos y usos de los mismos

<130> P038919EP

15 <140> Documento 08755645.2

<141> 16-05-2008

20 <150> Documento US 11/749686

<151> 16-05-2007

<160> 102

25 <170> PatentIn versión 3.5

<210> 1

30 <211> 8733

<212> ADN

<213> *Schizochytrium* sp.

35 <220>

<221> CDS

40 <222> (1)..(8733)

<400> 1

ES 2 467 918 T3

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

ES 2 467 918 T3

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 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 ggt gac tgc gac atg atg gtc acc ggt			720
Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly			
	225	230	235
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 gtt 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 Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr			
	305	310	315
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 ggt cac ggc acc			1056
Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr			
	340	345	350
ggt act ccc gtt 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 ggt ctc gcc			1200
Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala			

ES 2 467 918 T3

385					390					395					400	
ggt	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	ccg	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	ggt	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	gcg	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
Ala	Thr	Pro	Ala	Ala	Leu	Gln	Ser	Leu	Cys	Glu	Ala	Gln	Leu	Lys	Glu	
			500				505						510			
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						525				
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					540					
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					555					560	
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						575		
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					590			
aag	ggc	atc	gcc	acc	aac	ggc	gct	gtc	gcc	gcg	ctc	ttc	tcc	ggc	cag	1824
Lys	Gly	Ile	Ala	Thr	Asn	Gly	Ala	Val	Ala	Ala	Leu	Phe	Ser	Gly	Gln	
		595				600						605				
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					620					
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					635					640	
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	

ES 2 467 918 T3

645				650				655								
cgc	aag	ccg	tac	gag	cgt	gag	ccc	gag	cag	gac	cac	aag	aag	atc	tcc	2016
Arg	Lys	Pro	Tyr	Glu	Arg	Glu	Pro	Glu	Gln	Asp	His	Lys	Lys	Ile	Ser	
			660													
ctc	acc	gcc	tac	tcg	cag	ccc	tcg	acc	ctg	gcc	tgc	gct	ctc	ggg	gcc	2064
Leu	Thr	Ala	Tyr	Ser	Gln	Pro	Ser	Thr	Leu	Ala	Cys	Ala	Leu	Gly	Ala	
			675													
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													
cat	tcg	ctc	ggg	gag	ttc	gcc	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													
cgc	gac	gag	ctc	ttt	gag	ctt	gtc	tgc	cgc	cgc	gcc	cgc	atc	atg	ggc	2208
Arg	Asp	Glu	Leu	Phe	Glu	Leu	Val	Cys	Arg	Arg	Ala	Arg	Ile	Met	Gly	
			725													
ggc	aag	gac	gca	ccg	gcc	acc	ccc	aag	ggc	tgc	atg	gcc	gcc	gtc	att	2256
Gly	Lys	Asp	Ala	Pro	Ala	Thr	Pro	Lys	Gly	Cys	Met	Ala	Ala	Val	Ile	
			740													
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													
ggc	aac	tcc	aac	tcg	cct	tcg	cag	acc	gtc	atc	acc	ggc	tcc	gtc	gaa	2352
Gly	Asn	Ser	Asn	Ser	Pro	Ser	Gln	Thr	Val	Ile	Thr	Gly	Ser	Val	Glu	
			770													
ggg	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	
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	
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													
ccc	aag	gcc	gag	acc	aag	ctc	ttc	agc	aac	gtc	tct	ggc	gag	acc	tac	2544
Pro	Lys	Ala	Glu	Thr	Lys	Leu	Phe	Ser	Asn	Val	Ser	Gly	Glu	Thr	Tyr	
			835													
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													
aag	ttc	ctc	acc	cag	gtc	cgc	aac	atg	cac	cag	gcc	ggg	gcg	cgc	atc	2640
Lys	Phe	Leu	Thr	Gln	Val	Arg	Asn	Met	His	Gln	Ala	Gly	Ala	Arg	Ile	
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	
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	

ES 2 467 918 T3

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

ES 2 467 918 T3

1145		1150		1155		
aag cgt gtc gag atc ctc tcc gag gtc cag gcc atg ctc aat gtc						3519
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 gtc 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 gtt 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 gcc att gac tcc 4464	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 gtt 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 gcc tct 4599	Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser	1520	1525	1530		
gcc ccg gcg 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 cgt 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 gtt 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 gtt ggt gag gtc gtc 4914	Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val					

ES 2 467 918 T3

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 cgt gtc gag atc ctc tcc gag gtt 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 gtt ggc gag gtt 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			
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 gtt cag gcc 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 gcc atg aag gct gag atc gcc 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			

ES 2 467 918 T3

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 gcc gcc 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 gtt 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 gtt ggc gag gtt gtc gat gcc 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 gcc 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 gcc 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	
gtt 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 gtt 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 gcc gct gct cct 6264			
Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro 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 gcc aag gtg gtt 6354			
Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val Val			

ES 2 467 918 T3

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						2140					2145					
2135																
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						2155					2160					
2150																
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						2170					2175					
2165																
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						2185					2190					
2180																
aag gag gcc gag cag cgc ttt ggc gct ctc ggc gcc ttc atc tcg	6624															
Lys Glu Ala Glu Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser						2200					2205					
2195																
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						2215					2220					
2210																
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						2230					2235					
2225																
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						2245					2250					
2240																
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						2260					2265					
2255																
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						2275					2280					
2270																
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						2290					2295					
2285																
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						2305					2310					
2300																
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						2320					2325					
2315																
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						2335					2340					
2330																
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 ctg 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			
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 tgc 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 gtt 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			

ES 2 467 918 T3

2585	2590	2595	
gtg cag atc atc ccc cgc Val Gln Ile Ile Pro Arg 2600	gag ggc ggc gct gat Glu Gly Gly Ala Asp 2605	acc gtg gcg cgc Thr Val Ala Arg 2610	7839
atc gtg ctc ggc tcc tcg Ile Val Leu Gly Ser Ser 2615	ccg gct gag atc ctt Pro Ala Glu Ile Leu 2620	gtc ggc aac tgg Val Gly Asn Trp 2625	7884
cgc acc ccg tcc aag aag Arg Thr Pro Ser Lys Lys 2630	gtc ggc tcg gac acc Val Gly Ser Asp Thr 2635	atc acc ctg cac Ile Thr Leu His 2640	7929
cgc aag att tcc gcc aag Arg Lys Ile Ser Ala Lys 2645	tcc aac ccc ttc ctc Ser Asn Pro Phe Leu 2650	gag gac cac gtc Glu Asp His Val 2655	7974
atc cag ggc cgc cgc gtg Ile Gln Gly Arg Arg Val 2660	ctg ccc atg acg ctg Leu Pro Met Thr Leu 2665	gcc att ggc tcg Ala Ile Gly Ser 2670	8019
ctc gcg gag acc tgc ctc Leu Ala Glu Thr Cys Leu 2675	ggc ctc ttc ccc ggc Gly Leu Phe Pro Gly 2680	tac tcg ctc tgg Tyr Ser Leu Trp 2685	8064
gcc att gac gac gcc cag Ala Ile Asp Asp Ala Gln 2690	ctc ttc aag ggt gtc Leu Phe Lys Gly Val 2695	act gtc gac ggc Thr Val Asp Gly 2700	8109
gac gtc aac tgc gag gtg Asp Val Asn Cys Glu Val 2705	acc ctc acc ccg tcg Thr Leu Thr Pro Ser 2710	acg gcg ccc tcg Thr Ala Pro Ser 2715	8154
ggc cgc gtc aac gtc cag Gly Arg Val Asn Val Gln 2720	gcc accg ctc aag acc Ala Thr Leu Lys Thr 2725	ttt tcc agc ggc Phe Ser Ser Gly 2730	8199
aag ctg gtc ccg gcc tac Lys Leu Val Pro Ala Tyr 2735	cgc gcc gtc atc gtg Arg Ala Val Ile Val 2740	ctc tcc aac cag Leu Ser Asn Gln 2745	8244
ggc gcg ccc ccg gcc aac Gly Ala Pro Pro Ala Asn 2750	gcc acc atg cag ccg Ala Thr Met Gln Pro 2755	ccc tcg ctc gat Pro Ser Leu Asp 2760	8289
gcc gat ccg gcg ctc cag Ala Asp Pro Ala Leu Gln 2765	ggc tcc gtc tac gac Gly Ser Val Tyr Asp 2770	ggc aag acc ctc Gly Lys Thr Leu 2775	8334
ttc cac ggc ccg gcc ttc Phe His Gly Pro Ala Phe 2780	cgc gcc atc gat gac Arg Gly Ile Asp Asp 2785	gtg ctc tcg tgc Val Leu Ser Cys 2790	8379
acc aag agc cag ctt gtg Thr Lys Ser Gln Leu Val 2795	gcc aag tgc agc gct Ala Lys Cys Ser Ala 2800	gtc ccc ggc tcc Val Pro Gly Ser 2805	8424
gac gcc gct cgc ggc gag Asp Ala Ala Arg Gly Glu 2810	ttt gcc acg gac act Phe Ala Thr Asp Thr 2815	gac gcc cat gac Asp Ala His Asp 2820	8469
ccc ttc gtg aac gac ctg Pro Phe Val Asn Asp Leu 2810	gcc ttt cag gcc atg Ala Phe Gln Ala Met 2815	ctc gtc tgg gtg Leu Val Trp Val 2820	8514

ES 2 467 918 T3

2825	2830	2835	
cgc cgc acg ctc ggc 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 ggc 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 ggc 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 taa			8733
Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala Phe			
2900	2905	2910	

<210> 2

5 <211> 2910

<212> PRT

<213> *Schizochytrium* sp.

10

<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		
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				

ES 2 467 918 T3

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

ES 2 467 918 T3

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
 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
 465 470 475 480
 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

ES 2 467 918 T3

Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asp His 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
 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

ES 2 467 918 T3

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

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

ES 2 467 918 T3

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	

ES 2 467 918 T3

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			
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			

ES 2 467 918 T3

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			
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			

ES 2 467 918 T3

Ala Pro 1880	Ala Pro	Ala Ala	Ala Ala	Ala Ala	Ala Pro	Ala Val	Ser	Ser Glu Leu	1890
Leu Glu 1895	Lys Ala	Glu Thr	Val	Val Met	Glu Val	Leu	Ala Ala Lys	1905	
Thr Gly 1910	Tyr Glu	Thr Asp	Met	Ile Glu	Ser Asp	Met	Glu Leu Glu	1920	
Thr Glu 1925	Leu Gly	Ile Asp	Ser	Ile Lys	Arg Val	Glu	Ile Leu Ser	1935	
Glu Val 1940	Gln Ala	Met Leu	Asn	Val Glu	Ala Lys	Asp	Val Asp Ala	1950	
Leu Ser 1955	Arg Thr	Arg Thr	Val	Gly Glu	Val Val	Asp	Ala Met Lys	1965	
Ala Glu 1970	Ile Ala	Gly Gly	Ser	Ala Pro	Ala Pro	Ala	Ala Ala Ala	1980	
Pro Ala 1985	Pro Ala	Ala Ala	Ala	Ala	Pro Ala	Val Ser	Asn Glu Leu Leu	1995	
Glu Lys 2000	Ala Glu	Thr Val	Val	Met Glu	Val Leu	Ala	Ala Lys Thr	2010	
Gly Tyr 2015	Glu Thr	Asp Met	Ile	Glu Ser	Asp Met	Glu	Leu Glu Thr	2025	
Glu Leu 2030	Gly Ile	Asp Ser	Ile	Lys Arg	Val Glu	Ile	Leu Ser Glu	2040	
Val Gln 2045	Ala Met	Leu Asn	Val	Glu Ala	Lys Asp	Val	Asp Ala Leu	2055	
Ser Arg 2060	Thr Arg	Thr Val	Gly	Glu Val	Val Asp	Ala	Met Lys Ala	2070	
Glu Ile 2075	Ala Gly	Gly Ser	Ala	Pro Ala	Pro Ala	Ala	Ala Ala Pro	2085	
Ala Ser 2090	Ala Gly	Ala Ala	Pro	Ala Val	Lys Ile	Asp	Ser Val His	2100	
Gly Ala 2105	Asp Cys	Asp Asp	Leu	Ser Leu	Met His	Ala	Lys Val Val	2115	

ES 2 467 918 T3

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			

ES 2 467 918 T3

Gly	Gly	Ala	Arg	Gly	Ile	Thr	Pro	Leu	Cys	Ile	Arg	Glu	Ile	Thr
2360						2365					2370			
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			

ES 2 467 918 T3

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			
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

5 <211> 6180

<212> ADN

<213> *Schizochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(6180)

<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
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
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

ES 2 467 918 T3

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

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

ES 2 467 918 T3

tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg atc acc gag ggc aac Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn 645 650 655	1968
aac tcc gtc tac cgc tgc gcc gag ctc ggc aag tac ctc ctc gag acc Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr 660 665 670	2016
ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc gat ctc tgc ggc agt Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser 675 680 685	2064
gcc gaa aac ctt tac gtc aag tct cgc cgc ttc aag gtg tcc acc tcc Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser 690 695 700	2112
gat acc ccg cgc gcc agc ttt gac gcc gcc gcc gat ggc tac ttt gtc Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val 705 710 715 720	2160
ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt gag act agc tgc acc Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr 725 730 735	2208
aag gac gac cgt atc tac gct tgc atg gat gcc atc gtc cct ggc aac Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn 740 745 750	2256
gtc cct agc gcc tgc ttg cgc gag gcc ctc gac cag gcg cgc gtc aag Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys 755 760 765	2304
ccg ggc gat atc gag atg ctc gag ctc agc gcc gac tcc gcc cgc cac Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His 770 775 780	2352
ctc aag gac ccg tcc gtc ctg ccc aag gag ctc act gcc gag gag gaa Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu 785 790 795 800	2400
atc ggc ggc ctt cag acg atc ctt cgt gac gat gac aag ctc ccg cgc Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg 805 810 815	2448
aac gtc gca acg ggc agt gtc aag gcc acc gtc ggt gac acc ggt tat Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr 820 825 830	2496
gcc tct ggt gct gcc agc ctc atc aag gct gcg ctt tgc atc tac aac Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn 835 840 845	2544
cgc tac ctg ccc agc aac ggc gac gac tgg gat gaa ccc gcc cct gag Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu 850 855 860	2592
gcg ccc tgg gac agc acc ctc ttt gcg tgc cag acc tcg cgc gct tgg Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp 865 870 875 880	2640
ctc aag aac cct ggc gag cgt cgc tat gcg gcc gtc tcg ggc gtc tcc Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser 885 890 895	2688

ES 2 467 918 T3

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

ES 2 467 918 T3

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 gtt 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	
ggg 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 gtt 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 gtt 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
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 Asp Lys Gln Asn Glu Asp Ala Trp Thr	
1370 1375 1380	

acc atc gtc aag ctt gtg gct tgc 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 gag gct tgc tac gct gcg ctc tgc aag ggt gaa aag ccc aag	4284
Ala Glu 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 tgc 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 cgt 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	4554
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 cgc 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 cgc 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 gcc cgc	4869
Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Arg	
1610 1615 1620	

ES 2 467 918 T3

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	cgt	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	ggg	gtc	acc	5049
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	gtt	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				
ggg	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	ggg	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	ggg	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
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	
gtt 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 taa	6180
Leu	

<210> 4

5 <211> 2059

<212> PRT

<213> *Schizochytrium* sp.

10

<400> 4

ES 2 467 918 T3

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	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	
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	

ES 2 467 918 T3

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 Leu Val 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
 500 505 510

ES 2 467 918 T3

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
 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

ES 2 467 918 T3

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
1160 1165 1170

Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile
1175 1180 1185

Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val
1190 1195 1200

Arg Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser
1205 1210 1215

Lys Tyr Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu
1220 1225 1230

Ile Ser Gly Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu
1235 1240 1245

Gly Gly Asn Ile Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly
1250 1255 1260

ES 2 467 918 T3

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

ES 2 467 918 T3

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

ES 2 467 918 T3

Leu	Ile	Ala	Ser	Gly	Glu	Ile	Asn	Gln	Glu	Gln	Ala	Glu	Leu	Ala
1745						1750					1755			
Arg	Arg	Val	Pro	Val	Ala	Asp	Asp	Ile	Ala	Val	Glu	Ala	Asp	Ser
1760						1765					1770			
Gly	Gly	His	Thr	Asp	Asn	Arg	Pro	Ile	His	Val	Ile	Leu	Pro	Leu
1775						1780					1785			
Ile	Ile	Asn	Leu	Arg	Asp	Arg	Leu	His	Arg	Glu	Cys	Gly	Tyr	Pro
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

<210> 5

5 <211> 4509

<212> ADN

<213> *Schizochytrium* sp.

10 <220>

<221> CDS

15 <222> (1)..(4509)

<400> 5

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

ES 2 467 918 T3

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

ES 2 467 918 T3

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

ES 2 467 918 T3

ggt acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr 625 630 635 640	1920
aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu 645 650 655	1968
atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro 660 665 670	2016
ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn 675 680 685	2064
gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg 690 695 700	2112
aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val 705 710 715 720	2160
cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys 725 730 735	2208
ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln 740 745 750	2256
gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn 755 760 765	2304
aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly 770 775 780	2352
agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu 785 790 795 800	2400
aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val 805 810 815	2448
tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr 820 825 830	2496
gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser 835 840 845	2544
gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu 850 855 860	2592
gcc atc gcc gcc cac gag gat ctc gct ggc aag cac ggc att gcc aac Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys His Gly Ile Ala Asn 865 870 875 880	2640

ccc acc ttt gtg cac gcc ccg ggc aag atc agc tgg aag tac cgc ggc Pro Thr Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly 885 890 895	2688
cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc gtg Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val 900 905 910	2736
tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe 915 920 925	2784
ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val 930 935 940	2832
cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc tct Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala Ser 945 950 955 960	2880
gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc gct Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro Ala 965 970 975	2928
gtc gcc tcc ggc ccg gcc cag acc atc gac ctc aag cag ctc aag acc Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys Thr 980 985 990	2976
gag ctc ctc gag ctc gat gcc ccg ctc tac ctc tcg cag gac ccg acc Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr 995 1000 1005	3024
agc ggc cag ctc aag aag cac acc gac gtg gcc tcc ggc cag gcc Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln Ala 1010 1015 1020	3069
acc atc gtg cag ccc tgc acg ctc ggc gac ctc ggt gac cgc tcc Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg Ser 1025 1030 1035	3114
ttc atg gag acc tac ggc gtc gtc gcc ccg ctg tac acg ggc gcc Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly Ala 1040 1045 1050	3159
atg gcc aag ggc att gcc tcg gcg gac ctc gtc atc gcc gcc ggc Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly 1055 1060 1065	3204
aag cgc aag atc ctc ggc tcc ttt ggc gcc ggc ggc ctc ccc atg Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro Met 1070 1075 1080	3249
cac cac gtg cgc gcc gcc ctc gag aag atc cag gcc gcc ctg cct His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu Pro 1085 1090 1095	3294
cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct ttt gac agc Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser 1100 1105 1110	3339
aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag ggc gtc Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val 1115 1120 1125	3384

act	gtg	gtg	gag	gcc	tcg	gca	ttc	atg	acc	ctc	acc	ccg	cag	gtc	3429
Thr	Val	Val	Glu	Ala	Ser	Ala	Phe	Met	Thr	Leu	Thr	Pro	Gln	Val	
	1130					1135					1140				
gtg	cgc	tac	cgc	gcc	gcc	ggc	ctc	tcg	cgc	aac	gcc	gac	ggg	tcg	3474
Val	Arg	Tyr	Arg	Ala	Ala	Gly	Leu	Ser	Arg	Asn	Ala	Asp	Gly	Ser	
	1145					1150					1155				
gtc	aac	atc	cgc	aac	cgc	atc	atc	ggc	aag	gtc	tcg	cgc	acc	gag	3519
Val	Asn	Ile	Arg	Asn	Arg	Ile	Ile	Gly	Lys	Val	Ser	Arg	Thr	Glu	
	1160					1165					1170				
ctc	gcc	gag	atg	ttc	atc	cgc	ccg	gcc	ccg	gag	cac	ctc	ctc	gag	3564
Leu	Ala	Glu	Met	Phe	Ile	Arg	Pro	Ala	Pro	Glu	His	Leu	Leu	Glu	
	1175					1180					1185				
aag	ctc	atc	gcc	tcg	ggc	gag	atc	acc	cag	gag	cag	gcc	gag	ctc	3609
Lys	Leu	Ile	Ala	Ser	Gly	Glu	Ile	Thr	Gln	Glu	Gln	Ala	Glu	Leu	
	1190					1195					1200				
gcg	cgc	cgc	gtt	ccc	gtc	gcc	gac	gat	atc	gct	gtc	gag	gct	gac	3654
Ala	Arg	Arg	Val	Pro	Val	Ala	Asp	Asp	Ile	Ala	Val	Glu	Ala	Asp	
	1205					1210					1215				
tcg	ggc	ggc	cac	acc	gac	aac	cgc	ccc	atc	cac	gtc	atc	ctc	ccg	3699
Ser	Gly	Gly	His	Thr	Asp	Asn	Arg	Pro	Ile	His	Val	Ile	Leu	Pro	
	1220					1225					1230				
ctc	atc	atc	aac	ctc	cgc	aac	cgc	ctg	cac	cgc	gag	tgc	ggc	tac	3744
Leu	Ile	Ile	Asn	Leu	Arg	Asn	Arg	Leu	His	Arg	Glu	Cys	Gly	Tyr	
	1235					1240					1245				
ccc	gcg	cac	ctc	cgc	gtc	cgc	gtt	ggc	gcc	ggc	ggg	ggc	gtc	ggc	3789
Pro	Ala	His	Leu	Arg	Val	Arg	Val	Gly	Ala	Gly	Gly	Gly	Val	Gly	
	1250					1255					1260				
tgc	ccg	cag	gcc	gcc	gcc	gcc	gcg	ctc	acc	atg	ggc	gcc	gcc	ttc	3834
Cys	Pro	Gln	Ala	Ala	Ala	Ala	Ala	Leu	Thr	Met	Gly	Ala	Ala	Phe	
	1265					1270					1275				
atc	gtc	acc	ggc	act	gtc	aac	cag	gtc	gcc	aag	cag	tcc	ggc	acc	3879
Ile	Val	Thr	Gly	Thr	Val	Asn	Gln	Val	Ala	Lys	Gln	Ser	Gly	Thr	
	1280					1285					1290				
tgc	gac	aac	gtg	cgc	aag	cag	ctc	tcg	cag	gcc	acc	tac	tcg	gat	3924
Cys	Asp	Asn	Val	Arg	Lys	Gln	Leu	Ser	Gln	Ala	Thr	Tyr	Ser	Asp	
	1295					1300					1305				
atc	tgc	atg	gcc	ccg	gcc	gcc	gac	atg	ttc	gag	gag	ggc	gtc	aag	3969
Ile	Cys	Met	Ala	Pro	Ala	Ala	Asp	Met	Phe	Glu	Glu	Gly	Val	Lys	
	1310					1315					1320				
ctc	cag	gtc	ctc	aag	aag	gga	acc	atg	ttc	ccc	tcg	cgc	gcc	aac	4014
Leu	Gln	Val	Leu	Lys	Lys	Gly	Thr	Met	Phe	Pro	Ser	Arg	Ala	Asn	
	1325					1330					1335				
aag	ctc	tac	gag	ctc	ttt	tgc	aag	tac	gac	tcc	ttc	gac	tcc	atg	4059
Lys	Leu	Tyr	Glu	Leu	Phe	Cys	Lys	Tyr	Asp	Ser	Phe	Asp	Ser	Met	
	1340					1345					1350				
cct	cct	gcc	gag	ctc	gag	cgc	atc	gag	aag	cgt	atc	ttc	aag	cgc	4104
Pro	Pro	Ala	Glu	Leu	Glu	Arg	Ile	Glu	Lys	Arg	Ile	Phe	Lys	Arg	
	1355					1360					1365				

gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att aac 4149
Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile Asn
1370 1375 1380

ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac ccc 4194
Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro
1385 1390 1395

aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt gcc 4239
Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala
1400 1405 1410

agc cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac tac 4284
Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr
1415 1420 1425

cag gtc tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc atc 4329
Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile
1430 1435 1440

aag ggc acc tac ctc gac ccc gct gtc tcc aac gag tac ccc tgt 4374
Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys
1445 1450 1455

gtc gtc cag atc aac ctg caa atc ctc cgt ggt gcc tgc tac ctg 4419
Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu
1460 1465 1470

cgc cgt ctc aac gcc ctg cgc aac gac ccg cgc att gac ctc gag 4464
Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu
1475 1480 1485

acc gag gat gct gcc ttt gtc tac gag ccc acc aac gcg ctc taa 4509
Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu
1490 1495 1500

<210> 6

5 <211> 1502

<212> PRT

<213> *Schizochytrium* sp.

10

<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

ES 2 467 918 T3

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

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

ES 2 467 918 T3

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
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

ES 2 467 918 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

ES 2 467 918 T3

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 His Gly Ile Ala Asn
 865 870 875 880

Pro Thr Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly
 885 890 895

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

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

Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val
 930 935 940

Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala Ser
 945 950 955 960

Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro Ala
 965 970 975

Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys Thr
 980 985 990

Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr
 995 1000 1005

Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln Ala
 1010 1015 1020

Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg Ser
 1025 1030 1035

Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly Ala
 1040 1045 1050

Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly
 1055 1060 1065

Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro Met
 1070 1075 1080

His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu Pro
 1085 1090 1095

ES 2 467 918 T3

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

ES 2 467 918 T3

Lys	Leu	Tyr	Glu	Leu	Phe	Cys	Lys	Tyr	Asp	Ser	Phe	Asp	Ser	Met
1340						1345					1350			
Pro	Pro	Ala	Glu	Leu	Glu	Arg	Ile	Glu	Lys	Arg	Ile	Phe	Lys	Arg
1355						1360					1365			
Ala	Leu	Gln	Glu	Val	Trp	Glu	Glu	Thr	Lys	Asp	Phe	Tyr	Ile	Asn
1370						1375					1380			
Gly	Leu	Lys	Asn	Pro	Glu	Lys	Ile	Gln	Arg	Ala	Glu	His	Asp	Pro
1385						1390					1395			
Lys	Leu	Lys	Met	Ser	Leu	Cys	Phe	Arg	Trp	Tyr	Leu	Gly	Leu	Ala
1400						1405					1410			
Ser	Arg	Trp	Ala	Asn	Met	Gly	Ala	Pro	Asp	Arg	Val	Met	Asp	Tyr
1415						1420					1425			
Gln	Val	Trp	Cys	Gly	Pro	Ala	Ile	Gly	Ala	Phe	Asn	Asp	Phe	Ile
1430						1435					1440			
Lys	Gly	Thr	Tyr	Leu	Asp	Pro	Ala	Val	Ser	Asn	Glu	Tyr	Pro	Cys
1445						1450					1455			
Val	Val	Gln	Ile	Asn	Leu	Gln	Ile	Leu	Arg	Gly	Ala	Cys	Tyr	Leu
1460						1465					1470			
Arg	Arg	Leu	Asn	Ala	Leu	Arg	Asn	Asp	Pro	Arg	Ile	Asp	Leu	Glu
1475						1480					1485			
Thr	Glu	Asp	Ala	Ala	Phe	Val	Tyr	Glu	Pro	Thr	Asn	Ala	Leu	
1490						1495					1500			

<210> 7

5 <211> 1500

<212> ADN

<213> *Schizochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1500)

<400> 7

ES 2 467 918 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

ES 2 467 918 T3

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	ggc	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	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	ggt	gac	tgc	gac	atg	atg	gtc	acc	ggt	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	gtt	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	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	ggc	cac	ggc	acc		1056
Ala	Cys	Val	Asp	Pro	Ala	Thr	Val	Thr	Leu	Val	Glu	Gly	His	Gly	Thr		
			340					345					350				
ggt	act	ccc	gtt	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	ggc	ctc	gcc		1200
Ser	Ile	Lys	Ser	Ser	Ile	Gly	His	Leu	Lys	Ala	Val	Ala	Gly	Leu	Ala		
385					390					395					400		
ggt	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	ccg		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	ggt		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	gcg	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														1500
Ala	Thr	Pro	Ala														
			500														

<210> 8

5 <211> 500

<212> PRT

<213> *Schizochytrium* sp.

10

<400> 8

Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg

ES 2 467 918 T3

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
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
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
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			

ES 2 467 918 T3

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			
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				480
Thr	Ala	Tyr	Arg	Leu	Asn	Lys	Arg	Pro	Gln	Pro	Val	Leu	Met	Met	Ala
				485					490					495	
Ala	Thr	Pro	Ala												
			500												

<210> 9

5 <211> 1278

<212> ADN

<213> *Schizochytrium* sp.

5

<220>

<221> CDS

10 <222> (1)..(1278)

<400> 9

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 gac cac aag aag	288
Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asp His 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 ggc 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	

ES 2 467 918 T3

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			
	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			1278
Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro			
	420	425	

<210> 10

5 <211> 426

<212> PRT

<213> *Schizochytrium* sp.

10

<400> 10

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

ES 2 467 918 T3

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 Asp His 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
 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

ES 2 467 918 T3

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

<210> 11

5 <211> 5

<212> PRT

<213> *Schizochytrium* sp.

10

<220>

<221> MISC_FEATURE

15 <222> (4)..(4)

<223> Xaa puede ser cualquier aminoácido que se produce de manera natural

<400> 11

20

Gly His Ser Xaa Gly
1 5

<210> 12

25 <211> 258

<212> ADN

ES 2 467 918 T3

<213> *Schizochytrium* sp.

<220>

5 <221> CDS

<222> (1)..(258)

<400> 12

10
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 258
Lys Ala Glu Ile Ala Gly
85

<210> 13

15 <211> 86

<212> PRT

<213> *Schizochytrium* sp.

20

<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

<210> 14
 <211> 5
 5 <212> PRT
 <213> *Schizochytrium* sp.
 10 <400> 14
 Leu Gly Ile Asp Ser
 1 5
 <210> 15
 15 <211> 21
 <212> PRT
 20 <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
 25 <210> 16
 <211> 3006
 30 <212> ADN
 <213> *Schizochytrium* sp.
 <400> 16
 35

ES 2 467 918 T3

gcccccgccc	cggtcaaggc	tgctgcgcc	gccgcccccg	ttgcctcggc	cctgcccccg	60
gctgtctcga	acgagcttct	tgagaaggcc	gagactgtcg	tcattggaggt	cctcgccgcc	120
aagaccggct	acgagaccga	catgatcgag	gctgacatgg	agctcgagac	cgagctcggc	180
attgactcca	tcaagcgtgt	cgagatcctc	tccgaggtcc	aggccatgct	caatgtcgag	240
gccaaggatg	tcgatgccct	cagccgcact	cgcactgttg	gtgaggttgt	caacgccatg	300
aaggccgaga	tcgctggcag	ctctgccccg	gcgcctgctg	ccgctgctcc	ggctccggcc	360
aaggctgccc	ctgccgccgc	tgcgctgct	gtctcgaacg	agcttctcga	gaaggccgag	420
accgtcgtca	tggaggtcct	cgccgccaag	actggctacg	agactgacat	gatcgagtcc	480
gacatggagc	tcgagactga	gctcggcatt	gactccatca	agcgtgtcga	gacctctctc	540
gaggttcagg	ccatgctcaa	cgctcgaggcc	aaggacgtcg	acgctctcag	ccgcactcgc	600
actgtgggtg	aggtcgtcaa	cgccatgaag	gctgagatcg	ctggtggctc	tgccccggcg	660
cctgccgccg	ctgccccagg	tccggctgct	gccgcccctg	cgcttgccgc	cgccgcccct	720
gctgtctcga	acgagcttct	tgagaaggcc	gagaccgtcg	tcattggaggt	cctcgccgcc	780
aagactggct	acgagactga	catgatcgag	tccgacatgg	agctcgagac	cgagctcggc	840
attgactcca	tcaagcgtgt	cgagattctc	tccgaggtcc	aggccatgct	caacgtcgag	900
gccaaggacg	tcgacgtctc	cagccgcacc	cgcactgttg	gcgaggtcgt	cgatgccatg	960
aaggccgaga	tcgctgggtg	ctctgccccg	gcgcctgccg	ccgctgctcc	tgctccggct	1020

ES 2 467 918 T3

gctgccgccc ctgcgcctgc cgccectgcg cctgctgtct cgagcgagct tctcgagaag 1080
gccgagactg tcgtcatgga ggtcctcgcc gccaaagactg gctacgagac tgacatgata 1140
gagtcgcaca tggagctcga gaccgagctc ggcattgact ccatcaagcg tgtcgagatt 1200
ctctccgagg tccaggccat gctcaacgtc gaggccaagg acgtcgacgc tctcagccgc 1260
accgcactg ttggcgaggt cgtcgatgcc atgaaggccg agatcgctgg tggctctgcc 1320
ccggcgccctg ccgccgctgc tctgctccg gctgctgccg ccctgcgcc tgccgcccct 1380
gcgcctgccg ccctgcgcc tgctgtctcg agcgagcttc tcgagaaggc cgagactgtc 1440
gtcatggagg tctcgcgc caagactggc tacgagactg acatgattga gtccgacatg 1500
gagctcgaga ccgagctcgg cattgactcc atcaagcgtg tcgagattct ctccgaggtt 1560
caggccatgc tcaacgtcga ggccaaggac gtcgacgctc tcagccgcac tcgactgtt 1620
ggtgaggtcg tcgatgccat gaaggctgag atcgctggca gtcgcgcctc ggcgctgcc 1680
gccgctgctc ctgctccggc tgctgccgct cctgcgcccg ctgccgcgc 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 gttggcgagg ttgtcgatgc catgaaggcc 1980
gagatcgctg gtggctctgc cccggcgccct gccgcgcgtg ccctgctcc ggctgccgcc 2040
gcccctgctg tctcgaacga gcttctcgag aaggccgaga ctgtcgatcat ggaggtcctc 2100
gccgccaaga ctggctacga gaccgacatg atcgagtcg acatggagct cgagaccgag 2160
ctcggcattg actccatcaa gcgtgtcgag attctctccg aggttcaggc catgctcaac 2220
gtcgaggcca aggacgtcga tgctctcagc cgcactcgca ctgttggcga ggtcgtcgat 2280
gccatgaagg ctgagatcgc cggcagctcc gcccggcgcc ctgccgcgc tgctcctgct 2340
ccggctgctg ccgtcctgc gcccgctgcc gctgcccctg ctgtctcgag cgagcttctc 2400
gagaaggccg agaccgtcgt catggaggtc ctgcgcgcca agactggcta cgagactgac 2460
atgattgagt ccgacatgga gtcgagact gagctcgga ttgactccat caagcgtgtc 2520
gagatcctct ccgaggttca ggccatgctc aacgtcgagg ccaaggacgt cgatgccctc 2580
agccgcaccc gcactgttg cgaggttgct gatgccatga aggcgagat cgtggtggc 2640
tctgccccgg cgctgccgc cgtgcccct gctccggctg ccgcgcgcc tgctgtctcg 2700
aacgagcttc ttgagaaggc cgagaccgtc gtcattggagg tctcgcgc 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 ggcgcctgcc gccgctgctc ctgcctcggc tggcgccgcg . 3000

cctgcg 3006

<210> 17

5 <211> 2133

<212> ADN

<213> *Schizochytrium* sp.

10 <220>

<221> CDS

15 <222> (1)..(2133)

<400> 17

ES 2 467 918 T3

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

ES 2 467 918 T3

agc gaa ccg gca tgg tgc gct ggc atc act gac gag aag gct gtg caa Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln 195 200 205	624
aag gct gct acc cag gag ctc aag cgc gcc ttt agc gct ggc gag ggc Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly 210 215 220	672
ccc aag ccc acg ccc cgc gct gtc act aag ctt gtg ggc tct gtt ctt Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu 225 230 235 240	720
ggc gct cgc gag gtg cgc agc tct att gct gcg att gaa gcg ctc ggc Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly 245 250 255	768
ggc aag gcc atc tac tcg tcg tgc gac gtg aac tct gcc gcc gac gtg Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val 260 265 270	816
gcc aag gcc gtg cgc gat gcc gag tcc cag ctc ggt gcc cgc gtc tcg Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser 275 280 285	864
ggc atc gtt cat gcc tcg ggc gtg ctc cgc gac cgt ctc atc gag aag Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys 290 295 300	912
aag ctc ccc gac gag ttc gac gcc gtc ttt ggc acc aag gtc acc ggt Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly 305 310 315 320	960
ctc gag aac ctc ctc gcc gcc gtc gac cgc gcc aac ctc aag cac atg Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met 325 330 335	1008
gtc ctc ttc agc tcg ctc gcc ggc ttc cac ggc aac gtc ggc cag tct Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser 340 345 350	1056
gac tac gcc atg gcc aac gag gcc ctt aac aag atg ggc ctc gag ctc Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu 355 360 365	1104
gcc aag gac gtc tcg gtc aag tcg atc tgc ttc ggt ccc tgg gac ggt Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly 370 375 380	1152
ggc atg gtg acg ccg cag ctc aag aag cag ttc cag gag atg ggc gtg Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val 385 390 395 400	1200
cag atc atc ccc cgc gag ggc ggc gct gat acc gtg gcg cgc atc gtg Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val 405 410 415	1248
ctc ggc tcc tcg ccg gct gag atc ctt gtc ggc aac tgg cgc acc ccg Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro 420 425 430	1296
tcc aag aag gtc ggc tcg gac acc atc acc ctg cac cgc aag att tcc Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser 435 440 445	1344

ES 2 467 918 T3

gcc aag tcc aac ccc ttc ctc gag gac cac gtc atc cag gcc cgc cgc Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg 450 455 460	1392
gtg ctg ccc atg acg ctg gcc att ggc tcg ctc gcg gag acc tgc ctc Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu 465 470 475 480	1440
ggc ctc ttc ccc ggc tac tcg ctc tgg gcc att gac gac gcc cag ctc Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu 485 490 495	1488
ttc aag ggt gtc act gtc gac ggc gac gtc aac tgc gag gtg acc ctc Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu 500 505 510	1536
acc ccg tcg acg gcg ccc tcg ggc cgc gtc aac gtc cag gcc acg ctc Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu 515 520 525	1584
aag acc ttt tcc agc ggc aag ctg gtc ccg gcc tac cgc gcc gtc atc Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile 530 535 540	1632
gtg ctc tcc aac cag ggc gcg ccc ccg gcc aac gcc acc atg cag ccg Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro 545 550 555 560	1680
ccc tcg ctc gat gcc gat ccg gcg ctc cag ggc tcc gtc tac gac ggc Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly 565 570 575	1728
aag acc ctc ttc cac ggc ccg gcc ttc cgc ggc atc gat gac gtg ctc Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu 580 585 590	1776
tcg tgc acc aag agc cag ctt gtg gcc aag tgc agc gct gtc ccc ggc Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly 595 600 605	1824
tcc gac gcc gct cgc ggc gag ttt gcc acg gac act gac gcc cat gac Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp 610 615 620	1872
ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg cgc Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg 625 630 635 640	1920
cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc atc gtc Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val 645 650 655	1968
cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att acc ctc cgc Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg 660 665 670	2016
tcc aac cag tcg ggc ggt cac tcc cag cac aag cac gcc ctt cag ttc Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe 675 680 685	2064
cac aac gag cag ggc gat ctc ttc att gat gtc cag gct tcg gtc atc His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile 690 695 700	2112
gcc acg gac agc ctt gcc ttc Ala Thr Asp Ser Leu Ala Phe 705 710	2133

<210> 18

<211> 711

5

<212> PRT

<213> *Schizochytrium* sp.

10 <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
 180 185 190

Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln
 195 200 205

ES 2 467 918 T3

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

ES 2 467 918 T3

Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu
465 470 475 480

Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu
485 490 495

Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu
500 505 510

Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu
515 520 525

Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile
530 535 540

Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro
545 550 555 560

Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly
565 570 575

Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu
580 585 590

Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly
595 600 605

Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp
610 615 620

Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg
625 630 635 640

Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val
645 650 655

Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg
660 665 670

Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe
675 680 685

His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser~Val Ile
690 695 700

Ala Thr Asp Ser Leu Ala Phe
705 710

<210> 19

5 <211> 1350

<212> ADN

<213> *Schizochytrium* sp.

5

<220>

<221> CDS

10 <222> (1)..(1350)

<400> 19

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
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

ES 2 467 918 T3

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			
cgt	ctc	gat	gat	gcc	atc	cgc	gac	ggc	gac	cac	atc	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	cgt	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	cgt	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	ctc	gtc	gca	gcc	ggc	ttt	gcc	ggt	atg	tgc	aag	gtc	ctc	ctc	tcc	1152
Thr	Leu	Val	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	
					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	gcc	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															1350
Ala	Cys															
	450															

<210> 20

5 <211> 450

<212> PRT

<213> *Schizochytrium* sp.

5

<400> 20

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
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

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

ES 2 467 918 T3

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 Leu Val 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

<210> 21

5 $\langle 211 \rangle$ 1323

<212> ADN

<213> *Schizochytrium* sp.

<220>

<221> CDS

5 <222> (1)..(1323)

<400> 21

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
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 ggt 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 gtt 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 gcc 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 ggt 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 gtt gcg ggt 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	

ES 2 467 918 T3

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 ggt 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
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	400
gaa ccc gcc cct gag gcg ccc tgg gac agc acc ctc ttt gcg tgc 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

5 <211> 441

<212> PRT

<213> *Schizochytrium* sp.

10

<400> 22

Ser	Ala	Arg	Cys	Gly	Gly	Glu	Ser	Asn	Met	Arg	Ile	Ala	Ile	Thr	Gly
1				5				10						15	

ES 2 467 918 T3

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

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

ES 2 467 918 T3

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

5 <211> 1500

<212> ADN

<213> *Schizochytrium* sp.

10 <220>

<221> CDS

15 <222> (1)..(1500)

<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			

ES 2 467 918 T3

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

ES 2 467 918 T3

cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg cgc Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg 290 295 300	912
ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc aag tac Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr 305 310 315 320	960
gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc att agc ggc Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly 325 330 335	1008
aag ccc gac gcc tgc aag gct gcg atc gcg cgt ctc ggt ggc aac att Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile 340 345 350	1056
cct gcg ctt ccc gtg acc cag ggc atg tgc ggc cac tgc ccc gag gtg Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val 355 360 365	1104
gga cct tat acc aag gat atc gcc aag atc cat gcc aac ctt gag ttc Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe 370 375 380	1152
ccc gtt gtc gac ggc ctt gac ctc tgg acc aca atc aac cag aag cgc Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg 385 390 395 400	1200
ctc gtg cca cgc gcc acg ggc gcc aag gac gaa tgg gcc cct tct tcc Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser 405 410 415	1248
ttt ggc gag tac gcc ggc cag ctc tac gag aag cag gct aac ttc ccc Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro 420 425 430	1296
caa atc gtc gag acc att tac aag caa aac tac gac gtc ttt gtc gag Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu 435 440 445	1344
gtt ggg ccc aac aac cac cgt agc acc gca gtg cgc acc acg ctt ggt Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly 450 455 460	1392
ccc cag cgc aac cac ctt gct ggc gcc atc gac aag cag aac gag gat Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp 465 470 475 480	1440
gct tgg acg acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu 485 490 495	1488
gtt cct ggc gtc Val Pro Gly Val 500	1500

<210> 24

5 <211> 500

<212> PRT

<213> *Schizochytrium* sp.

10

<400> 24

ES 2 467 918 T3

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		
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	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			

ES 2 467 918 T3

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

<210> 25

5 <211> 1530

<212> ADN

<213> *Schizochytrium* sp.

5

<220>

<221> CDS

10 <222> (1)..(1530)

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

ES 2 467 918 T3

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

5 <211> 510

<212> PRT

<213> *Schizochytrium* sp.

10

<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

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 80

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

ES 2 467 918 T3

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

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

ES 2 467 918 T3

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

5 <211> 1350

<212> ADN

<213> *Schizochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1350)

<400> 27

ES 2 467 918 T3

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

ES 2 467 918 T3

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
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	
225					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	gcg	cgt	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

ES 2 467 918 T3

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	1350
Pro Ser	
450	

<210> 28

5 <211> 450

<212> PRT

<213> *Schizochytrium* sp.

10

<400> 28

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	

ES 2 467 918 T3

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						

ES 2 467 918 T3

	405		410		415										
Ile	Ser	Asp	Tyr	Gly	Lys	Gly	Asp	Leu	Asn	Lys	Lys	Ile	Val	Val	Asp
			420					425					430		

Phe	Lys	Gly	Ile	Ala	Leu	Lys	Met	Gln	Lys	Arg	Ser	Thr	Asn	Lys	Asn
		435					440					445			

Pro Ser
450

<210> 29

5 <211> 1497

<212> ADN

<213> *Schizochytrium* sp.

10 <220>

<221> CDS

15 <222> (1)..(1497)

<400> 29

ES 2 467 918 T3

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

ES 2 467 918 T3

145					150					155					160	
aag	cac	gtc	aac	tac	cgc	aac	atc	gac	ctc	gac	ccc	tcc	aag	ggc	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	ggc	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	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
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	ggc	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	ggc	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	

ES 2 467 918 T3

	405	410	415	
	gcc gcc cac gag gat ctc gct ggc aag cac ggc att gcc aac ccc acc			1296
	Ala Ala His Glu Asp Leu Ala Gly Lys His Gly Ile Ala Asn Pro Thr			
	420	425	430	
	ttt gtg cac gcc ccg ggc aag atc agc tgg aag tac cgc ggc cag ctc			1344
	Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly Gln Leu			
	435	440	445	
	acg ccc aag agc aag aag atg gac tcg gag gtc cac atc gtg tcc gtg			1392
	Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser Val			
	450	455	460	
	gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc ctc tgg			1440
	Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu Trp			
	465	470	475	480
	gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg cgc atc			1488
	Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg Ile			
	485	490	495	
	gcc tcc ggt			1497
	Ala Ser Gly			

<210> 30

5 <211> 499

<212> PRT

<213> *Schizochytrium* sp.

10

<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	

ES 2 467 918 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	

ES 2 467 918 T3

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 His Gly Ile Ala Asn Pro Thr
420 425 430

Phe Val His Ala Pro Gly Lys Ile Ser Trp Lys Tyr Arg Gly Gln Leu
435 440 445

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

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

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

Ala Ser Gly

<210> 31

5 <211> 1512

<212> ADN

<213> *Schizochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1512)

<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

ES 2 467 918 T3

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

ES 2 467 918 T3

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

5 <211> 504

<212> PRT

<213> *Schizochytrium* sp.

10

<400> 32

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

ES 2 467 918 T3

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

ES 2 467 918 T3

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

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

Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile
370 375 380

Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro
385 390 395 400

Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser
405 410 415

Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val
420 425 430

Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr
435 440 445

Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile
450 455 460

Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala
465 470 475 480

Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe
485 490 495

Val Tyr Glu Pro Thr Asn Ala Leu
500

<210> 33

5 <211> 714

<212> ADN

<213> *Nostoc* sp.

10

<400> 33

atgttgcagc atacttggct accaaaaccc ccaaatttaa ccttattgtc agatgaagtt	60
catctctggc gcattcccct tgaccaacca gaatcacagc tacaggattt agccgctacc	120
ttatctagtg acgaattagc ccgtgcaaac agattttatt ttcccgaaca tcgccggcgt	180

tttactgctg gtcgtgggtat tctccgcagt atcttggggg gctatttggg tgtggaacca 240
 gggcaagtta aatttgatta tgaatcccgt ggtaaaccac tattaggcga tcgctttgcc 300
 gagagtgggtt tattatttaa cttgtcacac tcccagaact tggccttggtg tgcagtcaat 360
 tacacgcgcc aaatcggcat cgatttagaa tatctccgcc ccacatctga tttagaatcc 420
 cttgccaaaaa gggtctttttt accgcgagaa tatgaattat tgcgatcgct acccgatgag 480
 caaaaacaaa aaattttctt tcgttactgg acttgtaaag aggttatct taaagcaacg 540
 ggtgacggca tcgctaaatt agaggaaatt gaaatagcac taactccac agaaccagct 600
 aagttacaga cagctccagc gtggagtctc ctagagctag tgccagatga taattgtgtt 660
 gctgctgttg ccgtggcggg ttttggctgg cagccaaaat tctggcatta ttga 714

<210> 34

5 <211> 237

<212> PRT

<213> *Nostoc* sp.

10

<400> 34

Met Leu Gln His Thr Trp Leu Pro Lys Pro Pro Asn Leu Thr Leu Leu
 1 5 10 15

Ser Asp Glu Val His Leu Trp Arg Ile Pro Leu Asp Gln Pro Glu Ser
 20 25 30

Gln Leu Gln Asp Leu Ala Ala Thr Leu Ser Ser Asp Glu Leu Ala Arg
 35 40 45

Ala Asn Arg Phe Tyr Phe Pro Glu His Arg Arg Arg Phe Thr Ala Gly
 50 55 60

Arg Gly Ile Leu Arg Ser Ile Leu Gly Gly Tyr Leu Gly Val Glu Pro
 65 70 75 80

Gly Gln Val Lys Phe Asp Tyr Glu Ser Arg Gly Lys Pro Ile Leu Gly
 85 90 95

Asp Arg Phe Ala Glu Ser Gly Leu Leu Phe Asn Leu Ser His Ser Gln
 100 105 110

Asn Leu Ala Leu Cys Ala Val Asn Tyr Thr Arg Gln Ile Gly Ile Asp
 115 120 125

Leu Glu Tyr Leu Arg Pro Thr Ser Asp Leu Glu Ser Leu Ala Lys Arg
 130 135 140

Phe Phe Leu Pro Arg Glu Tyr Glu Leu Leu Arg Ser Leu Pro Asp Glu

ES 2 467 918 T3

145		150		155		160
Gln Lys Gln Lys Ile Phe Phe Arg Tyr Trp Thr Cys Lys Glu Ala Tyr						
		165		170		175
Leu Lys Ala Thr Gly Asp Gly Ile Ala Lys Leu Glu Glu Ile Glu Ile						
		180		185		190
Ala Leu Thr Pro Thr Glu Pro Ala Lys Leu Gln Thr Ala Pro Ala Trp						
		195		200		205
Ser Leu Leu Glu Leu Val Pro Asp Asp Asn Cys Val Ala Ala Val Ala						
		210		215		220
Val Ala Gly Phe Gly Trp Gln Pro Lys Phe Trp His Tyr						
		225		230		235

<210> 35

5 <211> 8733

<212> ADN

<213> Secuencia artificial

10

<220>

<223> sintética

15

<400> 35

ES 2 467 918 T3

atggctgcta ggttgcaaga acaaaaaggt ggtgagatgg atactagaat tgctatcatt	60
ggaatgtctg ctatcttgcc atgtggtact actgttagag aatcttggga aactattaga	120
gctgggtattg attgtttgtc tgatttgcct gaagatagag ttgatgttac tgcttacttt	180
gatccagtta aaactactaa agataaaaatc tattgtaaga gaggtgggtt cattccagaa	240
tatgatcttg atgctagaga atttgggttg aatatgtttc agatggaaga ttctgatgct	300
aatcaaaacta tttctttgtt gaaagttaaa gaagcattgc aagatgctgg catcgatgct	360
ttgggttaaag agaagaagaa tattgggttg gttttgggta ttgggtgggg tcaaaaatct	420
tctcatgaat tttactcaag attgaattat gttgttggtg agaaggtatt gagaaaaatg	480
ggtatgccag aagaagatgt taaagttgct gttgaaaaat acaaagctaa ttttccagag	540
tggagattgg attcttttcc aggtttcttg ggaaatgtta ctgcaggaag atgtactaat	600
acttttaatc ttgatggcat gaattgtgtt gttgatgctg cttgtgcttc ttctttgatt	660
gctgttaaag ttgctattga tgaattgttg tacggtgatt gtgatatgat gggtactggg	720
gctacttgta ctgataattc tattggaatg tacatggctt tttctaaaac tccagttttc	780
tctactgatc catctgttag agcttatgat gaaaaaacta aaggaatgtt gattggtgaa	840
ggttctgcta tgttgggttt gaaaagatat gctgatgctg ttagagatgg tgatgaaatt	900
catgctgtta ttagagggtg tgcttcttct tctgatggta aagctgctgg tatctatact	960

ccaactat	ctgg	caaga	agaagcattg	agaagagctt	ataatagagc	ttgtgttgat	1020
ccagctactg	ttacttttgg	tgaagg	tcat	ggtactgg	ta	ctccagttgg	1080
gaattgactg	ctttgagaaa	tttgtttgat	aaagcatatg	gtgaaggtaa	tactgaaaaa		1140
gttgctgttg	gttctattaa	atcttctatt	ggtcatttga	aagctgttgc	tggtttggct		1200
ggaatgatta	aagttatcat	ggctttgaaa	cataaaaactt	tgccaggaac	tattaatg		1260
gataatccac	caaacttgta	cgataatact	ccaattaacg	aatcttcttt	gtacattaat		1320
actatgaata	gaccttgg	tccaccacca	ggtgttccaa	gaagagctgg	tatttcttct		1380
tttggttttg	gtgg	tgctaa	ttatcatgct	gttttggag	aagctgaacc	agaacatact	1440
actgcttata	ggttgaacaa	aagaccacaa	ccagttttga	tgatggctgc	tactccagct		1500
gctttgcaat	ctttgtgtga	agctcaattg	aaagaatttg	aagctgctat	taaagaaaac		1560
gaaactgtta	aaaatactgc	ttatattaaa	tgtgttaa	attggtgaaca	attcaaattc		1620
cctggtagta	ttccagctac	taatgctagg	ttgggtttct	tggttaaaga	tgctgaagat		1680
gcttgttcta	ctttgagagc	tatttgtgct	caatttgcta	aagatgttac	taaagaagca		1740
tggagattgc	caagagaagg	tgtttctttt	agagctaaag	gtattgctac	taatggtgct		1800
gttgctgctt	tgtttctgg	tcaagg	tgct	caatatactc	atatgttttc	tgaagttgct	1860
atgaattggc	cacaattcag	acaatctatt	gctgctatgg	atgctgctca	atctaaagtt		1920
gctggttctg	ataaagattt	tgaaagagtt	tctcaagttt	tgtatccaag	aaaaccatac		1980
gagagagaac	cagagcaaga	tcataagaag	atttctttga	ctgcttattc	tcaaccatct		2040
actttggctt	gtgctttggg	tgcttttgaa	atttttaaag	aagctggttt	tactccagat		2100
tttgctgctg	gtcattcttt	gggtgaattt	gctgctttgt	acgctgctgg	ttgtgttgat		2160
agagatgaat	tgtttgaatt	ggtttgtaga	agagctagaa	ttatgggtgg	taaagatgct		2220
ccagctactc	caaaagggtg	catggctgct	gttattgg	ta	caa	aatgctga	2280
gttcaagctg	ctaattgtttg	gttaggaaat	tctaattctc	catctcaa	ac	tgttattact	2340
ggttctgttg	aagg	tattca	agctgaatct	gctaggttgc	aaaaagaagg	ttttagagtt	2400
gttccattgg	cttgtgaatc	tgcttttcat	tctccacaga	tgga	aatgc	ttcttctgct	2460
tttaaagatg	ttatctctaa	agtttctttt	agaactccaa	aagctgaaac	taaattgttt		2520
tctaattgtt	ctgg	tgaac	ttatccaact	gatgctagag	aatgttgac	tcaacatatg	2580
acttcttctg	ttaaattttt	gactcaagtt	agaaatatgc	atcaagctgg	tgctagaatt		2640
tttgttgaat	tcgg	tccaaa	acaagttttg	tctaaattgg	tttctgaaac	tttgaaagat	2700
gatccatctg	ttgttactgt	ttctgttaat	ccagcttctg	gtactgattc	tgatattcaa		2760
ttgagagatg	ctgctgttca	attgggtgtt	gctgggtgtta	atttgcaagg	ttttgataaa		2820
tgggatgctc	cagatgctac	tagaatgcaa	gctattaaaa	aaaaaagaac	tactttgaga		2880

ttgtctgctg	ctacttatgt	ttctgataaa	actaagaaag	ttagagatgc	tgctatgaat	2940
gatggtagat	gtgttactta	cttgaaaggt	gctgctccat	tgattaaagc	tccagaacca	3000
gttgttgatg	aagctgctaa	aagagaagct	gaaagattgc	aaaaagaatt	gcaagatgct	3060
caaagacaat	tggatgatgc	taaaagagct	gctgctgaag	ctaattctaa	attggctgct	3120
gctaaagaag	aagctaaaac	tgctgctgct	tctgctaaac	cagctggtga	tactgctggt	3180
gttgaaaaac	atagagctat	tttgaaatct	atgttggtcg	aattggatgg	ttatggttct	3240
gttgatgctt	cttctttgca	acaacaacaa	caacaacaaa	ctgctccagc	tccagttaaa	3300
gctgctgctc	cagctgctcc	agttgcttct	gctccagcac	ccgcagttag	caacgaactc	3360
ttagaaaaag	ccgagacagt	agtgatggaa	gttcttgacg	ctaaaacggg	gtacgaaaca	3420
gatatgattg	aagcagatat	ggaacttgaa	actgaactgg	gcattgattc	gattaaacgc	3480
gtggaaattc	tgtcagaagt	gcaagctatg	ttaaatgttg	aagcgaaaga	tgttgatgca	3540
ctgtcacgca	cacgcaccgt	ggcggaagta	gtgaacgcca	tgaaagcaga	aattgcaggc	3600
tcctcagcac	ccgcgccggc	cgcagcagca	ccagcccccg	caaaagccgc	ccccgcagcg	3660
gcggctccag	ccgtttcaaa	cgaattactc	gaaaaagcag	aaaccgtagt	gatggaagtc	3720
cttgccgcca	aaacgggtta	tgagaccgat	atgatcgaaa	gcgatatgga	attagaaacc	3780
gaattagggg	ttgatagtat	taaacgcgta	gaaattctgt	ccgaagtaca	agctatgctg	3840
aatgtagaag	caaaagatgt	agatgcgtta	agccgcacac	gactgtttgg	tgaagtgtg	3900
aatgctatga	aagctgaaat	tgcaggaggt	tcagcaccgg	ccccagcagc	cgcagcccca	3960
ggctccagcag	cagccgcacc	ggcccccgcc	gccgccgcac	cggcagtatc	aaacgagttg	4020
ttagagaaaag	cggaaaccgt	tgtgatggaa	gtacttgccg	cgaagacagg	ttacgagacc	4080
gatatgatcg	aaagtgacat	ggaattagaa	accgaattgg	gcattgatag	cattaaacgc	4140
gtagaaattt	tatccgaagt	tcaagccatg	ttaaatgttg	aagccaaaga	tgtggatgcg	4200
ttatccccgca	cgcgtaccgt	cggagaagta	gtggacgcta	tgaaagcaga	gattgcagga	4260
ggaagtgcac	cggctccagc	agcagcagca	cccgccccag	cggcagcggc	gccggcaccg	4320
gccgctccgg	ccccagccgt	tagttcagaa	ctcctcgaaa	aagcagaaac	tgttgtcatg	4380
gaagtattag	ctgcaaaaac	aggttacgag	acggatatga	ttgaaagcga	tatggaatta	4440
gaaaccgaat	taggcattga	ttcaattaaa	cgtgttgaaa	tcttaagtga	agtccaagcc	4500
atgcttaatg	ttgaagccaa	agatgtagat	gcattatctc	gcacgcgtac	agtgggtgaa	4560
gttgtcgatg	cgatgaaagc	agaaatcgcg	ggaggatcag	cgccagcccc	ggcagcagca	4620
gcccccgcg	ccgccgcggc	cgcacctgcg	ccggccgccc	cagcccctgc	agcaccggcc	4680
ccagcagtgt	cgctcggaatt	actcgaaaaa	gctgaaacgg	tcgttatgga	agtacttgct	4740
gcaaagacgg	gctatgaaac	ggatatgatt	gaatcggata	tggaattaga	aacagaactt	4800

ggtattgact	ctattaaacg	cgtggaaatt	ctgagcgaag	tacaggcaat	gttaaacgta	4860
gaagccaaag	atgtagacgc	tttgtcacgc	acacggacgc	taggagaagt	tgtggatgcg	4920
atgaaagctg	aaattgccgg	ttcaagtget	agcgcccctg	ctgccgccgc	ccctgcccct	4980
gccgccgcag	caccggcccc	ggcagccgca	gctccagcag	ttagtaacga	attactcgaa	5040
aaagcagaaa	cgggtggcat	ggaagtgtta	gcagcaaaaa	ctggatatga	aacggacatg	5100
attgaaagcg	atatggaatt	agaaacagaa	ctgggaattg	atagtattaa	acgtgttgag	5160
attttatctg	aggttcaagc	tatgctgaat	gttgaagcga	aagatgtaga	cgcactgtct	5220
cggacccgca	cagtaggtga	agtgggtggac	gcatgaaag	cagaaatcgc	aggtggaagt	5280
gctccggccc	cggcggcagc	cgcacccgcg	cccgcggccg	cagccccagc	agttagcaac	5340
gaattactcg	agaaagcaga	aactgtagtg	atggaagtgt	tagccgcaaa	aacgggttat	5400
gaaacggata	tgattgaaag	cgatatggaa	ctggaaaccg	aactgggcat	tgattctatt	5460
aaacgtgtcg	aaatcttatc	ggaagtccaa	gcaatgctga	acgtagaggc	aaaggatggt	5520
gatgccctgt	cacgtacccg	taccgtaggt	gaagttgtag	atgccatgaa	agctgaaatc	5580
gcaggcagta	gcgccccggc	accagccgcc	gccgcccccg	cgccggcagc	cgccgcaccc	5640
gcgccagccg	cagctgctcc	agctgtatct	agtgagctgc	tcgaaaaagc	agaaaccgtg	5700
gttatggaag	tgctcgccgc	taaaacagga	tatgaaaccg	atatgattga	aagcgatatg	5760
gaattagaaa	ccgaactggg	tattgatagt	attaagcgtg	ttgaaatttt	gtcagaagtt	5820
caagctatgt	tgaatgtaga	agccaaagat	gtagacgctt	taagtcggac	gcgtactggt	5880
ggagaagtgc	tagacgcat	gaaagcagag	attgcaggcg	gaagtgcacc	ggccccggca	5940
gcagcagccc	cagcaccagc	ggccgccgct	cctgcagtgt	caaacgaact	tctggaaaaa	6000
gctgaaaccg	tcgtcatgga	agtgctggct	gcaaaaactg	gatatgaaac	agacatgatt	6060
gaatcagata	tggaactcga	aaccgaactg	gggattgata	gcattaaacg	tgtggaaatt	6120
ttatcggagg	tacaagcaat	gttaaattgt	gaagcaaaag	atgtggatgc	actgagccgt	6180
actcgtactg	ttggtgaggt	cgtggatgcg	atgaaagcag	aaattgctgg	agggagtgcg	6240
cctgcccccg	ccgccgccgc	acccgcgtct	gccggtgctg	ccccgcgtgt	caaaattgat	6300
tctgttcatg	gtgctgattg	tgatgatttg	tctttgatgc	atgctaaagt	tgttgatatt	6360
agaagaccag	atgaattgat	tttggaaga	ccagaaaata	gaccagtttt	ggttgttgat	6420
gatggttctg	aattgacttt	ggctttgggt	agagttttgg	gtgcttgtgc	tgttgttttg	6480
acttttgaag	gtttgcaatt	ggctcaaaga	gctggtgctg	ctgctattag	acatgttttg	6540
gctaaagatt	tgtctgctga	atctgctgaa	aaagctatta	aagaagctga	acaaagattt	6600
ggtgcttttg	gtggttttat	ctctcaacaa	gctgaaagat	ttgaaccagc	tgaaattttg	6660
ggttttactt	tgatgtgtgc	taaatttgct	aaagcatctt	tgtgcactgc	tgttgctggg	6720

ES 2 467 918 T3

ggtagaccag ctttcattgg tgttgctagg ttggatggta gggtggggtt tactttctcaa	6780
ggaacttctg atgctttgaa aagagctcaa agaggtgcta tttttgggtt gtgcaagact	6840
attggtttgg aatggctctga atctgatgtt ttctcaagag gtggtgatat tgctcaagg	6900
atgcatccag aagatgctgc tgttgctatt gttagagaaa tggcttgtgc tgatattaga	6960
attagagaag ttggtattgg tgctaatacaa caaagatgta ctattagagc tgctaaattg	7020
gaaactggaa atccacaaag acaaattgct aaagatgatg ttttggtgggt ttctgggtgg	7080
gctagaggaa ttactccatt gtgcattaga gaaattacta gacaaattgc tgggtggaaag	7140
tatatattgt tgggtaggtc taaagtttct gcttctgaac cagcttgggtg tgctgggtatt	7200
actgatgaaa aagctgttca aaaagctgct actcaagaat tgaaaagagc tttttctgct	7260
ggatgaaggc caaaaccaac tccaagagct gttactaaat tgggtgggtc tgttttgggt	7320
gctagagaag ttaggtcttc tattgctgct attgaagcat tgggtggaaa agctatctat	7380
tcttcttggt atgttaattc tgctgctgat gttgctaaag ctggttagaga tgctgaatct	7440
caattgggtg ctagagtctc tgggtattgt catgcttctg gtgttttgag agatagggtg	7500
attgaaaaaa aattgccaga tgaatttgat gctgtttttg gtactaaagt tactgggttg	7560
gaaaatttgt tggctgctgt tgatagagct aatttgaaac atatgggttt gttttcttct	7620
ttggctgggt ttcatggtaa tgttggtcaa tctgattatg ctatggctaa cgaagcattg	7680
aacaaaatgg gtttggaatt ggctaaagat gtttctgtta aatctatttg ttttggctct	7740
tgggatgggt gtatgggtac tccacaattg aaaaaacaat ttcaagaaat ggggtgttcaa	7800
attattccaa gagaagggtg tgctgatact gttgctagaa ttgttttggg ttcttctcca	7860
gctgaaattt tgggttgtaa ttggagaact ccatctaaaa aagttgggtc tgatactatt	7920
actttgcaca gaaaaatttc tgctaaatct aatccatttt tgggaagatca tgtcattcaa	7980
ggtagaagag ttttgccaat gactttggct attggttctt tggctgaaac ttgtttgggt	8040
ttgtttctct gatattcttt gtgggctatt gatgatgctc aattgtttta aggtgttact	8100
gttgatgggt atgttaattg tgaagttact ttgactccat ctactgctcc ttctggtaga	8160
gttaatgttc aagctacttt gaaaactttt tcttctggta aattgggttc agcttataga	8220
gctgttattg ttttgtctaa tcaagggtgc ccaccagcta atgctactat gcaaccacca	8280
tctttggatg ctgatccagc ttgcaagggt tctgtttatg atggaaagac tttgtttcat	8340
ggccagctt ttagagggtat tgatgatgtt ttgtcttgta ctaaatctca attggttgct	8400
aaatgttctg ctgttccagg ttctgatgct gctagagggt aatttgctac tgatactgat	8460
gctcatgatc catttggtta tgatttggct tttcaagcta tgttggtttg ggtagaaga	8520
actttgggtc aagctgcttt gccaaattct attcaaagaa ttgttcaaca cagaccagtt	8580
ccacaagata aaccattttt tattactttg agatctaata aatctgggtg tcattctcaa	8640

cataaacatg ctttgcaatt tcataacgaa caaggtgatt tgttcattga tgttcaagca 8700
tctggtattg ctactgattc tttggctttt taa 8733

<210> 36

5 <211> 6180

<212> ADN

<213> Secuencia artificial

10 <220>

<223> sintética

15 <400> 36

atggctgcta gaaatgtttc tgctgctcat gaaatgcatg atgaaaaaag aattgctgtt 60
gttggtatgg ctgttcaata tgctggttgt aagactaaag atgaattttg ggaagttttg 120
atgaatggta aagttgaatc taaagttatc tctgataaaa gattgggttc taattaccga 180
gctgaacatt acaaggctga aagatccaaa tacgctgata ctttttgtaa cgaaacttat 240
ggctactttg atgaaaacga aattgataac gaacatgaat tgttgttgaa tttggctaaa 300
caagcattgg ctgaaacttc tgttaaagat tctactagat gtggtattgt ttctgggtgt 360
ttgtcttttc ctatggataa tttgcaagggt gaattgttga atgtctatca aaatcatgtt 420
gagaagaaat tgggtgctag agttttttaa gatgcttctc attggtctga aagagaacaa 480
tctaacaaac cagaagctgg tgatagaaga attttcatgg acccagcttc ttttgttget 540
gaagaattga atttgggtgc tttgcattat tctgttgatg ctgcttgtgc tactgcttta 600
tacgttttga gattggctca agatcatttg gtttctggtg ctgctgatgt tatgttgtgt 660
ggtgctactt gtttgccaga accattcttt atcttgtctg gtttttctac ttttcaagct 720
atgccagttg gtactggtca aaatgtttct atgccattgc ataaagattc tcaaggtttg 780
actccaggtg aaggtggttc tatcatgggt ttgaaaagat tggatgatgc tattagagat 840
ggtgatcata tctatggtac tttgttgggt gctaattgtt ctaattctgg cactggtttg 900
ccattgaaac cattgttgcc atctgaaaaa aaatgtttga tggatactta tactagaatt 960
aatgttcac cacaataaat tcaatatgtt gaatgtcatg ctactggtac tccacaaggt 1020
gatagggttg aaattgatgc tgttaaagca tgttttgaag gaaaagttcc aagatttgggt 1080
actactaaag gaaacttttg tcatactttg gttgctgctg gttttgctgg aatgtgcaaa 1140
gttttgttgt ctatgaaaca tggatcatt ccaccaactc caggattga tgatgaaact 1200
aagatggacc cattggttgt ttctggtgaa gctattcctt ggccagaaac taatggtgaa 1260
ccaaaaagag ctggtttgtc tgcttttgggt tttggtggtg ctaatgctca tgctgttttt 1320
gaagaacatg atccatctaa tgctgcttgt actggtcatg attctatttc tgctttgtct 1380
gctagatgtg gtggtgaatc taatatgaga attgctatta ctgggatgga tgctactttt 1440
ggtgctttga aaggtttgga tgcttttgaa agagccatct acactggtgc tcatggtgct 1500

attccattgc cagaaaagag atggagattt ttgggcaaag ataaagattt cttggatttg 1560
 tgtggtgtta aagctactcc acatgggtgt tatattgaag atgttgaagt tgattttcaa 1620
 agattgagaa ctccaatgac tccagaagat atgttggtgc cacaacaatt gttggctggt 1680
 actactattg atagagctat tttggattct ggtatgaaaa aaggtggtaa tgttgctggt 1740
 tttgttggtt tgggtaccga tttggaattg tacagacata gagctagagt tgctttgaaa 1800
 gaaagagtta gaccagaagc atctaaaaaa ttgaatgata tgatgcagta cattaatgat 1860
 tgtggcacct ctacttctta tacttcttat attggaattt tggttgctac tagagtttct 1920
 tctcaatggg gttttactgg tccatctttt actattactg aagggaataa ctctgtttat 1980
 agatgtgctg aattgggaaa gtatttggtg gaaactggtg aagttgatgg tgttggtggt 2040
 gctggtggtg atttggtggtg ttctgctgaa aacttatacg ttaaatacaag aagattcaaa 2100
 gtttctactt ctgatactcc aagagcttct tttgatgctg ctgctgatgg ttactttggt 2160
 ggtgaagggt gtggtgcttt tgttttgaaa agagaaactt cttgtactaa agatgataga 2220
 atctatgctt gcatggatgc tattgttcca ggtaatgttc catctgcttg tttgagagaa 2280
 gcattggatc aagctagagt taaaccaggt gatattgaaa tgttggaatt gtctgctgat 2340
 tctgctagac atttgaaaga tccatctggt ttgccaaaag aattgactgc tgaagaagaa 2400
 attggtggtt tgcaaaactat tttgagagat gatgataaat tgccaagaaa tgttgctact 2460
 ggttctgtta aagctactgt tggtgatact gggtatgctt ctgggtgctgc ttctttgatt 2520
 aaagctgctt tgtgcatcta taataggtat ttgccatcta atggtgatga ttgggatgaa 2580
 ccagctccag aagctccttg ggattctact ttggttgctt gtcaaacttc aagagcttgg 2640
 ttgaaaaatc ctggagagag aagatatgct gctgtttctg gtgtttctga aactaggtct 2700
 tgttattctg ttttggtgtc tgaagctgaa ggtcattatg aaagagaaaa tagaatttct 2760
 ttggatgaag aagctccaaa attgattggt ttgagagctg attctcatga agaaattttg 2820
 ggtaggttgg ataaaattag agaaagattt ttgcaaccaa ctgggtgctgc tccaagagaa 2880
 tctgaattga aagctcaagc tagaagaatt ttcttggaat tgttggtgga aactttggct 2940
 caagatgctg cttcttctgg ttctcaaaaa ccattggctt tgtctttggt ttctactcca 3000
 tctaaattgc aaagagaagt tgaattggct gctaaaggta ttccaagatg tttgaaaatg 3060
 agaagagatt ggtcttctcc agctggttca agatatgctc cagaaccatt ggcttctgat 3120
 agagttgctt tcatgtacgg tgaaggaagg tctccatact atggaatcac tcaagatatt 3180
 catagaattt ggccagaatt gcatgaagtt attaacgaaa aaactaatag gttgtgggct 3240
 gaaggtgata gatgggttat gccaaagagct tcttttaaata ctgaattgga atctcaacaa 3300
 caagaatttg atagaaatat gattgaaatg tttaggttgg gtattttgac ttctattgct 3360
 tttactaatt tggctagaga tgttttgaat attactccaa aagctgcttt tgggttgtct 3420

ttgggtgaaa	tttctatgat	ttttgctttt	tctaaaaaaa	atggtttgat	ttctgatcaa	3480
ttgactaaag	atttgagaga	atctgatgtt	tggaacaaag	cattggctgt	tgaattcaat	3540
gctttgagag	aagcatgggg	tattccacaa	tctgttccaa	aagatgaatt	ttggcaaggt	3600
tatattgtta	gaggtactaa	acaagatatt	gaagctgcta	ttgctccaga	ttccaaatac	3660
gttaggttga	ctatcattaa	tgatgctaata	actgctttga	tttctggtaa	accagatgct	3720
tgtaaagctg	ctattgctag	gttgggtggt	aatattccag	ctttgccagt	tactcaagga	3780
atgtgtggtc	attgtccaga	agttgggtcca	tatactaaag	atattgctaa	aattcatgct	3840
aatttggaat	ttccagttgt	tgatggtttg	gatttgtgga	ctactattaa	tcaaaaaaga	3900
ttggttccaa	gagctactgg	tgctaaagat	gaatgggctc	catcttcttt	tggtgaatat	3960
gctgggtcaac	tttacgaaaa	acaagctaata	tttccacaaa	ttggtgaaac	tatctacaaa	4020
caaaattatg	atgtttttgt	tgaggttggt	ccaaacaacc	ataggtctac	tgctgttaga	4080
actactttgg	gtccacaaaag	aaatcatttg	gctgggtgcta	ttgataaaca	aaacgaagat	4140
gcttggtacta	ctattgttaa	attgggtgct	tctttgaaag	ctcatttggt	tccagggtgtt	4200
actattttctc	cattgtatca	ttctaaattg	gttgctgaag	ctgaagcatg	ttatgctgct	4260
ctgtgcaaag	gagaaaaacc	taagaagaac	aaatttggtta	gaaaaattca	attgaatggt	4320
aggttcaatt	ctaaagctga	tccaatttct	tctgctgatt	tggttctttt	tccaccagct	4380
gatccagcta	ttgaagctgc	tatttcttca	agaattatga	aaccagttgc	tccaaaattt	4440
tatgctaggt	tgaatattga	tgaacaagac	gaaacaagag	atccaatttt	gaacaaagat	4500
aatgctccat	ctagttcatc	tagttcctct	tcatctagtt	cttcatctag	ttctccatct	4560
ccagctcctt	ctgctccagt	tcaaaaaaaaa	gctgctccag	ctgctgaaac	taaagctggt	4620
gcttctgctg	atgctttgag	atctgctttg	ttggatttggt	attctatggt	ggctttgtct	4680
tctgcttctg	cttctggtaa	tttggttgaa	actgctccat	ctgatgcttc	tgttattggt	4740
ccaccatgta	atattgctga	tttgggttca	agagctttta	tgaaaactta	tggtgtttct	4800
gctccattgt	acactggtgc	tatggctaaa	ggtattgctt	ctgctgattt	ggttattgct	4860
gctggtagac	aaggcatttt	ggcttctttt	ggtgctgggtg	gtttgccaat	gcaagttggt	4920
agagaatcta	ttgaaaaaat	tcaagctgct	ttgccaaatg	gtccatatgc	tgtaattttg	4980
attcattctc	catttgattc	taatttgga	aaaggtaatg	ttgatttggt	tttggaaaaa	5040
ggtgttactt	ttgttgaaagc	atctgctttt	atgactttga	ctccacaagt	tgtaggttac	5100
agagctgctg	gtttgactag	aaatgctgat	ggttctgtta	atattagaaa	tagaattatc	5160
ggaaagggtt	caagaactga	attggctgaa	atgtttatga	gacctgccc	agaacacttg	5220
ttgcaaaaaat	tgattgcttc	tggtgaaatt	aatcaagaac	aagctgaatt	ggctagaaga	5280
gttccagttg	ctgatgatat	tgctgttgaa	gctgattctg	gtggtcatac	tgataataga	5340

```

ccaattcatg ttatcttgcc attgattatt aatttgagag acagattgca tagagaatgt 5400
ggttatccag ctaatttgag agttagagtt ggtgctggtg gtggtattgg ttgtccacaa 5460
gctgctttgg ctacttttaa tatgggtgct tctttcattg ttactggcac tgттаатcaa 5520
gttgctaaac aatctggtac ttgtgataat gttagaaaac aattggctaa agctacttat 5580
tctgatgttt gcatggctcc agctgctgat atgtttgaag aagggtgtaa attgcaagtt 5640
ttgaagaaag ggacaatgtt tccatcaaga gctaataagt tatacgaatt gttttgcaag 5700
tatgattctt ttgaatctat gccaccagct gaattggcta gagttgaaaa aagaattttc 5760
tcaagagctt tggaagaagt ttgggatgaa actaaaaatt ttacattaa taggttgcac 5820
aatccagaaa aaattcaaag agctgaaaga gatccaaaat tgaaaatgtc tttgtgtttt 5880
agatggtatt tgtctttggc ttcaagatgg gctaatactg gtgcttctga tagagttatg 5940
gattatcaag tttggtgtgg tccagctatt ggttctttta atgatttcat taaaggcacc 6000
tacttgacc cagctgttgc taacgaatat ccatgcgttg ttcaaattaa caaacaatt 6060
ttgagaggtg cttgtttcct cagaagattg gaaattttga gaaatgctag gttgtctgat 6120
ggtgctgctg ctttggttgc ttctattgat gatacttatg ttccagctga aaaattgtaa 6180

```

<210> 37

5 <211> 6180

<212> ADN

<213> Secuencia artificial

10

<220>

<223> sintética

15 <400> 37

ES 2 467 918 T3

atggccgctc gcaacgtgtc tgcagcgcat gagatgcacg atgaaaagcg catcgccgtc	60
gtcggcatgg ccgtccagta cgccggatgc aaaaccaagg acgagttctg ggaggtgctc	120
atgaacggca aggtcgagtc caaggtgatc agcgacaaac gactcggctc caactaccgc	180
gccgagcact acaaagcaga gcgcagcaag tatgccgaca ctttttgcaa cgaaacgtac	240
ggcacccttg acgagaacga gatcgacaac gagcacgaac tcctcctcaa cctcgccaag	300
caggcactcg cagagacatc cgtcaaagac tcgacacgct gcggcatcgt cagcggctgc	360
ctctcgttcc ccatggacaa cctccagggg gaactcctca acgtgtacca aaaccatgtc	420
gagaaaaagc tcggggcccg cgtcttcaag gacgcctccc attggtccga acgcgagcag	480
tccaacaaac ccgaggccgg tgaccgccgc atcttcatgg acccggcctc cttcgtcgcc	540
gaagaactca acctcggcgc ctttactac tccgtcgacg cagcatgcgc cacggcgctc	600
tacgtgctcc gcctcgcgca ggatcatctc gtctccggcg ccgccgacgt catgctctgc	660
ggtgccacct gcctgccgga gccctttttc atccttttcgg gcttttccac cttccaggcc	720
atgcccgtcg gcacgggcca gaacgtgtcc atgccgctgc acaaggacag ccagggcctc	780

accccggggtg	agggcgggctc	catcatgggtc	ctcaagcggtc	tcgatgatgc	catccgcgac	840
ggcgaccaca	tctacggcac	ccttctcggc	gccaatgtca	gcaactccgg	cacaggtctg	900
cccctcaagc	cccttctccc	cagcgagaaa	aagtgcctca	tggacaccta	cacgcgcatt	960
aacgtgcacc	cgcacaagat	tcagtacgtc	gagtgccacg	ccaccggcac	gccccaggg	1020
gatcgtgtgg	aaatcgacgc	cgtcaaggcc	tgctttgaag	gcaagggtccc	ccgtttcgg	1080
accacaaagg	gcaactttgg	acacaccctc	gtcgcagccg	gctttgccgg	tatgtgcaag	1140
gtcctcctct	ccatgaagca	tggcatcatc	ccgcccaccc	cggtatcga	tgacgagacc	1200
aagatggacc	ctctcgtcgt	ctccgggtgag	gccatcccat	ggccagagac	caacggcgag	1260
cccaagcgcg	ccggtctctc	ggcctttggc	tttgggtggca	ccaacgccc	tgccgtcttt	1320
gaggagcatg	acccctccaa	cgccgcctgc	acggggccacg	actccatttc	tgcgctctcg	1380
gcccgtcgcg	gcggtgaaa	caacatgcgc	atcgccatca	ctggtatgga	cgccaccttt	1440
ggcgctctca	agggactcga	cgccttcgag	cgcgccattt	acaccggcg	tcacggtgcc	1500
atcccactcc	cagaaaagcg	ctggcgcttt	ctcggaagg	acaaggactt	tcttgacctc	1560
tgcggcgtca	aggccacccc	gcacggctgc	tacattgaag	atgttgaggt	cgacttccag	1620
cgctccgca	cgcccatgac	ccctgaagac	atgctcctcc	ctcagcagct	tctggccgtc	1680
accaccattg	accgcgccat	cctcgactcg	ggaatgaaaa	agggtggcaa	tgctgcgcgtc	1740
tttgtcggcc	tcggcaccga	cctcgagctc	taccgtcacc	gtgctcgcgt	cgctctcaag	1800
gagcgcgctc	gccctgaagc	ctccaagaag	ctcaatgaca	tgatgcagta	cattaacgac	1860
tgcggcacat	ccacatcgta	cacctcgta	attggcaacc	tcgtcgccac	gcgcgtctcg	1920
tcgcagtggg	gcttcacggg	cccctccttt	acgatcaccg	agggaacaa	ctccgtctac	1980
cgctgcgcgc	agctcggcaa	gtacctcctc	gagaccggcg	aggtcgatgg	cgctcgtcgt	2040
gcgggtgtcg	atctctgcgg	cagtgccgaa	aacctttacg	tcaagtctcg	ccgcttcaag	2100
gtgtccacct	ccgatacccc	gcgcgccagc	tttgacgcgc	ccgccgatgg	ctactttgtc	2160
ggcgagggct	gcgggtgcctt	tgtgctcaag	cgtgagacta	gctgcaccaa	ggacgaccgt	2220
atctacgctt	gcatggatgc	catcgtccct	ggcaacgtcc	ctagcgcttg	cttgcgcgag	2280
gccctcgacc	aggcgcgcg	caagccgggc	gatatcgaga	tgctcgagct	cagcgccgac	2340
tccgcccgc	acctcaagga	cccgtccgtc	ctgcccagg	agctcactgc	cgaggaggaa	2400
atcgggcgcc	ttcagacgat	ccttcgtgac	gatgacaagc	tcccgcgcaa	cgctcgcaacg	2460
ggcagtgtca	aggccaccgt	cggtgacacc	ggttatgcct	ctggtgctgc	cagcctcatc	2520
aaggctgcgc	tttgcatcta	caaccgctac	ctgcccagca	acggcgacga	ctgggatgaa	2580
cccggccctg	aggcgccctg	ggacagcacc	ctctttgctg	gccagacctc	gcgcgcttgg	2640
ctcaagaacc	ctggcgagcg	tcgctatgcg	gccgtctcgg	gcgtctccga	gacgcgctcg	2700

tgctattccg	tgctcctctc	cgaagccgag	ggccactacg	agcgcgagaa	ccgcatctcg	2760
ctcgacgagg	aggcgcccaa	gctcattgtg	cttcgcgcgc	actcccacga	ggagatcctt	2820
ggtcgcctcg	acaagatccg	cgagcgcttc	ttgcagccca	cgggcgcgcg	cccgcgcgag	2880
tccgagctca	aggcgccagg	ccgcgcgcatc	ttcctcgagc	tcctcggcga	gacccttgcc	2940
caggatgccg	cttcttcagg	ctcgcaaaaag	cccctcgctc	tcagcctcgt	ctccacgccc	3000
tccaagctcc	agcgcgaggt	cgagctcgcg	gccaagggtta	tcccgcgctg	cctcaagatg	3060
cgcgcgcgatt	ggagctcccc	tgctggcagc	cgctacgcgc	ctgagccgct	cgccagcgac	3120
cgcgtcgcct	tcatgtacgg	cgaaggctcg	agcccttact	acggcatcac	ccaagacatt	3180
caccgcattt	ggccccgaact	ccacgaggctc	atcaacgaaa	agacgaaccg	tctctggggc	3240
gaaggcgacc	gctgggtcat	gccgcgcgcc	agcttcaagt	cggagctcga	gagccagcag	3300
caagagtttg	atcgcaacat	gattgaaatg	ttccgtcttg	gaatcctcac	ctcaattgcc	3360
ttcaccaatc	tggcgcgcga	cgttctcaac	atcacgcccc	aggccgcctt	tggcctcagt	3420
cttggcgaga	tttccatgat	ttttgccttt	tccaagaaga	acggtctcat	ctccgaccag	3480
ctcaccaagg	atcttcgcga	gtccgacgtg	tggacaagg	ctctggccgt	tgaatttaat	3540
gcgctgcgcg	aggcctgggg	cattccacag	agtgtcccca	aggacgagtt	ctggcaaggc	3600
tacattgtgc	gcggcaccaa	gcaggatata	gaggcggcca	tcgccccgga	cagcaagtac	3660
gtgcgcctca	ccatcatcaa	tgatgccaac	accgcccctca	ttagcggcaa	gcccgcgcgc	3720
tgcaaggctg	cgatcgcgcg	tctcgggtgg	aacattcctg	cgttcccgt	gaccagggc	3780
atgtgcggcc	actgccccga	ggtgggacct	tataccaagg	atatcgccaa	gatccatgcc	3840
aaccttgagt	tccccgttgt	cgacggcctt	gacctctgga	ccacaatcaa	ccagaagcgc	3900
ctcgtgccac	gcgccacggg	cgccaaggac	gaatggggcc	cttcttcctt	tggcgagtac	3960
gccggccagc	tctacgagaa	gcaggctaac	ttccccaaa	tcgtcgagac	catttacaag	4020
caaaactacg	acgtctttgt	cgaggttggg	cccaacaacc	accgtagcac	cgcagtgcgc	4080
accacgcttg	gtccccagcg	caaccacctt	gctggcgcca	tcgacaagca	gaacgaggat	4140
gcttgagcga	ccatcgtcaa	gcttgtggct	tcgtcaagg	cccaccttgt	tcctggcgtc	4200
acgatctcgc	cgctgtacca	ctccaagctt	gtggcggagg	ctgaggcttg	ctacgctgcg	4260
ctctgcaagg	gtgaaaagcc	caagaagaac	aagtttgtgc	gcaagattca	gctcaacggg	4320
cgttcaaca	gcaaggcggg	ccccatctcc	tcggccgata	ttgccagctt	tccgcctgcg	4380
gaccctgcc	ttgaagccgc	catctcgagc	cgcatacatga	agccggttgc	tccgaagttc	4440
tacgcgcgtc	tcaacattga	cgagcaggac	gagaccgctg	atccgatcct	caacaaggac	4500
aacgcgcgct	cttccagctc	tagctcctct	tccagctctt	ccagctcttc	cagcccgtcg	4560
ccagctccgt	ccgccccagt	gcaaaaagaag	gctgctccgg	ccgcggagac	caaggctggt	4620

gcttcggctg acgcacttcg cagtgccctg ctcgatctcg acagtatgct tgcgctgagc 4680
tctgccagtg cctccggcaa ccttggttag actgcgccta gcgacgcctc ggtcattgtg 4740
ccgccctgca acattgcgga tctcggcagc cgcgccttca tgaaaacgta cgggtgtttcg 4800
gcgcctctgt acacggggcg catggccaag ggcattgcct ctgcggacct cgtcattgcc 4860
gccggccgcc agggcattct tgcgtccttt ggcgccggcg gacttcccat gcaggttgtg 4920
cgtgagtcca tcgaaaagat tcaggccgcc ctgcccaatg gcccgtagcg tgtcaacctt 4980
atccattctc cctttgacag caacctcgaa aagggaatg tcgatctctt cctcgagaag 5040
ggtgtcacct ttgtcgaggc ctcggccttt atgacgctca ccccgaggt cgtgcggtac 5100
cgcgcggtg gcctcacgca caacgccgac ggctcggtca acatccgcaa ccgtatcatt 5160
ggcaaggctc cgcgcaccga gctcgccgag atgttcatgc gtctcgcc cgagcacctt 5220
cttcagaagc tcattgcttc cggcgagatc aaccaggagc aggccgagct cgcccgccgt 5280
gttcccgtcg ctgacgacat cgcggtcgaa gctgactcgg gtggccacac cgacaaccgc 5340
cccatccacg tcattctgcc cctcatcatc aaccttcgcg accgccttca ccgcgagtgc 5400
ggctaccgga ccaaccttcg cgtccgtgtg ggcgccggcg gtggcattgg gtgccccag 5460
gcggcgctgg ccaccttcaa catgggtgcc tcctttattg tcaccggcac cgtgaaccag 5520
gtcgccaagc agtcgggcac gtgcgacaat gtgcgcaagc agtcgcgaa ggccacttac 5580
tcggacgtat gcattggccc ggctgccgac atgttcgagg aaggcgtcaa gcttcaggtc 5640
ctcaagaagg gaaccatgtt tccctcgcg gcacaacaagc tctacgagct cttttgcaag 5700
tacgactcgt tcgagtccat gcccccgca gagcttgcg gcgtcgagaa gcgcatcttc 5760
agccgcgcg tcgaagaggt ctgggacgag accaaaaact ttacattaa ccgtcttcac 5820
aaccgggaga agatccagcg cgcgagcg gcacccaagc tcaagatgtc gctgtgcttt 5880
cgctggtacc tgagcctggc gagccgtgg gccaaactg gagcttccga tcggtcatg 5940
gactaccagg tctggtgcgg tctgccatt ggttcttca acgatttcat caagggaact 6000
taccttgatc cggccgtcg aaacgagtag ccgtgcgtcg ttcagattaa caagcagatc 6060
cttcgtggag cgtgcttctt gcgcgctctc gaaattctgc gcaacgcacg cctttccgat 6120
ggcgctgccg ctcttggtggc cagcatcgat gacacatacg tcccgccga gaagctgtaa 6180

<210> 38

5 <211> 8436

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1) .. (8433)

<400> 38

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

ES 2 467 918 T3

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

ES 2 467 918 T3

aca cat aaa aat gtt gac aaa gtt tgt ggc tac aag ttt att gac gaa	1584
Thr His Lys Asn Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu	
515 520 525	
ttt cag ctc caa gga agc tgt cct cca gaa aat ccg aga gta gga ttt	1632
Phe Gln Leu Gln Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe	
530 535 540	
tta gca aca ctg cct act tca aat atc att gtc gcg ctt aag gca att	1680
Leu Ala Thr Leu Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile	
545 550 555 560	
ctc gcg cag ctt gat gca aaa cca gat gcg aag aaa tgg gat ttg cct	1728
Leu Ala Gln Leu Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro	
565 570 575	
cat aaa aag gct ttt ggg gct acc ttc gca tcg tct tca gtg aaa ggc	1776
His Lys Lys Ala Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly	
580 585 590	
tct gtt gct gcg ctc ttc gca gga cag ggt acc cag tac tta aac atg	1824
Ser Val Ala Ala Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met	
595 600 605	
ttc tct gat gtg gca atg aac tgg cca ccg ttc cgt gac agc att gtc	1872
Phe Ser Asp Val Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val	
610 615 620	
gca atg gaa gaa gct caa act gag gta ttt gag ggc caa gtt gaa cca	1920
Ala Met Glu Glu Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro	
625 630 635 640	
att agc aaa gtt ctg ttt cca cga gag cgc tat gca tcc gaa agt gaa	1968
Ile Ser Lys Val Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu	
645 650 655	
cag ggg aat gaa ctt ctt tgc tta aca gag tac tct cag cca act acg	2016
Gln Gly Asn Glu Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr	
660 665 670	
ata gca gcc gca gta ggg gcc ttc gat att ttc aaa gcg gct ggc ttt	2064
Ile Ala Ala Ala Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe	
675 680 685	
aag cca gac atg gtt gga ggg cat tca ctt ggc gaa ttt gct gct ttg	2112
Lys Pro Asp Met Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu	
690 695 700	
tac gcg gct ggg tcc att tcg cgt gac gac ctg tac aag ctt gtg tgc	2160
Tyr Ala Ala Gly Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys	
705 710 715 720	
aaa cgg gca aag gca atg gcg aac gct agt gac gga gct atg gca gca	2208
Lys Arg Ala Lys Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala	
725 730 735	
gtg att ggc cca gat gca cgt cta gtt acg cca caa aat agt gac gtt	2256
Val Ile Gly Pro Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val	
740 745 750	
tat gtc gca aac ttc aac tcc gca act caa gta gtc atc agt ggc act	2304
Tyr Val Ala Asn Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr	
755 760 765	

ES 2 467 918 T3

gtt	caa	ggt	gtg	aaa	gaa	gag	tcg	aaa	ttg	ctc	att	tca	aag	ggg	ttc	2352
Val	Gln	Gly	Val	Lys	Glu	Glu	Ser	Lys	Leu	Leu	Ile	Ser	Lys	Gly	Phe	
	770					775					780					
cgc	gta	ctg	cca	ctt	aaa	tgc	cag	ggc	gcc	ttc	cat	tct	cct	ttg	atg	2400
Arg	Val	Leu	Pro	Leu	Lys	Cys	Gln	Gly	Ala	Phe	His	Ser	Pro	Leu	Met	
785					790				795						800	
ggg	cct	tct	gag	gat	agt	ttc	aaa	tca	ctt	gtg	gag	act	tgt	acc	atc	2448
Gly	Pro	Ser	Glu	Asp	Ser	Phe	Lys	Ser	Leu	Val	Glu	Thr	Cys	Thr	Ile	
				805					810					815		
tcg	ccg	cca	aaa	aat	gtg	aaa	ttc	ttt	tgc	aat	gtt	agt	ggc	aag	gaa	2496
Ser	Pro	Pro	Lys	Asn	Val	Lys	Phe	Phe	Cys	Asn	Val	Ser	Gly	Lys	Glu	
			820					825					830			
agc	cca	aac	cca	aaa	cag	acc	ctc	aag	tca	cac	atg	acg	tct	agc	gtt	2544
Ser	Pro	Asn	Pro	Lys	Gln	Thr	Leu	Lys	Ser	His	Met	Thr	Ser	Ser	Val	
			835				840					845				
cag	ttc	gag	gag	cag	att	cgt	aac	atg	tac	gat	gcc	gga	gca	cgt	gtt	2592
Gln	Phe	Glu	Glu	Gln	Ile	Arg	Asn	Met	Tyr	Asp	Ala	Gly	Ala	Arg	Val	
	850					855					860					
ttt	ctg	gag	ttt	gga	ccc	cgc	caa	gtc	ctt	gca	aag	ctt	atc	gcg	gaa	2640
Phe	Leu	Glu	Phe	Gly	Pro	Arg	Gln	Val	Leu	Ala	Lys	Leu	Ile	Ala	Glu	
865					870				875					880		
atg	ttt	ccc	tcg	tgt	aca	gct	atc	agc	gtt	aac	ccc	gcg	agc	agt	ggc	2688
Met	Phe	Pro	Ser	Cys	Thr	Ala	Ile	Ser	Val	Asn	Pro	Ala	Ser	Ser	Gly	
				885					890					895		
gac	agt	gac	gtg	caa	ctc	cgc	ctc	gcc	gcc	gta	aaa	ttc	gcg	gtc	tcg	2736
Asp	Ser	Asp	Val	Gln	Leu	Arg	Leu	Ala	Ala	Val	Lys	Phe	Ala	Val	Ser	
			900					905					910			
ggt	gca	gcc	ctt	agc	acc	ttt	gat	cca	tgg	gag	tat	cgc	aag	cca	caa	2784
Gly	Ala	Ala	Leu	Ser	Thr	Phe	Asp	Pro	Trp	Glu	Tyr	Arg	Lys	Pro	Gln	
		915					920					925				
gat	ctt	ctt	att	cga	aaa	cca	cga	aaa	act	gcc	ctt	gtt	cta	tca	gca	2832
Asp	Leu	Leu	Ile	Arg	Lys	Pro	Arg	Lys	Thr	Ala	Leu	Val	Leu	Ser	Ala	
	930					935					940					
gca	aca	tat	gtt	tcc	cca	aag	act	ctt	gca	gaa	cgt	aaa	aag	gct	atg	2880
Ala	Thr	Tyr	Val	Ser	Pro	Lys	Thr	Leu	Ala	Glu	Arg	Lys	Lys	Ala	Met	
945					950				955					960		
gaa	gat	atc	aag	cta	gta	tcc	att	aca	cca	aga	gat	agt	atg	gta	tca	2928
Glu	Asp	Ile	Lys	Leu	Val	Ser	Ile	Thr	Pro	Arg	Asp	Ser	Met	Val	Ser	
				965					970					975		
att	gga	aaa	atc	gcg	caa	gaa	gta	cgg	aca	gct	aaa	cag	cct	tta	gaa	2976
Ile	Gly	Lys	Ile	Ala	Gln	Glu	Val	Arg	Thr	Ala	Lys	Gln	Pro	Leu	Glu	
			980					985					990			
acc	gaa	att	cga	aga	ctc	aac	aaa	gaa	tta	gaa	cat	ctc	aag	aga	gag	3024
Thr	Glu	Ile	Arg	Arg	Leu	Asn	Lys	Glu	Leu	Glu	His	Leu	Lys	Arg	Glu	
		995					1000					1005				
cta	gca	gca	gcc	aaa	gcg	agt	gtc	aag	tct	gca	tca	aaa	agc	tct		3069
Leu	Ala	Ala	Ala	Lys	Ala	Ser	Val	Lys	Ser	Ala	Ser	Lys	Ser	Ser		
	1010					1015					1020					

ES 2 467 918 T3

aaa gag cga tct gtc cta tca aag cac cgc gct ttg ctt caa aac	3114
Lys Glu Arg Ser Val Leu Ser Lys His Arg Ala Leu Leu Gln Asn	
1025 1030 1035	
att ttg caa gac tac gat gat ctt cgt gtg gtg cca ttc gct gtt	3159
Ile Leu Gln Asp Tyr Asp Asp Leu Arg Val Val Pro Phe Ala Val	
1040 1045 1050	
cgt tct gtt gca gtg gac aac acc gcg ccg tat gct gac caa gtt	3204
Arg Ser Val Ala Val Asp Asn Thr Ala Pro Tyr Ala Asp Gln Val	
1055 1060 1065	
tcg acc cca gcg tca gag cgg tcg gct tca ccg ctt ttc gag aaa	3249
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys	
1070 1075 1080	
cgc agt tcg gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg	3294
Arg Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala	
1085 1090 1095	
gta ctg agc gtt ctc gca gac aag aca ggc tac gac agc tca atg	3339
Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met	
1100 1105 1110	
atc gag atg gac atg gac ctg gag agt gag ctt ggc gtt gat agc	3384
Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser	
1115 1120 1125	
atc aaa cgc gtg gag atc atg agc gag gtt caa acg ctg ctc agc	3429
Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser	
1130 1135 1140	
gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc aag act gtt	3474
Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val	
1145 1150 1155	
ggc gac gtc atc gag gcg atg aag ctg gaa ctc ggt gga ccc caa	3519
Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln	
1160 1165 1170	
ggc cag act ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tcc	3564
Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser	
1175 1180 1185	
gag cct gct gta ccg acc tca tcg tca agc agt att gct aat gtt	3609
Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val	
1190 1195 1200	
tcg tca gca cgc ctc gct gaa gct gaa gct gcg gta ctg agc gtt	3654
Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Val Leu Ser Val	
1205 1210 1215	
ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg gac	3699
Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp	
1220 1225 1230	
atg gac ctg gag agc gag ctt ggc gtt gat agc atc aaa cgc gtg	3744
Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val	
1235 1240 1245	
gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc	3789
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser	
1250 1255 1260	

ES 2 467 918 T3

gac gtt gac gct ctg tca aga act aag act gtt ggc gac gtc atc	3834
Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile	
1265 1270 1275	
gag gcg atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg	3879
Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu	
1280 1285 1290	
acc gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta	3924
Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val	
1295 1300 1305	
ccg acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc	3969
Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg	
1310 1315 1320	
ctc gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag	4014
Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys	
1325 1330 1335	
aca ggc tac gac agc tca atg atc gag atg gac atg gac ctg gag	4059
Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu	
1340 1345 1350	
agc gag ctt ggc gtc gac agc atc aaa cgc gtg gag atc atg agc	4104
Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser	
1355 1360 1365	
gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac gtt gac gct	4149
Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala	
1370 1375 1380	
ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag	4194
Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys	
1385 1390 1395	
ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa tcg	4239
Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser	
1400 1405 1410	
atc cgt cag cca ccg gtg tcc gag cct gct gta ccg acc tca tcg	4284
Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser	
1415 1420 1425	
tca agc agt att gct aat gtt ttg tca gca cgc ctc gct gaa gct	4329
Ser Ser Ser Ile Ala Asn Val Leu Ser Ala Arg Leu Ala Glu Ala	
1430 1435 1440	
gaa gcc gcg gta ctg agc gtt ctc gca gac aag aca ggc tac gac	4374
Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp	
1445 1450 1455	
agc tca atg atc gag atg gac atg gac ctg gag agc gag ctt ggc	4419
Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly	
1460 1465 1470	
gtt gat agc atc aaa cgc gtg gag atc atg agc gag gtt caa acg	4464
Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr	
1475 1480 1485	
ttg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc	4509
Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr	
1490 1495 1500	

ES 2 467 918 T3

aag Lys	act Thr	gtt Val	ggc Gly	gac Asp	gtc Val	atc Ile	gag Glu	gcg Ala	atg Met	aag Lys	ctg Leu	gaa Glu	ctc Leu	ggt Gly	4554
1505						1510					1515				
gga Gly	ccc Pro	caa Gln	ggc Gly	cag Gln	act Thr	ttg Leu	acc Thr	gcg Ala	gaa Glu	tcg Ser	atc Ile	cgt Arg	cag Gln	cca Pro	4599
1520						1525					1530				
ccg Pro	gtg Val	tct Ser	gag Glu	cct Pro	gct Ala	gta Val	ccg Pro	acc Thr	tca Ser	tcg Ser	tca Ser	agc Ser	agt Ser	att Ile	4644
1535						1540					1545				
gct Ala	aat Asn	gtt Val	tcg Ser	tca Ser	gca Ala	cgc Arg	ctc Leu	gct Ala	gaa Glu	gct Ala	gaa Glu	gcc Ala	gcg Ala	gta Val	4689
1550						1555					1560				
ctg Leu	agc Ser	gtt Val	ctc Leu	gca Ala	gac Asp	aag Lys	aca Thr	ggc Gly	tac Tyr	gac Asp	agc Ser	tca Ser	atg Met	atc Ile	4734
1565						1570					1575				
gag Glu	atg Met	gac Asp	atg Met	gac Asp	ctg Leu	gag Glu	agt Ser	gag Glu	ctt Leu	ggc Gly	gtc Val	gac Asp	agc Ser	atc Ile	4779
1580						1585					1590				
aaa Lys	cgc Arg	gtg Val	gag Glu	atc Ile	atg Met	agc Ser	gag Glu	gtt Val	caa Gln	acg Thr	ctg Leu	ctc Leu	agc Ser	gtg Val	4824
1595						1600					1605				
gaa Glu	gtc Val	tcc Ser	gac Asp	gtt Val	gac Asp	gct Ala	ctg Leu	tca Ser	aga Arg	acc Thr	aag Lys	act Thr	gtt Val	ggc Gly	4869
1610						1615					1620				
gac Asp	gtc Val	atc Ile	gag Glu	gcg Ala	atg Met	aag Lys	ctg Leu	gaa Glu	ctc Leu	ggg Gly	gga Gly	ccc Pro	caa Gln	ggc Gly	4914
1625						1630					1635				
cag Gln	act Thr	ttg Leu	acc Thr	tct Ser	gaa Glu	ccg Pro	atc Ile	cat His	cag Gln	cca Pro	cca Pro	gtg Val	tcc Ser	gag Glu	4959
1640						1645					1650				
cct Pro	gct Ala	gta Val	ccg Pro	acc Thr	tca Ser	tcg Ser	tca Ser	agc Ser	agt Ser	att Ile	gct Ala	aat Asn	gtt Val	tct Ser	5004
1655						1660					1665				
tca Ser	gca Ala	cgc Arg	ctc Leu	gct Ala	gaa Glu	gct Ala	gaa Glu	gcc Ala	gcg Ala	gta Val	ctg Leu	agc Ser	gtt Val	ctc Leu	5049
1670						1675					1680				
gca Ala	gac Asp	aag Lys	aca Thr	ggc Gly	tac Tyr	gac Asp	agc Ser	tca Ser	atg Met	atc Ile	gag Glu	atg Met	gac Asp	atg Met	5094
1685						1690					1695				
gac Asp	ctg Leu	gag Glu	agc Ser	gag Glu	ctt Leu	ggc Gly	gtt Val	gat Asp	agc Ser	atc Ile	aaa Lys	cgc Arg	gtg Val	gaa Glu	5139
1700						1705					1710				
atc Ile	atg Met	agc Ser	gag Glu	gtt Val	caa Gln	acg Thr	ctg Leu	ctc Leu	agc Ser	gtg Val	gaa Glu	gtc Val	tcc Ser	gac Asp	5184
1715						1720					1725				
gtt Val	gac Asp	gct Ala	ctg Leu	tca Ser	aga Arg	acc Thr	aag Lys	act Thr	gtt Val	ggc Gly	gac Asp	gtc Val	atc Ile	gag Glu	5229
1730						1735					1740				

gcg atg aag atg gaa ctc ggt gga ccc caa ggc cag act ttg acc	5274
Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr	
1745 1750 1755	
gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta ccg	5319
Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro	
1760 1765 1770	
acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc ctc	5364
Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu	
1775 1780 1785	
gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag aca	5409
Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr	
1790 1795 1800	
ggc tac gac agc tca atg atc gag atg gac atg gac ctg gag agc	5454
Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser	
1805 1810 1815	
gag ctt ggc gtt gat agc atc aaa cgc gtg gag atc atg agc gag	5499
Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu	
1820 1825 1830	
gtt caa gcg ctg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg	5544
Val Gln Ala Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu	
1835 1840 1845	
tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag atg	5589
Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Met	
1850 1855 1860	
gaa ctc ggt gga ccc caa ggc cag act ttg acc gca gaa tcg atc	5634
Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile	
1865 1870 1875	
cgt gag cca ccg gtg tct gag cct gct gta ccg acc tca tcg tca	5679
Arg Glu Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser	
1880 1885 1890	
agt agt atc gct aat gtt tct tca gct cgc ctc gct gaa gct gaa	5724
Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu	
1895 1900 1905	
gcc gcg gta ctg agc gtt ctc gca gac aag aca ggc tac gac agc	5769
Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser	
1910 1915 1920	
tca atg atc gag atg gac atg gac ctg gag agt gag ctt ggc gtc	5814
Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val	
1925 1930 1935	
gac agc atc aaa cgc gtg gag atc atg agc gag gtt caa acg ttg	5859
Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu	
1940 1945 1950	
ctc agc gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc aag	5904
Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys	
1955 1960 1965	
act gtt ggc gac gtc atc gag gcg atg aag ctg gaa ctt ggg gaa	5949
Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Glu	
1970 1975 1980	

tca tca agt att gag act ctc aat tgt acc gag gtt gag cac acg Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu His Thr 1985 1990 1995	5994
agc tac aaa agt gtc aag gct tca ggg tgt gag aat gta gat acc Ser Tyr Lys Ser Val Lys Ala Ser Gly Cys Glu Asn Val Asp Thr 2000 2005 2010	6039
cgt ttc gct aag gtt gta caa atc tcg ctt cct agc aag ctg aaa Arg Phe Ala Lys Val Val Gln Ile Ser Leu Pro Ser Lys Leu Lys 2015 2020 2025	6084
tcc act gtg tcg cac gat cga cct gta att gtt gta gat gat gga Ser Thr Val Ser His Asp Arg Pro Val Ile Val Val Asp Asp Gly 2030 2035 2040	6129
acg ccc tta acc acg gag ctt tgt aaa att ctt ggg ggt aat att Thr Pro Leu Thr Thr Glu Leu Cys Lys Ile Leu Gly Gly Asn Ile 2045 2050 2055	6174
gtg gtt ctc tct tat caa ggg aag ccc gct ggt cca cgg gga gtc Val Val Leu Ser Tyr Gln Gly Lys Pro Ala Gly Pro Arg Gly Val 2060 2065 2070	6219
gag gtg cca gat ctt tcc gag gaa gcc cta att caa gct ctt gca Glu Val Pro Asp Leu Ser Glu Glu Ala Leu Ile Gln Ala Leu Ala 2075 2080 2085	6264
ttg att cgg tct aca tat gga gtt cca att ggt ttt att tgt cag Leu Ile Arg Ser Thr Tyr Gly Val Pro Ile Gly Phe Ile Cys Gln 2090 2095 2100	6309
caa gtg tct aat gtg agc acc aag gca cag ctt tgt tgg gca ctc Gln Val Ser Asn Val Ser Thr Lys Ala Gln Leu Cys Trp Ala Leu 2105 2110 2115	6354
ctc gca gcg aag cat ctc aag aag gat ttg aat gct gtc tta ccc Leu Ala Ala Lys His Leu Lys Lys Asp Leu Asn Ala Val Leu Pro 2120 2125 2130	6399
gat tca aga tcc ttc ttc gtc gga gtt gta cgc ttg aac ggg aaa Asp Ser Arg Ser Phe Phe Val Gly Val Val Arg Leu Asn Gly Lys 2135 2140 2145	6444
ctt gga act ttc gaa aac atc agc gac ttc tct aaa ttt gat ttg Leu Gly Thr Phe Glu Asn Ile Ser Asp Phe Ser Lys Phe Asp Leu 2150 2155 2160	6489
acg aaa gcc cta gat tac gga cag cgt ggt tct ctc tta ggc ctg Thr Lys Ala Leu Asp Tyr Gly Gln Arg Gly Ser Leu Leu Gly Leu 2165 2170 2175	6534
tgc aag tca cta gac tta gaa tgg gaa cag gtg ttt tgc cgt gga Cys Lys Ser Leu Asp Leu Glu Trp Glu Gln Val Phe Cys Arg Gly 2180 2185 2190	6579
ata gat ctt gcg tgt gat ctt atg cca ctc cag gcc gca agg ata Ile Asp Leu Ala Cys Asp Leu Met Pro Leu Gln Ala Ala Arg Ile 2195 2200 2205	6624
ctc aga aat gag ctt cag tgt ccc aat atg cgc ctt cgc gag gtt Leu Arg Asn Glu Leu Gln Cys Pro Asn Met Arg Leu Arg Glu Val 2210 2215 2220	6669

ggg Gly	tac Tyr	gat Asp	att Ile	tct Ser	ggc Gly	gcc Ala	agg Arg	tac Tyr	acc Thr	att Ile	tca Ser	acc Thr	gat Asp	gac Asp	6714
2225						2230					2235				
ctg Leu	cta Leu	tgt Cys	gga Gly	ccc Pro	tcg Ser	aag Lys	gct Ala	aaa Lys	gta Val	gag Glu	gcc Ala	gca Ala	gac Asp	ttg Leu	6759
2240						2245					2250				
ttt Phe	ctt Leu	gtg Val	aca Thr	ggg Gly	ggc Gly	gca Ala	cga Arg	ggg Gly	att Ile	aca Thr	cct Pro	cat His	tgt Cys	gtt Val	6804
2255						2260					2265				
cgt Arg	gag Glu	att Ile	gca Ala	agt Ser	cga Arg	tcc Ser	ccc Pro	gga Gly	acc Thr	aca Thr	ttt Phe	gtg Val	ctg Leu	gtt Val	6849
2270						2275					2280				
gga Gly	aga Arg	agc Ser	gaa Glu	atg Met	tcc Ser	gac Asp	gag Glu	cct Pro	gac Asp	tgg Trp	gct Ala	gtt Val	ggc Gly	cac His	6894
2285						2290					2295				
tac Tyr	aat Asn	aaa Lys	gac Asp	ctg Leu	gac Asp	caa Gln	agc Ser	aca Thr	atg Met	aaa Lys	cac His	ttg Leu	aaa Lys	gca Ala	6939
2300						2305					2310				
acg Thr	cat His	gct Ala	gct Ala	gga Gly	ggg Gly	gta Val	aaa Lys	cct Pro	acg Thr	cct Pro	aaa Lys	gca Ala	cat His	cgt Arg	6984
2315						2320					2325				
gca Ala	ctt Leu	gtg Val	aac Asn	agg Arg	gtc Val	act Thr	ggc Gly	tca Ser	cgg Arg	gag Glu	gta Val	cga Arg	gaa Glu	tct Ser	7029
2330						2335					2340				
ctt Leu	aga Arg	gca Ala	atc Ile	cag Gln	gag Glu	gca Ala	ggg Gly	gca Ala	aat Asn	gtc Val	gaa Glu	tat Tyr	atc Ile	gcc Ala	7074
2345						2350					2355				
tgt Cys	gat Asp	gtt Val	tcg Ser	gat Asp	gaa Glu	aac Asn	aag Lys	gtc Val	cgc Arg	caa Gln	ctt Leu	gtg Val	caa Gln	aga Arg	7119
2360						2365					2370				
gtg Val	gag Glu	caa Gln	aag Lys	tat Tyr	ggc Gly	tgt Cys	gaa Glu	ata Ile	act Thr	ggg Gly	att Ile	tgg Trp	cat His	gca Ala	7164
2375						2380					2385				
agc Ser	ggg Gly	gtt Val	ctt Leu	cgt Arg	gac Asp	aaa Lys	ctt Leu	gtc Val	gag Glu	caa Gln	aag Lys	act Thr	aca Thr	gac Asp	7209
2390						2395					2400				
gac Asp	ttt Phe	gag Glu	gca Ala	gtt Val	ttt Phe	ggg Gly	acc Thr	aag Lys	gtg Val	act Thr	ggc Gly	ctt Leu	gta Val	aac Asn	7254
2405						2410					2415				
atc Ile	gtg Val	tca Ser	caa Gln	gtc Val	aat Asn	atg Met	tct Ser	aag Lys	cta Leu	cga Arg	cac His	ttc Phe	atc Ile	ctc Leu	7299
2420						2425					2430				
ttc Phe	agt Ser	tct Ser	ttg Leu	gct Ala	gga Gly	ttt Phe	cat His	ggg Gly	aac Asn	aag Lys	ggc Gly	caa Gln	acg Thr	gat Asp	7344
2435						2440					2445				
tat Tyr	gca Ala	att Ile	gct Ala	aat Asn	gaa Glu	gcc Ala	ttg Leu	aac Asn	aaa Lys	atc Ile	gcg Ala	cat His	act Thr	ctc Leu	7389
2450						2455					2460				

ES 2 467 918 T3

tca gcg ttt ttg ccc aaa ctg aat gca aag gtg cta gac ttc ggt	7434
Ser Ala Phe Leu Pro Lys Leu Asn Ala Lys Val Leu Asp Phe Gly	
2465 2470 2475	
ccg tgg gta ggt tca gga atg gta acc gaa aca ctt gag aag cat	7479
Pro Trp Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His	
2480 2485 2490	
ttt aaa gct atg ggg gtt cag act att cct ctc gag cca gga gca	7524
Phe Lys Ala Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala	
2495 2500 2505	
cgg act gtt gcg caa atc att ttg gca agt tcg cca ccg caa tcg	7569
Arg Thr Val Ala Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser	
2510 2515 2520	
ctt ttg ggg aac tgg ggc ttt cca gcc acc aaa ccg cta caa cgc	7614
Leu Leu Gly Asn Trp Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg	
2525 2530 2535	
tct aat gta gtc acg ggc aca ctc tct ccg gaa gag ata gaa ttc	7659
Ser Asn Val Val Thr Gly Thr Leu Ser Pro Glu Glu Ile Glu Phe	
2540 2545 2550	
atc gca gac cac aaa att caa ggc cgc aag gtg ctt ccc atg atg	7704
Ile Ala Asp His Lys Ile Gln Gly Arg Lys Val Leu Pro Met Met	
2555 2560 2565	
gct gca atc ggg ttc atg gcc tct att gcg gaa gga ctc tac ccg	7749
Ala Ala Ile Gly Phe Met Ala Ser Ile Ala Glu Gly Leu Tyr Pro	
2570 2575 2580	
ggg tac aat ctg caa ggc gtg gaa aat gct cag ctc ttt caa ggc	7794
Gly Tyr Asn Leu Gln Gly Val Glu Asn Ala Gln Leu Phe Gln Gly	
2585 2590 2595	
ttg act atc aac caa gag aca aaa ttt caa atc act ctc att gag	7839
Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln Ile Thr Leu Ile Glu	
2600 2605 2610	
gag cac aac tct gag gaa aac ctg gat gtc ctg aca tcc ctt ggt	7884
Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu Thr Ser Leu Gly	
2615 2620 2625	
gta atg ttg gaa agc ggg aag gtg ctt ccc gct tac cga tgt gtt	7929
Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr Arg Cys Val	
2630 2635 2640	
gta tgc ttg aat aca acc cag cag cag ccc aag cta tct cca aaa	7974
Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser Pro Lys	
2645 2650 2655	
att ctt aac ttg gaa gtt gac cct gca tgc gag gtt aac ccc tat	8019
Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro Tyr	
2660 2665 2670	
gat gga aag tcg ttg ttc cac ggt ccg ctt ttg caa ttc gtt caa	8064
Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln	
2675 2680 2685	
caa gtg ttg cac tca agt acc aaa ggc ctc gtt gcc aag tgc cgc	8109
Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg	
2690 2695 2700	

ES 2 467 918 T3

gcg ctt cca atc aaa gaa gcc atc cga ggg cca ttt atc aag caa 8154
Ala Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln
2705 2710 2715

aca ctc cat gat cca att cta gac gac gtc att ttt cag cta atg 8199
Thr Leu His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met
2720 2725 2730

ctc gtg tgg tgt cgt aat gct cta gga agt gca tcg cta ccc aac 8244
Leu Val Trp Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn
2735 2740 2745

aga att gaa aag atg tca tac ttt ggg aat gtc tca gaa ggt agc 8289
Arg Ile Glu Lys Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser
2750 2755 2760

act ttc ttt gcc tca gtt aca cct gtg gga cca aga gta cca aag 8334
Thr Phe Phe Ala Ser Val Thr Pro Val Gly Pro Arg Val Pro Lys
2765 2770 2775

gat ccc gtg atc aaa atg cag ttt ctt ctc caa gat gaa tcc ggc 8379
Asp Pro Val Ile Lys Met Gln Phe Leu Leu Gln Asp Glu Ser Gly
2780 2785 2790

aac aca ttt tca tcg ggg gag ggc tcg gtt gtg ctt agt gac gaa 8424
Asn Thr Phe Ser Ser Gly Glu Gly Ser Val Val Leu Ser Asp Glu
2795 2800 2805

ctc gtc ttt tga 8436
Leu Val Phe
2810

<210> 39

5 <211> 2811

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 39

Met Lys Asp Met Glu Asp Arg Arg Val Ala Ile Val Gly Met Ser Ala
1 5 10 15

His Leu Pro Cys Gly Thr Asp Val Lys Glu Ser Trp Gln Ala Ile Arg
20 25 30

Asp Gly Ile Asp Cys Leu Ser Asp Leu Pro Ala Asp Arg Leu Asp Val
35 40 45

Thr Ala Tyr Tyr Asn Pro Asn Lys Ala Thr Lys Asp Lys Ile Tyr Cys
50 55 60

Lys Arg Gly Gly Phe Ile Pro Asn Tyr Asp Phe Asp Pro Arg Glu Phe
65 70 75 80

Gly Leu Asn Met Phe Gln Met Glu Asp Ser Asp Ala Asn Gln Thr Leu
85 90 95

ES 2 467 918 T3

Thr Leu Leu Lys Val Lys Gln Ala Leu Glu Asp Ala Ser Ile Glu Pro
 100 105 110

Phe Thr Lys Glu Lys Lys Asn Ile Gly Cys Val Leu Gly Ile Gly Gly
 115 120 125

Gly Gln Lys Ala Ser His Glu Phe Tyr Ser Arg Leu Asn Tyr Val Val
 130 135 140

Val Glu Lys Val Leu Arg Lys Met Gly Leu Pro Asp Ala Asp Val Glu
 145 150 155 160

Glu Ala Val Glu Lys Tyr Lys Ala Asn Phe Pro Glu Trp Arg Leu Asp
 165 170 175

Ser Phe Pro Gly Phe Leu Gly Asn Val Thr Ala Gly Arg Cys Ser Asn
 180 185 190

Thr Phe Asn Met Glu Gly Met Asn Cys Val Val Asp Ala Ala Cys Ala
 195 200 205

Ser Ser Leu Ile Ala Ile Lys Val Ala Val Glu Glu Leu Leu Phe Gly
 210 215 220

Asp Cys Asp Thr Met Ile Ala Gly Ala Thr Cys Thr Asp Asn Ser Leu
 225 230 235 240

Gly Met Tyr Met Ala Phe Ser Lys Thr Pro Val Phe Ser Thr Asp Pro
 245 250 255

Ser Val Arg Ala Tyr Asp Glu Lys Thr Lys Gly Met Leu Ile Gly Glu
 260 265 270

Gly Ser Ala Met Phe Val Leu Lys Arg Tyr Ala Asp Ala Val Arg Asp
 275 280 285

Gly Asp Thr Ile His Ala Val Leu Arg Ser Cys Ser Ser Ser Ser Asp
 290 295 300

Gly Lys Ala Ala Gly Ile Tyr Thr Pro Thr Ile Ser Gly Gln Glu Glu
 305 310 315 320

Ala Leu Arg Arg Ala Tyr Ala Arg Ala Gly Val Cys Pro Ser Thr Ile
 325 330 335

Gly Leu Val Glu Gly His Gly Thr Gly Thr Pro Val Gly Asp Arg Ile
 340 345 350

ES 2 467 918 T3

Glu Leu Thr Ala Leu Arg Asn Leu Phe Asp Lys Ala Phe Gly Ser Lys
355 360 365

Lys Glu Gln Ile Ala Val Gly Ser Ile Lys Ser Gln Ile Gly His Leu
370 375 380

Lys Ser Val Ala Gly Phe Ala Gly Leu Val Lys Ala Val Leu Ala Leu
385 390 395 400

Lys His Lys Thr Leu Pro Gly Ser Ile Asn Val Asp Gln Pro Pro Leu
405 410 415

Leu Tyr Asp Gly Thr Gln Ile Gln Asp Ser Ser Leu Tyr Ile Asn Lys
420 425 430

Thr Asn Arg Pro Trp Phe Thr Gln Asn Lys Leu Pro Arg Arg Ala Gly
435 440 445

Val Ser Ser Phe Gly Phe Gly Gly Ala Asn Tyr His Ala Val Leu Glu
450 455 460

Glu Phe Glu Pro Glu His Glu Lys Pro Tyr Arg Leu Asn Thr Val Gly
465 470 475 480

His Pro Val Leu Leu Tyr Ala Pro Ser Val Glu Ala Leu Lys Val Leu
485 490 495

Cys Asn Asp Gln Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys
500 505 510

Thr His Lys Asn Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu
515 520 525

Phe Gln Leu Gln Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe
530 535 540

Leu Ala Thr Leu Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile
545 550 555 560

Leu Ala Gln Leu Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro
565 570 575

His Lys Lys Ala Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly
580 585 590

Ser Val Ala Ala Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met
595 600 605

ES 2 467 918 T3

Phe Ser Asp Val Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val
610 615 620

Ala Met Glu Glu Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro
625 630 635 640

Ile Ser Lys Val Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu
645 650 655

Gln Gly Asn Glu Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr
660 665 670

Ile Ala Ala Ala Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe
675 680 685

Lys Pro Asp Met Val Gly Gly His Ser Leu Gly Glu Phe Ala Ala Leu
690 695 700

Tyr Ala Ala Gly Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys
705 710 715 720

Lys Arg Ala Lys Ala Met Ala Asn Ala Ser Asp Gly Ala Met Ala Ala
725 730 735

Val Ile Gly Pro Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val
740 745 750

Tyr Val Ala Asn Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr
755 760 765

Val Gln Gly Val Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe
770 775 780

Arg Val Leu Pro Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met
785 790 795 800

Gly Pro Ser Glu Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile
805 810 815

Ser Pro Pro Lys Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu
820 825 830

Ser Pro Asn Pro Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val
835 840 845

Gln Phe Glu Glu Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val
850 855 860

ES 2 467 918 T3

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

ES 2 467 918 T3

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

ES 2 467 918 T3

Ser	Glu	Leu	Gly	Val	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Met	Ser
1355						1360					1365			
Glu	Val	Gln	Thr	Leu	Leu	Ser	Val	Glu	Val	Ser	Asp	Val	Asp	Ala
1370						1375					1380			
Leu	Ser	Arg	Thr	Lys	Thr	Val	Gly	Asp	Val	Ile	Glu	Ala	Met	Lys
1385						1390					1395			
Leu	Glu	Leu	Gly	Gly	Pro	Gln	Gly	Gln	Thr	Leu	Thr	Ala	Glu	Ser
1400						1405					1410			
Ile	Arg	Gln	Pro	Pro	Val	Ser	Glu	Pro	Ala	Val	Pro	Thr	Ser	Ser
1415						1420					1425			
Ser	Ser	Ser	Ile	Ala	Asn	Val	Leu	Ser	Ala	Arg	Leu	Ala	Glu	Ala
1430						1435					1440			
Glu	Ala	Ala	Val	Leu	Ser	Val	Leu	Ala	Asp	Lys	Thr	Gly	Tyr	Asp
1445						1450					1455			
Ser	Ser	Met	Ile	Glu	Met	Asp	Met	Asp	Leu	Glu	Ser	Glu	Leu	Gly
1460						1465					1470			
Val	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Met	Ser	Glu	Val	Gln	Thr
1475						1480					1485			
Leu	Leu	Ser	Val	Glu	Val	Ser	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr
1490						1495					1500			
Lys	Thr	Val	Gly	Asp	Val	Ile	Glu	Ala	Met	Lys	Leu	Glu	Leu	Gly
1505						1510					1515			
Gly	Pro	Gln	Gly	Gln	Thr	Leu	Thr	Ala	Glu	Ser	Ile	Arg	Gln	Pro
1520						1525					1530			
Pro	Val	Ser	Glu	Pro	Ala	Val	Pro	Thr	Ser	Ser	Ser	Ser	Ser	Ile
1535						1540					1545			
Ala	Asn	Val	Ser	Ser	Ala	Arg	Leu	Ala	Glu	Ala	Glu	Ala	Ala	Val
1550						1555					1560			
Leu	Ser	Val	Leu	Ala	Asp	Lys	Thr	Gly	Tyr	Asp	Ser	Ser	Met	Ile
1565						1570					1575			
Glu	Met	Asp	Met	Asp	Leu	Glu	Ser	Glu	Leu	Gly	Val	Asp	Ser	Ile
1580						1585					1590			

ES 2 467 918 T3

Lys	Arg	Val	Glu	Ile	Met	Ser	Glu	Val	Gln	Thr	Leu	Leu	Ser	Val
1595						1600					1605			
Glu	Val	Ser	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Lys	Thr	Val	Gly
1610						1615					1620			
Asp	Val	Ile	Glu	Ala	Met	Lys	Leu	Glu	Leu	Gly	Gly	Pro	Gln	Gly
1625						1630					1635			
Gln	Thr	Leu	Thr	Ser	Glu	Pro	Ile	His	Gln	Pro	Pro	Val	Ser	Glu
1640						1645					1650			
Pro	Ala	Val	Pro	Thr	Ser	Ser	Ser	Ser	Ser	Ile	Ala	Asn	Val	Ser
1655						1660					1665			
Ser	Ala	Arg	Leu	Ala	Glu	Ala	Glu	Ala	Ala	Val	Leu	Ser	Val	Leu
1670						1675					1680			
Ala	Asp	Lys	Thr	Gly	Tyr	Asp	Ser	Ser	Met	Ile	Glu	Met	Asp	Met
1685						1690					1695			
Asp	Leu	Glu	Ser	Glu	Leu	Gly	Val	Asp	Ser	Ile	Lys	Arg	Val	Glu
1700						1705					1710			
Ile	Met	Ser	Glu	Val	Gln	Thr	Leu	Leu	Ser	Val	Glu	Val	Ser	Asp
1715						1720					1725			
Val	Asp	Ala	Leu	Ser	Arg	Thr	Lys	Thr	Val	Gly	Asp	Val	Ile	Glu
1730						1735					1740			
Ala	Met	Lys	Met	Glu	Leu	Gly	Gly	Pro	Gln	Gly	Gln	Thr	Leu	Thr
1745						1750					1755			
Ala	Glu	Ser	Ile	Arg	Gln	Pro	Pro	Val	Ser	Glu	Pro	Ala	Val	Pro
1760						1765					1770			
Thr	Ser	Ser	Ser	Ser	Ser	Ile	Ala	Asn	Val	Ser	Ser	Ala	Arg	Leu
1775						1780					1785			
Ala	Glu	Ala	Glu	Ala	Ala	Val	Leu	Ser	Val	Leu	Ala	Asp	Lys	Thr
1790						1795					1800			
Gly	Tyr	Asp	Ser	Ser	Met	Ile	Glu	Met	Asp	Met	Asp	Leu	Glu	Ser
1805						1810					1815			
Glu	Leu	Gly	Val	Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Met	Ser	Glu
1820						1825					1830			

ES 2 467 918 T3

Val	Gln	Ala	Leu	Leu	Ser	Val	Glu	Val	Ser	Asp	Val	Asp	Ala	Leu
1835						1840					1845			
Ser	Arg	Thr	Lys	Thr	Val	Gly	Asp	Val	Ile	Glu	Ala	Met	Lys	Met
1850						1855					1860			
Glu	Leu	Gly	Gly	Pro	Gln	Gly	Gln	Thr	Leu	Thr	Ala	Glu	Ser	Ile
1865						1870					1875			
Arg	Glu	Pro	Pro	Val	Ser	Glu	Pro	Ala	Val	Pro	Thr	Ser	Ser	Ser
1880						1885					1890			
Ser	Ser	Ile	Ala	Asn	Val	Ser	Ser	Ala	Arg	Leu	Ala	Glu	Ala	Glu
1895						1900					1905			
Ala	Ala	Val	Leu	Ser	Val	Leu	Ala	Asp	Lys	Thr	Gly	Tyr	Asp	Ser
1910						1915					1920			
Ser	Met	Ile	Glu	Met	Asp	Met	Asp	Leu	Glu	Ser	Glu	Leu	Gly	Val
1925						1930					1935			
Asp	Ser	Ile	Lys	Arg	Val	Glu	Ile	Met	Ser	Glu	Val	Gln	Thr	Leu
1940						1945					1950			
Leu	Ser	Val	Glu	Val	Ser	Asp	Val	Asp	Ala	Leu	Ser	Arg	Thr	Lys
1955						1960					1965			
Thr	Val	Gly	Asp	Val	Ile	Glu	Ala	Met	Lys	Leu	Glu	Leu	Gly	Glu
1970						1975					1980			
Ser	Ser	Ser	Ile	Glu	Thr	Leu	Asn	Cys	Thr	Glu	Val	Glu	His	Thr
1985						1990					1995			
Ser	Tyr	Lys	Ser	Val	Lys	Ala	Ser	Gly	Cys	Glu	Asn	Val	Asp	Thr
2000						2005					2010			
Arg	Phe	Ala	Lys	Val	Val	Gln	Ile	Ser	Leu	Pro	Ser	Lys	Leu	Lys
2015						2020					2025			
Ser	Thr	Val	Ser	His	Asp	Arg	Pro	Val	Ile	Val	Val	Asp	Asp	Gly
2030						2035					2040			
Thr	Pro	Leu	Thr	Thr	Glu	Leu	Cys	Lys	Ile	Leu	Gly	Gly	Asn	Ile
2045						2050					2055			
Val	Val	Leu	Ser	Tyr	Gln	Gly	Lys	Pro	Ala	Gly	Pro	Arg	Gly	Val
2060						2065					2070			

ES 2 467 918 T3

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

ES 2 467 918 T3

Thr	His	Ala	Ala	Gly	Gly	Val	Lys	Pro	Thr	Pro	Lys	Ala	His	Arg
2315						2320					2325			
Ala	Leu	Val	Asn	Arg	Val	Thr	Gly	Ser	Arg	Glu	Val	Arg	Glu	Ser
2330						2335					2340			
Leu	Arg	Ala	Ile	Gln	Glu	Ala	Gly	Ala	Asn	Val	Glu	Tyr	Ile	Ala
2345						2350					2355			
Cys	Asp	Val	Ser	Asp	Glu	Asn	Lys	Val	Arg	Gln	Leu	Val	Gln	Arg
2360						2365					2370			
Val	Glu	Gln	Lys	Tyr	Gly	Cys	Glu	Ile	Thr	Gly	Ile	Trp	His	Ala
2375						2380					2385			
Ser	Gly	Val	Leu	Arg	Asp	Lys	Leu	Val	Glu	Gln	Lys	Thr	Thr	Asp
2390						2395					2400			
Asp	Phe	Glu	Ala	Val	Phe	Gly	Thr	Lys	Val	Thr	Gly	Leu	Val	Asn
2405						2410					2415			
Ile	Val	Ser	Gln	Val	Asn	Met	Ser	Lys	Leu	Arg	His	Phe	Ile	Leu
2420						2425					2430			
Phe	Ser	Ser	Leu	Ala	Gly	Phe	His	Gly	Asn	Lys	Gly	Gln	Thr	Asp
2435						2440					2445			
Tyr	Ala	Ile	Ala	Asn	Glu	Ala	Leu	Asn	Lys	Ile	Ala	His	Thr	Leu
2450						2455					2460			
Ser	Ala	Phe	Leu	Pro	Lys	Leu	Asn	Ala	Lys	Val	Leu	Asp	Phe	Gly
2465						2470					2475			
Pro	Trp	Val	Gly	Ser	Gly	Met	Val	Thr	Glu	Thr	Leu	Glu	Lys	His
2480						2485					2490			
Phe	Lys	Ala	Met	Gly	Val	Gln	Thr	Ile	Pro	Leu	Glu	Pro	Gly	Ala
2495						2500					2505			
Arg	Thr	Val	Ala	Gln	Ile	Ile	Leu	Ala	Ser	Ser	Pro	Pro	Gln	Ser
2510						2515					2520			
Leu	Leu	Gly	Asn	Trp	Gly	Phe	Pro	Ala	Thr	Lys	Pro	Leu	Gln	Arg
2525						2530					2535			
Ser	Asn	Val	Val	Thr	Gly	Thr	Leu	Ser	Pro	Glu	Glu	Ile	Glu	Phe
2540						2545					2550			

ES 2 467 918 T3

Ile	Ala	Asp	His	Lys	Ile	Gln	Gly	Arg	Lys	Val	Leu	Pro	Met	Met
2555						2560					2565			
Ala	Ala	Ile	Gly	Phe	Met	Ala	Ser	Ile	Ala	Glu	Gly	Leu	Tyr	Pro
2570						2575					2580			
Gly	Tyr	Asn	Leu	Gln	Gly	Val	Glu	Asn	Ala	Gln	Leu	Phe	Gln	Gly
2585						2590					2595			
Leu	Thr	Ile	Asn	Gln	Glu	Thr	Lys	Phe	Gln	Ile	Thr	Leu	Ile	Glu
2600						2605					2610			
Glu	His	Asn	Ser	Glu	Glu	Asn	Leu	Asp	Val	Leu	Thr	Ser	Leu	Gly
2615						2620					2625			
Val	Met	Leu	Glu	Ser	Gly	Lys	Val	Leu	Pro	Ala	Tyr	Arg	Cys	Val
2630						2635					2640			
Val	Cys	Leu	Asn	Thr	Thr	Gln	Gln	Gln	Pro	Lys	Leu	Ser	Pro	Lys
2645						2650					2655			
Ile	Leu	Asn	Leu	Glu	Val	Asp	Pro	Ala	Cys	Glu	Val	Asn	Pro	Tyr
2660						2665					2670			
Asp	Gly	Lys	Ser	Leu	Phe	His	Gly	Pro	Leu	Leu	Gln	Phe	Val	Gln
2675						2680					2685			
Gln	Val	Leu	His	Ser	Ser	Thr	Lys	Gly	Leu	Val	Ala	Lys	Cys	Arg
2690						2695					2700			
Ala	Leu	Pro	Ile	Lys	Glu	Ala	Ile	Arg	Gly	Pro	Phe	Ile	Lys	Gln
2705						2710					2715			
Thr	Leu	His	Asp	Pro	Ile	Leu	Asp	Asp	Val	Ile	Phe	Gln	Leu	Met
2720						2725					2730			
Leu	Val	Trp	Cys	Arg	Asn	Ala	Leu	Gly	Ser	Ala	Ser	Leu	Pro	Asn
2735						2740					2745			
Arg	Ile	Glu	Lys	Met	Ser	Tyr	Phe	Gly	Asn	Val	Ser	Glu	Gly	Ser
2750						2755					2760			
Thr	Phe	Phe	Ala	Ser	Val	Thr	Pro	Val	Gly	Pro	Arg	Val	Pro	Lys
2765						2770					2775			
Asp	Pro	Val	Ile	Lys	Met	Gln	Phe	Leu	Leu	Gln	Asp	Glu	Ser	Gly
2780						2785					2790			

Asn Thr Phe Ser Ser Gly Glu Gly Ser Val Val Leu Ser Asp Glu
 2795 2800 2805

Leu Val Phe
 2810

<210> 40

5 <211> 1500

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1500)

<400> 40

ES 2 467 918 T3

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

ES 2 467 918 T3

Ser	Phe	Pro	Gly	Phe	Leu	Gly	Asn	Val	Thr	Ala	Gly	Arg	Cys	Ser	Asn	
			180					185					190			
acc	ttc	aac	atg	gaa	ggc	atg	aac	tgc	gtt	gtg	gat	gct	gca	tgt	gcc	624
Thr	Phe	Asn	Met	Glu	Gly	Met	Asn	Cys	Val	Val	Asp	Ala	Ala	Cys	Ala	
		195					200				205					
agt	tct	cta	att	gca	atc	aag	gtt	gca	gtt	gaa	gag	cta	ctc	ttt	ggc	672
Ser	Ser	Leu	Ile	Ala	Ile	Lys	Val	Ala	Val	Glu	Glu	Leu	Leu	Phe	Gly	
	210					215				220						
gac	tgt	gac	acc	atg	att	gca	ggc	gcc	acc	tgc	acg	gac	aat	tca	ctt	720
Asp	Cys	Asp	Thr	Met	Ile	Ala	Gly	Ala	Thr	Cys	Thr	Asp	Asn	Ser	Leu	
225					230					235					240	
ggc	atg	tac	atg	gcc	ttc	tct	aaa	acg	cca	gtt	ttt	tct	act	gac	cca	768
Gly	Met	Tyr	Met	Ala	Phe	Ser	Lys	Thr	Pro	Val	Phe	Ser	Thr	Asp	Pro	
				245					250					255		
agt	gtc	cgc	gcg	tat	gat	gag	aaa	aca	aaa	ggg	atg	cta	att	gga	gaa	816
Ser	Val	Arg	Ala	Tyr	Asp	Glu	Lys	Thr	Lys	Gly	Met	Leu	Ile	Gly	Glu	
			260					265					270			
ggc	tca	gca	atg	ttc	gtt	ctt	aaa	cgc	tat	gcg	gat	gcc	gta	cgt	gat	864
Gly	Ser	Ala	Met	Phe	Val	Leu	Lys	Arg	Tyr	Ala	Asp	Ala	Val	Arg	Asp	
		275					280					285				
ggc	gac	aca	att	cac	gcg	gtt	ctg	cgt	tct	tgc	tct	tcg	tct	agt	gat	912
Gly	Asp	Thr	Ile	His	Ala	Val	Leu	Arg	Ser	Cys	Ser	Ser	Ser	Ser	Asp	
	290					295					300					
gga	aaa	gcg	gca	gga	att	tat	act	cct	act	ata	tct	gga	caa	gaa	gaa	960
Gly	Lys	Ala	Ala	Gly	Ile	Tyr	Thr	Pro	Thr	Ile	Ser	Gly	Gln	Glu	Glu	
305					310					315					320	
gct	ttg	cgt	cga	gcg	tat	gcc	cgt	gcg	ggg	gta	tgt	cca	tct	acg	atc	1008
Ala	Leu	Arg	Arg	Ala	Tyr	Ala	Arg	Ala	Gly	Val	Cys	Pro	Ser	Thr	Ile	
				325					330					335		
ggg	ctt	gtt	gag	ggc	cac	ggg	aca	ggg	acc	cct	gtt	gga	gat	cgc	att	1056
Gly	Leu	Val	Glu	Gly	His	Gly	Thr	Gly	Thr	Pro	Val	Gly	Asp	Arg	Ile	
			340					345					350			
gag	tta	aca	gct	ctg	cgg	aac	ttg	ttt	gac	aaa	gct	ttt	ggc	agc	aag	1104
Glu	Leu	Thr	Ala	Leu	Arg	Asn	Leu	Phe	Asp	Lys	Ala	Phe	Gly	Ser	Lys	
		355					360					365				
aag	gaa	caa	ata	gca	gtt	ggc	agc	ata	aag	tct	cag	ata	ggc	cac	ctg	1152
Lys	Glu	Gln	Ile	Ala	Val	Gly	Ser	Ile	Lys	Ser	Gln	Ile	Gly	His	Leu	
	370					375					380					
aaa	tct	gtt	gcc	ggc	ttt	gcc	ggc	ttg	gtc	aaa	gct	gtg	ctt	gcg	ctt	1200
Lys	Ser	Val	Ala	Gly	Phe	Ala	Gly	Leu	Val	Lys	Ala	Val	Leu	Ala	Leu	
385					390					395					400	
aaa	cac	aaa	acg	ctc	cca	ggc	tcg	att	aat	gtc	gac	cag	cca	cct	ttg	1248
Lys	His	Lys	Thr	Leu	Pro	Gly	Ser	Ile	Asn	Val	Asp	Gln	Pro	Pro	Leu	
				405					410					415		
ttg	tat	gac	ggc	act	caa	att	caa	gac	tct	tct	tta	tat	atc	aac	aag	1296
Leu	Tyr	Asp	Gly	Thr	Gln	Ile	Gln	Asp	Ser	Ser	Leu	Tyr	Ile	Asn	Lys	
			420					425					430			
aca	aat	aga	cca	tgg	ttt	acg	caa	aac	aag	ctt	ccg	cgt	cgg	gct	ggc	1344

ES 2 467 918 T3

Thr	Asn	Arg	Pro	Trp	Phe	Thr	Gln	Asn	Lys	Leu	Pro	Arg	Arg	Ala	Gly		
		435					440					445					
gtc	tca	agt	ttt	gga	ttt	gga	ggt	gca	aac	tac	cac	gcg	gtt	ctg	gaa	1392	
Val	Ser	Ser	Phe	Gly	Phe	Gly	Gly	Ala	Asn	Tyr	His	Ala	Val	Leu	Glu		
	450					455					460						
gaa	ttc	gag	ccc	gag	cat	gaa	aaa	cca	tac	cgc	ctc	aat	act	gtt	gga	1440	
Glu	Phe	Glu	Pro	Glu	His	Glu	Lys	Pro	Tyr	Arg	Leu	Asn	Thr	Val	Gly		
	465				470					475					480		
cat	cct	gtc	ctc	ttg	tac	gct	ccg	tct	gtg	gaa	gcc	ctc	aaa	gta	ctt	1488	
His	Pro	Val	Leu	Leu	Tyr	Ala	Pro	Ser	Val	Glu	Ala	Leu	Lys	Val	Leu		
				485					490					495			
tgc	aac	gac	cag													1500	
Cys	Asn	Asp	Gln														
			500														

<210> 41

5 <211> 500

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 41

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

ES 2 467 918 T3

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

ES 2 467 918 T3

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

<210> 42

5 <211> 1500

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1) .. (1500)

<400> 42

ES 2 467 918 T3

ctt gcg gag ctc aca att gca ttg gaa gag gca aaa aca cat aaa aat	48
Leu Ala Glu Leu Thr Ile Ala Leu Glu Glu Ala Lys Thr His Lys Asn	
1 5 10 15	
gtt gac aaa gtt tgt ggc tac aag ttt att gac gaa ttt cag ctc caa	96
Val Asp Lys Val Cys Gly Tyr Lys Phe Ile Asp Glu Phe Gln Leu Gln	
20 25 30	
gga agc tgt cct cca gaa aat ccg aga gta gga ttt tta gca aca ctg	144
Gly Ser Cys Pro Pro Glu Asn Pro Arg Val Gly Phe Leu Ala Thr Leu	
35 40 45	
cct act tca aat atc att gtc gcg ctt aag gca att ctc gcg cag ctt	192
Pro Thr Ser Asn Ile Ile Val Ala Leu Lys Ala Ile Leu Ala Gln Leu	
50 55 60	
gat gca aaa cca gat gcg aag aaa tgg gat ttg cct cat aaa aag gct	240
Asp Ala Lys Pro Asp Ala Lys Lys Trp Asp Leu Pro His Lys Lys Ala	
65 70 75 80	
ttt ggg gct acc ttc gca tcg tct tca gtg aaa ggc tct gtt gct gcg	288
Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly Ser Val Ala Ala	
85 90 95	
ctc ttc gca gga cag ggt acc cag tac tta aac atg ttc tct gat gtg	336
Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met Phe Ser Asp Val	

ES 2 467 918 T3

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

ES 2 467 918 T3

355	360	365	
gga ccc cgc caa gtc ctt	gca aag ctt atc gcg	gaa atg ttt ccc tcg	1152
Gly Pro Arg Gln Val Leu	Ala Lys Leu Ile Ala	Glu Met Phe Pro Ser	
370	375	380	
tgt aca gct atc agc gtt	aac ccc gcg agc agt	ggt gac agt gac gtg	1200
Cys Thr Ala Ile Ser Val	Asn Pro Ala Ser Ser	Gly Asp Ser Asp Val	
385	390	395	400
caa ctc cgc ctc gcc gcc	gta aaa ttc gcg gtc	tcg ggt gca gcc ctt	1248
Gln Leu Arg Leu Ala Ala	Val Lys Phe Ala Val	Ser Gly Ala Ala Leu	
405	410	415	
agc acc ttt gat cca tgg	gag tat cgc aag cca	caa gat ctt ctt att	1296
Ser Thr Phe Asp Pro Trp	Glu Tyr Arg Lys Pro	Gln Asp Leu Leu Ile	
420	425	430	
cga aaa cca cga aaa act	gcc ctt gtt cta tca	gca gca aca tat gtt	1344
Arg Lys Pro Arg Lys Thr	Ala Leu Val Leu Ser	Ala Ala Thr Tyr Val	
435	440	445	
tcc cca aag act ctt gca	gaa cgt aaa aag gct	atg gaa gat atc aag	1392
Ser Pro Lys Thr Leu Ala	Glu Arg Lys Lys Ala	Met Glu Asp Ile Lys	
450	455	460	
cta gta tcc att aca cca	aga gat agt atg gta	tca att gga aaa atc	1440
Leu Val Ser Ile Thr Pro	Arg Asp Ser Met Val	Ser Ile Gly Lys Ile	
465	470	475	480
gcg caa gaa gta cgg aca	gct aaa cag cct tta	gaa acc gaa att cga	1488
Ala Gln Glu Val Arg Thr	Ala Lys Gln Pro Leu	Glu Thr Glu Ile Arg	
485	490	495	
aga ctc aac aaa			1500
Arg Leu Asn Lys			
500			

<210> 43

5 <211> 500

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 43

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

ES 2 467 918 T3

Phe Gly Ala Thr Phe Ala Ser Ser Ser Val Lys Gly Ser Val Ala Ala
85 90 95

Leu Phe Ala Gly Gln Gly Thr Gln Tyr Leu Asn Met Phe Ser Asp Val
100 105 110

Ala Met Asn Trp Pro Pro Phe Arg Asp Ser Ile Val Ala Met Glu Glu
115 120 125

Ala Gln Thr Glu Val Phe Glu Gly Gln Val Glu Pro Ile Ser Lys Val
130 135 140

Leu Phe Pro Arg Glu Arg Tyr Ala Ser Glu Ser Glu Gln Gly Asn Glu
145 150 155 160

Leu Leu Cys Leu Thr Glu Tyr Ser Gln Pro Thr Thr Ile Ala Ala Ala
165 170 175

Val Gly Ala Phe Asp Ile Phe Lys Ala Ala Gly Phe Lys Pro Asp Met
180 185 190

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

Ser Ile Ser Arg Asp Asp Leu Tyr Lys Leu Val Cys Lys Arg Ala Lys
210 215 220

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

Asp Ala Arg Leu Val Thr Pro Gln Asn Ser Asp Val Tyr Val Ala Asn
245 250 255

Phe Asn Ser Ala Thr Gln Val Val Ile Ser Gly Thr Val Gln Gly Val
260 265 270

Lys Glu Glu Ser Lys Leu Leu Ile Ser Lys Gly Phe Arg Val Leu Pro
275 280 285

Leu Lys Cys Gln Gly Ala Phe His Ser Pro Leu Met Gly Pro Ser Glu
290 295 300

Asp Ser Phe Lys Ser Leu Val Glu Thr Cys Thr Ile Ser Pro Pro Lys
305 310 315 320

Asn Val Lys Phe Phe Cys Asn Val Ser Gly Lys Glu Ser Pro Asn Pro
325 330 335

ES 2 467 918 T3

Lys Gln Thr Leu Lys Ser His Met Thr Ser Ser Val Gln Phe Glu Glu
340 345 350

Gln Ile Arg Asn Met Tyr Asp Ala Gly Ala Arg Val Phe Leu Glu Phe
355 360 365

Gly Pro Arg Gln Val Leu Ala Lys Leu Ile Ala Glu Met Phe Pro Ser
370 375 380

Cys Thr Ala Ile Ser Val Asn Pro Ala Ser Ser Gly Asp Ser Asp Val
385 390 395 400

Gln Leu Arg Leu Ala Ala Val Lys Phe Ala Val Ser Gly Ala Ala Leu
405 410 415

Ser Thr Phe Asp Pro Trp Glu Tyr Arg Lys Pro Gln Asp Leu Leu Ile
420 425 430

Arg Lys Pro Arg Lys Thr Ala Leu Val Leu Ser Ala Ala Thr Tyr Val
435 440 445

Ser Pro Lys Thr Leu Ala Glu Arg Lys Lys Ala Met Glu Asp Ile Lys
450 455 460

Leu Val Ser Ile Thr Pro Arg Asp Ser Met Val Ser Ile Gly Lys Ile
465 470 475 480

Ala Gln Glu Val Arg Thr Ala Lys Gln Pro Leu Glu Thr Glu Ile Arg
485 490 495

Arg Leu Asn Lys
500

<210> 44

5 <211> 351

<212> ADN

<213> *Thraustochytrium* sp.

10 <220>

<221> CDS

15 <222> (1) .. (351)

<400> 44

tcg acc cca gcg tca gag cgg tcg gct tca ccg ctt ttc gag aaa cgc 48
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1 5 10 15

agt tcg gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg 96
Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Val Leu
20 25 30

ES 2 467 918 T3

agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg 144
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
35 40 45

gac atg gac ctg gag agt gag ctt ggc gtt gat agc atc aaa cgc gtg 192
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
50 55 60

gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac 240
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65 70 75 80

gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg 288
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
85 90 95

atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa 336
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
100 105 110

tcg atc cgt cag cca 351
Ser Ile Arg Gln Pro
115

<210> 45

5 <211> 117

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 45

Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
1 5 10 15

Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
20 25 30

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
35 40 45

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
50 55 60

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
65 70 75 80

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
85 90 95

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
100 105 110

Ser Ile Arg Gln Pro
115

15 <210> 46

<211> 5
 <212> PRT
 5 <213> *Thraustochytrium* sp.
 <220>
 10 <221> MISC_FEATURE
 <222> (1) .. (5)
 <223> Xaa puede ser cualquier aminoácido que se produce de manera natural
 15 <400> 46
 Leu Gly Xaa Asp Ser
 1 5
 20 <210> 47
 <211> 2790
 <212> ADN
 25 <213> *Thraustochytrium* sp.
 <220>
 30 <221> CDS
 <222> (1) .. (2790)
 <400> 47
 35

ES 2 467 918 T3

tcg acc cca gcg tca gag cgg tcg gct tca ccg ctt ttc gag aaa cgc	48
Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg	
1 5 10 15	
agt tcg gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg	96
Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Ala Val Leu	
20 25 30	
agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg	144
Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met	
35 40 45	
gac atg gac ctg gag agt gag ctt ggc gtt gat agc atc aaa cgc gtg	192
Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val	
50 55 60	
gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac	240
Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp	
65 70 75 80	
gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg	288
Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala	
85 90 95	
atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa	336
Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu	
100 105 110	
tcg atc cgt cag cca ccg gtg tcc gag cct gct gta ccg acc tca tcg	384
Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser	
115 120 125	
tca agc agt att gct aat gtt tcg tca gca cgc ctc gct gaa gct gaa	432
Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu	
130 135 140	

ES 2 467 918 T3

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

gag agc gag ctt ggc gtt gat agc atc aaa cgc gtg gag atc atg agc Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser 405 410 415	1248
gag gtt caa acg ttg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu 420 425 430	1296
tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag ctg gaa Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu 435 440 445	1344
ctc ggt gga ccc caa ggc cag act ttg acc gcg gaa tcg atc cgt cag Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln 450 455 460	1392
cca ccg gtg tct gag cct gct gta ccg acc tca tcg tca agc agt att Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile 465 470 475 480	1440
gct aat gtt tcg tca gca cgc ctc gct gaa gct gaa gcc gcg gta ctg Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Val Leu 485 490 495	1488
agc gtt ctc gca gac aag aca ggc tac gac agc tca atg atc gag atg Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met 500 505 510	1536
gac atg gac ctg gag agt gag ctt ggc gtc gac agc atc aaa cgc gtg Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val 515 520 525	1584
gag atc atg agc gag gtt caa acg ctg ctc agc gtg gaa gtc tcc gac Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp 530 535 540	1632
gtt gac gct ctg tca aga acc aag act gtt ggc gac gtc atc gag gcg Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala 545 550 555 560	1680
atg aag ctg gaa ctc ggt gga ccc caa ggc cag act ttg acc tct gaa Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ser Glu 565 570 575	1728
ccg atc cat cag cca cca gtg tcc gag cct gct gta ccg acc tca tcg Pro Ile His Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser 580 585 590	1776
tca agc agt att gct aat gtt tct tca gca cgc ctc gct gaa gct gaa Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu 595 600 605	1824
gcc gcg gta ctg agc gtt ctc gca gac aag aca ggc tac gac agc tca Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser 610 615 620	1872
atg atc gag atg gac atg gac ctg gag agc gag ctt ggc gtt gat agc Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser 625 630 635 640	1920
atc aaa cgc gtg gaa atc atg agc gag gtt caa acg ctg ctc agc gtg Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val 645 650 655	1968

gaa gtc tcc gac gtt gac gct ctg tca aga acc aag act gtt ggc gac Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp 660 665 670	2016
gtc atc gag gcg atg aag atg gaa ctc ggt gga ccc caa ggc cag act Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr 675 680 685	2064
ttg acc gcg gaa tcg atc cgt cag cca ccg gtg tct gag cct gct gta Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val 690 695 700	2112
ccg acc tca tcg tca agc agt att gct aat gtt tcg tca gca cgc ctc Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu 705 710 715 720	2160
gct gaa gct gaa gcg gcg gta ctg agc gtt ctc gca gac aag aca ggc Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly 725 730 735	2208
tac gac agc tca atg atc gag atg gac atg gac ctg gag agc gag ctt Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu 740 745 750	2256
ggc gtt gat agc atc aaa cgc gtg gag atc atg agc gag gtt caa gcg Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Ala 755 760 765	2304
ctg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg tca aga acc aag Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys 770 775 780	2352
act gtt ggc gac gtc atc gag gcg atg aag atg gaa ctc ggt gga ccc Thr Val Gly Asp Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro 785 790 795 800	2400
caa ggc cag act ttg acc gca gaa tcg atc cgt gag cca ccg gtg tct Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Glu Pro Pro Val Ser 805 810 815	2448
gag cct gct gta ccg acc tca tcg tca agt agt atc gct aat gtt tct Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser 820 825 830	2496
tca gct cgc ctc gct gaa gct gaa gcc gcg gta ctg agc gtt ctc gca Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala 835 840 845	2544
gac aag aca ggc tac gac agc tca atg atc gag atg gac atg gac ctg Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu 850 855 860	2592
gag agt gag ctt ggc gtc gac agc atc aaa cgc gtg gag atc atg agc Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser 865 870 875 880	2640
gag gtt caa acg ttg ctc agc gtg gaa gtc tcc gac gtt gac gct ctg Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu 885 890 895	2688
tca aga acc aag act gtt ggc gac gtc atc gag gcg atg aag ctg gaa Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu 900 905 910	2736

ES 2 467 918 T3

ctt ggg gaa tca tca agt att gag act ctc aat tgt acc gag gtt gag 2784
 Leu Gly Glu Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu
 915 920 925

cac acg 2790
 His Thr
 930

<210> 48

5 <211> 930

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 48

Ser Thr Pro Ala Ser Glu Arg Ser Ala Ser Pro Leu Phe Glu Lys Arg
 1 5 10 15

Ser Ser Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu
 20 25 30

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
 35 40 45

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
 50 55 60

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
 65 70 75 80

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
 85 90 95

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu
 100 105 110

Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
 115 120 125

Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
 130 135 140

Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
 145 150 155 160

Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
 165 170 175

Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
 180 185 190

ES 2 467 918 T3

Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
 195 200 205

Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr
 210 215 220

Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
 225 230 235 240

Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
 245 250 255

Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
 260 265 270

Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
 275 280 285

Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr
 290 295 300

Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
 305 310 315 320

Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu Leu Gly Gly Pro
 325 330 335

Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser
 340 345 350

Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Leu
 355 360 365

Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala
 370 375 380

Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
 385 390 395 400

Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
 405 410 415

Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
 420 425 430

Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
 435 440 445

ES 2 467 918 T3

Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Gln
 450 455 460

Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ile
 465 470 475 480

Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu Ala Val Leu
 485 490 495

Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met
 500 505 510

Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val
 515 520 525

Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp
 530 535 540

Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala
 545 550 555 560

Met Lys Leu Glu Leu Gly Gly Pro Gln Gly Gln Thr Leu Thr Ser Glu
 565 570 575

Pro Ile His Gln Pro Pro Val Ser Glu Pro Ala Val Pro Thr Ser Ser
 580 585 590

Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu Ala Glu Ala Glu
 595 600 605

Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly Tyr Asp Ser Ser
 610 615 620

Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu Gly Val Asp Ser
 625 630 635 640

Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Thr Leu Leu Ser Val
 645 650 655

Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys Thr Val Gly Asp
 660 665 670

Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro Gln Gly Gln Thr
 675 680 685

Leu Thr Ala Glu Ser Ile Arg Gln Pro Pro Val Ser Glu Pro Ala Val
 690 695 700

ES 2 467 918 T3

Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser Ser Ala Arg Leu
705 710 715 720

Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala Asp Lys Thr Gly
725 730 735

Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu Glu Ser Glu Leu
740 745 750

Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser Glu Val Gln Ala
755 760 765

Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu Ser Arg Thr Lys
770 775 780

Thr Val Gly Asp Val Ile Glu Ala Met Lys Met Glu Leu Gly Gly Pro
785 790 795 800

Gln Gly Gln Thr Leu Thr Ala Glu Ser Ile Arg Glu Pro Pro Val Ser
805 810 815

Glu Pro Ala Val Pro Thr Ser Ser Ser Ser Ser Ile Ala Asn Val Ser
820 825 830

Ser Ala Arg Leu Ala Glu Ala Glu Ala Ala Val Leu Ser Val Leu Ala
835 840 845

Asp Lys Thr Gly Tyr Asp Ser Ser Met Ile Glu Met Asp Met Asp Leu
850 855 860

Glu Ser Glu Leu Gly Val Asp Ser Ile Lys Arg Val Glu Ile Met Ser
865 870 875 880

Glu Val Gln Thr Leu Leu Ser Val Glu Val Ser Asp Val Asp Ala Leu
885 890 895

Ser Arg Thr Lys Thr Val Gly Asp Val Ile Glu Ala Met Lys Leu Glu
900 905 910

Leu Gly Glu Ser Ser Ser Ile Glu Thr Leu Asn Cys Thr Glu Val Glu
915 920 925

His Thr
930

<210> 49

5 <211> 2433

<212> ADN

<213> *Thraustochytrium* sp.

<220>

<221> CDS

5

<222> (1)..(2433)

<400> 49

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

10

ES 2 467 918 T3

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

ES 2 467 918 T3

gta ggt tca gga atg gta acc gaa aca ctt gag aag cat ttt aaa gct	1488
Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His Phe Lys Ala	
485 490 495	
atg ggg gtt cag act att cct ctc gag cca gga gca cgg act gtt gcg	1536
Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala Arg Thr Val Ala	
500 505 510	
caa atc att ttg gca agt tcg cca ccg caa tcg ctt ttg ggg aac tgg	1584
Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser Leu Leu Gly Asn Trp	
515 520 525	
ggc ttt cca gcc acc aaa ccg cta caa cgc tct aat gta gtc acg ggc	1632
Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg Ser Asn Val Val Thr Gly	
530 535 540	
aca ctc tct ccg gaa gag ata gaa ttc atc gca gac cac aaa att caa	1680
Thr Leu Ser Pro Glu Glu Ile Glu Phe Ile Ala Asp His Lys Ile Gln	
545 550 555 560	
ggc cgc aag gtg ctt ccc atg atg gct gca atc ggg ttc atg gcc tct	1728
Gly Arg Lys Val Leu Pro Met Met Ala Ala Ile Gly Phe Met Ala Ser	
565 570 575	
att gcg gaa gga ctc tac ccg ggg tac aat ctg caa ggc gtg gaa aat	1776
Ile Ala Glu Gly Leu Tyr Pro Gly Tyr Asn Leu Gln Gly Val Glu Asn	
580 585 590	
gct cag ctc ttt caa ggc ttg act atc aac caa gag aca aaa ttt caa	1824
Ala Gln Leu Phe Gln Gly Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln	
595 600 605	
atc act ctc att gag gag cac aac tct gag gaa aac ctg gat gtc ctg	1872
Ile Thr Leu Ile Glu Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu	
610 615 620	
aca tcc ctt ggt gta atg ttg gaa agc ggg aag gtg ctt ccc gct tac	1920
Thr Ser Leu Gly Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr	
625 630 635 640	
cga tgt gtt gta tgc ttg aat aca acc cag cag cag ccc aag cta tct	1968
Arg Cys Val Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser	
645 650 655	
cca aaa att ctt aac ttg gaa gtt gac cct gca tgc gag gtt aac ccc	2016
Pro Lys Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro	
660 665 670	
tat gat gga aag tcg ttg ttc cac ggt ccg ctt ttg caa ttc gtt caa	2064
Tyr Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln	
675 680 685	
caa gtg ttg cac tca agt acc aaa ggc ctc gtt gcc aag tgc cgc gcg	2112
Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg Ala	
690 695 700	
ctt cca atc aaa gaa gcc atc cga ggg cca ttt atc aag caa aca ctc	2160
Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln Thr Leu	
705 710 715 720	
cat gat cca att cta gac gac gtc att ttt cag cta atg ctc gtg tgg	2208
His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met Leu Val Trp	
725 730 735	

ES 2 467 918 T3

tgt cgt aat gct cta gga agt gca tcg cta ccc aac aga att gaa aag 2256
Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn Arg Ile Glu Lys
740 745 750

atg tca tac ttt ggg aat gtc tca gaa ggt agc act ttc ttt gcc tca 2304
Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser Thr Phe Phe Ala Ser
755 760 765

gtt aca cct gtg gga cca aga gta cca aag gat ccc gtg atc aaa atg 2352
Val Thr Pro Val Gly Pro Arg Val Pro Lys Asp Pro Val Ile Lys Met
770 775 780

cag ttt ctt ctc caa gat gaa tcc ggc aac aca ttt tca tcg ggg gag 2400
Gln Phe Leu Leu Gln Asp Glu Ser Gly Asn Thr Phe Ser Ser Gly Glu
785 790 795 800

ggc tcg gtt gtg ctt agt gac gaa ctc gtc ttt 2433
Gly Ser Val Val Leu Ser Asp Glu Leu Val Phe
805 810

<210> 50

5 <211> 811

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 50

Lys Ser Val Lys Ala Ser Gly Cys Glu Asn Val Asp Thr Arg Phe Ala
1 5 10 15

Lys Val Val Gln Ile Ser Leu Pro Ser Lys Leu Lys Ser Thr Val Ser
20 25 30

His Asp Arg Pro Val Ile Val Val Asp Asp Gly Thr Pro Leu Thr Thr
35 40 45

Glu Leu Cys Lys Ile Leu Gly Gly Asn Ile Val Val Leu Ser Tyr Gln
50 55 60

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

Glu Ala Leu Ile Gln Ala Leu Ala Leu Ile Arg Ser Thr Tyr Gly Val
85 90 95

Pro Ile Gly Phe Ile Cys Gln Gln Val Ser Asn Val Ser Thr Lys Ala
100 105 110

Gln Leu Cys Trp Ala Leu Leu Ala Ala Lys His Leu Lys Lys Asp Leu
115 120 125

Asn Ala Val Leu Pro Asp Ser Arg Ser Phe Phe Val Gly Val Val Arg
130 135 140

ES 2 467 918 T3

Leu Asn Gly Lys Leu Gly Thr Phe Glu Asn Ile Ser Asp Phe Ser Lys
 145 150 155 160

Phe Asp Leu Thr Lys Ala Leu Asp Tyr Gly Gln Arg Gly Ser Leu Leu
 165 170 175

Gly Leu Cys Lys Ser Leu Asp Leu Glu Trp Glu Gln Val Phe Cys Arg
 180 185 190

Gly Ile Asp Leu Ala Cys Asp Leu Met Pro Leu Gln Ala Ala Arg Ile
 195 200 205

Leu Arg Asn Glu Leu Gln Cys Pro Asn Met Arg Leu Arg Glu Val Gly
 210 215 220

Tyr Asp Ile Ser Gly Ala Arg Tyr Thr Ile Ser Thr Asp Asp Leu Leu
 225 230 235 240

Cys Gly Pro Ser Lys Ala Lys Val Glu Ala Ala Asp Leu Phe Leu Val
 245 250 255

Thr Gly Gly Ala Arg Gly Ile Thr Pro His Cys Val Arg Glu Ile Ala
 260 265 270

Ser Arg Ser Pro Gly Thr Thr Phe Val Leu Val Gly Arg Ser Glu Met
 275 280 285

Ser Asp Glu Pro Asp Trp Ala Val Gly His Tyr Asn Lys Asp Leu Asp
 290 295 300

Gln Ser Thr Met Lys His Leu Lys Ala Thr His Ala Ala Gly Gly Val
 305 310 315 320

Lys Pro Thr Pro Lys Ala His Arg Ala Leu Val Asn Arg Val Thr Gly
 325 330 335

Ser Arg Glu Val Arg Glu Ser Leu Arg Ala Ile Gln Glu Ala Gly Ala
 340 345 350

Asn Val Glu Tyr Ile Ala Cys Asp Val Ser Asp Glu Asn Lys Val Arg
 355 360 365

Gln Leu Val Gln Arg Val Glu Gln Lys Tyr Gly Cys Glu Ile Thr Gly
 370 375 380

Ile Trp His Ala Ser Gly Val Leu Arg Asp Lys Leu Val Glu Gln Lys
 385 390 395 400

ES 2 467 918 T3

Thr Thr Asp Asp Phe Glu Ala Val Phe Gly Thr Lys Val Thr Gly Leu
405 410 415

Val Asn Ile Val Ser Gln Val Asn Met Ser Lys Leu Arg His Phe Ile
420 425 430

Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Lys Gly Gln Thr Asp
435 440 445

Tyr Ala Ile Ala Asn Glu Ala Leu Asn Lys Ile Ala His Thr Leu Ser
450 455 460

Ala Phe Leu Pro Lys Leu Asn Ala Lys Val Leu Asp Phe Gly Pro Trp
465 470 475 480

Val Gly Ser Gly Met Val Thr Glu Thr Leu Glu Lys His Phe Lys Ala
485 490 495

Met Gly Val Gln Thr Ile Pro Leu Glu Pro Gly Ala Arg Thr Val Ala
500 505 510

Gln Ile Ile Leu Ala Ser Ser Pro Pro Gln Ser Leu Leu Gly Asn Trp
515 520 525

Gly Phe Pro Ala Thr Lys Pro Leu Gln Arg Ser Asn Val Val Thr Gly
530 535 540

Thr Leu Ser Pro Glu Glu Ile Glu Phe Ile Ala Asp His Lys Ile Gln
545 550 555 560

Gly Arg Lys Val Leu Pro Met Met Ala Ala Ile Gly Phe Met Ala Ser
565 570 575

Ile Ala Glu Gly Leu Tyr Pro Gly Tyr Asn Leu Gln Gly Val Glu Asn
580 585 590

Ala Gln Leu Phe Gln Gly Leu Thr Ile Asn Gln Glu Thr Lys Phe Gln
595 600 605

Ile Thr Leu Ile Glu Glu His Asn Ser Glu Glu Asn Leu Asp Val Leu
610 615 620

Thr Ser Leu Gly Val Met Leu Glu Ser Gly Lys Val Leu Pro Ala Tyr
625 630 635 640

Arg Cys Val Val Cys Leu Asn Thr Thr Gln Gln Gln Pro Lys Leu Ser
645 650 655

ES 2 467 918 T3

Pro Lys Ile Leu Asn Leu Glu Val Asp Pro Ala Cys Glu Val Asn Pro
660 665 670

Tyr Asp Gly Lys Ser Leu Phe His Gly Pro Leu Leu Gln Phe Val Gln
675 680 685

Gln Val Leu His Ser Ser Thr Lys Gly Leu Val Ala Lys Cys Arg Ala
690 695 700

Leu Pro Ile Lys Glu Ala Ile Arg Gly Pro Phe Ile Lys Gln Thr Leu
705 710 715 720

His Asp Pro Ile Leu Asp Asp Val Ile Phe Gln Leu Met Leu Val Trp
725 730 735

Cys Arg Asn Ala Leu Gly Ser Ala Ser Leu Pro Asn Arg Ile Glu Lys
740 745 750

Met Ser Tyr Phe Gly Asn Val Ser Glu Gly Ser Thr Phe Phe Ala Ser
755 760 765

Val Thr Pro Val Gly Pro Arg Val Pro Lys Asp Pro Val Ile Lys Met
770 775 780

Gln Phe Leu Leu Gln Asp Glu Ser Gly Asn Thr Phe Ser Ser Gly Glu
785 790 795 800

Gly Ser Val Val Leu Ser Asp Glu Leu Val Phe
805 810

<210> 51

5 <211> 5808

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(5805)

<220>

<221> misc_feature

20

<222> (743)..(743)

<223> n e s a, c, g o t

25 <400> 51

ES 2 467 918 T3

atg	caa	ctt	cct	cca	gcg	cat	tct	gcc	gat	gag	aat	cgc	atc	gcg	gtc	48
Met	Gln	Leu	Pro	Pro	Ala	His	Ser	Ala	Asp	Glu	Asn	Arg	Ile	Ala	Val	
1				5				10						15		
gtg	ggc	atg	gcc	gtc	aaa	tat	gcg	ggc	tgt	gac	aat	aaa	gaa	gag	ttt	96
Val	Gly	Met	Ala	Val	Lys	Tyr	Ala	Gly	Cys	Asp	Asn	Lys	Glu	Glu	Phe	
			20					25					30			

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

tat ggt gtg ctc ctt gaa gca aat tta agt aac gca ggt tgt ggg ctt Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu 290 295 300	912
cca ctc agc ccg cac tta ccg agc gaa gaa tca tgt att cgt gat acc Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr 305 310 315 320	960
tac cgc cgt gct gga gtt gct gca gat caa agt att cag tat att gag Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu 325 330 335	1008
tgc cac gct acg gga acc cct cga ggg gat gtc gtg gaa att gag gcg Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala 340 345 350	1056
gtt gaa aga gtt ttc aag aaa aac gtt cca cgc tta ggc tcg acg aaa Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys 355 360 365	1104
gga aat ttt ggt cac tcg tta gtt gcg gct ggt ttc gca ggt atg gca Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala 370 375 380	1152
aag ctt ctt ctt gca atg gaa cat gga gtg att cct ccc aca cca ggt Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly 385 390 395 400	1200
ctt gat gct tcg aac cag gca agt gag cac gtt gtg aca aag gct atc Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile 405 410 415	1248
act tgg cct gag aca cat ggg gct cca aaa cga gct ggc ctt tca gca Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala 420 425 430	1296
ttt gga ttt ggt ggg act aat gcg cat gca ctc ttc gaa gag ttt aat Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn 435 440 445	1344
gcc gag ggc ata agt tat cgc cct gga aag cct cca gtc gaa tcg aat Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Val Glu Ser Asn 450 455 460	1392
acc cgt cct tcc gtc gta ata act ggg atg gac tgt acc ttt ggg agc Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser 465 470 475 480	1440
ctt gaa ggg att gat gcg ttc gag act gcc ctg tac gag ggg cgt gac Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp 485 490 495	1488
gca gct cgt gac tta ccc gcc aaa cgt tgg agg ttc cta ggt gag gac Ala Ala Arg Asp Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp 500 505 510	1536
ttg gag ttt ctc cga gcc atc agg ctc aag gaa aag cct agg ggt tgt Leu Glu Phe Leu Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys 515 520 525	1584
ttt gtg gag agt gtt gac gtt aac ttt aga cgg ctg aaa acg ccc ttg Phe Val Glu Ser Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu 530 535 540	1632

ES 2 467 918 T3

aca cca gaa gat atg ttg cgg ccc caa caa ctc ttg gcg gtt tct acg Thr Pro Glu Asp Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr 545 550 555 560	1680
atg gac cga gca att atc gat gca ggt cta aag aag ggc caa cat gta Met Asp Arg Ala Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val 565 570 575	1728
gca gtt ctt gtt ggc cta gga act gac ctg gaa ctt tac cgt cat cga Ala Val Leu Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg 580 585 590	1776
gca aga gtc gcg ctt aaa gag gtt ttg cac ccg agc tta aag tca gac Ala Arg Val Ala Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp 595 600 605	1824
act gca att ctc cag aaa ata atg caa tat gtg aat gat gca gga act Thr Ala Ile Leu Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr 610 615 620	1872
tcg act tca tac aca tct tac att gga aac ctc gtt gcc acg cgt att Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile 625 630 635 640	1920
tcg tct cag tgg gga ttc aca ggg ccg tcc ttt act gtc aca gaa gga Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly 645 650 655	1968
aat aat tcc gtg tac aga tgt gca caa cta gcc aaa gat atg ctt cag Asn Asn Ser Val Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln 660 665 670	2016
gtt aac cga gtt gat gct gtc gtc atc gca ggc gtt gat ctc aac gga Val Asn Arg Val Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly 675 680 685	2064
agc gcc gaa agt ttt ttt gtc cga gca aat cgt caa aag ata tcc aag Ser Ala Glu Ser Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys 690 695 700	2112
cta agt cat cca tgt gca agc ttc gac aga gat gca gat gga ttt ttc Leu Ser His Pro Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe 705 710 715 720	2160
gca ggt gag ggc tgt ggt gcc cta gtt ttc aag agg tta gaa gac tgt Ala Gly Glu Gly Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys 725 730 735	2208
gct cct cag gaa aaa att tat gct agt ata gac tct atc gca ata gat Ala Pro Gln Glu Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp 740 745 750	2256
aaa gag cct act agc tca gct gtg aaa gct gtc tac caa agt gat tcg Lys Glu Pro Thr Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser 755 760 765	2304
agt ctc tcc gat att gag ctg tta gaa atc agt gga gac tcc aaa cgg Ser Leu Ser Asp Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg 770 775 780	2352
ttt gca gca ttc gaa ggc gct gtg gaa att caa tca agt gtg gaa gcc Phe Ala Ala Phe Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala 785 790 795 800	2400

cag cta aaa gga ctt tcc aaa gtc ctt gaa cct gca aaa ggc caa ggc Gln Leu Lys Gly Leu Ser Lys Val Leu Glu Pro Ala Lys Gly Gln Gly 805 810 815	2448
gta gcg gtg gga agt act cga gca acc gtt ggg gat ata ggg tat gct Val Ala Val Gly Ser Thr Arg Ala Thr Val Gly Asp Ile Gly Tyr Ala 820 825 830	2496
aca gga gcg gca agc ctg att aaa act gca ctc tgc tta tat aat cgc Thr Gly Ala Ala Ser Leu Ile Lys Thr Ala Leu Cys Leu Tyr Asn Arg 835 840 845	2544
tac ctt ccg gca tta gca aac tgg agt ggc cca tgt gaa cag tcc gcc Tyr Leu Pro Ala Leu Ala Asn Trp Ser Gly Pro Cys Glu Gln Ser Ala 850 855 860	2592
tgg ggc tca aac atg ttc gtt tgc cat gaa aca cgg ccg tgg atg aaa Trp Gly Ser Asn Met Phe Val Cys His Glu Thr Arg Pro Trp Met Lys 865 870 875 880	2640
aac cag aat gaa aag aga tgt gcc ctc att tct gga aca gat cca tct Asn Gln Asn Glu Lys Arg Cys Ala Leu Ile Ser Gly Thr Asp Pro Ser 885 890 895	2688
cat aca tgc ttt tcc ctc gta cta tcg gat act ggg tgt tat gaa gag His Thr Cys Phe Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu 900 905 910	2736
cac aat cga acg tgc ttt gat gtg caa gcg cca cag cta gtt ctg ata His Asn Arg Thr Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile 915 920 925	2784
cac gga ttc gat gga aaa act att gtg cgg cga ctt gaa gga tat ctc His Gly Phe Asp Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu 930 935 940	2832
ctt gaa ctt gtt gaa ggg cat gca agc cct tca gag tat ttc cac aaa Leu Glu Leu Val Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys 945 950 955 960	2880
ctg att gga caa agt cta ctt gag aac tcg aaa gaa agt aaa ctc aca Leu Ile Gly Gln Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr 965 970 975	2928
ctt tcg ctt gtg tgc aat ccg aac cag ctc caa aag gag ctc atg ctt Leu Ser Leu Val Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu 980 985 990	2976
gct atc aaa gga gta caa cga agc atg tta aca ggg aag gat tgg gtc Ala Ile Lys Gly Val Gln Arg Ser Met Leu Thr Gly Lys Asp Trp Val 995 1000 1005	3024
agt cca tca gga agt tgt ttt gcc cca aat ccg tta tca agc gca Ser Pro Ser Gly Ser Cys Phe Ala Pro Asn Pro Leu Ser Ser Ala 1010 1015 1020	3069
aaa gtg gca ttc atg tac gga gaa ggc cga agc ccg tac tgt ggt Lys Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr Cys Gly 1025 1030 1035	3114
gta ggc ttg ggt cta cat cgt ttg tgg ccc ggt ctc cat gaa aat Val Gly Leu Gly Leu His Arg Leu Trp Pro Gly Leu His Glu Asn 1040 1045 1050	3159

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

cac	ttt	gtg	gga	gag	gtc	tac	tcc	cgg	atc	gca	gac	ttt	cct	caa	3924
His	Phe	Val	Gly	Glu	Val	Tyr	Ser	Arg	Ile	Ala	Asp	Phe	Pro	Gln	
	1295					1300					1305				
atc	gtc	aac	acg	gtg	cat	tcg	gct	ggg	tat	gac	gta	ttt	ctt	gag	3969
Ile	Val	Asn	Thr	Val	His	Ser	Ala	Gly	Tyr	Asp	Val	Phe	Leu	Glu	
	1310					1315					1320				
ctt	ggc	tgt	gat	gct	tct	aga	tct	gca	gca	gtt	caa	aac	att	ctt	4014
Leu	Gly	Cys	Asp	Ala	Ser	Arg	Ser	Ala	Ala	Val	Gln	Asn	Ile	Leu	
	1325					1330					1335				
ggg	ggg	caa	gga	aag	ttc	ttg	tct	aca	gct	att	gac	aaa	aaa	gga	4059
Gly	Gly	Gln	Gly	Lys	Phe	Leu	Ser	Thr	Ala	Ile	Asp	Lys	Lys	Gly	
	1340					1345					1350				
cac	tcc	gcc	tgg	tca	caa	gta	ctt	cgg	gct	acc	gca	tca	tta	gct	4104
His	Ser	Ala	Trp	Ser	Gln	Val	Leu	Arg	Ala	Thr	Ala	Ser	Leu	Ala	
	1355					1360					1365				
gca	cat	cga	gta	ccg	gga	atc	tca	att	ttg	gat	ttg	ttt	cac	cca	4149
Ala	His	Arg	Val	Pro	Gly	Ile	Ser	Ile	Leu	Asp	Leu	Phe	His	Pro	
	1370					1375					1380				
aat	ttc	cga	gaa	atg	tgc	tgt	aca	atg	gca	acc	aca	cct	aaa	gtg	4194
Asn	Phe	Arg	Glu	Met	Cys	Cys	Thr	Met	Ala	Thr	Thr	Pro	Lys	Val	
	1385					1390					1395				
gaa	gat	aag	ttc	ctg	cgc	acg	att	caa	atc	aat	ggg	cgg	ttt	gaa	4239
Glu	Asp	Lys	Phe	Leu	Arg	Thr	Ile	Gln	Ile	Asn	Gly	Arg	Phe	Glu	
	1400					1405					1410				
aaa	gaa	atg	att	cac	cta	gaa	gat	aca	aca	tta	agt	tgc	tta	ccc	4284
Lys	Glu	Met	Ile	His	Leu	Glu	Asp	Thr	Thr	Leu	Ser	Cys	Leu	Pro	
	1415					1420					1425				
gct	cca	agt	gaa	gca	aat	atc	gca	gct	att	caa	tct	cgg	tca	att	4329
Ala	Pro	Ser	Glu	Ala	Asn	Ile	Ala	Ala	Ile	Gln	Ser	Arg	Ser	Ile	
	1430					1435					1440				
cga	tct	gct	gcg	gcg	cgt	tct	gga	caa	tcc	cat	gat	tgt	gca	tcc	4374
Arg	Ser	Ala	Ala	Ala	Arg	Ser	Gly	Gln	Ser	His	Asp	Cys	Ala	Ser	
	1445					1450					1455				
cat	agc	cat	gaa	gaa	aat	aag	gat	tca	tgc	cct	gaa	aag	ctg	aag	4419
His	Ser	His	Glu	Glu	Asn	Lys	Asp	Ser	Cys	Pro	Glu	Lys	Leu	Lys	
	1460					1465					1470				
ctt	gat	tct	gtg	tcc	gtc	gcc	ata	aat	ttc	gac	aat	gat	gac	cgc	4464
Leu	Asp	Ser	Val	Ser	Val	Ala	Ile	Asn	Phe	Asp	Asn	Asp	Asp	Arg	
	1475					1480					1485				
att	cag	ctt	ggg	cac	gcg	ggg	ttt	cgg	gag	atg	tac	aat	aca	aga	4509
Ile	Gln	Leu	Gly	His	Ala	Gly	Phe	Arg	Glu	Met	Tyr	Asn	Thr	Arg	
	1490					1495					1500				
tat	agc	ttg	tac	aca	ggg	gcg	atg	gca	aag	gga	att	gca	tct	gca	4554
Tyr	Ser	Leu	Tyr	Thr	Gly	Ala	Met	Ala	Lys	Gly	Ile	Ala	Ser	Ala	
	1505					1510					1515				
gat	ctt	gtc	att	gcc	gct	ggg	aaa	gag	ggc	atc	cta	gct	tcc	tat	4599
Asp	Leu	Val	Ile	Ala	Ala	Gly	Lys	Glu	Gly	Ile	Leu	Ala	Ser	Tyr	
	1520					1525					1530				

gga gct Gly Ala 1535	gga gga cta cct ctt Gly Gly Leu Pro 1540	gct act gtt cga aag Ala Thr Val Arg Lys 1545	gga ata gac Gly Ile Asp 4644
aaa att Lys Ile 1550	caa caa gcc ttg cca Gln Gln Ala Leu Pro 1555	agt ggc cca tat gct Ser Gly Pro Tyr Ala 1560	gta aat ctt Val Asn Leu 4689
att cac Ile His 1565	tct ccc ttt gac ggc Ser Pro Phe Asp Gly 1570	aac ttg gag cag gga Asn Leu Glu Gln Gly 1575	aac gtc gat Asn Val Asp 4734
ttg ttc Leu Phe 1580	ttg gaa aag aac gtc Leu Glu Lys Asn Val 1585	cgc gtg gcg gaa tgt Arg Val Ala Glu Cys 1590	tcc gcg ttt Ser Ala Phe 4779
aca acg Thr Thr 1595	cta aca gtg cca gta Leu Thr Val Pro Val 1600	gta cac tat cgt gct Val His Tyr Arg Ala 1605	gca ggg ctt Ala Gly Leu 4824
gtt cgg Val Arg 1610	cgc caa gat gga agc Arg Gln Asp Gly Ser 1615	att ttg atc aag aac Ile Leu Ile Lys Asn 1620	cga atc att Arg Ile Ile 4869
gct aaa Ala Lys 1625	gta tct agg aca gaa Val Ser Arg Thr Glu 1630	ctc gct gag atg ttc Leu Ala Glu Met Phe 1635	ctt cgt ccg Leu Arg Pro 4914
gca cct Ala Pro 1640	caa atc atc ctc gaa Gln Ile Ile Leu Glu 1645	aaa ctg gta gca gca Lys Leu Val Ala Ala 1650	gaa atc att Glu Ile Ile 4959
tca tct Ser Ser 1655	gac caa gcg cgt atg Asp Gln Ala Arg Met 1660	gca gcc aaa gtt ccc Ala Ala Lys Val Pro 1665	atg gcg gac Met Ala Asp 5004
gac atc Asp Ile 1670	gca gtc gaa gcc gac Ala Val Glu Ala Asp 1675	tct ggt ggg cac acg Ser Gly Gly His Thr 1680	gat aat cgg Asp Asn Arg 5049
cct atg Pro Met 1685	cac gtc att ttg ccc His Val Ile Leu Pro 1690	ctg ata att caa ctc Leu Ile Ile Gln Leu 1695	cgc aat act Arg Asn Thr 5094
ata ctt Ile Leu 1700	gca gag tat ggc tgt Ala Glu Tyr Gly Cys 1705	gcc acg gct ttt cgt Ala Thr Ala Phe Arg 1710	acc cgt ata Thr Arg Ile 5139
ggc gct Gly Ala 1715	gga gga ggc att ggt Gly Gly Gly Ile Gly 1720	tgt cct tca gcg gcc Cys Pro Ser Ala Ala 1725	ctc gca gcc Leu Ala Ala 5184
ttt gat Phe Asp 1730	atg ggt gcg agt ttt Met Gly Ala Ser Phe 1735	gtc gtg act gga agc Val Val Thr Gly Ser 1740	ata aat caa Ile Asn Gln 5229
att tgc Ile Cys 1745	cgc gag gca ggg act Arg Glu Ala Gly Thr 1750	tgc gat act gtt cgg Cys Asp Thr Val Arg 1755	gag cta ctt Glu Leu Leu 5274
gcc aac Ala Asn 1760	tca agc tac tcg gac Ser Ser Tyr Ser Asp 1765	gtg acg atg gcg cca Val Thr Met Ala Pro 1770	gca gca gac Ala Ala Asp 5319

atg ttt gac caa ggt gtg aaa ctc caa gtc tta aaa cga gga acg	5364
Met Phe Asp Gln Gly Val Lys Leu Gln Val Leu Lys Arg Gly Thr	
1775 1780 1785	
atg ttt cca agc aga gca aat aaa ctc cgg aag ctc ttt gtg aac	5409
Met Phe Pro Ser Arg Ala Asn Lys Leu Arg Lys Leu Phe Val Asn	
1790 1795 1800	
tac gaa tct cta gaa aca ctc ccg tcg aaa gag ttg aaa tac ctg	5454
Tyr Glu Ser Leu Glu Thr Leu Pro Ser Lys Glu Leu Lys Tyr Leu	
1805 1810 1815	
gaa aac atc ata ttc aag caa gca gta gac cag gtg tgg gag gaa	5499
Glu Asn Ile Ile Phe Lys Gln Ala Val Asp Gln Val Trp Glu Glu	
1820 1825 1830	
aca aag cgc ttt tac tgt gaa aaa ctg aac aat cca gat aaa att	5544
Thr Lys Arg Phe Tyr Cys Glu Lys Leu Asn Asn Pro Asp Lys Ile	
1835 1840 1845	
gca agg gcc atg aaa gat cct aaa ttg aag atg tcg ctt tgc ttt	5589
Ala Arg Ala Met Lys Asp Pro Lys Leu Lys Met Ser Leu Cys Phe	
1850 1855 1860	
cgg tgg tat ctc tcc aag agc tct ggg tgg gcc aac gca gga att	5634
Arg Trp Tyr Leu Ser Lys Ser Ser Gly Trp Ala Asn Ala Gly Ile	
1865 1870 1875	
aaa tct cgt gca ctc gac tac cag atc tgg tgt ggc ccg gca atg	5679
Lys Ser Arg Ala Leu Asp Tyr Gln Ile Trp Cys Gly Pro Ala Met	
1880 1885 1890	
ggc tcg ttc aac aat ttc gcc agc ggc aca tcc ctc gat tgg aaa	5724
Gly Ser Phe Asn Asn Phe Ala Ser Gly Thr Ser Leu Asp Trp Lys	
1895 1900 1905	
gtg act ggg gtt ttc cct ggc gtt gcg gaa gta aac atg gcc att	5769
Val Thr Gly Val Phe Pro Gly Val Ala Glu Val Asn Met Ala Ile	
1910 1915 1920	
tta gat ggc gcg cga gaa cta gct gct aaa cga aat taa	5808
Leu Asp Gly Ala Arg Glu Leu Ala Ala Lys Arg Asn	
1925 1930 1935	

<210> 52

5 <211> 1935

<212> PRT

<213> *Thraustochytrium* sp.

10 <220>

<221> MISC_FEATURE

15 <222> (248)..(248)

<223> Xaa puede ser Asp, Gly, Ala o Val

<400> 52

20

ES 2 467 918 T3

Met	Gln	Leu	Pro	Pro	Ala	His	Ser	Ala	Asp	Glu	Asn	Arg	Ile	Ala	Val
1				5					10					15	

Val	Gly	Met	Ala	Val	Lys	Tyr	Ala	Gly	Cys	Asp	Asn	Lys	Glu	Glu	Phe
			20					25					30		

ES 2 467 918 T3

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

ES 2 467 918 T3

Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
290 295 300

Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305 310 315 320

Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
325 330 335

Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
340 345 350

Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys
355 360 365

Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
370 375 380

Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385 390 395 400

Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
405 410 415

Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
420 425 430

Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
435 440 445

Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
450 455 460

Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465 470 475 480

Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
485 490 495

Ala Ala Arg Asp Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp
500 505 510

Leu Glu Phe Leu Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys
515 520 525

Phe Val Glu Ser Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu
530 535 540

ES 2 467 918 T3

Thr Pro Glu Asp Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr
545 550 555 560

Met Asp Arg Ala Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val
565 570 575

Ala Val Leu Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg
580 585 590

Ala Arg Val Ala Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp
595 600 605

Thr Ala Ile Leu Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr
610 615 620

Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile
625 630 635 640

Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly
645 650 655

Asn Asn Ser Val Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln
660 665 670

Val Asn Arg Val Asp Ala Val Val Ile Ala Gly Val Asp Leu Asn Gly
675 680 685

Ser Ala Glu Ser Phe Phe Val Arg Ala Asn Arg Gln Lys Ile Ser Lys
690 695 700

Leu Ser His Pro Cys Ala Ser Phe Asp Arg Asp Ala Asp Gly Phe Phe
705 710 715 720

Ala Gly Glu Gly Cys Gly Ala Leu Val Phe Lys Arg Leu Glu Asp Cys
725 730 735

Ala Pro Gln Glu Lys Ile Tyr Ala Ser Ile Asp Ser Ile Ala Ile Asp
740 745 750

Lys Glu Pro Thr Ser Ser Ala Val Lys Ala Val Tyr Gln Ser Asp Ser
755 760 765

Ser Leu Ser Asp Ile Glu Leu Leu Glu Ile Ser Gly Asp Ser Lys Arg
770 775 780

Phe Ala Ala Phe Glu Gly Ala Val Glu Ile Gln Ser Ser Val Glu Ala
785 790 795 800

ES 2 467 918 T3

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

ES 2 467 918 T3

Val	Asn	Asn	Lys	Thr	Val	Asp	Leu	Trp	Thr	Glu	Gly	Asp	Gly	Trp
1055						1060					1065			
Leu	Tyr	Pro	Arg	Thr	Leu	Thr	Arg	Glu	Glu	His	Thr	Lys	Ala	Ile
1070						1075					1080			
Glu	Ser	Phe	Asn	Ala	Asn	Gln	Ile	Glu	Met	Phe	Arg	Ala	Gly	Ile
1085						1090					1095			
Phe	Ile	Ser	Met	Cys	Gln	Thr	Asp	Tyr	Val	Met	Asn	Val	Leu	Gly
1100						1105					1110			
Val	Gln	Pro	Lys	Ala	Gly	Phe	Gly	Leu	Ser	Leu	Gly	Glu	Ile	Ser
1115						1120					1125			
Met	Leu	Phe	Ala	Met	Ser	Lys	Glu	Asn	Cys	Arg	Gln	Ser	Gln	Glu
1130						1135					1140			
Met	Thr	Asn	Arg	Leu	Arg	Gly	Ser	Pro	Val	Trp	Ser	Asn	Glu	Leu
1145						1150					1155			
Ala	Ile	Asn	Phe	Asn	Ala	Ile	Arg	Lys	Leu	Trp	Lys	Ile	Pro	Arg
1160						1165					1170			
Gly	Ala	Pro	Leu	Glu	Ser	Phe	Trp	Gln	Gly	Tyr	Leu	Val	His	Gly
1175						1180					1185			
Thr	Arg	Glu	Glu	Val	Glu	His	Ala	Ile	Gly	Leu	Ser	Glu	Pro	Tyr
1190						1195					1200			
Val	Arg	Leu	Leu	Ile	Val	Asn	Asp	Ser	Arg	Ser	Ala	Leu	Ile	Ala
1205						1210					1215			
Gly	Lys	Pro	Asp	Ala	Cys	Gln	Ala	Val	Ile	Ser	Arg	Leu	Asn	Ser
1220						1225					1230			
Lys	Phe	Pro	Ser	Leu	Pro	Val	Lys	Gln	Gly	Met	Ile	Gly	His	Cys
1235						1240					1245			
Pro	Glu	Val	Arg	Ala	Phe	Ile	Lys	Asp	Ile	Gly	Tyr	Ile	His	Glu
1250						1255					1260			
Thr	Leu	Arg	Ile	Ser	Asn	Asp	Tyr	Ser	Asp	Cys	Gln	Leu	Phe	Ser
1265						1270					1275			
Ala	Val	Thr	Lys	Gly	Ala	Leu	Asp	Ser	Ser	Thr	Met	Glu	Ile	Lys
1280						1285					1290			

ES 2 467 918 T3

His	Phe	Val	Gly	Glu	Val	Tyr	Ser	Arg	Ile	Ala	Asp	Phe	Pro	Gln
1295						1300					1305			
Ile	Val	Asn	Thr	Val	His	Ser	Ala	Gly	Tyr	Asp	Val	Phe	Leu	Glu
1310						1315					1320			
Leu	Gly	Cys	Asp	Ala	Ser	Arg	Ser	Ala	Ala	Val	Gln	Asn	Ile	Leu
1325						1330					1335			
Gly	Gly	Gln	Gly	Lys	Phe	Leu	Ser	Thr	Ala	Ile	Asp	Lys	Lys	Gly
1340						1345					1350			
His	Ser	Ala	Trp	Ser	Gln	Val	Leu	Arg	Ala	Thr	Ala	Ser	Leu	Ala
1355						1360					1365			
Ala	His	Arg	Val	Pro	Gly	Ile	Ser	Ile	Leu	Asp	Leu	Phe	His	Pro
1370						1375					1380			
Asn	Phe	Arg	Glu	Met	Cys	Cys	Thr	Met	Ala	Thr	Thr	Pro	Lys	Val
1385						1390					1395			
Glu	Asp	Lys	Phe	Leu	Arg	Thr	Ile	Gln	Ile	Asn	Gly	Arg	Phe	Glu
1400						1405					1410			
Lys	Glu	Met	Ile	His	Leu	Glu	Asp	Thr	Thr	Leu	Ser	Cys	Leu	Pro
1415						1420					1425			
Ala	Pro	Ser	Glu	Ala	Asn	Ile	Ala	Ala	Ile	Gln	Ser	Arg	Ser	Ile
1430						1435					1440			
Arg	Ser	Ala	Ala	Ala	Arg	Ser	Gly	Gln	Ser	His	Asp	Cys	Ala	Ser
1445						1450					1455			
His	Ser	His	Glu	Glu	Asn	Lys	Asp	Ser	Cys	Pro	Glu	Lys	Leu	Lys
1460						1465					1470			
Leu	Asp	Ser	Val	Ser	Val	Ala	Ile	Asn	Phe	Asp	Asn	Asp	Asp	Arg
1475						1480					1485			
Ile	Gln	Leu	Gly	His	Ala	Gly	Phe	Arg	Glu	Met	Tyr	Asn	Thr	Arg
1490						1495					1500			
Tyr	Ser	Leu	Tyr	Thr	Gly	Ala	Met	Ala	Lys	Gly	Ile	Ala	Ser	Ala
1505						1510					1515			
Asp	Leu	Val	Ile	Ala	Ala	Gly	Lys	Glu	Gly	Ile	Leu	Ala	Ser	Tyr
1520						1525					1530			

ES 2 467 918 T3

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

Met	Phe	Asp	Gln	Gly	Val	Lys	Leu	Gln	Val	Leu	Lys	Arg	Gly	Thr
1775						1780					1785			
Met	Phe	Pro	Ser	Arg	Ala	Asn	Lys	Leu	Arg	Lys	Leu	Phe	Val	Asn
1790						1795					1800			
Tyr	Glu	Ser	Leu	Glu	Thr	Leu	Pro	Ser	Lys	Glu	Leu	Lys	Tyr	Leu
1805						1810					1815			
Glu	Asn	Ile	Ile	Phe	Lys	Gln	Ala	Val	Asp	Gln	Val	Trp	Glu	Glu
1820						1825					1830			
Thr	Lys	Arg	Phe	Tyr	Cys	Glu	Lys	Leu	Asn	Asn	Pro	Asp	Lys	Ile
1835						1840					1845			
Ala	Arg	Ala	Met	Lys	Asp	Pro	Lys	Leu	Lys	Met	Ser	Leu	Cys	Phe
1850						1855					1860			
Arg	Trp	Tyr	Leu	Ser	Lys	Ser	Ser	Gly	Trp	Ala	Asn	Ala	Gly	Ile
1865						1870					1875			
Lys	Ser	Arg	Ala	Leu	Asp	Tyr	Gln	Ile	Trp	Cys	Gly	Pro	Ala	Met
1880						1885					1890			
Gly	Ser	Phe	Asn	Asn	Phe	Ala	Ser	Gly	Thr	Ser	Leu	Asp	Trp	Lys
1895						1900					1905			
Val	Thr	Gly	Val	Phe	Pro	Gly	Val	Ala	Glu	Val	Asn	Met	Ala	Ile
1910						1915					1920			
Leu	Asp	Gly	Ala	Arg	Glu	Leu	Ala	Ala	Lys	Arg	Asn			
1925						1930					1935			

<210> 53

5 <211> 1500

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1500)

<220>

<221> misc_feature

20

<222> (743)..(743)

<223> n e s a, c, g o t

25 <400> 53

ES 2 467 918 T3

atg	caa	ctt	cct	cca	gcg	cat	tct	gcc	gat	gag	aat	cgc	atc	gcg	gtc	48
Met	Gln	Leu	Pro	Pro	Ala	His	Ser	Ala	Asp	Glu	Asn	Arg	Ile	Ala	Val	
1				5				10					15			

ES 2 467 918 T3

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

ES 2 467 918 T3

atg gtg ctc aag cga ctg aaa gac gct atc aga gat gga aat cac att	864
Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile	
275 280 285	
tat ggt gtg ctc ctt gaa gca aat tta agt aac gca ggt tgt ggg ctt	912
Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu	
290 295 300	
cca ctc agc ccg cac tta ccg agc gaa gaa tca tgt att cgt gat acc	960
Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr	
305 310 315 320	
tac cgc cgt gct gga gtt gct gca gat caa agt att cag tat att gag	1008
Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu	
325 330 335	
tgc cac gct acg gga acc cct cga ggg gat gtc gtg gaa att gag gcg	1056
Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala	
340 345 350	
gtt gaa aga gtt ttc aag aaa aac gtt cca cgc tta ggc tcg acg aaa	1104
Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys	
355 360 365	
gga aat ttt ggt cac tcg tta gtt gcg gct ggt ttc gca ggt atg gca	1152
Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala	
370 375 380	
aag ctt ctt ctt gca atg gaa cat gga gtg att cct ccc aca cca ggt	1200
Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly	
385 390 395 400	
ctt gat gct tcg aac cag gca agt gag cac gtt gtg aca aag gct atc	1248
Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile	
405 410 415	
act tgg cct gag aca cat ggg gct cca aaa cga gct ggc ctt tca gca	1296
Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala	
420 425 430	
ttt gga ttt ggt ggg act aat gcg cat gca ctc ttc gaa gag ttt aat	1344
Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn	
435 440 445	
gcc gag ggc ata agt tat cgc cct gga aag cct cca gtc gaa tcg aat	1392
Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn	
450 455 460	
acc cgt cct tcc gtc gta ata act ggg atg gac tgt acc ttt ggg agc	1440
Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser	
465 470 475 480	
ctt gaa ggg att gat gcg ttc gag act gcc ctg tac gag ggg cgt gac	1488
Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp	
485 490 495	
gca gct cgt gac	1500
Ala Ala Arg Asp	
500	

<210> 54

5 <211> 500

<212> PRT

<213> *Thraustochytrium* sp.

<220>

5 <221> MISC_FEATURE

<222> (248)..(248)

<223> Xaa puede ser Asp, Gly, Ala o Val

10

<400> 54

Met Gln Leu Pro Pro Ala His Ser Ala Asp Glu Asn Arg Ile Ala Val
1 5 10 15

Val Gly Met Ala Val Lys Tyr Ala Gly Cys Asp Asn Lys Glu Glu Phe
20 25 30

Trp Lys Thr Leu Met Asn Gly Ser Ile Asn Thr Lys Ser Ile Ser Ala
35 40 45

Ala Arg Leu Gly Ser Asn Lys Arg Asp Glu His Tyr Val Pro Glu Arg
50 55 60

Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Arg Tyr Gly Cys Ile Gln
65 70 75 80

Gln Gly Thr Asp Asn Glu His Asp Leu Leu Leu Gly Leu Ala Gln Glu
85 90 95

Ala Leu Ala Asp Ala Ala Gly Arg Met Glu Lys Gln Pro Ser Glu Ala
100 105 110

Phe Asp Leu Glu Asn Thr Gly Ile Val Ser Gly Cys Leu Ser Phe Pro
115 120 125

Met Asp Asn Leu Gln Gly Glu Leu Leu Asn Leu Tyr Gln Ser His Val
130 135 140

Glu Lys Gln Leu Pro Pro Ser Ala Leu Val Glu Ala Val Lys Leu Trp
145 150 155 160

Ser Glu Arg Gln Lys Ser Thr Lys Ala His Ala Gly Asp Lys Arg Arg
165 170 175

Phe Ile Asp Pro Ala Ser Phe Val Ala Asp Lys Leu Asn Leu Gly Pro
180 185 190

Leu His Tyr Ala Ile Asp Ala Ala Cys Ala Ser Ala Leu Tyr Val Leu
195 200 205

Lys Leu Ala Gln Asp His Leu Val Ser Gly Ala Val Asp Met Met Leu
210 215 220

ES 2 467 918 T3

Cys Gly Ala Thr Cys Phe Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe
225 230 235 240

Ser Thr Phe Gln Ala Met Pro Xaa Gly Ala Asp Gly Val Ser Leu Pro
245 250 255

Leu His Lys Thr Ser Ala Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile
260 265 270

Met Val Leu Lys Arg Leu Lys Asp Ala Ile Arg Asp Gly Asn His Ile
275 280 285

Tyr Gly Val Leu Leu Glu Ala Asn Leu Ser Asn Ala Gly Cys Gly Leu
290 295 300

Pro Leu Ser Pro His Leu Pro Ser Glu Glu Ser Cys Ile Arg Asp Thr
305 310 315 320

Tyr Arg Arg Ala Gly Val Ala Ala Asp Gln Ser Ile Gln Tyr Ile Glu
325 330 335

Cys His Ala Thr Gly Thr Pro Arg Gly Asp Val Val Glu Ile Glu Ala
340 345 350

Val Glu Arg Val Phe Lys Lys Asn Val Pro Arg Leu Gly Ser Thr Lys
355 360 365

Gly Asn Phe Gly His Ser Leu Val Ala Ala Gly Phe Ala Gly Met Ala
370 375 380

Lys Leu Leu Leu Ala Met Glu His Gly Val Ile Pro Pro Thr Pro Gly
385 390 395 400

Leu Asp Ala Ser Asn Gln Ala Ser Glu His Val Val Thr Lys Ala Ile
405 410 415

Thr Trp Pro Glu Thr His Gly Ala Pro Lys Arg Ala Gly Leu Ser Ala
420 425 430

Phe Gly Phe Gly Gly Thr Asn Ala His Ala Leu Phe Glu Glu Phe Asn
435 440 445

Ala Glu Gly Ile Ser Tyr Arg Pro Gly Lys Pro Pro Val Glu Ser Asn
450 455 460

Thr Arg Pro Ser Val Val Ile Thr Gly Met Asp Cys Thr Phe Gly Ser
465 470 475 480

Leu Glu Gly Ile Asp Ala Phe Glu Thr Ala Leu Tyr Glu Gly Arg Asp
 485 490 495

Ala Ala Arg Asp
 500

<210> 55

5 <211> 1500

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1500)

<400> 55

tta ccc gcc aaa cgt tgg agg ttc cta ggt gag gac ttg gag ttt ctc 48
 Leu Pro Ala Lys Arg Trp Arg Phe Leu Gly Glu Asp Leu Glu Phe Leu
 1 5 10 15

cga gcc atc agg ctc aag gaa aag cct agg ggt tgt ttt gtg gag agt 96
 Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys Phe Val Glu Ser
 20 25 30

gtt gac gtt aac ttt aga cgg ctg aaa acg ccc ttg aca cca gaa gat 144
 Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu Thr Pro Glu Asp
 35 40 45

atg ttg cgg ccc caa caa ctc ttg gcg gtt tct acg atg gac cga gca 192
 Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr Met Asp Arg Ala
 50 55 60

att atc gat gca ggt cta aag aag ggc caa cat gta gca gtt ctt gtt 240
 Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val Ala Val Leu Val
 65 70 75 80

ggc cta gga act gac ctg gaa ctt tac cgt cat cga gca aga gtc gcg 288
 Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala
 85 90 95

ctt aaa gag gtt ttg cac ccg agc tta aag tca gac act gca att ctc 336
 Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp Thr Ala Ile Leu
 100 105 110

cag aaa ata atg caa tat gtg aat gat gca gga act tcg act tca tac 384
 Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr Ser Thr Ser Tyr
 115 120 125

aca tct tac att gga aac ctc gtt gcc acg cgt att tcg tct cag tgg 432
 Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile Ser Ser Gln Trp
 130 135 140

gga ttc aca ggg ccg tcc ttt act gtc aca gaa gga aat aat tcc gtg 480
 Gly Phe Thr Gly Pro Ser Phe Thr Val Thr Glu Gly Asn Asn Ser Val
 145 150 155 160

tac aga tgt gca caa cta gcc aaa gat atg ctt cag gtt aac cga gtt 528
 Tyr Arg Cys Ala Gln Leu Ala Lys Asp Met Leu Gln Val Asn Arg Val
 165 170 175

ES 2 467 918 T3

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

ES 2 467 918 T3

gga aaa act att gtg cgg cga ctt gaa gga tat ctc ctt gaa ctt gtt 1344
Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu Leu Glu Leu Val
435 440 445

gaa ggg cat gca agc cct tca gag tat ttc cac aaa ctg att gga caa 1392
Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys Leu Ile Gly Gln
450 455 460

agt cta ctt gag aac tcg aaa gaa agt aaa ctc aca ctt tcg ctt gtg 1440
Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr Leu Ser Leu Val
465 470 475 480

tgc aat ccg aac cag ctc caa aag gag ctc atg ctt gct atc aaa gga 1488
Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu Ala Ile Lys Gly
485 490 495

gta caa cga agc 1500
Val Gln Arg Ser
500

<210> 56

5 <211> 500

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 56

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

Arg Ala Ile Arg Leu Lys Glu Lys Pro Arg Gly Cys Phe Val Glu Ser
20 25 30

Val Asp Val Asn Phe Arg Arg Leu Lys Thr Pro Leu Thr Pro Glu Asp
35 40 45

Met Leu Arg Pro Gln Gln Leu Leu Ala Val Ser Thr Met Asp Arg Ala
50 55 60

Ile Ile Asp Ala Gly Leu Lys Lys Gly Gln His Val Ala Val Leu Val
65 70 75 80

Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala
85 90 95

Leu Lys Glu Val Leu His Pro Ser Leu Lys Ser Asp Thr Ala Ile Leu
100 105 110

Gln Lys Ile Met Gln Tyr Val Asn Asp Ala Gly Thr Ser Thr Ser Tyr
115 120 125

Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Ile Ser Ser Gln Trp
130 135 140

ES 2 467 918 T3

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

Ser Leu Val Leu Ser Asp Thr Gly Cys Tyr Glu Glu His Asn Arg Thr
405 410 415

Cys Phe Asp Val Gln Ala Pro Gln Leu Val Leu Ile His Gly Phe Asp
420 425 430

Gly Lys Thr Ile Val Arg Arg Leu Glu Gly Tyr Leu Leu Glu Leu Val
435 440 445

Glu Gly His Ala Ser Pro Ser Glu Tyr Phe His Lys Leu Ile Gly Gln
450 455 460

Ser Leu Leu Glu Asn Ser Lys Glu Ser Lys Leu Thr Leu Ser Leu Val
465 470 475 480

Cys Asn Pro Asn Gln Leu Gln Lys Glu Leu Met Leu Ala Ile Lys Gly
485 490 495

Val Gln Arg Ser
500

<210> 57

5 <211> 1500

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1500)

<400> 57

atg tta aca ggg aag gat tgg gtc agt cca tca gga agt tgt ttt gcc 48
Met Leu Thr Gly Lys Asp Trp Val Ser Pro Ser Gly Ser Cys Phe Ala
1 5 10 15

cca aat ccg tta tca agc gca aaa gtg gca ttc atg tac gga gaa ggc 96
Pro Asn Pro Leu Ser Ser Ala Lys Val Ala Phe Met Tyr Gly Glu Gly
20 25 30

cga agc ccg tac tgt ggt gta ggc ttg ggt cta cat cgt ttg tgg ccc 144
Arg Ser Pro Tyr Cys Gly Val Gly Leu Gly Leu His Arg Leu Trp Pro
35 40 45

ggt ctc cat gaa aat gtg aac aat aag aca gtc gat tta tgg acg gaa 192
Gly Leu His Glu Asn Val Asn Asn Lys Thr Val Asp Leu Trp Thr Glu
50 55 60

gga gat ggt tgg tta tat cct cga acg ttg aca cga gaa gag cat aca 240
Gly Asp Gly Trp Leu Tyr Pro Arg Thr Leu Thr Arg Glu Glu His Thr
65 70 75 80

aaa gcc atc gaa tct ttc aac gca aat caa att gaa atg ttt cgc gct 288
Lys Ala Ile Glu Ser Phe Asn Ala Asn Gln Ile Glu Met Phe Arg Ala
85 90 95

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

<211> 500

<212> PRT

5 <213> *Thraustochytrium* sp.

<400> 58

Met	Leu	Thr	Gly	Lys	Asp	Trp	Val	Ser	Pro	Ser	Gly	Ser	Cys	Phe	Ala
1				5				10					15		

Pro	Asn	Pro	Leu	Ser	Ser	Ala	Lys	Val	Ala	Phe	Met	Tyr	Gly	Glu	Gly
		20					25						30		

Arg	Ser	Pro	Tyr	Cys	Gly	Val	Gly	Leu	Gly	Leu	His	Arg	Leu	Trp	Pro
	35					40					45				

Gly	Leu	His	Glu	Asn	Val	Asn	Asn	Lys	Thr	Val	Asp	Leu	Trp	Thr	Glu
50						55				60					

ES 2 467 918 T3

Gly Asp Gly Trp Leu Tyr Pro Arg Thr Leu Thr Arg Glu Glu His Thr
 65 70 75 80

Lys Ala Ile Glu Ser Phe Asn Ala Asn Gln Ile Glu Met Phe Arg Ala
 85 90 95

Gly Ile Phe Ile Ser Met Cys Gln Thr Asp Tyr Val Met Asn Val Leu
 100 105 110

Gly Val Gln Pro Lys Ala Gly Phe Gly Leu Ser Leu Gly Glu Ile Ser
 115 120 125

Met Leu Phe Ala Met Ser Lys Glu Asn Cys Arg Gln Ser Gln Glu Met
 130 135 140

Thr Asn Arg Leu Arg Gly Ser Pro Val Trp Ser Asn Glu Leu Ala Ile
 145 150 155 160

Asn Phe Asn Ala Ile Arg Lys Leu Trp Lys Ile Pro Arg Gly Ala Pro
 165 170 175

Leu Glu Ser Phe Trp Gln Gly Tyr Leu Val His Gly Thr Arg Glu Glu
 180 185 190

Val Glu His Ala Ile Gly Leu Ser Glu Pro Tyr Val Arg Leu Leu Ile
 195 200 205

Val Asn Asp Ser Arg Ser Ala Leu Ile Ala Gly Lys Pro Asp Ala Cys
 210 215 220

Gln Ala Val Ile Ser Arg Leu Asn Ser Lys Phe Pro Ser Leu Pro Val
 225 230 235 240

Lys Gln Gly Met Ile Gly His Cys Pro Glu Val Arg Ala Phe Ile Lys
 245 250 255

Asp Ile Gly Tyr Ile His Glu Thr Leu Arg Ile Ser Asn Asp Tyr Ser
 260 265 270

Asp Cys Gln Leu Phe Ser Ala Val Thr Lys Gly Ala Leu Asp Ser Ser
 275 280 285

Thr Met Glu Ile Lys His Phe Val Gly Glu Val Tyr Ser Arg Ile Ala
 290 295 300

Asp Phe Pro Gln Ile Val Asn Thr Val His Ser Ala Gly Tyr Asp Val
 305 310 315 320

Phe Leu Glu Leu Gly Cys Asp Ala Ser Arg Ser Ala Ala Val Gln Asn
325 330 335

Ile Leu Gly Gly Gln Gly Lys Phe Leu Ser Thr Ala Ile Asp Lys Lys
340 345 350

Gly His Ser Ala Trp Ser Gln Val Leu Arg Ala Thr Ala Ser Leu Ala
355 360 365

Ala His Arg Val Pro Gly Ile Ser Ile Leu Asp Leu Phe His Pro Asn
370 375 380

Phe Arg Glu Met Cys Cys Thr Met Ala Thr Thr Pro Lys Val Glu Asp
385 390 395 400

Lys Phe Leu Arg Thr Ile Gln Ile Asn Gly Arg Phe Glu Lys Glu Met
405 410 415

Ile His Leu Glu Asp Thr Thr Leu Ser Cys Leu Pro Ala Pro Ser Glu
420 425 430

Ala Asn Ile Ala Ala Ile Gln Ser Arg Ser Ile Arg Ser Ala Ala Ala
435 440 445

Arg Ser Gly Gln Ser His Asp Cys Ala Ser His Ser His Glu Glu Asn
450 455 460

Lys Asp Ser Cys Pro Glu Lys Leu Lys Leu Asp Ser Val Ser Val Ala
465 470 475 480

Ile Asn Phe Asp Asn Asp Asp Arg Ile Gln Leu Gly His Ala Gly Phe
485 490 495

Arg Glu Met Tyr
500

<210> 59

5 <211> 1305

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1305)

<400> 59

ES 2 467 918 T3

aat	aca	aga	tat	agc	ttg	tac	aca	ggg	gcg	atg	gca	aag	gga	att	gca	48
Asn	Thr	Arg	Tyr	Ser	Leu	Tyr	Thr	Gly	Ala	Met	Ala	Lys	Gly	Ile	Ala	
1				5				10						15		
tct	gca	gat	ctt	gtc	att	gcc	gct	ggg	aaa	gag	ggc	atc	cta	gct	tcc	96

ES 2 467 918 T3

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

ES 2 467 918 T3

Asp	Met	Phe	Asp	Gln	Gly	Val	Lys	Leu	Gln	Val	Leu	Lys	Arg	Gly	Thr		
		275					280					285					
atg	ttt	cca	agc	aga	gca	aat	aaa	ctc	cgg	aag	ctc	ttt	gtg	aac	tac		912
Met	Phe	Pro	Ser	Arg	Ala	Asn	Lys	Leu	Arg	Lys	Leu	Phe	Val	Asn	Tyr		
	290					295					300						
gaa	tct	cta	gaa	aca	ctc	ccg	tcg	aaa	gag	ttg	aaa	tac	ctg	gaa	aac		960
Glu	Ser	Leu	Glu	Thr	Leu	Pro	Ser	Lys	Glu	Leu	Lys	Tyr	Leu	Glu	Asn		
305					310					315					320		
atc	ata	ttc	aag	caa	gca	gta	gac	cag	gtg	tgg	gag	gaa	aca	aag	cgc		1008
Ile	Ile	Phe	Lys	Gln	Ala	Val	Asp	Gln	Val	Trp	Glu	Glu	Thr	Lys	Arg		
				325					330					335			
ttt	tac	tgt	gaa	aaa	ctg	aac	aat	cca	gat	aaa	att	gca	agg	gcc	atg		1056
Phe	Tyr	Cys	Glu	Lys	Leu	Asn	Asn	Pro	Asp	Lys	Ile	Ala	Arg	Ala	Met		
			340					345					350				
aaa	gat	cct	aaa	ttg	aag	atg	tcg	ctt	tgc	ttt	cgg	tgg	tat	ctc	tcc		1104
Lys	Asp	Pro	Lys	Leu	Lys	Met	Ser	Leu	Cys	Phe	Arg	Trp	Tyr	Leu	Ser		
		355				360						365					
aag	agc	tct	ggg	tgg	gcc	aac	gca	gga	att	aaa	tct	cgt	gca	ctc	gac		1152
Lys	Ser	Ser	Gly	Trp	Ala	Asn	Ala	Gly	Ile	Lys	Ser	Arg	Ala	Leu	Asp		
	370					375					380						
tac	cag	atc	tgg	tgt	ggc	ccg	gca	atg	ggc	tcg	ttc	aac	aat	ttc	gcc		1200
Tyr	Gln	Ile	Trp	Cys	Gly	Pro	Ala	Met	Gly	Ser	Phe	Asn	Asn	Phe	Ala		
385					390					395					400		
agc	ggc	aca	tcc	ctc	gat	tgg	aaa	gtg	act	ggg	gtt	ttc	cct	ggc	gtt		1248
Ser	Gly	Thr	Ser	Leu	Asp	Trp	Lys	Val	Thr	Gly	Val	Phe	Pro	Gly	Val		
				405					410					415			
gcg	gaa	gta	aac	atg	gcc	att	tta	gat	ggc	gcg	cga	gaa	cta	gct	gct		1296
Ala	Glu	Val	Asn	Met	Ala	Ile	Leu	Asp	Gly	Ala	Arg	Glu	Leu	Ala	Ala		
			420					425					430				
aaa	cga	aat															1305
Lys	Arg	Asn															
		435															

<210> 60

5 <211> 435

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 60

ES 2 467 918 T3

Asn Thr Arg Tyr Ser Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
1 5 10 15

Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Glu Gly Ile Leu Ala Ser
20 25 30

Tyr Gly Ala Gly Gly Leu Pro Leu Ala Thr Val Arg Lys Gly Ile Asp
35 40 45

Lys Ile Gln Gln Ala Leu Pro Ser Gly Pro Tyr Ala Val Asn Leu Ile

ES 2 467 918 T3

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

ES 2 467 918 T3

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

<210> 61

5 <211> 4410

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(4410)

<400> 61

ES 2 467 918 T3

atg ggc ccg cga gtg gcg tca ggc aag gtg ccg gct tgg gag atg agc	48
Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser	
1 5 10 15	
aag tcc gag ctg tgt gat gac cgc acg gta gtc ttt gac tat gag gag	96
Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu	
20 25 30	
ctg ctg gag ttc gct gag ggc gat atc agt aag gtt ttt ggg ccg gag	144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu	
35 40 45	
ttc aaa gtg gtg gac ggg ttt agg cgc agg gtg agg ttg ccc gct cga	192
Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg	
50 55 60	
gag tac ctg ctg gtg acc cgg gtt acg ctg atg gat gcc gag gtg ggc	240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly	

ES 2 467 918 T3

65	70					75					80					
aac ttt cga gtg gga gca cgt atg gtg aca gag tat gac gta cct gtg	Asn Phe Arg Val Gly	Ala Arg Met Val Thr	Glu Tyr Asp Val Pro	Val												288
	85					90					95					
aac gga gag ctc tcg gaa ggg gga gat gtg ccg tgg gct gtg ttg gtg	Asn Gly Glu Leu Ser	Glu Gly Gly Asp	Val Pro Trp Ala Val	Leu Val												336
	100					105					110					
gaa gcc ggg cag tgc gac ttg ctg cta att tct tac atg ggc atc gat	Glu Ala Gly Gln Cys	Asp Leu Leu Leu	Ile Ser Tyr Met Gly	Ile Asp												384
	115					120					125					
ttc cag tgc aaa gga gag cgg gtc tac cgg ctg ctg aac acc acc ttg	Phe Gln Cys Lys Gly	Glu Arg Val Tyr	Arg Leu Leu Asn Thr	Thr Leu												432
	130					135					140					
acg ttt ttt ggc gtc gcg aaa gaa ggg gaa acg ctt gtg tac gat att	Thr Phe Phe Gly Val	Ala Lys Glu Gly Glu	Thr Leu Val Tyr Asp	Ile												480
	145					150					155					160
cgc gtc acg ggt ttc gcc aag agg ccg gac gga gat atc tcc atg ttc	Arg Val Thr Gly Phe	Ala Lys Arg Pro	Asp Gly Asp Ile Ser	Met Phe												528
						165					170					175
ttt ttc gaa tat gat tgc tac tgc aat ggc aag ctt ctc atc gaa atg	Phe Phe Glu Tyr Asp	Cys Tyr Cys Asn	Gly Lys Leu Leu Ile	Glu Met												576
						180										190
cga gat ggc tct gca ggc ttc ttc acg gac gaa gag ctc gct gcc ggc	Arg Asp Gly Ser Ala	Gly Phe Phe Thr	Asp Glu Glu Leu Ala	Ala Gly												624
						195										200
aaa gga gtg gtc gtc act cgt gca cag caa aac atg cgg gac aaa att	Lys Gly Val Val Val	Thr Arg Ala Gln Gln	Asn Met Arg Asp Lys	Ile												672
						210										215
gta cgg cag tcc att gag cct ttt gca ctg gcg gct tgc acg cac aaa	Val Arg Gln Ser Ile	Glu Pro Phe Ala Leu	Ala Ala Cys Thr His	Lys												720
						225										230
acg act ctg aac gag agt gac atg cag tcc ctt gtg gag cga aac tgg	Thr Thr Leu Asn Glu	Ser Asp Met Gln Ser	Leu Val Glu Arg Asn	Trp												768
						245										250
gca aac gtt ttt ggc acc agt aac aag atg gcg gag ctc aac tat aaa	Ala Asn Val Phe Gly	Thr Ser Asn Lys Met	Ala Glu Leu Asn Tyr	Lys												816
						260										265
att tgc gcc agg aaa atg ctc atg atc gac agg gtt acc cac att gac	Ile Cys Ala Arg Lys	Met Leu Met Ile Asp	Arg Val Thr His Ile	Asp												864
						275										280
cac cac ggt ggg gcg tat ggc ctc gga cta ctt gtt gga gag aag atc	His His Gly Gly Ala	Tyr Gly Leu Gly Leu	Leu Val Gly Glu Lys	Ile												912
						290										295
ttg gat cga aac cat tgg tac ttt cct tgt cac ttt gtc aat gat caa	Leu Asp Arg Asn His	Trp Tyr Phe Pro Cys	His Phe Val Asn Asp	Gln												960
						305										310
gtc atg gca ggg tca ctg gtc agc gat ggt tgc agc cag ctc tta aaa	Val Met Ala Gly Ser	Leu Val Ser Asp Gly	Cys Ser Gln Leu Leu	Lys												1008
						315										320

ES 2 467 918 T3

325								330				335				
ctc	tat	atg	atc	tgg	ctt	ggc	ctc	cac	ctg	aaa	atg	gag	gaa	ttt	gat	1056
Leu	Tyr	Met	Ile	Trp	Leu	Gly	Leu	His	Leu	Lys	Met	Glu	Glu	Phe	Asp	
340								345				350				
ttt	ctc	cca	gtt	agc	ggc	cac	aaa	aac	aag	gtg	cga	tgc	agg	gga	caa	1104
Phe	Leu	Pro	Val	Ser	Gly	His	Lys	Asn	Lys	Val	Arg	Cys	Arg	Gly	Gln	
355								360				365				
att	tca	ccg	cat	aaa	ggc	aag	ctt	gtc	tac	gtc	atg	gaa	atc	aaa	aag	1152
Ile	Ser	Pro	His	Lys	Gly	Lys	Leu	Val	Tyr	Val	Met	Glu	Ile	Lys	Lys	
370								375				380				
atg	ggc	tac	gat	caa	gca	tct	gga	agc	cca	tac	gcc	atc	gcg	gac	gtt	1200
Met	Gly	Tyr	Asp	Gln	Ala	Ser	Gly	Ser	Pro	Tyr	Ala	Ile	Ala	Asp	Val	
385				390				395				400				
gat	atc	att	gac	gtc	aac	gaa	gag	ctg	ggc	caa	agt	ttt	gac	atc	aac	1248
Asp	Ile	Ile	Asp	Val	Asn	Glu	Glu	Leu	Gly	Gln	Ser	Phe	Asp	Ile	Asn	
				405				410				415				
gac	ctt	gcg	agc	tac	gga	aaa	ggc	gac	ctg	agc	aaa	aaa	atc	gtg	gtt	1296
Asp	Leu	Ala	Ser	Tyr	Gly	Lys	Gly	Asp	Leu	Ser	Lys	Lys	Ile	Val	Val	
				420				425				430				
gac	ttc	aaa	gga	att	gct	ttg	cag	ctc	aaa	ggc	cgc	gct	ttt	tca	cgc	1344
Asp	Phe	Lys	Gly	Ile	Ala	Leu	Gln	Leu	Lys	Gly	Arg	Ala	Phe	Ser	Arg	
435								440				445				
atg	agt	tcc	agc	tcg	tcc	ttg	aac	gaa	gga	tgg	caa	tgt	gtt	cca	aaa	1392
Met	Ser	Ser	Ser	Ser	Ser	Leu	Asn	Glu	Gly	Trp	Gln	Cys	Val	Pro	Lys	
450								455				460				
cca	agc	cag	aga	atg	gaa	cac	gaa	cag	ccc	cct	gct	cac	tgc	ctt	gca	1440
Pro	Ser	Gln	Arg	Met	Glu	His	Glu	Gln	Pro	Pro	Ala	His	Cys	Leu	Ala	
465				470				475				480				
agc	gac	ccc	gaa	gcc	cct	tca	act	gtg	acc	tgg	cac	cca	atg	tca	aag	1488
Ser	Asp	Pro	Glu	Ala	Pro	Ser	Thr	Val	Thr	Trp	His	Pro	Met	Ser	Lys	
				485				490				495				
ctt	cct	ggc	aac	cct	acg	ccg	ttc	ttc	tcc	cct	tca	tct	tac	cct	ccg	1536
Leu	Pro	Gly	Asn	Pro	Thr	Pro	Phe	Phe	Ser	Pro	Ser	Ser	Tyr	Pro	Pro	
				500				505				510				
agg	gca	att	tgc	ttc	atc	cct	ttc	ccg	ggc	aat	ccc	ctt	gac	aac	aac	1584
Arg	Ala	Ile	Cys	Phe	Ile	Pro	Phe	Pro	Gly	Asn	Pro	Leu	Asp	Asn	Asn	
515								520				525				
tgc	aag	gct	gga	gaa	atg	ccc	ctg	aac	tgg	tac	aac	atg	tca	gag	ttc	1632
Cys	Lys	Ala	Gly	Glu	Met	Pro	Leu	Asn	Trp	Tyr	Asn	Met	Ser	Glu	Phe	
530								535				540				
atg	tgt	ggc	aag	gtt	tct	aac	tgc	ttg	ggc	cca	gaa	ttc	gca	cgc	ttt	1680
Met	Cys	Gly	Lys	Val	Ser	Asn	Cys	Leu	Gly	Pro	Glu	Phe	Ala	Arg	Phe	
545				550				555				560				
gac	aag	tcg	aac	acc	agc	cgg	agc	cct	gct	ttt	gac	ttg	gct	ctg	gtg	1728
Asp	Lys	Ser	Asn	Thr	Ser	Arg	Ser	Pro	Ala	Phe	Asp	Leu	Ala	Leu	Val	
				565				570				575				
acc	cga	gtt	gtt	gaa	gtc	aca	aac	atg	gaa	cac	ggc	aag	ttt	cta	aac	1776
Thr	Arg	Val	Val	Glu	Val	Thr	Asn	Met	Glu	His	Gly	Lys	Phe	Leu	Asn	

ES 2 467 918 T3

580	585	590	
gtt gat tgc aat cca agc aaa ggc aca atg gtg ggg gag ttt gac tgt Val Asp Cys Asn Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys 595 600 605			1824
ccc caa gac gcg tgg ttc ttt gat ggt tcg tgc aac gac ggc cat atg Pro Gln Asp Ala Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met 610 615 620			1872
ccg tat tcc att atc atg gaa atc gga ctg caa acc tca ggt gtt ctc Pro Tyr Ser Ile Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu 625 630 635 640			1920
acc tcg gtg ttg aag gca ccg ctg act atg gac aag gat gac att ctc Thr Ser Val Leu Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu 645 650 655			1968
ttt cga aac ctc gat gca agt gct gaa atg gtg cgt cca gac gtg gat Phe Arg Asn Leu Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp 660 665 670			2016
gtt cgc ggc aaa acg att cga aac gtg acc aag tgt acc ggc tat gca Val Arg Gly Lys Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala 675 680 685			2064
atg ttg gga aag atg ggg att cac cgg ttc acg ttt gag ttg agc gtt Met Leu Gly Lys Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val 690 695 700			2112
gac ggc gtg gta ttt tat aaa gga tcc act tcc ttt gga tgg ttc act Asp Gly Val Val Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr 705 710 715 720			2160
ccc gag gtg ttt gct cag caa gct gga ctc gac aac ggg aaa aag acg Pro Glu Val Phe Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr 725 730 735			2208
gag ccc tgg tgc aag act aac aac acc tcg gtt cga aga gtt gaa atc Glu Pro Trp Cys Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile 740 745 750			2256
gca tcc gcc aaa gga aaa gag cag ctg act gag aag ctt ccc gac gca Ala Ser Ala Lys Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala 755 760 765			2304
act aat gct caa gtt ctt cgg cgt tca gag cag tgt gaa tac ctc gat Thr Asn Ala Gln Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp 770 775 780			2352
tac ctc aat att gcc cct gac tct ggg ctg cat ggg aag ggc tac gcc Tyr Leu Asn Ile Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala 785 790 795 800			2400
cac gga cac aaa gac gtt aac ccg caa gac tgg ttc ttc tct tgc cac His Gly His Lys Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His 805 810 815			2448
ttt tgg ttc gat cct gta atg cca gga tct tta gga att gaa tca atg Phe Trp Phe Asp Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met 820 825 830			2496
ttc cag ctt atc gag gcc ttt gcg gtg gac caa aac att cct gga gag Phe Gln Leu Ile Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu			2544

ES 2 467 918 T3

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

1085	1090	1095	
tgc agc gcg ttc atg gaa ctc acg gcc cag gtt gtc cgg tac cgc Cys Ser Ala Phe Met Glu Leu Thr Ala Gln Val Val Arg Tyr Arg 1100 1105 1110			3339
gcg tct ggt cta agc aaa agt gcg gac ggt tcg att cgc att gct Ala Ser Gly Leu Ser Lys Ser Ala Asp Gly Ser Ile Arg Ile Ala 1115 1120 1125			3384
cac cgt att att ggc aag gtt tcc aga acc gag ctg gca gaa atg His Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu Ala Glu Met 1130 1135 1140			3429
ttt att cgt cca gca cca cag cac ctc ctc caa aaa ctc gta gcc Phe Ile Arg Pro Ala Pro Gln His Leu Leu Gln Lys Leu Val Ala 1145 1150 1155			3474
tcc ggc gag ctg aca gct gag caa gcc gag ctt gca aca cag gtt Ser Gly Glu Leu Thr Ala Glu Gln Ala Glu Leu Ala Thr Gln Val 1160 1165 1170			3519
ccg gtg gcg gat gac att gcg gtc gaa gcc gac tcg ggg ggg cat Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His 1175 1180 1185			3564
acc gac aac agg cct att cac gtc att ctt cct cta atc atc aac Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn 1190 1195 1200			3609
cta cgc aac cgt ttg cat aaa gag ctt gac tac cct tcg cat ctc Leu Arg Asn Arg Leu His Lys Glu Leu Asp Tyr Pro Ser His Leu 1205 1210 1215			3654
cgg gta cgt gtg ggt gct ggt ggt ggt att gga tgt cct caa gcc Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala 1220 1225 1230			3699
gct ctt gca gca ttt caa atg ggg gca gcg ttt tta atc act gga Ala Leu Ala Ala Phe Gln Met Gly Ala Ala Phe Leu Ile Thr Gly 1235 1240 1245			3744
acg gtg aac cag ctt gct cgt gaa agt ggc act tgt gac aac gtc Thr Val Asn Gln Leu Ala Arg Glu Ser Gly Thr Cys Asp Asn Val 1250 1255 1260			3789
cgg tta cag ctc tca aag gcc acg tat agc gac gtg tgt atg gct Arg Leu Gln Leu Ser Lys Ala Thr Tyr Ser Asp Val Cys Met Ala 1265 1270 1275			3834
cct gct gcc gat atg ttt gac caa ggc gtg gag ctg caa gta ttg Pro Ala Ala Asp Met Phe Asp Gln Gly Val Glu Leu Gln Val Leu 1280 1285 1290			3879
aag aaa ggc acg ctg ttc cca agt cgt gct aag aag ctg tac gag Lys Lys Gly Thr Leu Phe Pro Ser Arg Ala Lys Lys Leu Tyr Glu 1295 1300 1305			3924
ctg ttc tgc aag tat gac tcg ttt gag gca atg ccg gct gaa gaa Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ala Met Pro Ala Glu Glu 1310 1315 1320			3969
ttg caa cgg gtt gaa aag cgg att ttt caa aag tcg ctt gct gaa Leu Gln Arg Val Glu Lys Arg Ile Phe Gln Lys Ser Leu Ala Glu 4014			

<210> 62

<211> 1470

<212> PRT

5

<213> *Thraustochytrium* sp.

<400> 62

Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1 5 10 15

Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
35 40 45

Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

10 Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Gly
65 70 75 80

ES 2 467 918 T3

Asn Phe Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Val Pro Val
 85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Val Pro Trp Ala Val Leu Val
 100 105 110

Glu Ala Gly Gln Cys Asp Leu Leu Leu Ile Ser Tyr Met Gly Ile Asp
 115 120 125

Phe Gln Cys Lys Gly Glu Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
 130 135 140

Thr Phe Phe Gly Val Ala Lys Glu Gly Glu Thr Leu Val Tyr Asp Ile
 145 150 155 160

Arg Val Thr Gly Phe Ala Lys Arg Pro Asp Gly Asp Ile Ser Met Phe
 165 170 175

Phe Phe Glu Tyr Asp Cys Tyr Cys Asn Gly Lys Leu Leu Ile Glu Met
 180 185 190

Arg Asp Gly Ser Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Ala Gly
 195 200 205

Lys Gly Val Val Val Thr Arg Ala Gln Gln Asn Met Arg Asp Lys Ile
 210 215 220

Val Arg Gln Ser Ile Glu Pro Phe Ala Leu Ala Ala Cys Thr His Lys
 225 230 235 240

Thr Thr Leu Asn Glu Ser Asp Met Gln Ser Leu Val Glu Arg Asn Trp
 245 250 255

Ala Asn Val Phe Gly Thr Ser Asn Lys Met Ala Glu Leu Asn Tyr Lys
 260 265 270

Ile Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr His Ile Asp
 275 280 285

His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
 290 295 300

Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
 305 310 315 320

Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
 325 330 335

ES 2 467 918 T3

Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
 340 345 350
 Phe Leu Pro Val Ser Gly His Lys 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 Lys
 370 375 380
 Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
 385 390 395 400
 Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
 405 410 415
 Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
 420 425 430
 Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
 435 440 445
 Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
 450 455 460
 Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
 465 470 475 480
 Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
 485 490 495
 Leu Pro Gly Asn Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro
 500 505 510
 Arg Ala Ile Cys Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn
 515 520 525
 Cys Lys Ala Gly Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe
 530 535 540
 Met Cys Gly Lys Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe
 545 550 555 560
 Asp Lys Ser Asn Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val
 565 570 575
 Thr Arg Val Val Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn
 580 585 590

ES 2 467 918 T3

Val Asp Cys Asn Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys
 595 600 605
 Pro Gln Asp Ala Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met
 610 615 620
 Pro Tyr Ser Ile Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu
 625 630 635 640
 Thr Ser Val Leu Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu
 645 650 655
 Phe Arg Asn Leu Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp
 660 665 670
 Val Arg Gly Lys Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala
 675 680 685
 Met Leu Gly Lys Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val
 690 695 700
 Asp Gly Val Val Phe Tyr Lys Gly Ser Thr Ser Phe Gly Trp Phe Thr
 705 710 715 720
 Pro Glu Val Phe Ala Gln Gln Ala Gly Leu Asp Asn Gly Lys Lys Thr
 725 730 735
 Glu Pro Trp Cys Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile
 740 745 750
 Ala Ser Ala Lys Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala
 755 760 765
 Thr Asn Ala Gln Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp
 770 775 780
 Tyr Leu Asn Ile Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala
 785 790 795 800
 His Gly His Lys Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His
 805 810 815
 Phe Trp Phe Asp Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met
 820 825 830
 Phe Gln Leu Ile Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu
 835 840 845

Tyr Asn Val Ser Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala
 850 855 860

Trp Lys Tyr Arg Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys
 865 870 875 880

Glu Val His Ile Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr
 885 890 895

Val Asp Ile Val Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val
 900 905 910

Tyr Glu Ala Lys Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln
 915 920 925

Ala Ile Pro Asp Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro
 930 935 940

Gly Lys Thr Gly Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu
 945 950 955 960

Leu Glu Val Asp Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu
 965 970 975

Val Gln Phe Glu Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile
 980 985 990

Lys Pro Cys Ser Ile Ser Asp Leu Gly Asp Lys Ser Phe Met Glu Thr
 995 1000 1005

Tyr Asn Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly
 1010 1015 1020

Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Arg Lys Ile
 1025 1030 1035

Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro Ile Ser Ile Val Arg
 1040 1045 1050

Glu Ala Leu Glu Lys Ile Gln Gln His Leu Pro His Gly Pro Tyr
 1055 1060 1065

Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn Leu Glu Lys
 1070 1075 1080

Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val Val Glu
 1085 1090 1095

ES 2 467 918 T3

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

Val Trp Gln Glu Thr Ser Asp Phe Tyr Ile His Arg Ile Lys Asn
1340 1345 1350

Pro Glu Lys Ile Asn Arg Ala Ala Ser Asp Gly Lys Leu Lys Met
1355 1360 1365

Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ser Ser Phe Trp Ala
1370 1375 1380

Asn Ser Gly Ala Gln Asp Arg Val Met Asp Tyr Gln Ile Trp Cys
1385 1390 1395

Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Thr Lys Gly Thr Tyr
1400 1405 1410

Leu Asp Val Thr Val Ala Lys Ser Tyr Pro Cys Val Ala Gln Ile
1415 1420 1425

Asn Leu Gln Ile Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly
1430 1435 1440

Val Ile Arg Phe Asp Arg Met Leu Leu Gln Ala Val Asp Ile Asp
1445 1450 1455

Asp Pro Val Phe Thr Tyr Val Pro Thr Gln Pro Leu
1460 1465 1470

<210> 63

5 <211> 1500

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

15 <222> (1)..(1500)

<400> 63

atg ggc ccg cga gtg gcg tca ggc aag gtg ccg gct tgg gag atg agc 48
Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1 5 10 15

aag tcc gag ctg tgt gat gac cgc acg gta gtc ttt gac tat gag gag 96
Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
20 25 30

ctg ctg gag ttc gct gag ggc gat atc agt aag gtt ttt ggg ccg gag 144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ser Lys Val Phe Gly Pro Glu
35 40 45

ttc aaa gtg gtg gac ggg ttt agg cgc agg gtg agg ttg ccc gct cga 192
Phe Lys Val Val Asp Gly Phe Arg Arg Arg Val Arg Leu Pro Ala Arg
50 55 60

ES 2 467 918 T3

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

ES 2 467 918 T3

```

gtc atg gca ggg tca ctg gtc agc gat ggt tgc agc cag ctc tta aaa 1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
325 330 335

ctc tat atg atc tgg ctt ggc ctc cac ctg aaa atg gag gaa ttt gat 1056
Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
340 345 350

ttt ctc cca gtt agc ggc cac aaa aac aag gtg cga tgc agg gga caa 1104
Phe Leu Pro Val Ser Gly His Lys Asn Lys Val Arg Cys Arg Gly Gln
355 360 365

att tca ccg cat aaa ggc aag ctt gtc tac gtc atg gaa atc aaa aag 1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Lys
370 375 380

atg ggt tac gat caa gca tct gga agc cca tac gcc atc gcg gac gtt 1200
Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
385 390 395

gat atc att gac gtc aac gaa gag ctg ggt caa agt ttt gac atc aac 1248
Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
405 410 415

gac ctt gcg agc tac gga aaa ggt gac ctg agc aaa aaa atc gtg gtt 1296
Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
420 425 430

gac ttc aaa gga att gct ttg cag ctc aaa ggc cgc gct ttt tca cgc 1344
Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
435 440 445

atg agt tcc agc tcg tcc ttg aac gaa gga tgg caa tgt gtt cca aaa 1392
Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
450 455 460

cca agc cag aga atg gaa cac gaa cag ccc cct gct cac tgc ctt gca 1440
Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
465 470 475 480

agc gac ccc gaa gcc cct tca act gtg acc tgg cac cca atg tca aag 1488
Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
485 490 495

ctt cct ggc aac
Leu Pro Gly Asn 1500
500

```

<210> 64

5 <211> 500

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 64

```

Met Gly Pro Arg Val Ala Ser Gly Lys Val Pro Ala Trp Glu Met Ser
1 5 10 15

Lys Ser Glu Leu Cys Asp Asp Arg Thr Val Val Phe Asp Tyr Glu Glu
20 25 30

```

ES 2 467 918 T3

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

ES 2 467 918 T3

His His Gly Gly Ala Tyr Gly Leu Gly Leu Leu Val Gly Glu Lys Ile
 290 295 300
 Leu Asp Arg Asn His Trp Tyr Phe Pro Cys His Phe Val Asn Asp Gln
 305 310 315 320
 Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Leu Leu Lys
 325 330 335
 Leu Tyr Met Ile Trp Leu Gly Leu His Leu Lys Met Glu Glu Phe Asp
 340 345 350
 Phe Leu Pro Val Ser Gly His Lys 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 Lys
 370 375 380
 Met Gly Tyr Asp Gln Ala Ser Gly Ser Pro Tyr Ala Ile Ala Asp Val
 385 390 395 400
 Asp Ile Ile Asp Val Asn Glu Glu Leu Gly Gln Ser Phe Asp Ile Asn
 405 410 415
 Asp Leu Ala Ser Tyr Gly Lys Gly Asp Leu Ser Lys Lys Ile Val Val
 420 425 430
 Asp Phe Lys Gly Ile Ala Leu Gln Leu Lys Gly Arg Ala Phe Ser Arg
 435 440 445
 Met Ser Ser Ser Ser Ser Leu Asn Glu Gly Trp Gln Cys Val Pro Lys
 450 455 460
 Pro Ser Gln Arg Met Glu His Glu Gln Pro Pro Ala His Cys Leu Ala
 465 470 475 480
 Ser Asp Pro Glu Ala Pro Ser Thr Val Thr Trp His Pro Met Ser Lys
 485 490 495
 Leu Pro Gly Asn
 500

<210> 65

5 <211> 1500

<212> ADN

<213> *Thraustochytrium* sp.

10

<220>

<221> CDS

<222> (1)..(1500)

<400> 65

5

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

aag act aac aac acc tcg gtt cga aga gtt gaa atc gca tcc gcc aaa Lys Thr Asn Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys 245 250 255	768
gga aaa gag cag ctg act gag aag ctt ccc gac gca act aat gct caa Gly Lys Glu Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln 260 265 270	816
gtt ctt cgg cgt tca gag cag tgt gaa tac ctc gat tac ctc aat att Val Leu Arg Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile 275 280 285	864
gcc cct gac tct ggg ctg cat ggg aag ggc tac gcc cac gga cac aaa Ala Pro Asp Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys 290 295 300	912
gac gtt aac ccg caa gac tgg ttc ttc tct tgc cac ttt tgg ttc gat Asp Val Asn Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp 305 310 315 320	960
cct gta atg cca gga tct tta gga att gaa tca atg ttc cag ctt atc Pro Val Met Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile 325 330 335	1008
gag gcc ttt gcg gtg gac caa aac att cct gga gag tac aac gta tcc Glu Ala Phe Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser 340 345 350	1056
aat ccg acc ttt gcc cat gca cca ggc aaa acg gcg tgg aaa tac cga Asn Pro Thr Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg 355 360 365	1104
ggc cag ctc aca cca aag aac cgt gcg atg gac tgc gag gtg cat atc Gly Gln Leu Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile 370 375 380	1152
gtt tca att acc gcc tcc ccc gag aac ggg ggc tac gtt gac atc gtg Val Ser Ile Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val 385 390 395 400	1200
gcc gat gga gcg ctt tgg gta gat gga ctt cgc gtg tac gaa gcc aaa Ala Asp Gly Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys 405 410 415	1248
gag ctt cga gtt cgt gtc gtt tcg gca aaa cct caa gca att ccg gat Glu Leu Arg Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp 420 425 430	1296
gta caa caa cag cca cct agc gca aag gcg gac ccg ggg aaa aca gga Val Gln Gln Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly 435 440 445	1344
gtt gca ctt tcg ccc act cag cta cgc gac gtc ctg ctt gaa gtg gac Val Ala Leu Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp 450 455 460	1392
aat cca ttg tat ctt ggt gta gag aac tcc aat ttg gtg cag ttt gag Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu 465 470 475 480	1440
tcg aaa cct gca act tct tca cgt atc gtt tcg atc aaa ccg tgc tcg Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser 485 490 495	1488
att agt gac ctt Ile Ser Asp Leu 500	1500

<210> 66

<211> 500

5

<212> PRT

<213> *Thraustochytrium* sp.

10

<400> 66

```

Pro Thr Pro Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys
1          5          10          15

Phe Ile Pro Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly
          20          25          30

Glu Met Pro Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys
          35          40          45

Val Ser Asn Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn
          50          55          60

Thr Ser Arg Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val
65          70          75          80

Glu Val Thr Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn
          85          90          95

Pro Ser Lys Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala
          100          105          110

Trp Phe Phe Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile
          115          120          125

Ile Met Glu Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu
          130          135          140

Lys Ala Pro Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu
          145          150          155          160

Asp Ala Ser Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys
          165          170          175

Thr Ile Arg Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys
          180          185          190

Met Gly Ile His Arg Phe Thr Phe Glu Leu Ser Val Asp Gly Val Val
          195          200          205

```

ES 2 467 918 T3

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

ES 2 467 918 T3

Asn Pro Leu Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu
465 470 475 480

Ser Lys Pro Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser
485 490 495

Ile Ser Asp Leu
500

<210> 67

5 <211> 1410

<212> ADN

<213> *Thraustochytrium* sp.

10 <220>

<221> CDS

15 <222> (1)..(1410)

<400> 67

ES 2 467 918 T3

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

ES 2 467 918 T3

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

ES 2 467 918 T3

Thr	Val	Ala	Lys	Ser	Tyr	Pro	Cys	Val	Ala	Gln	Ile	Asn	Leu	Gln	Ile		
			420					425					430				
ttg	caa	gga	gct	gcg	tat	ctg	aaa	cgc	ctt	ggg	gtc	att	cgt	ttt	gac	1344	
Leu	Gln	Gly	Ala	Ala	Tyr	Leu	Lys	Arg	Leu	Gly	Val	Ile	Arg	Phe	Asp		
		435					440					445					
cgc	atg	ctg	ctg	cag	gcc	gtc	gat	atc	gac	gat	cct	gta	ttt	act	tac	1392	
Arg	Met	Leu	Leu	Gln	Ala	Val	Asp	Ile	Asp	Asp	Pro	Val	Phe	Thr	Tyr		
		450				455					460						
gtg	ccg	acc	cag	cca	ctt											1410	
Val	Pro	Thr	Gln	Pro	Leu												
465					470												

<210> 68

5 <211> 470

<212> PRT

<213> *Thraustochytrium* sp.

10 <400> 68

ES 2 467 918 T3

Gly Asp Lys Ser Phe Met Glu Thr Tyr Asn Val Ser Ala Pro Leu Tyr
1 5 10 15

Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala
20 25 30

Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro
35 40 45

Ile Ser Ile Val Arg Glu Ala Leu Glu Lys Ile Gln Gln His Leu Pro
50 55 60

His Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn
65 70 75 80

Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Met Gly Val Thr Val
85 90 95

Val Glu Cys Ser Ala Phe Met Glu Leu Thr Ala Gln Val Val Arg Tyr
100 105 110

Arg Ala Ser Gly Leu Ser Lys Ser Ala Asp Gly Ser Ile Arg Ile Ala
115 120 125

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

Ile Arg Pro Ala Pro Gln His Leu Leu Gln Lys Leu Val Ala Ser Gly
145 150 155 160

Glu Leu Thr Ala Glu Gln Ala Glu Leu Ala Thr Gln Val Pro Val Ala

ES 2 467 918 T3

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

420	425	430
Leu Gln Gly Ala Ala Tyr Leu Lys Arg Leu Gly Val Ile Arg Phe Asp		
435	440	445
Arg Met Leu Leu Gln Ala Val Asp Ile Asp Asp Pro Val Phe Thr Tyr		
450	455	460
Val Pro Thr Gln Pro Leu		
465	470	

<210> 69

5 <211> 6180

<212> ADN

<213> Secuencia artificial

10 <220>

<223> sintética

15 <400> 69

atggccgctc gcaacgtgag cgccgcgcat gagatgcacg atgaaaagcg catcgccgctc	60
gtcggcatgg ccgtccagta cgccggatgc aaaaccaagg acgagttctg ggaggtgctc	120
atgaacggca aggtcgagtc caaggtgatc agcgacaaac gactcggctc caactaccgc	180
gccgagcact acaaagcaga gcgcagcaag tatgccgaca ctttttgcaa cgaaacgtac	240
ggcacccttg acgagaacga gatcgacaac gagcacgaac tcctcctcaa cctcgccaag	300
caggcactcg cagagacatc cgtcaaagac tcgacacgct gcggcatcgt cagcggctgc	360
ctctcgttcc ccatggacaa cctccagggt gaactcctca acgtgtacca aaaccatgtc	420
gagaaaaagc tcggggcccg cgtcttcaag gacgcctccc attggtccga acgcgagcag	480
tccaacaaac ccgaggccgg tgaccgccgc atcttcatgg acccggcctc cttcgtcgcc	540
gaagaactca acctcggcgc ctttactac tccgtcgacg cagcatgcgc cacggcgctc	600
tacgtgctcc gcctcgcga ggatcatctc gtctccggcg ccgccgacgt catgctctgc	660
ggtgccacct gcctgccgga gccctttttc atccttttcgg gcttttccac cttccaggcc	720
atgcccgtcg gcacggggcca gaacgtgtcc atgccgctgc acaaggacag ccagggcctc	780
accccggtg agggcggctc catcatggtc ctcaagcgtc tcgatgatgc catccgcgac	840
ggcgaccaca tctacggcac ctttctcggc gccaatgtca gcaactccgg cacaggtctg	900
ccctcaagc cccttctccc cagcgagaaa aagtgcctca tggacaccta cacgcgcatt	960
aacgtgcacc cgcacaagat tcagtacgtc gagtgccacg ccaccggcac gccccagggt	1020
gatcgtgtgg aaatcgacgc cgtcaaggcc tgctttgaag gcaaggctcc ccgtttcggg	1080
accacaaagg gcaacttttg acacaccctc gtcgcagccg gctttgccgg tatgtgcaag	1140
gtcctcctct ccatgaagca tggcatcatc ccgcccaccc cgggtatcga tgacgagacc	1200

aagatggacc	ctctcgtcgt	ctccggtgag	gccatcccat	ggccagagac	caacggcgag	1260
cccaagcgcg	ccggtctctc	ggcctttggc	tttggtggca	ccaacgcccc	tgccgtcttt	1320
gaggagcatg	acccctccaa	cgccgcctgc	acgggccaag	actccatttc	tgcgctctcg	1380
gcccgtgcg	gcggtgaaag	caacatgcgc	atcgccatca	ctggtatgga	cgccaccttt	1440
ggcgctctca	agggactcga	cgccttcgag	cgcgccatth	acaccggcgc	tcacggtgcc	1500
atcccactcc	cagaaaagcg	ctggcgctth	ctcggaagg	acaaggactt	tcttgacctc	1560
tgcggtctca	aggccacccc	gcacggctgc	tacattgaag	atgttgaggt	cgacttccag	1620
cgcctccgca	cgcccatgac	ccctgaagac	atgctcctcc	ctcagcagct	tctggccgtc	1680
accaccattg	accgcgccat	cctcgactcg	ggaatgaaaa	agggtggcaa	tgctgcgctc	1740
tttgctggcc	tcggcaccga	cctcgagctc	taccgtcacc	gtgctcgcgt	cgctctcaag	1800
gagcgctcc	gccctgaagc	ctccaagaag	ctcaatgaca	tgatgcagta	cattaacgac	1860
tgcggtacat	ccacatcgta	cacctcgta	attggcaacc	tcgtcgccac	gcgcgtctcg	1920
tcgcagtggt	gcttcacggg	cccctcctth	acgatcaccg	agggaacaaa	ctccgtctac	1980
cgctgcgccg	agctcgga	gtacctcctc	gagaccggcg	aggtcgatgg	cgctcgcttt	2040
gcgggtgtcg	atctctgcgg	cagtgccgaa	aacctttacg	tcaagtctcg	ccgcttcaag	2100
gtgtccacct	ccgatacccc	gcgcgccagc	tttgacgccg	ccgccgatgg	ctactttgtc	2160
ggcgagggct	gcggtgcctt	tgtgtctcaag	cgtgagacta	gctgcaccaa	ggacgaccgt	2220
atctacgctt	gcatggatgc	catcgctccct	ggcaacgtcc	ctagcgcttg	cttgccgcgag	2280
gccctcgacc	aggcgcgctg	caagccgggc	gatatcgaga	tgctcgagct	cagcgccgac	2340
tccgccccgc	acctcaagga	cccgtccgtc	ctgcccagg	agctcactgc	cgaggaggaa	2400
atcggtggcc	ttcagacgat	ccttcgtgac	gatgacaagc	tcccgcgcaa	cgctcgcaacg	2460
ggcagtgctca	aggccaccgt	cggtagacac	ggttatgcct	ctggtgctgc	cagcctcatc	2520
aaggctgcgc	tttgcatcta	caaccgctac	ctgcccagca	acggcgacga	ctgggatgaa	2580
cccgccccctg	aggcgccctg	ggacagcacc	ctctttgctg	gccagacctc	gcgcgcttgg	2640
ctcaagaacc	ctggcgagcg	tcgctatgcg	gccgtctcgg	gcgtctccga	gacgcgctcg	2700
tgctattccg	tgctcctctc	cgaagccgag	ggccactacg	agcgcgagaa	ccgcatctcg	2760
ctcgacgagg	aggcgcccaa	gctcattgtg	cttcgcgccg	actcccacga	ggagatcctt	2820
ggtcgcctcg	acaagatccg	cgagcgcttc	ttgcagccca	cgggcgccgc	cccgcgcgag	2880
tccgagctca	aggcgagggc	ccgccgcatc	ttcctcgagc	tcctcggcga	gacccttgcc	2940
caggatgccg	cttcttcagg	ctcgcaaaag	cccctcgctc	tcagcctcgt	ctccacgccc	3000
tccaagctcc	agcgcgaggt	cgagctcgcg	gccaagggtg	tcccgcgctg	cctcaagatg	3060
cgccgcgatt	ggagctcccc	tgctggcagc	cgctacgcgc	ctgagccgct	cgccagcgac	3120

cgcgtcgcct tcatgtacgg cgaaggctgc agcccttact acggcatcac ccaagacatt 3180
 caccgcattt ggcccgaact ccacgaggtc atcaacgaaa agacgaaccg tctctgggcc 3240
 gaaggcgacc gctgggtcat gccgcgcgcc agcttcaagt cggagctcga gagccagcag 3300
 caagagtttg atcgcaacat gattgaaatg ttccgtcttg gaatcctcac ctcaattgcc 3360
 ttcaccaatc tggcgcgcga cgttctcaac atcacgcccaggccgcctt tggcctcagt 3420
 cttggcgaga tttccatgat ttttgccttt tccaagaaga acggtctcat ctccgaccag 3480
 ctcaccaagg atcttcgcga gtccgacgtg tggaacaagg ctctggccgt tgaatttaat 3540
 gcgctgcgcg aggcctgggg cattccacag agtgtcccca aggacgagtt ctggcaaggc 3600
 tacattgtgc gcggcaccaa gcaggatata gaggcggcca tcgccccgga cagcaagtac 3660
 gtgcgcctca ccatcatcaa tgatgccaac accgcctca ttagcggcaa gcccgacgcc 3720
 tgcaaggctg cgatcgcgcg tctcgggtggc aacattcctg cgcttcccggt gacccagggc 3780
 atgtgcggcc actgccccga ggtgggacct tataccaagg atatcgccaa gatccatgcc 3840
 aaccttgagt tccccgttgt cgacggcctt gacctctgga ccacaatcaa ccagaagcgc 3900
 ctctgtccac gcgccacggg cgccaaggac gaatgggccc cttcttcctt tggcgagtac 3960
 gccggccagc tctacgagaa gcaggctaac ttccccaaa tcgtcgagac catttacaag 4020
 caaaactacg acgtctttgt cgaggttggg cccaacaacc accgtagcac cgcagtgcgc 4080
 accacgcttg gtccccagcg caaccacctt gctggcgcca tcgacaagca gaacgaggat 4140
 gcttggacga ccatcgtcaa gcttgtggct tcgctcaagg cccaccttgt tcttggcgtc 4200
 acgatctcgc cgctgtacca ctccaagctt gtggcggagg ctgaggcttg ctacgctgcg 4260
 ctctgcaagg gtgaaaagcc caagaagaac aagtttgtgc gcaagattca gctcaacggt 4320
 cgcttcaaca gcaaggcgga ccccatctcc tcggccgata ttgccagctt tccgcctgcg 4380
 gaccctgcca ttgaagccgc catctcgagc cgcacatga agccggttgc tccgaagttc 4440
 tacgcgcgtc tcaacattga cgagcaggac gagaccctg atccgatcct caacaaggac 4500
 aacgcgccgt cttccagctc tagctcctct tccagctctt ccagctcttc cagcccgtcg 4560
 ccagctccgt ccgccccagt gcaaaagaag gctgctccgg ccgcggagac caaggctgtt 4620
 gcttcggctg acgcacttcg cagtgccttg ctcgatctcg acagtatgct tgcgctgagc 4680
 tctgccagtg cctccggcaa cttgttgag actgcgccta gcgacgcctc ggtcattgtg 4740
 ccgccctgca acattgcgga tctcggcagc cgcgccttca tgaaaacgta cgggtgtttcg 4800
 gcgcctctgt acacggggcg catggccaag ggcattgcct ctgcggacct cgtcattgcc 4860
 gccggccgcc agggcatcct tgcgtccttt ggcgccggcg gacttcccat gcaggttgtg 4920
 cgtgagtgca tcgaaaagat tcaggccgcc ctgcccgaatg gcccgtagcg tgtcaacctt 4980
 atccattctc cctttgacag caacctcgaa aagggaatg tcgatctctt cctcgagaag 5040

ggtgtcacct ttgtcgaggc ctcggccttt atgacgctca ccccgaggt cgtgcggtac 5100
 cgcgcggtg gcctcacgcg caacgccgac ggctcggtca acatccgcaa ccgtatcatt 5160
 ggcaaggtct cgcgcaccca gctcgccgag atgttcatgc gtctgcgcc cgagcacctt 5220
 cttcagaagc tcattgcttc cggcgagatc aaccaggagc aggccgagct cgcccgccgt 5280
 gttcccgctg ctgacgacat cgcggtcgaa gctgactcgg gtggccacac cgacaaccgc 5340
 cccatccacg tcattctgcc cctcatcatc aaccttcgcg accgccttca ccgcgagtgc 5400
 ggctacccgg ccaaccttcg cgtccgtgtg ggcgcggcg gtggcattgg gtgccccag 5460
 gcggcgctgg ccaccttcaa catgggtgcc tcctttattg tcaccggcac cgtgaaccag 5520
 gtgcgcaagc agtcgggcac gtgcgacaat gtgcgcaagc agctcgcgaa ggccacttac 5580
 tcggacgtat gcatggcccc ggctgccgac atgttcgagg aaggcgtcaa gcttcaggtc 5640
 ctcaagaagg gaaccatgtt tccttcgctc gccacaagc tctacgagct cttttgcaag 5700
 tacgactcgt tcgagtccat gcccccgca gagcttgctc gcgtcgagaa gcgcactctc 5760
 agccgcgcgc tcgaagaggt ctgggacgag accaaaaact ttacattaa ccgtcttcac 5820
 aaccgggaga agatccagcg cgccgagcgc gacccaagc tcaagatgtc gctgtgcttt 5880
 cgctgggtacc tgagcctggc gagccgctgg gccaaactg gagcttccga tcgcgtcatg 5940
 gactaccagg tctggtgcgg tcctgccatt ggttccttca acgatttcat caagggaaact 6000
 taccttgatc cggccgtcgc aaacgagtac ccgtgcgtcg ttcagattaa caagcagatc 6060
 cttcgtggag cgtgcttctt gcgccgtctc gaaattctgc gcaacgcacg cctttccgat 6120
 ggcgctgccg ctcttggtggc cagcatcgat gacacatacg tcccggccga gaagctgtaa 6180

<210> 70

5 <211> 8394

<212> ADN

<213> Secuencia artificial

10 <220>

<223> sintética

15 <220>

<221> misc_feature

<222> (227)..(227)

20 <223> n es a, c, g o t

<400> 70

ES 2 467 918 T3

ccgcgaatca agaaggtagg cgcgctgcga ggcgcggcgg cggagcggag cgagggagag	60
ggagaggggag agagaggggag ggagacgtcg ccgcggcggg gcctggcctg gcctgggttg	120
gcttggtcag cgcggccttg tccgagcgtg cagctggagt tgggtggatt catttggatt	180
ttcttttggt tttgtttttc tctctttccc ggaaagtgtt ggccggncgg tgttctttgt	240
tttgatttct tcaaaagttt tggtggttgg ttctctctct tggctctctg tcaggcggtc	300

cggtccacgc	cccggcctct	cctctcctct	cctctcctct	cctctccgtg	cgtatacgtg	360
cgtacgtttg	tatacgtaca	tacatcccgc	ccgccgtgcc	ggcgaggggt	tgctcagcct	420
ggagcaatgc	gatgcgatgc	gatgcgatgc	gacgcgacgc	gacgcgagtc	actggttcgc	480
gctgtggctg	tggcttgctt	gcttacttgc	tttcgagctc	tcccgttttc	ttctttcctt	540
ctcacgccac	caccaacgaa	agaagatcgg	ccccggcacg	ccgctgagaa	gggctggcgg	600
cgatgacggc	acgcgcgcgc	gctgccacgt	tggcgctcgc	tgctgctgct	gctgctgctg	660
ctgctgctgc	tgctgctgct	gctgctgctt	ctgcgcgcag	gctttgccac	gaggccggcg	720
tgctggccgc	tgccgcttcc	agtccgcgtg	gagagatcga	atgagagata	aactggatgg	780
attcatcgag	ggatgaatga	acgatggttg	gatgcctttt	tcctttttca	ggccacacgc	840
gggaagcagg	agcgcgtgaa	tctgccgcca	tccgcatacg	tctgcatcgc	atcgcacgcg	900
atgcacgcat	cgctcgcccg	gagccacaga	cgggcgacag	ggcggccagc	cagccaggca	960
gccagccagg	caggcaccag	agggccagag	agcgcgcctc	acgcacgcgc	cgcagtgcgc	1020
gcatcgctcg	cagtgcagac	cttgattccc	cgcgcggatc	tccgcgagcc	cgaaacgaag	1080
agcgcgtac	gggcccattc	tagcgtcgcc	tcgcaccgca	tcgcatcgca	tcgcgttccc	1140
tagagagtag	tactcgacga	aggcaccatt	tccgcgctcc	tcttcggcgc	gatcgaggcc	1200
cccggcgccg	cgacgatcgc	ggcggccgcg	gcgctggcgg	cggccctggc	gctcgcgctg	1260
gcggccgcgc	cgggcgtctg	gccctggcgc	gcgcggggcg	cgcaggagga	gcggcagcgg	1320
ctgctcgccg	ccagagaagg	agcgcgccgg	gcccggggag	ggacggggag	gagaaggaga	1380
aggcgcgcaa	ggcggccccg	aaagagaaga	ccctggactt	gaacgcgaag	aagaagaaga	1440
aggagaagaa	gttgaagaag	aagaagaaga	aggagaggaa	gttgaagaag	acgaggagca	1500
ggcgcgttcc	aaggcgcgtt	ctcttcggga	ggcgcgttcc	agctgcggcg	gcggggcggg	1560
ctgcggggcg	ggcgcggggc	cgggtgcggg	cagaggggac	gcgcgcgcgc	aggcggaggg	1620
ggccgagcgg	gagcccctgc	tgctgcgggg	cgcgcggggc	gcaggtgtgg	cgcgcgcgac	1680
gacggaggcg	acgacgccag	cggccgcgac	gacaaggccg	gcggcgctcg	cgggcggaag	1740
gccccgcgcg	gagcaggggc	gggagcagga	caaggcgcag	gagcaggagc	agggccggga	1800
gcgggagcgg	gagcggggcg	cggagcccga	ggcagaacct	aatcgagatc	cagagcgagc	1860
agaggccggc	cgcgagcccc	agcccgcgcc	gcagatcact	agtaccgctg	cggaatcaca	1920
gcagcagcag	cagcagcagc	agcagcagca	gcagcagcag	cagccacgag	agggagataa	1980
agaaaaagcg	gcagagacga	tgggtccccg	tggtgcttcc	ggtaagggtc	cggcctggga	2040
gatgtcgaag	tcggagctct	gcgatgaccg	cacggctcgt	tttgattacg	aggagctcct	2100
tgagttcgcc	gagggtgaca	tctccaaggt	gttcgggtccg	gagtttaaag	tcgttgacgg	2160
ctttcgtcgc	cgcgtccgcc	tccccgcgcg	tgagtacctc	cttgtcaccc	gcgtcacgct	2220

tatggacgcc	gaggtgggta	actttcgcgt	gggcgctcgc	atggttactg	agtacgacgt	2280
tccggtcaac	ggcgagctgt	ctgagggcgg	cgacgtgccc	tgggccgttc	ttgtcgaggc	2340
cggccagtgc	gacctcctcc	tcattctcgt	catgggcata	gacttccagt	gcaagggaga	2400
gcgcgtctac	cgctcctca	acaccacct	tacgttcttc	ggcgtcgcca	aggagggcga	2460
gactctcgtg	tacgacatcc	gcgtcactgg	tttcgccaag	cgccctgacg	gcgacatctc	2520
gatgttcttc	tttgagtacg	actgctactg	caacggtaag	ctcctcatcg	agatgcgcga	2580
tggttctgcc	ggctttttca	ctgatgagga	gctcgtgcc	ggcaagggcg	tcgtcgtcac	2640
gcgtgcccag	caaaacatgc	gcgacaagat	cgtccgtcag	tccatcgagc	cgtttgcgct	2700
cgcgcctgc	actcacaaga	cgacctcaa	cgagtcgcac	atgcagtccc	tcgtcgagcg	2760
caactgggcg	aacgtgttcg	gcaccagcaa	caagatggcc	gagctcaact	acaagatttg	2820
cgcgcgcaag	atgctcatga	tcgaccgcgt	gacctacata	gaccatcatg	gcggcgctta	2880
cggcctcggc	ctccttgctg	gcgaaaagat	cctcgatcgt	aaccattggg	acttccccctg	2940
ccactttgtg	aatgaccaag	ttatggccgg	ctcccttgtc	tctgacgggt	gttcgcagct	3000
gcttaagctt	tacatgatct	ggctcggcct	ccacctcaag	atggaggagt	tcgattttct	3060
ccccgtcagc	ggccacaaga	acaaggctcg	ctgccgcgga	cagatttcgc	cccacaaggg	3120
taagctcgtc	tacgtgatgg	agattaagaa	gatgggctac	gaccaggctt	cgggcagccc	3180
ctacgccatt	gcggagcttg	atatcattga	tgtcaacgag	gagctcggcc	agtcgtttga	3240
catcaacgat	ctcgcgtcgt	acggtaaggg	cgacctctcg	aagaagattg	ttgtcgattt	3300
caaaggtatc	gctctccagc	tgaaaggccg	cgccttctcc	cgcattgtct	cctcgtcctc	3360
cctcaacgag	ggctggcagt	gcgtgcccga	gccgtcgcag	cgcattggagc	acgagcaacc	3420
ccctgcccat	tgctcgcct	ccgatccgga	ggcccttagc	acggtgacgt	ggcaccctcat	3480
gagcaagctc	cccggcaacc	ccacccccct	cttttctccc	tcgtcgtatc	cgcctcgtgc	3540
tatctgcttc	atcccttttc	cgggcaacc	gctcgacaac	aactgcaagg	ccggcgagat	3600
gcccccaac	tggtacaaca	tgagcgaatt	catgtgcggg	aaggtcagca	actgcctcgg	3660
cccggagttc	gcccgttttg	acaagagcaa	cacctccgc	tccccgcct	tcgaccttgc	3720
gctcgtcacg	cgcgtcgtcg	aggtcaccaa	catggagcac	ggcaagttcc	tcaacgtcga	3780
ctgtaaccct	tcgaagggca	cgatgggtggg	agagtctgac	tgcccgagg	acgcctgggt	3840
ctttgacggc	tcctgcaatg	acggctcatat	gccgtactcc	atcatcatgg	agatcggcct	3900
tcagacgtct	gggtgcctca	cgtcgggtgct	caaggctcct	ctcacgatgg	acaaggacga	3960
catcctcttt	cgcaaccttg	atgcctcggc	tgagatgggt	cgccctgacg	tcgatgtccg	4020
cggcaagacc	atccgcaacg	tgaccaagtg	taccggctac	gccatgcttg	gtaagatggg	4080
tatccatcgc	tttacctttg	agctctccgt	ggacggcgct	gtcttttaca	agggcagcac	4140

cagcttcggc tggtttacgc cggaagtgtt tgcccagcag gccggtctgg acaacggcaa 4200
 gaagaccgag ccctggtgca agaccaacaa cacctccgtc cgccgcgtcg agatcgccag 4260
 cgctaagggg aaggagcagc tcacggagaa gctccctgat gctaccaacg cccaggttct 4320
 ccgccgctcc gagcagtgcg agtacctcga ctacctcaac atcgcccccg actccggcct 4380
 tcacggcaag ggttacgcgc acggccacaa ggacgtcaac cccaggact ggtttttctc 4440
 gtgccacttc tggttcgacc cggatgatgc cggctcgctt ggcattgagt ccatgttcca 4500
 gctcatcgaa gcctttgctg tcgaccagaa tatccctggc gagtacaatg tttcgaaccc 4560
 gacctttgcc catgcccccg gtaagaccgc ttggaagtat cgccggccagc tcaactccaa 4620
 gaaccgcgct atggactgcg aggtccacat cgtcagcadc acggcgagcc cggagaacgg 4680
 cggctacgtc gacatcgctc ccgacggcgc cctttgggtc gatggcctcc gcgtttacga 4740
 ggccaaggag ctccgtgtcc gcgtcgtcag cgccaagccg caagctatcc cggatgtcca 4800
 gcagcagcct ccctcggcta aggccgaccc tggcaagacc ggtgtcgccc tcagccccac 4860
 ccagctccgc gatgtcctcc tggaggtcga caaccctctg tacctcgggtg tcgagaactc 4920
 caaccttgct cagttcgaaa gcaagccggc gacgtcctcc cgcacgtgtt ccattaagcc 4980
 gtgtttctatc tccgacctcg gcgacaagtc tttcatggag acttacaacg tgtctgcgcc 5040
 tctttacacc ggcgcgatgg ccaaggggat tgccagcgtc gacctcgtca tcgccgctgg 5100
 caagcgtaag atcctcgggt cgttcgggtg cgggtggcctc ccgatctcca tcgttcgga 5160
 ggcccttgag aagatccagc agcacctccc ccacggccct tacgccgtca accttatcca 5220
 ctccccgttc gactcgaacc tggagaaggg caatgtcgac ctttttcttg aaatgggtgt 5280
 caccgtcgtc gagtgctcgg ctttcatgga gctcacggct caggctcgtt gctaccgcgc 5340
 ttccggcctc tccaagagcg cggacggcag catccgcadc gcgcaccgca tcatcggcaa 5400
 ggtgagccgc actgagcttg ccgagatgtt catccgtcct gccctcagc accttctcca 5460
 gaagctgggtc gccagcggcg agctcaccgc cgaacaggcc gagcttgcca cgcaagtccc 5520
 ggtcgccgat gacatcgcg tcgaggccga ttcggggcgt cacaccgaca accgccgat 5580
 ccacgtgatt ctccccctta ttatcaacct ccgcaaccgc cttcacaagg agctcgacta 5640
 cccctcgcac ctgcgcgtgc gcgtgggagc cggcggtggt atcggtgcc ctcaggcggc 5700
 cctggcggcc tttcagatgg gcgtgcttt cctgattacc ggcaccgtca accagctcgc 5760
 ccgtgagtcg ggcacctgcg acaacgtccg ccttcagctc tccaaggcca cctacagcga 5820
 cgtgtgcatg gccctgccc cggacatgtt tgaccagggc gtggagcttc aggtgtctaa 5880
 gaagggcacg ctgttcccca gccgcgcaa gaagctctac gagctcttct gcaagtacga 5940
 ctccctcgag gccatgccgg ccgaggagct tcagcgtgtc gagaagcgca tctttcagaa 6000
 gtcgctggcc gaggtttggc aggagacttc ggatttctac attcacgca ttaagaaccc 6060

tgagaagatc aatcgcgctg ccagcgcgag taaactcaag atgtcgcttt gcttccgctg	6120
gtatcttggc ctctcgtcgt tctgggcca ctcgggtgcc caggaccgcg tcatggacta	6180
ccagatttgg tgcgggtccc cgatcggagc cttcaacgat ttaccaagg gcacctacct	6240
cgacgtcacg gttgccaaagt cttaccgctg cgttgcccag atcaacctcc aaatccttca	6300
gggtgcggct tacctcaagc gctcgggtgt gatccgcttc gaccgcatgc tcctccaggc	6360
ggtcgacatc gatgaccggt ttttcacgta cgtccccacc cagcccctct aagaaagtga	6420
accttgctct aacccgacag cgaatggcgg gagggggcgg gctaaaagat cgtattacat	6480
agtatttttc ccctactctt tgtgtttgtc tttttttttt tttgaacgca ttcaagccac	6540
ttgtctbggt ttacttgttt gtttgcttgc ttgcttgctt gcttgctgc ttcttggtca	6600
gacggcccaa aaaaggga aaattcattc atggcacaga taagaaaaag aaaargtttg	6660
tcgaccaccg tcatcagaaa gcaagagaag agaaacactc gcgctcacat tctcgctcgc	6720
gtaagaatct tagccacgca tacgaagtaa tttgtccatc tggcgaatct ttacatgagc	6780
gttttcaagc tggagcgtga gatcatacct ttcttgatcg taatgttcca acctgcata	6840
ggcctcgctg cgatccgcta gcaatgcgtc gtactcccgt tgcacctgcg ccctcgcttc	6900
attgtgacgt gagttcagat tcttctcgag accttcgagc gctgctaatt tcgctgacg	6960
ctccttcttt tgtgcttcca tgacacgcgc cttcacgctg cgttccactt ctctctcaga	7020
catgcccttg gctgcctcga cctgctcggt aagcttcgtc gtaatctcct cgatctcgga	7080
attcttcttg ccctccatcc actcggcacc atacttgga gctgttcaa cacgctcatt	7140
gaaaaacttt tcattctctt ccagctccgc aaccgcgct cgaagctcat tcacttccgc	7200
caccacggct tcggcatcga gcgccgaatc agtcgccgaa ctttccgaaa gatacaccac	7260
ggcccctccg ctgctgctgc gcagcgtcat catcagtcgc gtgttatckt cgcgcagatt	7320
ctccacctgc tccgtaagca gcttcacggt ggctcttga ttctgagggc tcacgtcgtg	7380
gattagcgtc tgcagctctt gcagctccgt cagcttgga gagctcgtaa tcatggcttt	7440
gcacttgctc agacgtcgca gagcgttcga gagccgcttc gcgttatctg ccatggacgc	7500
ttctgcgctc gcggcctccc tgacgacagt ctcttgagc ttactagat catgtccaat	7560
cagcttgcggt tgcagctctc caatcacgtt ctgcatcttg tttgtgtgtc cgggcgcgcg	7620
ctcgctctgc gatctgcgaa tttcctctc gagctcgctg tcgagctcca rggcgccttt	7680
aagtagctcg aagtcagccg ccgttagccc cagctccgtc gccgcgttca gacagtcggt	7740
tagcttgatt cgattccgct tttccatggc aagtttaaga tcctggccca gctgcacctc	7800
ctgcgccttg cgcacatgc gcggttcgc ctggcgcaaa agcttcgagt cgtatcctgc	7860
ctgccatgcc agcgcaatgg cacgcacgag cgacttgagt tgccmactat tcatcgccga	7920
gatgagcagc attttgatct gcatgaacac ctcgtcagag tcgtcatcct ctgcctctc	7980

ES 2 467 918 T3

```

cagctctgcg ggcgagcgac gctctccttg cagatgaagc gagggccgca ggcctccgaa 8040
gagcacctct tgcgcgagat cctcctccgt cgtcgccctc cgcaggatyg cggycgtgtc 8100
cgccatcttg ccgccacagc agcttttgst cgtctcgac cttcaatttc tggcgccgct 8160
ggtgcccgtg gtgcccgttg tgctggtgct ggtgctggtg ctggtgctgg tgccttgtgc 8220
tggtgctgcc acagacaccg ccgctcctgc tgctgctctt ccggccccct cgcgcgcgcc 8280
gcgagccccc gccgcgcgcc gtgcctgggc tctcgcgct ctccgcgggc tctcggcct 8340
cggcctcgcc gtccgcgacg acgtctgcgc ggccgatggt gcggatctgc tcta 8394

```

<210> 71

5 <211> 12495

<212> ADN

<213> Secuencia artificial

10

<220>

<223> sintética

15 <400> 71

ES 2 467 918 T3

cgactgcgag ggctcttgcg tcatccgcct gatgcggccg ccggccggaa gaggccgccc	60
cgcagcctcc ccggtgcggg gccgcgccgg gatgcctcgc ctgcgtcccc agcgatgcga	120
ggcggcgacg tctgtgctgc cgtcaggtcc tgtgcgaggt cccactgtta atattaagta	180
cgtaccgcac aggacccgga gtagcgcggc gcggggacga cctccgacgc ggccggaaga	240
cggcgcccac agcacgcgct cgggttcgcg ggccggcgcc gcgcgaagtg cgtaagcaga	300
cggcgcgcgg tccggctgcg ggggaagaag gtgtggaagg tgcgggggga acccgtggcc	360
gcctacgtgt gcctgcgacg gaggcgcggc ggccgccggc acctaggaac tgtctaagag	420
ggcttacaac ggacagcccg tcacaggcgc agaggcggcg gcggcgggtc ttcccttctc	480
tcccggcaga aagagaggga gggagagaga gagaaagcct gaatgtggct gcgatggaga	540
cttgggcagc gataatgact tgggtgggctg aagattactc tcagcattgc tgcgcgctta	600
gcattcttgc ggcattggcc aggtcggcgt cgggtttcgc ttgcggcgac gaccgagggc	660
cgttgatcg cgcggcagat tagacaggca gctgggacgc acgcacgcac gcacgaacga	720
acaaacaaac gagagaacgt gcgtacgggc ctgctggccg cggatgcaga aaagtgaatg	780
gaattttcca ctacagccag cgagccctgt cagccagcta gctgcagctt gagcggagcg	840
gcggctgttg ctgcgtctct tgaacttgaa ttcgaatgga tgtatgtact tgcttgccctc	900
ttgatgctgc gttgacttgt ggatgatgtt caaagtggca ggcacaccac ctaaaatcaa	960
aaaatgtaaa ggcgcggcgc agaaccgtgc atctgccgag gagcgcgtgc accgctgccg	1020
tgcgtgcggt gcgtgcgtgc tcgcctgcat ccgcgtcgcc ggcaaggtct ctgtgctgta	1080
tctatcgctc gcattcgaat cgctcgcgcc ttctgctccg acgtgcccta aaacatagag	1140
tccatcatta ggctagtgat actctatgtg acttgcttcc ttccctggaga ggaagaatga	1200

ataccgtatg	gtatcaggta	tgaagaaaac	caacctcgcg	cctccggacc	caccgcccgcg	1260
cggccagcca	gatccgcgcg	gccagctcgt	ccatgccagc	ctcgggaggt	atccatcccc	1320
tccgcgccgg	cagctgtcgg	tctctcgcca	tcgcctgtgc	cgctcgaaga	agtttcagca	1380
acctctccgt	cctgcgccat	ggcaccatct	tatcttctcc	gcgtcccact	cttctctccc	1440
aaagcctgca	cctcaaaggc	gaggccgggc	caggccgcag	ccgcgccacg	agaagcccca	1500
tcacgcgcgc	tcgcacgcgc	ccgcgcgcgt	ctgcgcgaga	cggcgctggt	atcttctttt	1560
cttttctttt	cttttcttct	ctcctccccg	ctctcctcgc	ttcctgcgcg	ccgctgcggg	1620
ggcgaaggaa	gacagggggg	ctcgggagga	agaggaggcg	aggggaaggaa	ggagggcggc	1680
gacggcatcg	ccagacgacg	acggcggcgc	cgggggacgc	ggggcgcgcg	tccccccgtg	1740
cgtggcagca	tgatgcatgc	gatgcgcgcg	atgcgtgcgc	cgcgaggagg	cgaggccccga	1800
tgggcggagg	cgagcggggg	ggatacgggc	ggcggcgggc	gggagcggga	ggcaggcaga	1860
ccccgaaccc	cgagggtaaa	actacacaaa	cagacagaca	gcctgcggct	gcctggcaac	1920
aagactcacg	cacgcacgca	cgcaccgatc	ggtcgagcga	gcgagcgctt	gatcgaacgg	1980
acgtgagaac	cagcaaggaa	gcagcagcag	cagcagcagc	agcaacagct	cgagagagag	2040
aaaatgaagg	atatggagga	tcgccgcgtc	gccattgtcg	gtatgtcggc	ccatctcccc	2100
tcgggcaccg	acgtgaaggga	gtcgtggcag	gccattcgtg	atggatatcga	ttgcctgtct	2160
gacctccccg	ccgaccgcct	cgacgtcacg	gcgtactaca	acccaacaa	ggccactaag	2220
gataagatth	actgcaagcg	cgggtggctth	attcccaact	acgattthga	tcctcgcgag	2280
tttggcctta	acatgttcca	gatggaggac	tccgacgcca	accagactct	cactcttctg	2340
aagggtcaagc	aggccctgga	ggacgcttcc	attgagccct	ttaccaagga	gaagaagaac	2400
atcggctgcg	tcctcggaat	cggcgggtggc	cagaaggcca	gccacgagtt	ttactctcgc	2460
ctcaactacg	ttgtcgtcga	gaaggtcctc	cgcaagatgg	gcctccctga	tgccgacgtt	2520
gaggaggccg	tggagaagta	caaggccaat	tttcttgagt	ggcgctcga	cagcttccct	2580
ggcttcctcg	gcaacgtcac	tgctggccgc	tgctccaaca	cctttaacat	ggagggtatg	2640
aactgcgtgg	tggacgccgc	ctgcgcctcc	tcctcatcgc	ccattaaggt	ggcgggtggag	2700
gagctcctct	tcggcgactg	cgacaccatg	attgccgggtg	ccacctgcac	ggataactcg	2760
ctcggcatgt	acatggctth	ttccaagacg	ccggtgttca	gcaccgacct	ctcgtccgc	2820
gcctacgatg	aaaagactaa	gggcatgctt	atcggcgagg	gctccgcat	gttcgtgctc	2880
aagcgtacg	ccgacgcggt	ccgtgatggc	gacactatcc	acgccgttct	ccgtagctgc	2940
tcgtccagca	gcgacggcaa	ggccgctgga	atttacacgc	ctacgatctc	gggccaggag	3000
gaggccctcc	gccgtgccta	cgcccgcgct	ggcgtctgcc	cctcgacct	cggcctcgth	3060
gagggtcacg	gcaccggcac	gccggctcggc	gaccgcatcg	agctcactgc	cctccgcaat	3120

ctcttcgaca aggccttttg ctccaagaag gaacaaatcg ctgtcgggtc catcaagtct 3180
 cagatcggcc acctcaagag cgtggccggc tttgctggcc tggtaaggc ggtcctggcc 3240
 ctcaagcaca agacgctccc tggtctgatt aacgtggacc agcctcccct cctttacgac 3300
 ggcacgcaga tccaggactc gtccctgtac attaacaaga ctaaccgccc ctggtttacc 3360
 caaaacaagc tccctcgccg tgcgggtgtc agctccttcg gttttggtgg cgccaactac 3420
 cacgccgtcc tcgaagagtt tgagcccag cagagaagc cttaccgcct caacacggtc 3480
 ggccaccccg tcctcctcta cgcctccagc gtcgaggccc tcaaggctct ttgcaacgac 3540
 cagctcgccg aactcacat cgcgttgag gaggccaaga cccacaagaa cgtggacaag 3600
 gtttgcggtt acaagtttat cgacgagttc cagctccagg gtagctgccc tcctgagaac 3660
 ccccgctcg gctttctcgc cactctcccg acgtccaaca tcatcgctcg tctcaaggcg 3720
 atcctcgctc agctcgatgc caagcctgac gccagaagt gggacctccc gcacaagaag 3780
 gctttcgggtg ctaccttcgc ctgctcgctg gtcaagggtc cggttgccgc gctgttcgcc 3840
 ggccagggca ctcagtacct caacatgttt tccgatgtcg cgatgaactg gcccccttc 3900
 cgcgattcga tcgtggccat ggaggaggcc cagaccgagg tgttcgaggg ccagggtgaa 3960
 cccatttcga aggtcctctt tccgcgcgag cgttacgcct cggagtcgga gcagggaac 4020
 gagctgctct gcctcacga gtactcgag ccactacaa ttgccgcgcg tgtcggcgcc 4080
 tttgacatct tcaaggcggc tggctttaag cccgatatgg tcggaggcca ctcccttggc 4140
 gagtttctg cctgtacgc cgctggctcg atctctcgcg atgacctcta caagctcgtc 4200
 tgcaagcgtg ccaaggcgat ggcgaacgcg agcgatggag ccatggccgc tgtgatcggt 4260
 cccgatgtc gcctcgtcac tcctcagaac tccgacgtct acgtcgcaa tttcaatagc 4320
 gcgacgcagg tcgtcatcag cggcacggtg cagggcgtca aagaggagag caagctcctt 4380
 attagcaagg gttttcgcgt cctgcctctg aagtgccagg gcgccttcca ctccccgctc 4440
 atgggacctt ccgaggactc cttcaagtcg ctcgttgaga cttgtactat ctccccgcct 4500
 aagaacgtca agttcttttg caacgtgagc ggcaaggagt cgcctaattc gaagcagact 4560
 ctcaagtctc acatgaccag ctccgtccag tttgaggagc agatccgcaa catgtacgat 4620
 gccggtgccc gtgtctttct ggagtttggc cctcgccagg tgcctgcaa gctcatcgcc 4680
 gagatgttcc cctcgtgcac cgccatctcg gtcaaccggc cctcctcggg cgactcggac 4740
 gtccagctcc gcctcgcggc cgtcaagttt gcggtgtccg gcgctgccct ctcgacgttt 4800
 gacccctggg agtaccgcaa gcccaggac ctcttatcc gcaagccccg caagaccgcg 4860
 ctcgctcctc ctgctgccac gtacgtttcg cctaagacct tcgccgagcg caagaaggcc 4920
 atggaggaca tcaagctcgt gtcgatcacc cctcgcgact cgatggtttc gattggtaag 4980
 atcgctcaag aggtccgcac tgctaagcag cccctcgaaa ccgagattcg ccgcctcaac 5040

aaggagctcg	aacacctcaa	gcgcgagctc	gcggctgcc	aagcttcggt	caagtctgcc	5100
tccaagtcct	ctaaggagcg	cagcgtcctc	tccaagcacc	gcgccctcct	ccagaacatc	5160
ctccaagact	acgacgacct	ccgcgtcgtc	cccttcgccg	tccgctcggt	cgccgtggac	5220
aacaccgcgc	cgtacgccga	ccaggtcagc	acccctgcct	cggagcgtag	cgcgtcgccc	5280
cttttcgaga	agcgttcctc	cgtttcgtcg	gctcgccctgg	ccgaggcgga	ggccgccgtc	5340
ctctctgtgc	tcgccgacaa	gacgggttac	gacagctcga	tgatcgagat	ggacatggac	5400
ctggagagcg	agcttggcgt	ggactccatc	aagcgcgtcg	agatcatgtc	cgagggtccag	5460
actcttctct	cggtcgaggt	cagcgatgtt	gacgctctca	gccgcaccaa	aacggtcggc	5520
gacgtgatcg	aggctatgaa	gctcgaactt	ggcggacccc	aaggccagac	cctcaccgcc	5580
gagtcgatcc	gtcagcctcc	cgtgtcggag	cccgcgcgtc	cgacgtcctc	ctcttcctcc	5640
atcgctaacg	tctcttctgc	tcgcctcgcc	gaggctgaag	ctgctgttct	ttctgtcctc	5700
gccgacaaga	ccggctacga	ctcgtccatg	attgagatgg	acatggacct	cgaaagcgag	5760
ctcggcgtcg	acagcattaa	gcgcgtcgag	atcatgtccg	aggtgcagac	ccttctcagc	5820
gtggagggtt	cggacgtcga	tgccctctcc	cgcactaaga	ccgtcggcga	tgtcatcgag	5880
gccatgaagc	tcgaactcgg	cggtcctcag	ggtcaaacgc	tgaccgccga	gtcgattcgc	5940
cagccccctg	tttccgagcc	ggcgtgccg	acgtcgtcgt	cgagctcgat	cgccaacggt	6000
tccagcgcgc	gcctcgccga	ggccgaagcc	gccgtcctct	ccgtcctcgc	ggacaagacc	6060
ggttacgact	cctcgatgat	tgaaatggac	atggatcttg	agtctgagct	tggcgtcgat	6120
tccatcaagc	gcgttgagat	catgtcggaa	gtgcaaactc	tcctttctgt	tgaagtctct	6180
gatgtcgatg	ctctcagccg	cacgaaaacc	gtcggtgacg	tcatcgaggc	catgaagctc	6240
gaactcggtg	gccctcaggg	ccagaccctc	accgccgagt	ccatccgtca	accgcccggt	6300
tccgagcccc	ccgtcccagc	ctcctcgtcg	agctcgattg	cgaacgttct	ctcggcgcggt	6360
ctcgcggagg	ccgaggccgc	ggtcctctcg	gtcctcgccg	acaagacggg	ctacgactcc	6420
agcatgatcg	agatggacat	ggacctcgaa	tctgagctcg	gcgtcgacag	cattaagcgc	6480
gtcgagatta	tgagcgaggt	ccagaccctc	ctctccgtgg	aggtttcgga	cgtggacgcg	6540
ctctcgcgta	ccaagaccgt	tggcgacgtt	attgaggcca	tgaaactcga	actcgggtggc	6600
cctcagggac	agaccctcac	tgccgagtcg	attcgccagc	ctccggtttc	tgagcctgct	6660
gttcctacgt	cctcttcttc	ttctattgct	aatgtcagct	cggcgcgcct	cgcggaggcc	6720
gaggctgccg	tgctctcggt	tctcgcggac	aagaccggct	atgactcgtc	catgattgag	6780
atggatatgg	atcttgagtc	cgagctgggc	gtcgatagca	tcaagcgtgt	ggaaatcatg	6840
tccgaagtcc	agaccctcct	ctcggtcgag	gtcagcgatg	tcgacgccct	ctcgcgcacc	6900
aagaccgtcg	gcgatgtcat	cgaggccatg	aaacttgagc	tcggcggacc	tcagggacag	6960

ES 2 467 918 T3

accctcacgt	ccgagccgat	ccaccagccg	cccgtagcgc	agccggccgt	cccgaacttcg	7020
agctcctcgt	ccatcgccaa	tgtctctagc	gctcgccctcg	ccgaggccga	ggccgcccgtg	7080
ctgtccgtcc	tcgccgataa	gaccggatat	gactccagca	tgatcgagat	ggacatggac	7140
cttgagtcgc	agctgggtgt	tgattctatt	aagcgtgttg	agattatgtc	ggaggtgcag	7200
acccttctct	cggtcgaagt	ctctgatgtc	gatgccctct	ctcgtaacaa	gactgtcggc	7260
gacgtcatcg	aggctatgaa	gatggagctc	ggcggacccc	agggccagac	tcttaccgcc	7320
gaatctatcc	gccagccgcc	tgtcagcgag	cctgccgtcc	ctacctctc	gtcctcgtcg	7380
attgccaacg	tttcttctgc	ccgtcttgcc	gaggccgagg	cgcccgctct	ctcggtctct	7440
gcggacaaga	ccggttatga	ttcgtcgatg	attgagatgg	acatggacct	tgagtcggag	7500
ctcgagagtc	acagcatcaa	gcgcgtcgag	atcatgtcgg	aggtccaggc	tctcctctcc	7560
gtcgaggtta	gcgatgtcga	cgccctctcg	cgcaccaaga	ccgttggcga	tgttattgag	7620
gccatgaaaa	tggagctcgg	tggtccgcag	ggccagaccc	tactgccga	gtcgattcgc	7680
gagccgccgg	tttccgagcc	ggccgtgccc	acgagcagct	cgtcgagcat	cgctaacgtc	7740
agctcggccc	gccttgccga	ggccgaggct	gctgtcctct	ccgtgctcgc	cgataagact	7800
ggttatgact	cgtccatgat	tgagatggac	atggacctgg	agtccgagct	cggcgtggac	7860
tcgatcaagc	gcgttgagat	catgtctgag	gtgcagaccc	tgctttccgt	cgaggtgagc	7920
gacgtcgatg	ccctgtcgcg	taccaagact	gtgggtgacg	tcacgaggc	gatgaagctg	7980
gagctcggag	agtcgagctc	gattgagact	cttaattgca	ccgaggtcga	gcacacgtcc	8040
tacaagagcg	tcaaggccag	cggatgcgag	aacgtggaca	cccgtttcgc	caaagtggtc	8100
cagatcagcc	tcccgtccaa	gcttaagagc	accgtcagcc	atgaccgccc	ggtcattggt	8160
gtggatgacg	gcaccccgct	taccacggag	ctctgcaaga	ttctcggtgg	caacattgtc	8220
gttctcagct	accagggcaa	gcctgccggc	cctcgtggcg	tcgaggttcc	cgacctctcg	8280
gaggaggccc	tcacccaggc	gctcgccctc	attcgctcga	cctacggcgt	ccccattggt	8340
ttcatctgcc	agcaggtttc	caacgtcagc	accaaggccc	agctctgctg	ggcgtcgtg	8400
gccgctaagc	acctcaagaa	ggacctcaat	gccgtcctcc	ctgactcccg	ctcgtttttc	8460
gtgggcgttg	tccgcctcaa	cggcaagctg	ggcacctttg	agaacatttc	cgacttttagc	8520
aagttcgatc	tcactaaggc	gctcgactac	ggccagcgcg	gctccctcct	cggcctgtgc	8580
aagagcctcg	atcttgagtg	ggagcagggt	ttctgccgcg	gcatcgacct	ggcctgcgac	8640
ctcatgcctc	tccaagccgc	gcgtatcctc	cgcaacgagc	tccagtgccc	gaacatgcgc	8700
ctgcgcgagg	tcggctacga	tatctcgggc	gctcgctaca	ccatctccac	cgacgacctt	8760
ctttgcggcc	ctagcaaggc	caaggtcgaa	gcggcggaac	tctttcttgt	cacgggtggt	8820
gccccgcggt	tactcccca	ctgcgtccgc	gagatcgctc	cgcgtagccc	cggcacgacg	8880

tttgttctcg	tgggccgctc	cgagatgagc	gacgagcccg	actgggccgt	gggtcactac	8940
aacaaggatc	tgcaccagtc	cacgatgaag	cacctgaagg	ccacgcatgc	tgccggaggc	9000
gtgaagccga	ctcccaaggc	gcaccgcgc	cttgtcaacc	gtgtcaccgg	ctcgcgcgag	9060
gtccgcgaga	gcctgcgcg	cattcaggag	gcggtgcca	acgttgagta	catcgccctgc	9120
gacgtcagcg	acgagaacaa	ggtgcccag	ctcgtccagc	gcgtggagca	gaagtacggt	9180
tgcgagatta	cgggcatctg	gcacgcctcg	ggtgttctcc	gcgacaagct	cgtggagcaa	9240
aagaccacgg	acgacttcga	ggcctgttt	ggcactaagg	tcacgggctt	cgtgaacatc	9300
gtcagccagg	tcaacatgtc	caagctgcgc	catttcctcc	tcttctcttc	gctcgcgggc	9360
tttcacggca	acaagggcca	gaccgactac	gccatcgcca	acgaggccct	caacaagatc	9420
gcccacaccc	tttccgcctt	tctccccaag	ctcaacgcga	aggtcctcga	ctttggccccg	9480
tgggtcggct	ccggcatggt	cacggagact	cttgagaagc	actttaaggc	catgggcgtc	9540
cagaccatcc	ccctcgaacc	gggtgcccgc	accgtggccc	agatcatcct	cgccagctct	9600
ccgccccagt	ccctcctcgg	taactggggc	ttccccgcca	cgaaccctt	tcagcgctcc	9660
aacgtggtca	ccggcacctt	ctcgccggag	gagatcgaat	ttatcgccga	tcacaagatc	9720
cagggccgca	aggttctccc	gatgatggcc	gccatcggtt	tcattggcgtc	gatcgccgag	9780
ggcctctacc	ctggctacaa	ccttcagggc	gttgagaacg	ctcagctttt	tcagggcctc	9840
accatcaacc	aggaaaccaa	gttcagatc	acgctcatcg	aggagcacia	ctccgaggag	9900
aacctcgacg	tcctcaccag	cctcggtgtc	atgctcgaaa	gcggcaagg	gcttcccgc	9960
taccgttgcg	tcgtctgcct	caacactacc	cagcagcagc	ccaagctcag	ccccaaagatc	10020
ctcaacctgg	aggttgaccc	cgctgcgag	gtcaaccctt	acgacggcaa	gtccctcttt	10080
cacggccccgc	tcctccaatt	tgtccagcag	gtcctccact	cgtcgacgaa	aggcctcgtg	10140
gcgaagtgcc	gtgccctgcc	gattaaggag	gccatccgtg	gcccgttcat	caagcagacg	10200
ctccacgacc	cgatcctcga	cgacgtcatc	tttcagctca	tgctcgtgtg	gtgccgcaac	10260
gccctcggct	cggcctccct	cccgaaccgc	attgagaaga	tgagctactt	cggtaatgtg	10320
tcggagggca	gcactttctt	cgccagcgtt	accccgctcg	gcccccggtg	cccgaaggat	10380
cccgtgatca	agatgcagtt	tctcctccag	gacgagtccg	gcaacacctt	ttcgtccggc	10440
gagggcagcg	tcgtcctcag	cgatgagctc	gtgttctaaa	tctaaactag	tgaatcgcca	10500
caagttgtct	ttttgttact	ctacgtactg	ctctttctaa	ttttacgacg	tatgcttctg	10560
ctcttgacga	cgacaaccat	ggcaataaaa	gtaaggcaag	aaagtgagtg	cgtgctcgct	10620
cacaaggctct	aggccgaaac	gaggcgcctt	aggattgggt	gtctccgtcg	taagtcatga	10680
cggagcgtag	gacaccgacg	gcctgccgcg	caaataacg	aaccgctgca	caattctttt	10740
cgttgagaac	aacctcgaac	ggcctgcctt	gctgagctgc	cgacacggtg	gaaggaagca	10800

tagcggccaa tcgaggggat gctattaatt aggcttggcg ctgctgccgc cgtgcctgaa 10860
gatgtgcga agcgtggcaa caccgtccac cttgccatta aagtcatgca ctccaaatac 10920
ctcgaacaga gcctctcgca tgtttggcca agcaccttcg agagagtctg gactctcact 10980
ccttcctcac cgttgctcgcc attctcagcg agctcatgca atcaccagga cggagagacg 11040
gccagctgct tttctgcttt ccattgttat tagagaaacg cttctcgctc tcatcgctct 11100
agtagacatt ccgatggctt cgttcgccaa tttgtcacct aagtaagcta gagtgttaag 11160
tctaaatgcc tttgaccgcg gtacggcgct acgtagatgc ctgtccttgc cagcaaacgc 11220
tagttcgcgg tgtgcgtaat ttggcccgca ttatgctggc tctcaaaatc aaccgccacc 11280
actcgcggtt gcacgatgat tttcgtgcac tcatgacatg agaaatgtga tactcaaact 11340
agtatagacc tcctactcct actgctgctt ttctcgtcag agctgtctcc ggaaaagttg 11400
acaagttgtt ggccttcttc ttctctgcta gtaagtagat catcatggat gagacgatga 11460
tgataatgat gatgatgata gcgaagagct tttgggactc gagcttgagc tgatgtggct 11520
tattttatta ctcaaaaacc aatgcaccgt tgcgagctct cgagagtgga gcaaagttgc 11580
aacagcgcac acacaaagaa cacatagaaa tagacatcaa aaatagtact ttacaatcc 11640
agccgaggca gcgcgaaaag tcagcatctg cgaactgcc aaaaagccgt agatcaagtt 11700
ttgacacaac gctcatgcac gtctacgtgc tcgcgcatat aggaaaacaa atcacatcga 11760
gagattccca ctcaggctct gcgtggcctt ttgggcatac tcttgccgcg cagcagccca 11820
caacgcctca tcttctccca taccaaggct aaaggatggg tgtattttgt cgtgaagttg 11880
gggcggcatt ttctctaaaa gcattgatgc aatagatttc gggctgtagc gcgcggaaaa 11940
atgcatgagc agaagcgtat cgacgcgatc aaacaagcct cgatgaccga acgctgcaag 12000
ttggtcaaca tgaatatggc cgtggtcttc ggccttggaa atgtcgccat cgaaaaaggt 12060
cacctccacg cagagcgtct ttgcacggag aaaggagat tcagggtccc gttccaaagc 12120
agcagcgacc gtgtcgcccg tgtaagcaaa gtcgacaacc tcgtgtatga cgtagtcctc 12180
aggctttggg ggcgaggtct tgaagcgagt gcgctttgtc gcaacgggga gctccgcaag 12240
ttccggacgt aactcctttt tgatattaat gaccgcatat cctacagaag gcacacgatg 12300
ctcggttgca aggactgcaa cgcgcgctatt cccgttcaac tctacgtgtg cgccgggggt 12360
gacgcccgtg atcttcatat tgaaggggtc ctgggtgtcg tcaacatctc ctctccgtc 12420
gagggttgc atcgctcgcc ctgcactgcg aaaatcctcc acagcctccg caggcatgta 12480
gacgcgcgtc tcctt 12495

<210> 72

5 <211> 8647

<212> ADN

<213> Secuencia artificial

10

<220>

<223> sintética

5 <400> 72

ctatactggt gctcttcggt gctgatttag ctcgacttgg gtccacacca ctgcccctct	60
actgtgaatg gatcaatgga cgcacgacgg gccgacgaaa gtgcgcgagt gaggtaacct	120
aagcaacggc ggtcttcaga ggggacgcac gccctccgtc gcagtcagtc cagacaggca	180
gaaaagcgtc ttagggacca cgcacgcacg cacgcacgca cgcacgcccg cacgcacgct	240
ccctccctcg cgtgcctatt tttttaggct tccttccgca cgggcctacc tctcgtctcc	300
tcgcctcgcc gcaccaggcg gcagcagcga tacctgccgg tgccgcctcc gtcacgcgct	360
cagccgcagc tcagcccagc cgcgagctag ggtttgttcg tcctgaattg tttgatttga	420
tttgatttga tttgatccga tccgatccga tctgatctga tttgctttgc tttgctttgt	480
ctccctcccg gcgcggacca agcgtccgtc tgccgcgccg agcttccctt cttctcccag	540
ccctccttct gctcccgct ctcgcgcaag cacgcagctt cgccgcgcga tccggtcggt	600
cggtcggtcg atcgaccgc ctgccgctgc tgcgtgtggc gggcttttct ccatcggcga	660
ctctttcttc tccatacgtc ctactacgta catacatact gccggcttcc tcctcttcca	720
gcgcggcgac ggcggcaggc tgcgacgtcg tcgccgcgcg gggcgccgcg cgcgcgcgcg	780
ccgcgcgcgc cgtcgcaggg cctcgtcgcc gccgcgcgtc cgctccgctc cgaggccgcg	840
agagggccgc ggcggcgcga tggatggatg gatggatgga tggatggatg gattttgttg	900
atcgatggcg gcgcatgggc ggagatgagc gaggacgagc gcgcgagcgc ggcagccgga	960
ttcgcagggc ctcgctcgcc tcgcgcccgc tgccgcgcgc gcccttgcgag cctgcgcgcg	1020
gagcgagcga gcgagcgagc ggggctttct ttgtctcgcg cgccgcttgg cctcgtgtgt	1080
cttgtgcttg cgtagcgggc gccgcggtgg aagatggctc attcaatcga cccattcacg	1140
cacgcactcc ggcgcgcaga gaaggccgag gaggagcagc aagcaaacca aaagctctcg	1200
cgctcgcggt ctcgggctcg agcggctctg gagagagagt cttgcggcga ccaccggcag	1260
cagcagcagc agcagcagcg ctgtcgagca cgagcacgag cacgagcacg agcacgagca	1320
ttcgagcaag aggacagaca cggttgtcag cgcctagctc gctcgataca gaaagaggcg	1380
ggttgggcgt aaaaaaaaaag gagcacgcaa gccgccagcc agccagctag ctagccagcc	1440
tgccctgccaa aatgcagctc ccgcccgtc actccgctga tgagaaccgc attgccgtgg	1500
tcggcatggc cgtgaagtac gctggctgcg acaacaagga ggaattcttg aaaaccctca	1560
tgaacggttc gattaacacc aaaagcatta gcgcggctcg ccttggctcg aacaagcgcg	1620
acgagcacta cgtcccggag cgcagcaagt acgccgacac gttctgcaac gagcgttacg	1680
gctgcatcca gcagggcacg gataacgagc acgacctcct cctcggcctc gccaggagg	1740
ccctcgccga tgctgccggc cgtatggaga agcagccgtc ggaggccttc gatcttgaga	1800

ataccggcat	cgtgtcgggc	tgctgagct	ttccgatgga	taacctccag	ggcgagctcc	1860
tcaacctcta	ccagtccac	gtcgagaagc	agcttcccc	gtctgctctc	gtggaggccg	1920
tcaagctctg	gagcgagcgc	cagaagtcca	ctaaggccca	cgcgggcgac	aagcgccgct	1980
ttatcgatcc	ggcgtctttc	gtcgccgaca	agctgaacct	cggccccctc	cactacgcca	2040
tcgacgccgc	ctgcgcctcg	gcgctctacg	tgctgaagct	cgcccaggac	cacctcgtea	2100
gcggcgctgt	cgatatgatg	ctctgcggcg	ccacttgttt	cccggagccg	ttcttcatcc	2160
tctcgggctt	cagcaccttc	caagccatgc	cggtcggtgc	ggacggcgtg	tcgctcccc	2220
tccacaagac	ctcggctggc	ctcaccctcg	gtgagggcgg	ctccatcatg	gttctcaagc	2280
gcctcaagga	cgccatccgc	gacggcaacc	acatctacgg	cgtgctcctt	gaggccaacc	2340
tctcgaacgc	cggatgcggc	ctcccgtct	cgccgcacct	cccttccgag	gagtcgtgca	2400
ttcgcgacac	ctatcgccgc	gctggcgctg	ctgccgacca	gagcatccag	tacatcgagt	2460
gccacgccac	cggcaccctg	cgtggcgacg	tcgtcgagat	cgaggctgtc	gagcgcgctc	2520
ttaagaagaa	tgtgccccgc	cttggttcta	ccaagggcaa	cttcggccac	tcctcgtcg	2580
ctgccggctt	cgcgggtatg	gcgaagctgc	tcctcgctat	ggagcacggc	gtcatcccg	2640
ccacccttg	cctggacgcc	agcaatcagg	cttccgagca	cgtcgtcact	aaggcgatta	2700
cctggccgga	aaccacggc	gctcctaagc	gcgcggcct	ctccgccttc	ggctttggcg	2760
gcacgaacgc	gcatgccctt	tttgaggagt	tcaatgccga	gggcatttcc	taccgccccg	2820
gtaagcctcc	ggtcgagtcg	aacaccgcc	cctctgtggt	catcaactggc	atggactgca	2880
cgtttggcag	cctggaggga	atcgacgcct	ttgagactgc	cctctacgag	ggcgtgatg	2940
ccgcccgtga	tcttctcgcc	aagcgttggc	gcttctcgg	cgaggatctg	gagtttctcc	3000
gcgccatccg	cctgaaggag	aagccccgtg	gctgctttgt	tgagtccgtt	gacgtgaact	3060
ttcgccgcct	caagaccct	ctcaccctcg	aggatatgct	ccgcccgcag	cagctcctcg	3120
ctgtctccac	tatggaccgc	gcgatcatcg	acgcggcct	caagaagggc	cagcatgtcg	3180
cgtctctcgt	cggcctcggc	accgatctgg	agctttaccg	ccaccgtgcc	cgtgtggccc	3240
tcaaggaggt	gctccacccc	tcgctgaaga	gcgacaccgc	catcctccag	aagatcatgc	3300
agtacgtcaa	cgacgcgggc	acctcgacct	cctacacctc	gtacatcggc	aacctcgtcg	3360
ccaccctgat	cagctcgag	tggggcttta	ccggcccag	ctttactgtc	acggagggtg	3420
acaacacgct	ctaccgttgc	gccagctcg	ctaaggacat	gctgcaagtc	aaccgcgtcg	3480
atgcgcgtcg	cattgccggt	gtggacctga	acggctccgc	cgagagcttc	ttcgtccgcg	3540
ctaaccgtca	aaagatcagc	aagctttccc	accgctgcgc	cagcttcgac	cgcgacgccg	3600
acggcttctt	tgccggcgaa	ggctgcggcg	ctctcgtgtt	caagcgctt	gaggactgcg	3660
cccctcagga	gaagatttac	gcttccatcg	acagcatcgc	catcgacaag	gagcctacct	3720

cctcggccgt	caaggctgtc	taccagtccg	actcctccct	ctccgacatc	gagcttcttg	3780
agatcagcgg	cgactccaag	cgttttgcgg	ccttcgaggg	cgctgtcgag	atccagtcct	3840
ctgtggaggc	gcagctcaag	ggcctttcca	aggttcttga	gcccgcgaag	ggccaaggcg	3900
tgcgcgttgg	ctcgacccgc	gccactgtcg	gtgatatcgg	ctacgccact	ggcgcggtt	3960
cgctcatcaa	gacggccctg	tgcctctaca	accgttacct	gcctgccctc	gccaactgga	4020
gcggccccctg	cgaacagagc	gcctggggta	gcaacatgtt	cgtctgccac	gagactcgtc	4080
cttggatgaa	gaaccagaac	gagaagcgct	gcgcgctcat	ctcgggcacg	gacccgtccc	4140
acacctgctt	cagcctcgtg	ctctccgata	ctggttgcta	cgaggagcac	aaccgcacgt	4200
gctttgacgt	gcaggccccg	cagctcgttc	tcattcatgg	ttttgatggc	aagaccattg	4260
tgcgccgcct	tgagggtat	ctccttgagc	ttgtcgaggg	ccatgcctcg	ccgtccgagt	4320
actttcacia	gctcatcggc	cagagcctcc	ttgagaactc	caaggagtcg	aagctcacgc	4380
tctcgctcgt	ctgcaacccc	aaccgaactc	agaaggagct	catgctcgcc	atcaagggtg	4440
tgcagcgctc	tatgctgacg	ggtaaggact	gggtttcccc	ctcgggctct	tgctttgctc	4500
cgaaccctct	ctcgcccgcc	aaggctcgctt	tcatgtacgg	cgaaggccgc	tccccctact	4560
gcggtgtggg	ccttggcctc	caccgcctct	ggcctggcct	ccacgaaaac	gtgaacaaca	4620
agacggtgga	cctctggacg	gagggcgacg	gttggtctta	cccgcgcacc	ctcactcgty	4680
aggagcacac	taaggccatc	gagtcgttta	acgctaacca	gatcgagatg	ttccgtgccg	4740
gtatctttat	ctccatgtgc	cagacggact	acgtgatgaa	cgtcctcggc	gtgcagccca	4800
aggccggctt	tggcctctcg	ctcggcgaga	tcagcatgct	cttcgccatg	tccaaggaga	4860
actgccgcca	gagccaggag	atgaccaacc	gccttcgtgg	cagccccgtc	tggctgaacg	4920
agcttgccat	taacttcaat	gcgatccgta	agctctggaa	gattccccgc	ggtgccccgc	4980
ttgagtcctt	ctggcagggc	tacctcgccc	acggcacccg	cgaggaggtc	gagcacgcca	5040
tgcgcctcag	cgagccctac	gtccgcctgc	ttatcgtgaa	tgacagccgc	tgcggcctta	5100
tgcggggtaa	gcctgacgcc	tgccaggccg	tgatctcgcg	cctcaactcg	aagttccctt	5160
ctctccctgt	gaagcagggc	atgatcggcc	actgcccgga	ggtgcgcgcc	ttcatcaagg	5220
acattggtta	catccacgag	actctccgta	tctctaacga	ttactcggac	tgccagctct	5280
tttcggccgt	caccaagggg	gcccttgact	cgtctactat	ggagatcaag	cacttcgtcg	5340
gtgagggtcta	cagccgcatt	gccgattttc	cccagattgt	caacacgggtg	cactcggccg	5400
gctacgacgt	gtttctcgaa	ctcggtcgcg	acgcctcccc	cagcgccgct	gttcagaaca	5460
tcctcggcgg	tcagggcaag	tttctgtcta	ctgccattga	caagaagggg	cactctgcct	5520
ggtcgcaggt	tctccgcgct	accgcctcgc	tcgctgccca	ccgcgtgccc	ggcatctcga	5580
tcctcgatct	ctttcaccgc	aatttccgcg	agatgtgctg	cacgatggct	acgacccccca	5640

aggtggagga	caagtttctc	cgcaccatcc	agattaacgg	ccgtttcgag	aaggagatga	5700
tccacctcga	agataccact	ctctcctgcc	ttcccgcgcc	ttcggaggcc	aacattgccg	5760
ccattcagtc	gcgctcgatt	cgcagcgctg	cggcccgtag	cggccagtcg	cacgactgcg	5820
cgtcccacag	ccatgaggag	aataaggact	cctgcccgga	gaagctcaaa	cttgactcgg	5880
tcagcgtcgc	gatcaacttt	gacaacgacg	accgtattca	gctcggccac	gcgggttttc	5940
gcgagatgta	caacacccgc	tactcgctct	acacgggcgc	tatggctaaa	ggcatcgctt	6000
cggccgacct	cgtcatcgct	gcgggcaagg	agggatcct	cgccagctac	ggtgctggcg	6060
gcctcccgt	cgccaccgtc	cgcaagggtg	tcgacaagat	ccagcaggcg	ctccccagcg	6120
gtccttacgc	cgtcaacctc	atccactccc	cgtttgacgg	caacctggag	cagggcaacg	6180
tcgacctctt	tctcgaaaag	aacgtccgcg	tcgccgagtg	ttctgccttt	acgacctca	6240
ccgtgcccgt	cgtccattat	cgcgccgctg	gcctcgttcg	cgccaggac	ggctcgatcc	6300
tcattaagaa	ccgcatcatc	gccaaaggta	gccgcaccga	gctcgctgag	atgttcctcc	6360
gccctgcgcc	gcagattatc	ctcgaaaagc	tggtcgccgc	cgagatcatc	tccagcgatc	6420
aggcgcgtat	ggccgcgaag	gtgccgatgg	ccgacgacat	cgccgtcgag	gccgacagcg	6480
gcggccacac	ggacaaccgc	ccgatgcatg	ttattctccc	cctcattatc	cagctccgca	6540
acaccattct	cgcggagtac	ggctgcgcta	ctgcctttcg	cacgcgtatt	ggcgtggtg	6600
gcggcatcgg	ttgcccgcgc	gccgctctcg	ccgccttcga	catgggtgcc	tctttcgctg	6660
tcaccggctc	catcaatcag	atttgccgcg	aggcgggcac	ctgcgacacg	gttcgcgagc	6720
tcctcgccaa	ctcgtcgta	agcgatgtca	cgatggctcc	tgccgccgac	atgttcgatc	6780
agggcgtaga	gctccaagtc	ctcaagcgcg	gcaccatgtt	cccctcgcg	gctaacaagc	6840
ttcgcaagct	ctttgtgaac	tacgagtccc	tcgaaactct	cccctccaag	gagctcaagt	6900
accttgagaa	cattattttc	aagcaggctg	tggaccaggt	gtgggaggag	actaagcgct	6960
tctactgcga	gaagctgaac	aaccccgaca	agatcgctcg	cgtatgaag	gaccccaagc	7020
tcaagatgtc	gctctgcttc	cgctgggtatc	tcagcaagtc	gtcgggctgg	gccaacgccg	7080
gcatcaagtc	tcgcgccttc	gactaccaa	tctgggtgcg	accgcgatg	ggcagcttta	7140
acaacttcgc	ctccggcacg	agcctcgact	ggaaagtgc	cgggtgtctt	cccggcgctg	7200
ccgaggtaga	catggccatc	ctcgatggcg	cccgcgagct	cgtgccaa	cgcaactaag	7260
taagctctca	tatatgttag	ttgcgtgaga	ccgacacgaa	gataatatca	catacgcttt	7320
tgtttgttct	ttcaattatt	tgtctgtgct	tcattgtgct	cctcagtatc	tagctggcgg	7380
ctcttatctt	cttttaaaat	atctggacaa	ggacaaaaac	aagaataaag	gcgagaagat	7440
gtgaatttca	tttcgacttg	agaactcgaa	gagcattgat	gcggttagta	tatgggtatt	7500
ttccagacac	ttttcatcat	catcatcatc	atcatcatta	tgaagaagta	gtagctgata	7560

ES 2 467 918 T3

aagtagactc actgtttgca gcgagaacca gggcctttcg gtccgtaacc ttggtgaaaa 7620
atcgaggagc gatcatgcct tgtgcgctgc cggggctaag aatgaccccg ttgtctgatt 7680
cgaaaaatth gatgccagcc tcggcagctg cagcagcagc gtctacatac acgaagtacg 7740
tcacaaggac ttgcgcatcc cgctgatcac acccgtctcc gaggacggaa gcccgtgagc 7800
aaagtgaatg tgattccgtt tcattcgca caaaccttg tcctggatct gagcccaggc 7860
agaagtatth gttccgtgca agcacgtctc aatgcttgct ggatcgagtg gacgcaacaa 7920
actcgtctct ttgacaactg acatggagtg tccctgattc gcacgaatca accagcttcc 7980
gtccgtgtcc tggcgcaacg tgaaccgctg cttgtcgttg tctcgacga cagcctcaat 8040
gtccgaaaag gagatcccct tgaagcgctt gagcttgccc tgtagctcag ccagacgaac 8100
aaatccatct tcgcgcaact cgagaccctc ctgaacggcc gagtgcgcga gcacacggct 8160
tagcatgtgc gaaagcttct tgttatgcgc cttggctttt gaaggcagcg ttggcgcgct 8220
gcgagagtgt ttacttggtt gcgcagcttc actatcacga gcagcacttc gtttattcat 8280
ttccatgttg tttgcagggt cagtggtagc cccatccggc acggggccaga gcaccctgtg 8340
agagcgccgt tcctcgaccc tgaggaaaag ctccgtaggt atggctcctg tatcgcccag 8400
accaggggta agtaggacgc cattgcttga ttggaacatt tccaagcccg catcatgtgc 8460
gcgcgcggta tcgaggaaga tgagcacctc tgcggagccg cgaaagccac tgagcactcc 8520
gtcctgacca ggaagtcccg aggcaaatg aatgtgattt cggcccatgc ggcggaggcc 8580
ctggcgttgg attgcaggcc agtggcgaag atacgttccg tgaatacaca gcggaagcaa 8640
tgagaga 8647

<210> 73

5 <211> 4479

<212> ADN

<213> Secuencia artificial

10 <220>

<223> sintética

15 <220>

<221> CDS

<222> (1)..(4479)

20 <400> 73

ES 2 467 918 T3

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	

ES 2 467 918 T3

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

ES 2 467 918 T3

ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln 305 310 315 320	960
gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys 325 330 335	1008
atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp 340 345 350	1056
ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln 355 360 365	1104
atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu 370 375 380	1152
atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn 385 390 395 400	1200
atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg 405 410 415	1248
atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp 420 425 430	1296
ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn 435 440 445	1344
ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly 450 455 460	1392
ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala 465 470 475 480	1440
gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro 485 490 495	1488
gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys 500 505 510	1536
gag atg tcc tgg cac ccc atg tca aag ctt cct ggc aac cct acg ccg Glu Met Ser Trp His Pro Met Ser Lys Leu Pro Gly Asn Pro Thr Pro 515 520 525	1584
ttc ttc tcc cct tca tct tac cct ccg agg gca att tgc ttc atc cct Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys Phe Ile Pro 530 535 540	1632
ttc ccg ggc aat ccc ctt gac aac aac tgc aag gct gga gaa atg ccc Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly Glu Met Pro 545 550 555 560	1680

ES 2 467 918 T3

ctg aac tgg tac aac atg tca gag ttc atg tgt ggc aag gtt tct aac Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys Val Ser Asn 565 570 575	1728
tgc ttg ggc cca gaa ttc gca cgc ttt gac aag tcg aac acc agc cgg Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn Thr Ser Arg 580 585 590	1776
agc cct gct ttt gac ttg gct ctg gtg acc cga gtt gtt gaa gtc aca Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val Glu Val Thr 595 600 605	1824
aac atg gaa cac ggc aag ttt cta aac gtt gat tgc aat cca agc aaa Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn Pro Ser Lys 610 615 620	1872
ggc aca atg gtg ggg gag ttt gac tgt ccc caa gac gcg tgg ttc ttt Gly Thr Met Val Gly Glu Phe Asp Cys Pro Gln Asp Ala Trp Phe Phe 625 630 635 640	1920
gat ggt tcg tgc aac gac ggc cat atg ccg tat tcc att atc atg gaa Asp Gly Ser Cys Asn Asp Gly His Met Pro Tyr Ser Ile Ile Met Glu 645 650 655	1968
atc gga ctg caa acc tca ggt gtt ctc acc tcg gtg ttg aag gca ccg Ile Gly Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro 660 665 670	2016
ctg act atg gac aag gat gac att ctc ttt cga aac ctc gat gca agt Leu Thr Met Asp Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Ser 675 680 685	2064
gct gaa atg gtg cgt cca gac gtg gat gtt cgc ggc aaa acg att cga Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys Thr Ile Arg 690 695 700	2112
aac gtg acc aag tgt acc ggc tat gca atg ttg gga aag atg ggg att Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys Met Gly Ile 705 710 715 720	2160
cac cgg ttc acg ttt gag ttg agc gtt gac ggc gtg gta ttt tat aaa His Arg Phe Thr Phe Glu Leu Ser Val Asp Gly Val Val Phe Tyr Lys 725 730 735	2208
gga tcc act tcc ttt gga tgg ttc act ccc gag gtg ttt gct cag caa Gly Ser Thr Ser Phe Gly Trp Phe Thr Pro Glu Val Phe Ala Gln Gln 740 745 750	2256
gct gga ctc gac aac ggg aaa aag acg gag ccc tgg tgc aag act aac Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys Lys Thr Asn 755 760 765	2304
aac acc tcg gtt cga aga gtt gaa atc gca tcc gcc aaa gga aaa gag Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys Gly Lys Glu 770 775 780	2352
cag ctg act gag aag ctt ccc gac gca act aat gct caa gtt ctt cgg Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln Val Leu Arg 785 790 795 800	2400
cgt tca gag cag tgt gaa tac ctc gat tac ctc aat att gcc cct gac Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile Ala Pro Asp 805 810 815	2448

ES 2 467 918 T3

tct ggg ctg cat ggg aag ggc tac gcc cac gga cac aaa gac gtt aac	2496
Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys Asp Val Asn	
820 825 830	
ccg caa gac tgg ttc ttc tct tgc cac ttt tgg ttc gat cct gta atg	2544
Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Pro Val Met	
835 840 845	
cca gga tct tta gga att gaa tca atg ttc cag ctt atc gag gcc ttt	2592
Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile Glu Ala Phe	
850 855 860	
gcg gtg gac caa aac att cct gga gag tac aac gta tcc aat ccg acc	2640
Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser Asn Pro Thr	
865 870 875 880	
ttt gcc cat gca cca ggc aaa acg gcg tgg aaa tac cga ggc cag ctc	2688
Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg Gly Gln Leu	
885 890 895	
aca cca aag aac cgt gcg atg gac tgc gag gtg cat atc gtt tca att	2736
Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile Val Ser Ile	
900 905 910	
acc gcc tcc ccc gag aac ggg ggc tac gtt gac atc gtg gcc gat gga	2784
Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val Ala Asp Gly	
915 920 925	
gcg ctt tgg gta gat gga ctt cgc gtg tac gaa gcc aaa gag ctt cga	2832
Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys Glu Leu Arg	
930 935 940	
gtt cgt gtc gtt tcg gca aaa cct caa gca att ccg gat gta caa caa	2880
Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp Val Gln Gln	
945 950 955 960	
cag cca cct agc gca aag gcg gac ccg ggg aaa aca gga gtt gca ctt	2928
Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly Val Ala Leu	
965 970 975	
tcg ccc act cag cta cgc gac gtc ctg ctt gaa gtg gac aat cca ttg	2976
Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp Asn Pro Leu	
980 985 990	
tat ctt ggt gta gag aac tcc aat ttg gtg cag ttt gag tcg aaa cct	3024
Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu Ser Lys Pro	
995 1000 1005	
gca act tct tca cgt atc gtt tcg atc aaa ccg tgc tcg att agt	3069
Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser Ile Ser	
1010 1015 1020	
gac ctt ggc gat aag tct ttt atg gaa acg tac aac gtg tca gca	3114
Asp Leu Gly Asp Lys Ser Phe Met Glu Thr Tyr Asn Val Ser Ala	
1025 1030 1035	
cct ctg tat act ggc gcc atg gcc aag ggc att gcc tcg gcg gac	3159
Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp	
1040 1045 1050	
ctc gtc atc gcc gcc ggc aag cgc aag atc ctc ggc tcc ttt ggc	3204
Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly	
1055 1060 1065	

gcc ggc ggc ctc ccc atg cac cac gtg cgc gcc gcc ctc gag aag Ala Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys 1070 1075 1080	3249
atc cag gcc gcc ctg cct cag ggc ccc tac gcc gtc aac ctc atc Ile Gln Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile 1085 1090 1095	3294
cac tcg cct ttt gac agc aac ctc gag aag ggc aac gtc gat ctc His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu 1100 1105 1110	3339
ttc ctc gag aag ggc gtc act gtg gtg gag gcc tcg gca ttc atg Phe Leu Glu Lys Gly Val Thr Val Val Glu Ala Ser Ala Phe Met 1115 1120 1125	3384
acc ctc acc ccg cag gtc gtg cgc tac cgc gcc gcc ggc ctc tcg Thr Leu Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Ser 1130 1135 1140	3429
cgc aac gcc gac ggt tcg gtc aac atc cgc aac cgc atc atc ggc Arg Asn Ala Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly 1145 1150 1155	3474
aag gtc tcg cgc acc gag ctc gcc gag atg ttc atc cgc ccg gcc Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe Ile Arg Pro Ala 1160 1165 1170	3519
ccg gag cac ctc ctc gag aag ctc atc gcc tcg ggc gag atc acc Pro Glu His Leu Leu Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr 1175 1180 1185	3564
cag gag cag gcc gag ctc gcg cgc cgc gtt ccc gtc gcc gac gat Gln Glu Gln Ala Glu Leu Ala Arg Arg Val Pro Val Ala Asp Asp 1190 1195 1200	3609
atc gct gtc gag gct gac tcg ggc ggc cac acc gac aac cgc ccc Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg Pro 1205 1210 1215	3654
atc cac gtc atc ctc ccg ctc atc atc aac ctc cgc aac cgc ctg Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu 1220 1225 1230	3699
cac cgc gag tgc ggc tac ccc gcg cac ctc cgc gtc cgc gtt ggc His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val Arg Val Gly 1235 1240 1245	3744
gcc gga ggt ggc gtc ggc tgc ccg cag gcc gcc gcc gcc gcc ggc ctc Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu 1250 1255 1260	3789
acc atg ggc gcc gcc ttc atc gtc acc ggc act gtc aac cag gtc Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln Val 1265 1270 1275	3834
gcc aag cag tcc ggc acc tgc gac aac gtg cgc aag cag ctc tcg Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser 1280 1285 1290	3879
cag gcc acc tac tcg gat atc tgc atg gcc ccg gcc gcc gac atg Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met 1295 1300 1305	3924

ttc gag gag ggc gtc aag ctc cag gtc ctc aag aag gga acc atg	3969
Phe Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met	
1310 1315 1320	
ttc ccc tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac	4014
Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr	
1325 1330 1335	
gac tcc ttc gac tcc atg cct cct gcc gag ctc gag cgc atc gag	4059
Asp Ser Phe Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu	
1340 1345 1350	
aag cgt atc ttc aag cgc gca ctc cag gag gtc tgg gag gag acc	4104
Lys Arg Ile Phe Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr	
1355 1360 1365	
aag gac ttt tac att aac ggt ctc aag aac ccg gag aag atc cag	4149
Lys Asp Phe Tyr Ile Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln	
1370 1375 1380	
cgc gcc gag cac gac ccc aag ctc aag atg tcg ctc tgc ttc cgc	4194
Arg Ala Glu His Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg	
1385 1390 1395	
tgg tac ctt ggt ctt gcc agc cgc tgg gcc aac atg ggc gcc ccg	4239
Trp Tyr Leu Gly Leu Ala Ser Arg Trp Ala Asn Met Gly Ala Pro	
1400 1405 1410	
gac cgc gtc atg gac tac cag gtc tgg tgt ggc ccg gcc att ggc	4284
Asp Arg Val Met Asp Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly	
1415 1420 1425	
gcc ttc aac gac ttc atc aag ggc acc tac ctc gac ccc gct gtc	4329
Ala Phe Asn Asp Phe Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val	
1430 1435 1440	
tcc aac gag tac ccc tgt gtc gtc cag atc aac ctg caa atc ctc	4374
Ser Asn Glu Tyr Pro Cys Val Val Gln Ile Asn Leu Gln Ile Leu	
1445 1450 1455	
cgt ggt gcc tgc tac ctg cgc cgt ctc aac gcc ctg cgc aac gac	4419
Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp	
1460 1465 1470	
ccg cgc att gac ctc gag acc gag gat gct gcc ttt gtc tac gag	4464
Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu	
1475 1480 1485	
ccc acc aac gcg ctc	4479
Pro Thr Asn Ala Leu	
1490	

<210> 74

5 <211> 1493

<212> PRT

<213> Secuencia artificial

10

<220>

<223> Constructo sintético

15 <400> 74

ES 2 467 918 T3

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 2 467 918 T3

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
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

ES 2 467 918 T3

Glu Met Ser Trp His Pro Met Ser Lys Leu Pro Gly Asn Pro Thr Pro
 515 520 525

Phe Phe Ser Pro Ser Ser Tyr Pro Pro Arg Ala Ile Cys Phe Ile Pro
 530 535 540

Phe Pro Gly Asn Pro Leu Asp Asn Asn Cys Lys Ala Gly Glu Met Pro
 545 550 555 560

Leu Asn Trp Tyr Asn Met Ser Glu Phe Met Cys Gly Lys Val Ser Asn
 565 570 575

Cys Leu Gly Pro Glu Phe Ala Arg Phe Asp Lys Ser Asn Thr Ser Arg
 580 585 590

Ser Pro Ala Phe Asp Leu Ala Leu Val Thr Arg Val Val Glu Val Thr
 595 600 605

Asn Met Glu His Gly Lys Phe Leu Asn Val Asp Cys Asn Pro Ser Lys
 610 615 620

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

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

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

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

Ala Glu Met Val Arg Pro Asp Val Asp Val Arg Gly Lys Thr Ile Arg
 690 695 700

Asn Val Thr Lys Cys Thr Gly Tyr Ala Met Leu Gly Lys Met Gly Ile
 705 710 715 720

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

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

Ala Gly Leu Asp Asn Gly Lys Lys Thr Glu Pro Trp Cys Lys Thr Asn
 755 760 765

ES 2 467 918 T3

Asn Thr Ser Val Arg Arg Val Glu Ile Ala Ser Ala Lys Gly Lys Glu
 770 775 780

Gln Leu Thr Glu Lys Leu Pro Asp Ala Thr Asn Ala Gln Val Leu Arg
 785 790 795 800

Arg Ser Glu Gln Cys Glu Tyr Leu Asp Tyr Leu Asn Ile Ala Pro Asp
 805 810 815

Ser Gly Leu His Gly Lys Gly Tyr Ala His Gly His Lys Asp Val Asn
 820 825 830

Pro Gln Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Pro Val Met
 835 840 845

Pro Gly Ser Leu Gly Ile Glu Ser Met Phe Gln Leu Ile Glu Ala Phe
 850 855 860

Ala Val Asp Gln Asn Ile Pro Gly Glu Tyr Asn Val Ser Asn Pro Thr
 865 870 875 880

Phe Ala His Ala Pro Gly Lys Thr Ala Trp Lys Tyr Arg Gly Gln Leu
 885 890 895

Thr Pro Lys Asn Arg Ala Met Asp Cys Glu Val His Ile Val Ser Ile
 900 905 910

Thr Ala Ser Pro Glu Asn Gly Gly Tyr Val Asp Ile Val Ala Asp Gly
 915 920 925

Ala Leu Trp Val Asp Gly Leu Arg Val Tyr Glu Ala Lys Glu Leu Arg
 930 935 940

Val Arg Val Val Ser Ala Lys Pro Gln Ala Ile Pro Asp Val Gln Gln
 945 950 955 960

Gln Pro Pro Ser Ala Lys Ala Asp Pro Gly Lys Thr Gly Val Ala Leu
 965 970 975

Ser Pro Thr Gln Leu Arg Asp Val Leu Leu Glu Val Asp Asn Pro Leu
 980 985 990

Tyr Leu Gly Val Glu Asn Ser Asn Leu Val Gln Phe Glu Ser Lys Pro
 995 1000 1005

Ala Thr Ser Ser Arg Ile Val Ser Ile Lys Pro Cys Ser Ile Ser
 1010 1015 1020

ES 2 467 918 T3

Asp	Leu	Gly	Asp	Lys	Ser	Phe	Met	Glu	Thr	Tyr	Asn	Val	Ser	Ala
1025						1030					1035			
Pro	Leu	Tyr	Thr	Gly	Ala	Met	Ala	Lys	Gly	Ile	Ala	Ser	Ala	Asp
1040						1045					1050			
Leu	Val	Ile	Ala	Ala	Gly	Lys	Arg	Lys	Ile	Leu	Gly	Ser	Phe	Gly
1055						1060					1065			
Ala	Gly	Gly	Leu	Pro	Met	His	His	Val	Arg	Ala	Ala	Leu	Glu	Lys
1070						1075					1080			
Ile	Gln	Ala	Ala	Leu	Pro	Gln	Gly	Pro	Tyr	Ala	Val	Asn	Leu	Ile
1085						1090					1095			
His	Ser	Pro	Phe	Asp	Ser	Asn	Leu	Glu	Lys	Gly	Asn	Val	Asp	Leu
1100						1105					1110			
Phe	Leu	Glu	Lys	Gly	Val	Thr	Val	Val	Glu	Ala	Ser	Ala	Phe	Met
1115						1120					1125			
Thr	Leu	Thr	Pro	Gln	Val	Val	Arg	Tyr	Arg	Ala	Ala	Gly	Leu	Ser
1130						1135					1140			
Arg	Asn	Ala	Asp	Gly	Ser	Val	Asn	Ile	Arg	Asn	Arg	Ile	Ile	Gly
1145						1150					1155			
Lys	Val	Ser	Arg	Thr	Glu	Leu	Ala	Glu	Met	Phe	Ile	Arg	Pro	Ala
1160						1165					1170			
Pro	Glu	His	Leu	Leu	Glu	Lys	Leu	Ile	Ala	Ser	Gly	Glu	Ile	Thr
1175						1180					1185			
Gln	Glu	Gln	Ala	Glu	Leu	Ala	Arg	Arg	Val	Pro	Val	Ala	Asp	Asp
1190						1195					1200			
Ile	Ala	Val	Glu	Ala	Asp	Ser	Gly	Gly	His	Thr	Asp	Asn	Arg	Pro
1205						1210					1215			
Ile	His	Val	Ile	Leu	Pro	Leu	Ile	Ile	Asn	Leu	Arg	Asn	Arg	Leu
1220						1225					1230			
His	Arg	Glu	Cys	Gly	Tyr	Pro	Ala	His	Leu	Arg	Val	Arg	Val	Gly
1235						1240					1245			
Ala	Gly	Gly	Gly	Val	Gly	Cys	Pro	Gln	Ala	Ala	Ala	Ala	Ala	Leu
1250						1255					1260			

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

<210> 75

5 <211> 4479

<212> ADN

<213> Secuencia artificial

5

<220>

<223> sintética

10 <400> 75

atggcgctcc	gtgtcaagac	gaacaagaag	ccatgctggg	agatgaccaa	ggaggagctg	60
accagcggca	agaccgaggt	gttcaactat	gaggaactcc	tcgagttcgc	agagggcgac	120
atcgccaagg	tcttcggacc	cgagttcgcc	gtcatcgaca	agtaccgcg	ccgcgtgcgc	180
ctgcccgcgc	gcgagtacct	gctcgtgacc	cgcgtcaccc	tcattggacgc	cgaggtcaac	240
aactaccgcg	tcggcgcccc	catggtcacc	gagtacgac	tccccgtcaa	cggagagctc	300
tccgagggcg	gagactgccc	ctgggcccgc	ctggtcgaga	gtggccagtg	cgatctcatg	360
ctcatctcct	acatgggcat	tgacttccag	aaccagggcg	accgcgtcta	ccgcctgctc	420
aacaccacgc	tcacctttta	cggcgtggcc	cacgagggcg	agaccctcga	gtacgacatt	480
cgcgtcaccc	gcttcgcca	gcgtctcgac	ggcggcatct	ccatgttctt	cttcgagtac	540
gactgctacg	tcaacggccg	cctcctcatc	gagatgcgcg	atgggtgcgc	cggcttcttc	600
accaacgagg	agctcgacgc	cggcaagggc	gtcgtcttca	cccgcggcga	cctcgccgcc	660
cgcgccaaga	tcccaaagca	ggacgtctcc	ccctacgcgc	tcgccccctg	cctccacaag	720
accaagctca	acgaaaagga	gatgcagacc	ctcgtcgaca	aggactgggc	atccgtcttt	780
ggctccaaga	acggcatgcc	ggaaatcaac	tacaaactct	gcgcgcgtaa	gatgctcatg	840
attgaccgcg	tcaccagcat	tgaccacaag	ggcgggtgtct	acggcctcgg	tcagctcgtc	900
ggtgaaaaga	tcctcgagcg	cgaccactgg	tactttccct	gccactttgt	caaggatcag	960
gtcatggccg	gatccctcgt	ctccgacggc	tgagccaga	tgctcaagat	gtacatgac	1020
tggctcggcc	tccacctcac	caccggaccc	tttgacttcc	gcccgggtcaa	cggccacccc	1080
aacaaggctc	gctgccgcgg	ccaaatctcc	ccgcacaagg	gcaagctcgt	ctacgtcatg	1140
gagatcaagg	agatgggctt	cgacgaggac	aacgaccctg	acgccattgc	cgacgtcaac	1200
atcattgatg	tcgacttcga	aaagggccag	gacttttagc	tcgaccgcat	cagcgactac	1260
ggcaagggcg	acctcaacaa	gaagatcgtc	gtcgacttta	agggcatcgc	tctcaagatg	1320
cagaagcgct	ccaccaacaa	gaaccctccc	aaggttcagc	ccgtctttgc	caacggcgcc	1380
gccactgtcg	gccccgaggc	ctccaaggct	tcctccggcg	ccagcgccag	cgccagcgcc	1440
gccccggcca	agcctgcctt	cagcgccgat	gttcttgccg	ccaagcccgt	tgcccttccc	1500
gagcacatcc	tcaagggcga	cgccctcgcc	ccaaggaga	tgtcctggca	ccccatgagc	1560
aagctccccg	gcaacccac	ccccttcttt	tctccctcgt	cgtatccgcc	ccgtgctatc	1620
tgcttcatcc	cttttccggg	caaccgcgtc	gacaacaact	gcaaggccgg	cgagatgccc	1680

ctcaactggg	acaacatgag	cgaattcatg	tgcggtaagg	tcagcaactg	cctcggcccc	1740
gagttcgcgc	gctttgacaa	gagcaacacc	tcccgcctccc	ccgccttcga	ccttgcgctc	1800
gtcacgcgcg	tcgtcgaggt	caccaacatg	gagcacggca	agttcctcaa	cgtcgactgt	1860
aacccttcga	agggcacgat	ggtgggagag	ttcgactgcc	cgcaggacgc	ctggttcttt	1920
gacggctcct	gcaatgacgg	tcatatgccg	tactccatca	tcatggagat	cggccttcag	1980
acgtctgggt	tcctcacgtc	ggtgctcaag	gctcctctca	cgatggacaa	ggacgacatc	2040
ctcttttcga	accttgatgc	ctcggctgag	atggtgcgcc	ctgacgtcga	tgtccgcggc	2100
aagaccatcc	gcaacgtgac	caagtgtacc	ggctacgcca	tgcttggtaa	gatgggtatc	2160
catcgcttta	cctttgagct	ctccgtggac	ggcgtcgtct	tttacaaggg	cagcaccagc	2220
ttcggctggg	ttacgccgga	agtgtttgcc	cagcaggccg	gtctggacaa	cggcaagaag	2280
accgagccct	ggtgcaagac	caacaacacc	tccgtccgcc	gcgtcgagat	cgccagcgct	2340
aagggttaagg	agcagctcac	ggagaagctc	cctgatgcta	ccaacgcccc	ggttctccgc	2400
cgtcccgagc	agtgcgagta	cctcgactac	ctcaacatcg	cccccgactc	cggccttcac	2460
ggcaaggggt	acgcgcacgg	ccacaaggac	gtcaaccccc	aggactgggt	tttctcgtgc	2520
cacttctggg	tcgacccggt	gatgcccggc	tcgcttggca	ttgagtccat	gttccagctc	2580
atcgaagcct	ttgctgtcga	ccagaatatc	cctggcgagt	acaatgtttc	gaaccgcacc	2640
tttgcccatg	cccccggtaa	gaccgcttgg	aagtatcgcg	gccagctcac	tcccaagaac	2700
cgcgctatgg	actgcgaggt	ccacatcgtc	agcatcacgg	cgaagccgga	gaacggcggc	2760
tacgtcgaca	tcgtcgccga	cggcgccctt	tgggtcgatg	gcctccgcgt	ttacgaggcc	2820
aaggagctcc	gtgtccgcgt	cgtcagcgcc	aagccgcaag	ctatcccgga	tgtccagcag	2880
cagcctccct	cggctaaggc	cgaccctggc	aagaccggtg	tcgccctcag	ccccaccag	2940
ctccgcgatg	tcctcctgga	ggtcgacaac	cctctgtacc	tcggtgtcga	gaactccaac	3000
cttgtccagt	tcgaaagcaa	gccggcgacg	tcctcccgca	tcgtgtccat	taagccgtgt	3060
tctatctccg	acctcggcga	caagtctttc	atggagactt	acaacgtgtc	tgcgcctctg	3120
tatactggcg	ccatggccaa	gggcattgcc	tcggcggaac	tcgtcatcgc	cgcgggcaag	3180
cgaagatcc	tcggctcctt	tggcgccggc	ggcctcccca	tgcaccacgt	gcgcgcgcgc	3240
ctcgagaaga	tccaggccgc	cctgcctcag	ggccccctacg	ccgtcaacct	catccactcg	3300
ccttttgaca	gcaacctcga	gaagggaac	gtcgatctct	tcctcgagaa	gggcgtcact	3360
gtgggtggagg	cctcggcatt	catgaccctc	accccgaggg	tcgtgcgcta	ccgcgcgcgc	3420
ggcctctcgc	gcaacgccga	cggttcggtc	aacatccgca	accgcatcat	cggcaaggtc	3480
tcgcgcaccg	agctcgccga	gatgttcatc	cgcgcggccc	cggagcacct	cctcgagaag	3540
ctcatcgccct	cgggcgagat	caccagggag	caggccgagc	tcgcgcgcgc	cgttcccgtc	3600

ES 2 467 918 T3

gccgacgata tcgctgtcga ggctgactcg ggcggccaca ccgacaaccg ccccatccac 3660
gtcatcctcc cgctcatcat caacctccgc aaccgcctgc accgcgagtg cggctacccc 3720
gcgcacctcc gcgtccgcgt tggcgccgga ggtggcgctcg gctgcccga ggccgcccgc 3780
gccgcgctca ccatggggcg cgccttcacg gtcaccggca ctgtcaacca ggtcgccaag 3840
cagtccggca cctgcgacaa cgtgcgcaag cagctctcgc aggccaccta ctcgatatc 3900
tgcattggccc cggccgcccga catgttcgag gagggcgctca agctccaggt cctcaagaag 3960
ggaaccatgt tccccctcgc cgccaacaag ctctacgagc tcttttgcaa gtacgactcc 4020
ttcgactcca tgcctcctgc cgagctcgag cgcctcgaga agcgtatctt caagcgcgca 4080
ctccaggagg tctgggagga gaccaaggac ttttacatta acggtctcaa gaaccggag 4140
aagatccagc gcgccgagca cgaccccaag ctcaagatgt cgctctgctt ccgctgggtac 4200
cttgggtcttg ccagccgctg ggccaacatg ggcgccccgg accgcgtcat ggactaccag 4260
gtctgggtgtg gcccgggcat tggcgccctc aacgacttca tcaagggcac ctacctcgac 4320
cccgtgtgtt ccaacgagta cccctgtgtc gtccagatca acctgcaaat cctccgtggt 4380
gcctgctacc tgcgcccgtt caacgccctg cgcaacgacc cgcgcattga cctcgagacc 4440
gaggatgctg cctttgtcta cgagcccacc aacgcgctc 4479

<210> 76

5 <211> 4200

<212> ADN

<213> *Schizochytrium* sp.

10

<400> 76

tttctctctc tcgagctgtt gctgctgctg ctgctgctgc tgcttccttg ctggttctca 60
cgtccgttcg atcaagcgtt cgctcgctcg accgatcggg gcgtgctgct gtgctgaggt 120
cttgttgcca ggcagccgca ggctgtctgt ctgtttgtgt agttttaccc tcgggggttcg 180
gggtctgcct gcctcccgtt cccgcccgcg gccgcccgtg tccacccgcg tcgcctccgc 240
ccatcgggcc tcgcctcctc gcgccgcacg catcgcgctg atcgcatgca tcagctgccc 300
acgcacgggg ggacgcgcgc cccgcgtccc ccgcccgcgc cgtcgtcgtc tggcgatgcc 360
gtcgccgccc tccttccttc cctcgctccc tcttcctccc gagccccctt gtcttccttc 420
gccccgcag cggcgcgag gaagcgagga gagcggggag gagagaagaa aagaaaagaa 480
aagaaaagaa aataacagcg ccgtctcgcg cagacgcgcg cggccgcgtg cgaggcggcg 540
tgatggggct tctcgtggcg cggctgcggc ctggcccggc ctgcctttg aggtgcaggc 600
tttgggagag aagagtggga cgcggagaag ataagatggt gccatggcgc aggacggaga 660
ggttgctgaa acttcttcga gcggcacagg cgatggcgag agaccgacag ctgccggcg 720
ggaggggatg gatacctccc gaggctggca tggacgagct ggccgcgcgg atctggctgg 780

ES 2 467 918 T3

ccgcgcggcg gtgggtccgg aggcgcgagg ttggttttct tcatacctga taccatacgg 840
 tattcattct tcctctccag gaaggaagca agtcacatag agtatcacta gcctaatagat 900
 ggactctatg ttttagggca cgtcggagca gaaggcgcga gcgattcgaa tgcgagcgat 960
 agatacagca cagagacctt gccggcgacg cggatgcagg cgagcacgca cgcaccgcac 1020
 gcacggcagc ggtgcacgcg ctctcggca gatgcacggg tctgcgccgc gcctttacat 1080
 tttttgattt taggtggtgt gcctgccact ttgaacatca tccacaagtc aacgcagcat 1140
 caagaggcaa gcaagtacat acatccattc gaattcaagt tcaagagacg cagcaacagc 1200
 cgccgctccg ctcaagctgc agctagctgg ctgacagggc tcgctggctg tagtggaata 1260
 ttccattcac ttttctgcat ccgcggccag caggcccgta cgcacgttct ctcgtttgtt 1320
 tgttcgttcg tgcgtgcgtg cgtgcgtccc agctgcctgt ctaatctgcc gcgcgatcca 1380
 acgacctcg gtcgtcgccg caagcgaaac ccgacgccga cctggccaat gccgcaagaa 1440
 tgctaagcgc gcagcaatgc tgagagtaat cttcagccca ccaagtcatt atcgctgccc 1500
 aagtctccat cgcagccaca ttcaggcttt ctctctctct cctccctct ctttctgccc 1560
 ggagagaagg aaagaccgc cgccgccgc tctgcgcctg tgacgggctg tccgttgtaa 1620
 gccctcttag acagttccta ggtgccgggc gccgcgcgc ctccgtcgca ggcacacgta 1680
 ggccggccac ggttcccccc gcaccttcca caccttcttc cccgcagcc ggaccgcgcg 1740
 ccgtctgctt acgcacttcg cgcggccgc gccgcggaac ccgagcgcgt gctgtgggcg 1800
 ccgtcttccg gccgcgtcgg aggtcgtccc cgcgcgcgc tactccgggt cctgtgcggg 1860
 acgtacttaa tattaacagt gggacctcgc acaggacctg acggcagcac agacgtcgcc 1920
 gcctcgcatc gctggggacg caggcgaggc atcccggcgc ggccccgcac cggggaggct 1980
 gcggggcggc ctcttccggc cggcgccgc atcaggcgga tgacgcaaga gccctcgag 2040
 tcgctcgctc gcgggagcgc agcgcggcgc cagcgtggcc aagctccgc ccttctggc 2100
 tggtcgcatg cctgcctgcc tgcctgcctg cgtgcgtgcg tgcgtgcgtg ccttcgtgcg 2160
 tgctgcctt cgtgcgtgcg tgcgtgagt cgcggaaga gggatcatgc gaggatcaat 2220
 cccccgccgc acctcgactt ttgaagaagc cgcgatgcga tgcgatgcga tgcgatgcga 2280
 cgcgataccg tgcgaggcta cgaagcgagt ctggccggcc gtcatacaac gcacgttttc 2340
 gagaaggagg gctggcgagg gcgtgcatgc cggcgaccat tgcgaacgcg gcgtctcgtg 2400
 gctggcgaag gtgcctggag gatctaacga tcgctgctat gatgctatag ctgtgctgat 2460
 ccccggtcca ttccaccacg tctgtgcctg ccgcctgacc tgcgcttggc tttccttcaa 2520
 gttctcctcc gccgggcctt caggaccgag acgagacctg cagctgcagc tagactcgcg 2580
 ctgcctcgcg gaggattcgc cggccgccgg gccggacggg actcgcgagg tcacacggcc 2640
 gccggcgatc gcgatggctg tgctgacgta ctcgctgcgtg gcagccgtac gtcagcgacg 2700

ES 2 467 918 T3

```

ccgcctccgt atttgtgatt cgttagttgg ttgttggttg atttgttgat taatTTTTTT 2760
gttcgtaggc ttggttatag ctaatagttt agtttatact ggtgctcttc ggtgctgatt 2820
tagctcgact tgggtccaca ccactgcccc tctactgtga atggatcaat ggacgcacga 2880
cgggcccgcg aaagtgcgcg agtgaggtaa cctaagcaac ggcggtcttc agaggggacg 2940
cacgccctcc gtcgcagtca gtccagacag gcagaaaagc gtcttaggga ccacgcacgc 3000
acgcacgcac gcacgcacgc ccgcacgcac gctccctccc tcgcgtgcct atTTTTTTtag 3060
gcttccttcc gcacgggcct acctctcgct ccctcgcttc gccgcaccag gcggcagcag 3120
cgatacctgc cggtgccgcc tccgtcacgc gctcagccgc agctcagccc agccgcgagc 3180
tagggtttgt tcgtcctgaa ttgtttgatt tgatttgatt tgatttgatc cgatccgatc 3240
cgatctgatc tgatttgctt tgctttgctt tgtctccctc ccggcgcgga ccaagcgtcc 3300
gtctgcgcgc cgcagcttcc cttcttctcc cagccctcct tctgctcccg cctctcgcgc 3360
aagcacgcag cttcgccgcc gcatccggtc ggtcggtcgg tcgatcgacc cgcttgcgcg 3420
tgctgctgtg gccgggcttt tctccatcgg cgactctttc ttctccatac gtctactac 3480
gtacatacat actgccggct tcctcctctt ccagcgcggc gacggcgga ggctgcgacg 3540
tcgtcgccgc cgcgggcgcc gcgcgcgccg ccgcgcgcgc ccgcgtcgca gggcctcgtc 3600
gccgccgcgc ctccgctccg ctccgaggcc gcgagagggc cgcggcgcg cgatggatgg 3660
atggatggat ggatggatgg atggattttg ttgatcgatg gcggcgcatg ggcggagatg 3720
agcgaggacg agcgcgcgag cgcggcagcc ggattcgagc ggcttcgctc gcctcgcgcc 3780
cgctgccgcg cccgccttgc gagcctgcgc cgcgagcgag cgagcgagcg agcggggctt 3840
tctttgtctc gcgcgcgcgt tggcctcgct tgtcttgtgc ttgcgtagcg ggcgcgcgcg 3900
tggaagatgg ctcatcatt caacccattc acgcacgcac tccggcgcg agagaaggcc 3960
gaggaggagc agcaagcaaa ccaaaagctc tcgcgctcgc ggtctcgggc tcgagcggtc 4020
tcggagagag agtcttgcg cgaccaccgg cagcagcagc agcagcagca gcgctgtcga 4080
gcacgagcac gagcacgagc acgagcacga gcattcgagc aagaggacag acacggttgt 4140
cagcgccatg ctcgctcgat acagaaagag gcgggttggg cgtaaaaaaa aaggagcacg 4200

```

<210> 77

5 <211> 3886

<212> ADN

<213> *Schizochytrium* sp.

10

<220>

<221> misc_feature

15 <222> (2115)..(2115)

<223> n e s a , c , g o t

<400> 77

gatcttgatt gccaaagtctt ggattgtcga ttccgatgaa tcgagctctt tgttgtcgag	60
ctctggcttg ccgagcttcc agaaatagac aaaattgccg agttcctgat tgcggggctc	120
tcgattgcca aggtctggtg gattctcgaa ctctcgattg tcaaaatctt ggtcgtctcg	180
tcggattctt tcttgatttg ttttgtcaag accttgagat tgtgcaaaac cttgatcggt	240
gacaaaccct tgatcgacag cagcctttca tcacgctcag ctcttgatcat tgattatatt	300
ccccctgaca gccaacacct tgatgcaggg tctcaacctt gatcttttga ggccatcatc	360
agcatcacgc cccggcactc accctcaaca ttcgacagcc aacgcttttt tttcttcgac	420
taggatctga gaataaaagc aggtcaccac gaccgtaggc caacgcgaca accatggaaa	480
taaagtgaca acgaacgact tgcaagttta aatgtaaaga gcagcaattg cccgcccaca	540
gacaaatgaa agcaggcgcc gagtcttatt tgaggagggtg ggctgtggc aatgggcgaa	600
agaaaatcaa ggacaaggag agcagggttac gtaccggtat actggtatac gtacatggat	660
ggttcttggc aagttgacgg gatgtgtgcg agtgaccgtg gtagttaacg aaagagccgc	720
aagggaagg aaagcaagag aatgcagact tttccacagg atggatgggt ccgcagcttg	780
ccgcatgatg aaacgctgta tttcacctgg cacgtgggtg cgcacgcgc cacaatgat	840
cgcgggcgcg ggtgtattat acatcttccc cctcaggctc actgccatcc ctccatgcgt	900
cgctcgtgcg aacgacgcaa gccttttcgca tcgtgcagcc tctttctggt aaggcaagag	960
ctaaacccaa acctaaacga aagaacatct ttacctctct ctctctccca ttggtcgcgt	1020
gcgctccgcc gctcgtcctt cctcctgcca gtgtcgcgcc ctaacttccc cctccctcc	1080
ctccctccct cctccctct ctctctccac cgcctctctc tccgcgctgc gtgcggtgct	1140
gcccctggacc aatggcatgc tgctgcacgc tcggcgggatg acgcaagccg ctctcgcaatt	1200
tccggatcag atctcggcgg ggcgtgcgcc gcgggggtcac tgcggacctg ccgcggcccc	1260
tgttctcttc acatccatca tgtcctccaa acctccgcct cctccacgca cgtacgcacg	1320
cccgtctgca cgcgcgcact gccgctgcga aagcaagcgc ccgcccgcg cccggcgacg	1380
ggaaggcggc cgcggtctcc ctccgcggtt gcctcgtctc cgcgcggggc tgggggggca	1440
gcagaaggcg ggtggcgcg gcggcttccg tcttcgtcag cggcctacgt cggcgggcg	1500
gcgcgagact acgcatgccc ttgcgtcatg cgctcgcagg tagccgccgc gggcctagcg	1560
tttccgctgg cgcgcgcctt aagcccccg cgcgcacggg attgccgcga taccgtacgg	1620
ccaagaccgc cgcagacgtc ggccctctcg cggccagcca gccagcagcg cagcggagga	1680
agagcgcgca ggcgcggcgg gagggcgggc gcggagcagc gcagagcggg gcggagcagc	1740
gcggagcaga acgggcagac tcggagcggg caggggcgggc agagctttgg ggtttaagga	1800
ccgggttacc ggcaagtga gcggctgcgg ggagcggctg tgggaggggt gagtacgcaa	1860
gcacgatgcg agcgagagag agacgctgcc gcgaatcaag aaggtaggcg cgctgcgagg	1920
cgcgggcgcg gagcggagcg agggagaggg agaggagag agaggagagg agacgtcgcc	1980

5

gcggcggggc ctggcctggc ctggtttggc ttggtcagcg cggccttgtc cgagcgtgca 2040
 gctggagttg ggtggattca tttggatttt cttttgtttt tgtttttctc tctttcccgg 2100
 aaagtgttgg ccggnccggtg ttctttgttt tgattttcttc aaaagttttg gtggttgggtt 2160
 ctctctcttg gctctctgtc aggccgtccg gtccacgccc cggcctctcc tctcctctcc 2220
 tctcctctcc tctccgtgcg tatacgtacg tacgtttgta tacgtacata catcccgcgc 2280
 gccgtgcccg cgagggtttg ctacgcctgg agcaatgcga tgcgatgcga tgcgatgcga 2340
 cgcgacgcga cgcgagtcac tggttcgcgc tgtggctgtg gcttgcttgc ttacttgctt 2400
 tcgagctctc ccgctttctt ctttccttct cacgccacca ccaacgaaag aagatcggcc 2460
 ccggcacgcc gctgagaagg gctggcggcg atgacggcac gcgcgcccgc tgccacgttg 2520
 gcgctcgctg ctgctgctgc tgcctgctgt gctgctgctg ctgctgctgc tgcctgcttct 2580
 gcgcgcaggc tttgccacga ggccggcgtg ctggccgctg ccgcttccag tccgcgtgga 2640
 gagatcgaat gagagataaa ctggatggat tcatcgaggg atgaatgaac gatggttgga 2700
 tgccctttttc ctttttcagg tccacagcgg gaagcaggag cgcgtgaatc tgccgccatc 2760
 cgcatacgtc tgcacgcat cgcatcgcat gcacgcatcg ctgcgccgga gccacagacg 2820
 ggcgacaggg cggccagcca gccaggcagc cagccaggca ggcaccagag ggccagagag 2880
 cgcgcctcac gcacgcgccg cagtgcgcgc atcgctcgca gtgcagacct tgattccccg 2940
 cgcggatctc cgcgagcccg aaacgaagag cgcggtacgg gcccatccta gcgtgcctc 3000
 gcaccgcatc gcatcgcatc gcgttcccta gagagtagta ctgcaggaag gcaccatttc 3060
 cgcgctcctc ttcggcgcgga tcgaggcccc cggcgccgcg acgatcgcgg cggccgcggc 3120
 gctggcggcg gccctggcgc tcgcgctggc ggccgcgcg ggcgtctggc cctggcgcg 3180
 gcgggcgccc caggaggagc ggcagcggct gctcgccgcc agagaagagc gcgccgggcc 3240
 cggggaggga cggggaggag aaggagaagg cgcgcaaggc ggccccgaaa gagaagacct 3300
 tggacttgaa cgcgaagaag aagaagaagg agaagaagtt gaagaagaag aagaagaagg 3360
 agaggaaagtt gaagaagacg aggagcaggc gcgttccaag gcgcgttctc ttccggaggc 3420
 gcgttccagc tgcggcgcg ggccgggctg cggggcgggc gcgggcgcg gtgcgggcag 3480
 aggggacgcg cgcgcggagg cggagggggc cgagcgggag cccctgctgc tgcggggcgc 3540
 ccgggcccga ggtgtggcgc gcgcgacgac ggaggcgacg acgccagcgg ccgcgacgac 3600
 aaggccggcg gcgtcggcgg gcggaaggcc ccgcgcggag caggggcggg agcaggacaa 3660
 ggcgaggag caggagcagg gccgggagcg ggagcgggag cgggcggcg agcccaggc 3720
 agaaccat cgagatccag agcgagcaga ggccggccgc gagccgagc ccgcgccgca 3780
 gatcactagt accgctgcgg aatcacagca gcagcagcag cagcagcagc agcagcagca 3840
 gcagcagcag ccacgagagg gagataaaga aaaagcggca gagacg 3886

<211> 20
 <212> ADN
 5 <213> Secuencia artificial
 <220>
 <223> sintética
 10 <400> 78
 catatggcgc tccgtgtcaa 20
 15 <210> 79
 <211> 35
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 25 <223> sintética
 <400> 79
 gccaggaagc ttgacatgg ggtgccagga catct 35
 30 <210> 80
 <211> 37
 35 <212> ADN
 <213> Secuencia artificial
 <220>
 40 <223> sintética
 <400> 80
 45 tcttggcacc ccatgtcaaa gcttctctggc aacccta 37
 <210> 81
 <211> 20
 50 <212> ADN
 <213> Secuencia artificial
 55 <220>
 <223> sintética
 <400> 81
 60 agtatacaga ggtgctgaca 20
 <210> 82
 65 <211> 28

<212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> sintética
 <400> 82
 10 gcaccccatg agcaagctcc ccggcaac 28
 <210> 83
 15 <211> 28
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> sintética
 25 <400> 83
 gtatacagag gcgcagacac gttgtaag 28
 <210> 84
 30 <211> 28
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> sintética
 40 <400> 84
 ctgcagccag atgctcaaga tgtacatg 28
 45 <210> 85
 <211> 31
 <212> ADN
 50 <213> Secuencia artificial
 <220>
 55 <223> sintética
 <400> 85
 ggagcttgct catggggtgc caggacatct c 31
 60 <210> 86
 <211> 10
 65 <212> ADN

<213> Secuencia artificial
 <220>
 5 <223> sintética
 <400> 86
 ggatccatgg 10
 10 <210> 87
 <211> 31
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> sintética
 <400> 87
 25 aagctgtgac agtcaagtc gcaaaacat g 31
 <210> 88
 <211> 15
 30 <212> ADN
 <213> Secuencia artificial
 35 <220>
 <223> sintética
 <400> 88
 40 taaccgggt ctaga 15
 <210> 89
 45 <211> 31
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> sintética
 55 <400> 89
 gactgctagc ttaagcaagg attttctaa c 31
 <210> 90
 60 <211> 31
 <212> ADN
 65 <213> Secuencia artificial

<220>
 <223> sintética
 5 <400> 90
 gactggatcc tctgatgcg gtatttctc c 31
 <210> 91
 10 <211> 12
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> sintética
 20 <400> 91
 ctaggcggcc gc 12
 25 <210> 92
 <211> 31
 <212> ADN
 30 <213> Secuencia artificial
 <220>
 35 <223> sintética
 <400> 92
 gactactagt ctaagaaacc attattatca t 31
 40 <210> 93
 <211> 31
 45 <212> ADN
 <213> Secuencia artificial
 <220>
 50 <223> sintética
 <400> 93
 55 gactggatcc agctttaa atcgggtgc a 31
 <210> 94
 <211> 46
 60 <212> ADN
 <213> Secuencia artificial
 65 <220>

<223> sintética
 <400> 94
 5 gactggatcc gccacatgt tgcagcatatc ttggctacca aaaccc 46
 <210> 95
 <211> 35
 10 <212> ADN
 <213> Secuencia artificial
 15 <220>
 <223> sintética
 <400> 95
 20 gacttctaga tcaataatgc cagaattttg gctgc 35
 <210> 96
 25 <211> 5
 <212> PRT
 <213> Secuencia artificial
 30 <220>
 <223> Motivo de sitio activo
 35 <220>
 <221> misc_feature
 <222> (2) .. (2)
 40 <223> Xaa puede ser cualquier aminoácido que se produce de manera natural
 <220>
 45 <221> misc_feature
 <222> (4) .. (4)
 <223> Xaa puede ser cualquier aminoácido que se produce de manera natural
 50 <400> 96
 Gly Xaa Ser Xaa Gly
 1 5
 55 <210> 97
 <211> 13
 <212> PRT
 60 <213> Secuencia artificial
 <220>
 65 <223> Motivo de sitio activo

<220>
 <221> MISC_FEATURE
 5 <222> (1) .. (1)
 <223> Puede ser Ile
 10 <220>
 <221> misc_feature
 <222> (2) .. (3)
 15 <223> Xaa puede ser cualquier aminoácido que se produce de manera natural
 <220>
 20 <221> misc_feature
 <222> (5) .. (7)
 <223> Xaa puede ser cualquier aminoácido que se produce de manera natural
 25 <220>
 <221> misc_feature
 30 <222> (9) .. (12)
 <223> Xaa puede ser cualquier aminoácido que se produce de manera natural
 <400> 97
 35 Leu Xaa Xaa His Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa Pro
 1 5 10
 <210> 98
 40 <211> 4
 <212> PRT
 <213> *Schizochytrium* sp.
 45 <220>
 <221> misc_feature
 50 <222> (2) .. (2)
 <223> Xaa puede ser cualquier aminoácido que se produce de manera natural
 <400> 98
 55 Asp Xaa Ala Cys
 1
 <210> 99
 60 <211> 4
 <212> PRT
 <213> *Schizochytrium* sp.

<400> 99
 Gly Phe Gly Gly
 1
 5 <210> 100
 <211> 15
 10 <212> PRT
 <213> *Schizochytrium* sp.
 <400> 100
 15 His Gly Ile Ala Asn Pro Thr Phe Val His Ala Pro Gly Lys Ile
 1 5 10 15
 <210> 101
 20 <211> 10
 <212> PRT
 <213> Secuencia artificial
 25 <220>
 <223> Constructo sintético
 30 <400> 101
 Gly Cys Ser Gln Met Leu Lys Met Tyr Met
 1 5 10
 <210> 102
 35 <211> 11
 <212> PRT
 40 <213> Secuencia artificial
 <220>
 <223> Constructo sintético
 45 <400> 102
 Glu Met Ser Trp His Pro Met Ser Lys Leu Pro
 1 5 10

REIVINDICACIONES

1. Sistema de poliketido sintasa (PKS) de ácidos grasos poliinsaturados (AGPI) quimérico, en el que un dominio deshidrasa-2 (DH2) de la proteína transportadora de β -hidroxiacil-acilo similar a FabA de un primer sistema de PKS de AGPI se sustituye por un dominio DH2 de un segundo sistema de PKS de AGPI diferente, para producir un sistema de PKS de AGPI quimérico que produce una razón diferente de AGPI omega-3 con respecto a omega-6 en comparación con el primer sistema de PKS de AGPI.
2. Sistema de PKS de AGPI quimérico según la reivindicación 1, en el que una proteína que comprende el dominio DH2 del primer sistema de PKS de AGPI se sustituye por una proteína homóloga que comprende el dominio DH2 del segundo sistema de PKS de AGPI.
3. Sistema de PKS de AGPI quimérico según la reivindicación 1 o la reivindicación 2, en el que el dominio DH2 del sistema de PKS de AGPI primero o segundo comprende un dominio DH2 de *Schizochytrium* o *Thraustochytrium*.
4. Sistema de PKS de AGPI quimérico según cualquiera de las reivindicaciones 1-3, en el que:
 - (a) el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y el segundo sistema de PKS de AGPI es un sistema de PKS de AGPI de *Thraustochytrium*,
 - (b) el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y el segundo sistema de PKS de AGPI es de un traustokitridio diferente, en el que un OrfC que comprende un dominio DH2 del sistema de PKS de AGPI de *Schizochytrium* se sustituye por un OrfC que comprende un dominio DH2 del traustokitridio diferente,
 - (c) el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y el segundo sistema de PKS de AGPI es de *Thraustochytrium* 23B, en el que un OrfC que comprende un dominio DH2 del sistema de PKS de AGPI de *Schizochytrium* se sustituye por un OrfC que comprende un dominio DH2 de *Thraustochytrium* 23B, o
 - (d) el primer sistema de PKS de AGPI es un sistema de PKS de AGPI de *Schizochytrium*, y el segundo sistema de PKS de AGPI es de *Thraustochytrium* 23B, en el que el dominio DH2 del OrfC del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el dominio DH2 de *Thraustochytrium* 23B para producir una proteína quimérica que comprende el dominio DH2 de *Thraustochytrium* 23B.
5. Sistema de PKS de AGPI quimérico según la reivindicación 4, en el que el OrfC de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*.
6. Sistema de PKS de AGPI quimérico según la reivindicación 5, en el que la secuencia de ácido nucleico comprende SEQ ID NO: 70.
7. Sistema de PKS de AGPI quimérico según cualquiera de las reivindicaciones 4-6, en el que el OrfA del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el OrfA de *Thraustochytrium* 23B.
8. Sistema de PKS de AGPI quimérico según la reivindicación 7, en el que el OrfA de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*.
9. Sistema de PKS de AGPI quimérico según la reivindicación 8, en el que la secuencia de ácido nucleico que codifica para OrfA comprende SEQ ID NO: 71.
10. Sistema de PKS de AGPI quimérico según cualquiera de las reivindicaciones 4-6, en el que el OrfB del sistema de PKS de AGPI de *Schizochytrium* se sustituye por el OrfB de *Thraustochytrium* 23B.
11. Sistema de PKS de AGPI quimérico según la reivindicación 10, en el que el OrfB de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*.
12. Sistema de PKS de AGPI quimérico según la reivindicación 11, en el que la secuencia de ácido nucleico que codifica para OrfB comprende SEQ ID NO: 72.
13. Sistema de PKS de AGPI quimérico según la reivindicación 4, en el que la proteína quimérica que comprende el dominio DH2 de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que comprende SEQ ID NO: 73.
14. Sistema de PKS de AGPI quimérico según la reivindicación 4, en el que el DH2 de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que está optimizada para el uso de codones de *Schizochytrium*.
15. Sistema de PKS de AGPI quimérico según la reivindicación 4 o la reivindicación 14, en el que la proteína

quimérica que comprende el dominio DH2 de *Thraustochytrium* 23B se codifica por una secuencia de ácido nucleico que comprende SEQ ID NO: 75.

16. Sistema de PKS de AGPI quimérico según cualquiera de las reivindicaciones 1-4 ó 7-15, en el que el sistema de PKS de AGPI quimérico comprende una proteína que comprende: una secuencia de aminoácidos que es idéntica en al menos el 95%, 96%, 97%, 98% o el 99% a SEQ ID NO: 74, o una secuencia de aminoácidos de SEQ ID NO: 74.

17. Sistema de PKS de AGPI quimérico según la reivindicación 1, en el que el sistema de PKS de AGPI quimérico:

(a) comprende SEQ ID NO: 2, SEQ ID NO: 4 y SEQ ID NO: 74,

(b) comprende SEQ ID NO: 39, SEQ ID NO: 4 y SEQ ID NO: 62,

(c) comprende SEQ ID NO: 39, SEQ ID NO: 4 y SEQ ID NO: 74,

(d) se codifica por moléculas de ácido nucleico que comprenden SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 70,

(e) se codifica por moléculas de ácido nucleico que comprenden SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 73,

(f) se codifica por moléculas de ácido nucleico que comprenden SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 75, o

(g) se codifica por moléculas de ácido nucleico que comprenden SEQ ID NO: 71, SEQ ID NO: 3 y SEQ ID NO: 70.

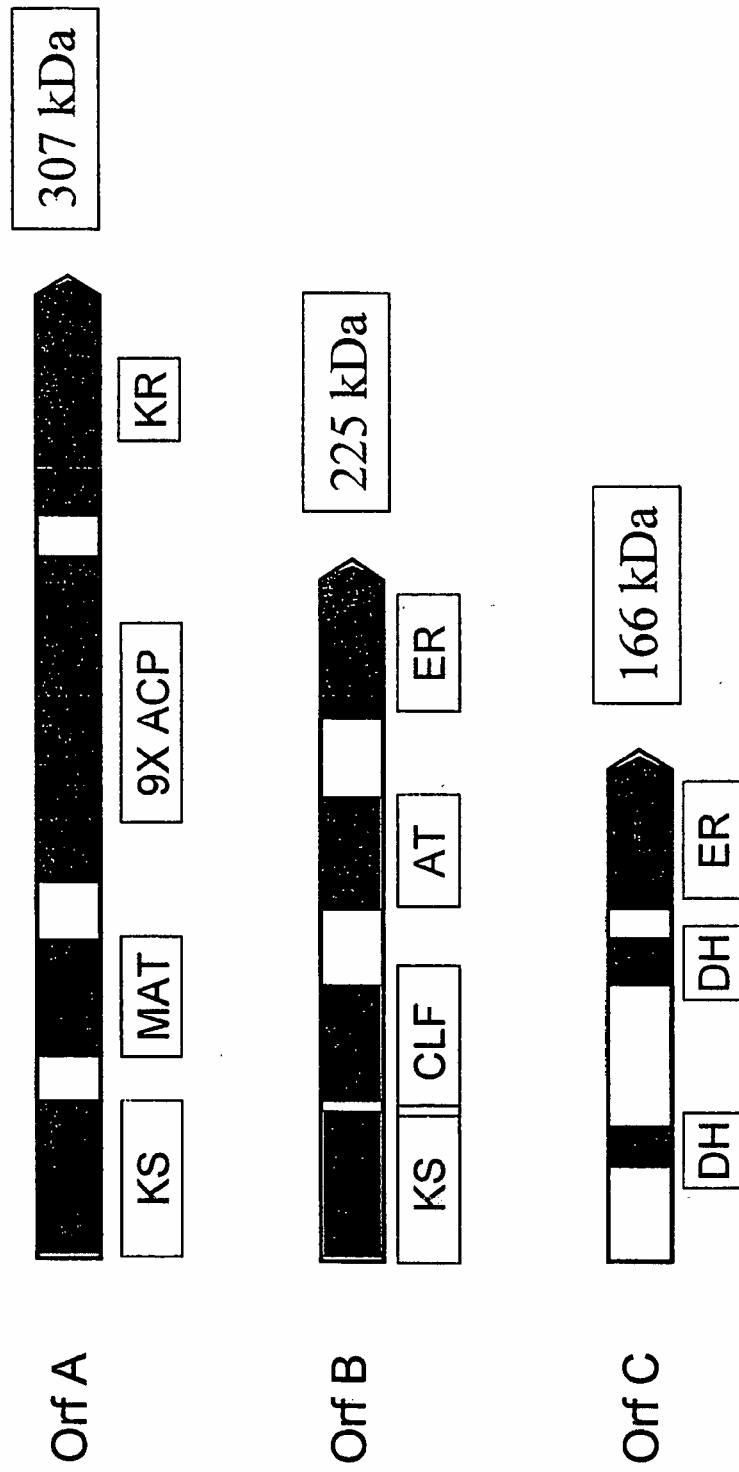
18. Método de alteración de la razón de omega-3 con respecto a omega-6 de AGPI producidos por un primer sistema de PKS de AGPI, que comprende expresar el sistema de PKS de AGPI quimérico según cualquiera de las reivindicaciones 1-17 en un organismo.

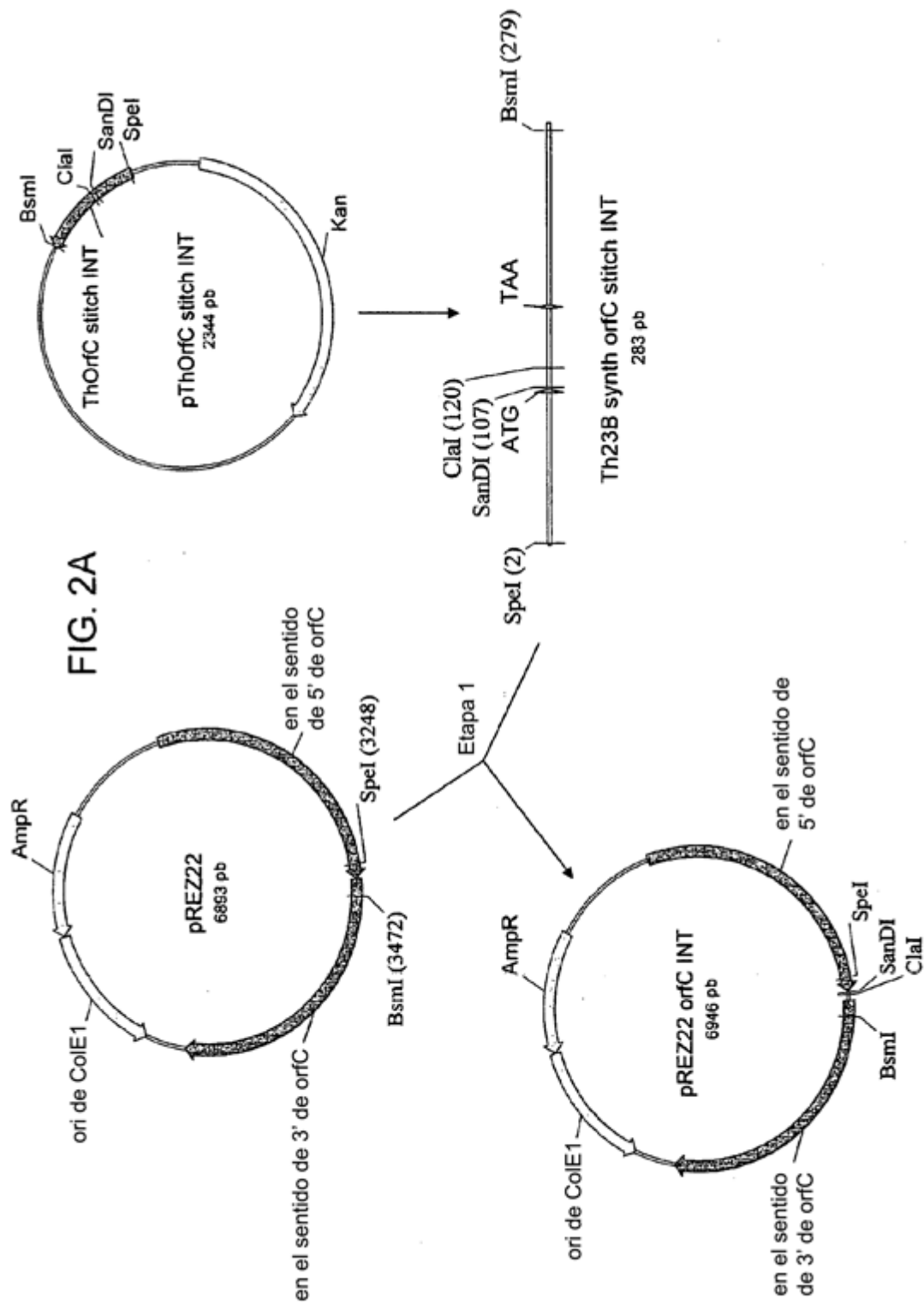
19. Método según la reivindicación 18, en el que el sistema de PKS de AGPI quimérico se expresa por un microorganismo o una planta.

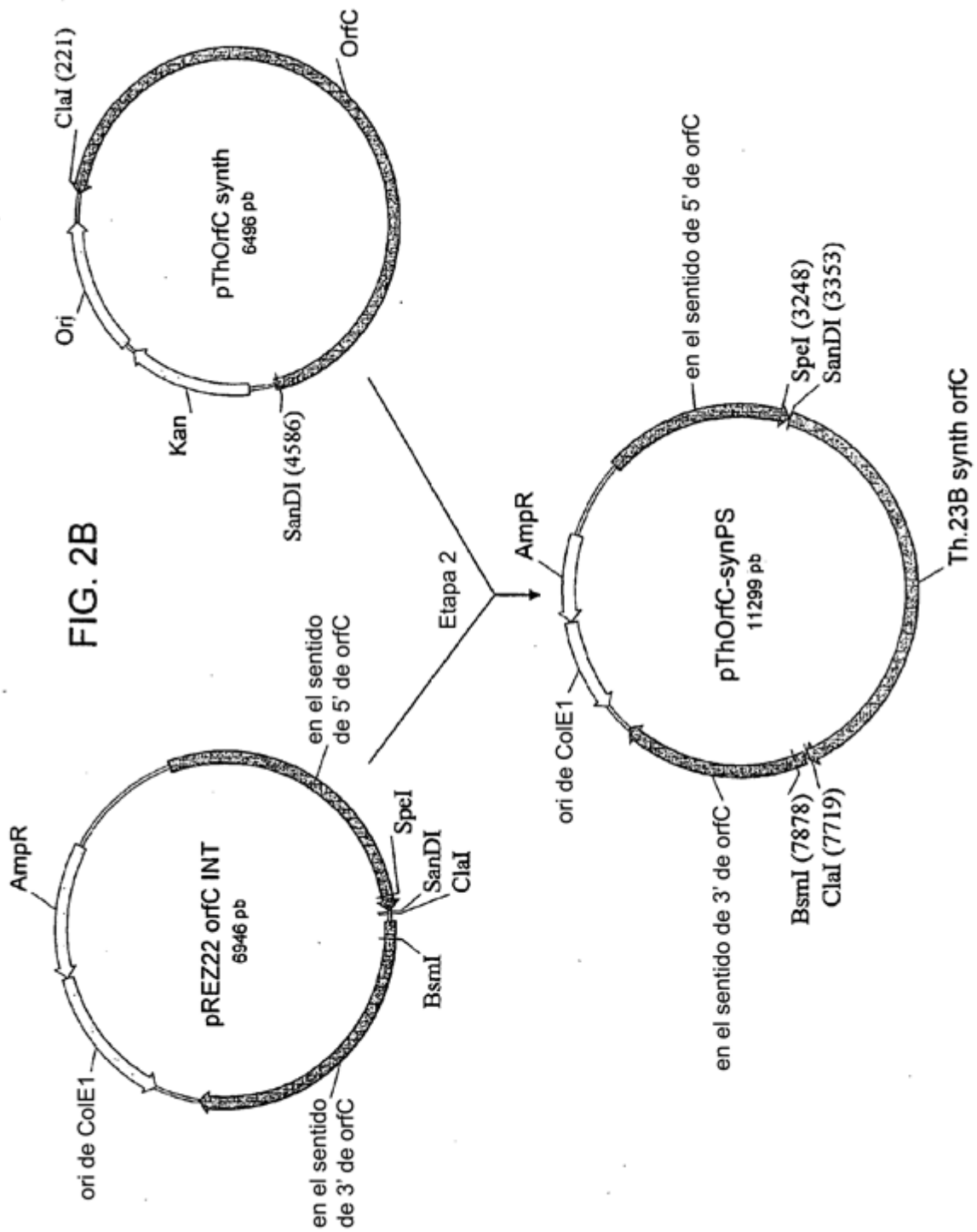
20. Microorganismo o planta o parte de la planta modificado genéticamente, que comprende un sistema de PKS de AGPI quimérico según cualquiera de las reivindicaciones 1-17.

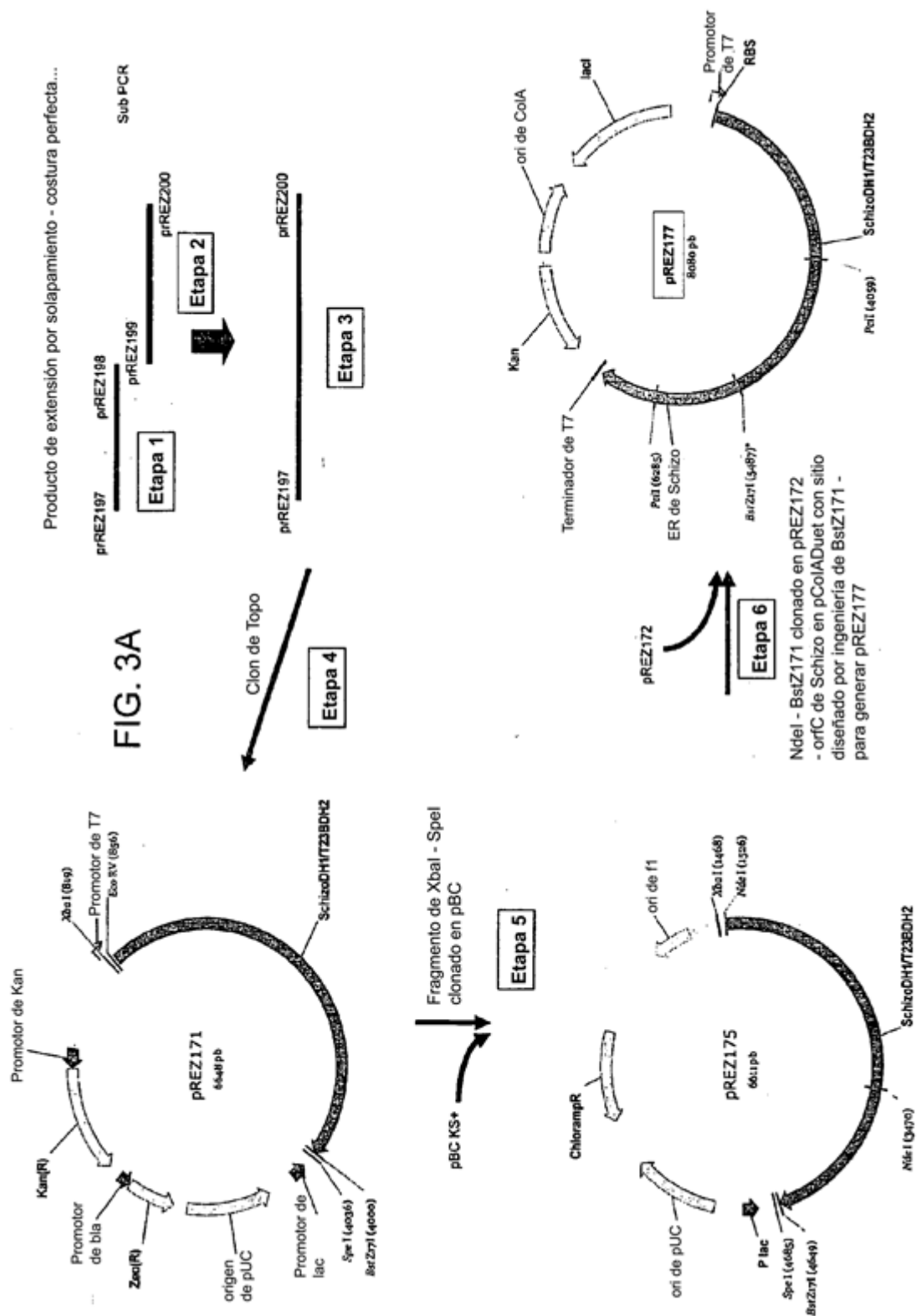
21. Método de aumento de la producción de AGPI y de alteración de la razón de omega-3 con respecto a omega-6 de AGPI producidos por un primer sistema de PKS de AGPI, que comprende expresar un sistema de PKS de AGPI quimérico en un organismo, en el que el dominio DH2 de un primer sistema de PKS de AGPI se sustituye por un dominio DH2 de un segundo sistema de PKS de AGPI diferente, para producir un sistema de PKS de AGPI quimérico que produce una razón diferente de AGPI omega-3 con respecto a omega-6 en comparación con el primer sistema de PKS de AGPI, y en el que el dominio DH2 del segundo sistema de PKS de AGPI está optimizado para el uso de codones del organismo del que se deriva el primer sistema de PKS de AGPI.

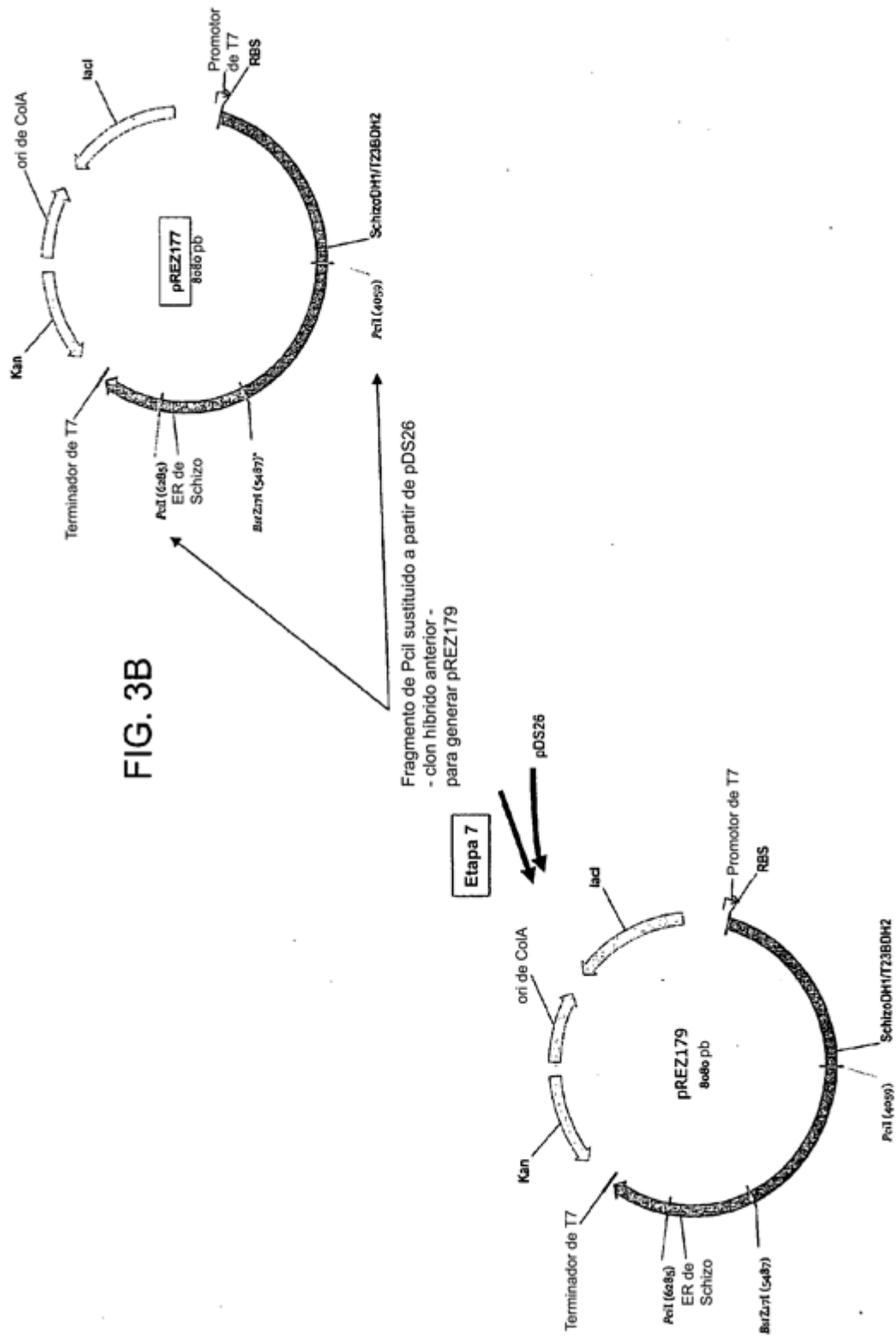
FIG. 1

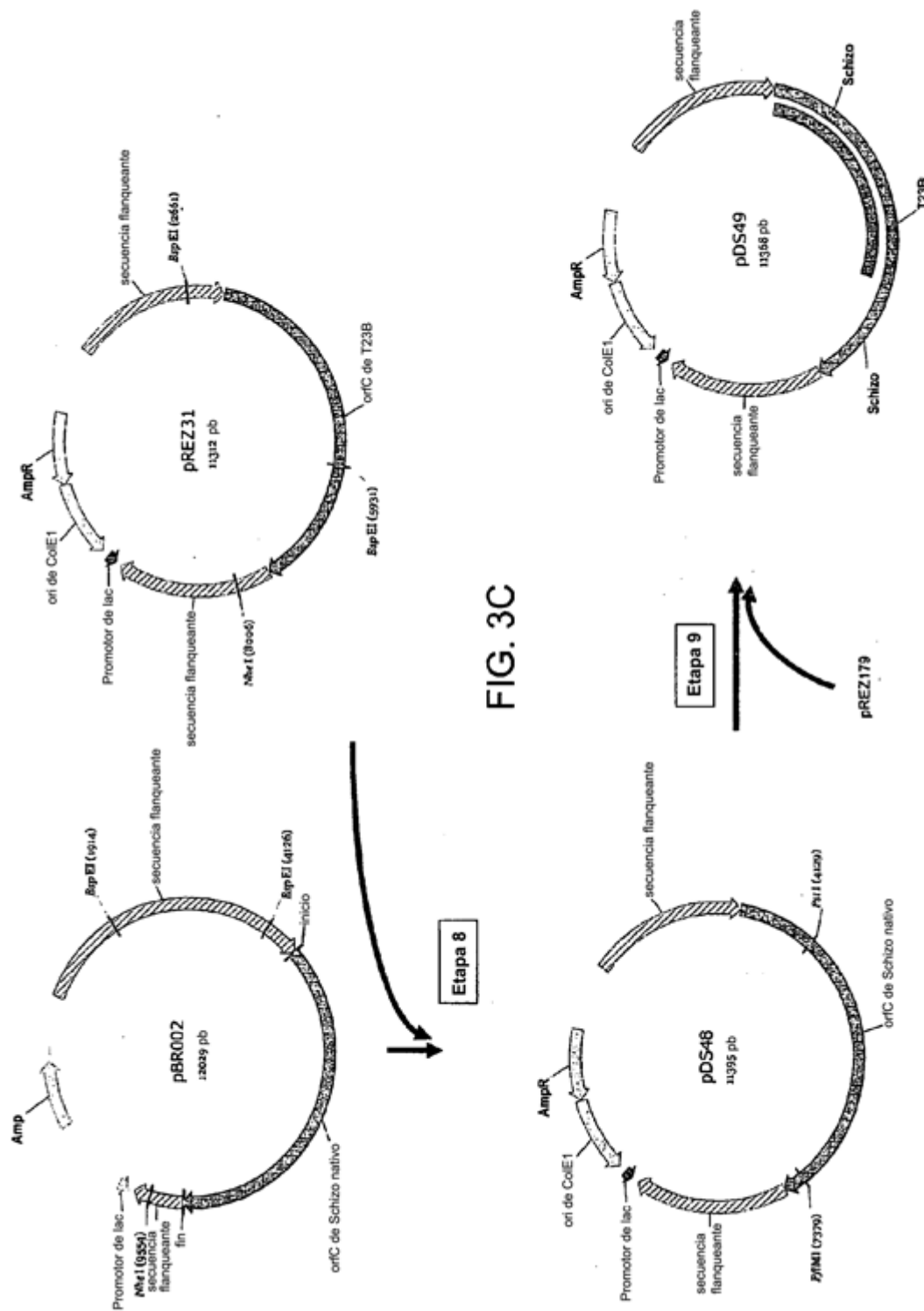






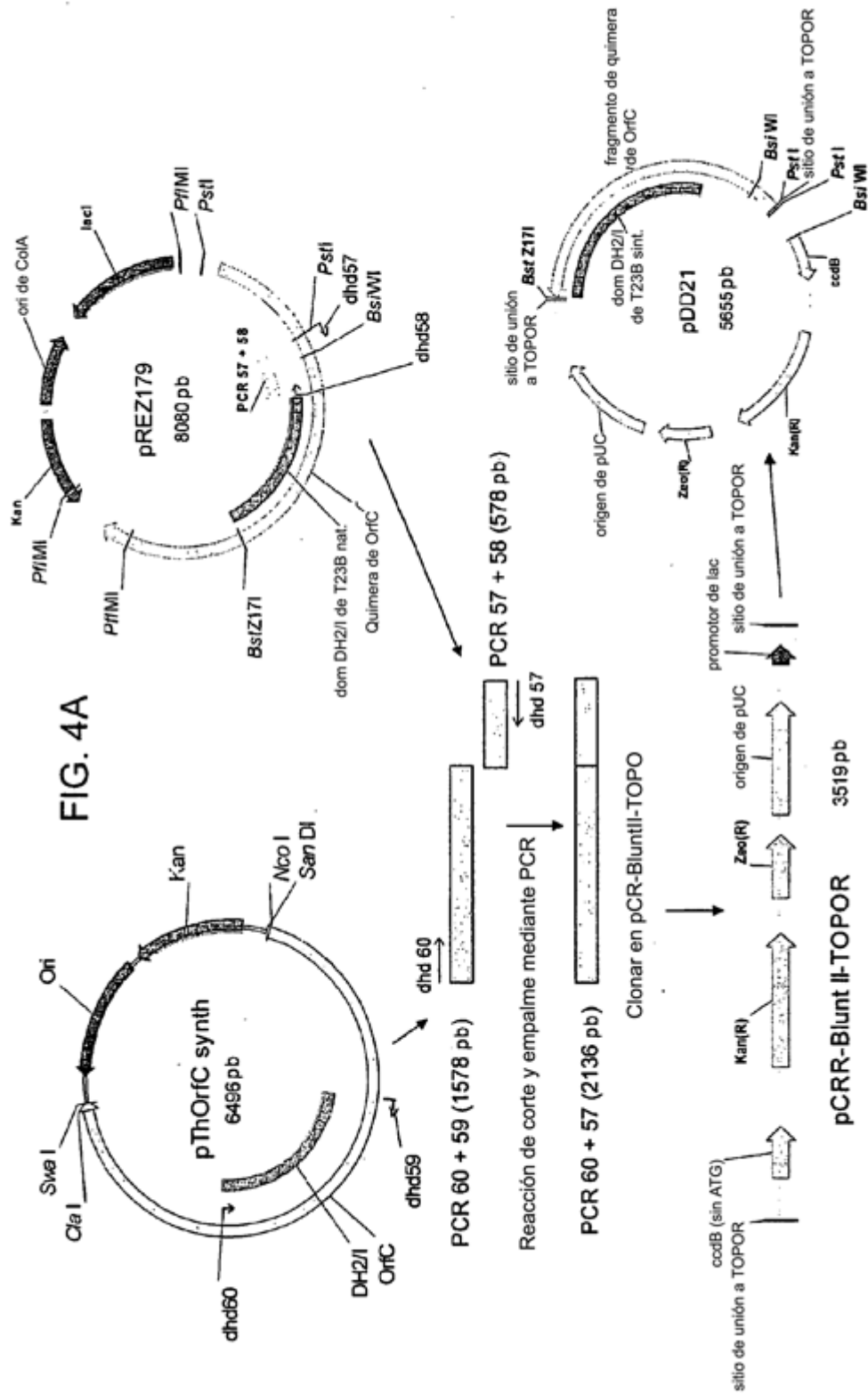






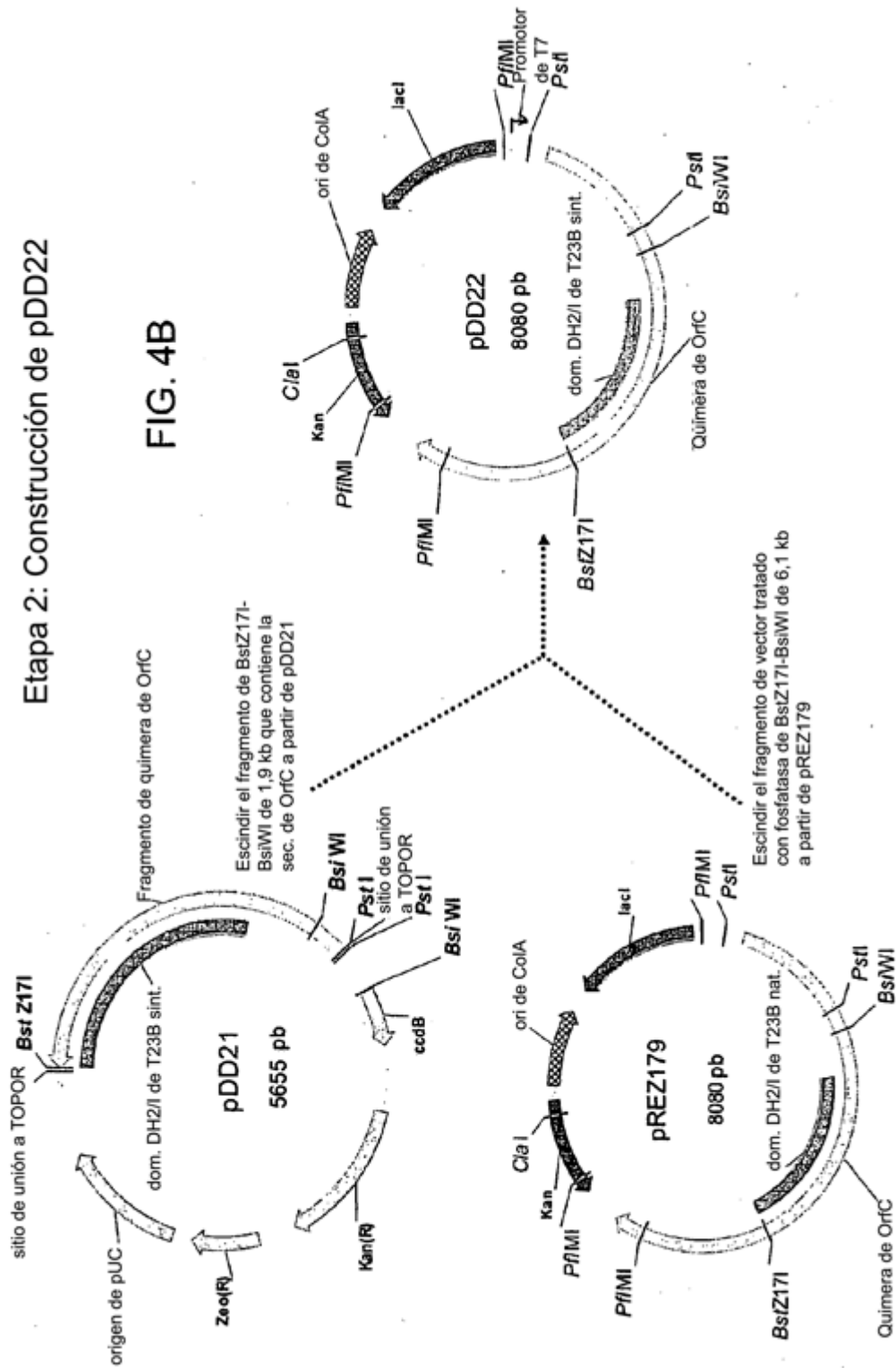
Etapa 1: Construcción de pDD21

FIG. 4A



Etapa 2: Construcción de pDD22

FIG. 4B



Etapa 3: Construcción de pDD24

FIG. 4C

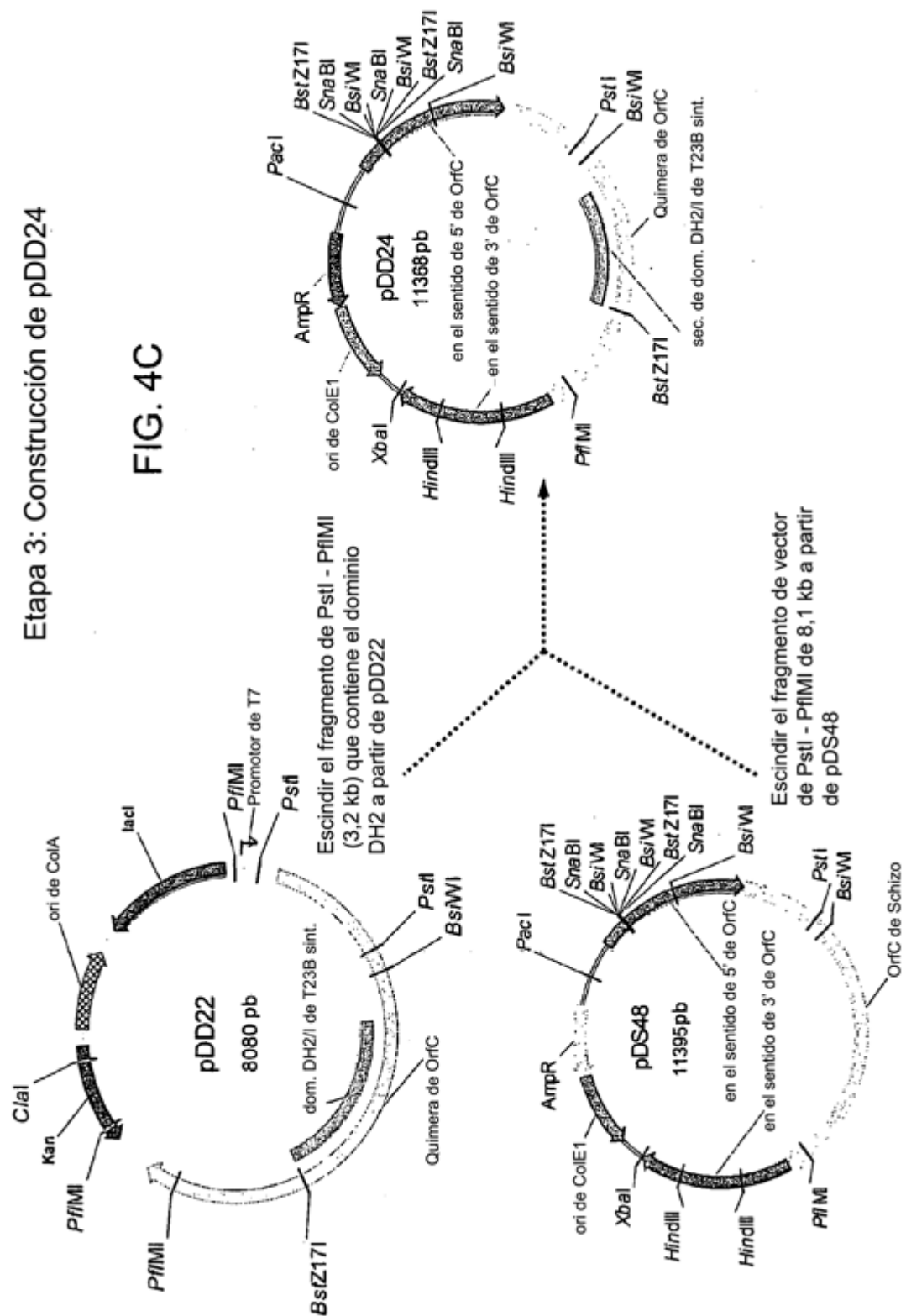


FIG. 5

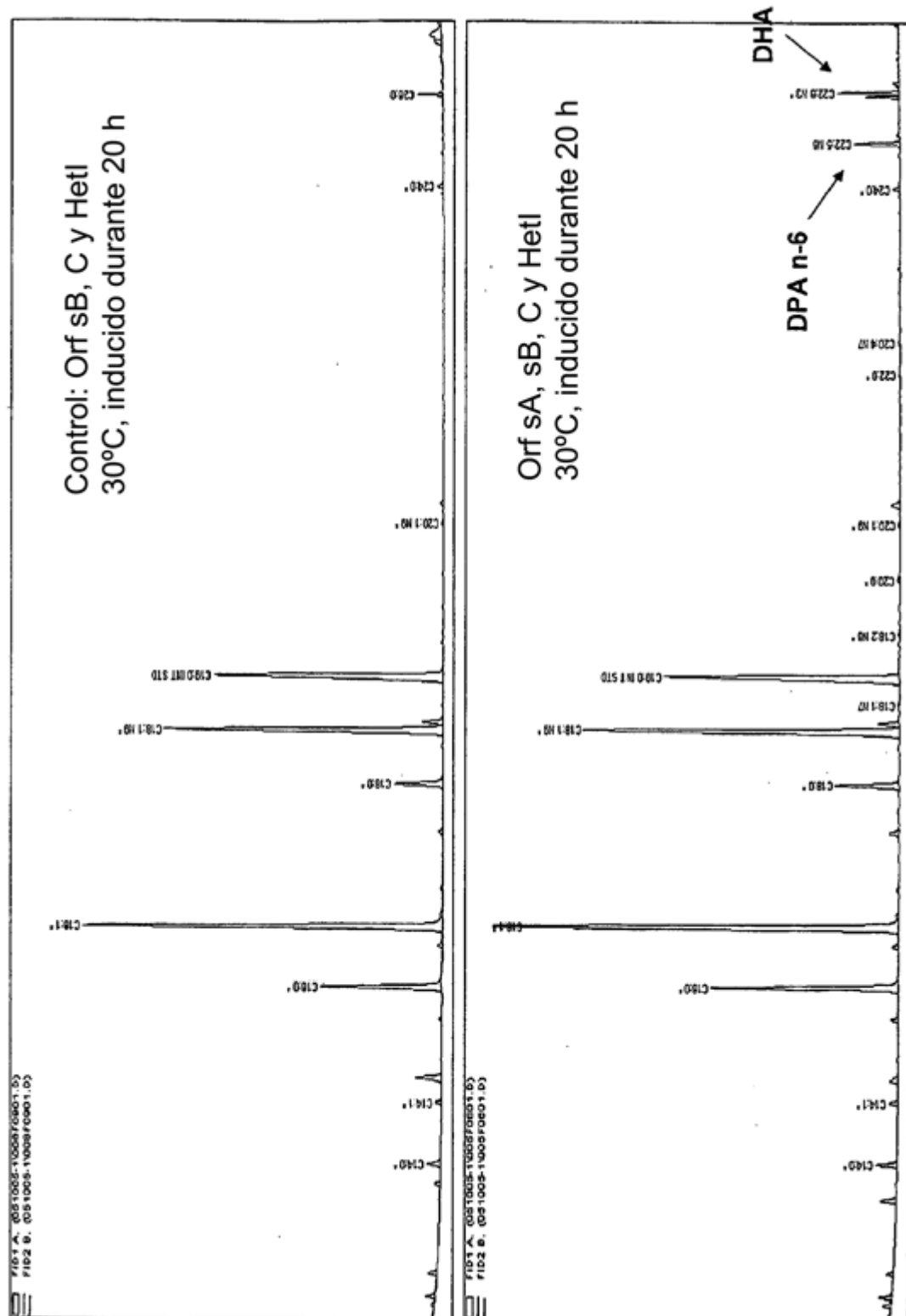


FIG. 6

