

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 698 473**

51 Int. Cl.:

C12N 1/12 (2006.01)

C12N 15/79 (2006.01)

C12P 21/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.04.2002 E 10178911 (3)**

97 Fecha y número de publicación de la concesión europea: **12.09.2018 EP 2314706**

54 Título: **Promotor y procedimiento para la transformación de microorganismos Traustoquitriales**

30 Prioridad:

16.04.2001 US 284116 P

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

04.02.2019

73 Titular/es:

DSM IP ASSETS B.V. (100.0%)

Het Overloon 1

6411 TE Heerlen, NL

72 Inventor/es:

ROESSLER, PAUL G.;

MATTHEWS, T. DAVE;

RAMSEIER, TOM M. y

METZ, JAMES G.

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 698 473 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

Promotor y procedimiento para la transformación de microorganismos Traustoquitriales

5 Campo de la invención

La presente invención se refiere de manera general a una molécula aislada de ácidos nucleicos codificante de una acetolactato sintasa de Traustoquitriales, incluyendo acetolactato sintasas que proporcionan una sensibilidad reducida a compuestos de sulfonilurea, inhibidores de clase imidazolinona y/o oxibenzoatos de pirimidinilo, en microorganismos del orden Traustoquitriales; a moléculas de ácidos nucleicos que comprenden marcadores seleccionables útiles para la transformación de microorganismos del orden Traustoquitriales y a métodos de transformación de dichos microorganismos utilizando moléculas de ácidos nucleicos recombinantes de la presente invención. La presente invención se refiere además a promotores génicos útiles en los sistemas de expresión de los Traustoquitriales. Las moléculas de ácidos nucleicos recombinantes de la presente invención pueden utilizarse para la expresión de ácidos nucleicos foráneos en un microorganismo Traustoquitriales, así como para la delección, mutación o inactivación de genes en microorganismos Traustoquitriales.

15 Antecedentes de la invención

20 Los desarrollos han resultado en una revisión de la taxonomía de los traustoquitridios. Algunos taxónomos teóricos sitúan los traustoquitridios en el grupo de las algas o de los protistas similares a algas. Sin embargo, debido a la incertidumbre taxonómica, sería mejor para los fines de la presente invención considerar las cepas indicadas en la presente invención como traustoquitridios (Orden: Traustoquitriales; Familia: *Thraustochytriaceae*; Género: *Thraustochytrium*, *Schizochytrium*, *Labyrinthuloides*, o *Japonochytrium*). Los cambios de taxonomía se resumen a continuación.

25 Las cepas de determinados microorganismos unicelulares dados a conocer y reivindicados en la presente memoria son miembros del orden Traustoquitriales. Los traustoquitridios son eucariotas marinos con una historia taxonómica problemática. Se proporciona una revisión de los problemas de posicionamiento taxonómico de los traustoquitridios en Moss (1986), Bahnweb y Jackie (1986) y Chamberlain y Moss (1988).

30 Por conveniencia, inicialmente los taxónomos situaron a los traustoquitridios junto con otros eucariotas zoospóricos incoloros dentro de los Ficomisetos (hongos similares a algas). Sin embargo, el nombre Ficomisetos eventualmente perdió su estatus taxonómico y los traustoquitridios se incluyeron en los Oomicetos (hongos zoospóricos biflagelados). Inicialmente se supuso que los Oomicetos estaban relacionados con las algas heterocontas y eventualmente un amplio abanico de estudios ultraestructurales y bioquímicos, resumidos por Barr (1983), apoyan esta premisa. De hecho, los Oomicetos han sido aceptados por Leedale (1974) y otros ficólogos como parte de las algas heterocontas. Sin embargo, por razones de conveniencia resultante de su naturaleza heterotrófica, los Oomicetos y los traustoquitridios han sido estudiados en gran parte por micólogos (científicos que estudian los hongos) y no por ficólogos (científicos que estudian las algas).

35 Desde otra perspectiva taxonómica, los biólogos evolutivos han desarrollado dos escuelas generales de pensamiento sobre cómo han evolucionado los eucariotas. Una teoría propone un origen exógeno de orgánulos unidos a membranas mediante una serie de endosimbiosis (Margulis, 1970); p.ej., las mitocondrias se habrían derivado de endosimbiontes bacterianos; los cloroplastos, de cianofitos, y los flagelos, de espiroquetas. Otra teoría sugiere una evolución gradual de los orgánulos unidos a membranas a partir de sistemas no unidos a membranas del ancestro procariota mediante un proceso autógeno (Cavalier-Smith, 1975). Sin embargo, ambos grupos de biólogos evolutivos han eliminado los Oomicetos y los traustoquitridios de los hongos y los sitúan en las algas cromófitas en el reino de los Chromophyta (Cavalier-Smith, 1981) (este reino recientemente se ha expandido para incluir otros protistas y los miembros de este reino ahora se denominan Stramenopiles) o con todas las algas en el reino Protista (Margulis y Sagan, 1985).

40 Con el desarrollo de la microscopía electrónica, algunos estudios sobre la ultraestructura de las zoosporas de dos géneros de traustoquitridios, *Thrautochytrium* y *Schizochytrium* (Perkins, 1976; Kazama, 1980; Barr, 1981) han proporcionado evidencia firme de que *Thraustochytriaceae* se relaciona sólo distantemente a los Oomicetos. Además, los datos genéticos de un análisis de correspondencias (una forma de estadística multivariante) de las secuencias de ARN ribosómico 5S indican que los *Traustoquitriales* son claramente un grupo único de los eucariotas, completamente separado de los hongos, y más estrechamente relacionado con las algas rojas y pardas y con miembros de los Oomicetos (Mannella et al., 1987). La mayoría de taxónomos ha acordado eliminar los traustoquitridios de los Oomicetos (Bartnicki-García, 1988).

45 En resumen, utilizando el sistema taxonómico de Cavalier-Smith (1981, 1983), los traustoquitridios se clasifican con las algas cromófitas en el reino Chromophyta (Stramenopiles). Esto las sitúa en un reino completamente diferente de los hongos, que están dentro del reino Eufungi. Por lo tanto, la posición taxonómica de los traustoquitridios se resume a continuación:

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

Algunos de los primeros taxónomos separaron unos cuantos miembros originales del género *Thraustochytrium* (aquellos con un estadio vital amebode) en un género separado llamado *Ulkenia*. Sin embargo, ahora se conoce que la mayoría, si no todos, los traustoquitrídios (incluyendo *Thraustochytrium* y *Schizochytrium*) muestran estadios ameboides y, como tales, *Ulkenia* no es considerado por algunos un género válido. Tal como se utiliza en la presente memoria, el género *Thraustochytrium* incluirá *Ulkenia*.

A pesar de la incertidumbre en las posiciones taxonómicas dentro de las clasificaciones superiores de Filo y Reino, los traustoquitrídios siguen siendo un grupo claramente diferenciado y característico cuyos miembros siguen siendo clasificables dentro del orden Traustoquitriales.

Schizochytrium y otros microorganismos traustoquitriales presentan un valor comercial real y potencial sustancial debido a su capacidad de producir grandes cantidades de compuestos lipoides, incluyendo ácidos grasos altamente insaturados (HDEA, por sus siglas en inglés) y diversos carotenoides (p.ej., astaxantina). Los ácidos grasos altamente insaturados omega-3 son de interés comercial significativo en el aspecto de que recientemente se han reconocido como importantes compuestos en la dieta para prevenir la arterioesclerosis y la enfermedad cardíaca coronaria, para aliviar condiciones inflamatorias y para retrasar el crecimiento de las células tumorales. Estos efectos beneficiosos son el resultado tanto de que los HUPA omega-3 causan inhibición competitiva de compuestos producidos a partir de ácidos grasos omega-6 como de los compuestos beneficiosos producidos directamente a partir de los HUFA omega-3 mismos (Simopoulos et al., 1986). Los ácidos grasos omega-6 son los HUFA predominantes observados en plantas y animales. Por lo tanto, el desarrollo ulterior de los microorganismos Traustoquitriales como organismos de producción comerciales se beneficiará significativamente de la capacidad de producir cambios genéticos específicos en los organismos mediante tecnología de DNA recombinante, incluyendo la potenciación de la producción de los altamente valiosos HUFA y carotenoides por dichos organismos. Además, la capacidad de obtener una mejor comprensión de la bioquímica y biología molecular de este poco caracterizado grupo de organismos proporcionará información valiosa que puede utilizarse para guiar futuros esfuerzos de desarrollo de cepas. Sin embargo, anteriormente a la presente invención, no se disponía de métodos y constructos recombinantes adecuados para transformar traustoquitrídios, incluyendo miembros de los géneros *Schizochytrium* y *Thraustochytrium*. Resulta importante que el desarrollo de marcadores seleccionables que resultan particularmente útiles para transformar organismos Traustoquitriales y la identificación de secuencias de promotor específicas de Traustoquitriales no se encontraban disponibles antes de la presente invención.

Se propuso *Schizochytrium* como huésped recombinante en la patente US nº 6.194.167 y en el documento nº WO 98/46763 se describe la identificación de desaturasas en *Schizochytrium*.

Investigadores anteriores han descrito métodos de transformación y reactivos para la utilización en diversos microorganismos, incluyendo en microalgas que no pertenecen al orden Thraustochytriales. La patente US nº 6.027.900 de Allnutt et al. da a conocer fusiones genéticas para la utilización en ingeniería genética de algas eucarióticas, y particularmente, *Phaeodactylum tricornutum*, utilizando un promotor para un gen fotosintético algal recolector de luz y el gen *Sh ble* de *Streptoalloteichus hindustanus* como marcador seleccionable. Las células se cultivan a altas concentraciones de sal (p.ej., 10 a 35 g/l) y Zeocin™ para la selección de transformantes. Las células microalgales adecuadas para la transformación utilizando dicho método son microalgas fotosintéticas que pueden cultivarse bajo condiciones de alta salinidad. La patente US nº 5.661.017 de Dunahay et al. da a conocer un método para transformar algas que contienen clorofila C (p.ej., diatomeas) utilizando un constructo recombinante que comprende un marcador seleccionable operablemente ligado a una secuencia de control regulador adecuada para la expresión del marcador en las algas que contienen clorofila C. El marcador seleccionable se da a conocer como cualquier marcador adecuado, incluyendo marcadores aislados a partir de fuentes bacterianas y fúngicas, y preferentemente es la neomicina fosfotransferasa. La secuencia de control reguladora puede incluir cualquier secuencia reguladora derivada de un alga que contiene clorofila C y, preferentemente, de *Cyclotella cryptica* (p.ej., una secuencia reguladora de acetil-CoA carboxilasa de *C. cryptica*).

Sin embargo, dichos métodos no son fácilmente transferibles a la transformación de microorganismos Traustoquitriales debido a que, antes de la presente invención, la transformación de microorganismos tales como los Traustoquitriales (p.ej., microalgas) distaba de ser rutinaria. Los marcadores y sistemas de transformación que están bien desarrollados para bacterias y levaduras no son necesariamente fáciles de adaptar a otros microorganismos. En efecto, la patente US nº 5.661.017 observa que "se ha tenido poco éxito en el desarrollo de sistemas de transformación para microalgas eucarióticas" (col. 1, líneas 49 a 51), lo que se debe en parte a la dificultad de introducción de ADN foráneo en dichos microorganismos, y en parte debido a la falta de marcadores y vectores adecuados para la utilización en dicha transformación. El sistema descrito en la patente US nº 5.661.017 se desarrolló específicamente para las algas que contienen clorofila C debido a que sus inventores creían que eran susceptibles de transformación genética, en

particular en comparación con otras algas. De manera similar, la patente US nº 6.027.900, que enseña un método de transformación que es específico de las microalgas fotosintéticas, apoya que la mayoría de las algas son refractarias a cualquier tipo de manipulación genética (col. 1, líneas 39 a 47). Los sistemas adaptados para bacterias, levaduras, y células de insecto y animales no se han adaptado con facilidad a las microalgas. Por lo tanto, anteriormente a la presente invención, todavía existía una necesidad en la técnica de sistemas de transformación eficaces y métodos que sean específicos para las microalgas.

Además, aunque el orden Traustoguitriales ahora se agrupa con las algas cromófitas en el grupo Stramenopiles, todavía hay autores que defienden la opinión de que estos microorganismos son bastante diferentes a la mayoría de microalgas y algunos de estos expertos en la materia cree que miembros del orden Traustoguitriales podrían no ser clasificables en absoluto como microalgas. Los microorganismos considerados microalgas han evolucionado por lo menos cuatro veces separadas durante la evolución, conduciendo a que los microorganismos de tipo "microalgal" se sitúen en diferentes reinos (p.ej., las algas rojas, las algas verdes y las algas doradas (Chromophyta) se encuentran todas en reinos separados). Como resultado, los sistemas de transformación que se ha demostrado que resultan útiles en otras microalgas no se espera que resulten útiles para los Traustoguitriales. Por lo tanto, a pesar del valor comercial de los microorganismos Traustoguitriales, la capacidad de aprovechar todo el potencial de dichos microorganismos mediante ingeniería genética no se ha alcanzado todavía. Antes de la presente invención, los presentes inventores no conocían ningún promotor, marcador seleccionable o vector útil para la transformación de los microorganismos Traustoguitriales ni se disponía de ningún conocimiento sobre qué sistemas de selección podían utilizarse con, o adaptarse a, los Traustoguitriales.

En resumen, existe una necesidad en la técnica de desarrollar métodos para transformar microorganismos Traustoguitriales, proporcionando de esta manera un medio para crear cepas de valor comercial incrementado. Además, existe una necesidad en la técnica de desarrollar métodos para la mutación o inactivación de genes específicos mediante recombinación homóloga o no homóloga en microorganismos Traustoguitriales, proporcionando una nueva manera de alterar el metabolismo celular y de identificar las funciones de genes específicos en los Traustoguitriales.

Sumario de la invención

La presente invención proporciona una molécula aislada de ácidos nucleicos que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en:

- a. SEC ID nº 4;
- b. nucleótidos 441 a 894 de SEC ID nº 9;
- c. una secuencia de ácidos nucleicos que es por lo menos aproximadamente 95% idéntica a los nucleótidos 441 a 894 de la SEC ID nº 9 a lo largo de la longitud completa de los nucleótidos 441 a 894 de la SEC ID nº 9, en la que dicha secuencia de ácidos nucleicos presenta una actividad transcripcional por lo menos basal del promotor α -tubulina, y
- d. un polinucleótido aislado que comprende una secuencia de ácidos nucleicos que es totalmente complementaria a dicho polinucleótido de (a), (b) o (c).

La invención proporciona además un vector recombinante para la transformación de microorganismos del Orden Traustoguitriales, que comprende una secuencia de ácidos nucleicos codificante de una proteína de unión a bleomicina operablemente ligada a un promotor α -tubulina de Traustoguitriales que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en los nucleótidos 441 a 894 de la SEC ID nº 9, y una secuencia de ácidos nucleicos que es por lo menos aproximadamente 95% idéntica a los nucleótidos 441 a 894 de la SEC ID nº 9 a lo largo de la longitud completa de los nucleótidos 441 a 894 de la SEC ID nº 9, en la que dicha secuencia de ácidos nucleicos presenta una actividad transcripcional por lo menos basal del promotor α -tubulina.

Breve descripción de los dibujos

- La FIG. 1 ilustra la construcción del plásmido recombinante pTUBZEO-11.
- La FIG. 2 ilustra la construcción del plásmido recombinante pTUBZEO11-2.
- La FIG. 3A ilustra el plásmido recombinante pMON50200.
- La FIG. 3B ilustra el plásmido recombinante pMON50201.
- La FIG. 3C ilustra el plásmido recombinante pMON50202.
- La FIG. 3D ilustra el plásmido recombinante pMON50203.

Descripción detallada de la invención

La presente invención comprende métodos y materiales relacionados para transformar genéticamente microorganismos del orden Traustoguitriales. Específicamente, la invención proporciona una molécula aislada de ácidos nucleicos tal como se define en la reivindicación 1 y un vector recombinante tal como se define en la reivindicación 2. Todas las cepas de microorganismos unicelulares dadas a conocer en la presente memoria para la

utilización como transformante de los constructos recombinantes de la presente invención, que también pueden denominarse generalmente traustozoitridios, son miembros del orden Traustozoitriales. Según la presente invención, las expresiones "traustozoitridio", "microorganismo Traustozoitriales" y "microorganismo del orden Traustozoitriales" pueden utilizarse intercambiabilmente. Los presentes inventores no conocen ningún informe anterior que describa un sistema de transformación para *Schizozoytrium* o cualquier otro microorganismo Traustozoitriales. Los sistemas de transformación descritos en la presente memoria pueden utilizarse para introducir genes foráneos en microorganismos del orden Traustozoitriales, proporcionando de esta manera un medio para crear cepas con un valor comercial incrementado. Además, la presente invención permite la mutación o inactivación de genes específicos mediante recombinación homóloga o no homóloga, proporcionando una nueva manera de alterar el metabolismo celular y de identificar las funciones de genes específicos en los microorganismos Traustozoitriales.

Más específicamente, los presentes inventores han demostrado la transformación genética de un microorganismo Traustozoitriales del género *Schizozoytrium* {Orden, Traustozoitriales; Familia: *Thraustozoytriaceae*; Género: *Schizozoytrium*), mediante la utilización de dos tipos de vectores de transformación. Estos vectores pueden introducirse en las células mediante métodos estándares, seguido de la identificación y aislamiento de células recombinantes basándose en su capacidad de crecer en presencia de compuestos selectivos. Los presentes inventores han demostrado la eficacia de estos vectores mediante la introducción de los mismos mediante bombardeo de partículas, aunque también pueden utilizarse y se conocen en la técnica otros medios para introducir los vectores (p.ej., electroporación) y se pretende que se encuentren comprendidos en la presente invención.

Para un vector de transformación, ejemplificado en la presente memoria mediante el vector recombinante denominado pTUBZEO11-2, se creó un gen quimérico en el que el gen *ble* (que codifica una "proteína de unión a bleomicina") de *Streptoalloteichus hindustanus* se introdujo cadena abajo de un promotor del gen de tubulina de *Schizozoytrium*. Se introdujo un terminador de SV40 cadena abajo del gen *ble* en este constructo. Este vector permite la expresión del gen *ble* en *Schizozoytrium*, proporcionando de esta manera resistencia a Zeocin™ y compuestos relacionados, que resultan tóxicos para las células de tipo salvaje al incluirlas en el medio de crecimiento a niveles apropiados. La fuente del gen *ble* y del terminador de SV40 en este constructo fue un vector disponible comercialmente denominado pSV40/Zeo, que se obtuvo de Invitrogen Corporation (Carlsbad, CA) (manual técnico 180202, VersionB, "ZeoCassette Vectors", Invitrogen Corporation, 1600 Faraday Ave., Carlsbad, CA 92008). Se aisló el promotor del gen de tubulina mediante la reacción en cadena de la polimerasa; uno de los cebadores utilizados para la reacción se basaba en los datos de secuencia obtenidos mediante un proyecto aleatorio de secuenciación de ADNc de *Schizozoytrium*. Se muestra el mapa de pTUBZEO11-2 en la fig. 2 y la secuencia de nucleótidos de pTUBZEO11-2 se representa mediante SEC ID nº 9. La transformación de *Schizozoytrium* con este vector se confirmó mediante la utilización de la reacción en cadena de la polimerasa y análisis de transferencia southern para detectar la presencia de secuencias de vector integradas en el genoma de *Schizozoytrium*.

El gen *ble* ha sido utilizado por investigadores anteriores como marcador seleccionable para la transformación genética de una diversidad de organismos, incluyendo bacterias, microalgas no traustozoitridas, hongos, protozoos, plantas y células animales (ver, por ejemplo, la patente US nº 6.027.900; Lumbreras et al., Plant J. 14:441-447, 1998; Rohe et al., Curr. Genet. 29:587-590, 1996; Messina et al., Gene 165:213-217, 1995; Guerrero et al., Appl. Microbiol. Biotechnol. 36:759-762, 1992; Pérez et al., Plant Mol. Biol. 13:365-373, 1989; Gattino et al., Gene 91:35-41, 1990). El gen *ble* codifica una "proteína de unión a bleomicina" que proporciona resistencia a varios antibióticos, incluyendo bleomicina, fleomicina y Zeocin™ (Drocourt et al., Nucleic Acids Res. 18:4009, 1990). Este gen se encuentra disponible comercialmente de Invitrogen Corporation, que era la fuente del gen que los presentes inventores utilizaron para crear el vector de transformación de *Schizozoytrium*, pTUBZBO11-2. Sin embargo, los presentes inventores se cree que son los primeros en producir un vector de transformación en el que el gen *ble* se encuentra unido a un promotor de traustozoitridio de una manera que permite la expresión del gen en los traustozoitridios.

Se creó un segundo juego de vectores de transformación mediante mutagénesis dirigida a sitio in vitro de un gen de acetolactato sintasa (*als*) que los presentes inventores aislaron a partir de una biblioteca genómica de *Schizozoytrium*. Estas mutaciones cambian la secuencia de aminoácidos del enzima codificado (ALS) de manera que es mucho menos sensible al sulfometurón metilo y otros compuestos sulfonilurea, así como los inhibidores de la clase imidazolinona y los oxibenzoatos de pirimidinilo, a los que son sensibles los microorganismos del orden Traustozoitriales. Los compuestos sulfonilurea, tales como el sulfometurón-metilo (SMM) con frecuencia resultan tóxicos para las células debido a que son capaces de unirse e inactivar el enzima acetolactato sintasa (ALS) de una diversidad de organismos. ALS cataliza la primera etapa en la biosíntesis de los aminoácidos valina, leucina e isoleucina, las imidazolinonas, las triazolopirimidinas y otros compuestos también se ha demostrado que se unen e inactivan ALS de determinados organismos. Anteriormente se han utilizado formas mutantes de genes que codifican la acetolactato sintasa (también conocida como acetohidroxi ácido sintasa) de otros organismos a modo de marcadores seleccionables para la transformación de levadura y plantas (Annu. Rev. Plant Physiol. Plant Mol. Biol. 40:441-470, 1989). Sin embargo, no hay informes anteriores a la presente invención que describan la secuencia o propiedades del gen *als* de *Schizozoytrium* o cualquier otro miembro de los Traustozoitriales, o la utilización de genes *als* de Traustozoitriales mutantes para proporcionar resistencia a compuestos de sulfonilurea, imidazolinona u oxibenzoato de pirimidinilo. De hecho, según el conocimiento de los presentes inventores, no se ha producido ningún informe publicado con respecto a la sensibilidad de los microorganismos Traustozoitriales a estos agentes selectivos, incluyendo el sulfometurón-

metilo y, por lo tanto, no es conocido de antes de la presente invención si resultaría ni siquiera posible utilizar dicho marcador seleccionable en un sistema de transformación de traustozófito. Cabe destacar que los genes con homología sustancial a genes *als* conocidos se observan en diversos organismos, pero no codifican enzimas que son capaces de catalizar la síntesis de acetolactato (Biochimica et Biophysica Acta 1385:401-419, 1998). Por lo tanto, no habría resultado evidente que un homólogo de *als* clonado de hecho codifica ALS. Con el fin de determinar definitivamente si el gen de *Schizochytrium* clonado era un gen *als* verdadero, los presentes inventores demostraron, mediante experimentos de transformación, una correlación positiva entre la resistencia a sulfometurón-metilo y la expresión del gen *als* putativo de *Schizochytrium* mutado.

Los presentes inventores han producido tres vectores de transformación diferentes que contienen genes *als* mutantes: un gen *als* mutante codifica un enzima con una valina en la posición 595 en lugar de un triptófano (plásmido pMON50201 o ALSmut1-7), otro codifica una glutamina en la posición 192 en lugar de una prolina (plásmido pMON50202, o ALSmut2-2), y una tercera forma contiene ambas mutaciones (plásmido pMON50203, o ALSmut3-5). En estos vectores, la expresión de los genes *als* mutantes recombinantes se encuentra bajo el control del promotor y terminador del gen *als* nativo. Los mapas de estos vectores, junto con un vector que contiene el gen *als* de *Schizochytrium* de tipo salvaje (plásmido pMON50200, o AE-5) se muestran en las figs. 3A-3D. La transformación de *Schizochytrium* con estos vectores codificantes de ALS mutantes mediante la utilización de la reacción en cadena de la polimerasa y análisis de transferencia southern para detectar la presencia de secuencias de vector integradas en el genoma de *Schizochytrium*. Tal como se indica en detalle posteriormente, ahora que los presentes inventores han identificado la secuencia completa del gen *als*, también pueden generarse otras mutaciones, especificadas posteriormente.

Se han utilizado sistemas de transformación para introducir genes foráneos en las células de los Traustozófitos mediante cotransformación. Se describen genes foráneos que se sitúan entre diversos promotores de *Schizochytrium* y un terminador apropiado (p.ej., SV40 o una región de terminador génico de *Schizochytrium*). Se describe un gen sintético que codifica un ácido-graso ω -3 desaturasa del nemátodo *Caenorhabditis elegans*, representado en la presente memoria mediante SEC ID n° 29, para incrementar los niveles de ácido docosahexaenoico en la SEC ID n° 30 de *Schizochytrium* que representa la secuencia de aminoácidos de la desaturasa codificada por la SEC ID n° 29. También pueden introducirse casetes de expresión que contienen genes foráneos en células de Traustozófitos mediante inclusión directa dentro del vector de transformación que contiene el marcador seleccionable.

Además, los presentes inventores también han demostrado con los vectores codificantes de ALS mutante que se produce la recombinación homóloga en *Schizochytrium*, indicando la viabilidad de la utilización de medios recombinantes para inactivar o mutar genes de *Schizochytrium* nativos específicos.

Con respecto a las secuencias de promotor de los Traustozófitos descritas en la presente memoria, una búsqueda en una base de datos de secuencias (GenBank) de todas las secuencias de nucleótidos y proteínas informadas de miembros del orden Traustozófitos, indica que, en el momento de la presente invención, no se han publicado secuencias de promotor de *Schizochytrium* o de cualquier otro miembro de los Traustozófitos. El único gen que ha sido informado de cualquier especie de *Schizochytrium* es del ARN ribosómico 5S de *S. aggregatum* (GenBank n° de acceso X06104 y M13616). Se ha informado de las secuencias de ARN ribosómico 5S y 18S de los miembros de los Thrastozófitos, especie *Ulkenia* y géneros *Labyrinthuloides* y *Thraustochytrium*, aunque ello no tiene consecuencias sobre la presente invención. Se ha descrito una región codificante parcial de un gen de "ARN polimerasa de tipo T3/T7 putativa" de *Thraustochytrium aureum* (Nucleic Acids Research 15:648-654, 1996), pero no se ha descrito una secuencia de promotor para este gen.

La presente invención puede utilizarse para introducir cualesquiera genes u otras secuencias de nucleótidos que resultan de interés en un microorganismo del orden Traustozófitos. Entre tales secuencias de nucleótidos se incluyen, aunque sin limitarse a ellas, ácidos nucleicos codificantes de proteínas (p.ej., enzimas) asociados a la síntesis de ácidos grasos (p.ej., los ácidos grasos ácido docosahexaenoico (DHA), ácido docosapentaenoico (DFA), ácido eicosapentaenoico (EPA) y ácido araquidónico (ARA). Entre dichas proteínas se incluyen, aunque sin limitarse a ellas, ácido graso sintetas, ácido graso desaturasas y ácido graso elongasas, así como proteínas asociadas a un complejo de poliquétido sintasa y/o proteínas asociadas a la incorporación de dichos ácidos grasos en fosfolípidos o en moléculas de triacilglicerol. Por ejemplo, la invención se ha utilizado para introducir genes codificantes de diversos ácido-graso ω -3 desaturasas en *Schizochytrium* en un esfuerzo por incrementar el nivel de ácido docosahexaenoico (DHA) en las células mediante desaturación ω -3 del ácido docosapentaenoico (DPA). A título de otro ejemplo, también se ha demostrado la expresión de una isomerasa putativa de ácido graso polienoico del alga roja *Ptilota* en *Schizochytrium*. Los genes codificantes de un complejo de poliquétido sintasa de *Schizochytrium* (es decir, un sistema de poliquétido sintasa) se han depositado en GenBank, n° de acceso AF378329 (ORFA), AF378328 (ORFB) y AF378329 (ORFC).

La presente invención también resulta útil para introducir genes de microorganismos Traustozófitos y otras secuencias de nucleótidos codificantes de proteínas asociadas a la ruta biosintética de isoprenoides. Entre tales proteínas se incluyen, aunque sin limitarse a ellas, la HMG-CoA sintasa y la HMG-CoA reductasa. Entre otras proteínas adecuadas se incluyen proteínas asociadas a la síntesis de moléculas derivadas de subunidades de isoprenoide,

incluyendo, aunque sin limitarse a ellas, diversos compuestos esteroides y diversos compuestos carotenoides. Entre las proteínas asociadas a la síntesis de diversos compuestos carotenoides se incluyen, aunque sin limitarse a ellos, escualeno sintasa, fitoeno sintasa, fitoeno desaturasa y diversas carotenoides ciclasas, hidroxilasa y quetolasas.

5 La presente invención también resulta útil para introducir en los Traustoquitriales una o más secuencias de ácidos nucleicos codificantes de proteínas asociadas a la síntesis de compuestos antioxidantes, incluyendo, aunque sin limitación, vitamina E y ácido lipoico.

Además, la presente invención puede utilizarse para introducir cualesquiera genes u otros vectores de secuencias de nucleótidos en microorganismos Traustoquitriales con el fin de inactivar o deletar genes (es decir, "inactivación génica" o "disrupción génica dirigida"). La inactivación o delección de genes se utiliza típicamente con el fin de potenciar el valor comercial del microorganismo. Por ejemplo, puede resultar deseable eliminar genes que codifican enzimas (o ácidos nucleicos que regulan la expresión de tales genes) de las rutas de síntesis de ácidos grasos saturados y poliinsaturados. En otro aspecto, puede resultar deseable inhibir o inactivar genes codificantes de proteínas que participan en la degradación de otros compuestos valiosos producidos por el microorganismo de los Traustoquitriales o que reducen el valor del compuesto deseado. Por ejemplo, los genes codificantes de lipasas, enzimas de oxidación de los ácidos grasos y proteínas que presentan sabores u olores desagradables pueden ser dianas de inactivación deseables para la presente invención. En todavía otro aspecto, puede resultar deseable inactivar genes codificantes de proteínas que están asociadas a la síntesis de compuestos la síntesis de los cuales compite con otras moléculas de interés. Por ejemplo, entre dichos genes se incluyen, aunque sin limitarse a ellos, genes codificantes de proteínas que participan en la biosíntesis de carbohidratos, genes codificantes de proteínas participantes en la síntesis de diversos productos de rutas de isoprenoides (p.ej., esteroides o compuestos carotenoides específicos) y genes codificantes de proteínas que participan en la síntesis de componentes de la pared celular. A título de ejemplo, se han introducido genes en células de *Schizochytrium* mediante la utilización de la presente invención en un esfuerzo por inactivar genes que son homólogos a los genes de poliquétido sintasa de *Shewanella* con el fin de evaluar su papel en la producción de ácidos grasos altamente insaturados (HUFA). Tal como se ejemplifica en el Ejemplo 6, la presente invención también puede utilizarse para inactivar, deletar o mutar genes nativos que participan en la producción de ácidos grasos, carotenoides, esteroides, vitaminas u otros compuestos con el fin de mejorar la economía o aceptabilidad de productos que no están relacionados con estos compuestos. Se indica que, en algunas realizaciones, tal como se ha comentado anteriormente, puede resultar deseable potenciar la producción de una proteína dada, mientras que en otras realizaciones puede resultar deseable inhibir la producción de la misma proteína. Tales determinaciones se basan en el uso dado y en los objetivos de producción para el microorganismo específico. La presente invención también resulta útil para determinar el procedimiento de recombinación genética en *Schizochytrium*.

35 Se describen posteriormente diversas realizaciones de la presente invención inicialmente con respecto a un gen *als* y/o proteína ALS de Traustoquitriales. En la presente memoria se dan a conocer la identificación, aislamiento y producción de secuencias de ácidos nucleicos codificantes de marcadores seleccionables que resultan adecuados para la utilización en constructos recombinantes para la transformación de microorganismos traustoquitridos. Tales marcadores seleccionables permiten la selección de microorganismos que han sido transformados con éxito con los constructos recombinantes de la presente invención. Un marcador seleccionable útil para la transformación de Traustoquitriales según la presente descripción es una acetolactato sintasa de Traustoquitriales (es decir, ALS). Preferentemente, la acetolactato sintasa ha sido modificada, mutada o de otro modo seleccionada para ser resistente a la inhibición por compuestos sulfonilurea, inhibidores de clase imidazolinona y/o oxibenzoatos de pirimidinilo (es decir, tal ALS es un homólogo de una acetolactato sintasa natural).

45 Una acetolactato sintasa es una proteína que presenta actividad biológica de acetolactato sintasa, incluyendo proteínas de longitud completa, proteínas de fusión o cualquier homólogo de una acetolactato sintasa natural. Un homólogo de una acetolactato sintasa incluye proteínas que difieren de una acetolactato sintasa natural en que por lo menos uno o unos cuantos, aunque sin limitarse a uno o unos cuantos, aminoácidos han sido deletados (p.ej., una versión truncada de la proteína, tal como un péptido o fragmento), insertados, invertidos, sustituidos y/o derivatizados (p.ej., mediante glucosilación, fosforilación, acetilación, miristoilación, prenilación, palmitación, amidación y/o adición de glucosilfosfatidil inositol). Se describen en detalle posteriormente los homólogos preferentes de una acetolactato sintasa natural.

55 Una proteína aislada, tal como una acetolactato sintasa aislada, es una proteína que ha sido extraída de su medio natural (es decir, que ha sido sometida a manipulación humana) y puede incluir proteínas purificadas, proteínas parcialmente purificadas, proteínas producidas recombinantemente y proteínas producidas sintéticamente, por ejemplo. De esta manera, "aislado" no refleja el grado en que se ha purificado la proteína. Preferentemente, se produce recombinantemente una acetolactato sintasa aislada. Una "acetolactato sintasa de Traustoquitriales" se refiere a una acetolactato sintasa (incluyendo un homólogo de una acetolactato sintasa natural) procedente de un microorganismo Traustoquitriales o que ha sido producida de otro modo a partir de los conocimientos de la estructura (p.ej., secuencia) de una acetolactato sintasa natural a partir de un microorganismo Traustoquitriales. En otras palabras, una acetolactato sintasa de Traustoquitriales incluye cualquier acetolactato sintasa que presente la estructura y función de una acetolactato sintasa natural procedente de un microorganismo Traustoquitriales o que es un homólogo biológicamente activo (es decir, que presenta actividad biológica) de una acetolactato sintasa natural procedente de

un microorganismo Traustouitriales tal como se describe en detalle en la presente memoria. De esta manera, la acetolactato sintasa de Traustouitriales puede incluir proteínas purificadas, parcialmente purificadas, recombinantes, mutadas/modificadas y sintéticas.

5 En general, la actividad biológica o la acción biológica de una proteína se refiere a cualquier función o funciones mostradas o realizadas por la proteína que se atribuyen a la forma natural de la proteína según se mide u observa in vivo (es decir, en el medio fisiológico natural de la proteína) o in vitro (es decir, bajo condiciones de laboratorio). Por ejemplo, una actividad biológica de una acetolactato sintasa incluye actividad enzimática de acetolactato sintasa. Las modificaciones de una proteína, tal como en un homólogo o mimético (comentado posteriormente) pueden resultar en
10 proteínas que presentan la misma actividad biológica que la proteína natural, o en proteínas que presentan una actividad biológica reducida o incrementada en comparación con la proteína natural. Las modificaciones que resultan en una reducción de la expresión de la proteína o en una reducción de la actividad de la proteína, puede denominarse inactivación (completa o parcial), regulación negativa o acción reducida de una proteína. De manera similar, las modificaciones que resultan en un incremento de la expresión de proteína o en un incremento de la actividad de la proteína pueden denominarse amplificación, sobreproducción, activación, potenciación, regulación positiva o acción incrementada de una proteína.

Con respecto a la acetolactato sintasa indicada, resulta preferente que las modificaciones presentes en los homólogos de acetolactato sintasa, en comparación con una acetolactato sintasa natural, no modifiquen sustancialmente, o por lo menos no reduzcan sustancialmente, la actividad biológica básica de la sintasa en comparación con la proteína natural. Sin embargo, tales homólogos pueden presentar diferencias en características aparte de la actividad funcional o enzimática de la proteína en comparación con la forma natural, tal como una sensibilidad reducida a la inhibición por determinados compuestos en comparación con la proteína natural. Preferentemente, un homólogo de una acetolactato sintasa natural presenta una sensibilidad reducida (es decir, reducida, disminuida) a compuestos que se unen e inactivan las acetolactato sintasas naturales en comparación con la acetolactato sintasa natural a partir de la que se deriva el homólogo. Por ejemplo, los compuestos sulfonilurea, tales como el sulfometurón-metilo (SMM), con frecuencia resultan tóxicos para las células debido a que son capaces de unirse e inactivar la acetolactato sintasa (ALS). Las imidazolinonas, triazolopirimidinas, y otros compuestos similares (denominados en general en la presente memoria, inhibidores de clase imidazolinona) también se ha demostrado que se unen e inactivan ALS. Por lo tanto,
30 un homólogo de una acetolactato sintasa natural preferentemente presenta una sensibilidad reducida a los compuestos sulfonilurea, así como a los inhibidores de clase imidazolinona (p.ej., al presentar sitios de unión alterados para dichos inhibidores o sitios de unión con afinidad reducida para el inhibidor) y a oxibenzoatos de pirimidinilo, manteniendo simultáneamente la actividad enzimática de acetolactato sintasa.

35 Tal como se utiliza en la presente memoria, una proteína que presenta "actividad biológica de acetolactato sintasa" o a la que se hace referencia como "acetolactato sintasa", se refiere a una proteína que cataliza la primera etapa en la biosíntesis de los aminoácidos valina, leucina e isoleucina. Más específicamente, una acetolactato sintasa aislada de la presente invención, incluyendo proteínas de longitud completa, proteínas truncadas, proteínas de fusión y homólogos, puede identificarse de una manera sencilla a partir de la capacidad de la proteína de catalizar la síntesis de acetolactato a partir de piruvato. El experto en la materia podrá evaluar a actividad biológica de acetolactato sintasa mediante cualquier ensayo in vitro o in vivo de actividad enzimática.

Se describe una acetolactato sintasa que presenta una secuencia de aminoácidos que es por lo menos aproximadamente 65% idéntica a una secuencia de aminoácidos seleccionada del grupo de SEC ID nº 15, SEC ID nº 19, SEC ID nº 22, SEC ID nº 24 a lo largo de por lo menos aproximadamente 600 aminoácidos de cualquiera de dichas secuencias, en donde la proteína es una acetolactato sintasa (es decir, presenta actividad biológica de acetolactato sintasa). Más preferentemente, una acetolactato sintasa tal como se describe presenta una secuencia de aminoácidos que es por lo menos aproximadamente 70% idéntica, y más preferentemente por lo menos aproximadamente 75% idéntica, y todavía más preferentemente por lo menos aproximadamente 80% idéntica y todavía más preferentemente por lo menos aproximadamente 90% idéntica y todavía más preferentemente por lo menos aproximadamente 95% idéntica, y todavía más preferentemente por lo menos aproximadamente 96% idéntica, y todavía más preferentemente por lo menos aproximadamente 97% idéntica, y todavía más preferentemente por lo menos aproximadamente 98% idéntica, y todavía más preferentemente por lo menos aproximadamente 99% idéntica a cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24 a lo largo de por lo menos aproximadamente 600 aminoácidos de cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, en las que la proteína presenta actividad biológica de acetolactato sintasa.

Se describe además una acetolactato sintasa que presenta una secuencia de aminoácidos que es por lo menos aproximadamente 75% idéntica a una secuencia de aminoácidos seleccionada del grupo de SEC ID nº 15, SEC ID nº 19, SEC ID nº 22, SEC ID nº 24, a lo largo de por lo menos 95 aminoácidos de cualquiera de dichas secuencias, en las que la proteína es una acetolactato sintasa (es decir, presenta actividad biológica de acetolactato sintasa). Más preferentemente, una acetolactato sintasa tal como se describe presenta una secuencia de aminoácidos que es por lo menos aproximadamente 80% idéntica, y todavía más preferentemente por lo menos 85% idéntica, y todavía más preferentemente por lo menos 90% idéntica, y todavía más preferentemente por lo menos 95% idéntica, y todavía más preferentemente por lo menos 96% idéntica, y todavía más preferentemente por lo menos 97% idéntica, y todavía más

preferentemente por lo menos 98% idéntica, y todavía más preferentemente por lo menos 99% idéntica, a cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24 a lo largo de por lo menos 95 aminoácidos de cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, en las que la proteína presenta actividad biológica de acetolactato sintasa. Todavía más preferentemente, una acetolactato sintasa tal como se describe presenta una secuencia de aminoácidos que presenta cualquiera de los porcentajes de identidad anteriormente indicados respecto a cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24 a lo largo de por lo menos 100 aminoácidos, y más preferentemente 125, y más preferentemente 150, y más preferentemente 175, y más preferentemente 200, y más preferentemente 225, y más preferentemente 250, y más preferentemente 275, y más preferentemente 300, y más preferentemente 325, y más preferentemente 350, y más preferentemente 375, y más preferentemente 400, y más preferentemente 425, y más preferentemente 450, y más preferentemente 475, y más preferentemente 500, y más preferentemente 525, y más preferentemente 550, y más preferentemente 575, de cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, en las que la proteína presenta actividad biológica de acetolactato sintasa.

Tal como se utiliza en la presente memoria, a menos que se indique lo contrario, la referencia a un porcentaje (%) de identidad se refiere a una evaluación de homología que se lleva a cabo utilizando: (1) una búsqueda de homología de BLAST con BLAST 2.0 Basic utilizando blastp para las búsquedas de aminoácidos y blastn para las búsquedas de ácidos nucleicos con parámetros por defecto estándares, en las que la secuencia problema se filtra para regiones de baja complejidad por defecto (descrita en Altschul, S.F., Madden, T.L., Schääffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:33 89-3402, 1997); (2) una alineación de BLAST 2 (utilizando los parámetros indicados posteriormente), (3) y/o PSI-BLAST con los parámetros por defecto estándares (Position-Specific Iterated BLAST) Se indica que debido a algunas diferencias en los parámetros estándares entre BLAST 2.0 Basic BLAST y BLAST 2, dos secuencias específicas podrían reconocerse como con homología significativa utilizando el programa BLAST2, mientras que una búsqueda realizada en BLAST 2.0 Basic BLAST utilizando una de las secuencias como secuencia problema podría no identificar la segunda secuencia dentro de las mejores coincidencias. Además, PSI-BLAST proporciona una versión fácil de utilizar automatizada de una búsqueda "de perfil", que es un modo sensible de buscar homólogos de secuencia. El programa en primer lugar lleva a cabo una búsqueda en la base de datos BLAST con huecos. El programa PSI-BLAST utiliza la información de cualesquiera alineaciones significativas que devuelve para construir una matriz de puntuación específica de posición que sustituye la secuencia problema para la siguiente ronda de búsquedas en la base de datos. Por lo tanto, debe entenderse que el porcentaje de identidad puede determinarse mediante la utilización de cualquiera de estos programas.

Pueden alinearse dos secuencias específicas entre sí utilizando la secuencia de BLAST2 tal como se indica en Tatusova y Madden, "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", *FEMS Microbiol Lett.* 174: 247-250, 1999). La alineación de secuencias de BLAST2 se lleva a cabo en blastp o blastn utilizando el algoritmo de BLAST 2.0 para llevar a cabo 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. En aras de la claridad en la presente memoria, se ha llevado a cabo una alineación de secuencias de BLAST2 utilizando los parámetros por defecto estándares siguientes:

Para blastn, utilizando 0 matriz BLOSUM62:

recompensa por coincidencia=1
penalización por no coincidencia=-2
penalizaciones por hueco abierto (5) y por extensión de hueco (2)
hueco x_dropoff (50) expect (10) tamaño de palabra (11) filtro (on)

Para blastp, utilizando 0 matriz BLOSUM62:

penalizaciones por hueco abierto (11) y por extensión de hueco (1)
hueco x_dropoff (50) expect (10) tamaño de palabra (3) filtro (on).

Una acetolactato sintasa tal como se describe también puede incluir proteínas que presentan una secuencia de aminoácidos que comprende por lo menos 30 residuos aminoácidos contiguos de cualquiera de entre SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24 (es decir, 30 residuos aminoácidos contiguos que presentan una identidad de 100% respecto a 30 aminoácidos contiguos de cualquiera de entre SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24). Se describe una acetolactato sintasa que incluye proteínas que presentan secuencias de aminoácidos que comprenden por lo menos 50, y más preferentemente por lo menos 75, y más preferentemente por lo menos 100, y más preferentemente por lo menos 115, y más preferentemente por lo menos 130, y más preferentemente por lo menos 150, y más preferentemente por lo menos 200, y más preferentemente por lo menos 250, y más preferentemente por lo menos 300, y más preferentemente por lo menos 350, y más preferentemente por lo menos 400, y más preferentemente por lo menos 450, y más preferentemente por lo menos 500, y más preferentemente por lo menos 550, y más preferentemente por lo menos 600, y más preferentemente por lo menos 650 residuos

aminoácidos contiguos de cualquiera de entre SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24. Dicha proteína presenta actividad biológica de acetolactato sintasa.

El término "contiguo" o "consecutivo" con respecto a las secuencias de ácidos nucleicos o aminoácidos descritas en la presente memoria se refiere a conectarse en una secuencia no interrumpida. Por ejemplo, que una primera secuencia comprenda 30 aminoácidos contiguos (o consecutivos) de una segunda secuencia, se refiere a que la primera secuencia incluye una secuencia no interrumpida de 30 residuos aminoácidos que es 100% idéntica a una secuencia no interrumpida de 30 residuos aminoácidos en la segunda secuencia. De manera similar, que una primera secuencia presente "identidad de 100%" respecto a una segunda secuencia se refiere a que la primera secuencia se corresponde exactamente con la segunda secuencia sin huecos entre nucleótidos o aminoácidos.

En otro ejemplo descrito, un ejemplo de acetolactato sintasa, que incluye un homólogo de acetolactato sintasa, incluye una proteína que presenta una secuencia de aminoácidos que es suficientemente similar a una secuencia de aminoácidos de acetolactato sintasa natural para que una secuencia de ácidos nucleicos codificante del homólogo sea capaz de hibridarse bajo condiciones de astringencia moderada, elevada o muy elevada (indicadas posteriormente) a (es decir, con) una molécula de ácidos nucleicos codificante de la acetolactato sintasa natural (es decir, al complemento de la cadena de ácidos nucleicos codificante de la secuencia de aminoácidos de la acetolactato sintasa natural). Preferentemente, una acetolactato sintasa descrita está codificada por una secuencia de ácidos nucleicos que se hibrida bajo condiciones de astringencia moderada, elevada o muy elevada con el complemento de una secuencia de ácidos nucleicos que codifica una proteína que comprende una secuencia de aminoácidos representada por SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24. Todavía más preferentemente, una acetolactato sintasa tal como se describe está codificada por una secuencia de ácidos nucleicos que se hibrida bajo condiciones de astringencia moderada, elevada o muy elevada con el complemento de los nucleótidos 1.260 a 3.314 de SEC ID nº 15, los nucleótidos 1.260 a 3.314 de SEC ID nº 18, 1.260 a 3.314 de SEC ID nº 21 o los nucleótidos 1.260 a 3.314 de SEC ID nº 23. Tales condiciones de hibridación se describen en detalle posteriormente. Un complemento de secuencia de ácidos nucleicos de la secuencia de ácidos nucleicos codificante de una acetolactato sintasa tal como se describe se refiere a la secuencia de ácidos nucleicos de la cadena de ácidos nucleicos que es complementaria a la cadena que codifica la acetolactato sintasa. Se apreciará que un ADN de doble cadena que codifica una secuencia de aminoácidos dada comprende un ADN monocatenario y su cadena complementaria que presenta una secuencia que es un complemento del ADN monocatenario. De esta manera, las moléculas de ácidos nucleicos indicadas pueden ser de bicatenarias o monocatenarias, y entre ellas se incluyen las moléculas de ácidos nucleicos que forman híbridos estables bajo condiciones de hibridación restrictivas con una secuencia de ácidos nucleicos que codifica la secuencia de aminoácidos SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, y/o con el complemento de la secuencia de ácidos nucleicos que codifica cualquiera de dichas secuencias de aminoácidos. Los métodos para deducir una secuencia complementaria son conocidos por el experto en la materia. Debe indicarse que debido a que las tecnologías de secuenciación de aminoácidos y de secuenciación de ácidos nucleicos no están totalmente exentas de errores, las secuencias presentadas en la presente memoria en el mejor de los casos representan secuencias aparentes de una acetolactato sintasa.

Los homólogos de la acetolactato sintasa pueden ser el resultado de variación alélica natural o mutación natural. Los homólogos de acetolactato sintasa descritos también pueden producirse utilizando técnicas conocidas de la técnica, incluyendo, aunque sin limitarse a ellas, modificaciones directas de la proteína o modificaciones del gen codificante de la proteína utilizando, por ejemplo, técnicas de ADN clásicas o recombinantes para llevar a cabo mutagénesis aleatoria o dirigida. Una variante alélica natural de un ácido nucleico codificante de una acetolactato sintasa es un gen que se encuentra en esencialmente el mismo locus (o loci) en el genoma que el gen que codifica una secuencia de aminoácidos SEC ID nº 15 pero que, debido a variaciones naturales causadas por, por ejemplo, mutación o recombinación, presenta una secuencia similar, aunque no idéntica. Las variantes alélicas naturales típicamente codifican proteínas que presentan una actividad similar a la de la proteína codificada por el gen con el que se están comparando. Una clase de variantes alélicas puede codificar la misma proteína, pero presentar diferentes secuencias de ácidos nucleicos debido a la degeneración del código genético. Las variantes alélicas pueden comprender además alteraciones en las regiones 5' o 3' no traducidas del gen (p.ej., en regiones de control regulatorio). Las variantes alélicas son bien conocidas por el experto en la materia.

Las proteínas acetolactato sintasa tal como se describen también incluyen productos de expresión de fusiones génicas (por ejemplo, utilizadas para sobreexpresar formas activas solubles de la proteína recombinante), de genes mutagenizados (tales como genes que presentan modificaciones de codones para potenciar la transcripción génica y la traducción) y de genes truncados (tales como genes en los que se han eliminado dominios de unión a membrana para generar formas solubles de una proteína membranal, o genes en los que se han eliminado secuencias de señal que resultan mal tolerados en un huésped recombinante particular).

El tamaño mínimo de una proteína y/o homólogo tal como se describe es un tamaño suficiente para presentar actividad biológica de acetolactato sintasa. Preferentemente, una proteína de la presente invención presenta una longitud de por lo menos 30 aminoácidos, y más preferentemente, por lo menos aproximadamente 50, y más preferentemente por lo menos 75, y más preferentemente por lo menos 100, y más preferentemente por lo menos 115, y más preferentemente por lo menos 130, y más preferentemente por lo menos 150, más preferentemente por lo menos 200,

y más preferentemente por lo menos 250, y más preferentemente por lo menos 300, y más preferentemente por lo menos 350, y más preferentemente por lo menos 400, y más preferentemente por lo menos 450, y más preferentemente por lo menos 500, y más preferentemente por lo menos 550, y más preferentemente por lo menos 600, y más preferentemente por lo menos 650, y más preferentemente por lo menos 684 aminoácidos de longitud. No existe un límite, aparte del límite práctico, al tamaño máximo de dicha proteína en el aspecto de que la proteína puede incluir una parte de una proteína acetolactato sintasa o una acetolactato sintasa de longitud completa, más secuencias adicionales (p.ej., una secuencia de proteína de fusión), si se desea.

También se describe una proteína de fusión que incluye un dominio que contiene acetolactato sintasa (es decir, una secuencia de aminoácidos de una acetolactato sintasa según la presente invención) unida a uno o más segmentos de fusión. Entre los segmentos de fusión descritos se incluyen, aunque sin limitarse a ellos, segmentos que pueden: potenciar la estabilidad de la proteína; proporcionar otra actividad biológica deseable y/o ayudar en la purificación de una acetolactato sintasa (p.ej., mediante cromatografía de afinidad). Un segmento de fusión adecuado puede ser un dominio de cualquier tamaño que presenta la función deseada (p.ej., proporciona una estabilidad, solubilidad, acción o actividad biológica incrementadas, y/o simplifica la purificación de la proteína).

Pueden unirse segmentos de fusión a los extremos amino y/o carboxilo del dominio que contiene acetolactato sintasa de la proteína y puede ser susceptible de corte con el fin de permitir la recuperación sencilla de una acetolactato sintasa. Las proteínas de fusión preferentemente se producen mediante el cultivo de una célula recombinante transfectada con una molécula de ácidos nucleicos de fusión que codifica una proteína que incluye el segmento de fusión unido al extremo carboxilo-terminal y/o amino-terminal de un dominio que contiene acetolactato sintasa.

Se describe además un mimético de una acetolactato sintasa. Tal como se utiliza en la presente memoria, el término "mimético" se utiliza para referirse a cualquier compuesto péptido o no péptido que es capaz de mimetizar la acción biológica de un péptido natural, con frecuencia debido a que el mimético presenta una estructura básica que mimetiza la estructura básica del péptido natural y/o que presenta propiedades biológicas destacadas del péptido natural. Entre los miméticos pueden incluirse, aunque sin limitarse a ellos, péptidos que presentan modificaciones sustanciales respecto del prototipo, tales como la falta de similitud respecto a cadenas laterales con el péptido natural (tales modificaciones, por ejemplo, pueden reducir su susceptibilidad a la degradación), anticuerpos antiidiotípicos y/o catalíticos, o fragmentos de los mismos; partes no proteicas de una proteína aislada (p.ej., estructuras de carbohidrato), o moléculas orgánicas sintéticas o naturales, incluyendo ácidos nucleicos y fármacos identificados mediante química combinatorial, por ejemplo.

Dichos miméticos pueden diseñarse, seleccionarse y/o de otro modo identificarse utilizando una diversidad de métodos conocidos de la técnica. Se dan a conocer diversos métodos de diseño de fármacos, útiles para diseñar miméticos u otros compuestos terapéuticos, en Maulik et al., *Molecular Biotechnology: Therapeutic Applications and Strategies*, Wiley-Liss, Inc., 1997. Puede obtenerse un mimético de acetolactato sintasa por ejemplo a partir de estrategias de diversidad molecular (una combinación de estrategias relacionadas que permiten la construcción rápida de bibliotecas grandes de moléculas químicamente diversas), bibliotecas de compuestos naturales o sintéticos, en particular de bibliotecas químicas o combinatoriales (es decir, bibliotecas de compuestos que difieren en secuencia o tamaño pero que presentan bloques constructivos similares) o mediante diseño de fármacos racional dirigido o aleatorio. Ver, por ejemplo, Maulik et al., supra.

En una estrategia de diversidad molecular, se sintetizan bibliotecas grandes de compuestos, por ejemplo, a partir de péptidos, oligonucleótidos, carbohidratos y/o moléculas orgánicas sintéticas, utilizando enfoques biológicos, enzimáticos y/o químicos. Entre los parámetros críticos para desarrollar una estrategia de diversidad molecular se incluyen la diversidad de subunidades, el tamaño molecular y la diversidad de la biblioteca. El objetivo general del cribado de tales bibliotecas es utilizar la aplicación secuencial de la selección combinatorial para obtener ligandos de afinidad elevada para una diana deseada y después optimizar las moléculas de cabeza mediante estrategias de diseño aleatorio o dirigido. Los métodos de diversidad molecular se describen en detalle en Maulik et al., *ibid.*

Maulik et al. Dan a conocer además, por ejemplo, métodos de diseño dirigido en los que el usuario dirige el procedimiento de creación de nuevas moléculas a partir de una biblioteca de fragmentos, de fragmentos apropiadamente seleccionados; el diseño aleatorio, en el que el usuario utiliza un algoritmo genético o de otro tipo para mutar aleatoriamente fragmentos y sus combinaciones, aplicando simultáneamente un criterio de selección para evaluar el ajuste de los ligandos candidatos, y un enfoque basado en una matriz en el que el usuario calcular la energía de interacción entre estructuras receptoras tridimensionales y fragmentos pequeños sonda, seguido de la unión entre sí de los sitios de sonda favorables.

Se describen acetolactato sintasas que pueden derivarse de cualquier microorganismo Traustoquitriales y, en particular, de cualquier organismo *Schizochytrium*. La acetolactato sintasa descrita de la presente invención presenta una secuencia de aminoácidos seleccionada del grupo de SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 y SEC ID nº 24. La proteína que presenta una secuencia de aminoácidos representada por SEC ID nº 15 es una acetolactato sintasa natural (es decir, de tipo salvaje) procedente de un microorganismo Traustoquitriales y, específicamente, es una acetolactato sintasa de *Schizochytrium*. Las secuencias de aminoácidos representadas por SEC ID nº 19, SEC ID nº

- 22 y SEC ID nº 24 son secuencias que han sido modificadas, de manera que los enzimas resultantes presentan una sensibilidad reducida a los compuestos sulfonilurea, así como a inhibidores de clase imidazolinona y oxibenzoatos de pirimidinilo, en comparación con la proteína natural representada por la secuencia de aminoácidos SEC ID nº 15. Se indica que las proteínas representadas por SEC ID nº 19, SEC ID nº 22 y SEC ID nº 24 presentan actividad biológica de acetolactato sintasa. Las acetolactato sintasas con sensibilidad reducida a los compuestos sulfonilurea, así como a inhibidores de clase imidazolinona y oxibenzoatos de pirimidinilo son acetolactato sintasas descritas, debido a que las secuencias de ácidos nucleicos codificantes de tales sintasas pueden utilizarse en vectores recombinantes de la presente invención a modo de marcadores seleccionables.
- Por lo tanto, un ejemplo descrito se refiere a una acetolactato sintasa modificada, incluyendo cualquier homólogo de cualquiera de entre SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 y SEC ID nº 24, en el que el homólogo presenta actividad biológica de acetolactato sintasa, y en particular, en el que el homólogo presenta una sensibilidad reducida a compuestos sulfonilurea, así como a inhibidores de clase imidazolinona y oxibenzoatos de pirimidinilo, en comparación con la proteína natural representada por la secuencia de aminoácidos SEC ID nº 15. En un ejemplo descrito, entre dichos homólogos de acetolactato sintasa se incluyen proteínas que presentan una secuencia de aminoácidos que difieren de la SEC ID nº 15 por una delección, inserción o sustitución de aminoácido en una o más de las posiciones siguientes: 116G, 117A, 192P, 200A, 251K, 358M, 383D, 592V, 595W, o 599F. Estas posiciones corresponden a sitios de mutación de ALS conocidos en una acetolactato sintasa de levadura (es decir, 116G, 117A, 192P, 200A, 251K, 354M, 379D, 583V, 586W y 590F, respectivamente) (ver Mazur y Falco, Annu. Rev. Plant Physiol. Plant Mol. Biol. 40:441-470, 1989). Otros sitios de mutación posibles son conocidos por el experto en la materia basados en mutaciones de aminoácido con éxito en ALS de otros organismos. La aplicación de dichos sitios a los sitios correspondientes en ALS de Traustoguitriales se encuentra comprendida en la presente invención.
- Se describe una molécula aislada de ácidos nucleicos que comprende una secuencia de ácidos nucleicos que codifica una acetolactato sintasa de Traustoguitriales y una secuencia de ácidos nucleicos totalmente complementaria a la misma. Una molécula de ácidos nucleicos descrita codificante de una acetolactato sintasa incluye una molécula de ácidos nucleicos codificante de cualquiera de las proteínas acetolactato sintasa, incluyendo homólogos, comentados anteriormente. Más particularmente un ejemplo descrito se refiere a una molécula aislada de ácidos nucleicos que comprende una secuencia de ácidos nucleicos que codifica una proteína que presenta una secuencias de aminoácidos que es por lo menos aproximadamente 65% idéntica a una secuencia de aminoácidos seleccionada del grupo de entre SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 y SEC ID nº 24, a lo largo de por lo menos aproximadamente 600 aminoácidos de cualquiera de tales secuencias, en el que la proteína es una acetolactato sintasa (es decir, como actividad biológica de acetolactato sintasas). Más preferentemente, una molécula aislada de ácidos nucleico tal como se describe presenta una secuencia de ácidos nucleicos codificante de una secuencia de aminoácidos que es por lo menos aproximadamente 70% idéntica, y más preferentemente por lo menos aproximadamente 75% idéntica, y todavía más preferentemente por lo menos aproximadamente 80% idéntica, y todavía más preferentemente por lo menos aproximadamente 85% idéntica y todavía más preferentemente por lo menos aproximadamente 90% idéntica y todavía más preferentemente por lo menos aproximadamente 95% idéntica, y todavía más preferentemente por lo menos aproximadamente 96% idéntica, y todavía más preferentemente por lo menos aproximadamente 97% idéntica, y todavía más preferentemente por lo menos aproximadamente 98% idéntica, y todavía más preferentemente por lo menos aproximadamente 99% idéntica a cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24 a lo largo de por lo menos aproximadamente 600 aminoácidos de cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, en las que la proteína presenta actividad biológica de acetolactato sintasa.
- En otro ejemplo descrito, una molécula aislada de ácidos nucleicos presenta una secuencia de ácidos nucleicos codificante de una secuencia de aminoácidos que es por lo menos 75% idéntica a una secuencia de aminoácidos seleccionada del grupo de SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 y SEC ID nº 24 a lo largo de por lo menos 95 aminoácidos de cualquiera de dichas secuencias, en donde la proteína es una acetolactato sintasa (es decir, presenta actividad biológica de acetolactato sintasa). Más preferentemente, una molécula aislada de ácidos nucleicos de la presente invención presenta una secuencia de ácidos nucleicos codificante de una secuencia de aminoácidos que es por lo menos aproximadamente 80% idéntica, y todavía más preferentemente por lo menos aproximadamente 85% idéntica, y todavía más preferentemente por lo menos aproximadamente 90% idéntica y todavía más preferentemente por lo menos 95% idéntica, y todavía más preferentemente por lo menos 96% idéntica, y todavía más preferentemente por lo menos 97% idéntica, y todavía más preferentemente por lo menos 98% idéntica, y todavía más preferentemente por lo menos 99% idéntica a cualquiera de entre SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, a lo largo de por lo menos 95 aminoácidos de cualquiera de SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, en donde la proteína presenta actividad biológica de acetolactato sintasa.
- En todavía otro ejemplo descrito, una molécula aislada de ácidos nucleicos presenta una secuencia de ácidos nucleicos codificante de una secuencia de aminoácidos que presenta cualquiera de los porcentajes de identidad anteriormente indicados respecto a cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24 a lo largo de por lo menos 100 aminoácidos, y más preferentemente 125, y más preferentemente 150, y más preferentemente 175, y más preferentemente 200, y más preferentemente 225, y más preferentemente 250, y más preferentemente 275, y más preferentemente 300, y más preferentemente 325, y más preferentemente 350, y más

preferentemente 375, y más preferentemente 400, y más preferentemente 425, y más preferentemente 450, y más preferentemente 475, y más preferentemente 500, y más preferentemente 525, y más preferentemente 550, y más preferentemente 575 de cualquiera de las secuencias SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24, en donde la proteína presenta actividad biológica de acetolactato sintasa. Se determinó el porcentaje de identidad utilizando los parámetros por defecto de BLAST 2.0 Basic BLAST, tal como se ha indicado anteriormente.

En un ejemplo descrito, entre las moléculas de ácidos nucleicos codificantes de una acetolactato sintasa se incluyen moléculas aisladas de ácidos nucleicos que se hibridan bajo condiciones de astringencia moderada, y todavía más preferentemente bajo condiciones de astringencia elevada, y todavía más preferentemente bajo condiciones de astringencia muy elevada con el complemento de una secuencia de ácidos nucleicos codificante de una acetolactato sintasa natural. Preferentemente, una molécula aislada de ácidos nucleicos codificante de una acetolactato sintasa tal como se describe comprende una secuencia de ácidos nucleicos que se hibrida bajo condiciones de astringencia moderada o elevada con el complemento de una secuencia de ácidos nucleicos que codifica una proteína que comprende una secuencia de aminoácidos representada por SEC ID nº 15, SEC ID nº 19, SEC ID nº 22 o SEC ID nº 24. En un ejemplo descrito, una molécula aislada de ácidos nucleicos comprende una secuencia de ácidos nucleicos que se hibrida bajo condiciones de astringencia moderada, elevada o muy elevada con el complemento de una secuencia de ácidos nucleicos representada por los nucleótidos 1.260 a 3.314 de SEC ID nº 14, los nucleótidos 1.260 a 3.314 de SEC ID nº 18, los nucleótidos 1.260 a 3.314 de SEC ID nº 21 o los nucleótidos 1.260 a 3.314 de SEC ID nº 23.

Tal como se utiliza en la presente memoria, las condiciones de hibridación se refieren a condiciones de hibridación estándares bajo las que se utilizan las moléculas de ácidos nucleicos para identificar moléculas de ácidos nucleicos similares. Tales condiciones estándares se dan a conocer en, por ejemplo, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989. Sambrook et al., *ibid.* (ver específicamente las páginas 9.31 a 9.62). Además, las fórmulas para calcular las condiciones apropiadas de hibridación y lavado para conseguir la hibridación permitiendo diversos grados de no correspondencia de los nucleótidos se dan a conocer en, por ejemplo, Meinkoth et al., *Anal. Biochem.* 138:267-284, 1984; Meinkoth et al., *ibid.*

Más particularmente, las condiciones de hibridación y lavado de astringencia moderada, tal como se denominan en la presente memoria, se refieren a condiciones que permiten el aislamiento de moléculas de ácidos nucleicos que presentan una identidad de secuencia de ácidos nucleicos de por lo menos aproximadamente 70% respecto a la molécula de ácidos nucleicos que se utiliza para sondear la reacción de hibridación (es decir, condiciones que permiten una no correspondencia de aproximadamente 30% o menos de los nucleótidos). Las condiciones de hibridación y lavado de astringencia elevada, tal como se denominan en la presente memoria, se refieren a condiciones que permiten el aislamiento de moléculas de ácidos nucleicos que presentan una identidad de secuencia de ácidos nucleicos de por lo menos aproximadamente 80% respecto a la molécula de ácidos nucleicos que se utiliza para sondear la reacción de hibridación (es decir, condiciones que permiten una no correspondencia de aproximadamente 20% o menos de los nucleótidos). Las condiciones de hibridación y lavado de astringencia muy elevada, tal como se denominan en la presente memoria, se refieren a condiciones que permiten el aislamiento de moléculas de ácidos nucleicos que presentan una identidad de secuencia de ácidos nucleicos de por lo menos aproximadamente 90% respecto a la molécula de ácidos nucleicos que se utiliza para sondear la reacción de hibridación (es decir, condiciones que permiten una no correspondencia de aproximadamente 10% o menos de los nucleótidos). Tal como se ha comentado anteriormente, el experto en la materia puede utilizar las fórmulas en Meinkoth et al., *ibid.* Para calcular las condiciones de hibridación y lavado apropiadas para conseguir estos niveles particulares de no correspondencia de nucleótidos. Tales condiciones variarán, dependiendo de si se forman híbridos de ADN:ARN o de ADN:ADN. Las temperaturas de fusión calculadas para los híbridos de ADN:ADN son 10°C inferiores a las de los híbridos de ADN:ARN. En ejemplos particulares, las condiciones de hibridación restrictivas para los híbridos de ADN:ADN incluyen la hibridación a una fuerza iónica de 6X SSC (Na⁺ 0,9 M) a una temperatura de entre aproximadamente 20°C y aproximadamente 35°C (astringencia más baja), más preferentemente de entre aproximadamente 28°C y aproximadamente 40°C (mayor astringencia) y todavía más preferentemente, de entre aproximadamente 35°C y aproximadamente 45°C (todavía más astringencia), con condiciones de lavado apropiadas. En ejemplos particulares, entre las condiciones de hibridación restrictivas para los híbridos de ADN:ARN se incluyen la 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 preferentemente de entre 38°C y aproximadamente 50°C, y todavía más preferentemente, de entre aproximadamente 45°C y aproximadamente 55°C, con condiciones de lavado de astringencia similar. Estos valores se basan en cálculos de una temperatura de fusión para moléculas de más de aproximadamente 100 nucleótidos, 0% de formamida y un contenido de G+C de aproximadamente 40%. Alternativamente, la T_m puede calcularse empíricamente tal como se indica en Sambrook et al., *supra*, páginas 9.31 a 9.62. En general, las condiciones de lavado deberían ser tan astringentes como resulte posible y deben resultar apropiadas para las condiciones de hibridación seleccionadas. Por ejemplo, las condiciones de hibridación pueden incluir una combinación de condiciones salinas y de temperatura que sean aproximadamente 20-25°C inferiores a la T_m calculada de un híbrido particular, y las condiciones de lavado típicamente incluyen una combinación de condiciones salinas y de temperatura que son aproximadamente 12-20°C inferiores a la T_m calculada del híbrido particular. Un ejemplo de condiciones de hibridación adecuadas para la utilización con híbridos de ADN:ADN incluye una hibridación de 2 a 24 horas en 6X SSC (formamida al 50%) a aproximadamente 42°C, seguido de etapas de lavado que incluyen uno o más lavados a temperatura ambiente en

aproximadamente 2X SSC, seguido de lavados adicionales a temperaturas más altas y una fuerza iónica más baja (p.ej., por lo menos un lavado a aproximadamente 37°C en aproximadamente 0,1X-0,5X SSC, seguido de por lo menos un lavado a aproximadamente 68°C en aproximadamente 0,1X-0,5X SSC).

5 En otro ejemplo descrito, las moléculas de ácidos nucleicos codificantes de una acetolactato sintasa tal como se ha descrito incluyen moléculas de ácidos nucleicos que comprenden una secuencia de ácidos nucleicos codificante de una proteína que presenta una secuencia de aminoácidos que comprende por lo menos 30 residuos aminoácidos contiguos de cualquiera de SEC ID n° 15, SEC ID n° 19, SEC ID n° 22 o SEC ID n° 24 (es decir, 30 residuos aminoácidos contiguos que presentan una identidad de 100% respecto a 30 aminoácidos contiguos de cualquiera de SEC ID n° 15, SEC ID n° 19, SEC ID n° 22 o SEC ID n° 24). En un ejemplo descrito, una molécula aislada de ácidos nucleicos comprende una secuencia de ácidos nucleicos codificante de una proteína que presenta una secuencia de aminoácidos que comprende por lo menos 50, y más preferentemente por lo menos 75, y más preferentemente por lo menos 100, y más preferentemente por lo menos 150, y más preferentemente por lo menos 200, y más preferentemente por lo menos 250, y más preferentemente por lo menos 300, y más preferentemente por lo menos 350, y más preferentemente por lo menos 400, y más preferentemente por lo menos 450, y más preferentemente por lo menos 500, y más preferentemente por lo menos 550, y más preferentemente por lo menos 600, y más preferentemente por lo menos 650 residuos aminoácidos contiguos de cualquiera de SEC ID n° 15, SEC ID n° 19, SEC ID n° 22 o SEC ID n° 24. Dicha proteína presenta actividad biológica de acetolactato sintasa. En un ejemplo descrito, una molécula aislada de ácidos nucleicos codificante de una acetolactato sintasa comprende una secuencia de ácidos nucleicos que presenta por lo menos 60 nucleótidos contiguos, y más preferentemente por lo menos 150, y más preferentemente por lo menos 225, y más preferentemente por lo menos 300, y más preferentemente por lo menos 345, y más preferentemente por lo menos 390, y más preferentemente por lo menos 450, y más preferentemente por lo menos 525, y más preferentemente por lo menos 600, y más preferentemente por lo menos 750, y más preferentemente por lo menos 900, y más preferentemente por lo menos 1.050, y más preferentemente por lo menos 1.200, y más preferentemente por lo menos 1.350, y más preferentemente por lo menos 1.500, y más preferentemente por lo menos 1.650, y más preferentemente por lo menos 1.800, y todavía más preferentemente por lo menos 1.950 nucleótidos contiguos de los nucleótidos 1.260-3.314 de la SEC ID n° 15, de los nucleótidos 1.260-3.314 de la SEC ID n° 18, de los nucleótidos 1.260-3.314 de la SEC ID n° 21 o de los nucleótidos 1.260-3.314 de la SEC ID n° 23.

10 Entre las moléculas de ácidos nucleicos descritas se incluyen los nucleótidos 1.260-3.314 de la SEC ID n° 14 (codifica la SEC ID n° 15), los nucleótidos 1.260-3.314 de la SEC ID n° 18 (codifica la SEC ID n° 19), los nucleótidos 1.260-3.314 de la SEC ID n° 21 (codifica la SEC ID n° 22) o los nucleótidos 1.260-3.314 de la SEC ID n° 23 (codifica la SEC ID n° 24), SEC ID n° 14, SEC ID n° 18, SEC ID n° 21 o SEC ID n° 23. Según la presente invención, una molécula aislada de ácidos nucleicos es una molécula de ácidos nucleicos que ha sido extraída de su medio natural (es decir, que ha sido sometida a manipulación humana), siendo su medio natural el genoma o cromosoma en la que se encuentra naturalmente la molécula de ácidos nucleicos. De esta manera, "aislada" no refleja necesariamente el grado en que se ha purificado la molécula de ácidos nucleicos, sino que indica que la molécula no incluye un genoma entero o un cromosoma entero en el que se encuentra la molécula de ácidos nucleicos en la naturaleza. Una molécula aislada de ácidos nucleicos puede incluir un gen, tal como un gen de acetolactato sintasa descrito en la presente memoria. Una molécula aislada de ácidos nucleicos que incluye un gen no es un fragmento de un cromosoma que incluye dicho gen, sino que por el contrario incluye la región codificante y las regiones reguladoras asociadas con el gen, pero ningún gen adicional naturalmente presente en el mismo cromosoma. Una molécula aislada de ácidos nucleicos puede incluir además una secuencia de ácidos nucleicos especificada, flanqueada (es decir, en el extremo 5' y/o 3' de la secuencia) por ácidos nucleicos adicionales que normalmente no flanquean la secuencia de ácidos nucleicos especificada en la naturaleza (es decir, son secuencias heterólogas). Molécula aislada de ácidos nucleicos puede incluir ADN, ARN (p.ej., ARNm) o derivados de ADN o ARN (p.ej., ADNc). Aunque la expresión "molécula de ácidos nucleicos" principalmente se refiere a la molécula física de ácidos nucleicos y la expresión "secuencia de ácidos nucleicos" se refiere principalmente a la secuencia de nucleótidos en la molécula de ácidos nucleicos, las dos expresiones pueden utilizarse intercambiablemente, especialmente con respecto a que una molécula de ácidos nucleicos, o una secuencia de ácidos nucleicos, es capa de codificar una proteína.

15 Preferentemente, una molécula aislada de ácidos nucleicos de la presente invención se produce utilizando tecnología de ADN recombinante (p.ej., reacción en cadena de la polimerasa (PCR), amplificación, clonación) o síntesis química. Entre las moléculas aisladas de ácidos nucleicos se incluyen moléculas de ácidos nucleicos naturales y homólogos de las mismas, incluyendo, aunque sin limitarse a ellas, variantes alélicas naturales y moléculas de ácidos nucleicos modificados en las que se han insertado, delecionados, sustituido y/o invertido, nucleótidos de tal manera que dichas modificaciones proporcionan el efecto deseado sobre la actividad biológica de la proteína. Las variantes alélicas y homólogos de proteínas (p.ej., proteínas codificadas por homólogos de ácidos nucleicos) han sido comentadas en detalle anteriormente.

20 Una molécula de ácidos nucleicos puede producirse utilizando varios métodos conocidos por el experto en la materia (ver, por ejemplo, Sambrook et al., *ibid*). Por ejemplo, pueden modificarse moléculas de ácidos nucleicos utilizando una diversidad de técnicas, incluyendo, aunque sin limitarse a ellas, técnicas de mutagénesis clásicas y técnicas de ADN recombinante, tales como la mutagénesis dirigida a sitio, el tratamiento químico de una molécula de ácidos

nucleicos para inducir mutaciones, el corte con enzimas de restricción de un fragmento de ácidos nucleicos, la ligación de fragmentos de ácidos nucleicos, la amplificación por PCR y/o la mutagénesis de regiones seleccionadas de una secuencia de ácidos nucleicos, la síntesis de mezclas de oligonucleótidos y la ligación de grupos de mezcla para "construir" una mezcla de moléculas de ácidos nucleicos, y combinaciones de los mismos. Los homólogos de molécula de ácidos nucleicos pueden seleccionarse de una mezcla de ácidos nucleicos modificados mediante cribado para la función de la proteína codificada por el ácido nucleico y/o mediante hibridación con un gen de tipo salvaje.

De manera similar, el tamaño mínimo de una molécula de ácidos nucleicos de la presente invención es un tamaño suficiente para codificar una proteína que presenta la actividad biológica deseada, o suficiente para formar una sonda o cebador oligonucleótido que es capaz de formar un híbrido estable con la secuencia complementaria de una molécula de ácidos nucleicos codificante de la proteína natural (p.ej., bajo condiciones de astringencia moderada, elevada o muy elevada). De esta manera, el tamaño de la molécula de ácidos nucleicos codificante de dicha proteína puede depender de la composición de ácidos nucleicos y el porcentaje de homología o identidad entre la molécula de ácidos nucleicos y la secuencia complementaria, así como de las condiciones de hibridación de por sí (p.ej., la temperatura, la concentración salina y la concentración de formamida). El tamaño mínimo de una molécula de ácidos nucleicos que se utiliza como cebador oligonucleótido o como sonda típicamente es de por lo menos aproximadamente 12 a aproximadamente 15 nucleótidos de longitud en el caso de que las moléculas de ácidos nucleicos sean ricas en GC y de por lo menos aproximadamente 15 a aproximadamente 18 bases de longitud en el caso de que sean ricas en AT. No existe ningún límite, aparte del límite práctico, para el tamaño máximo de una molécula de ácidos nucleicos de la presente invención, en el aspecto de que la molécula de ácidos nucleicos puede incluir una parte de una secuencia codificante de proteína (p.ej., una secuencia codificante de acetolactato sintasa) o una secuencia de ácidos nucleicos codificante de una proteína de longitud completa.

Una realización de la presente invención incluye un vector recombinante para la utilización para la transformación de un microorganismo Traustoquitriales. Según la presente invención, un vector recombinante es una molécula de ácidos nucleicos manipulada (es decir, producida artificialmente) que se utiliza como herramienta para manipular una secuencia de ácidos nucleicos seleccionada y para introducir dicha secuencia de ácidos nucleicos en una célula huésped. Por lo tanto, el vector recombinante resulta adecuado para la utilización en la clonación, secuenciación y/o, de otro modo, manipulación de la secuencia de ácidos nucleicos seleccionada, tal como mediante la expresión y/o introducción de la secuencia de ácidos nucleicos seleccionada en una célula huésped para formar una célula recombinante. Tal vector típicamente contiene secuencias de ácidos nucleicos heterólogos, es decir, secuencias de ácidos nucleicos que naturalmente no se observa que sean contiguas a la secuencia de ácidos nucleicos que debe introducirse, aunque el vector puede contener también secuencias de ácidos nucleicos reguladoras (p.ej., promotores, regiones no traducidas) que son naturalmente contiguas a las moléculas de ácidos nucleicos de la presente invención (comentadas en detalle posteriormente). El vector puede ser de ARN o ADN, procariótico o eucariótico, y típicamente es un plásmido. El vector puede mantenerse como elemento extracromosómico (p.ej., un plásmido) o puede integrarse en el cromosoma del microorganismo recombinante. El vector entero puede mantenerse en su sitio dentro de una célula huésped o, bajo determinadas condiciones, puede delecionarse el ADN plasmídico, reteniendo la molécula de ácidos nucleicos de la presente invención. La molécula de ácidos nucleicos integrada puede encontrarse bajo el control de un promotor cromosómico, bajo el control de un promotor nativo o plasmídico, o bajo una combinación de varios controles de promotor. Pueden integrarse en el cromosoma una única copia o múltiples copias de la molécula de ácidos nucleicos. Un vector recombinante tal como se ha descrito contiene por lo menos un marcador seleccionable para microorganismos Traustoquitriales según la presente invención, tal como una secuencia de ácidos nucleicos codificante de una acetolactato sintasa de Traustoquitriales (proteína natural u homólogo) o una secuencia de ácidos nucleicos codificante del gen *ble* (descrito posteriormente). Tal como se utiliza en la presente memoria, la expresión "molécula de ácidos nucleicos recombinante" se utiliza principalmente para referirse a un vector recombinante en el que se ha ligado la secuencia de ácidos nucleicos que debe clonarse, manipularse y transformarse en la célula huésped (es decir, el inserto).

La invención proporciona un vector recombinante para la transformación de microorganismos del Orden Traustoquitriales, que comprende una secuencia de ácidos nucleicos codificante de una proteína de unión a bleomicina operablemente ligada a un promotor α -tubulina de Traustoquitriales que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en los nucleótidos 441 a 894 de la SEC ID n° 9, y una secuencia de ácidos nucleicos que es por lo menos aproximadamente 95% idéntica a los nucleótidos 441 a 894 de la SEC ID n° 9 a lo largo de la longitud completa de los nucleótidos 441 a 894 de la SEC ID n° 9, en la que dicha secuencia de ácidos nucleicos presenta una actividad transcripcional por lo menos basal del promotor α -tubulina, o una molécula de ácidos nucleicos recombinante que comprende la molécula aislada de ácidos nucleicos según la reivindicación 1, operablemente ligada a una secuencia de control transcripcional. Tal como se utiliza en la presente memoria, la expresión "molécula recombinante" o "molécula de ácidos nucleicos recombinante" principalmente se refiere a una molécula de ácidos nucleicos o secuencia de ácidos nucleicos operablemente ligada a una secuencia de control transcripcional, aunque puede utilizarse intercambiamente con la expresión "molécula de ácidos nucleicos" en el caso de que dicha molécula de ácidos nucleicos sea una molécula recombinante tal como se comenta en la presente memoria. Según la presente invención, la expresión "operablemente ligada" se refiere a la unión de una molécula de ácidos nucleicos a una secuencia de control transcripcional de manera tal que la molécula sea capaz de expresarse al transfectarse (es decir, transformarse, transducirse, transfectarse, conjugarse o conducirse) dentro de una célula huésped. Las

secuencias de control transcripcional son secuencias que controlan el inicio, elongación o terminación de la transcripción. Son secuencias de control de la transcripción particularmente importantes aquellas que controlan el inicio de la transcripción, tales como las secuencias de promotor, intensificador, operador y represor. Entre las secuencias de control transcripcional adecuadas se incluye cualquier secuencia de control transcripcional que pueda funcionar en un microorganismo del orden Traustoquitriales. Los presentes inventores se cree que son los primeros en aislar e identificar por lo menos tres de tales promotores, descritos en detalle en otros sitios de la presente memoria.

El promotor puede ser un promotor de acetolactato sintasa de Traustoquitriales (representado en la presente memoria por los nucleótidos 1 a 1.259 de la SEC ID nº 14), un promotor de α -tubulina de Traustoquitriales (representado en la presente memoria por los nucleótidos 441-894 de la SEC ID nº 9). La clonación y secuenciación del promotor de α -tubulina se describe en la sección de Ejemplos. El promotor de α -tubulina comprende la secuencia de promotor de α -tubulina de Traustoquitriales natural (nucleótidos 441-894 de la SEC ID nº 9) o una secuencia de ácidos nucleicos que es por lo menos 95% idéntica a los nucleótidos 441-894 de la SEC ID nº 9, en la que el promotor presenta una actividad transcripcional por lo menos basal del promotor de α -tubulina. Los métodos para determinar el porcentaje de identidad se han descrito anteriormente en la presente memoria para las secuencias de la acetolactato sintasa y se encuentran comprendidas en la presente memoria.

En una realización, un vector recombinante de la presente invención es un vector de expresión. Tal como se utiliza en la presente memoria, la expresión "vector de expresión" se utiliza para referirse a un vector que resulta adecuado para la producción de un producto codificado (p.ej., una proteína de interés). En la presente realización, una secuencia de ácidos nucleicos codificante del producto que debe producirse se inserta en el vector recombinante para producir una molécula de ácidos nucleicos recombinante. La secuencia de ácidos nucleicos codificante de la proteína que debe producirse se inserta en el vector de una manera que liga operablemente la secuencia de ácidos nucleicos con secuencias reguladoras en el vector (p.ej., un promotor de Traustoquitriales de la presente invención), permitiendo la transcripción y traducción de la secuencia de ácidos nucleicos dentro del microorganismo recombinante. Los marcadores seleccionables de la presente invención permiten la selección de un microorganismo recombinante en el que se ha introducido con éxito una molécula de ácidos nucleicos recombinante de la presente invención.

En otra realización, un vector recombinante de la presente invención es un vector de localización. Tal como se utiliza en la presente memoria, la expresión "vector de localización" se utiliza para referirse a un vector que se utiliza para introducir una molécula particular de ácidos nucleicos en una célula recombinante, en la que la molécula de ácidos nucleicos se utiliza para deleccionar o inactivar un gen endógeno dentro de la célula huésped (es decir, se utiliza para la tecnología de disrupción génica dirigida o de inactivación génica). Dicho vector también puede conocerse en la técnica como un vector "knock-out". En un aspecto de la presente realización, una parte del vector, aunque más típicamente, la molécula de ácidos nucleicos insertada en el vector (es decir, el inserto), presenta una secuencia de ácidos nucleicos que es homóloga respecto a una secuencia de ácidos nucleicos de un gen diana en la célula huésped (es decir, un gen que es la diana para la deleción o inactivación). La secuencia de ácidos nucleicos del inserto vector se diseña para unirse al gen diana de manera que el gen diana y el inserto experimenten recombinación homóloga, de manera que el gen diana endógeno resulta deleccionado, inactivado o atenuado (es decir, por la mutación o deleción de como mínimo una parte del gen diana endógeno).

El vector recombinante de la presente invención es un vector recombinante que resulta adecuado para la utilización en un microorganismo Traustoquitriales.

El experto en la materia apreciará que la utilización de tecnologías de ADN recombinante pueden mejorar el control de la expresión de las moléculas de ácidos nucleicos transformadas mediante la manipulación, por ejemplo, del número de copias de las moléculas de ácidos nucleicos dentro de la célula huésped, la eficiencia con la que se transcriben dichas moléculas de ácidos nucleicos, la eficiencia con la que se traducen los transcritos resultantes y la eficiencia de las modificaciones post-traduccionales. Además, la secuencia de promotor puede manipularse genéticamente para mejorar el nivel de expresión en comparación con el promotor nativo. Entre las técnicas recombinantes útiles para controlar la expresión de moléculas de ácidos nucleicos se incluyen, aunque sin limitarse a ellas, la integración de las moléculas de ácidos nucleicos en una o más cromosomas de la célula huésped, la adición de secuencias de estabilidad del vector a plásmidos, las sustituciones o modificaciones de señales de control transcripcional (p.ej., promotores, operadores e intensificadores), las sustituciones o modificaciones de señales de control traduccional (p.ej., sitios de unión ribosómica, secuencias de Shine-Dalgarno), la modificación de moléculas de ácidos nucleicos para que se correspondan con el uso de codones de la célula huésped y la deleción de secuencias que desestabilizan los transcritos.

En una realización de la presente invención, un vector recombinante adecuado para la utilización en la transformación de microorganismos Traustoquitriales contiene el gen *Ble* Sh de *Streptoalloteichus hindustanus* como marcador seleccionable (que codifica una "proteína de unión a bleomicina") en combinación con un promotor de Traustoquitriales tal como se ha indicado anteriormente en la presente memoria. Un vector recombinante de la invención comprende el gen *ble* y un promotor de Traustoquitriales incluye, por ejemplo, la secuencia de vector representada por la SEC ID nº 9. La secuencia de aminoácidos de la proteína de unión a bleomicina de *Streptoalloteichus hindustanus* se representa en la presente memoria como SEC ID nº 10.

- Las moléculas de ácidos nucleicos recombinantes de la presente invención, que pueden ser de ADN o ARN, también pueden contener secuencias reguladoras adicionales, tales como secuencias reguladoras de la traducción, orígenes de replicación y otras secuencias reguladoras que son compatibles con la célula recombinante. En una realización, una molécula recombinante de la presente invención, incluyendo aquellas que están integradas en el cromosoma de la célula huésped, también contienen señales secretorias (es decir, secuencias de ácidos nucleicos de segmento de señal) para permitir que la proteína expresada sea secretada de la célula que produce la proteína. Entre los segmentos de señal adecuados se incluyen un segmento de señal que está asociado naturalmente a la proteína que debe expresarse o a cualquier segmento de señal heterólogo capaz de dirigir la secreción de la proteína según la presente invención. En otra realización, una molécula recombinante de la presente invención comprende una secuencia líder para permitir que una proteína expresada sea transportada e insertada en la membrana de una célula huésped. Entre las secuencias líder adecuadas se incluyen una secuencia líder que está asociada naturalmente a la proteína, o cualquier secuencia líder heteróloga capaz de dirigir el transporte e inserción de la proteína en la membrana de una célula.
- En una realización, la molécula de ácidos nucleicos recombinante comprende además una secuencia de ácidos nucleicos codificante de una proteína que debe producir la célula, en la que la secuencia de ácidos nucleicos codificante de la proteína está operablemente ligada a una secuencia de control transcripcional. Las proteínas que pueden resultar deseables para la producción en un Traustoquitriales son conocidas por el experto en la materia y todas pretender estar comprendidas en la presente invención. Entre las proteínas particularmente preferentes se incluyen, aunque sin limitarse a ellas, proteínas asociadas a la síntesis de un ácido graso seleccionado del grupo que consiste en ácido docosahexaenoico (DHA), ácido docosapentaenoico (DPA), ácido eicosapentaenoico (EPA) y ácido araquidónico (ARA). Entre dichas proteínas se incluyen, por ejemplo, una ácido-graso sintasa, una ácido graso desaturasa, una ácido graso elongasa, una proteína asociada a un complejo de poliquétido sintasa y una proteína asociada a la incorporación de ácidos grasos en fosfolípidos o en moléculas de triacilglicerol. En un aspecto, la proteína es una ácido-graso ω -3 desaturasa codificada por la secuencia de ácidos nucleicos SEC ID n° 29. La SEC ID n° 30 representa la secuencia de aminoácidos de la desaturasa. En otro aspecto, la proteína es una isomerasa de ácido graso polienoico. En una realización, entre las proteínas que pueden producirse en microorganismos Traustoquitriales utilizando el presente método se incluyen proteínas asociadas a las rutas biosintéticas de isoprenoide. Entre dichas proteínas se incluyen, aunque sin limitarse a ellas, HMG-CoA sintasa, HMG-CoA reductasa, escualeno sintasa, fitoeno sintasa, fitoeno desaturasa, una carotenoide ciclasa, una carotenoide hidroxilasa y una carotenoide quetolasa. En todavía otra realización, las proteínas que pueden producirse en los microorganismos Traustoquitriales utilizando el presente método se incluyen, aunque sin limitarse a ellos, vitamina E y ácido lipoico.
- En una realización, la molécula de ácidos nucleicos recombinante útil en el método de la presente invención incluye una secuencia de ácidos nucleicos que se hibrida con una secuencia diana de ácidos nucleicos en el microorganismo, de manera que un gen que comprende la secuencia de ácidos nucleicos diana resulta mutado o inactivado mediante recombinación homóloga con la segunda secuencia de ácidos nucleicos. Dicha secuencia de ácidos nucleicos puede ser homóloga respecto a genes que codifican enzimas (o ácidos nucleicos que regulan la expresión de dichos genes) de las rutas de síntesis de ácido graso saturado y poliinsaturado, genes codificantes de proteínas que participan en la degradación de otros compuestos valiosos producidos por el microorganismo Traustoquitriales o que, de otra manera, disminuyen el valor del compuesto deseado, o genes codificantes de proteínas que están asociados a la síntesis de compuestos la síntesis de los cuales compite con otras moléculas de interés. Por ejemplo, entre las secuencias de ácidos nucleicos diana se incluyen, aunque sin limitarse a ellas, secuencias codificantes de lipasas, enzimas de oxidación de ácidos grasos, proteínas que participan en la síntesis de carbohidratos, proteínas que participan en la síntesis de productos de las rutas de isoprenoide, proteínas que participan en la síntesis de componentes de la pared celular, proteínas que participan en las rutas de síntesis de ácidos grasos saturados, proteínas que participan en las rutas de síntesis de ácidos grasos poliinsaturados, proteínas asociadas con un complejo de poliquétido sintasa y proteínas asociadas a la incorporación de ácidos grasos en fosfolípidos o moléculas de triacilglicerol.
- En una realización de la presente invención, el método para la transformación de microorganismos Traustoquitriales incluye una etapa de introducción en la célula de por lo menos una molécula adicional de ácidos nucleicos recombinante que comprende una secuencia de ácidos nucleicos codificante de una proteína que debe expresarse, estando operablemente ligada a la secuencia de ácidos nucleicos a una secuencia de control transcripcional. Alternativamente, la molécula adicional de ácidos nucleicos recombinante puede incluir una segunda secuencia de ácidos nucleicos que se hibrida con una secuencia de ácidos nucleicos diana en el microorganismo, de manera que un gen que comprende la secuencia de ácidos nucleicos diana resulta mutada o inactivada mediante recombinación homóloga con la segunda secuencia de ácidos nucleicos. De esta manera, pueden introducirse múltiples proteínas en la célula, pueden inactivarse múltiples genes o resultan posibles combinaciones de los dos. La molécula adicional de ácidos nucleicos recombinante puede introducirse en el microorganismo Traustoquitriales simultáneamente con la primera molécula de ácidos nucleicos recombinante (es decir, cotransformarse), o como una transformación posterior (p.ej., a fin de "acumular" características).
- En una realización, el método incluye además la etapa de introducir en la célula por lo menos una molécula adicional de ácidos nucleicos recombinante que comprende una secuencia de ácidos nucleicos codificante de una proteína de unión a bleomicina. En esta realización, la molécula adicional de ácidos nucleicos recombinante se introduce

preferentemente en una etapa posterior, en lugar de cotransformarse. Preferentemente, la molécula de ácidos nucleicos recombinante que comprende una secuencia de ácidos nucleicos codificante de una proteína de unión a bleomicina que comprende además una secuencia de ácidos nucleicos codificante de una segunda proteína que debe ser expresada por la célula, en la que la secuencia de ácidos nucleicos codificante de la segunda proteína se liga operablemente a una secuencia de control transcripcional. Alternativamente, o adicionalmente, la molécula de ácidos nucleicos recombinante que comprende una secuencia de ácidos nucleicos codificante de una proteína de unión a bleomicina comprende además una segunda secuencia de ácidos nucleicos que se hibrida con una secuencia diana de ácidos nucleicos en el microorganismo, de manera que un gen que comprende la secuencia diana de ácidos nucleicos resulta mutada o inactivada mediante recombinación homóloga con la segunda secuencia de ácidos nucleicos. En una realización, dicha molécula de ácidos nucleicos recombinante comprende una secuencia de ácidos nucleicos SEC ID nº 9.

Entre las células huésped adecuadas para la transformación utilizando el método de la presente invención se incluyen, aunque sin limitarse a ellas, cualquier microorganismo del orden Traustoquiritales. Las células huésped pueden ser células no transformadas o células que ya se encuentran transfectadas con por lo menos una molécula de ácidos nucleicos. Las células huésped preferentes para la utilización en la presente invención se incluyen microorganismos de un género, incluyendo, aunque sin limitación: *Thraustochytrium*, *Labyrinthuloides*, *Japonochytrium* y *Schizochytrium*. Entre las especies preferentes dentro de estos géneros se incluyen, aunque sin limitarse a ellos, cualquier especie de *Schizochytrium*, *Schizochytrium limacinum*, *Schizochytrium minutum*; cualquier especie de *Thraustochytrium* species (incluyendo especies de *Ulkenia* anteriores, tales como *U. visurgensis*, *U. amoebaida*, *U. sarkariana*, *U. profunda*, *U. radiata* y *U. minuta*), e incluyendo *Thraustochytrium striatum*, *Thraustochytrium aureum*, *Thraustochytrium roseum* y cualquier especie de *Japonochytrium*. Entre las cepas particularmente preferentes de Traustoquiritales se incluyen, aunque sin limitarse a ellas: *Schizochytrium* sp. (S31)(ATCC 20888); *Schizochytrium* sp. (S8) (ATCC 20889); *Schizochytrium* sp. (LC-RM)(ATCC 18915); *Schizochytrium* sp. (SR21); *Schizochytrium aggregatum* (Goldstein et Belsky)(ATCC 28209); *Schizochytrium limacinum* (Honda et Yokochi)(IFO 32693); *Thraustochytrium* sp.(23B)(ATCC 20891); *Thraustochytrium striatum* (Schneider)(ATCC 24 473); *Thraustochytrium aureum* (Goldstein)(ATCC 34304); *Thraustochytrium roseum* (Goldstein) (ATCC 28210) y *Japonochytrium* sp. (L1)(ATCC 28207).

Según la presente invención, el término "transformación" se utiliza para referirse a cualquier método por el que una molécula de ácidos nucleicos exógena (es decir, una molécula de ácidos nucleicos recombinante) puede insertarse en las células microbianas, tales como las células microbianas de Traustoquiritales. En sistemas microbianos, el término "transformación" se utiliza para describir un cambio hereditario debido a la adquisición de ácidos nucleicos exógenos por los microorganismos y es esencialmente sinónimo al término "transfección". Entre las técnicas de transformación adecuadas se incluyen, aunque sin limitarse a ellas, el bombardeo de partículas, la electroporación, la microinyección, la lipofección, la adsorción, la infección y la fusión de protoplastos.

En una realización, una proteína que debe producirse utilizando un método de la presente invención se produce mediante el cultivo de una célula que expresa la proteína (es decir, un microorganismo Traustoquiritales recombinante) bajo condiciones eficaces para producir la proteína. En algunos casos, la proteína puede recuperarse, y en otros, el microorganismo puede recolectarse completo o como lisado y utilizarse como una "biomasa". En otra realización, un gen diana se deleciona o se inactiva mediante el cultivo de una célula que ha sido transformada con una molécula recombinante que comprende un vector de localización de la presente invención bajo condiciones eficaces para permitir la recombinación dentro de la célula, resultando en la delección o inactivación de un gen diana. Una célula preferente para el cultivo es una célula recombinante de la presente invención. Entre las condiciones de cultivo eficaces se incluyen, aunque sin limitarse a ellas, medios eficaces, biorreactor, temperatura, pH y condiciones de oxígeno que permiten la producción de proteína y/o la recombinación. Un medio eficaz se refiere a cualquier medio en el que típicamente se cultiva una célula Traustoquiritales. Dicho medio típicamente comprende un medio acuoso que presenta fuentes asimilables de carbono, nitrógeno y fosfato, y sales, minerales, metales y otros nutrientes apropiados, tales como vitaminas. Se comentan en detalle ejemplos de medios y condiciones de cultivo adecuados en la sección de Ejemplos. Las condiciones de cultivo adecuadas para los microorganismos Traustoquiritales también se describen en la patente US nº 5.340.742, publicada el 23 de agosto de 1994, de Barclay. Las células de la presente invención pueden cultivarse en biorreactores de fermentación convencionales, matraces de agitación, probetas, placas de microtitulación y placas Petri. El cultivo puede llevarse a cabo a una temperatura, pH y contenido de oxígeno apropiados para una célula recombinante. Tales condiciones de cultivo se encuentran comprendidas dentro de los conocimientos del experto ordinario en la materia.

Dependiendo del vector y sistema huésped utilizado para la producción, las proteínas resultantes de la presente invención pueden permanecer dentro de la célula recombinante, secretarse al medio de fermentación, secretarse a un espacio entre dos membranas celulares, o retenerse sobre la superficie externa de una membrana celular. La expresión "recuperación de la proteína" se refiere a recolectar el medio de fermentación completa que contiene la proteína y no requiere necesariamente etapas adicionales de separación o purificación. Las proteínas producidas mediante el método de la presente invención pueden purificarse utilizando una diversidad de técnicas estándares de purificación de proteínas, tales como, aunque sin limitarse a ellas, la cromatografía de afinidad, la cromatografía de intercambio iónico, la filtración, la electroforesis, la cromatografía de interacción hidrofóbica, la cromatografía de

filtración en gel, la cromatografía de fase inversa, la cromatografía de cancanavalina A, el cromatoenfoco y la solubilización diferencial. Las proteínas producidas mediante el método de la presente invención preferentemente se recuperan en forma "sustancialmente pura". Tal como se utiliza en la presente memoria, "sustancialmente puro" se refiere a una pureza que permite la utilización eficaz de la proteína como producto comercial.

5 Todavía otra realización de la presente invención se refiere a un microorganismo recombinante del orden Traustozytriales que ha sido transformado con una molécula de ácidos nucleicos recombinante que comprende una secuencia de ácidos nucleicos codificante de una acetolactato sintasa de la presente invención. Preferentemente, la acetolactato sintasa proporciona al microorganismo una sensibilidad reducida a compuestos seleccionados del grupo que consiste en compuestos sulfonilurea, inhibidores de clase imidazolinona y oxibenzoatos de pirimidinilo. Las moléculas de ácidos nucleicos recombinantes adecuadas y las secuencias para la utilización en la transformación de dicho microorganismo han sido descritas en detalle anteriormente. Dicho microorganismo puede transformarse adicionalmente con otras moléculas de ácidos nucleicos recombinantes, incluyendo moléculas de ácidos nucleicos recombinantes que comprenden un marcador seleccionable gen *ble* y secuencias de control transcripcional de Traustozytriales, tal como se ha indicado anteriormente en la presente memoria. Los microorganismos Traustozytriales recombinantes según la presente invención se describen en la sección de Ejemplos. Según la presente invención, un microorganismo Traustozytriales recombinante de la presente invención se manipula genéticamente para expresar una proteína de interés (se han comentado anteriormente ejemplos de tales proteínas) utilizando los vectores recombinantes indicados en la presente memoria y/o se manipulan genéticamente para una delección o inactivación dirigida de un gen diana utilizando los vectores recombinantes indicados en la presente memoria.

25 Tal como se utiliza en la presente memoria, un microorganismo recombinante presenta un genoma que ha sido modificado (es decir, mutado o modificado) respecto a su forma normal (es decir, de tipo salvaje o natural) utilizando tecnología recombinante. Un microorganismo recombinante según la presente invención puede incluir un microorganismo en el que se han insertado, delecionado o modificado moléculas de ácidos nucleicos (es decir, se han mutado, p.ej., mediante inserción, delección, sustitución y/o inversión de nucleótidos), de manera que tales modificaciones proporcionen el efecto deseado dentro del microorganismo. Tal como se utiliza en la presente memoria, las modificaciones genéticas que resultan en una reducción de la expresión génica, en la función del gen, o en la función del producto génico (es decir, la proteína codificada por el gen) puede denominarse inactivación (completa o parcial), delección, interrupción, bloqueo o regulación negativa de un gen. Por ejemplo, una modificación genética en un gen que resulta en una reducción de la función de la proteína codificada por dicho gen puede resultar en una delección completa del gen (es decir, el gen no existe y, por lo tanto, la proteína no existe), una mutación en el gen que resulta en una traducción incompleta o nula de la proteína (p.ej., la proteína no se expresa) o una mutación en el gen que reduce o anula la función natural de la proteína (p.ej., se expresa una proteína que presenta una actividad o acción enzimática reducida o nula). Puede hacerse referencia a las modificaciones genéticas que resultan en un incremento de la expresión o función génica como amplificación, sobreproducción, sobreexpresión, activación, potenciación, adición o regulación positiva de un gen.

40 Según la presente invención, puede producirse un microorganismo Traustozytriales recombinante utilizando cualquier microorganismo del orden Traustozytriales. Entre los géneros preferentes de Traustozytriales se incluyen, aunque sin limitarse a ellos: *Thraustochytrium*, *Labyrinthuloides*, *Japonochoytrium*, y *Schizochytrium*. Entre las especies preferentes dentro de dichos géneros se incluyen, aunque sin limitarse a ellas: cualquier especie de *Schizochytrium* incluyendo *Schizochytrium aggregatum* y *Schizochytrium limacinum*; cualquier especie de *Thraustochytrium* (incluyendo cualquier de las especies anteriores de *Ulkenia*, tales como *U. visurgensis*, *U. amoeboida*, *U. sarkariana*, *U. profunda*, *U. radiata*, *U. minuta* y *Ulkenia* sp. BP-5601), *Thraustochytrium striatum*, *Thraustochytrium aureum*, *Thraustochytrium roseum*; y cualquier especie de *Japonochoytrium*. Entre las cepas particularmente preferentes de Traustozytriales se incluyen, aunque sin limitarse a ellas: *Schizochytrium* sp. (S31)(ATCC 20888); *Schizochytrium* sp. (S8) (ATCC 20889); *Schizochytrium* sp. (LC-RM)(ATCC 18915); *Schizochytrium* sp. (SR21); *Schizochytrium aggregatum* (Goldstein et Belsky)(ATCC 28209); *Schizochytrium limacinum* (Honda et Yokochi)(IFO 32693); *Thraustochytrium* sp.(23B)(ATCC 20891); *Thraustochytrium striatum* (Schneider)(ATCC 24 24473); *Thraustochytrium aureum* (Goldstein)(ATCC 34304); *Thraustochytrium roseum* (Goldstein)(ATCC 28210) y *Japonochoytrium* sp. (L1)(ATCC 28207).

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

Ejemplos

60 Ejemplo 1

El presente ejemplo describe la producción del plásmido recombinante pTUBZEO11-2. La construcción del plásmido recombinante pTUBZEO11-2 se ilustra en las figs. 1 y 2. Este plásmido contiene el gen *ble* de *Streptoaloteichus hindustanus* funcionalmente acoplado con un promotor de gen de α -tubulina aislado a partir de *Schizochytrium* sp. Este plásmido se produjo de la manera siguiente: Se aisló un clon de ADNc (CGNE0002-001-B6) a partir de una

5 biblioteca de ADNc de *Schizochytrium* sp. y se secuenció parcialmente (SEC ID nº 1) como parte de un proyecto de secuenciación de ADNc a gran escala de *Schizochytrium*. Se determinó que la secuencia de nucleótidos codificaba la α -tubulina mediante búsqueda de homología de BLASTX (Gish W. y D. States, Nat. Genet. 3:266-272, 1993). La secuencia de aminoácidos deducida a partir de las bases 116 a 550 era 93% idéntica a los primeros 145 aminoácidos de la α -tubulina de *Pelvetica fastigiata* (GenBank nº de acceso U58642).

10 Con el fin de aislar el promotor asociado a dicho gen, se aisló el ADN genómico a partir de células de *Schizochytrium* sp. y se procesó mediante la utilización de un kit "GenomeWalker™" (Clontech Laboratories, Inc., Palo Alto, CA), que implica la digestión enzimática de ADN genómico con endonucleasas de restricción para generar extremos romos, seguido de la ligación del ADN digerido a moléculas adaptadoras específicas de ADN bicatenario proporcionadas en el kit. A continuación, se amplificó el ADN cadena arriba de la secuencia codificante de α -tubulina mediante la reacción en cadena de polimerasa (PCR) utilizando el cebador adaptado externo (AP1) proporcionado en el kit y el cebador específico de α -tubulina PGR20 (SEC ID nº 2). Se llevó a cabo una amplificación adicional del gen utilizando el cebador adaptador anidado (AP2) proporcionado en el kit y el cebador anidado específico de α -tubulina PGR19 (SEC ID nº 3).
15 Los productos de PCR resultantes se subclonaron en el plásmido pCR2.1-TOPO (Invitrogen Corp., Carlsbad, CA). Se secuenció uno de los fragmentos subclonados: la secuencia de 725 pb inmediatamente anteriores al codón de inicio del gen de α -tubulina se proporciona como SEC ID nº 4.

20 Utilizando cebadores oligonucleótidos basándose en la secuencia de ADN obtenida de esta manera, se utilizó la PCR utilizando la ADN polimerasa Taq (Perkin-Elmer Corp., Norwalk, CT) para generar una región de promotor de α -tubulina modificada en la que se había incorporado un sitio de restricción NcoI en el extremo 3' del fragmento de ADN; este sitio NcoI contenía un codón de inicio que se encontraba en la misma posición que en la región codificante de α -tubulina. Los cebadores utilizados en esta reacción eran PGR33 (SEC ID nº 5) y PGR34 (SEC ID nº 6) y el molde era ADN genómico aislado a partir de células de *Schizochytrium* sp. Se utilizaron las condiciones de reacción siguientes:
25 94°C durante 4 min; (94°C durante 1 min, 54°C durante 45 s, 72°C durante 2 min) x 30; 72°C durante 7 min. Este fragmento se clonó en el plásmido pCR2.1-TOPO para formar el plásmido p7TUB (SEC ID nº 7). Se digirió el plásmido p7TUB con NcoI y un fragmento de 463 pb resultante que contenía la región de promotor de α -tubulina de *Schizochytrium* se aisló mediante purificación en gel de agarosa. El plásmido pSV40/Zeo (Invitrogen Corp., Carlsbad, CA), que contiene el gen *ble* de *Streptoalloteichus hindustanus* flanqueado por un promotor de SV40 y terminador,
30 también se digirió con NcoI, rindiendo un fragmento de 3.201 pb y un fragmento de 314 pb. El fragmento de 3.201 pb se purificó en gel de agarosa y se ligó en el fragmento NcoI de 463 pb a partir de p7TUB, rindiendo pTUBZEO-11 (SEC ID nº 8), ilustrado en la fig. 1.

35 A continuación, el plásmido pTUBZEO-11 se digirió con *SphI* y un fragmento de 1.122 pb resultante que contenía el gen *ble* flanqueado por el promotor de α -tubulina de *Schizochytrium* y el terminador de SV40 se purificó en gel de agarosa y se ligó al plásmido pUC19 (Messing, J., *Meth. Enzymol.* 101:20, 1983) que había sido linearizado mediante digestión con *SphI*. El plásmido resultante se denominó pTUBZEO11-2 (SEC ID nº 9) y se ilustra en las figs. 2 y 4. El plásmido pTUBZBO11-2 también se denomina pMON50000. En la SEC ID nº 9, se encuentra contenido el promotor de α -tubulina de *Schizochytrium*, dentro de los nucleótidos 441-894; la región codificante del gen *ble* se encuentra
40 contenida dentro de los nucleótidos 895-1.269, y el terminador de SV40 se encuentra contenido dentro de los nucleótidos 1.270-1.524.

Ejemplo 2

45 El presente ejemplo describe la producción de los plásmidos recombinantes pMON50200, pMON50201, pMON50202 y pMON50203.

50 Se aisló el gen codificante de acetolactato sintasa nativa (*als*) de *Schizochytrium* sp. de la manera siguiente. Se aisló un clon de ADNc (LIB81-028-Q1-E1-D9) a partir de una biblioteca de ADNc de *Schizochytrium* y se secuenció parcialmente (SEC ID nº 11) como parte de un proyecto de secuenciación de ADNc a gran escala de *Schizochytrium* sp. Se determinó mediante homología de BLASTX que la secuencia de nucleótidos codificaba acetolactato sintasa, p.ej., la secuencia de aminoácidos deducida de las bases 154 a 378 era 68% idéntica a los aminoácidos 313 a 387 de ALS de *Schizosaccharomyces pombe* (GenBank, nº de acceso P36620). A continuación, se obtuvo la secuencia de longitud completa de dicho ADNc clonado, que indicaba que el clon de ADNc no contenía la región codificante de *als* entera.
55 Con el fin de obtener el gen *als* de longitud completa, se sondeó una biblioteca genómica de lambda de *Schizochytrium* utilizando protocolos estándares (ver, p.ej., Sambrook et al., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 1989) con una sonda de ADN de 372 pb marcada con digoxigenina (DIG, denominada sonda ALS2). Se generó la sonda LAS2 mediante PCR utilizando una mezcla de nucleótidos que incluía DIG-11-UTP (Boehringer Mannheim Biochemicals GmbH, Alemania), utilizando el cebador directo PGR38 (SEC ID nº 12) y el cebador inverso PGR39 (SEC ID nº 13), que estaban basados en la secuencia de ADNc clon LIB81-028-Q1-E1-D9. Se aisló uno de los clones genómicos identificados con la sonda ALS2, denominado ALS-4A, y se caracterizó adicionalmente mediante transferencias de hibridación southern utilizando la sonda ALS2 marcada con DIG. Se encontró que un fragmento de 4,9 kpb procedente de ADN de lambda ALS-4A digerido con *AhdI* se hibridaba con la sonda ALS2. Se aisló este fragmento mediante purificación en gel de agarosa, se trató con ADN polimerasa de T4 con
65 el fin de generar extremos romos y después se ligó en pBluescriptII KS+ digerido con *SmaI* (Stratagene Corp., La

Jolla, CA) para formar el plásmido pMON50200 (ilustrado en la fig. 3-A). La secuencia de pMON50200 se proporciona como SEC ID n° 14. La secuencia del enzima acetolactato sintasa codificada por el gen *als* de *Schizochytrium* se proporciona como SEC ID n° 15.

5 Se produjeron los plásmidos pMON50201, pMON50202 y pMON50203 (ilustrados en las figs. 3-B, 3-C y 3-D, respectivamente) a partir del plásmido pMON50200 mediante mutagénesis dirigida a sitio, de manera que los enzimas acetolactato sintasa codificados ya no resultan inhibidos por determinados compuestos, incluyendo sulfometurón-metilo (SMM). Estos plásmidos se construyeron de la manera siguiente. El kit de mutagénesis dirigida a sitio "Transformera" (Clontech Laboratories Inc., Palo Alto, CA) se utilizó para introducir las mutaciones siguientes en el
10 plásmido pMON50200 siguiendo las instrucciones del fabricante. Se utilizó un cebador oligonucleótido de selección, DM19 (SEC ID n° 16) en los tres constructos; este cebador conduce a la conversión de un sitio *EcoRV* único en el sitio de clonación múltiple de pMON50200 en un sitio *AuflI*. Se utilizó el cebador DM14 (SEC ID n° 17) para modificar el residuo aminoácido número 595 en el enzima ALS codificado de triptófano a valina, introduciendo simultáneamente un sitio *AcII* en la secuencia génica; el plásmido resultante se denominó pMON50201 (SEC ID n° 18). De manera similar, el cebador DM15 (SEC ID n° 20) se utilizó para modificar el residuo aminoácido número 192 en el enzima ALS codificado de una prolina a una glutamina y para introducir un sitio *BsgI* en el gen *als*, resultando en el plásmido pMON50202 (SEC ID n° 21). Para la construcción del plásmido pMON50203 (SEC ID n° 23), se utilizaron ambos
15 cebadores DM14 y DM15, resultando en un enzima ALS codificado que contenía ambas sustituciones de residuos aminoácidos descritos anteriormente. Las secuencias de los enzimas acetolactato sintasa mutante codificados por los plásmidos pMON50201, pMON50202 y pMON50203 se proporcionan como SEC ID n° 19, SEC ID n° 22 y SEC ID n° 24, respectivamente.

Ejemplo 3

25 El presente ejemplo describe la transformación genética de *Schizochytrium* sp. con las moléculas recombinantes descritas en los Ejemplos 1 y 2.

La cepa utilizada en el presente ejemplo en *Schizochytrium* sp. N230D, un derivado de la cepa 20888 de la American Type Culture Collection (ATCC, Manassas, VA). Para los cultivos líquidos, las células se cultivaron axénicamente en
30 medio M50-3 a 30°C bajo agitación a 200-300 rpm. El medio M50-3 contenía los componentes siguientes: NaCl, 12,5 g; MgSO₄·7H₂O, 2,5 g; KCl, 0,5 g; CaCl₂, 0,05 g; glucosa, 30 g; Na-glutamato, 3 g; KH₂PO₄, 0,4 g; extracto de levadura, 1 g; NaHCO₃, 0,4 g; Na₂EDTA, 30 mg; FeCl₃·6H₂O, 1,2 mg; H₃BO₃, 34,2 mg; ZnSO₄·7H₂O, 0,67 mg; CoCl₂·6H₂O, 0,13 mg; NaMoO₄·2H₂O, 25 µg; CuSO₄·5H₂O, 10 µg; NiSO₄·6H₂O, 0,26 mg; tiamina·HCl, 100 µg; biotina, 0,5 µg; cianocobalamina, 0,5 µg y agua desionizada (enrase a un litro); pH final ajustado a 7,0. Para el crecimiento en medio sólido, las células se cultivaron a 30°C sobre medio M50-3 o medio M1E-3 solidificado mediante la adición de agar al 1,5% (p/v). El medio M1E-3 contiene los componentes siguientes: glucosa, 4 g; (NH₄)₂SO₄, 0,75 g; Na₂SO₄, 5 g; MgSO₄·7H₂O, 2 g; KH₂PO₄, 0,5 g; KCl, 0,5 g; CaCl₂·2 H₂O, 0,1 g; tampón MOPS, 20,9 g; FeSO₄·4H₂O, 0,3 mg; MnCl₂·4H₂O, 0,1 mg; ZnSO₄·7H₂O, 80 µg; CoCl₂·6H₂O, 2 µg; NaMoO₄·2H₂O, 1 µg; CuSO₄·5H₂O, 60 µg; NiSO₄·6H₂O, 80 µg; tiamina·HCl, 320 µg; CA-pantotenato, 320 µg; cianocobalamina, 8 µg, y agua desionizada (enrase a un litro);
40 pH final ajustado a 7,0.

La sensibilidad de *Schizochytrium* sp. a Zeocin™ y SMM se determinó mediante la inclusión de estos inhibidores en medio M1E-3 solidificado a diversas concentraciones y extendiendo las células sobre las placas a densidades similares a las presentes durante los procedimientos utilizados para la selección de células recombinantes.

45 Se llevó a cabo la transformación genética de las células de *Schizochytrium* mediante bombardeo de partículas (Sanford J.C., F.D. Smith y J.A. Russell, Meth. Enzymol. 217:483-509, 1993) utilizando un sistema de inyección de partículas Bio-Rad Biolistic PDS-1000/He (Bio-Rad Laboratories, Hercules, Ca). Se cultivaron células *Schizochytrium* sp. N230D en medio M50-3 líquido hasta una densidad óptica de 680 nm (DO₆₈₀) de 0,4 a 0,8 (longitud del camino óptico: 10 mm). Se centrifugó brevemente una alícuota de células correspondiente a 1,0 DO₆₈₀, se eliminó la solución sobrenadante y las células peletizadas se resuspendieron en 100 µl de agua estéril. A continuación, las células resuspendidas se extendieron en un círculo de 4 a 6 cm sobre una placa Petri que contenía medio agar solidificado (p.ej., medio M50-3 o M1E-3) y se dejó en reposo durante 30 a 60 min de manera que pudiese absorberse el exceso de agua en el medio sólido; a esto se le denomina placa diana.

55 Se utilizó una alícuota de 1,5 mg de microportadores de oro (diámetro nominal de 0,6 µm, disponible de Bio-Rad Laboratories, Inc., Hercules, CA) con 2,5 µg de ADN plasmídico de transformación (es decir, plásmido TUBZEO11-2, pMON50201, pMON50202 o pMON50203) siguiendo las instrucciones del fabricante (manual de instrucciones del sistema de inyección de partículas Biolistic® PDS-1000/He; Bio-Rad Laboratories, Hercules, CA). Las células se bombardearon con los microportadores de oro recubiertos con ADN utilizando las condiciones siguientes: disco de ruptura de 1.100 psi, cámara de vacío de 25° Hg, ensamblaje de lanzamiento de microportador en el estante superior y placas diana en el estante intermedio, proporcionando una distancia de disco de ruptura a malla de retención de 1,5 a 2 cm y una distancia de malla de retención a diana de aproximadamente 7 cm. Tras el bombardeo, se dejó que las células se recuperasen sobre las placas diana durante 4 a 6 horas a 30°C. A continuación, se desprendieron las células de las placas diana mediante enjuague con 1,5 ml de agua estéril, se recolectaron en un tubo de microcentrífuga, se
65

centrifugaron brevemente y se resuspendieron en 400 µl de agua estéril. Se extendieron cien microlitros de la suspensión sobre cuatro placas M1E-3 que contenía 150 a 200 µg/ml de Zeocin™ (Invitrogen Corp., Carlsbad, CA) o 25 µg/ml de SMM. Las placas que contenían Zeocin™ se utilizaron para seleccionar las células que habían sido transformadas con plásmido pTUBZEO11-2, mientras que las placas que contenían SMM se utilizaron para seleccionar las células que habían sido transformadas con los plásmidos pMON50201, pMON50202 o pMON50203. A continuación, las placas se incubaron durante 7 a 10 días a 30°C. Las colonias que aparentemente eran resistentes al agente selectivo seguidamente se sembraron sobre placas de M1E-3 fresco que contenía el mismo agente selectivo, a fin de confirmar la resistencia. Este protocolo resulta típicamente en la generación de 100 a 1.000 cepas resistentes a Zeocin™ o a SMM por cada bombardeo.

Ejemplo 4

El ejemplo a continuación demuestra el análisis por PCR de células de *Schizochytrium* transformadas.

Se utilizó la PCR para confirmar la presencia de secuencias de plásmido en las cepas putativamente transformadas que eran resistentes a los agentes selectivos Zeocin™ o SMM. Se obtuvo ADN molde de transformantes putativos y células de *Schizochytrium* N230D no recombinantes, mediante la utilización de un asa de inoculación de 1 µl de plástico de un solo uso a fin de extraer una pequeña cantidad de células (1 a 2 mm³) de las colonias resistentes que habían sido sembradas sobre placas de agar (tal como se indica en el Ejemplo 3). Las células se resuspendieron en 15 a 20 µl de Triton X-100 al 1% en un tubo de microcentrifuga, se introdujeron en un baño de agua en ebullición durante 10 minutos y después se centrifugaron durante 5 minutos a 14.000xg. Se utilizaron partes de estos extractos (1 a 3 µl) para proporcionar el ADN molde para reacciones de PCR de 25 µl utilizando ADN polimerasa *Taq*. Para detectar la presencia de secuencias de pTUBZEO11-2 en el ADN de *Schizochytrium*, se utilizaron los cebadores DM20 (SEC ID nº 25) y DM21 (SEC ID nº 26); estos cebadores se hibridan con el gen *ble* en el plásmido pTUBZEO11-2 y amplifican un fragmento de ADN de 346 pb. El perfil térmico utilizado fue el siguiente: 94°C durante 4 min; (94°C durante 45 s, 52°C durante 45 s, 72°C durante 2 min) x 30; 72°C durante 7 min. Con el fin de detectar la presencia de secuencias de pMON50201, pMON50202 o pMON50203 en el ADN de *Schizochytrium*, se utilizaron los cebadores BLA1 (SEC ID nº 27) y BLA2 (SEC ID nº 28); estos cebadores se hibridan con el gen *bla* (resistencia a ampicilina) presente en el esqueleto de vector y amplifican un fragmento de ADN de 1.229 pb. El perfil térmico utilizado fue el siguiente: 94°C durante 4 min (94°C durante 45 s, 55°C durante 45 s, 72°C durante 2 min) x 30; 72°C durante 7 min. Los productos de PCR se analizaron mediante electroforesis en gel de agarosa estándar, seguido de la tinción con bromuro de etidio.

Los resultados de estos análisis confirman que la amplia mayoría de cepas seleccionadas bajo estas condiciones son transformantes verdaderos que contiene ADN plasmídico. No se generaron productos de PCR del tamaño correcto al utilizar ADN molde de las células de *Schizochytrium* sp. N230D de control que no habían sido bombardeadas con los plásmidos de transformación.

Ejemplo 5

El ejemplo a continuación describe los análisis de transferencia southern de las células de *Schizochytrium* transformadas.

Se llevaron a cabo transferencias de hibridación southern utilizando ADN aislado a partir de células de *Schizochytrium* N230D parental y varios transformantes putativos con el fin de confirmar la presencia de secuencias de ADN de vector de transformación dentro de las células transformadas; se llevó a cabo transferencia southern utilizando técnicas conocidas por el experto en la materia (ver, p.ej., Sambrook et al., Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, 1989). Se aisló ADN mediante la utilización de un kit de purificación de ADN "QIAamp" (Qiagen Inc., Valencia, CA) digerido con diversos enzimas de restricción, se separaron mediante electroforesis a través de geles de agarosa (0,8%-1,2% p/v) y después se transfirieron a membranas de nilón mediante transferencia capilar alcalina.

La detección del ADN del vector en las células transformadas con pTUBZEO11-2 se llevó a cabo mediante la utilización del sistema a base de DIG "Genius" (Boehringer Mannheim Biochemicals GmbH, Alemania), utilizando como sonda de hibridación, un fragmento de gen *ble* marcado con DIG de 346 pb generado mediante PCR con los cebadores DM20 (SEC ID nº 25) y DM21 (SEC ID nº 26) y una mezcla de nucleótidos que incluía DIG-11-UTP. La prehibridación de la membrana se llevó a cabo a 68°C durante 1 h en el tampón de hibridación suministrado en el kit Genius. La hibridación se llevó a cabo a 68°C durante 18 h en tampón de hibridación que contenía la sonda del gen *ble* que había sido desnaturalizada por calor durante 5 min a 94°C. A continuación, las membranas se lavaron dos veces durante 5 min con 50 ml de 2X SSC/SDS al 0,1% y dos veces durante 15 min en 50 ml de 0,1XSSC/SDS al 0,1%. La detección quimioluminiscente del ADN hibridante se llevó a cabo tal como se indica en las instrucciones del kit Genius.

El ADN procedente de las células de *Schizochytrium* N230D no transformadas no se hibridó con la sonda de gen *ble*. A la inversa, el ADN procedente de las células transformadas se hibridó con la sonda, de la manera siguiente:

SphI: el ADN digerido con *SphI* procedente de células de *Schizochytrium* transformadas contenía un fragmento de ADN de ~1.100 pb que se hibridaba con la sonda génica *ble*; este fragmento, que también se observó en el ADN de pTUBZEO11-2 digerido con *SphI*, representa el casete de expresión del gen *ble* entero (que incluye el promotor del gen de tubulina y el terminador de SV40).

XhoI: para cada uno de los transformantes sometidos a ensayo, la digestión con *XhoI* del ADN resultó en fragmentos hibridantes de tamaño superior a 15-20 kpb. *XhoI* no corta dentro de pTUBZEO11-2 y, por lo tanto, estos resultados indican que pTUBZEO11-2 aparentemente no existe en forma de elemento extracromosómico en las células transformadas, sino que por el contrario se integra en el cromosoma de *Schizochytrium*.

NcoI o *HindIII*: ambas enzimas cortan una vez dentro de pTUBZEO11-2. La digestión del ADN transformante con cualquiera de estas enzimas típicamente conduce a un fragmento hibridante prominente que comigra con el vector pTUBZEO11-2 linearizado (es decir, ~3,8 kpb). Ello sugiere que el vector puede integrarse en el cromosoma en forma de repeticiones en tándem.

Ejemplo 6

El presente ejemplo demuestra la recombinación homóloga en *Schizochytrium*.

Los experimentos siguientes se llevaron a cabo para demostrar que puede producirse la recombinación homóloga en *Schizochytrium* entre las secuencias de ADN nativo endógenas y las secuencias de ADN homólogas presentes en las moléculas de ADN recombinante introducidas en las células. Este tipo de recombinación homóloga puede resultar muy beneficioso para producir cepas recombinantes con propiedades deseables. Por ejemplo, puede utilizarse la recombinación homóloga para inactivar genes endógenos mediante la inserción dirigida de secuencias genéticas foráneas. Además, puede utilizarse la recombinación homóloga para sustituir un gen endógeno o parte del mismo por una forma alterada del gen de manera que las células recombinantes muestran nuevas propiedades.

Se demostró que la recombinación homóloga se produce en las células de *Schizochytrium* transformadas con el plásmido pMON50202, que contiene una mutación en el gen *als* de *Schizochytrium*. Esta mutación introduce un sitio *BsgI* en la posición de pb 571 de la región codificante *als*. Existe un sitio *BsgI* natural en la posición de pb 1.324 de la región codificante *als*. Por lo tanto, pueden utilizarse transferencias southern de ADN de *Schizochytrium* digerido con *BsgI* para diferenciar el gen *als* nativo del gen *als* mutante recombinante. Para estos experimentos, se produjo una sonda de hibridación específica de *als* mediante PCR utilizando una mezcla de nucleótidos que incluía DIG-11-UTP (Boehringer Mannheim Biochemicals GmbH, Alemania), el cebador directo PGR28 (SEC ID nº 32), el cebador inverso PGR30 (SEC ID nº 33) y una pequeña cantidad de pMON50200 como el molde. La sonda de hibridación marcada con DIG de 323 pb resultante se denominó ALS1.

Se digirió ADN procedente de células de *Schizochytrium* N230D no recombinantes con *BsgI* y *AhdI* separadamente, se sometieron a electroforesis en gel de agarosa, se transfirieron a una membrana de nilón y después se sondearon con la sonda ALS1 utilizando procedimientos esencialmente iguales a los indicados en el Ejemplo 5. La sonda ALS1 marcó un fragmento de 1,76 kpb de ADN digerido con *BsgI* y un fragmento de 4,9 kpb de ADN digerido con *AhdI*.

Las transferencias southern de DHA digerido con *BshI* y *AhdI* procedente de diversas cepas recombinantes que habían sido transformadas con pMON50202 también se sondearon con la sonda ALS1. En algunos casos, el fragmento *BsgI* de 1,76 kpb no se encontraba presente y en su lugar se marcaba un fragmento de 0,75 kpb correspondiente al fragmento *BsgI* de 753 pb presente en pMON50202. Se marcó un fragmento *AgdI* de 4,9 kpb en estas cepas recombinantes, sin embargo, indicando que el gen *als* mutante recombinante se había recombinado con el gen *als* nativo mediante recombinación homóloga de doble entrecruzamiento.

También se observó que se producía recombinación homóloga de un solo entrecruzamiento en cepas recombinantes transformadas con pMON50202. En estos casos, se marcaron los fragmentos *BsgI* tanto de 1,76 kpb como de 0,75 kpb en las transferencias southern de ADN procedentes de las cepas recombinantes, pero el fragmento *AhdI* de 4,9 kpb se sustituyó por fragmentos marcados de mayor tamaño, indicando que el vector pMON50202 completo se había insertado en el gen *als* nativo, como una copia única o como repeticiones en tándem.

Se obtuvo evidencia adicional de recombinación homóloga en *Schizochytrium* mediante la introducción de moléculas de ADN recombinantes que contenían un gen *als* mutante truncado, de manera que el enzima ALS incompleto codificado por el gen truncado no era funcional. Este gen truncado se produjo mediante digestión de pMON50202 con *Clal* e *HindIII*, rindiendo un fragmento de 2,8 kpb, eliminando de esta manera los últimos 388 pb de la secuencia codificante *als* junto con la región de terminador de *als*. Este fragmento de 2,8 kpb se ligó en pBluescriptII KS+ (Stratagene Corp., La Jolla, CA) que había sido digerido con *Clal* e *HindIII*, rindiendo el plásmido pAR2. Sólo se esperaba que el plásmido pAR2 proporcionase resistencia a SMM en las células de *Schizochytrium* transformadas en el caso de que un gen *als* mutante funcional se restaurase en las cepas transformadas mediante recombinación homóloga entre el gen *als* nativo y el gen *als* mutante truncado presente en pAR2. Se introdujo este constructo en las células de *Schizochytrium* N230D mediante bombardeo de partículas y se aislaron las cepas resistentes a SMM tal como se indica en el Ejemplo 3. El análisis de transferencia western del ADN digerido con *BsgI* procedente de los

transformantes realizado tal como se ha indicado anteriormente en el presente ejemplo, indicó que la recombinación homóloga se había producido claramente en estas cepas; es decir, un fragmento *Basgl* de 1,76 kpb se había hibridado con la sonda ALS1 en las células no recombinante, pero que esto había sido sustituido por un fragmento hibridante de 0,75 kpb en las células que habían sido transformadas con pAR2.

Ejemplo 7

El presente ejemplo describe la utilización del vector de transformación pTUBZEO11-2 o pMON50202 para producir mediante cotransformación cepas que contiene moléculas de ADN foráneo adicionales que no están unidas a un gen marcador seleccionable.

Se llevó a cabo la cotransformación mediante introducción simultánea de pTUBZEO11-2 y un plásmido adicional que contenía cualquiera de entre varios genes. Los plásmidos se coprecipitaron sobre las partículas de oro tal como se indica en el Ejemplo 3 utilizando 2,5 µg de cada plásmido. Tras el bombardeo de las células diana con las partículas de oro recubiertas con plásmido, se seleccionaron las cepas recombinantes en placas de agar que contenían Zeocin™ tal como se indica en el Ejemplo 1. La presencia del segundo plásmido no seleccionado seguidamente se confirmó mediante análisis de PCR o mediante hibridación de transferencia southern. Típicamente se alcanzan frecuencias de cotransformación muy elevadas (p.ej., 50% a 90%). Por ejemplo, el plásmido pTR202, que contiene el gen *fat-1* de *Caenorhabditis elegans* (Spychalla et al., Proc. Natl. Acad. Sci. USA 94:1142-1147, 1997) unido al promotor y terminador del gen de tubulina de *Schizochytrium* se introdujo mediante el método proporcionado en el presente ejemplo y aproximadamente 68% de las cepas resistentes a Zeocin™ resultantes se mostró mediante PCR que contenían el gen *fat-1* (ver la Tabla 1). Se observaron resultados similares al cointroducir pMON50202 y un plásmido adicional, seguido de la selección de las células transformadas sobre un medio sólido que contenía SMM. Este método de cotransformación puede utilizarse para introducir cualquier ADN foráneo deseado.

Tabla 1. Eficiencias de cotransformación utilizando el plásmido marcador seleccionable pTubZeo11-2 y plásmidos que contiene diversos genes *fad*. Se cribaron transformantes resistentes a Zeocin^R para secuencias de ADN *fad* mediante PCR.

Gen <i>fad</i> introducido	Nº que contiene genes <i>fad</i> Nº de cepas Zeocin ^R sometidas a ensayo	Eficiencia de cotransformación
syn <i>fat1</i>	17/25	68%
nat <i>fat1</i>	24/55	96%
mut <i>fat1</i>	21/25	84%
desB	20/25	80%

Los sistemas de transformación descritos en los presentes ejemplos representan un avance significativo en la capacidad de manipular genéticamente *Schizochytrium*, que es el organismo más productivo que se conoce para la producción fermentativa de compuestos a base de lípidos. La disponibilidad de dos sistemas de transformación independientes, junto con las elevadas eficiencias de cotransformación que se produce, deberían permitir la acumulación de múltiples características en las cepas manipuladas. Además, la aparente presencia de recombinación homóloga en este microalga debería permitir desarrollar procedimientos de inactivación génica con el fin de identificar las funciones de genes desconocidos y de eliminar características no deseables en las cepas de producción. Los presentes inventores están utilizando actualmente estos sistemas para alterar el metabolismo de los ácidos grasos en *Schizochytrium* y están explorando posibilidades de utilizar esta especie y microalgas relacionadas (p.ej., *Thraustochytrium*) para la producción de carotenoides, esteroides y otros compuestos lipídicos.

Aunque se han descrito en detalle diversas realizaciones de la presente invención, resulta evidente que el experto en la materia concebirá modificaciones y adaptaciones de dichas realizaciones. Sin embargo, debe entenderse expresamente que tales modificaciones y adaptaciones se encuentran comprendidas dentro del alcance de la presente invención, tal como se proporciona en las reivindicaciones, posteriormente.

LISTADO DE SECUENCIAS

<110> Roessler, Paul
Matthews, T. Dave
Ramseier, Tom
Metz, James

<120> Producto y procedimiento para la transformación de microorganismos Traustoquitriales

<130> 2997-23-PCT

<150> 60/284,116

<151> 2001-04-16

<160> 35

<170> Patent In versión 3.1

5 <210> 1
 <211> 551
 <212> ADN
 <213> Schizochytrium sp.

10 <220>
 <221> misc_feature
 <222> (520)..(520)
 <223> n = a, c, g, o t

15 <220>
 <221> misc_feature
 <222> (541)..(541)
 <223> n = a, c, g, o t

20 <400> 1
 gtcgtgccta acaacacgcc gttctacccc gccttcttcg cgcaccttcg cgtccaagca 60
 tccttcaagt ttatctctct agttcaactt caagaagaac aacaccacca acaagatgcg 120
 tgaggtcatc tocatccaca tcggccaggc cgggtgtcag gtcggtaacg cctgctggga 180
 gctctactgc ctcgagcatg gcatccagcc ggacggccag atgcctcgg acaagaccat 240
 tggcggcggc gatgatgcct tcaaacacctt cttctccgag actgggcgcy gcaagcacgt 300
 gccccgcgcc gtgctctgct atctcgagcc caccgtctgt gacgaggctc gcaccggcac 360
 ctaccgcgct ctttaccacc ccgagcagat catcaccggc aaggaggacy ctgccaacaa 420
 ctacgctcgt ggccaactaca ccacccgcaa ggagatcgtc gacctcgtcc tcgaccgcat 480
 ccgcaagctc gccgacaact gcactggctt tcagggtctn ctctgcttca acgcccgtcg 540
 nngtggtacc g 551

25 <210> 2
 <211> 27
 <212>ADN
 <213>Schizochytrium sp.

30 <400> 2
 gcgccagtct cggagaagaa ggtgttg 27

<210> 3
 <211> 27
 <212>ADN
 <213>Schizochytrium sp.

35 <400> 3
 agtcccagc aggcgttacc gacctga 27

40 <210> 4
 <211> 725
 <212> ADN
 <213> Schizochytrium sp.

45 <400> 4
 gagacgtgct togcaagacc gctgtgctcg cgcgcacgc tctgtgtgtt acattaattt 60
 tttttagat gaagtttctc tattctctcg aaattctgta gaatgttata gtctcttcac 120
 tcccgtgatt ggagaggatt cttgcttgtt cctcccggc cgggtagcgc ttggagcaac 180
 gcttgagcgc gcgctcgaaa gcggacggcg caacgagcgc tttcacgccc cgctgtocaa 240
 gtcccatttt tctccttacc ccacggccgt tgcagccaa ttttaggccc cccactgacc 300
 gaggtctgct gataatccac ttttccattg atcttccagg tttcgttaac tcatgccact 360
 gagcaaaact tcggcttttc ctaacaaaag ctctcctcac aaagcatggt gcggcaacgg 420
 acgtgtcctc atactccact gccacacaag gtcgataaac taagctcctc acaaatagag 480
 gagaattcca ctgacaactg aaaacaatgt atgagagacg atcaccactg gagcggcgcg 540
 gcggttgggc gcggaggctc gcagcaaaaa caagcgactc gccgagcaaa cccgaatcag 600
 ccttcagacg gtcgtgccta acaacacgcc gttctacccc gccttcttcg cgcaccttcg 660
 cgtccaagca tccttcaagt ttatctctct agttcaactt caagaagaac aacaccacca 720
 acaag 725

ES 2 698 473 T3

<210> 5
 <211> 24
 <212> ADN
 <213> Schizochytrium sp.
 5
 <400> 5
 cacccatggt gttggtggtg ttgt 24
 <210> 6
 <211> 24
 <212> ADN
 <213> Schizochytrium sp.
 10
 <400> 6
 aaactcgaga cgtgcttcgc aaga 24
 <210> 7
 <211> 4646
 <212> ADN
 <213> Schizochytrium sp.
 20
 <400> 7
 aaactcgaga cgtgcttcgc aagaccgctg tgctcgcgcc gcaogctctg tgtgttacat 60
 taattttttt gtagatgaag tttctctatt ctctcgaaat tctgtagaat gttatagtct 120
 ottcactccc gtgattggag aggattcttg ctgtttccct cccgcccggg tagcgcttgg 180
 agcaacgctt gagcgcgcgc tcgaaagcgg acggcgcaac gagccgtttc acgcccgcgt 240
 gtccaagtcc cttttttctc cttaccccat ggcggttgca tgccaatttt aggcccccga 300
 ctgaccgagg tctgtcgata atccactttt ccattgatct tccagggtttc gttactcat 360
 gccactgagc aaaacttggg tctttcctaa caaaagctct cctcacaag catggcgcg 420
 caacggagct gtcctcatac tccactgcca cacaaggtcg ataaactaag ctctcaca 480
 atagaggaga attcactga caactgaaa caatgtatga gagacgatca ccactggagc 540
 ggcgcggcgg ttggcgcggg aggtcggcag caaaaacaag cgactcgccg agcaaacccg 600
 aatcagcctt cagacggtcg tgcctaaca caogccgttc taccocgct tcttcgcgcc 660
 ccttcgctc caagcatcct tcaagtttat ctctctagt caacttcaag aagaacaaca 720
 ccaccaacac catgggtgaa gggcgaattc tgcaatatac catcacactg gcggccgctc 780
 gagcatgcat ctagagggcc caattcgccc tatagttagt cgtattaca ttactggcc 840
 gtcgttttac aacgtcgtga ctgggaaaac cctggcgtta cccaactta tcgccttgca 900
 gcacatccc ctttcgccag ctggcgtaat agcgaagagg cccgcaccga tcgccttcc 960
 caacagttgc gcagcctgaa tggcgaatgg gacgcgcct gtagcggcgc attaagcgcg 1020

ES 2 698 473 T3

gccccgtgtg tggttacgcg cagcgtgacc gctacacttg ccagcgcctt agcgcocgct 1080
cctttcgctt tccctttagg gttccgattt acggttcgocg gctttccocg tcaagctcta 1140
aatcgggggg tccctttagg gttccgattt agagctttac ggcacctcga ccgcaaaaaa 1200
cttgatttgg gtgatggttc acgtagtggg ccatcgocct gatagacggt ttttcgocct 1260
ttgacgttgg agtccacggt ctttaatagt ggactcttgt tccaaactgg aacaacactc 1320
aacocctatcg cggctctattc ttttgattta taagggattt tgccgatttc ggctatttgg 1380
ttaaaaaatg agctgattta acaaattcag ggcgcaaggg ctgctaaagg aaccggaaca 1440
cgtagaaaagc cagtccgcag aaacgggtgct gaccccgat gaatgtcagc tactgggcta 1500
tctggacaag ggaaaaacgca agcgcacaaga gaaagcaggt agcttgagc gggtttacat 1560
ggcgatagct agactgggoc gttttatgga cagcaagcga accggaattg ccagctgggg 1620
cgccctctgg taaggttggg aagcctgca aagtaaaactg gatggctttc ttgocgccaa 1680
ggatctgatg gcgcagggga tcaagatctg atcaagagac aggatgagga tcgtttcgca 1740
tgattgaaca agatggattg cacgcagggt ctccggocgc ttgggtggag aggtatttgc 1800
gctatgactg ggcaacaacag acaatcggct gctctgatgc cgccgtgttc cgctgtcag 1860
cgagggggcg cccggttctt tttgtcaaga ccgacctgtc cgtgocctg aatgaaactg 1920
aggacgaggg agcgcgggcta tcgtggctgg ccacgacggg cgttccttgc gcagctgtgc 1980
tcgacgttgt cactgaagcg ggaagggact ggctgtctatt gggcgaagtg ccggggcag 2040
atctcctgtc atctcgcctt gctcctgccc agaaaagtatc catcatggct gatgaaactg 2100
ggcggctgca tacgcttgat cccgtacctt gccattcga ccaccaagcg aaacatcgca 2160
tcgagcagagc acgtactcgg atggaagccc gtcttgcga tcaggatgat ctggacgaa 2220
agcatcaggg gctcgcgcca gccgaactgt tcgccaggct caaggcgcgc atgcccagc 2280
gcgaggatct cgtcgtgatt catggcgatg cctgcttgc gaatatcat gtggaaaatg 2340
gocgcttttc tggattcaac gactgtggcc ggctgggtgt ggccgaccgc tatcaggaca 2400
tagcgttggg taccctgatg attgctgaag agcttggcgg cgaatggct gaccgcttcc 2460
tcgtgcttta cggatcggc gctccgattt cgcagcgcac cgccttctat cgccttctc 2520
acgagttctt ctgaattgaa aaaggaagag tatgagtatt caacatttcc gtgtcgccct 2580
tattcccttt tttgocgcat tttgoccttcc tgtttttgc caaccagaaa cgtggtgaa 2640
agtaaaagat gctgaagatc agttgggtgc acgagtggtt tacatcgaa tgatctcaa 2700
cagcggtaag atccttgaga gtttgcgcc cgaagaacgt tttccaatga tgagcactt 2760
taaagttctg ctatgtcata cactattatc ccgatttgc gccgggcaag agcaactcg 2820
tcgcccggcg cggatattctc agaatgactt ggttgagtac tcaccagtca cagaaaagca 2880
tcttaccggt ggcatgacag taagagaatt atgcagtgc gccataacca tgagtataa 2940
cactgcccgc aacttacttc tgacaacgat cggaggaccg aaggagctaa ccgcttttt 3000
gcacaacatg ggggatcatg taactcgcct tgatcgttgg gaaccggagc tgaatgaagc 3060
cataccaaac gacgagagt acaccacgat gcctgtagca atgccaaca cgttgcgcaa 3120
actattaact ggcgaactac ttactctagc ttcccgcaa caattaatag actggatgga 3180
ggcggataaa gttgcaggac cacttctgoc ctccggcctt ccggtggct ggtttattgc 3240
tgataaatct ggagccgggt agcgtgggtc tcgcccgtatc attgcagcac tggggccaga 3300
tggttaagccc tcccgatcag tagttatcta cagcagggg agtcaggcaa ctatggatg 3360
acgaaataga cagatcgtg agataggtgc ctcaactgatt aagcattggt aactgtcaga 3420
ccaagtttac tcatatatac tttagattga tttaaaactt catttttaat ttaaaaggat 3480
ctagggtgaa atcctttttg ataatctcat gaccaaaatc cottaacgtg agttttcgtt 3540
ccactgagcg tcagaccocg tagaaaagat caaaggatct tcttgagatc cttttttct 3600
gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcgggtg tttgtttgc 3660
ggatcaagag ctaccaactc tttttccgaa ggtaactggc tcagcagag cgcagatacc 3720
aaatactgtc ctctagtgt agccgtagtt aggecaccac ttcaagaact ctgtagcacc 3780
gctacatac ctgcctctgc taatcctggt accagtggct gctgocagtg gcgataagtc 3840
gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 3900
aacggggggt tcgtgcacac agcccagctt ggagcgaacg acctacacc aactgagata 3960
cctacagcgt gagcattgag aaagcggcac gcttcccga gggagaagg cggacaggta 4020
tcocgtaagc ggcaggtcgc gaacaggaga gcgcacgagg gagctccag ggggaaacgc 4080
ctggatctct tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 4140
atgctcgtca gggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 4200
cotggccttt tgctggcctt ttgctcacat gttcttctc gogttatccc ctgattctgt 4260
ggataaccgt attaccgct ttgagtgagc tgataccgct cggccagacc gaacgaccga 4320
gcgcagcgag tcagtgagcg aggaagcggga agagcggcca ataccgaaac cgcctctccc 4380
cgcgcgttgg ccgattcatt aatgcagctg gcacgacagg tttcccgact ggaaagcggg 4440
cagttagcgc aacgcaatta atgtgagtta gctcactcat taggcacccc aggtttaca 4500
ctttatgctt ccggctcgta tgttgtgtg aatttgtgag ggataacaat ttcacacagg 4560

aaacagctat gaccatgatt acgccaagct tggtaaccgag ctccgatcca ctagtaacgg 4620
ccgocagctg gctggaattc gcocctt 4646

5

- <210> 8
- <211> 3664
- <212> ADN
- <213> Schizochytrium sp.

ES 2 698 473 T3

<400> 8

ccatggcctg	tgcattgcaa	tttagggccc	ccactgacc	gaggtctgtc	gataatccac	60
ttttccattg	atcttccagg	tttogttaac	tcatgccact	gagcaaaaact	tcggtctttc	120
ctaacaaaag	ctctctcac	aaagcatggc	gogggcaacg	acgtgtcctc	atactccact	180
gccacacaag	gtcgataaac	taagctcctc	acaaaatagag	gagaattcca	ctgacaactg	240
aaaacaatgt	atgagagacg	atcaccactg	gagcggcgcg	gcggttgggc	gcgagggtcg	300
gcagcaaaaa	caagcgaactc	gocgagcaaa	ccogaatcag	ccttcagacg	gtcgtgccta	360
acaacacggc	gttctacccc	gccttcttog	cgccccttcg	cgccaagca	tccttcaagt	420
ttatctctct	agttcaactt	caagaagaac	aacaccacca	acaccatggc	caagttgacc	480
agtgcctgtc	cggtgctcac	cgcgcgagac	gtcgccggag	cggtcgagtt	ctggaccgac	540
cggtcgggtg	tctcccggga	cttcgtggag	gaogacttcg	cggtgtggtg	ccgggacgac	600
gtgaccctgt	tcattcagcgc	ggtccaggac	caggtggtgc	cgacaacac	octggcctgg	660
gtgtgggtgc	goggcctgga	cgagctgtac	gocgagtggt	cgagggtcgt	gtccacgaac	720
ttccgggacg	cctccggggc	ggcoactgac	gagatcggcg	agcagccgtg	ggggcgggag	780
ttcgcccctg	gagaccgggc	cgcaactgc	gtgcaacttcg	tggccgagga	gcaggactga	840
cacgtgctac	gcagatttoga	ttccaccgcc	gccttctatg	aaaggttggg	cttcggaatc	900
gttttccggg	acgcccggctg	gatgatcctc	cagcgggggg	atctcatgct	ggagttcttc	960
gcccacccca	acttgtttat	tgcagcttat	aatggttaca	aataaagcaa	tagcatcaca	1020
aatttcacaa	ataaagcatt	tttttcactg	cattctagtt	gtggtttgtc	caaaactcatc	1080
aatgtatctt	atcatgtctg	aattcccggg	gatcctctag	agtcgacctg	caggcatgca	1140
agcttggcac	tggccctgtg	tttacaacgt	cgtagctggg	aaaaccctgg	cgttacccaa	1200
cttaaatcgcc	ttgcagcaca	tcccccttc	gccagctggc	gtaatagcga	agaggcccgc	1260
accgatcgcc	cttcccaca	gttgccgagc	ctgaatggcg	aatggcgcc	gatcggtgat	1320
tttctcctta	cgcatctgtg	cggtatttca	caccgcata	atgggtgact	ctcagtacaa	1380
tctgctctga	tgccgcatag	ttaagccagc	cccagacacc	gccaacaccc	gctgacggga	1440
cctgacgggg	ttgctctgctc	ccggcatccg	cttacagaca	agctgtgacc	gtctccggga	1500
gctgcatgtg	tcagaggttt	tcaccgctcat	caccgaaacg	cgcgagacga	aaggccctcg	1560
tgatacgcct	atTTTTatag	gttaatgtca	tgataaata	ggtttcttag	acgtcaggtg	1620
gcacttttctg	gggaaatgtg	cgcggaacc	ctatttgttt	atTTTTctaa	atacattcaa	1680
atatgtatcc	gctcatgaga	caataaccct	gataaatgct	tcaataatat	tgaaaaagga	1740
agagtatgag	tattcaaat	ttccgtgtcg	cccttattcc	ctTTTTtgcg	gcattttgcc	1800
ttcctgtttt	tgctcaccca	gaaacgctgg	tgaaagttaa	agatgctgaa	gatcagttgg	1860
gtccacgagt	gggttacatc	gaactggatc	tcaacagcgg	taagatcctt	gagagttttc	1920
gcccogaaga	acgttttcca	atgatgagca	cttttaaagt	tctgctatgt	ggcgggttat	1980
tatcccgtat	tgaccgcccgg	caagagcaac	toggctggcg	catacactat	tctcagaatg	2040
acttggttga	gtactcacca	gtcacagaaa	agcatcttac	ggatggcatg	acagtaagag	2100
aattatgcag	tgctgccata	accatgagtg	ataacactgc	ggccaactta	cttctgacaa	2160
cgatcggagg	accgaaggag	ctaaccgctt	ttttgcacaa	catgggggat	catgtaactc	2220
gccttgatgc	ttgggaaccg	gagctgaatg	aagccatacc	aaacgacgag	cgtgacacca	2280
cgatgcctgt	agcaatggca	acaacgttgc	gcaaacattt	aaactggcga	ctacttactc	2340
tagcttcccg	gcaacaatta	atagactgga	tggaggcgga	taaagttgca	ggaccacttc	2400
tgcgctcggc	ccttccggct	ggctggttta	ttgctgataa	atctggagcc	ggtgagcgtg	2460
ggtctcggcg	tatcatttga	gcactggggc	cagatggtaa	gcctcccctg	atcgtagtta	2520
tctacacgac	ggggagtcag	gcaactatgg	atgaacgaaa	tagacagatc	gctgagatag	2580
gtgcctcact	gattaagcat	tggttaactgt	cagaccaagt	ttactcatal	atactttaaga	2640
ttgatttaaa	acttcaatctt	taatttaaaa	ggatctaggt	gaagatcctt	tttgataatc	2700
tcatgaccaa	aatcccttaa	cgtaggtttt	cgttccactg	agcgtcagac	ccogtagaaa	2760
agatcaaaag	atcttcttga	gatccttttt	ttctgcccgt	aatctgctgc	ttgcaaaaa	2820
aaaaaccacc	gctaccagcg	gtggtttgtt	tgccggatca	agagctacca	actctttttc	2880
cgaaaggtaac	tggcttcagc	agagcgcaga	taccaaatac	tgtccttcta	gtgtagccgt	2940
agttaggcca	ccacttcaag	aactctgtag	caccgcctac	atacctcgct	ctgctaattc	3000

tgttaccagt	ggctgctgcc	agtggcgata	agtcgtgtct	taccgggttg	gactcaagac	3060
gatagttacc	ggataaggcg	cagcggtcgg	gctgaacggg	gggttcgtgc	acacagccca	3120
gcttgagagc	aacgacctac	accgaactga	gatacctaca	gogtgagcat	tgagaagcg	3180
ccacgccttc	cgaagggaga	aaggcggaca	ggtatccggt	aagcggcagg	gtcggaacag	3240
gagagcgcac	gagggagcct	ccagggggaa	acgocctgta	tctttatagt	cctgtcgggt	3300
ttcgccaact	ctgacttgag	cgctgatttt	tgtgatgctc	gtcagggggg	cggagcctat	3360
ggaaaaacgc	cagcaacgcg	gcctttttac	ggttcctggc	cttttgctgg	ccttttgctc	3420
acatgtgtgc	tggggccagc	cgccagatc	tgagctcggc	gocgcgatat	cgctagctcg	3480
aggcaggcag	aagtatgcaa	agcatgcac	tcaattagtc	agcaaccagg	tgtggaaagt	3540
ccccaggctc	cccagcaggc	agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	3600
tagtcccggc	cctaactccg	cccattccgc	ccctaactcc	gcccagttcc	gcccattctc	3660
cgcc						3664

5

<210> 9

<211> 3808

<212> ADN

<213> Secuencia artificial

10

<220>

<223> Schizochytrium sp. y Streptoalloteichus hindustanus

ES 2 698 473 T3

<220>
 <221> promotor
 <222> (441)..(894)
 <223>

5

<220>
 <221> CDS
 <222> (895)..(1269)
 <223>

10

<220>
 <221> terminador
 <222> (1270)..(1524)
 <223>

15

<400> 9

tgcgcggttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	60
cagcttgtct	gtaagcggat	gccggggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcagggcgc	240
attcgccatt	caggctgctg	aactgttggg	aagggcgatc	ggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaagggg	ggatgtgctg	caagggcatt	aagttgggta	acgccagggg	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgaatt	cgagctcgg	accgggggat	420
cctctagagt	cgacctgcag	gcatgccaat	tttaggcccc	ccactgaccg	aggtctgtcg	480
ataatccact	tttccattga	ttttccaggt	ttcgtaact	catgccactg	agcaaaactt	540
cggtctttcc	taacaaaagc	tctcctcaca	aagcatggcg	cggcaacgga	cgtgtcctca	600
tactccactg	ccacacaagg	tcgataaact	aagctcctca	caaatagagg	agaattccac	660
tgacaactga	aaacaatgta	tgagagacga	tcaccactgg	agcggcgcg	cggttggggc	720
cggaggtcgg	cagcaaaaac	aagcgactcg	cogagcaaac	cogaatcagc	cttcagacgg	780
tcgtgcctaa	caacacgocg	ttctaccccg	ccttcttcgc	gccccttcgc	gtccaagcat	840
ccttcaagtt	tatctctcta	gttcaacttc	aagaagaaca	acaccaccaa	cacc atg	897
					Met	
					1	
gcc aag ttg acc agt gcc gtt ccg gtg ctc acc gcg cgc gac gtc gcc						945
Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val Ala						
	5		10		15	
gga gcg gtc gag ttc tgg acc gac cgg ctc ggg ttc tcc cgg gac ttc						993
Gly Ala Val Glu Phe Trp Thr Asp Arg Leu Gly Phe Ser Arg Asp Phe						
	20		25		30	

ES 2 698 473 T3

gtg gag gac gac ttc gcc ggt gtg gtc cgg gac gac gtg acc ctg ttc 1041
 Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Asp Val Thr Leu Phe
 35 40 45
 atc agc gog gtc cag gac cag gtg gtg ccg gac aac acc ctg gcc tgg 1089
 Ile Ser Ala Val Gln Asp Gln Val Val Pro Asp Asn Thr Leu Ala Trp
 50 55 60 65
 gtg tgg gtg cgc ggc ctg gac gag ctg tac gcc gag tgg tcg gag gtc 1137
 Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu Val
 70 75 80
 gtg tcc acg aac ttc cgg gac gcc tcc ggg ccg gcc atg acc gag atc 1185
 Val Ser Thr Asn Phe Arg Asp Ala Ser Gly Pro Ala Met Thr Glu Ile
 85 90 95
 ggc gag cag ccg tgg ggg cgg gag ttc gcc ctg cgc gac ccg gcc ggc 1233
 Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala Gly
 100 105 110
 aac tgc gtg cac ttc gtg gcc gag gag cag gac tga cacgtgctac 1279
 Asn Cys Val His Phe Val Ala Glu Glu Gln Asp
 115 120
 gagatttcga ttccacggcc gccttctatg aaagggtggg ctctcggaatc gttttccggg 1339
 acgcccggctg gatgatcctc cagcgcgggg atctcatgct ggagttcttc gcccaaccca 1399
 acttgtttat tgcagcttat aatggttaca aataaaagcaa tagcatcaca aatttcacaa 1459
 ataaagcatt tttttcactg cattctagtt gtgggttgtc caaactcatc aatgtatctt 1519
 atcatgtctg aattccgggg gatcctctag agtcgacctg caggcatgca agcttggcgt 1579
 aatcatggtc atagctgttt cctgtgtgaa attgttatcc gtcacaatt ccacacaaca 1639
 tacgagccgg aagcataaaag tgtaaagcct ggggtgccta atgagtgagc taactcaat 1699
 taattgctgt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt 1759
 aatgaatcgg ccaacgcgcg ggggagggcg gtttgcgtat tggggcgtct tccgcttctc 1819
 cgtcactga ctgcctgccc tcggtcgttc ggctgcggcg agcggatca gctcactcaa 1879
 agggcgtaac acggttatcc acagaatcag gggataacgc aggaaagaac atgtgagcaa 1939
 aaggccagca aaaggccagg aaccgtaaaa aggcgcgctt gctggcgctt tcccataggc 1999
 tccgcccccc tgaagagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaccoga 2059
 caggactata aagataccag gcgcttcccc ctggaagctc cctcgtgccc tctcctgttc 2119
 cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttogggaagc gtggcgctt 2179
 ctcaatgctc accgtgtagg tatctcagtt cgggtgtaggt cgctcgtccc aagctggcgt 2239
 gtgtgcacga acccccogtt cagccogacc gctgcccctt atccggtaac tatcgtcttg 2299
 agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta 2359
 gcagagcgag gtatgtaggc ggtgtacag agttcttgaa gtgggtggcct aactacggct 2419
 aactagaag gacagtattt ggtatctgag ctctgctgaa gccagttacc ttoggaaaaa 2479
 gatttggtag ctcttgatcc ggcaaaacaaa ccaccgctgg tagcgtggt tttttgttt 2539
 gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta 2599
 cggggtctga cgctcagtg aacgaaaact cacgttaagg gattttggtc atgagattat 2659
 caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaa tcaatctaaa 2719
 gtatatatga gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacotatct 2779
 cagcgatctg tctatctgt tcatccatag ttgcctgact ccccgctgtg tagataacta 2839
 cgatacggga gggcttacca tctggcccca gtgctgcaat gatacccgga gaccacgct 2899
 caccggctcc agatttatca gcaataaacc agccagccgg aagggcccg cgcagaagtg 2959
 ctctgcaac tttatccgcc tccatccagt ctattaattg ttgcccggaa gctagagtaa 3019
 gtatgtcgcc agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt 3079
 cagcctcgtc gtttggtatg gcttcattca gctccggctc ccaacgatca aggcagatta 3139
 catgatcccc catgtttgtc aaaaaagcgg tttagctcctt cggctcctcc atcgtttgtc 3199
 gaagtaagtt ggccgcagtg ttatcactca tgggtatggc agcactgcat aattctctta 3259
 ctgtcatgcc atccgtaaga tgcttttctg tgactgggta gtactcaacc aagtcattct 3319
 gagaatagtg tatgcccgca ccgagttgct cttgcccggc gtcaatacgg gataatacgg 3379
 cgccacatag cagaacttta aaagtgcctc tcaattggaaa acgttcttcg gggcgaaaaac 3439
 tctcaaggat cttaccgctg ttgagatcca gttcogatga acccactcgt gcacccaact 3499
 gatcttcagc atcttttact ttcaccagcg tttctggggt agcaaaaaa ggaagcaca 3559
 atgcccgaaa aaagggaata agggcgacac ggaaatgtt aatactcata ctcttcttt 3619
 ttcaatatta ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat 3679
 gtatttagaa aaataacaa ataggggttc cgcgcacatt tccccgaaa gtgccacctg 3739

acgtotaaga aaccattatt atcatgacat taacctataa aaataggcgt atcacgagggc 3799
 cctttogtc 3808

5 <210> 10
 <211> 124
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> Schizochytrium sp. y Streptoalloteichus hindustanus

ES 2 698 473 T3

<400> 10

```

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

```

<210> 11

<211> 1416

<212> ADN

<213> Schizochytrium sp.

<400> 11

```

gcaaaaggct gagcttttcc acaaggagcg cattggcgct cctggcacg cgcacttcaa 60
gctcattgcc gagatgatca accgtgogga gcgaccgctc atctatgctg gccaggggtg 120
catgcagagc ccgttgaatg gcccggtgt gctcaaggag ttcgcggaga aggccaacat 180
tcccgtagcc accaccatgc aggttctcgg cggctttgac gagcgtagtc ccctctccct 240
caagatgctc ggcattgcaag gctctgccta cgcacaactac tcgatgcaga acgccgatct 300
tatcctggcg ctccgtgccc gctttgatga tcgtgtgacg gcccgcttg acgcctttgc 360
tccggaggct cgcggtgccc agcgcgaggg ccgcggtggc atcgttcact ttgagatttc 420
ccccaaagaac ctccacaagg tcgtccagcc caccgtcgcg gtctcggcg acgtgggtcga 480
gaacctcgcc aacgtcacgc cccacgtgca gcgccaggag cgcgagccgt ggtttgcgca 540
gatcgccgat tggaaaggaga agcaaccitt tctgctcgag tctggtgatt cggacgacaa 600
ggtttctcaag ccgcagcagg tcctcacgga gcttaacaag cagattctcg agattcagga 660
gaaggagccc gaccaggagg tctacatcac cacgggctgc ggaagccacc agatgcaggc 720
agcgcagttc cttacctgga ccaagccgcg ccagtggatc tcctcgggtg gcgccggcac 780
tatgggttac ggccttccct cggccattgg cgccaagatt gcccaagccc atgctattgt 840
tatgacatc gatggtgatg cttcttattc gatgaccggt atggaattga tcacagcagc 900
cgaattcaag gttggcgtga agattcttct tttgcagaac aactttcagg gcatgggtcaa 960
gaactggcag gatctctttt acgacaagcg ctactcggcg accgccatgt tcaaccgcg 1020
cttcgacaag gtcgccgatg cgatgcgtgc caagggtctc tactgcgcga aacagtccga 1080

```

```

gctcaaggac aagatcaagg agtttctcga gtacgatgag ggtcccgtcc tcctcagagt 1140
tttcgtggac aaggacagc tcgtcttggc catggtcccc gctggctttc cgctccacga 1200
gatggtcctc gagcctccta agcccaagga cgcctaagtt cttttttcca tggcgggoga 1260
gcgagcagc gcgcgagcgc gcaagtgcgc aagcgccttg ccttgctttg cttcgcttcg 1320
ctttgctttg cttcacacaa cctaagtatg aattcaagtt ttcttgcttg tcggcgaaaa 1380
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa

```

<210> 12

<211> 20

<212> ADN

<213> Schizochytrium sp.

<400> 12

ggatctcttt tacgacaagc 20

<210> 13

<211> 18

<212> ADN

<213> Schizochytrium sp.

ES 2 698 473 T3

<400> 13
 gggtgtgtga agcaaagc 18

5
 <210> 14
 <211> 7847
 <212>ADN
 <213>Schizochytrium sp.

10
 <220>
 <221> prim_transcript
 <222> (1)..(1259)
 <223>

15
 <220>
 <221> CDS
 <222> (1260)..(3314)
 <223>

20
 <220>
 <221> prim_transcript
 <222> (3315)..(4887)
 <223>

<400> 14
 ttgtcgacag caacttgcaa gttatacgcg accaccaggc aatctcagca cgcccagcga 60
 gcacggagct tgccaagagg gtttacaagt cgtcgttcat tcgctctcaa gctttgcctc 120
 aacgcaacta ggcccaggcc tactttcact gtgtccttgc ttgcctttca caccgaccga 180
 gtgtgcacaa ccgtgttttg cacaaagcgc aagatgctca ctcgactgtg aagaaaggtt 240
 gcgcgcaagc gactgcgact gcgaggatga ggatgactgg cagcctgttc aaaaactgaa 300
 aatccgcgat gggtcagtg ctttcgcgca tgacgcctgc gagagacaag ttaactcgtg 360
 tcaactggcat gtcctagcat ctttacgcga gcaaaattca atcgctttat ttttccagtt 420
 tcgtaacctt ctcgcaaccg cgaatcgccg tttcagcctg actaatctgc agctgcgtgg 480
 cactgtcagt cagtcagtca gtcgtgcgcg ctgttccagc accgaggtcg cgcgtcgccg 540
 cgcctggacc gctgctgcta ctgctagtgg cacggcaggt aggagcttgt tgccggaaca 600
 ccagcagccg ccagtcgacg ccagccaggg gaaagtccgg cgtcgaaggg agaggaaggc 660
 ggcgtgtgca aactaacggt gaccactggc gcccgccgac acgagcagga agcagggcagc 720
 tgcaagagcg agcgcgcaag tgcagaatgc gcgaaagatc cacttgccgc cggcgggccc 780
 gcacttgccg gcgcggcgcg gaacagtgcg gaaaggagcg gtgcagacgg cgcgcagtga 840
 cagtggggcg aaagccgcgc agtaagcagc ggcgggggaa gggtatacga gtgccgcggg 900

25

ES 2 698 473 T3

ccgcccacaca cagaagtata cgcgggccga agtggggcgt cgcgcgcggg aagtgcggaa 960
 tggcgggcaa ggaaggagg agacggaaag agggcgggaa agagagagag agagagttaa 1020
 aaaagaaaaga aaaaagaaa gaaagaaaaga aagctcggag ccacgccgag gggagagaga 1080
 gaaatgaaag cacggcacgg caaagcaaag caaagcagac ccagccagac ccagccgagg 1140
 gaggagcgcg cgcaggaccc gcgcggcgag cgagcagaca cggcgcgcga gcgagcgcg 1200
 gagcgcgcgc gcgagcgcgc aaggcttgct gcgagcgcgc gagcgcgcga gcgggaagg 1259
 atg agc gcg acc cgc gcg gcg acg agg aca gcg gcg gcg ctg tcc tcg 1307
 Met Ser Ala Thr Arg Ala Ala Thr Arg Thr Ala Ala Ala Leu Ser Ser
 1 5 10 15
 gcg ctg acg acg cct gta aag cag cag cag cag cag cag ctg cgc gta 1355
 Ala Leu Thr Thr Pro Val Lys Gln Gln Gln Gln Gln Leu Arg Val
 20 25 30
 ggc gcg gcg tcg gca cgg ctg gcg gcc gcg gcg ttc tcg tcc ggc acg 1403
 Gly Ala Ala Ser Ala Arg Leu Ala Ala Ala Ala Phe Ser Ser Gly Thr
 35 40 45
 ggc gga gac gcg gcc aag aag gcg gcc gcg gcg agg gcg ttc tcc acg 1451
 Gly Gly Asp Ala Ala Lys Lys Ala Ala Ala Ala Arg Ala Phe Ser Thr
 50 55 60
 gga cgc gcc ccc aac gcg aca cgc gag aag agc tcg ctg gcc acg gtc 1499
 Gly Arg Gly Pro Asn Ala Thr Arg Glu Lys Ser Ser Leu Ala Thr Val
 65 70 75 80
 cag gcg gcg acg gac gat gcg cgc ttc gtc ggc ctg acc ggc gcc caa 1547
 Gln Ala Ala Thr Asp Asp Ala Arg Phe Val Gly Leu Thr Gly Ala Gln
 85 90 95
 atc ttt cat gag ctc atg cgc gag cac cag gtg gac acc atc ttt ggc 1595
 Ile Phe His Glu Leu Met Arg Glu His Gln Val Asp Thr Ile Phe Gly
 100 105 110
 tac cct ggc ggc gcc att ctg ccc gtt ttt gat gcc att ttt gag agt 1643
 Tyr Pro Gly Gly Ala Ile Leu Pro Val Phe Asp Ala Ile Phe Glu Ser
 115 120 125
 gac gcc ttc aag ttc att ctc gct cgc cac gag cag ggc gcc ggc cac 1691
 Asp Ala Phe Lys Phe Ile Leu Ala Arg His Glu Gln Gly Ala Gly His
 130 135 140 145
 atg gcc gag ggc tac gcg cgc gcc acg gcc aag ccc ggc gtt gtc ctc 1739
 Met Ala Glu Gly Tyr Ala Arg Ala Thr Gly Lys Pro Gly Val Val Leu
 145 150 155 160
 gtc acc tcg ggc cct gga gcc acc aac acc atc acc ccg atc atg gat 1787
 Val Thr Ser Gly Pro Gly Ala Thr Asn Thr Ile Thr Pro Ile Met Asp
 165 170 175
 gct tac atg gac ggt acg ccg ctg ctc gtg ttc acc ggc cag gtg ccc 1835
 Ala Tyr Met Asp Gly Thr Pro Leu Leu Val Phe Thr Gly Gln Val Pro
 180 185 190
 acc tct gct gtc ggc acg gac gct ttc cag gag tgt gac att gtt ggc 1883
 Thr Ser Ala Val Gly Thr Asp Ala Phe Gln Glu Cys Asp Ile Val Gly
 195 200 205
 atc agc cgc gcg tgc acc aag tgg aac gtc atg gtc aag gac gtg aag 1931
 Ile Ser Arg Ala Cys Thr Lys Trp Asn Val Met Val Lys Asp Val Lys
 210 215 220
 gag ctc ccg cgc cgc atc aat gag gcc ttt gag att gcc atg agc ggc 1979
 Glu Leu Pro Arg Arg Ile Asn Glu Ala Phe Glu Ile Ala Met Ser Gly
 225 230 235 240
 cgc ccg ggt ccc gtg ctc gtc gat ctt cct aag gat gtg acc gcc gtt 2027
 Arg Pro Gly Pro Val Leu Val Asp Leu Pro Lys Asp Val Thr Ala Val
 245 250 255
 gag ctc aag gaa atg ccc gac agc tcc ccc cag gtt gct gtg cgc cag 2075
 Glu Leu Lys Glu Met Pro Asp Ser Ser Pro Gln Val Ala Val Arg Gln
 260 265 270
 aag caa aag gtc gag ctt ttc cac aag gag cgc att ggc gct cct ggc 2123
 Lys Gln Lys Val Glu Leu Phe His Lys Glu Arg Ile Gly Ala Pro Gly

ES 2 698 473 T3

acg gcc gac ttc aag ctc att gcc gag atg atc aac cgt gcg gag cga	2171
Thr Ala Asp Phe Lys Leu Ile Ala Glu Met Ile Asn Arg Ala Glu Arg	
290 295 300	
ccc gtc atc tat gct ggc cag ggt gtc atg cag agc ccg ttg aat ggc	2219
Pro Val Ile Tyr Ala Gly Gln Gly Val Met Gln Ser Pro Leu Asn Gly	
305 310 315 320	
ccg gct gtg ctc aag gag ttc gcg gag aag gcc aac att ccc gtg acc	2267
Pro Ala Val Leu Lys Glu Phe Ala Glu Lys Ala Asn Ile Pro Val Thr	
325 330 335	
acc acc atg cag ggt ctc ggc gcc ttt gac gag cgt agt ccc ctc tcc	2315
Thr Thr Met Gln Gly Leu Gly Gly Phe Asp Glu Arg Ser Pro Leu Ser	
340 345 350	
ctc aag atg ctc ggc atg cac gcc tct gcc tac gcc aac tac tcg atg	2363
Leu Lys Met Leu Gly Met His Gly Ser Ala Tyr Ala Asn Tyr Ser Met	
355 360 365	
cag aac gcc gat ctt atc ctg gcg ctc ggt gcc cgc ttt gat gat cgt	2411
Gln Asn Ala Asp Leu Ile Leu Ala Leu Gly Ala Arg Phe Asp Asp Arg	
370 375 380	
gtg acg ggc cgc gtt gac gcc ttt gct ccg gag gct cgc cgt gcc gag	2459
Val Thr Gly Arg Val Asp Ala Phe Ala Pro Glu Ala Arg Arg Ala Glu	
385 390 395 400	
cgc gag ggc cgc ggt ggc atc gtt cac ttt gag att tcc ccc aag aac	2507
Arg Glu Gly Arg Gly Gly Ile Val His Phe Glu Ile Ser Pro Lys Asn	
405 410 415	
ctc cac aag gtc gtc cag ccc acc gtc gcg gtc ctc ggc gac gtg gtc	2555
Leu His Lys Val Val Gln Pro Thr Val Ala Val Leu Gly Asp Val Val	
420 425 430	
gag aac ctc gcc aac gtc acg ccc cac gtg cag cgc cag gag cgc gag	2603
Glu Asn Leu Ala Asn Val Thr Pro His Val Gln Arg Gln Glu Arg Glu	
435 440 445	
ccg tgg ttt gcg cag atc gcc gat tgg aag gag aag cac cct ttt ctg	2651
Pro Trp Phe Ala Gln Ile Ala Asp Trp Lys Glu Lys His Pro Phe Leu	
450 455 460	
ctc gag tct gtt gat tcg gac gac aag gtt ctc aag ccg cag cag gtc	2699
Leu Glu Ser Val Asp Ser Asp Asp Lys Val Leu Lys Pro Gln Gln Val	
465 470 475 480	
ctc acg gag ctt aac aag cag att ctc gag att cag gag aag gac gcc	2747
Leu Thr Glu Leu Asn Lys Gln Ile Leu Glu Ile Gln Glu Lys Asp Ala	
485 490 495	
gac cag gag gtc tac atc acc acg gcc gtc gga agc cac cag atg cag	2795
Asp Gln Glu Val Tyr Ile Thr Thr Gly Val Gly Ser His Gln Met Gln	
500 505 510	
gca gcg cag ttc ctt acc tgg acc aag ccg cgc cag tgg atc tcc tcg	2843
Ala Ala Gln Phe Leu Thr Trp Thr Lys Pro Arg Gln Trp Ile Ser Ser	
515 520 525	
ggt ggc gcc ggc act atg gcc tac gcc ctt ccc tcg gcc att ggc gcc	2891
Gly Gly Ala Gly Thr Met Gly Tyr Gly Leu Pro Ser Ala Ile Gly Ala	
530 535 540	
aag att gcc aag ccc gat gct att gtt att gac atc gat ggt gat gct	2939
Lys Ile Ala Lys Pro Asp Ala Ile Val Ile Asp Ile Asp Gly Asp Ala	
545 550 555 560	
tct tat tcg atg acc ggt atg gaa ttg atc aca gca gcc gaa ttc aag	2987
Ser Tyr Ser Met Thr Gly Met Glu Leu Ile Thr Ala Ala Glu Phe Lys	
565 570 575	
gtt ggc gtg aag att ctt ctt ttg cag aac aac ttt cag gcc atg gtc	3035
Val Gly Val Lys Ile Leu Leu Leu Gln Asn Asn Phe Gln Gly Met Val	
580 585 590	
aag aac tgg cag gat ctc ttt tac gac aag cgc tac tcg gcc acc gcc	3083

ES 2 698 473 T3

Lys	Asn	Trp	Gln	Asp	Leu	Phe	Tyr	Asp	Lys	Arg	Tyr	Ser	Gly	Thr	Ala		
		595					600					605					
atg	ttc	aac	ccg	cgc	ttc	gac	aag	gtc	gcc	gat	gcg	atg	cgt	gcc	aag	3131	
Met	Phe	Asn	Pro	Arg	Phe	Asp	Lys	Val	Ala	Asp	Ala	Met	Arg	Ala	Lys		
	610				615					620							
ggt	ctc	tac	tgc	gcg	aaa	cag	tcg	gag	ctc	aag	gac	aag	atc	aag	gag	3179	
Gly	Leu	Tyr	Cys	Ala	Lys	Gln	Ser	Glu	Leu	Lys	Asp	Lys	Ile	Lys	Glu		
625				630						635					640		
ttt	ctc	gag	tac	gat	gag	ggt	ccc	gtc	ctc	ctc	gag	ggt	ttc	gtg	gac	3227	
Phe	Leu	Glu	Tyr	Asp	Glu	Gly	Pro	Val	Leu	Leu	Glu	Val	Phe	Val	Asp		
			645					650					655				
aag	gac	acg	ctc	gtc	ttg	ccc	atg	gtc	ccc	gct	ggc	ttt	cag	ctc	cac	3275	
Lys	Asp	Thr	Leu	Val	Leu	Pro	Met	Val	Pro	Ala	Gly	Phe	Pro	Leu	His		
		660						665					670				
gag	atg	gtc	ctc	gag	cct	cct	aag	ccc	aag	gac	gcc	taa	gttctttttt			3324	
Glu	Met	Val	Leu	Glu	Pro	Pro	Lys	Pro	Lys	Asp	Ala						
		675					680										
ccatggcggg	cgagcgagcg	agcgcgcgag	cgcgcaagt	cgcaagcgc	ttgccttgct											3384	
ttgcttcgct	tcgctttgct	ttgcttcaca	caacctaat	atgaattcaa	gtttctttg											3444	
ttgtcggcga	tgctgcctg	ccaaccagcc	agccatccg	ccggccgtcc	ttgacgcct											3504	
cgcttcggcg	gcggccatcg	attcaatca	cccattccg	acgttccg	ccctcacgtc											3564	
cgtctgcgca	cgaccctgc	acgaccaag	caaggccaac	gcgcgctca	gctcagctg											3624	
tcgacgagtc	gcacgtcaca	tatctcagat	gcatttggac	tgtgagtgt	attatgccac											3684	
tagcacgcaa	cgatcttcg	ggtcctcgt	cattgcatcc	gttcgggccc	tcgagcggt											3744	
gacgcgagtc	gcccgcgaga	cgctgcagca	ggccgctcc	acgcgaggg	tcgagctcg											3804	
cgcccccggg	cgatgtctg	ctggcgccga	ctgactctg	gagcgcaag	aagacacgg											3864	
gacgcgagga	ggaccgaaga	gagacgctg	ggtatgcag	atataccc	ggcgggacat											3924	
tcgttcggca	tacactcccc	cattcagct	tgctcgtcct	tgccagagcc	gagcgogaac											3984	
ggttcogaac	gcccgaagga	ttttggctct	ggtgggtgga	ctccgatcga	ggcgcaggt											4044	
ctccgcagtc	tctcgcaggg	cgccagtggt	cgttagaat	agggagtgc	gagctottg											4104	
cgccgccttag	ctcactctcc	gcccacgcgc	gcctcgcgc	catgcccgc	tcccgtctgt											4164	
cgctcgcctg	gcccgcgacc	gctcgcgac	agtacgcag	tgggacagag	ctcagggcga											4224	
cgccgaatcgc	tcgggttgta	agggtttcaa	gggtcggcg	tcgtcgcgtg	ccaaagtga											4284	
aatagtaggg	gggggggggg	gtacccaccc	cgggcaggt	ctcctcgcca	gcctaagtgc											4344	
ctaaggaggc	gtaggggttt	cgttgaccag	agaagcggg	aacctgccgc	ggcggggaga											4404	
acctatcgcc	ggagaacttg	ccagggcga	ggcagttct	caatttgcg	acagcggcgc											4464	
gcccacgcga	ggcggccg	tgccgataca	gcgagcgac	cgccggggc	cgctggcga											4524	
cacagctgcg	cgccggagtc	gctgcgagaa	ggcttctcg	tggttggtt	ggggtcgcg											4584	
gtgagcggg	atggatgccc	aggtaactgc	gcgtgcgcgc	gcccaggag	aaaaggacag											4644	
acgcgcggg	ctgcgatgcg	agcacgcgat	gcgagcacgc	gatgcgagca	cgcgatgcga											4704	
gcacgcgagc	gagcgcgca	gcaaatgcca	cggaacacgc	gttttttgt	tggtgatttc											4764	
tatgtatgcg	gggagacttc	gatggccgaa	aggggtgcaa	ggccaaaaga	tgctgacagc											4824	
ttcgatcggt	ctacggcgcg	agcaggaaag	ggagcaagg	gcgggaattct	tctgccttga											4884	
cccgggggat	ccactagttc	tagacgggcc	gccaccgcg	tgagctcca	attcgccta											4944	
tagtgagtcg	tattacgcgc	gctcactggc	cgctcgttta	caactcgtg	actgggaaaa											5004	
ccctggcggt	acccaactta	atcgccttgc	agcacatccc	ccttccgcca	gctggcgtaa											5064	
tagcgaagag	gcccgcaccg	atcgccttcc	ccaacagttg	cgagcctga	atggcgaaatg											5124	
ggaacgcgccc	tgtagcggcg	cattaagcgc	ggcgggtgtg	tggtttacgc	gcagcgtgac											5184	
cgctacactt	gccagcgcgc	tagcgcgcgc	tcctttcgt	ttcttccctt	cctttctcgc											5244	
cacgttcgcc	ggctttcccc	gtcaagctct	aaatcgggg	ctccccttag	ggttccgatt											5304	
tagtgcttta	cgccacctcg	accccaaaaa	acttgattag	ggtgatggtt	cacgtagtg											5364	
gcccctgcgcc	tgatagcgg	tttttcgccc	tttgacgttg	gagtcacgt	tctttaatag											5424	
tggaactctt	ttccaaactg	gaacaacact	caaccctatc	tcggtctatt	cctttgatit											5484	
ataagggatt	ttgccgattt	cgccctattg	gttaaaaaat	gagctgattt	aacaaaaaat											5544	
taacgcgaat	tttaacaaaa	tattaacgct	tacaatttag	gtggcacttt	tcggggaat											5604	
gtgcgcggaa	cccctatttg	tttatttttc	taaatacatt	caaataatgta	tccgctcatg											5664	
agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	gagtattcaa											5724	
catttcgctg	tcgccttat	tccccttttt	cgggcatttt	gccttccctg	ttttgctcac											5784	
ccagaaacgc	tggtgaaagt	aaaagatgct	gaagatcagt	tgggtgcacg	agtgggttac											5844	

ES 2 698 473 T3

```

atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt 5904
ccaatgatga gcactttttaa agttctgcta tgtggcgcgg tattatcccg tattgacgcc 5964
gggcaagagc aactcggctc ccgcatacac tattctcaga atgacttggg tgagtactca 6024
ccagtcacag aaaagpatct tacggatggc atgacagtaa gagaattatg cagtgtctcc 6084
ataaccatga gtgataacac tgcggccaac ttacttctga caacgatcgg aggaccgaa 6144
gagctaaccg cttttttgca caacatgggg gatcatgtaa ctccgcttga tcgttgggaa 6204
ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg 6264
gcaacaacgt tgcgcaaac attaactggc gaactactta ctctagcttc ccggcaacaa 6324
ttaatagact ggatggaggc ggataaagt gcaggaccac ttctgcgctc ggoccttccg 6384
gctggctggg ttattgtgta taaatctgga gccggtgagc gtgggtctcg cggtatcatt 6444
gcagcactgg ggcagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 6504
caggcaacta tggatgaacg aaatgacag atcgtgaga taggtgcctc actgattaag 6564
cattggtaac tgtcagacca agttactca tatatacttt agattgattt aaaacttcat 6624
ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct 6684
taacgtgagt ttctgttcca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct 6744
tgagatcctt tttttctgcg cgtaatctgc tgcctgcaaa caaaaaaacc accgctacca 6804
ggcgtggttt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt aactggcttc 6864
agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg ccaccacttc 6924
aagaactctg tagcaccgcc tacataacct gctctgctaa tcctgttacc agtggctgct 6984
gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag 7044
cgcagcgggt cgggtgaac ggggggttcg tgcacacagc ccagcttggg gcgaacgacc 7104
acaaacgaa tgagatacct acagcgtgag ctatgagaaa gcgccacgct tccggaagg 7164
agaaagggg acaggtatcc ggtaaagcgc agggctcgaa caggagagcg cacgaggag 7224
cttccagggg gaacgcctg gtatctttat agtctgtcgg ggttctcgca cctctgactt 7284
gagcgtcgat ttttgtgatg ctctcaggg gggcggagcc tatggaaaaa cgccagcaac 7344
gcggcctttt tacggttccg gcccttttgc tggccttttg ctcacatggt ctttctgcg 7404
ttatccctg attctgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc 7464
cgcagccgaa cgaccgagcg cagcagatca gtgagcagag aagcgggaaga gcgcccaata 7524
ccgactgga aagcgggcag tgagcgcgcaac gcaattaatg gcagctggca cgacaggttt 7584
gcacccagc ctttacactt tatgcttccg gctcgtatgt tgtgtggaat tgtgagcggg 7644
taacaatttc acacagggaa cagctatgac catgattacg ccaagcgcgc aattaacctt 7704
cactaaaggc aacaaaagct ggtaccggg cccccctcg aggtcgacgg tatcgataag 7764
cttgataatc aattcctgca gcc 7824

```

<210> 15

<211> 684

<212> PRT

<213> Schizochytrium sp.

<400> 15

```

Met Ser Ala Thr Arg Ala Ala Thr Arg Thr Ala Ala Ala Leu Ser Ser
1          5          10          15

Ala Leu Thr Thr Pro Val Lys Gln Gln Gln Gln Gln Gln Leu Arg Val
20          25          30

Gly Ala Ala Ser Ala Arg Leu Ala Ala Ala Ala Phe Ser Ser Gly Thr
35          40          45

Gly Gly Asp Ala Ala Lys Lys Ala Ala Ala Ala Arg Ala Phe Ser Thr
50          55          60

Gly Arg Gly Pro Asn Ala Thr Arg Glu Lys Ser Ser Leu Ala Thr Val
65          70          75          80

Gln Ala Ala Thr Asp Asp Ala Arg Phe Val Gly Leu Thr Gly Ala Gln
85          90          95

```

5

10

ES 2 698 473 T3

Ile Phe His Glu Leu Met Arg Glu His Gln Val Asp Thr Ile Phe Gly
100 105 110

Tyr Pro Gly Gly Ala Ile Leu Pro Val Phe Asp Ala Ile Phe Glu Ser
115 120 125

Asp Ala Phe Lys Phe Ile Leu Ala Arg His Glu Gln Gly Ala Gly His
130 135 140

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

Val Thr Ser Gly Pro Gly Ala Thr Asn Thr Ile Thr Pro Ile Met Asp
165 170 175

Ala Tyr Met Asp Gly Thr Pro Leu Leu Val Phe Thr Gly Gln Val Pro
180 185 190

Thr Ser Ala Val Gly Thr Asp Ala Phe Gln Glu Cys Asp Ile Val Gly
195 200 205

Ile Ser Arg Ala Cys Thr Lys Trp Asn Val Met Val Lys Asp Val Lys
210 215 220

Glu Leu Pro Arg Arg Ile Asn Glu Ala Phe Glu Ile Ala Met Ser Gly
225 230 235 240

Arg Pro Gly Pro Val Leu Val Asp Leu Pro Lys Asp Val Thr Ala Val
245 250 255

Glu Leu Lys Glu Met Pro Asp Ser Ser Pro Gln Val Ala Val Arg Gln
260 265 270

Lys Gln Lys Val Glu Leu Phe His Lys Glu Arg Ile Gly Ala Pro Gly
275 280 285

Thr Ala Asp Phe Lys Leu Ile Ala Glu Met Ile Asn Arg Ala Glu Arg
290 295 300

Pro Val Ile Tyr Ala Gly Gln Gly Val Met Gln Ser Pro Leu Asn Gly
305 310 315 320

Pro Ala Val Leu Lys Glu Phe Ala Glu Lys Ala Asn Ile Pro Val Thr
325 330 335

Thr Thr Met Gln Gly Leu Gly Gly Phe Asp Glu Arg Ser Pro Leu Ser
340 345 350

Leu Lys Met Leu Gly Met His Gly Ser Ala Tyr Ala Asn Tyr Ser Met
355 360 365

Gln Asn Ala Asp Leu Ile Leu Ala Leu Gly Ala Arg Phe Asp Asp Arg
370 375 380

Val Thr Gly Arg Val Asp Ala Phe Ala Pro Glu Ala Arg Arg Ala Glu
385 390 395 400

Arg Glu Gly Arg Gly Gly Ile Val His Phe Glu Ile Ser Pro Lys Asn
405 410 415

ES 2 698 473 T3

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

<210> 16
 <211> 30
 <212> ADN
 <213> Schizochytrium sp.

5

<400> 16
 tatcgataag cttgacgtcg aattcctgca 30

10

<210> 17
 <211> 32
 <212> ADN
 <213> Schizochytrium sp.

15

<210> 18
 <211> 7847
 <212> ADN
 <213> Schizochytrium sp.

20

ES 2 698 473 T3

```

<220>
<221> prim_transcript
<222> (1)..(1259)
<223>
5

<220>
<221> CDS
<222> (1260)..(3314)
<223>
10

<220>
<221> prim_transcript
<222> (3315)..(4887)
<223>
15

<220>
<221> mutación
<222> (3042)..(3044)
<223>
20

<400> 18
ttgtcgacag caacttgcaa gttatacgcg accaccaggc aatctcagca cgcccagcga 60
gcacggagct tgcgaagagg gtttacaagt cgtcgttcat tcgctctcaa gctttgcctc 120
aacgcaacta ggcccaggcc tactttcact gtgtcttgtc ttgcctttca caccgaccga 180
gtgtgcacaa ccgtgttttg cacaaagcgc aagatgctca ctcgactgtg aagaagggtt 240
gcgcgcaagc gactgcgact gcgaggatga ggatgactgg cagcctgttc aaaaactgaa 300
aatcccgcat gggtcagtgc cattcgcgca tgacgcctgc gagagacaag ttaactcgtg 360
tcaactggcat gtccctagcat ctttacgcga gcaaaattca atogctttat tttttcagtt 420
tcgtaacctt ctcgcaaccg cgaatcgccg tttcagcctg actaatctgc agctgctgtg 480
cactgtcagt cagtcagtca gtctgtgctg ctgttccagc accgaggtcg cgcgtcgccg 540
cgcctggacc gctgctgcta ctgctagtgg cacggcaggt aggagcttgt tgccggaaca 600
ccagcagccg ccagtcgacg ccagccaggg gaaagtccgg cgtcgaaggg agaggaaggc 660
ggcgtgtgca aactaacgtt gacctggc gcccgccgac acgagcagga agcaggcagc 720
tgcagagcgc agcgcgcaag tgcagaatgc gcgaaagatc cacttgcgcg cggcgggccc 780
gcacttgcgg gcgcggcgcg gaacagtgcg gaaaggagcg gtgcagacgg cgcgcagtga 840
cagtgggctc aaagccgctc agtaagcagc gccggggaac ggtatacgca gtgctcgctg 900
ccgcccacac cagaagtata cgcgggcccga agtggggcgt cgcgcgcggg aagtgcggaa 960
tggcgggcaa ggaaaggagg agacgaaaag agggcgggaa agagagagag agagagtgaa 1020
aaaagaaaga aagaagaaa gaaagaaaaga aagctcggag ccacgcgcg gggagagaga 1080
gaaatgaaag cacggcacgg caaagcaaag caaagcagac ccagccagac ccagccgagg 1140
gaggagcgcg cgcaggaccc gcgcgccgag cgagcgagca cggcgcgcga gcgagcagc 1200
gagcagcgcg gcgagcagc aaggcttctt gcgagcagc gagcagcga gcgggaagg 1259
atg agc gcg acc cgc gcg gcg acg agg aca gcg gcg gcg ctg tcc tcg 1307
Met Ser Ala Thr Arg Ala Ala Thr Arg Thr Ala Ala Ala Leu Ser Ser
1 5 10 15
gcg ctg acg acg cct gta aag cag cag cag cag cag cag ctg cgc gta 1355

```

ES 2 698 473 T3

Ala	Leu	Thr	Thr	Pro	Val	Lys	Gln	Gln	Gln	Gln	Gln	Leu	Arg	Val			
			20					25				30					
ggc	gcg	gcg	tcg	gca	cgg	ctg	gcg	gcc	gcg	gcg	ttc	tcg	tcc	ggc	acc	1403	
Gly	Ala	Ala	Ser	Ala	Arg	Leu	Ala	Ala	Ala	Ala	Phe	Ser	Ser	Gly	Thr		
			35				40				45						
ggc	gga	gac	gcg	gcc	aag	aag	gcg	gcc	gcg	gog	agg	gcg	ttc	tcc	acc	1451	
Gly	Gly	Asp	Ala	Ala	Lys	Lys	Ala	Ala	Ala	Ala	Arg	Ala	Phe	Ser	Thr		
			50			55					60						
gga	gcg	ggc	ccc	aac	gcg	aca	cgc	gag	aag	agc	tcg	ctg	gcc	acc	gtc	1499	
Gly	Arg	Gly	Pro	Asn	Ala	Thr	Arg	Glu	Lys	Ser	Ser	Leu	Ala	Thr	Val		
					70					75					80		
cag	gcg	gcg	acc	gac	gat	gcg	cgc	ttc	gtc	ggc	ctg	acc	ggc	gcc	caa	1547	
Gln	Ala	Ala	Thr	Asp	Asp	Ala	Arg	Phe	Val	Gly	Leu	Thr	Gly	Ala	Gln		
				85					90						95		
atc	ttt	cat	gag	ctc	atg	cgc	gag	cac	cag	gtg	gac	acc	atc	ttt	ggc	1595	
Ile	Phe	His	Glu	Leu	Met	Arg	Glu	His	Gln	Val	Asp	Thr	Ile	Phe	Gly		
			100					105							110		
tac	cct	ggc	ggc	gcc	att	ctg	ccc	gtt	ttt	gat	gcc	att	ttt	gag	agt	1643	
Tyr	Pro	Gly	Gly	Ala	Ile	Leu	Pro	Val	Phe	Asp	Ala	Ile	Phe	Glu	Ser		
			115				120					125					
gac	gcc	ttc	aag	ttc	att	ctc	gct	cgc	cac	gag	cag	ggc	gcc	ggc	cac	1691	
Asp	Ala	Phe	Lys	Phe	Ile	Leu	Ala	Arg	His	Glu	Gln	Gly	Ala	Gly	His		
			130			135					140						
atg	gcc	gag	ggc	tac	gcg	cgc	gcc	acc	ggc	aag	ccc	ggc	ggt	gtc	ctc	1739	
Met	Ala	Glu	Gly	Tyr	Ala	Arg	Ala	Thr	Gly	Lys	Pro	Gly	Val	Val	Leu		
					150					155					160		
gtc	acc	tcg	ggc	cct	gga	gcc	acc	aac	acc	atc	acc	ccg	atc	atg	gat	1787	
Val	Thr	Ser	Gly	Pro	Gly	Ala	Thr	Asn	Thr	Ile	Thr	Pro	Ile	Met	Asp		
				165					170						175		
gct	tac	atg	gac	ggt	acc	ccg	ctg	ctc	gtg	ttc	acc	ggc	cag	gtg	ccc	1835	
Ala	Tyr	Met	Asp	Gly	Thr	Pro	Leu	Leu	Val	Phe	Thr	Gly	Gln	Val	Pro		
				180					185						190		
acc	tct	gct	gtc	ggc	acc	gac	gct	ttc	cag	gag	tgt	gac	att	gtt	ggc	1883	
Thr	Ser	Ala	Val	Gly	Thr	Asp	Ala	Phe	Gln	Glu	Cys	Asp	Ile	Val	Gly		
				195			200					205					
atc	agc	cgc	gcg	tgc	acc	aag	tgg	aac	gtc	atg	gtc	aag	gac	gtg	aag	1931	
Ile	Ser	Arg	Ala	Cys	Thr	Lys	Trp	Asn	Val	Met	Val	Lys	Asp	Val	Lys		
						215					220						
gag	ctc	ccg	gcg	cgc	atc	aat	gag	gcc	ttt	gag	att	gcc	atg	agc	ggc	1979	
Glu	Leu	Pro	Arg	Arg	Ile	Asn	Glu	Ala	Phe	Glu	Ile	Ala	Met	Ser	Gly		
					230					235					240		
cgc	ccg	ggt	ccc	gtg	ctc	gtc	gat	ctt	cct	aag	gat	gtg	acc	gcc	ggt	2027	
Arg	Pro	Gly	Pro	Val	Leu	Val	Asp	Leu	Pro	Lys	Asp	Val	Thr	Ala	Val		
				245					250						255		
gag	ctc	aag	gaa	atg	ccc	gac	agc	tcc	ccc	cag	ggt	gct	gtg	cgc	cag	2075	
Glu	Leu	Lys	Glu	Met	Pro	Asp	Ser	Ser	Pro	Gln	Val	Ala	Val	Arg	Gln		
				260				265							270		
aag	caa	aag	gtc	gag	ctt	ttc	cac	aag	gag	cgc	att	ggc	gct	cct	ggc	2123	
Lys	Gln	Lys	Val	Glu	Leu	Phe	His	Lys	Glu	Arg	Ile	Gly	Ala	Pro	Gly		
				275			280								285		
acc	gcc	gac	ttc	aag	ctc	att	gcc	gag	atg	atc	aac	cgt	gcg	gag	cga	2171	
Thr	Ala	Asp	Phe	Lys	Leu	Ile	Ala	Glu	Met	Ile	Asn	Arg	Ala	Glu	Arg		
					295						300						
ccc	gtc	atc	tat	gct	ggc	cag	ggt	gtc	atg	cag	agc	ccg	ttg	aat	ggc	2219	
Pro	Val	Ile	Tyr	Ala	Gly	Gln	Gly	Val	Met	Gln	Ser	Pro	Leu	Asn	Gly		
					310					315					320		
ccg	gct	gtg	ctc	aag	gag	ttc	gcg	gag	aag	gcc	aac	att	ccc	gtg	acc	2267	
Pro	Ala	Val	Leu	Lys	Glu	Phe	Ala	Glu	Lys	Ala	Asn	Ile	Pro	Val	Thr		
				325					330						335		

ES 2 698 473 T3

acc acc atg cag ggt ctc ggc ggc ttt gac gag cgt agt ccc ctc tcc	2315
Thr Thr Met Gln Gly Leu Gly Gly Phe Asp Glu Arg Ser Pro Leu Ser	
340 345 350	
ctc aag atg ctc ggc atg cac ggc tct gcc tac gcc aac tac tcg atg	2363
Leu Lys Met Leu Gly Met His Gly Ser Ala Tyr Ala Asn Tyr Ser Met	
355 360 365	
cag aac gcc gat ctt atc ctg gcg ctc ggt gcc cgc ttt gat gat cgt	2411
Gln Asn Ala Asp Leu Ile Leu Ala Leu Gly Ala Arg Phe Asp Asp Arg	
370 375 380	
gtg acg ggc cgc gtt gac gcc ttt gct cgg gag gct cgc cgt gcc gag	2459
Val Thr Gly Arg Val Asp Ala Phe Ala Pro Glu Ala Arg Arg Ala Glu	
385 390 395 400	
cgc gag ggc cgc ggt ggc atc gtt cac ttt gag att tcc ccc aag aac	2507
Arg Glu Gly Arg Gly Gly Ile Val His Phe Glu Ile Ser Pro Lys Asn	
405 410 415	
ctc cac aag gtc gtc cag ccc acc gtc gcg gtc ctc gcc gac gtg gtc	2555
Leu His Lys Val Val Gln Pro Thr Val Ala Val Leu Gly Asp Val Val	
420 425 430	
gag aac ctc gcc aac gtc acg ccc cac gtg cag cgc cag gag cgc gag	2603
Glu Asn Leu Ala Asn Val Thr Pro His Val Gln Arg Gln Glu Arg Glu	
435 440 445	
ccg tgg ttt gcg cag atc gcc gat tgg aag gag aag cac cct ttt ctg	2651
Pro Trp Phe Ala Gln Ile Ala Asp Trp Lys Glu Lys His Pro Phe Leu	
450 455 460	
ctc gag tct gtt gat tcg gac gac aag gtt ctc aag ccg cag cag gtc	2699
Leu Glu Ser Val Asp Ser Asp Lys Val Leu Lys Pro Gln Gln Val	
465 470 475 480	
ctc acg gag ctt aac aag cag att ctc gag att cag gag aag gac gcc	2747
Leu Thr Glu Leu Asn Lys Gln Ile Leu Glu Ile Gln Glu Lys Asp Ala	
485 490 495	
gac cag gag gtc tac atc acc acg ggc gtc gga agc cac cag atg cag	2795
Asp Gln Glu Val Tyr Ile Thr Thr Gly Val Gly Ser His Gln Met Gln	
500 505 510	
gca cgc cag ttc ctt acc tgg acc aag ccg cgc cag tgg atc tcc tcg	2843
Ala Ala Gln Phe Leu Thr Trp Thr Lys Pro Arg Gln Trp Ile Ser Ser	
515 520 525	
ggt ggc gcc ggc act atg ggc tac ggc ctt ccc tcg gcc att ggc gcc	2891
Gly Gly Ala Gly Thr Met Gly Tyr Gly Leu Pro Ser Ala Ile Gly Ala	
530 535 540	
aag att gcc aag ccc gat gct att gtt att gac atc gat ggt gat gct	2939
Lys Ile Ala Lys Pro Asp Ala Ile Val Ile Asp Ile Asp Gly Asp Ala	
545 550 555 560	
tct tat tcg atg acc ggt atg gaa ttg atc aca gca gcc gaa ttc aag	2987
Ser Tyr Ser Met Thr Gly Met Glu Leu Ile Thr Ala Ala Glu Phe Lys	
565 570 575	
gtt ggc gtg aag att ctt ctt ttg cag aac aac ttt cag gcc atg gtc	3035
Val Gly Val Lys Ile Leu Leu Leu Gln Asn Asn Phe Gln Gly Met Val	
580 585 590	
aag aac gtt cag gat ctc ttt tac gac aag cgc tac tcg gcc acc gcc	3083
Lys Asn Val Gln Asp Leu Phe Tyr Asp Lys Arg Tyr Ser Gly Thr Ala	
595 600 605	
atg ttc aac ccg cgc ttc gac aag gtc gcc gat cgc atg cgt gcc aag	3131
Met Phe Asn Pro Arg Phe Asp Lys Val Ala Asp Ala Met Arg Ala Lys	
610 615 620	
ggt ctc tac tgc cgc aaa cag tcg gag ctc aag gac aag atc aag gag	3179
Gly Leu Tyr Cys Ala Lys Gln Ser Glu Leu Lys Asp Lys Ile Lys Glu	
625 630 635 640	
ttt ctc gag tac gat gag ggt ccc gtc ctc ctc gag gtt ttc gtg gac	3227
Phe Leu Glu Tyr Asp Glu Gly Pro Val Leu Leu Glu Val Phe Val Asp	

ES 2 698 473 T3

	645		650		655	
aag gac acg ctc gtc ttg ccc atg gtc ccc gct ggc ttt ccg ctc cac						3275
Lys Asp Thr Leu Val Leu Pro Met Val Pro Ala Gly Phe Pro Leu His						
	660		665		670	
gag atg gtc ctc gag cct cct aag ccc aag gac gcc taa gttctttttt						3324
Glu Met Val Leu Glu Pro Pro Lys Pro Lys Asp Ala						
	675		680			
ccatggcggg cgagcgcgag agcgcgcgag cgcgcgaagt cgcaagcgcc ttgccttgct						3384
ttgcttgcgt tcgctttgct ttgcttcaca caacctaat atgaattcaa gttttcttgc						3444
ttgtcggcga tgctgcctg ccaaccagcc agccatccgg ccggccgtcc ttgacgcctt						3504
cgcttcgggc gggccatcg attcaattca cccatccgat acggtccgcc coctcaagtc						3564
cgctctgcga cgacccttc acgaccacgc caaggccaac gcgccgctca gctcagcttg						3624
tcgacgcgac gcacgtaca tatctcagat gcatttggac tgtgagtggt attatgccac						3684
tagcacgcaa cgatcttogg ggtccctcgt cattgcatec gttcggggccc tgcaggcgtg						3744
gacgcgagtc gccgcgaga cgtgcagca ggcogctccg acgcgagggc tcgagctcgc						3804
cgcgcccgcg cgatgtctgc ctggcgccga ctgatctctg gagcgcaagg aagacacggc						3864
gacgcgagga ggaccgaaga gagacgctgg ggtatgcagg atatacccg ggccggacat						3924
tcgttcggca tacactccc cattcgagct tgctcgtcct tggcagagcc gagcgcgaac						3984
ggttccgaac gcggcaagga ttttgctct ggtgggtgga ctccgatcga ggccaggtt						4044
ctccgcaggt tctcgcagge cggcagtggt cgttagaaat agggagtgcc ggagcttga						4104
cgcgcccttag ctactctcc gccacgcgc gcattcgcgc catgcgcgcg tcccgtctgt						4164
cgctgcgctg gccgcgaccg gctgcgccag agtaogacag tgggacagag ctcgaggoga						4224
cgcgaatcgc tcgggttgta agggtttcaa ggtcggggcg tcgtcgcgtg ccaaagtga						4284
aatagtaggg gggggggggg gtaccacccc cgggcagggt ctccctcgcca gcctaagtgc						4344
ctaagggagc gtagggggtt cgttgaccag agaagcggag aacctgccgc ggccgggaga						4404
acctatcggc ggagaacttg ccaggcgcga ggcagttctc caatttgcgg acagcggcgc						4464
gccacgcgga ggccggccgc tgccgataca gcgagggcgc cgcgcggggc cgcgtggcga						4524
cacagctcgc cgcggagtcg gctgcgagaa ggcttctcgc tggcttgggt ggggtcgcgg						4584
gtggcagggg atggatgccc aggtacgtcg gcgtgcgcgc gccacgggag aaaaggacag						4644
acgcgcgggc ctgcgatcgc agcacgcgat gcgagcacgc gatcgcgagca cgcgatcga						4704
gcacgcgagc gagcccccga gcaaatgcc cggaaacgc gttttttgt tggatattc						4764
tatgtatcgc gggagacttc gatggccgaa aggggtgcaa ggccaaaaga tgctgacagc						4824
ttcagctcgt ctaogggcgc agcaggaagg ggagcaagg gcggaattct tctgccttga						4884
cccggggat ccaactagtc tagagcggcc gccacgcgc tggagotcca attcgcctta						4944
tagtgagtcg tattacgcgc gctcactggc cgtcgtttta caacgtcgtg actgggaaaa						5004
ccctggcggt acccaactta atcgccttgc agcacatccc cctttcgcca gctggcgtaa						5064
tagcgaagag gcccgaccg atcgccttc ccaacagttg cgcagcctga atggcgaatg						5124
ggacgcgccc tgtagcggcg cattaagcgc ggccgggtgt gtggttacgc gcagcgtgac						5184
cgctacactt gccagcgcgc tagcgcgccgc tcccttccgt ttcttcccct cctttctcgc						5244
cagtttcgcc ggctttcccc gtaagctct aaatcggggg ctccctttag ggttccgatt						5304
tagtgcctta cggcacctcg acccaaaaa acttgattag ggtgatggt cacgtatgg						5364
gccatgccc tgatagacgg tttttcgccc tttgacgttg gagtccacgt tctttaatag						5424
ttgactcttg ttccaaactg gaacaacact caaccctatc tcggtctatt cttttgatt						5484
ataagggatt ttgcccattt cggcctattg gttaaaaaat gagctgattt aacaaaaatt						5544
taacgcgaat ttttaaaaa tattaacgct tacaatttag gtggcacttt tcggggaaat						5604
gtgcgcggaa ccctatttg ttatttttc taaatacatt caaatatgta tccgctcatg						5664
agacaatac cctgataaat gtttcaataa tattgaaaaa ggaagagtat gagtattcaa						5724
catttcgctg tcgcccttat tccctttttt gcggcatttt gccttcctgt ttttgctcac						5784
ccgaaaacgc ttggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac						5844
atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt						5904
ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg tattgacgcc						5964
gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttgg tgagtactca						6024
ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtctgcc						6084
ataaccatga gtgataaac tcgggccaac ttacttctga caacgatcgg aggaccgaag						6144
gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga tcgttgggaa						6204
ccggagctga atgaagcat accaaaacgc gagcgtgaca ccacgatgcc tgtagcaatg						6264
gcaacaacgt tgcgcaact attaactggc gaactactta ctctagcttc ccggcaacaa						6324
ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttccg						6384
gctggctggt ttattgctga taaatctgga gccggtgagc gttggctctc cggtatcatt						6444

ES 2 698 473 T3

```

gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 6504
caggcaacta tggatgaaacg aaatagacag atcgcgtgaga taggtgcctc actgattaag 6564
cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaacttcat 6624
ttttaattta aaagatccta ggtgaagatc ctttttgata atctcatgac caaaatccct 6684
taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct 6744
tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaac accgctacca 6804
gcggtgggtt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt aactggcttc 6864
agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg ccaccacttc 6924
aagaactctg tagcacccgc tacatacctc gctctgctaa tcctgttacc agtggctgct 6984
gccagtggcg ataagtcgtg tcttaccogg ttggactcaa gacgatagtt accggataag 7044
gcgcagcggg cgggctgaac ggggggttcg tgcacacagc ccagcttggg gccaacgacc 7104
tacaccgaac tgagatcctt acagcgtgag ctatgagaaa gcccacgct tcccgaaggg 7164
agaaaggcgg acaggtatcc ggtaagcggc agggctcggaa caggagagcg cacgaggag 7224
cttcaggggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca cctctgactt 7284
gagcgtcgat ttttgatgat ctgcgcaggg gggcggagcc tatggaaaaa cgccagcaac 7344
gcgccctttt tacggttccct ggccttttgc tggccttttg ctacatgtt ctttctcgcg 7404
ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc 7464
cgcagccgaa cgaccgagcg cagcgcgtca gtgagcggag aagcgggaaga ggcccaata 7524
cgcaaaccgc ctctcccgcg gcgttgcccg attcattaat gcagctggca cgacagggtt 7584
cccactgga aagcgggcag tgagcgcac gcaattaatg tgagttagct cactcattag 7644
gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat tgtgagcgga 7704
taacaatttc acacaggaaa cagctatgac catgattacg ccaagcgcgc aattaaccct 7764
cactaaaggg aacaaaagct ggtaccogg cccccctcg aggtcgacgy tatcgataag 7824
cttgacgtcg aattctcgca gcc 7847

```

<210> 19
 <211> 684
 <212> PRT
 <213> Schizochytrium sp.

5

```

<400> 19
Met Ser Ala Thr Arg Ala Ala Thr Arg Thr Ala Ala Ala Leu Ser Ser
1 5 10 15

Ala Leu Thr Thr Pro Val Lys Gln Gln Gln Gln Gln Gln Leu Arg Val
20 25 30

Gly Ala Ala Ser Ala Arg Leu Ala Ala Ala Ala Phe Ser Ser Gly Thr
35 40 45

Gly Gly Asp Ala Ala Lys Lys Ala Ala Ala Ala Arg Ala Phe Ser Thr
50 55 60

Gly Arg Gly Pro Asn Ala Thr Arg Glu Lys Ser Ser Leu Ala Thr Val
65 70 75 80

Gln Ala Ala Thr Asp Asp Ala Arg Phe Val Gly Leu Thr Gly Ala Gln
85 90 95

Ile Phe His Glu Leu Met Arg Glu His Gln Val Asp Thr Ile Phe Gly
100 105 110

Tyr Pro Gly Gly Ala Ile Leu Pro Val Phe Asp Ala Ile Phe Glu Ser
115 120 125

Asp Ala Phe Lys Phe Ile Leu Ala Arg His Glu Gln Gly Ala Gly His
130 135 140

Met Ala Glu Gly Tyr Ala Arg Ala Thr Gly Lys Pro Gly Val Val Leu

```

10

ES 2 698 473 T3

Leu Glu Ser Val Asp Ser Asp Asp Lys Val Leu Lys Pro Gln Gln Val
 465 470 475 480
 Leu Thr Glu Leu Asn Lys Gln Ile Leu Glu Ile Gln Glu Lys Asp Ala
 485 490 495
 Asp Gln Glu Val Tyr Ile Thr Thr Gly Val Gly Ser His Gln Met Gln
 500 505 510
 Ala Ala Gln Phe Leu Thr Trp Thr Lys Pro Arg Gln Trp Ile Ser Ser
 515 520 525
 Gly Gly Ala Gly Thr Met Gly Tyr Gly Leu Pro Ser Ala Ile Gly Ala
 530 535 540
 Lys Ile Ala Lys Pro Asp Ala Ile Val Ile Asp Ile Asp Gly Asp Ala
 545 550 555 560
 Ser Tyr Ser Met Thr Gly Met Glu Leu Ile Thr Ala Ala Glu Phe Lys
 565 570 575
 Val Gly Val Lys Ile Leu Leu Leu Gln Asn Asn Phe Gln Gly Met Val
 580 585 590
 Lys Asn Val Gln Asp Leu Phe Tyr Asp Lys Arg Tyr Ser Gly Thr Ala
 595 600 605
 Met Phe Asn Pro Arg Phe Asp Lys Val Ala Asp Ala Met Arg Ala Lys
 610 615 620
 Gly Leu Tyr Cys Ala Lys Gln Ser Glu Leu Lys Asp Lys Ile Lys Glu
 625 630 635 640
 Phe Leu Glu Tyr Asp Glu Gly Pro Val Leu Leu Glu Val Phe Val Asp
 645 650 655
 Lys Asp Thr Leu Val Leu Pro Met Val Pro Ala Gly Phe Pro Leu His
 660 665 670
 Glu Met Val Leu Glu Pro Pro Lys Pro Lys Asp Ala
 675 680

5 <210> 20
 <211> 26
 <212>ADN
 <213>Schizochytrium sp.

10 <400> 20
 ccggccaggt gcagacctct gctgtc 26

15 <210> 21
 <211> 7847
 <212>ADN
 <213>Schizochytrium sp.

20 <220>
 <221> prim_transcript
 <222> (1)..(1259)
 <223>

25 <220>
 <221> mutación
 <222> (1334)..(1835)
 <223>

30 <220>
 <221> CDS
 <222> (1260)..(3314)
 <223>

<220>
 <221> prim_transcript
 <222> (3315)..(4887)
 <223>

5

<400> 21
 ttgtcgacag caacttgcaa gttatacgcg accaccaggc aatctcagca cgcccagcga 60
 gcacggagct tgccaagagg gtttacaagt cgtcgttcat tcgctctcaa gctttgcctc 120
 aacgcaacta ggcccagggc tactttcact gtgtcttctc ttgcctttca caccgaccga 180
 gtgtgcacaa ccgtgttttg cacaaagcgc aagatgctca ctgactgtg aagaaagggt 240
 gcgcgcaagc gactgcgact gcgaggatga ggatgactgg cagcctgttc aaaaactgaa 300
 aatccgcgat gggtcagtgc cattcgcgca tgacgcctgc gagagacaag ttaactogtg 360
 tcactggcat gtcctagcat ctttacgcga gcaaaattca atcgctttat tttttcagtt 420
 tcgtaacctt ctgcgcaacgc cgaatcgcgc tttcagcctg actaatctgc agctgcgtgg 480
 cactgtcagt cagtcagtca gtctgtcgcg ctgttccagc accgaggtcg cgcgtcgcgg 540
 cgcctggacc gctgctgcta ctgctagtgg cacggcaggt aggagcttgt tgccggaaca 600
 ccagcagccg ccagtcgacg ccagccaggg gaaagtccgg cgtcgaaggg agaggaaggc 660
 ggcgtgtgca aactaacggt gaccactggc gcccgccgac acgagcagga agcaggcagc 720
 tgcagagcgc agcgcgcaag tgcagaatgc gcgaaagatc cacttgccgcg cggcggggcgc 780
 gcacttgccg gcgcgcccgc gaacagtgcg gaaaggagcg gtgcagacgg cgcgcagtga 840
 cagtggggcg aaagccgcgc agtaagcagc ggcgggggaa ggtatacgcg gtgccggggg 900
 ccgcccacaca cagaagtata cgcggggcga agtggggcgt cgcgcgcggg aagtgcggaa 960
 tggcggggcaa ggaaaggagg agacggaaag agggcgggaa agagagagag agagagtgaa 1020
 aaaagaaaga aagaaagaaa gaaagaaaga aagctcggag ccacgcccg gggagagaga 1080
 gaaatgaaaag cacggcacgg caaagcaaag caaagcagac ccagccagac ccagccgagg 1140
 gaggagcgcg cgcaggaccc gcgcggcagc cgagcgagca cggcgcgcga gcgagcgagc 1200
 gagcagcgcg gcgagcagc aaggcttgcg gcgagcgatc gagcagcga gcgggaagg 1259
 atg agc gcg acc cgc gcg gcg acg agg aca gcg gcg gcg ctg tcc tcg 1307
 Met Ser Ala Thr Arg Ala Ala Thr Arg Thr Ala Ala Ala Leu Ser Ser
 1 5 10 15
 gcg ctg acg acg cct gta aag cag cag cag cag cag cag ctg cgc gta 1355
 Ala Leu Thr Thr Pro Val Lys Gln Gln Gln Gln Gln Gln Leu Arg Val
 20 25 30
 ggc gcg gcg tcg gca cgg ctg gcg gcc gcg gcg ttc tcg tcc ggc acg 1403
 Gly Ala Ala Ser Ala Arg Leu Ala Ala Ala Phe Ser Ser Gly Thr
 35 40 45
 ggc gga gac gcg gcc aag aag gcg gcc gcg gcg agg gcg ttc tcc acg 1451
 Gly Gly Asp Ala Ala Lys Lys Ala Ala Ala Arg Ala Phe Ser Thr
 50 55 60
 gga cgc ggc ccc aac gcg aca cgc gag aag agc tcg ctg gcc acg gtc 1499
 Gly Arg Gly Pro Asn Ala Thr Arg Glu Lys Ser Ser Leu Ala Thr Val
 65 70 75 80
 cag gcg gcg acg gac gat gcg cgc ttc gtc ggc ctg acc ggc gcc caa 1547
 Gln Ala Ala Thr Asp Asp Ala Arg Phe Val Gly Leu Thr Gly Ala Gln
 85 90 95
 atc ttt cat gag ctc atg cgc gag cac cag gtg gac acc atc ttt ggc 1595
 Ile Phe His Glu Leu Met Arg Glu His Gln Val Asp Thr Ile Phe Gly
 100 105 110
 tac cct ggc ggc gcc att ctg ccc gtt ttt gat gcc att ttt gag agt 1643

ES 2 698 473 T3

Tyr	Pro	Gly	Gly	Ala	Ile	Leu	Pro	Val	Phe	Asp	Ala	Ile	Phe	Glu	Ser		
		115				120					125						
gac	gcc	ttc	aag	ttc	att	ctc	gct	cgc	cac	gag	cag	ggc	gcc	ggc	cac		1691
Asp	Ala	Phe	Lys	Phe	Ile	Leu	Ala	Arg	His	Glu	Gln	Gly	Ala	Gly	His		
		130				135					140						
atg	gcc	gag	ggc	tac	gcg	cgc	gcc	acg	ggc	aag	ccc	ggc	ggt	gtc	ctc		1739
Met	Ala	Glu	Gly	Tyr	Ala	Arg	Ala	Thr	Gly	Lys	Pro	Gly	Val	Val	Leu		
145					150					155					160		
gtc	acc	tcg	ggc	cct	gga	gcc	acc	aac	acc	atc	acc	cgc	atc	atg	gat		1787
Val	Thr	Ser	Gly	Pro	Gly	Ala	Thr	Asn	Thr	Ile	Thr	Pro	Ile	Met	Asp		
					165					170					175		
gct	tac	atg	gac	ggt	acg	ccg	ctg	ctc	gtg	ttc	acc	ggc	cag	gtg	cag		1835
Ala	Tyr	Met	Asp	Gly	Thr	Pro	Leu	Leu	Val	Phe	Thr	Gly	Gln	Val	Gln		
			180							185					190		
acc	tot	gct	gtc	ggc	acg	gac	gct	ttc	cag	gag	tgt	gac	att	ggt	ggc		1883
Thr	Ser	Ala	Val	Gly	Thr	Asp	Ala	Phe	Gln	Glu	Cys	Asp	Ile	Val	Gly		
			195				200					205					
atc	agc	cgc	gcg	tgc	acc	aag	tgg	aac	gtc	atg	gtc	aag	gac	gtg	aag		1931
Ile	Ser	Arg	Ala	Cys	Thr	Lys	Trp	Asn	Val	Met	Val	Lys	Asp	Val	Lys		
			210				215				220						
gag	ctc	ccg	cgc	cgc	atc	aat	gag	gcc	ttt	gag	att	gcc	atg	agc	ggc		1979
Glu	Leu	Pro	Arg	Arg	Ile	Asn	Glu	Ala	Phe	Glu	Ile	Ala	Met	Ser	Gly		
225					230					235					240		
cgc	ccg	ggt	ccc	gtg	ctc	gtc	gat	ctt	cct	aag	gat	gtg	acc	gcc	ggt		2027
Arg	Pro	Gly	Pro	Val	Leu	Val	Asp	Leu	Pro	Lys	Asp	Val	Thr	Ala	Val		
					245					250					255		
gag	ctc	aag	gaa	atg	ccc	gac	agc	tcc	ccc	cag	ggt	gct	gtg	cgc	cag		2075
Glu	Leu	Lys	Glu	Met	Pro	Asp	Ser	Ser	Pro	Gln	Val	Ala	Val	Arg	Gln		
			260						265						270		
aag	caa	aag	gtc	gag	ctt	ttc	cac	aag	gag	cgc	att	ggc	gct	cct	ggc		2123
Lys	Gln	Lys	Val	Glu	Leu	Phe	His	Lys	Glu	Arg	Ile	Gly	Ala	Pro	Gly		
			275				280					285					
acg	gcc	gac	ttc	aag	ctc	att	gcc	gag	atg	atc	aac	cgt	gcg	gag	cga		2171
Thr	Ala	Asp	Phe	Lys	Leu	Ile	Ala	Glu	Met	Ile	Asn	Arg	Ala	Glu	Arg		
			290				295				300						
ccc	gtc	atc	tat	gct	ggc	cag	ggt	gtc	atg	cag	agc	ccg	ttg	aat	ggc		2219
Pro	Val	Ile	Tyr	Ala	Gly	Gln	Gly	Val	Met	Gln	Ser	Pro	Leu	Asn	Gly		
305					310					315					320		
ccg	gct	gtg	ctc	aag	gag	ttc	gcg	gag	aag	gcc	aac	att	ccc	gtg	acc		2267
Pro	Ala	Val	Leu	Lys	Glu	Phe	Ala	Glu	Lys	Ala	Asn	Ile	Pro	Val	Thr		
					325					330					335		
acc	acc	atg	cag	ggt	ctc	ggc	ggc	ttt	gac	gag	cgt	agt	ccc	ctc	tcc		2315
Thr	Thr	Met	Gln	Gly	Leu	Gly	Gly	Phe	Asp	Glu	Arg	Ser	Pro	Leu	Ser		
			340					345							350		
ctc	aag	atg	ctc	ggc	atg	cac	ggc	tct	gcc	tac	gcc	aac	tac	tcg	atg		2363
Leu	Lys	Met	Leu	Gly	Met	His	Gly	Ser	Ala	Tyr	Ala	Asn	Tyr	Ser	Met		
			355				360								365		
cag	aac	gcc	gat	ctt	atc	ctg	gcg	ctc	ggt	gcc	cgc	ttt	gat	gat	cgt		2411
Gln	Asn	Ala	Asp	Leu	Ile	Leu	Ala	Leu	Gly	Ala	Arg	Phe	Asp	Asp	Arg		
			370				375								380		
gtg	acg	ggc	cgc	ggt	gac	gcc	ttt	gct	ccg	gag	gct	cgc	cgt	gcc	gag		2459
Val	Thr	Gly	Arg	Val	Asp	Ala	Phe	Ala	Pro	Glu	Ala	Arg	Arg	Ala	Glu		
385					390					395					400		
cgc	gag	ggc	cgc	ggt	ggc	atc	ggt	cac	ttt	gag	att	tcc	ccc	aag	aac		2507
Arg	Glu	Gly	Arg	Gly	Gly	Ile	Val	His	Phe	Glu	Ile	Ser	Pro	Lys	Asn		
					405					410					415		
ctc	cac	aag	gtc	gtc	cag	ccc	acc	gtc	gcg	gtc	ctc	ggc	gac	gtg	gtc		2555
Leu	His	Lys	Val	Val	Gln	Pro	Thr	Val	Ala	Val	Leu	Gly	Asp	Val	Val		
			420							425					430		

ES 2 698 473 T3

gag aac ctc gcc aac gtc acg ccc cac gtg cag cgc cag gag cgc gag 2603
 Glu Asn Leu Ala Asn Val Thr Pro His Val Gln Arg Gln Glu Arg Glu
 435 440 445

cgg tgg ttt gcg cag atc gcc gat tgg aag gag aag cac cct ttt ctg 2651
 Pro Trp Phe Ala Gln Ile Ala Asp Trp Lys Glu Lys His Pro Phe Leu
 450 455 460

ctc gag tct gtt gat tgc gac gac aag gtt ctc aag ccg cag cag gtc 2699
 Leu Glu Ser Val Asp Ser Asp Asp Lys Val Leu Lys Pro Gln Gln Val
 465 470 475 480

ctc acg gag ctt aac aag cag att ctc gag att cag gag aag gac gcc 2747
 Leu Thr Glu Leu Asn Lys Gln Ile Leu Glu Ile Gln Glu Lys Asp Ala
 485 490 495

gac cag gag gtc tac atc acc acg ggc gtc gga agc cac cag atg cag 2795
 Asp Gln Glu Val Tyr Ile Thr Thr Gly Val Gly Ser His Gln Met Gln
 500 505 510

gca gcg cag ttc ctt acc tgg acc aag ccg cgc cag tgg atc tcc tcg 2843
 Ala Ala Gln Phe Leu Thr Trp Thr Lys Pro Arg Gln Trp Ile Ser Ser
 515 520 525

ggt ggc gcc ggc act atg ggc tac ggc ctt ccc tcg gcc att ggc gcc 2891
 Gly Gly Ala Gly Thr Met Gly Tyr Gly Leu Pro Ser Ala Ile Gly Ala
 530 535 540

aag att gcc aag ccc gat gct att gtt att gac atc gat ggt gat gct 2939
 Lys Ile Ala Lys Pro Asp Ala Ile Val Ile Asp Ile Asp Gly Asp Ala
 545 550 555 560

tct tat tcg atg acc ggt atg gaa ttg atc aca gca gcc gaa ttc aag 2987
 Ser Tyr Ser Met Thr Gly Met Glu Leu Ile Thr Ala Ala Glu Phe Lys
 565 570 575

ggt ggc gtg aag att ctt ctt ttg cag aac aac ttt cag gcc atg gtc 3035
 Val Gly Val Lys Ile Leu Leu Leu Gln Asn Asn Phe Gln Gly Met Val
 580 585 590

aag aac tgg cag gat ctc ttt tac gac aag cgc tac tcg gcc acc gcc 3083
 Lys Asn Trp Gln Asp Leu Phe Tyr Asp Lys Arg Tyr Ser Gly Thr Ala
 595 600 605

atg ttc aac ccg cgc ttc gac aag gtc gcc gat gcg atg cgt gcc aag 3131
 Met Phe Asn Pro Arg Phe Asp Lys Val Ala Asp Ala Met Arg Ala Lys
 610 615 620

ggt ctc tac tgc gcg aaa cag tcg gag ctc aag gac aag atc aag gag 3179
 Gly Leu Tyr Cys Ala Lys Gln Ser Glu Leu Lys Asp Lys Ile Lys Glu
 625 630 635 640

ttt ctc gag tac gat gag ggt ccc gtc ctc gag gtt ttc gtg gac 3227
 Phe Leu Glu Tyr Asp Glu Gly Pro Val Leu Leu Glu Val Phe Val Asp
 645 650 655

aag gac acg ctc gtc ttg ccc atg gtc ccc gct ggc ttt ccg ctc cac 3275
 Lys Asp Thr Leu Val Leu Pro Met Val Pro Ala Gly Phe Pro Leu His
 660 665 670

gag atg gtc ctc gag cct cct aag ccc aag gac gcc taa gttctttttt 3324
 Glu Met Val Leu Glu Pro Pro Lys Pro Lys Asp Ala
 675 680

ccatggcggg cgagcgagcg agcgcgcgag cgcgcaagtg cgcaagcgcc ttgccttget 3384
 ttgcttcgct tcgctttgct ttgcttcaca caacctaatg atgaattcaa gttttcttgc 3444
 ttgtcggcga tgcctgcctg ccaaccagcc agccatccgg ccggccgtcc ttgacgcctt 3504
 cgcttcggc gcggccatcg attcaattca cccatccgat acgttccgcc ccctcacgctc 3564
 cgtctgcgca cgaccctgc acgaccagcc caaggccaac gcgccgctca gctcagcttg 3624
 tcgacgagtc gcacgtcaca tatctcagat gcatttggac tgtgagtgtt attatgccac 3684
 tagcacgcaa cgatcttcgg ggtcctcgct cattgcaccc gttcgggcc tgcaggcgtg 3744
 gacgcgagtc gccgccgaga cgctgcagca ggccgctccg acgcgagggc tcgagctcgc 3804
 cgcgcccgcg cgatgtctgc ctggcgcgga ctgatctctg gagcgcaagg aagacacggc 3864
 gacgcgagga ggaccgaaga gagacgctgg ggtatgcagg atatacccgg ggcgggacat 3924
 tcgttcggca tacactcccc cattcgagct tgctcgtcct tggcagagcc gagcgcgaac 3984

ES 2 698 473 T3

ggttccgaac	goggcaagga	ttttggctct	ggtgggtgga	ctccgatcga	ggcgaggtt	4044
ctccgcagg	tctcgcaagg	cggcagtggt	cgttagaaa	agggagtgcc	ggagtcttga	4104
cgcgccctag	ctcactctcc	gccacgcgcg	gcatcgccgc	catgcccgcg	tcccgtctgt	4164
cgctgocgctg	gocgcaaccg	gctgocccag	agtagcagag	tgggacagag	ctcgaggcga	4224
cgcgaaatcgc	togggtgta	agggtttcaa	gggtcgggcg	tcgtcgcgtg	ccaaagtga	4284
aatagtaggg	gggggggggg	gtaccacccc	cgggcaggtt	ctcctcgcca	gcctaagtgc	4344
ctaagggagc	gtaggggttt	cgttgaccag	agaagcggag	aacctgcccg	ggcgcgga	4404
acctatcggc	ggagaacttg	ccagggcgga	ggcagttctc	caatttgccg	acagcgccgc	4464
gccccacgca	ggcgcccgcg	tggcgataca	gcgaggcgac	cgcgccgggc	cgcgtagcga	4524
cacagctgcg	cgcgagtgcg	gctgcgagaa	ggcttctcgc	tggcttggtt	ggggtcgcgg	4584
gtggcaaggg	atggatgccc	aggtacgtcg	gctgocgcgc	gcccagggag	aaaaggacag	4644
acgcgccggc	ctgcgatgcg	agcacgcgat	gcgagcacgc	gatgcgagca	cgcgatgcga	4704
gcacgcgagc	gagcgcccga	gcaaatgcca	cggaaacacg	gttttttgtt	tgggtgatttc	4764
tatgtatgcg	gggagacttc	gatggccgaa	aggggtgcaa	ggccaaaaga	tgctgacagc	4824
ttcgatcggg	ctagcgcgcg	agcaggaag	ggagcaaggg	gcggaattct	tctgccttga	4884
cccgggggat	ccactagtcc	tagagcgccc	gccaccgccc	tggagctcca	attcgcccta	4944
tagtgagctg	tattacgcgc	gctcactggc	cgtcgtttta	caacgtcgtg	actgggaaaa	5004
ccctggcgtt	accacaactta	atcgcccttc	agcacatccc	cctttcgcca	gctggcgtaa	5064
tagcgaagag	gcccgcaccg	atcgcccttc	ccaacagtgc	cgcagcctga	atggcgatg	5124
ggacgcgccc	tgtagcggcg	cattaagcgc	ggcggtgtg	gtggttacgc	gcagcgtgac	5184
cgctacactt	gccagcggcc	tagcgcggcg	tcctttcgc	ttcttcccct	cctttctcgc	5244
cacgttcgccc	ggctttcccc	gtcaagctct	aaatcggggg	ctccctttag	ggttccgatt	5304
tagtgcttta	cgccacctcg	acccccaaaa	acttgattag	ggtgatgggt	cacgtagtgg	5364
gccatcgccc	tgatagacgg	tttttcgccc	tttgacgttg	gagtccacgt	tctttaatag	5424
tggactcttg	ttccaaactg	gaacaacact	caaccctatc	tcggtctatt	cttttgattt	5484
ataagggtat	ttgcccattt	cggcctattg	gttaaaaaat	gagctgattt	aacaaaaatt	5544
taacgcgaat	tttaacaaaa	tattaacgct	tacaatttag	gtggcacttt	tccgggaaat	5604
gtgocgga	cccctatttg	tttatttttc	taaatacatt	caaatatgta	tccgctcatg	5664
agacaataac	ctgataaaat	gcttcaataa	tattgaaaaa	ggaagagtat	gagtatcaaa	5724
catttccggt	tgcgcccttat	tccttttttt	goggcatttt	gccttccctg	ttttgctcac	5784
ccagaaacgc	tgggtgaaagt	aaaagatgct	gaagatcagt	tgggtgcaag	agtgggttac	5844
atcgaactgg	atctcaacag	cggtaaagat	cttgagagtt	ttcgccccga	agaacgtttt	5904
ccaatgatga	gcaactttaa	agttctgcta	tgtggcggcg	tattatcccc	tattgacgcc	5964
gggcaagagc	aactcggctg	ccgcatacac	tattctcaga	atgacttgg	tgagtactca	6024
ccagtcacag	aaaagcatct	tacggatggc	atgacagtaa	gagaattatg	cagtgcctgc	6084
ataaccatga	gtgataaac	tgcggccaac	ttacttctga	caacgatcgg	aggaaccgaag	6144
gagctaaacc	cttttttgca	caacatgggg	gatcatgtaa	ctcgccctga	tcgttgggaa	6204
ccggagctga	atgaagccat	accaaaacgc	gagcgtgaca	ccacgatgcc	tgtagcaatg	6264
gcaacaacgt	tgcgcaaaact	attaactggc	gaactactta	ctctagcttc	ccggcaacaa	6324
ttaatagact	ggatggaggg	ggataaaagt	gcaggaccac	ttctgcgctc	ggcccttccg	6384
gctggctgg	ttattgctga	taaatctgga	gccggtgagc	gtgggtctcg	oggtatcatt	6444
gcagcactgg	ggccagatgg	taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	6504
caggcaacta	tggatgaacg	aaatagacag	atcgctgaga	taggtgcctc	actgattaa	6564
catttggtaac	tgtcagacca	agtttactca	tatatacttt	agattgattt	aaaaactcat	6624
ttttaattta	aaaggatcta	ggtgaagatc	citttttgata	atctcatgac	caaaatccct	6684
taacgtgagt	tttcgttcca	ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	6744
tgagatcctt	tttttctgcg	cgtaatctgc	tgcttgcaaa	caaaaaaac	accgctacca	6804
gcggtgggtt	gtttgocgga	tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	6864
agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	6924
aagaactctg	tagcaccggc	tacatacctc	gctctgctaa	tcctggtacc	agtggctgct	6984
gccagtggcg	ataagtctgt	tcttaccggg	ttggactcaa	gacgatagtt	accggataag	7044
gcgcagcgg	cgggctgaac	ggggggtcgc	tgcacacagc	ccagcttgg	gcgaacgacc	7104
tacaccgaac	tgagatacct	acagcgtgag	ctatgagaaa	gcgccacgct	tcccgaaggg	7164
agaaaggcgg	acaggtatcc	ggtaaagcgc	agggctggaa	caggagagcg	cacgagggag	7224
cctccagggg	gaaacgcctg	gtatctttat	agtcctgtcg	ggtttcgcca	cctctgactt	7284
gagcgtogat	ttttgtgatg	ctcgtcaggg	ggcgggagcc	tatggaaaa	cgccagcaac	7344
gcgccctttt	tacggttctc	ggccttttgc	tggccttttg	ctcacatgct	ctttcctgcg	7404
ttatcccctg	attctgtgga	taaccgtatt	accgctttg	agtgagctga	taccgctcgc	7464
cgcagccgaa	cgaccgagcg	cagcgagtca	gtgagcaggg	aaagcgaaga	gcccacaata	7524

cgcaaacgcg	ctctccccgc	gcgttggcgc	attcattaat	gcagctggca	cgacaggttt	7584
cccgactgga	aaagcggcag	tgagcctaac	gcaattaaat	tgagttagct	cactcattag	7644
gcaccccagg	ctttacactt	tatgcttccg	gctcgtatgt	tgtgtggaat	tgtgagcga	7704
taacaatttc	acacagga	cagctatgac	catgattaoc	ccaagcgcgc	aattaacoc	7764
cactaaaggg	aacaaaagct	gggtaccggg	ccccccctcg	aggtcgacgg	tatcgataag	7824
cttgacgtcg	aattcctgca	gcc				7847

<210> 22
 <211> 684
 <212> PRT
 <213> Schizochytrium sp.

ES 2 698 473 T3

<400> 22

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

ES 2 698 473 T3

				245						250						255
Glu	Leu	Lys	Glu	Met	Pro	Asp	Ser	Ser	Pro	Gln	Val	Ala	Val	Arg	Gln	
			260					265					270			
Lys	Gln	Lys	Val	Glu	Leu	Phe	His	Lys	Glu	Arg	Ile	Gly	Ala	Pro	Gly	
		275					280					285				
Thr	Ala	Asp	Phe	Lys	Leu	Ile	Ala	Glu	Met	Ile	Asn	Arg	Ala	Glu	Arg	
	290					295					300					
Pro	Val	Ile	Tyr	Ala	Gly	Gln	Gly	Val	Met	Gln	Ser	Pro	Leu	Asn	Gly	
305					310					315					320	
Pro	Ala	Val	Leu	Lys	Glu	Phe	Ala	Glu	Lys	Ala	Asn	Ile	Pro	Val	Thr	
				325					330					335		
Thr	Thr	Met	Gln	Gly	Leu	Gly	Gly	Phe	Asp	Glu	Arg	Ser	Pro	Leu	Ser	
			340					345						350		
Leu	Lys	Met	Leu	Gly	Met	His	Gly	Ser	Ala	Tyr	Ala	Asn	Tyr	Ser	Met	
		355					360					365				
Gln	Asn	Ala	Asp	Leu	Ile	Leu	Ala	Leu	Gly	Ala	Arg	Phe	Asp	Asp	Arg	
	370					375						380				
Val	Thr	Gly	Arg	Val	Asp	Ala	Phe	Ala	Pro	Glu	Ala	Arg	Arg	Ala	Glu	
385					390					395					400	
Arg	Glu	Gly	Arg	Gly	Gly	Ile	Val	His	Phe	Glu	Ile	Ser	Pro	Lys	Asn	
				405					410					415		
Leu	His	Lys	Val	Val	Gln	Pro	Thr	Val	Ala	Val	Leu	Gly	Asp	Val	Val	
			420					425					430			
Glu	Asn	Leu	Ala	Asn	Val	Thr	Pro	His	Val	Gln	Arg	Gln	Glu	Arg	Glu	
		435				440						445				
Pro	Trp	Phe	Ala	Gln	Ile	Ala	Asp	Trp	Lys	Glu	Lys	His	Pro	Phe	Leu	
	450					455					460					
Leu	Glu	Ser	Val	Asp	Ser	Asp	Asp	Lys	Val	Leu	Lys	Pro	Gln	Gln	Val	
465				470						475					480	
Leu	Thr	Glu	Leu	Asn	Lys	Gln	Ile	Leu	Glu	Ile	Gln	Glu	Lys	Asp	Ala	
				485					490					495		
Asp	Gln	Glu	Val	Tyr	Ile	Thr	Thr	Gly	Val	Gly	Ser	His	Gln	Met	Gln	
			500					505					510			
Ala	Ala	Gln	Phe	Leu	Thr	Trp	Thr	Lys	Pro	Arg	Gln	Trp	Ile	Ser	Ser	
		515					520					525				
Gly	Gly	Ala	Gly	Thr	Met	Gly	Tyr	Gly	Leu	Pro	Ser	Ala	Ile	Gly	Ala	
	530					535					540					
Lys	Ile	Ala	Lys	Pro	Asp	Ala	Ile	Val	Ile	Asp	Ile	Asp	Gly	Asp	Ala	
545					550					555					560	

ES 2 698 473 T3

Ser Tyr Ser Met Thr Gly Met Glu Leu Ile Thr Ala Ala Glu Phe Lys
 565 570 575
 Val Gly Val Lys Ile Leu Leu Leu Gln Asn Asn Phe Gln Gly Met Val
 580 585 590
 Lys Asn Trp Gln Asp Leu Phe Tyr Asp Lys Arg Tyr Ser Gly Thr Ala
 595 600 605
 Met Phe Asn Pro Arg Phe Asp Lys Val Ala Asp Ala Met Arg Ala Lys
 610 615 620
 Gly Leu Tyr Cys Ala Lys Gln Ser Glu Leu Lys Asp Lys Ile Lys Glu
 625 630 635 640
 Phe Leu Glu Tyr Asp Glu Gly Pro Val Leu Leu Glu Val Phe Val Asp
 645 650 655
 Lys Asp Thr Leu Val Leu Pro Met Val Pro Ala Gly Phe Pro Leu His
 660 665 670
 Glu Met Val Leu Glu Pro Pro Lys Pro Lys Asp Ala
 675 680

5 <210> 23
 <211> 7847
 <212> ADN
 <213> Schizochytrium sp.

10 <220>
 <221> prim_transcript
 <222> (1)..(1259)
 <223>

15 <220>
 <221> CDS
 <222> (1260)..(3314)
 <223>

20 <220>
 <221> prim_transcript
 <222> (3315)..(4887)
 <223>

25 <220>
 <221> mutación
 <222> (1834)..(1835)
 <223>

30 <220>
 <221> mutación
 <222> (3042)..(3044)
 <223>

35 <400> 23
 ttgtcgacag caacttgcaa gttatacgcg accaccaggc aatctcagca cgcccagcga 60
 gcaocggagct tgcgaagagg gtttacaagt cgtcgttcat tgcgtctcaa gctttgcctc 120
 aacgcaacta ggcccaggcc tactttcact gtgtcttgtc ttgcctttca caccgaccga 180
 gtgtgcacaa ccgtgttttg cacaaagcgc aagatgctca ctcgactgtg aagaaaggtt 240

ES 2 698 473 T3

```

gcgcgcaagc gactgcgact gcgaggatga ggatgactgg cagcctgttc aaaaactgaa 300
aatccgcgat gggtcagatgc ctttcgcgca tgacgcctgc gagagacaag ttaactcgtg 360
tcactggcat gtcctagcat ctttacgcga gcaaaattca atcgctttat tttttcagtt 420
tcgtaacctt ctgcgaaccg cgaatcgccg tttcagcctg actaatctgc agctgctgtg 480
cactgtcagt cagtcagtca gtcgtgcccg ctgttccagc accgaggtcg cgcgtcgccg 540
cgcttgacc gctgctgcta ctgctagtgg cacggcaggt aggagcttgt tgccggaaca 600
ccagcagccg ccagtcgacg ccagccaggg gaaagtccgg cgtcgaaggg agaggaaggg 660
ggcgtgtgca aactaacgtt gaccactggc gcccgccgac acgagcagga agcagggcagc 720
tgcagagcgc agcgcgcaag tgcagaatgc gcgaaagatc cacttgccgc cggcgggccc 780
gcacttgccg gcgcccgcgc gaacagtgcg gaaaggagcg gtgcagacgg cgcgcagtg 840
cagtgggccc aaagcccgcg agtaagcagc ggcggggaac ggtatacgca gtgcccgggg 900
ccgcccacac cagaagtata cgcgggcccga agtggggcgt cgcgcgcccgg aagtgcggaa 960
tggcgggcaa ggaaaggagg agacggaaag agggcgggaa agagagagag agagagtga 1020
aaaagaaaga aagaaagaaa gaaagaaaga aagctcggag ccacgcccgc gggagagaga 1080
gaaatgaaag cacggcacgg caaagcaaag caaagcagac ccagccagac ccagccaggg 1140
gaggagcgcg cgcaggacc cgcgggcccga cgaagcagca cgcgcgcccga gcgagcagac 1200
gagcagcgcg cgcagcagac aaggcttgcg cgcagcagc gagcagcagca gcgggaagg 1259
atg agc gcg acc cgc gcg gcg acg agg aca gcg gcg gcg ctg tcc tcg 1307
Met Ser Ala Thr Arg Ala Ala Thr Arg Thr Ala Ala Ala Leu Ser Ser
1 5 10 15
gcg ctg acg acg cct gta aag cag cag cag cag cag cag ctg cgc gta 1355
Ala Leu Thr Thr Pro Val Lys Gln Gln Gln Gln Gln Gln Leu Arg Val
20 25 30
ggc gcg gcg tcg gca cgg ctg gcg gcc gcg gcg ttc tcg tcc ggc acg 1403
Gly Ala Ala Ser Ala Arg Leu Ala Ala Ala Ala Phe Ser Ser Gly Thr
35 40 45
ggc gga gac gcg gcc aag aag gcg gcc gcg gcg agg gcg ttc tcc acg 1451
Gly Gly Asp Ala Ala Lys Lys Ala Ala Ala Ala Arg Ala Phe Ser Thr
50 55 60
gga cgc gcc ccc aac gcg aca cgc gag aag agc tcg ctg gcc acg gtc 1499
Gly Arg Gly Pro Asn Ala Thr Arg Glu Lys Ser Ser Leu Ala Thr Val
65 70 75 80
cag gcg gcg acg gac gat gcg cgc ttc gtc gcc ctg acc ggc gcc caa 1547
Gln Ala Ala Thr Asp Ala Arg Phe Val Gly Leu Thr Gly Ala Gln
85 90 95
atc ttt cat gag ctc atg cgc gag cac cag gtg gac acc atc ttt ggc 1595
Ile Phe His Glu Leu Met Arg Glu His Gln Val Asp Thr Ile Phe Gly
100 105 110
tac cct gcc ggc gcc att ctg ccc gtt ttt gat gcc att ttt gag agt 1643
Tyr Pro Gly Gly Ala Ile Leu Pro Val Phe Asp Ala Ile Phe Glu Ser
115 120 125
gac gcc ttc aag ttc att ctc gct cgc cac gag cag gcc gcc gcc cac 1691
Asp Ala Phe Lys Phe Ile Leu Ala Arg His Glu Gln Gly Ala Gly His
130 135 140
atg gcc gag ggc tac gcg cgc gcc acg ggc aag ccc gcc gtt gtc ctc 1739
Met Ala Glu Gly Tyr Ala Arg Ala Thr Gly Lys Pro Gly Val Val Leu
145 150 155 160
gtc acc tcg ggc cct gga gcc acc aac acc atc acc ccg atc atg gat 1787
Val Thr Ser Gly Pro Gly Ala Thr Asn Thr Ile Thr Pro Ile Met Asp
165 170 175
gct tac atg gac ggt acg ccg ctg ctc gtg ttc acc gcc cag gtg cag 1835
Ala Tyr Met Asp Gly Thr Pro Leu Leu Val Phe Thr Gly Gln Val Gln
180 185 190
acc tct gct gtc ggc acg gac gct ttc cag gag tgt gac att gtt ggc 1883
Thr Ser Ala Val Gly Thr Asp Ala Phe Gln Glu Cys Asp Ile Val Gly
195 200 205
atc agc cgc gcg tgc acc aag tgg aac gtc atg gtc aag gac gtg aag 1931
Ile Ser Arg Ala Cys Thr Lys Trp Asn Val Met Val Lys Asp Val Lys
210 215 220

```

ES 2 698 473 T3

gag ctc ccg cgc cgc atc aat gag gcc ttt gag att gcc atg agc ggc 1979
 Glu Leu Pro Arg Arg Ile Asn Glu Ala Phe Glu Ile Ala Met Ser Gly
 225 230 235 240
 cgc ccg ggt ccc gtg ctc gtc gat ctt cct aag gat gtg acc gcc gtt 2027
 Arg Pro Gly Pro Val Leu Val Asp Leu Pro Lys Asp Val Thr Ala Val
 245 250 255
 gag ctc aag gaa atg ccc gac agc tcc ccc cag gtt gct gtg cgc cag 2075
 Glu Leu Lys Glu Met Pro Asp Ser Ser Pro Gln Val Ala Val Arg Gln
 260 265 270
 aag caa aag gtc gag ctt ttc cac aag gag cgc att ggc gct cct ggc 2123
 Lys Gln Lys Val Glu Leu Phe His Lys Glu Arg Ile Gly Ala Pro Gly
 275 280 285
 acg gcc gac ttc aag ctc att gcc gag atg atc aac cgt cgc gag cga 2171
 Thr Ala Asp Phe Lys Leu Ile Ala Glu Met Ile Asn Arg Ala Glu Arg
 290 295 300
 ccc gtc atc tat gct ggc cag ggt gtc atg cag agc ccg ttg aat ggc 2219
 Pro Val Ile Tyr Ala Gly Gln Gly Val Met Gln Ser Pro Leu Asn Gly
 305 310 315 320
 ccg gct gtg ctc aag gag ttc gcg gag aag gcc aac att ccc gtg acc 2267
 Pro Ala Val Leu Lys Glu Phe Ala Glu Lys Ala Asn Ile Pro Val Thr
 325 330 335
 acc acc atg cag ggt ctc ggc gcc ttt gac gag cgt agt ccc ctc tcc 2315
 Thr Thr Met Gln Gly Leu Gly Gly Phe Asp Glu Arg Ser Pro Leu Ser
 340 345 350
 ctc aag atg ctc gcc atg cac gcc tct gcc tac gcc aac tac tcg atg 2363
 Leu Lys Met Leu Gly Met His Gly Ser Ala Tyr Ala Asn Tyr Ser Met
 355 360 365
 cag aac gcc gat ctt atc ctg gcg ctc ggt gcc cgc ttt gat gat cgt 2411
 Gln Asn Ala Asp Leu Ile Leu Ala Leu Gly Ala Arg Phe Asp Asp Arg
 370 375 380
 gtg acg ggc cgc gtt gac gcc ttt gct ccg gag gct cgc cgt gcc gag 2459
 Val Thr Gly Arg Val Asp Ala Phe Ala Pro Glu Ala Arg Arg Ala Glu
 385 390 395 400
 cgc gag ggc cgc ggt ggc atc gtt cac ttt gag att tcc ccc aag aac 2507
 Arg Glu Gly Arg Gly Ile Val His Phe Glu Ile Ser Pro Lys Asn
 405 410 415
 ctc cac aag gtc gtc cag ccc acc gtc gcg gtc ctc ggc gac gtg gtc 2555
 Leu His Lys Val Val Gln Pro Thr Val Ala Val Leu Gly Asp Val Val
 420 425 430
 gag aac ctc gcc aac gtc acg ccc cac gtg cag cgc cag gag cgc gag 2603
 Glu Asn Leu Ala Asn Val Thr Pro His Val Gln Arg Gln Glu Arg Glu
 435 440 445
 ccg tgg ttt gcg cag atc gcc gat tgg aag gag aag cac cct ttt ctg 2651
 Pro Trp Phe Ala Gln Ile Ala Asp Trp Lys Glu Lys His Pro Phe Leu
 450 455 460
 ctc gag tct gtt gat tcg gac gac aag gtt ctc aag ccg cag cag gtc 2699
 Leu Glu Ser Val Asp Ser Asp Asp Lys Val Leu Lys Pro Gln Gln Val
 465 470 475 480
 ctc acg gag ctt aac aag cag att ctc gag att cag gag aag gac gcc 2747
 Leu Thr Glu Leu Asn Lys Gln Ile Leu Glu Ile Gln Glu Lys Asp Ala
 485 490 495
 gac cag gag gtc tac atc acc acg ggc gtc gga agc cac cag atg cag 2795
 Asp Gln Glu Val Tyr Ile Thr Thr Gly Val Gly Ser His Gln Met Gln
 500 505 510
 gca gcg cag ttc ctt acc tgg acc aag ccg cgc cag tgg atc tcc tcg 2843
 Ala Ala Gln Phe Leu Thr Trp Thr Lys Pro Arg Gln Trp Ile Ser Ser
 515 520 525
 ggt ggc gcc ggc act atg ggc tac ggc ctt ccc tcg gcc att ggc gcc 2891
 Gly Gly Ala Gly Thr Met Gly Tyr Gly Leu Pro Ser Ala Ile Gly Ala

ES 2 698 473 T3

530	535	540	
aag att gcc aag ccc gat gct att gtt att gac atc gat ggt gat gct			2939
Lys Ile Ala Lys Pro Asp Ala Ile Val Ile Asp Ile Asp Gly Asp Ala			
545	550	555	560
tct tat tcg atg acc ggt atg gaa ttg atc aca gca gcc gaa ttc aag			2987
Ser Tyr Ser Met Thr Gly Met Glu Leu Ile Thr Ala Ala Glu Phe Lys			
565	570	575	
ggt ggc gtg aag att ctt ctt ttg cag aac aac ttt cag ggc atg gtc			3035
Val Gly Val Lys Ile Leu Leu Leu Gln Asn Asn Phe Gln Gly Met Val			
580	585	590	
aag aac gtt cag gat ctc ttt tac gac aag cgc tac tcg ggc acc gcc			3083
Lys Asn Val Gln Asp Leu Phe Tyr Asp Lys Arg Tyr Ser Gly Thr Ala			
595	600	605	
atg ttc aac ccg cgc ttc gac aag gtc gcc gat gcg atg cgt gcc aag			3131
Met Phe Asn Pro Arg Phe Asp Lys Val Ala Asp Ala Met Arg Ala Lys			
610	615	620	
ggt ctc tac tgc gcg aaa cag tcg gag ctc aag gac aag atc aag gag			3179
Gly Leu Tyr Cys Ala Lys Gln Ser Glu Leu Lys Asp Lys Ile Lys Glu			
625	630	635	640
ttt ctc gag tac gat gag ggt ccc gtc ctc ctc gag gtt ttc gtg gac			3227
Phe Leu Glu Tyr Asp Glu Gly Pro Val Leu Leu Glu Val Phe Val Asp			
645	650	655	
aag gac acg ctc gtc ttg ccc atg gtc ccc gct ggc ttt ccg ctc cac			3275
Lys Asp Thr Leu Val Leu Pro Met Val Pro Ala Gly Phe Pro Leu His			
660	665	670	
gag atg gtc ctc gag cct cct aag ccc aag gac gcc taa gttctttttt			3324
Glu Met Val Leu Glu Pro Pro Lys Pro Lys Asp Ala			
675	680		
ccatggcggg cgagcgcgag agcgcgcgag cgcgcaagt gcgaagcgcc ttgccttgc			3384
ttgcttcgct tcgctttgct ttgcttcaca caacctaat atgaattcaa gttttcttgc			3444
ttgtcggcga tgcctgcctg ccaaccagcc agccatccgg ccggccgctc ttgacgcctt			3504
cgcttcgggc gggccatcg attcaattca cccatccgat acgttcggcc ccctcagctc			3564
cgctcgcgca cgaccctgc acgaccacgc caaggccaac gcgccgctca gctcagcttg			3624
tcgacgagtc gcacgtcaca tatctcagat gcatttgac tgtgagtgtt attatgccac			3684
tagcagcaaa cgatcttcgg ggtcctcgtt cattgcatcc gttcggggcc tcgagcgtg			3744
gacgcgagtc gccgcgaga cgctgcagca gcccgctccg acgcgagggc tcgagctcgc			3804
cgcgcccgcg cgatgtctgc ctggcgccga ctgatctctg gagcgcaagg aagacacggc			3864
gacgcgagga ggaccgaaga gagacgctgg ggtatgcagg atatacccgg ggcgggacat			3924
tcgctccgca tacactcccc cattcgagct tgcctgtcct tggcagagcc gagcgcgaac			3984
ggttcogaac gcggcaagga ttttgctct ggtgggtgga ctccgatcga ggcgcaggtt			4044
ctccgcaggt tctcgcaggc cggcagtggt cgttagaaat agggagtgcc ggagtcttga			4104
cgcgcttag ctcaactctcc gccacgcgc gcacgcgcgc catgccgcgc tcccgtctgt			4164
cgctgcgctg gccgcgaccg gctgcgcca agtacgacag tgggacagag ctccgagcga			4224
cgcaaatcgc tcgggttgta agggtttcaa ggttcggcg tcgtcgcgtg ccaaagtga			4284
aatagtaggg gggggggggg gtaccacccc cgggcaggtt ctctcgcga gcctaagtgc			4344
ctaagggagc gtaggggttt cgttgaccag agaagcggg aaacctgccg ggcgcggaga			4404
acctatcggc ggagaacttg ccagggcgcga ggcagttctc caatttgcg acagcggcg			4464
gccacgcgca ggcggccgcg tggcgataca gcgagggcgc cgcgccggggc cgcgtggcga			4524
cacagctcgc cgcgagtcg gctgcgagaa gcttctcgc tggcttggtt ggggtcgcgg			4584
gtggcagggg atggatgccc aggtacgtcg gcgtgcgcgc gccagggag aaaaggacag			4644
acgcgcgggc ctgcgatgc agcacgcgat gcgagcacgc gatgcgagca cgcgatgcga			4704
gcacgcgagc gagcgcocga gcaaatgcca cggaacacgc gttttttgtt tgggtatttc			4764
tatgtatgcg gggagacttc gatggccgaa aggggtgcaa gcccaaaaaga tgctgacagc			4824
ttcgatcggc ctacggcgc agcaggaagg ggagcaagg gcggaattct tctgccttga			4884
cccgggggat ccaactagttc tagagcggcc gccaccgcgg tggagctcca attcgcctta			4944
tagtgagtcg tattacgcgc gctcactggc cgtcgtttta caacgtcgtg actgggaaaa			5004
ccctggcgtt acccaactta atcgccttgc agcacatccc cttttcgcca gctggcgtaa			5064
tagcgaagg gccccaccg atcgccttc ccaacagttg cgcagcctga atggcgaatg			5124
ggacgcgccc tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac			5184

```

cgctacactt gccagcgccc tagcgcccgc tcctttogct ttcttccctt cctttctcgc 5244
cacgttcgcc ggctttcccc gtcaagctct aaatcggggg ctccctttag ggttccgatt 5304
tagtgcttta oggcacctcg accccaaaaa acttgattag ggtgatggtt cacgtagtgg 5364
gccatcgccc tgatagacgg tttttcgccc tttgacgttg gagtccacgt tctttaatag 5424
tggactcttg ttccaaactg gaacaacact caaccctatc tgggtctatt cttttgattt 5484
ataagggatt ttgccgattt cggcctattg gttaaaaaat gagctgattt aacaaaaatt 5544
taacgcgaat tttaacaaaa tattaacgct tacaatttag gtggcacttt tcggggaatt 5604
gtgcgcgaa cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg 5664
agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa 5724
catttccgtg tcggccttat tccctttttt ggggcatttt gccttcctgt ttttgctcac 5784
ccagaaacgc tgggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac 5844
atcgaaactg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaactgttt 5904
ccaatgatga gcacttttaa agttctgcta tgtggcggcg tattatcccc tattgaagcc 5964
gggcaagagc aactcggctg ccgcatacac tattctcaga atgacttggg tgagtactca 6024
ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtctgccc 6084
ataaccatga gtgataaac tgcggccaac ttacttctga caacgatcgg aggaccgaag 6144
gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga tcggtgggaa 6204
ccggagctga atgaagccat accaaaagac gagcgtgaca ccacgatgcc tgtagcaatg 6264
gcaacaacgc tgcgcaact attaactggc ctaactactta ctctagcttc ccggcaacaa 6324
ttaatagact ggatggaggc ggataaagt ttgaggaccac ttctgcgctc ggcccttccg 6384
gctggctggg ttattgctga taaatctgga gccgggtgagc gtgggtctcg cggtatcatt 6444
gcagcaactg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 6504
caggcaacta tggatgaacg aatagacag atcgtgaga taggtgcctc actgattaag 6564
cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaaacttcat 6624
ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaaacctc 6684
taacgtgagt tttcgttcca ctgagcgtca gaccocgtag aaaagatcaa aggatcttct 6744
tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaacc accgctacca 6804
gcggtgggtt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt aactggcttc 6864
agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg ccaccacttc 6924
aagaactctg tagcaccgcc tacatacctc gctctgctaa tccgtttacc agtggctgct 6984
gccagtggcg ataagtctgt tottaccggg ttggactcaa gacgatagtt accggataag 7044
gcgcaagcgg cgggctgaac ggggggttcg tgcacacagc ccagcttggg gcgaaacgacc 7104
tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct tccggaagg 7164
agaaaggcgg acaggatccc ggtaagcggc agggctcggaa caggagagcg cacgagggag 7224
cttccagggg gaaacgcctg gtatctttat agtccctgct ggtttcgcca cctctgactt 7284
gagcgtcgat tttttgtgat ctcgctcagg gggcggagcc tatggaaaaa cgccagcaac 7344
gcggcctttt tacggttctt ggccctttgc tggccttttg ctacatggt ctttctgog 7404
ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc 7464
cgcagccgaa cgaccgagcg cagcagatca gtgagcgagg aagcggaaga gcgcccaata 7524
cgcaaacgcg cctccccgcg cggttggccg atcattaat gcagctggca cgacaggttt 7584
cccgaactga aagcgggcag tgagcgcaac gcaattaatg tgagttagct cactcattag 7644
gcaccocagc ctttacactt tatgcttccg gctcgtatgt tgtgtggaat tgtgagcggg 7704
taacaatttc acacagggaa cagctatgac catgattacg ccaagcgcgc aattaacctt 7764
cactaaaggg aacaaaagct gggtaaccgg cccccctcg aggtcgacgg tatcgataag 7824
cttgacgtcg aattcctgca gcc 7847

```

<210> 24

<211> 684

5 <212> PRT

<213> Schizochytrium sp.

<400> 24

Met Ser Ala Thr Arg Ala Ala Thr Arg Thr Ala Ala Ala Leu Ser Ser
1 5 10 15

Ala Leu Thr Thr Pro Val Lys Gln Gln Gln Gln Gln Gln Leu Arg Val
20 25 30

Gly Ala Ala Ser Ala Arg Leu Ala Ala Ala Ala Phe Ser Ser Gly Thr

10

ES 2 698 473 T3

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

ES 2 698 473 T3

Leu Lys Met Leu Gly Met His Gly Ser Ala Tyr Ala Asn Tyr Ser Met
 355 360 365

Gln Asn Ala Asp Leu Ile Leu Ala Leu Gly Ala Arg Phe Asp Asp Arg
 370 375 380

Val Thr Gly Arg Val Asp Ala Phe Ala Pro Glu Ala Arg Arg Ala Glu
 385 390 395 400

Arg Glu Gly Arg Gly Gly Ile Val His Phe Glu Ile Ser Pro Lys Asn
 405 410 415

Leu His Lys Val Val Gln Pro Thr Val Ala Val Leu Gly Asp Val Val
 420 425 430

Glu Asn Leu Ala Asn Val Thr Pro His Val Gln Arg Gln Glu Arg Glu
 435 440 445

Pro Trp Phe Ala Gln Ile Ala Asp Trp Lys Glu Lys His Pro Phe Leu
 450 455 460

Leu Glu Ser Val Asp Ser Asp Asp Lys Val Leu Lys Pro Gln Gln Val
 465 470 475 480

Leu Thr Glu Leu Asn Lys Gln Ile Leu Glu Ile Gln Glu Lys Asp Ala
 485 490 495

Asp Gln Glu Val Tyr Ile Thr Thr Gly Val Gly Ser His Gln Met Gln
 500 505 510

Ala Ala Gln Phe Leu Thr Trp Thr Lys Pro Arg Gln Trp Ile Ser Ser
 515 520 525

Gly Gly Ala Gly Thr Met Gly Tyr Gly Leu Pro Ser Ala Ile Gly Ala
 530 535 540

Lys Ile Ala Lys Pro Asp Ala Ile Val Ile Asp Ile Asp Gly Asp Ala
 545 550 555 560

Ser Tyr Ser Met Thr Gly Met Glu Leu Ile Thr Ala Ala Glu Phe Lys
 565 570 575

Val Gly Val Lys Ile Leu Leu Leu Gln Asn Asn Phe Gln Gly Met Val
 580 585 590

Lys Asn Val Gln Asp Leu Phe Tyr Asp Lys Arg Tyr Ser Gly Thr Ala
 595 600 605

Met Phe Asn Pro Arg Phe Asp Lys Val Ala Asp Ala Met Arg Ala Lys
 610 615 620

Gly Leu Tyr Cys Ala Lys Gln Ser Glu Leu Lys Asp Lys Ile Lys Glu
 625 630 635 640

Phe Leu Glu Tyr Asp Glu Gly Pro Val Leu Leu Glu Val Phe Val Asp
 645 650 655

Lys Asp Thr Leu Val Leu Pro Met Val Pro Ala Gly Phe Pro Leu His
 660 665 670

Glu Met Val Leu Glu Pro Pro Lys Pro Lys Asp Ala
 675 680

5 <210> 25
 <211> 18
 <212> ADN
 <213> Schizochytrium sp.

10 <400> 25
 gttgaccagt gccgttcc 18

ES 2 698 473 T3

<210> 26
 <211> 18
 <212> ADN
 <213> Schizochytrium sp.
 5
 <400> 26
 cgaagtcac gcagttgc 18
 <210> 27
 <211> 31
 <212> ADN
 <213> Schizochytrium sp.
 10
 <400> 27
 gcgccatgg gacgtcaggt ggcactttc g 31
 <210> 28
 <211> 36
 <212> ADN
 <213> Schizochytrium sp.
 20
 <400> 28
 gcgccatgg ccgcggaag cagcagatta cgcgca 36
 <210> 29
 <211> 1263
 <212> ADN
 <213> Caenorhabditis elegans
 25
 <220>
 <221> CDS
 <222> (23)..(1228)
 <223>
 30
 <400> 29
 gcggcgcga attcagatct cc atg gtc gct cac tcg tcg gag ggt ctc tcg 52
 Met Val Ala His Ser Ser Glu Gly Leu Ser
 1 5 10
 gcc acc gcc ccg gtc acc ggc ggc gac gtc ctc gtc gac gcc cgc gcc 100
 Ala Thr Ala Pro Val Thr Gly Gly Asp Val Leu Val Asp Ala Arg Ala
 15 20 25
 tcg ctc gag gag aag gag gcc ccg cgc gac gtc aac gcc aac acc aag 148
 Ser Leu Glu Glu Lys Glu Ala Pro Arg Asp Val Asn Ala Asn Thr Lys
 30 35 40
 cag gcc acc acc gag gag ccc cgc atc cag ctg ccc acc gtc gac gcc 196
 Gln Ala Thr Thr Glu Glu Pro Arg Ile Gln Leu Pro Thr Val Asp Ala
 45 50 55
 ttc cgc cgc gcc atc ccc gcc cac tgc ttc gag cgc gac ctc gtc aag 244

ES 2 698 473 T3

Phe Arg Arg Ala Ile Pro Ala His Cys Phe Glu Arg Asp Leu Val Lys
60 65 70
tcg atc cgc tac ctc gtc cag gac ttc gcc gcc ctc acc atc ctc tac 292
Ser Ile Arg Tyr Leu Val Gln Asp Phe Ala Ala Leu Thr Ile Leu Tyr
75 80 85 90
ttc gcc ctc ccc gcc ttc gag tac ttc ggc ctc ttc ggc tac ctc gtc 340
Phe Ala Leu Pro Ala Phe Glu Tyr Phe Gly Leu Phe Gly Tyr Leu Val
95 100 105
tgg aac atc ttc atg ggc gtc ttc ggc ttc gcc ctc ttc gtc gtc ggc 388
Trp Asn Ile Phe Met Gly Val Phe Gly Phe Ala Leu Phe Val Val Gly
110 115 120
cac gac tgc ctc cac gga agc ttc tcg gac aac cag aac ctc aac gac 436
His Asp Cys Leu His Gly Ser Phe Ser Asp Asn Gln Asn Leu Asn Asp
125 130 135
ttc atc ggc cac atc gcc ttc tcg ccc ctc ttc tcg ccc tac ttc ccc 484
Phe Ile Gly His Ile Ala Phe Ser Pro Leu Phe Ser Pro Tyr Phe Pro
140 145 150
tgg cag aag tcg cac aag ctc cac cac gcc ttc acc aac cac atc gac 532
Trp Gln Lys Ser His Lys Leu His His Ala Phe Thr Asn His Ile Asp
155 160 165 170
aag gac cac ggc cac gtc tgg atc cag gac aag gac tgg gag gcc atg 580
Lys Asp His Gly His Val Trp Ile Gln Asp Lys Asp Trp Glu Ala Met
175 180 185
ccc tcg tgg aag cgc tgg ttc aac ccc atc ccc ttc tcg ggc tgg ctc 628
Pro Ser Trp Lys Arg Trp Phe Asn Pro Ile Pro Phe Ser Gly Trp Leu
190 195 200
aag tgg ttc ccc gtc tac acc ctc ttc ggc ttc tgc gac ggc tcg cac 676
Lys Trp Phe Pro Val Tyr Thr Leu Phe Gly Phe Cys Asp Gly Ser His
205 210 215
ttc tgg ccc tac tcg tcg ctc ttc gtc cgc aac tcg gac cgc gtc cag 724
Phe Trp Pro Tyr Ser Ser Leu Phe Val Arg Asn Ser Asp Arg Val Gln
220 225 230
tgc gtg atc agc ggc atc tgc tgc gtc tgc gcc tac atc gcc ctc 772
Cys Val Ile Ser Gly Ile Cys Cys Cys Val Cys Ala Tyr Ile Ala Leu
235 240 245 250
acc atc gcc ggc tcg tac tcg aac tgg ttc tgg tac tac tgg gtc ccc 820
Thr Ile Ala Gly Ser Tyr Ser Asn Trp Phe Trp Tyr Tyr Trp Val Pro
255 260 265
ctc tcg ttc ttc ggc ctc atg ctc gtc atc gtc acc tac ctg cag cac 868
Leu Ser Phe Phe Gly Leu Met Leu Val Ile Val Thr Tyr Leu Gln His
270 275 280
gtc gac gac gtc gcc gag gtc tac gag gcc gac gag tgg tcg ttc gtc 916
Val Asp Asp Val Ala Glu Val Tyr Glu Ala Asp Glu Trp Ser Phe Val
285 290 295
cgc ggc cag acc cag acc atc gac cgc tac tac ggc ctc ggc ctc gac 964
Arg Gly Gln Thr Gln Thr Ile Asp Arg Tyr Tyr Gly Leu Gly Leu Asp
300 305 310
acc acc atg cac cac atc acc gac ggc cac gtg gcc cac cac ttc ttc 1012
Thr Thr Met His His Ile Thr Asp Gly His Val Ala His His Phe Phe
315 320 325 330
aac aag atc ccg cac tac cac ctc atc gag gcc acc gag ggc gtc aag 1060
Asn Lys Ile Pro His Tyr His Leu Ile Glu Ala Thr Glu Gly Val Lys
335 340 345
aag gtc ctc gag ccc ctc tcg gac acc cag tac ggc tac aag tcg cag 1108
Lys Val Leu Glu Pro Leu Ser Asp Thr Gln Tyr Gly Tyr Lys Ser Gln
350 355 360
gtc aac tac gac ttc ttc gcc cgc ttc ctc tgg ttc aac tac aag ctc 1156
Val Asn Tyr Asp Phe Phe Ala Arg Phe Leu Trp Phe Asn Tyr Lys Leu
365 370 375
gac tac ctc gtg cac aag acc gcc ggc atc atg cag ttc cgc acc acc 1204
Asp Tyr Leu Val His Lys Thr Ala Gly Ile Met Gln Phe Arg Thr Thr
380 385 390
ctc gag gag aag gcc aag gcc aag taacccgggg gtacccttaa ggcatgcgcg 1258
Leu Glu Glu Lys Ala Lys Ala Lys
395 400
gccgc 1263

5

<210> 30
<211> 402
<212> PRT
<213> Caenorhabditis elegans

ES 2 698 473 T3

<400> 30

Met Val Ala His Ser Ser Glu Gly Leu Ser Ala Thr Ala Pro Val Thr
1 5 10 15
Gly Gly Asp Val Leu Val Asp Ala Arg Ala Ser Leu Glu Glu Lys Glu
20 25 30
Ala Pro Arg Asp Val Asn Ala Asn Thr Lys Gln Ala Thr Thr Glu Glu
35 40 45
Pro Arg Ile Gln Leu Pro Thr Val Asp Ala Phe Arg Arg Ala Ile Pro
50 55 60
Ala His Cys Phe Glu Arg Asp Leu Val Lys Ser Ile Arg Tyr Leu Val
65 70 75 80
Gln Asp Phe Ala Ala Leu Thr Ile Leu Tyr Phe Ala Leu Pro Ala Phe
85 90 95
Glu Tyr Phe Gly Leu Phe Gly Tyr Leu Val Trp Asn Ile Phe Met Gly
100 105 110
Val Phe Gly Phe Ala Leu Phe Val Val Gly His Asp Cys Leu His Gly
115 120 125
Ser Phe Ser Asp Asn Gln Asn Leu Asn Asp Phe Ile Gly His Ile Ala
130 135 140
Phe Ser Pro Leu Phe Ser Pro Tyr Phe Pro Trp Gln Lys Ser His Lys
145 150 155 160
Leu His His Ala Phe Thr Asn His Ile Asp Lys Asp His Gly His Val
165 170 175
Trp Ile Gln Asp Lys Asp Trp Glu Ala Met Pro Ser Trp Lys Arg Trp
180 185 190
Phe Asn Pro Ile Pro Phe Ser Gly Trp Leu Lys Trp Phe Pro Val Tyr
195 200 205
Thr Leu Phe Gly Phe Cys Asp Gly Ser His Phe Trp Pro Tyr Ser Ser
210 215 220
Leu Phe Val Arg Asn Ser Asp Arg Val Gln Cys Val Ile Ser Gly Ile
225 230 235 240
Cys Cys Cys Val Cys Ala Tyr Ile Ala Leu Thr Ile Ala Gly Ser Tyr
245 250 255
Ser Asn Trp Phe Trp Tyr Tyr Trp Val Pro Leu Ser Phe Phe Gly Leu
260 265 270
Met Leu Val Ile Val Thr Tyr Leu Gln His Val Asp Asp Val Ala Glu
275 280 285
Val Tyr Glu Ala Asp Glu Trp Ser Phe Val Arg Gly Gln Thr Gln Thr
290 295 300
Ile Asp Arg Tyr Tyr Gly Leu Gly Leu Asp Thr Thr Met His His Ile
305 310 315
Thr Asp Gly His Val Ala His His Phe Phe Asn Lys Ile Pro His Tyr
325 330 335
His Leu Ile Glu Ala Thr Glu Gly Val Lys Lys Val Leu Glu Pro Leu
340 345 350
Ser Asp Thr Gln Tyr Gly Tyr Lys Ser Gln Val Asn Tyr Asp Phe Phe
355 360 365
Ala Arg Phe Leu Trp Phe Asn Tyr Lys Leu Asp Tyr Leu Val His Lys
370 375 380
Thr Ala Gly Ile Met Gln Phe Arg Thr Thr Leu Glu Glu Lys Ala Lys
385 390 395 400
Ala Lys

ES 2 698 473 T3

<210> 31
 <211> 570
 <212> ADN
 <213> *Caenorhabditis elegans*
 5
 <400> 31
 agcagcagat tgcccgaggg cggcggaagg gacgaggccc aggcgggctcg tgaaagcgca 60
 tttccgaagg cgggctcggc gacgacgccg gcgcggcgac gacggccctg ccggaccggg 120
 cctgggggtgg acggcgaggg taactaggac ttgggggaagc cgagctgagc gacttgagcg 180
 ggttgagggg acgaactggt taggcgcggc cgagtcgtca gagccagcct gtggagaaag 240
 aggcgccgcc gagtgcgacg gggaacgctg cgcgcgacct gcattgcacc gcatcgcawt 300
 cgcaccgcaw tcgcaccgca ccgcatcgca ccgcatcgca tcgagaccgg acgcagcgag 360
 acccgacgct gggccttccc ggcgaaaaaa agtgatctgg cttacaaatc ccgagacgag 420
 acagacgtcg gcagcagaaa cgaatcagtc gagcagcagc tgcagcagca gcagcagcag 480
 cagcagccca tcgcgagcaa gggctcagcc agcagaacac caatcaggcc aagaatcgca 540
 cggaaagcaag ccttgacatc ctttgccaac 570

 <210> 32
 <211> 18
 <212> ADN
 <213> *Schizochytrium* sp.
 10

 <400> 32
 15 gaccgctcat ctatgctg 18

 <210> 33
 <211> 18
 <212> ADN
 <213> *Schizochytrium* sp.
 20

 <400> 33
 ctcaaagtga acgatgcc 18
 25
 <210> 34
 <211> 4244
 <212> ADN
 <213> *Schizochytrium* sp.

ES 2 698 473 T3

<400> 34

tittctctctc	tcgagctggt	gctgctgctg	ctgctgctgc	tgcttctctg	ctggttctca	60
cgctccgttcg	atcaagcgct	cgctcgctcg	accgatcggg	gcgtgcgtgc	gtgcgtgagt	120
ottgttgcca	ggcagccgca	ggctgctgtg	ctggttggtg	agttttacc	tcgggggtcg	180
gggtctgcct	gcctcccgc	cccgcgccgc	gcccgcgta	tcaccccgc	tcgcctcgc	240
ccatcggggc	tcgcctcctc	gcccgcacg	catcgccgc	atcgcatgca	tcgatgctgcc	300
acgcacgggg	ggacgcgcgc	cccgcgtccc	ccgcgcgcgc	cgctcgtcgc	tgccgatgcc	360
gtcgcgcgcc	tccttctctc	cctgcctcc	tccttctccc	gagccccct	gtcttctctc	420
gccccgcag	cgggcgcag	gaagcgagga	gagcggggag	gagagaagaa	aagaaaaaaa	480
aagaaaaaa	aataacagcg	cctctcgcg	cagacgcgcg	cgcccgctg	cgaggcggcg	540
tgatggggct	tcctcgtggc	cggctcgggc	ctggcccggc	ctcgccttg	aggtgcaggc	600
tttgggagag	aagagtggga	cgcggaagag	ataagatggt	gccatggcg	aggacggaga	660
ggttgctgaa	acttctctga	gcggcacagg	cgatggcgag	agaccgacag	ctgcccgggc	720
ggaggggatg	gatacctccc	gaggctggca	tgacgagct	ggccgcgcg	atctggctgg	780
ccgcgcggcg	gtgggtccgg	aggcgcgagg	ttggttttct	tcatacctga	taaccatacgg	840
tattcattct	tcctctccag	gaaggagca	agtacatag	agtatcacta	gcctaattgat	900
ggactctatg	ttttaggcca	cgtcggagca	gaaggcgcga	gcgattcgaa	tgccgagcgt	960
agatacagca	cagagaccct	gcccgcagc	cgatgcagg	cgagcagca	cgacccgcac	1020
gcacggcagc	ggtgacgcg	ctcctcggca	gatgcacgg	tcctgcgcgc	gcctttacat	1080
tttttgattt	taggtggtgt	gcctgcact	ttgaacatca	tcacaagtc	aacgcagcat	1140
caagaggcaa	gcaagtagat	acatccattc	gaattcaagt	tcagagacg	cagcaaacgc	1200
cgccgctccg	ctcaagctgc	agctagctgg	ctgacagggc	tcgctggctg	tagtggaaaa	1260
ttccattcac	ttttctgcat	cccgccagc	caggcccgt	cgcacgttct	ctcgtttggt	1320
tgcttcgttcg	tcgctgcgtg	cgtcgcctcc	agctgcctgt	ctaactctgc	gcgcatcca	1380
acgacccctc	gtcgtcgcgc	caagcgaaac	ccgacgcgc	cctggccaat	gcccgaagaa	1440
tgtaagcgc	gcagcaatgc	tgagagtaat	cttcagcca	ccaagtctt	atcgctgccc	1500
aagtctccat	cgacgcaca	ttcaggctt	ctctctctct	ccctccctct	ctttctgccc	1560
ggagagaagg	aaagaccgc	cgccgcgcc	tcctgcgcctg	tgacgggctg	tcggttgtaa	1620
gcccctctag	acagttccta	ggtgcggggc	gcccgcgcgc	ctcctgcgca	ggcacaagta	1680
ggcgccacg	ggttcccccc	gcaccttcca	caccttcttc	ccccgcagcc	ggaccgcg	1740
ccgtctgctt	acgcacttgc	cgccgcgcc	gcccgcgaac	ccgagcgcgt	gctgtggggc	1800
ccgtcttccg	gcccgcctcg	agtgcttccc	cgccgcgcgc	tactccgggt	cctgtgcgg	1860
acgtacttaa	tattaacagt	gggacctcgc	acaggacctg	acggcagcac	agacgtcgc	1920
gcctcgcac	gtcgggacg	caggcgaggc	atcccgcgc	ggccccgcac	cggggagget	1980
gcccggggggc	ctcttccggc	cgccggccgc	atcaggcgga	tgacgcaaga	gcccctcgcag	2040
tcgctcgcctc	gcccggagcgc	agccggggc	cagcgtggcc	aagctcccgc	cccttctggc	2100
tgctgcgcat	cctgcctgcc	tgctgcctg	cgctgcctgc	tgctgcctg	ccttctgctg	2160
tgctgcctct	cgctgcctg	tgctgcctg	cgccggaaaga	gggatcatgc	gaggtcaaat	2220
caccgccgcg	acctcgcact	ttgaagaagc	cgcatgcga	tgcatgcga	tgcatgcga	2280
cgcataccg	tgcatggcta	cgaagcgagt	ctggccggcc	gtcacaaca	gcacgttttc	2340
gagaaggagg	gctggcggag	cgctgcctgc	cgccgaccat	tgcaacgcgc	gcgtctcgtg	2400
gctggcgaag	gatctaacga	tgctgcctg	tgctgcctg	gatgcataag	ctgtgcctgat	2460
ccccggtcca	ttccaccacg	tcctgtgcctg	ccgctgacc	tgctctggc	tttcttcaa	2520
gttctcctcc	gcccggcctt	caggaccgag	acgagacctg	cagctgcagc	tagactcgcg	2580
ctcgcctcgc	gaggtatcgc	cgccgcgcgc	gcgggacggg	actcgcgagg	tcacaagccc	2640
gcccgcgcatc	gcgatggctg	tgctgcctg	ctcgtgcctg	gcagccgtac	gtcagcagc	2700
ccgcctccgt	attgtggatt	cgttagttgg	ttgttgggtg	atttggttgat	taattttttt	2760

gttcgtaggc	ttggttatag	ctaatagttt	agtttatact	ggtgctcttc	ggtgctgatt	2820
tagctcgact	tggttccaca	ccactgcccc	tctactgtga	atggatcaat	ggacgcacga	2880
cgggccgacg	aaagtgcgcg	agtgaggtaa	ctcaagcaac	ggcgtctctc	agaggggacg	2940
cacgcctccc	gtcgcagta	gtccagacag	gcagaaaaagc	gtcttaggga	ccacgcacgc	3000
acgcacgcac	gcacgcagc	ccgcacgcac	gctccctccc	tcgctgcct	atttttttag	3060
gcttctctcc	gcacgggccc	acctctcgt	ccctgcctc	gcccacaccg	gcccgcagcag	3120
cgatacctgc	cggtgcgcgc	tcgctcaagc	gctcagccgc	agctcagccc	agccgcagc	3180
tagggtttgt	tcgtctgtaa	ttggttgatt	tgatttgatt	tgatttgatc	cgatccgatc	3240
cgatctgato	tgatttgott	tgctttgctt	tgtctccctc	ccgcgcggga	ccaagcgtcc	3300
gtctgcgcgc	cgagcttccc	cttcttctcc	cagccctccc	tcctgcctcc	cctctcgcgc	3360
aagcagcag	cttcgcgcgc	gcactccggtc	ggtcggctcg	tcgatcgacc	cgccctgccc	3420
tgctgctgtg	gcccggcttt	tcctccatcgc	cgactcttcc	ttctccatac	gtcctaactac	3480
gtacatacat	actgcccggc	tcctccctct	ccagcgcggc	gacggcggca	ggctgcagc	3540
tcctgcgcgc	cgccggcgc	gcccgcgcgc	ccgcgcgcgc	cgccgcgcgc	ggccctcgtc	3600
gcccgcgcgc	ctccgctccg	ctccgagccc	gcgaggggc	cgccgcgcgc	cgatggatgg	3660
atggatggat	ggatggatgg	atggatggg	ttgatcgatg	gcccgcgcgc	ggcggagatg	3720
agcagggacg	agcgcgcgag	cgccgcagcc	ggattcgcag	ggccctcgtc	gcctcgcgc	3780
cgctgcgcgc	cccgccttgc	gagcctgcgc	cgagcgagcg	cgagcgagcg	agcgggctt	3840
tccttctctc	gcccgcgcgc	tgccctcgtg	tgtcttctgc	ttgcttagcg	ggccgcgcgc	3900
tggaagatgg	ctcattcaat	cgacccattc	acgcacgcac	tcggcgcgc	agagaaggcc	3960
gaggaggagc	agcaagcaaa	ccaaaagctc	tcgctcgcgc	ggtctcgggc	tcgagcggtc	4020
tcggagagag	agcttctcgc	gcacccagc	cagcagcagc	agcagcagca	gcgctctca	4080
gcacgagcac	gagcagcagc	acgagcagca	gcattcagc	aagaggacag	acacggttg	4140
cagcgcctag	ctcgcctcgc	acagaaagag	gcccgttggg	cgtaaaaaaa	aaggagcagc	4200
caagccgcca	gcccgcgcgc	tagctagcca	gctcgcctgc	caaa		4244

ES 2 698 473 T3

<210> 35
 <211> 4244
 <212> ADN
 <213> Schizochytrium sp.

5

<400> 35
 tttctctctc tcgagctggt gctgctgctg ctgctgctgc tgcttccttg ctggttctca 60
 cgtccggttc atcaagcgct cgctcgctcg accgatcggg gcgtgcgtgc gtgcgtgagt 120
 cttgttgcca ggcagccgca ggctgtctgt ctgtttgtgt agttttacc tccggggttcg 180
 gggctctgct gcctcccgtc cccgcccgcg gccgcccgtg tccaccccgc tcgcctccgc 240
 ccatcgggcc tcgcctcctc gcgcgcgacg catcgccgcg atcgcatgca tcatgctgcc 300
 acgcaocggg ggacgcgcgc cccgcgtccc ccgcgcgcgc cgtcgtcgtc tgggatgcc 360
 gtcgcgcgcc tccttccttc cctcgcctcc tcttcctccc gagccccctt gtcttccttc 420
 gccccgcag cggcgcgcag gaagcgagga gagcggggag gagagaagaa aagaaaagaa 480
 aagaaaagaa aataacagcg ccgtctcgcg cagacgcgcg cggccgcgtg cgaggcggcg 540
 tgatggggct tctcgtggcg cggctgcggc ctggcccggc ctgcctttg aggtgcaggc 600
 tttgggagag aagagtggga cgcgggagaa ataagatggt gccatggcgc aggacggaga 660
 ggttgctgaa acttcttcga gcggcacagg cgatggcgag agaccgacag ctgccggcgc 720
 ggaggggatg gatacctccc gaggctggca tggacgagct ggccgcgcgc atctggctgg 780
 ccgcgcggcg gtgggtcccg aggcgcgagg ttggttttct tcatacctga taccatacgg 840
 tattcattct tcctctccag gaaggaagca agtcacatag agtatcacta gcctaataatg 900
 ggactctatg ttttagggca cgtcggagca gaaggcgcga gcgattcgaa tgcgagcgat 960
 agatacagca cagagacctt gccggcgacg cggatgcagg cgagcacgca cgcacccgac 1020
 gcaacggcagc ggtgcacgcg ctctcggca gatgcacggg tctgcgccgc gcctttacat 1080
 tttttgatth taggtggtgt gcctgccact ttgaacatca tccacaagt c aacgcagcat 1140
 caagaggcaa gcaagtacat acatccattc gaattcaagt tcaagagacg cagcaacagc 1200
 ccgcgctccg ctcaagctgc agctagctgg ctgacagggc tcgctggctg tagtggaaaa 1260
 ttccattcac ttttctgcat ccgcggccag caggcccgtg cgcacgttct ctcgtttgtt 1320
 tgttcgcttc tgcgtgcgtg cgtgcgtccc agctgcctgt ctaatctgcc gcgcatcca 1380
 acgacctcg gtogtcgccc caagcgaaac ccgacgccga cctggccaat gccgcaagaa 1440
 tgctaagcgc gcagcaatgc tgagagtaat ctfcagcca ccaagtcatt atcgtgcc 1500
 aagtctccat cgcagccaca ttcaggcttt ctctctctct cctccctct ctttctgcc 1560
 ggagagaagg aaagacccgc cgcgcgcgcc tctgcgcctg tgacgggctg tccgttgtaa 1620

ES 2 698 473 T3

gccctcttag	acagttccta	ggtgccgggc	gcccgcgcgc	ctccgtcgca	ggcacacgta	1680
ggcggccaog	ggttcccccc	gcaccttcca	caccttcttc	ccccgcagcc	ggaccgcgcg	1740
ccgtctgctt	acgcacttcg	cgcggccgcc	gcccgcgaac	ccgagcgcgt	gctgtgggcg	1800
ccgtcttccg	gocgcgtcgg	aggtcgtccc	cgcgcgcgcg	tactccgggt	cctgtgcggt	1860
acgtacttaa	tattaacagt	gggacctcgc	acaggacctg	acggcagcac	agacgtcgcc	1920
gcctcgcate	gctggggacg	caggcgaggc	atcccggcgc	ggccccgcac	cggggaggct	1980
gogggggcgc	ctcttccggc	cggggcgcgc	atcaggcgga	tgacgcaaga	gccctcgcag	2040
tcgctcgctc	gcgggagcgc	agcgcggcgc	cagcgtggcc	aagctcccgc	cccttctggc	2100
tggtcgcag	ccctgcctgc	tgctgcctg	cgtgcgtgog	tgctgcgtg	ccttctgctg	2160
tcctcgcctt	cgtgcgtgcy	tcgtgagtg	cggcggaaga	gggatcatgc	gaggatcaat	2220
caccocgcgc	acctcgactt	ttgaagaagc	cgcgatgcga	tgcgatgcga	tcgatgcga	2280
cgcgataccg	tgcgaggcta	cgaagcgagt	ctggccggcc	gtcatacaac	gcacgttttc	2340
gagaaggagg	gctggcggag	gcctgcatgc	cgcgcacccat	tgcaacgcgc	gcgtctcgtg	2400
gctggogaag	gtgcctggag	gatctaacga	tcgctgctat	gatgctatag	ctgtgctgat	2460
ccccggtcca	ttocaccaog	tctgtgcctg	ccgcctgacc	tgcgcttggc	tttccctcaa	2520
gttctcctcc	gcccggcctt	caggaccgag	acgagacctg	cagctgcagc	tagactcgcg	2580
ctcgcctcgc	gaggattcgc	cggccgcgcg	gcccgaacggg	actcgcgagg	tcacacggcc	2640
gcccggcagc	gcgatggctg	tgctgacgta	ctcgtgcgtg	gcagccgtac	gtcagcgacg	2700
ccgcctccgt	attgtggatt	cgttagttgg	ttgttggttg	atltgttgat	taattttttt	2760
gttcgtaggc	ttggttatag	ctaatagttt	agttataact	ggtgctcttc	ggtgctgatt	2820
tagctcgact	tggttccaca	ccaatgcccc	tctactgtga	atggatcaat	ggacgcacga	2880
cgggcccagc	aaagtgcgog	agtgaggtaa	cctaagcaac	ggcggctctc	agaggggacg	2940
cacgcctccc	gtcgcagtca	gtccagacag	gcagaaaaagc	gtcttaggga	ccacgcacgc	3000
acgcacgcac	gcacgcacgc	ccgcacgcac	gtccctccc	tcgcgtgcct	atlttttttag	3060
gcttccctcc	gcacgggcct	acctctcgtc	ccctcgcctc	gcccaccag	gcccgcagcag	3120
cgatacctgc	cgggtcccgc	tcctgcacgc	gtccagccgc	agctcagccc	agcccgcgagc	3180
tagggtttgt	tcgtcctgaa	ttgtttgatt	tgatttgatt	tgatttgatc	cgatccgatc	3240
cgatctgac	tgatttgctt	tgctttgctt	tgtctccctc	cggcgcgga	ccaagcgtcc	3300
gtctgcgcgc	cgcagcttcc	cttcttctcc	cagccctcct	tctgctcccg	cctctcgcgc	3360
aagcacgcag	cttcgcgcgc	gcacccggtc	ggtcggctcg	tcgatcgacc	cgcctgccgc	3420
tgctgctgtg	gcccggcttt	tcctccatog	cgactcttcc	ttctccatac	gtcctactac	3480
gtacatacat	actgccggct	tctcctctct	ccagcgcggc	gacgggggca	ggctgcgacg	3540
tcgtcgcgcg	cgcgggcgcg	gcgcgcgcgc	ccgcgcgcgc	ccgcgtcgca	ggcctcgtc	3600
gcccgcgcgc	ctccgctccg	ctccgagggc	gcgagagggc	cgcggcggcg	cgatggatgg	3660
atggatggat	ggatggatgg	atggattttg	ttgatcgatg	gcccgcgatg	ggcggagatg	3720
agcagggagc	agcgcgcgag	cgcggcagcc	ggattcgcag	ggcctcgcctc	gcctcgcgcc	3780
cgctgcgcgc	cccgccttgc	gagcctgcgc	cgcgagcgcg	cgagcgcgag	agcggggctt	3840
tctttgtctc	gcgcgcgcgc	tgccctcgtg	tgtcttctgc	ttgcgtagcg	ggcgcgcgcg	3900
tggaagatgg	ctcattcaat	cgaccattc	acgcacgcac	tcgggcgcgc	agagaaggcc	3960
gaggaggagc	agcaagcaaa	ccaaaagctc	tcgcgctcgc	ggtctcgggc	tcgagcggctc	4020
tcggagagag	agtcttgcgg	cgaccaccgg	cagcagcagc	agcagcagca	gcgctgtcga	4080
gcacgagcac	gagcacgagc	acgagcaaga	gcattcgcgc	aagaggacag	acacggttgt	4140
cagcgcctag	ctcgcctcgt	acagaaagag	gcgggttggg	cgtaaaaaaa	aaggagcacg	4200
caagccgccca	gccagccagc	tagctagcca	gcctgcctgc	caaa		4244

REIVINDICACIONES

- 5
1. Molécula aislada de ácidos nucleicos que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en:
- 10
- a. SEC ID nº 4;
 - b. nucleótidos 441 a 894 de SEC ID nº 9;
 - c. una secuencia de ácidos nucleicos que es por lo menos aproximadamente 95% idéntica a los nucleótidos 441 a 894 de la SEC ID nº 9 a lo largo de la longitud completa de los nucleótidos 441 a 894 de la SEC ID nº 9, en la que dicha secuencia de ácidos nucleicos presenta una actividad transcripcional por lo menos basal del promotor α -tubulina, y
 - d. un polinucleótido aislado que comprende una secuencia de ácidos nucleicos que es totalmente complementaria a dicho polinucleótido de (a), (b) o (c).
- 15
2. Vector recombinante para la transformación de microorganismos del Orden Traustozoa, que comprende una secuencia de ácidos nucleicos codificante de una proteína de unión a bleomicina operablemente ligada a un promotor α -tubulina de Traustozoa que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en los nucleótidos 441 a 894 de la SEC ID nº 9, y una secuencia de ácidos nucleicos que es por lo menos aproximadamente 95% idéntica a los nucleótidos 441 a 894 de la SEC ID nº 9 a lo largo de la longitud completa de los nucleótidos 441 a 894 de la SEC ID nº 9, en la que dicha secuencia de ácidos nucleicos presenta una actividad transcripcional por lo menos basal del promotor α -tubulina.
- 20
- 25
- 30

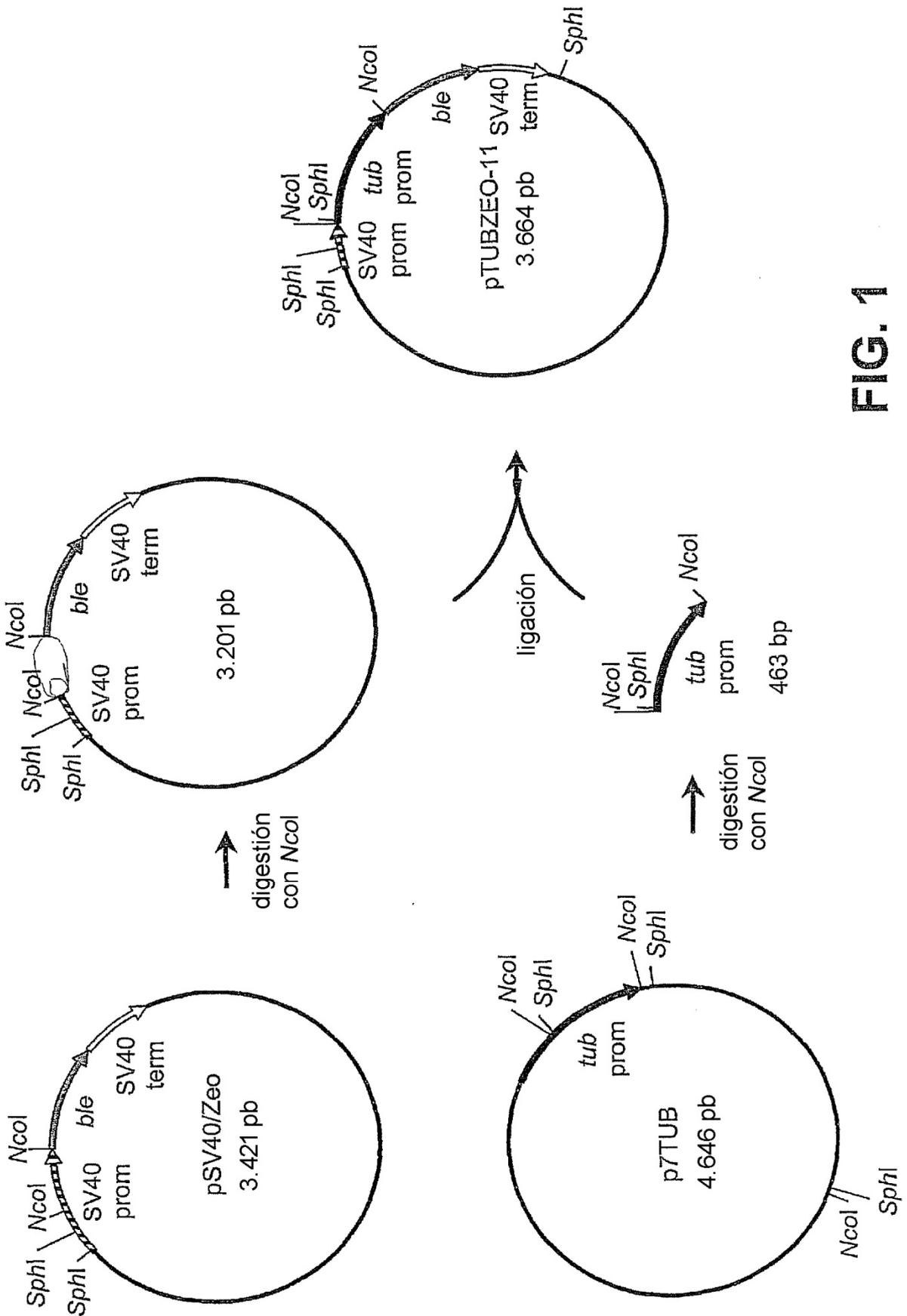


FIG. 1

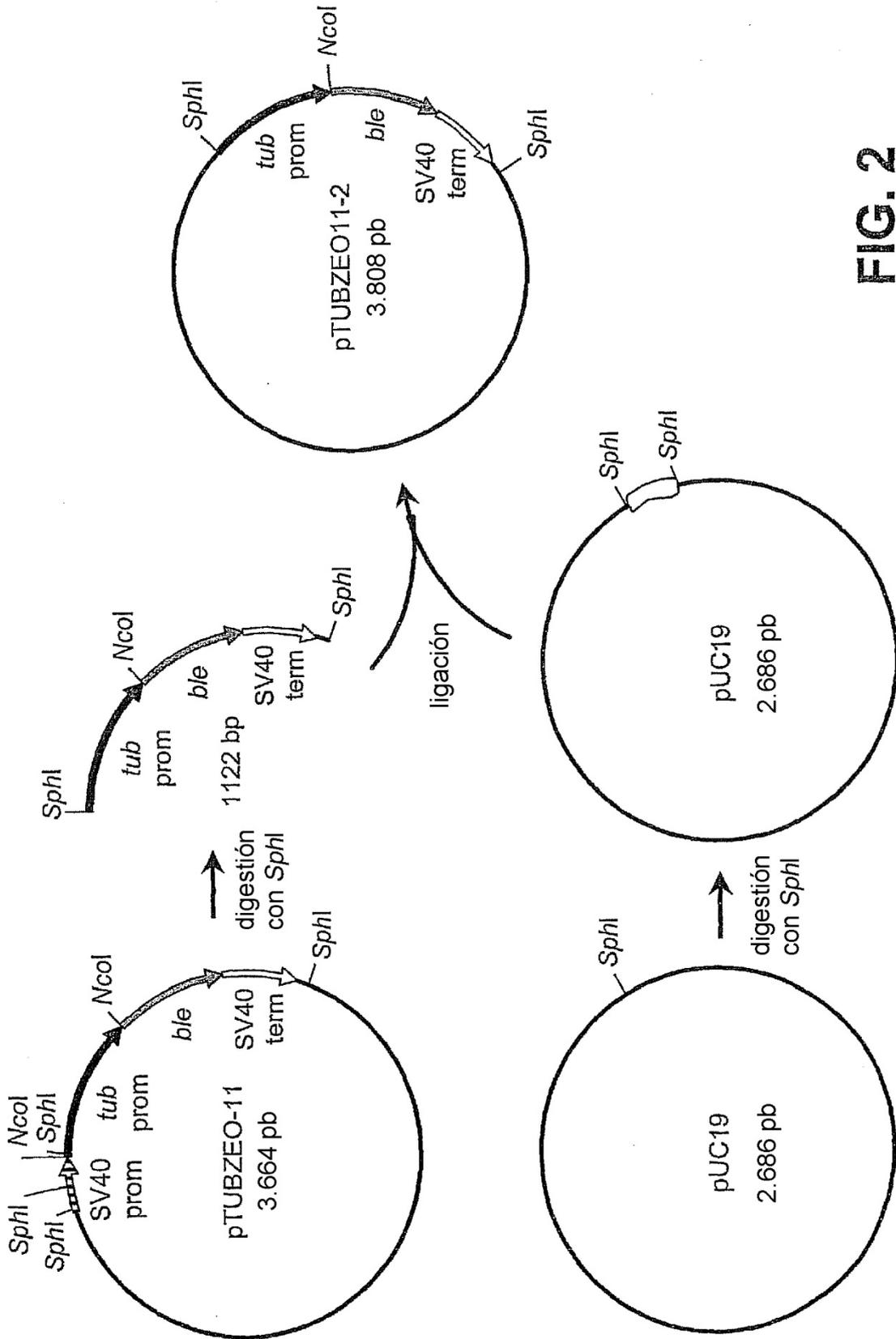


FIG. 2

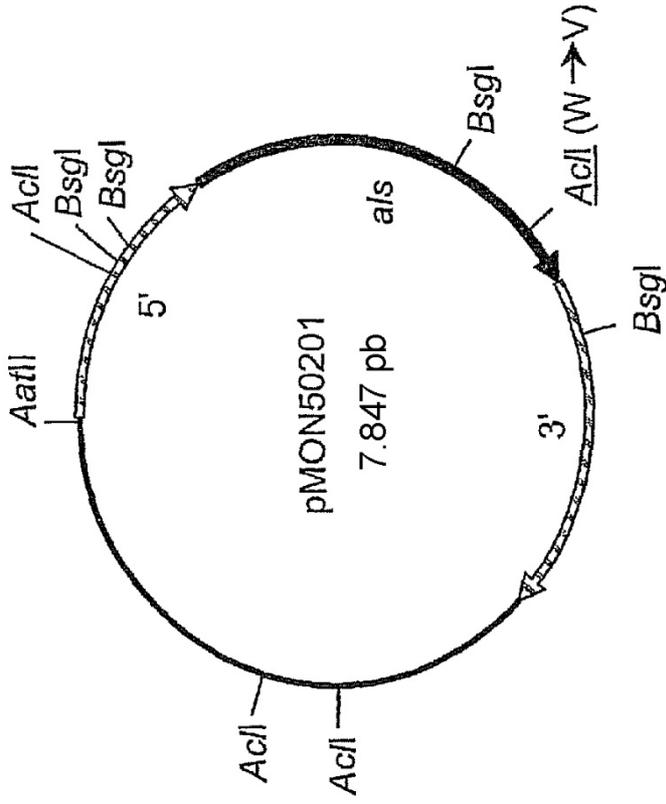


FIG. 3B

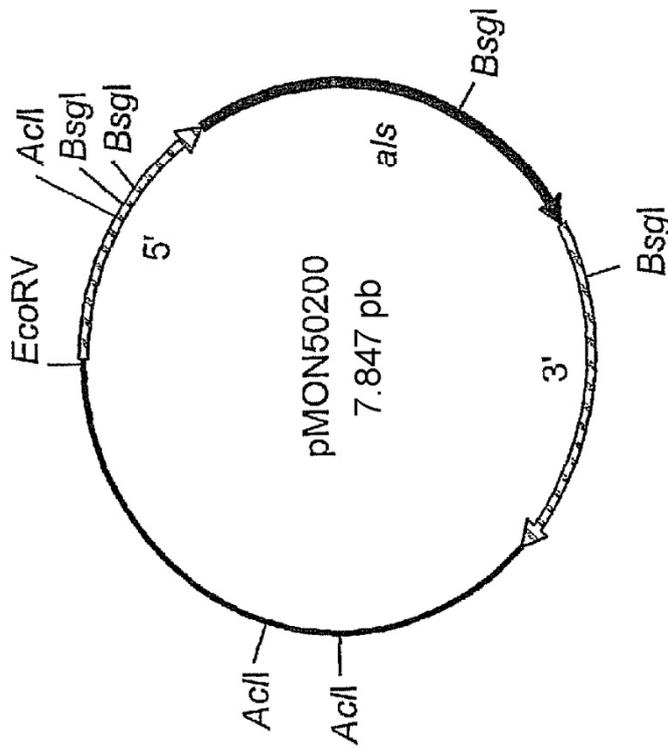


FIG. 3A

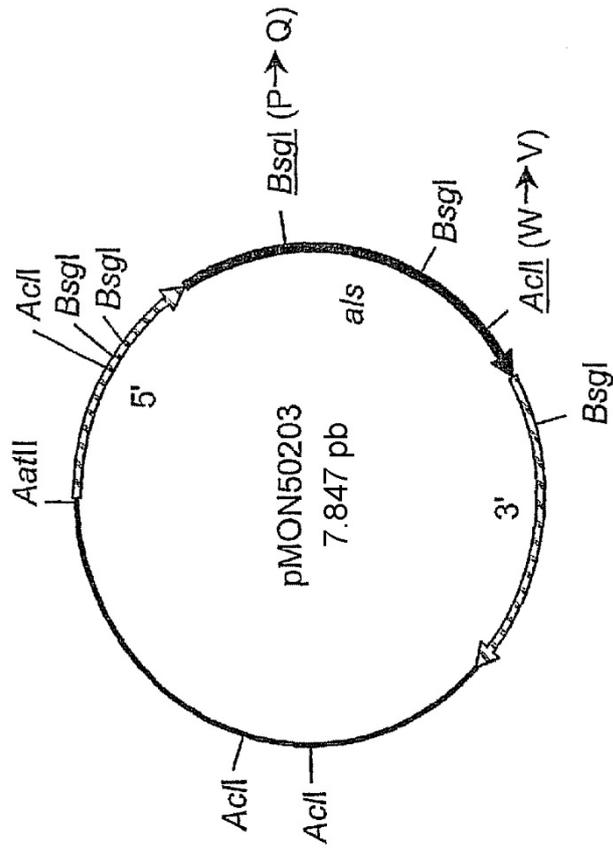


FIG. 3D

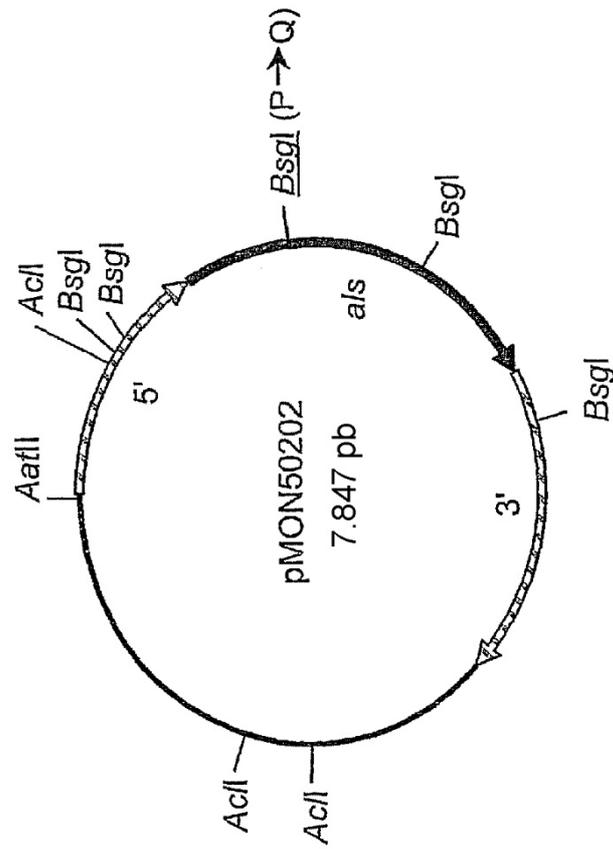


FIG. 3C