

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 682 279**

51 Int. Cl.:

C12N 1/12 (2006.01)
C12N 1/36 (2006.01)
C12N 15/01 (2006.01)
C12N 15/11 (2006.01)
C12N 15/82 (2006.01)
C12N 15/87 (2006.01)
C07K 14/415 (2006.01)
C07K 14/405 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **06.12.2013 PCT/US2013/073741**
 87 Fecha y número de publicación internacional: **12.06.2014 WO14089533**
 96 Fecha de presentación y número de la solicitud europea: **06.12.2013 E 13860017 (6)**
 97 Fecha y número de publicación de la concesión europea: **25.07.2018 EP 2929029**

54 Título: **Mutantes de algas que tienen un fenotipo aclimatado a la luz intensa en compartimentos**

30 Prioridad:

06.12.2012 US 201261733956 P
09.04.2013 US 201361810216 P
23.08.2013 US 201361869590 P
23.09.2013 US 201361881342 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
19.09.2018

73 Titular/es:

SYNTHETIC GENOMICS, INC. (100.0%)
11149 N. Torrey Pines Rd, Suite 100
La Jolla, CA 92037, US

72 Inventor/es:

BAILEY, SHAUN;
MCCARREN, JAY;
LIEBERMAN, SOYAN, LEUNG;
MEUSER, JONATHAN, E.;
ROMANO, ANNA, E.;
YEE, DANIEL;
SORIAGA, LEAH;
BROWN, ROBERT, C.;
WEISSMAN, JOSEPH, C.;
PRINCE, ROGER, C.;
NIELSEN, ROBERT, D. y
SCHWARTZ, ARIEL, S.

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 682 279 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

Mutantes de algas que tienen un fenotipo aclimatado a la luz intensa en compartimentos

Referencia cruzada con solicitudes relacionadas

5 Esta solicitud reivindica prioridad bajo U.S.C. § 119 (e) de la solicitud de patente provisional de EE.UU. No. 61/733.956, presentada el 6 de diciembre de 2012; de la solicitud de patente provisional de EE.UU. No. 61/810.216, presentada el 9 de abril de 2013; de la Solicitud de Patente Provisional de EE.UU. Nº 61/869.590, presentada el 23 de agosto de 2013; y del documento de patente 61/881.342 presentado el 23 de septiembre de 2013.

Referencia a un listado de secuencias

10 Esta solicitud contiene referencias a secuencias de ácido nucleico que se han presentado al mismo tiempo con este documento en el archivo de texto de listado de secuencia "SGI1700-4WO Sequence Listing_ST25.txt", tamaño de archivo 331 kilobytes (kb), creado el 6 de diciembre de 2013.

Campo de la invención

15 La presente invención se refiere a mutantes de algas que tienen un contenido reducido de clorofila y una mayor eficacia fotosintética. La presente invención también se refiere, en algunas realizaciones, a genes que codifican reguladores de aclimatación a la luz, a constructos que incluyen al menos una porción de los genes reguladores, y a métodos de ingeniería de microorganismos fotosintéticos que usan dichos constructos.

Antecedentes

La presente invención se refiere a cepas de algas mutantes que tienen nuevos rasgos fotosintéticos, a métodos para generar, identificar y/o aislar tales mutantes y a genes que codifican proteínas que regulan la fotosíntesis.

20 En sistemas de crecimiento de algas abiertos a gran escala, los cultivos deben crecer a densidades y profundidades de cultivo razonablemente altas en la región de 20-30 cm. Este entorno de cultivo proporciona un auto-oscurecimiento significativo, lo que garantiza que cada célula experimente solo un bajo nivel de irradiación promedio. El estado foto-fisiológico predominante de una célula de alga en estas condiciones es un estado aclimatado a poca luz, que se caracteriza por un sistema de antena de recolección de luz auxiliar relativamente grande asociado con los centros de reacción fotosintética. Sin embargo, una antena de recolección de luz más grande en cada célula individual agrava el auto-oscurecimiento del cultivo, lo que lleva a un nivel de irradiación promedio incluso más bajo, lo que provoca como respuesta un aumento adicional en el tamaño de la antena de las células de algas. El resultado general es un cultivo con muy poca penetración de la luz, lo que garantiza que la mayoría del sistema de crecimiento abierto esté a oscuras. Además, la gran y eficiente antena de recolección de luz impulsa la saturación de la fotosíntesis a intensidades de luz relativamente bajas. Por lo tanto, en la capa superficial de los estanques, donde hay luz disponible, una parte importante de la luz incidente supera la luz requerida para impulsar las tasas fotosintéticas máximas. Este exceso de irradiancia se disipa a través de canales térmicos y se pierde en forma de calor. La eficiencia del uso de la luz en los sistemas de crecimiento abierto es muy baja y se ha sugerido que hasta el 80% de la irradiancia activa fotosintética, incidente sobre la superficie del estanque, se pierde en forma de calor.

35 Por lo tanto, la respuesta de aclimatación con poca luz disminuye la eficiencia general del uso de la luz de un cultivo en estanque al aumentar el auto-oscurecimiento y disminuir el nivel de irradiancia de saturación para la fotosíntesis. Los métodos anteriores para disminuir el tamaño de antena de recolección de luz en algas se han enfocado únicamente en la antena, dirigiendo directamente la biosíntesis de polipéptidos de recolección de luz, o reduciendo su ensamblaje o función indirectamente al interrumpir la biosíntesis de la clorofila, el control de traducción de proteínas o los mecanismos de localización de proteínas. Como resultado, las cepas obtenidas de pigmento reducido a menudo se desequilibran en la recolección de luz, el transporte de electrones y la fijación de carbono, lo que puede afectar adversamente a la productividad del cultivo.

45 La mayoría de los fotoautótrofos se aclimatan a diferentes niveles de irradiancia para maximizar la captura de la luz en condiciones de poca luz o para evitar los efectos potencialmente nocivos de la recolección de energía de excitación en exceso bajo alta irradiación. La característica más obvia de la respuesta de aclimatación a la irradiancia es un cambio en el nivel de pigmentación, típicamente asociado con cambios en la abundancia de la antena de recolección de luz auxiliar. Navakoudis et al. (BIOCHIMICA ET BIOPHYSICA ACTA. BIOENERGETICS, 2007, Vol. 1767(7): 948-955) describe un mutante de alga que tiene un menor contenido de clorofila y que tiene minimizada la inactivación no fotoquímica.

50 La aclimatación a la irradiancia es, sin embargo, una respuesta, en gran parte, pleiotrópica, que implica cambios en la composición y la función en múltiples niveles dentro de la maquinaria fotosintética y en todo el organismo. La regulación de la aclimatación a la irradiancia en fotoautótrofos oxigenados está mal definida y una mayor comprensión de la red reguladora subyacente puede permitir la manipulación beneficiosa de la composición y la función de la maquinaria fotosintética.

Sumario

La protección buscada para esta invención es como se define en las reivindicaciones.

La presente invención describe los resultados de un procedimiento de selección inclinado hacia el aislamiento de mutantes que retienen tantas características como sea posible del estado foto-fisiológico aclimatado de luz intensa natural, que incluye una menor antena de captación de luz, en equilibrio con todos los demás aspectos del proceso fotosintético. Esta pantalla genética avanzada (denominada pantalla "aclimatada a la luz intensa en compartimentos" (LIHLA, por sus siglas en inglés, "Locked In High Light Acclimated")), se diseñó específicamente para aislar los componentes reguladores globales asociados con la aclimatación fotosintética a la irradiancia. Con base en la implementación de la pantalla LIHLA, se proporcionan genes que codifican nuevos reguladores de la aclimatación de luz (por sus siglas en inglés, "Light Acclimation Regulators", LAR1, LAR2 y LAR3). Estos genes han demostrado ser, mediante análisis transcriptómicos y fotofisiológicos, reguladores globales de la aclimatación fotosintética a la irradiancia.

En un aspecto, se proporciona en este documento un mutante de algas desregulado en la aclimatación a poca luz, en donde el mutante de algas exhibe al menos una reducción del 20% en clorofila en condiciones de poca luz con respecto a un alga de tipo silvestre de la misma cepa y en donde, además, el mutante de alga exhibe una mayor inactivación fotoquímica (qP) en todas las irradiancias fisiológicamente relevantes por encima de $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ con respecto a un alga de tipo silvestre de la misma cepa; exhibe inicio de inactivación no fotoquímica (NPQ) a una intensidad de luz más alta que la intensidad de luz a la que aparece NPQ en un alga de tipo silvestre de la misma cepa y exhibe una NPQ más baja en todas las irradiancias fisiológicas por encima de $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ que es exhibida por un alga de tipo silvestre de la misma cepa, en donde, además, el mutante está mutado en un gen que codifica un polipéptido que comprende un dominio de dedo de cinc PF02135 TAZ, en donde el polipéptido comprende la proteína LAR1 (SEQ ID NO: 4), la proteína No-LAR1 (SEQ ID NO: 8), o un homólogo de la proteína LAR1 (SEQ ID NO: 4) o la proteína No-LAR1 (SEQ ID NO: 8) que tiene al menos un 80% de identidad con SEQ ID NO: 4 o SEQ ID NO: 8.

La descripción describe mutantes de algas que tienen un fenotipo "aclimatado a la luz intensa en compartimentos" o "LIHLA", en el que los mutantes exhiben propiedades fotosintéticas de células de algas con aclimatación a la luz intensa en ambas altas y bajas intensidades de luz. Los mutantes se caracterizan por una cantidad reducida de clorofila por célula y pueden exhibir, por ejemplo, una reducción del 20% o mayor en clorofila por célula, y preferiblemente un 25% o más, un 30% o más, 40% o más, o una reducción del 50% o superior de la clorofila por célula en condiciones de poca luz, en comparación con las células de tipo silvestre. Además, los mutantes LIHLA tienen al menos una de las siguientes propiedades fotosintéticas con respecto a las células de tipo silvestre, cuando tanto los mutantes LIHLA como las células de tipo silvestre se aclimatan a la poca luz: un aumento de la inactivación fotoquímica (qP) respecto a todas las intensidades de luz fisiológicamente relevantes mayores de $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ (por ejemplo, sobre todas las intensidades de luz fisiológicamente relevantes mayores que $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayores que $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayores que $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o mayores que $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$); un aumento de la tasa fotosintética máxima (P_{max}) sobre una base de clorofila, teniendo los mutantes al menos un 70% de P_{max} de las células de tipo silvestre por célula, y preferiblemente un 75% o más, 80% o más, 85% o más, 90% o más, o sustancialmente la misma tasa fotosintética máxima (P_{max}) que las células de tipo silvestre o una P_{max} más alta en comparación con las células de tipo silvestre por célula; saturación de la fotosíntesis a niveles de irradiancia más altos (mayor E_k); retraso en el inicio de la inactivación no fotoquímica (NPQ) en respuesta al aumento de la intensidad de la luz; niveles reducidos de NPQ en todas las irradiancias fisiológicamente relevantes mayores de $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por ejemplo, mayor que $450 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o mayor que $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. Además, un mutante LIHLA puede tener una velocidad de transporte de electrones máxima para el fotosistema II (ETR_{PSII}) que es al menos tan alta como la ETR_{PSII} de una célula de tipo silvestre y, preferiblemente, al menos aproximadamente 1,5 veces la velocidad de la ETR_{PSII} de tipo silvestre, o al menos aproximadamente 2 veces la tasa de la ETR_{PSII} de tipo silvestre.

En algunos ejemplos, un mutante LIHLA tiene al menos un 20%, 25%, 30%, 35%, 40%, 45% o 50% de reducción en clorofila con respecto a una célula de tipo silvestre, exhibe una mayor inactivación fotoquímica (qP) respecto a todas las intensidades de luz fisiológicamente relevantes mayores que $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ con respecto a una célula de tipo silvestre; tiene una mayor tasa fotosintética máxima (P_{max}) por cada base de clorofila (por ejemplo, al menos 1,5 veces la P_{max} de células de tipo silvestre, por ejemplo, al menos 2 veces la P_{max} de células de tipo silvestre), con al menos un 75% o al menos 80% de la P_{max} de células de tipo silvestre por célula; experimenta la saturación de la fotosíntesis a niveles de irradiancia más altos (mayor E_k) que el tipo silvestre; exhibe una aparición retrasada de NPQ en respuesta a la intensidad de luz creciente en comparación con las células de tipo silvestre; y tiene niveles más bajos de NPQ respecto a todas las irradiancias mayores de aproximadamente $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ que las células de tipo silvestre, cuando tanto los mutantes LIHLA como las células de tipo silvestre se aclimatan a poca luz. Adicionalmente, un mutante LIHLA puede tener una velocidad máxima de transporte de electrones PSII (ETR_{PSII}) que sea al menos equivalente a la velocidad de transporte de electrones PSII de tipo silvestre y que puede ser, por ejemplo, aproximadamente 1,5 veces la tasa de transporte de electrones PSII de tipo silvestre o superior. Además, un mutante LIHLA puede tener una tasa máxima de transporte de electrones PSI (ETR_{PSI}) que es sustancialmente equivalente o mayor que la tasa de transporte de electrones PSI de tipo silvestre. Además, un cultivo que

comprende un mutante LIHLA, donde el cultivo está expuesto a la luz de una fuente de luz, puede tener una mayor cantidad de penetración de luz en el cultivo que un cultivo que comprenda un alga de tipo silvestre comparable (es decir, un alga silvestre de la cepa progenitora).

5 Adicionalmente a cualquiera de los rasgos anteriores, un mutante LIHLA puede tener cualquier combinación de los siguientes rasgos con respecto a células tipo silvestre o control: mayor Fv/Fm, mayor Φ PSII, y una α menor, o pendiente inicial de la curva P/I. Por ejemplo, un mutante LIHLA aclimatado a poca luz puede tener un Fv/Fm mayor, un Φ PSII aumentado, y, opcionalmente, una α menor con respecto a las células tipo silvestre o control aclimatadas a poca luz.

10 La descripción describe mutantes que tienen un fenotipo LIHLA (p. ej., con respecto a las células de tipo silvestre aclimatadas a poca luz, los mutantes aclimatados a poca luz tienen la clorofila reducida en al menos un 20%, un qP más alto en irradiaciones superiores a $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, al menos 1,5 veces o al menos 2 veces la P_{max} de las células de tipo silvestre en una base de clorofila y al menos 70% o al menos 80% de P_{max} de las células de tipo silvestre por célula, la aparición de NPQ a mayor irradiancia y NPQ inferior a todas las irradiaciones superiores a $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, E_k superior, y preferiblemente, una ETR_{PSII} superior, en la que los mutantes se desregulan en la expresión de al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, al menos sesenta, al menos setenta, al menos ochenta, al menos noventa, o al menos 100 genes que están regulados en respuesta a la intensidad de la luz en células de tipo silvestre. Por ejemplo, los mutantes proporcionados en este documento pueden desregularse en la expresión de genes que se expresan diferencialmente cuando se aclimatan células de tipo silvestre aclimatadas de luz intensa a poca intensidad de luz.

20 Además, en este documento se proporcionan mutantes que tienen al menos una mutación en un gen que codifica una proteína reguladora, en la que una proteína reguladora puede ser una proteína que afecta directa o indirectamente la transcripción de múltiples genes, por ejemplo, al menos diez, al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, al menos sesenta, al menos ochenta, o al menos 100 genes. En algunos ejemplos no limitantes, ese mutante está mutado en un gen que codifica un factor de transcripción o un activador transcripcional.

25 Un mutante, como se proporciona en este documento, puede ser un mutante que surge espontáneamente, derivado de una mutación clásica (por ejemplo, UV, irradiación gamma o mutagénesis química) o puede ser obtenido mediante ingeniería genética (por ejemplo, recombinación homóloga, atenuación génica por antisentido o RNAi, o modificación del genoma, por ejemplo, usando meganucleasas, nucleasas con dedos de cinc, talens o sistemas CRISPR/cas).

30 También se proporciona en este documento una cepa de alga mutante o recombinante que exhibe propiedades fotosintéticas alteradas con respecto a una cepa silvestre o de control, por ejemplo, la cepa progenitora de la que se deriva el mutante, en la que la cepa incluye un gen mutado o atenuado que en la cepa de tipo silvestre codifica un polipéptido que incluye un dominio de dedo de cinc TAZ. El polipéptido puede incluir una secuencia de aminoácidos que puede tener al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, por ejemplo, al menos 80%, al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 9 o SEQ ID NO: 10. En diversos ejemplos, el polipéptido puede tener al menos 30%, al menos 35%, al menos 40%, al menos 45%, o al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia de aminoácidos con la secuencia de aminoácidos codificada por los genes LAR1 como se proporciona en este documento, por ejemplo, SEQ ID NO: 4 o SEQ ID NO: 8. En algunos ejemplos, una cepa de alga mutante puede tener un gen mutado o atenuado que en una cepa de tipo silvestre codifica un polipéptido que tiene al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia de aminoácidos con SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 o SEQ ID NO: 18. El polipéptido puede incluir un dominio de dedo de cinc TAZ. La cepa de alga mutante o recombinante que tiene un gen mutado o atenuado que codifica un polipéptido que tiene un dominio de dedo de cinc TAZ puede exhibir aclimatación alterada a poca luz con respecto a células de tipo silvestre de la misma cepa, y puede tener un fenotipo LIHLA como se describe en este documento.

50 También se proporciona en este documento una cepa de alga mutante o recombinante que exhibe propiedades fotosintéticas alteradas con respecto a células de tipo silvestre o células de control de la misma cepa de fondo, por ejemplo, la cepa de tipo silvestre que es un progenitor de la cepa mutante o recombinante, en donde la cepa mutante o recombinante incluye un gen mutado o atenuado que en la cepa tipo silvestre o control codifica un polipéptido que incluye un dominio de unión al ADN de tipo myb. El polipéptido puede incluir una secuencia de aminoácidos que puede tener al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, o al menos un 95% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 22 o SEQ ID NO: 23. En varios ejemplos, el polipéptido puede tener al menos 50%, al menos 55%, al menos 60%, al menos 65%, o al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con un polipéptido codificado por un gen LAR2 como se proporciona en este documento, por ejemplo, SEQ ID NO: 6 o SEQ ID NO: 21. En algunos ejemplos, una cepa de alga mutante puede tener un gen mutado o atenuado en un gen que codifica en una cepa de tipo silvestre un polipéptido que tiene al menos 65%, al

menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia de aminoácidos con SEQ ID NO: 6 o SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, o SEQ ID NO: 32. El polipéptido puede incluir un dominio de unión a ADN de tipo myb. La cepa de alga mutante o recombinante que tiene un gen mutado o atenuado que codifica un polipéptido que tiene un dominio de unión al ADN de tipo myb puede exhibir una aclimatación alterada a poca luz, y puede tener un fenotipo LIHLA como se describe en el presente documento.

Se proporciona adicionalmente en este documento una cepa de alga mutante o recombinante que exhibe propiedades fotosintéticas alteradas con respecto a células de tipo silvestre o células de control de la misma cepa de fondo, por ejemplo la cepa de tipo silvestre que es un progenitor de la cepa mutante o recombinante, en donde la cepa mutante o recombinante incluye un gen mutado o atenuado que en la cepa tipo silvestre o de control codifica un polipéptido que tiene un dominio que tiene al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 64. En varios ejemplos, la cepa mutante o recombinante incluye un gen mutado o atenuado que codifica un polipéptido que tiene al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia de aminoácidos con un polipéptido seleccionado del grupo que consiste en SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76 y SEQ ID NO: 78. En algunos ejemplos, una cepa de alga mutante puede tener un gen mutado o atenuado que en una cepa de tipo silvestre codifica un polipéptido que tiene al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90% o, al menos 95% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 63 o SEQ ID NO: 66. La cepa de alga mutante o recombinante que tiene un gen mutado o atenuado que codifica un polipéptido al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90% o al menos 95% de identidad de secuencia de aminoácidos con SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76 o SEQ ID NO: 78 puede exhibir una aclimatación alterada a poca luz, y puede tener un fenotipo LIHLA como se describe en el presente documento.

La descripción describe un mutante, como se proporciona en este documento, que tiene un gen mutado que codifica una proteína reguladora o una célula recombinante diseñada para tener una estructura alterada o expresión de un gen que codifica una proteína reguladora que puede mutarse en un gen que codifica un regulador global de la respuesta de aclimatación a la luz, en la que, por ejemplo, al menos diez, al menos quince, al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, o al menos 100 genes pueden desregularse en la cepa mutante en condiciones de poca luz (p. ej., menor o igual a aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, menor o igual a aproximadamente $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, menor o igual a aproximadamente $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o menor o igual a aproximadamente $50 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) con respecto a la cepa de tipo natural o control en condiciones de poca luz. Por ejemplo, el nivel de expresión de al menos diez, al menos quince, al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, o al menos 100 genes puede diferir en un \log_2 veces de 1 o mayor en la cepa mutante en condiciones de poca luz con respecto a la cepa tipo natural o control en condiciones de poca luz. La célula mutante o recombinante puede tener, por ejemplo, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, al menos diez, al menos once, al menos doce, al menos trece, al menos catorce, al menos quince, al menos dieciséis, al menos diecisiete, al menos dieciocho, al menos diecinueve, al menos veinte, al menos treinta, al menos cuarenta, o al menos cincuenta genes que están regulados negativamente en la cepa mutante con respecto a la cepa tipo silvestre en condiciones de poca luz. Por ejemplo, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, al menos diez, al menos once o al menos doce genes de la proteína de unión de clorofila de recolección de luz (LHC) pueden estar regulados negativamente en un mutante LIHLA con poca luz en comparación con una célula de tipo natural o de control con poca luz. Además, al menos dos, al menos tres, al menos cuatro, al menos cinco, o al menos seis genes de proteínas no LHC que codifican proteínas que funcionan en la fotosíntesis pueden estar reguladas negativamente en un mutante LIHLA. Además, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve o al menos diez genes que codifican proteínas que no funcionan en la fotosíntesis se pueden regular positivamente en el mutante con respecto a la cepa tipo silvestre o de control en condiciones de poca luz. Al menos cinco, al menos diez, al menos veinte o al menos treinta genes desregulados pueden tener un nivel de expresión (p. ej., un nivel de abundancia de transcripción) que difiere en un \log_2 veces de al menos 1, por ejemplo, puede estar presente en el mutante LIHLA a un nivel de dos veces o más del nivel de tipo silvestre, o 50% o menos del nivel tipo silvestre o de control, en las mismas condiciones.

Un mutante LIHLA que tiene un gen mutado que codifica una proteína reguladora puede ser además un mutante que permita una mayor penetración de la luz en el cultivo. Por ejemplo, un cultivo de un mutante LIHLA puede permitir la entrada de más luz al cultivo que es capaz de penetrar en un cultivo de una cepa de control o de tipo silvestre comparable. Por ejemplo, al menos un 50% más de luz, al menos un 60% más de luz, al menos un 70% más de luz, al menos un 80% más de luz, al menos un 90% más de luz o al menos un 100% más de luz, puede penetrar aproximadamente 2 cm por debajo de la superficie de un estanque de un cultivo mutante LIHLA que tiene una densidad de aproximadamente $4,5 \times 10^7$ células/ml en comparación con la cantidad de luz que puede penetrar aproximadamente 2 cm por debajo de la superficie de un estanque de un cultivo o células tipo silvestre o control que tengan aproximadamente la misma densidad de células.

Además, un mutante LIHLA que tiene un gen mutado que codifica una proteína reguladora en algunos ejemplos puede crecer hasta una densidad celular más alta que una cepa de control o de tipo silvestre. Por ejemplo, un

mutante LIHLA puede crecer hasta una densidad celular más alta que una cepa tipo silvestre o de control después de al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, al menos diez, al menos once, al menos doce, al menos trece, al menos catorce, o al menos quince días en el cultivo.

5 En varios ejemplos, un mutante LIHLA que se rompe en la expresión o función de un regulador global de la respuesta de aclimatación de luz tiene, al menos, una reducción del 25% en la clorofila con respecto a las células de tipo silvestre con poca luz (por ejemplo, menos de aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$), y está desregulada en la expresión de múltiples genes regulados por la luz, por ejemplo, exhibe desregulación de al menos diez, al menos quince, o al menos veinte genes que están regulados en respuesta a la intensidad de la luz en una célula de tipo silvestre, donde la diferencia en el nivel de expresión de los genes bajo condiciones de poca luz entre el mutante
10 LIHLA y una célula de tipo silvestre es al menos el doble. Además, en estos ejemplos, el mutante LIHLA exhibe un qP mayor que el tipo silvestre en todas las intensidades de luz mayores que $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, exhibe una aparición retrasada de NPQ con respecto a la intensidad de la luz en comparación con las células de tipo silvestre y NPQ inferior a todas las irradiancias superiores a aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, y exhibe un P_{max} mayor por clorofila y un P_{max} por célula que no es inferior al 70%, 75% u 80% del P_{max} natural. Además, el mutante LIHLA puede tener
15 una ETR_{PSII} máxima que sea mayor que la ETR_{PSII} máxima de las células de tipo silvestre cuando tanto el mutante LIHLA como las células de tipo silvestre se aclimatan a poca luz.

Un mutante LIHLA de algas puede ser, por ejemplo, una microalga tal como un género seleccionado del grupo que consiste en Achnanthes, Amphiprora, Amphora, Ankistrodesmus, Asteromonas, Boekelovia, Bolidomonas, Borodinella, Botrydium, Botryococcus, Bracteococcus, Chaetoceros, Carteria, Chlamydomonas, Chlorococcum,
20 Chlorogonium, Chlorella, Chroomonas, Chrysosphaera, Cricosphaera, Cryptocodinium, Cryptomonas, Cyanidioschyzon, Cyclotella, Cylindrotheca, Cymatopleura, Dixoniella, Dunaliella, Elipsoidon, Emiliana, Entomoneis, Eremosphaera, Ernodesmius, Euglena, Eustigmatos, Franceia, Fragilaria, Fragilariopsis, Gloeothamnion, Haematococcus, Halocafeteria, Hantzschia, Heterosigma, Hymenomonas, Isochrysis, Lepocinclis, Micractinium, Monodus, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Neochloris, Nephrochloris, Nephroselmis,
25 Nitzschia, Ochromonas, Oedogonium, Oocystis, Ostreococcus, Parachlorella, Parietochloris, Pascheria, Pavlova, Pelagomonas, Phæodactylum, Phagus, Picochlorum, Platymonas, Pleurochrysis, Pleurococcus, Prototheca, Pseudochlorella, Pseudoneochloris, Pseudostaurastrum, Pyramimonas, Pyrobotrys, Scenedesmus, Schizochlamydelta, Skeletonema, Spyrogyra, Stichococcus, Tetrachlorella, Tetraselmis, Thalassiosira, Tribonema, Vaucheria, Viridiella, Vischeria y Volvox. Por ejemplo, el mutante puede ser una diatomea (Bacillariófito) tal como, pero no limitada a, una especie de Achnanthes, Amphora, Chaetoceros, Cyclotella, Cylindrotheca, Cymatopleura, Entomoneis, Fragilaria, Fragilariopsis, Navicula, Nitzschia, Phæodactylum o Thalassiosira. Alternativamente, un mutante LIHLA puede ser un eustigmatofito, tal como, por ejemplo, una especie de Eustigmatos, Monodus, Nannochloropsis o Vischeria.

También se proporciona un método para aislar un mutante de algas desregulado en aclimatación de poca luz. El
35 método incluye: mutagenizar una población de algas; seleccionar la población mutagenizada de algas respecto a una baja fluorescencia de clorofila; seleccionar mutantes que retengan una baja fluorescencia de clorofila cuando se aclimatan a condiciones de poca luz; y seleccionar los mutantes seleccionados por fluorometría para identificar mutantes de algas de fluorescencia baja en clorofila y estables con poca luz que tienen coeficientes de extinción fotoquímicos (qP) al menos tan altos como las algas tipo silvestre. El método puede incluir identificar mutantes de
40 algas de fluorescencia baja en clorofila y estables con poca luz que tienen coeficientes de extinción fotoquímicos (qP) que son más altos que los de las algas de tipo silvestre a intensidades de luz mayores que aproximadamente $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayores que aproximadamente $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o mayores que aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. El cribado para baja fluorescencia de clorofila puede hacerse por clasificación celular activada por fluorescencia (FACS). El método incluye además seleccionar las algas con bajo contenido en clorofila para reducir la clorofila, por
45 ejemplo, una reducción del 20% o superior, una reducción del 30% o superior, una reducción del 40% o superior, o una reducción del 50% o superior, en clorofila por célula, en condiciones de poca luz. El método incluye además seleccionar mutantes de algas con baja clorofila para P_{max} aumentada por clorofila con respecto a células de tipo silvestre, y para P_{max} por célula a al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90% o al menos 95% de P_{max} por célula de las células de tipo silvestre. El método incluye además seleccionar mutantes de
50 algas con bajo contenido en clorofila para la aparición retardada de NPQ en respuesta al aumento de la intensidad de la luz en comparación con la aparición de NPQ en respuesta a la intensidad de la luz en células de tipo silvestre, y para niveles inferiores de NPQ con respecto a las células de tipo silvestre aclimatadas a poca luz a todas las irradiancias superiores a aproximadamente $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, superiores a aproximadamente $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, superiores a aproximadamente $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o superiores a aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

55 En varios ejemplos, el método puede incluir la selección de uno o más de Ek superior, ETR_{PSII} máximo superior, Fv/Fm más alto (eficacia fotosintética) y un mayor rendimiento cuántico fotosintético del fotosistema II, ΦPSII , con respecto a las células tipo silvestre. En algunos ejemplos, el método puede incluir la selección para una pendiente disminuida (alfa) de la curva de irradiancia fotosintética (P/I).

También se proporcionan métodos para producir productos de algas, que comprenden cultivar un mutante de algas
60 que se desregula en condiciones de poca luz como se proporciona en este documento y aislar al menos un producto del cultivo. El producto puede ser un lípido, un terpenoide, un policétido, una proteína, un péptido, uno o más aminoácidos, un carbohidrato, un alcohol, un ácido nucleico, uno o más nucleótidos, nucleósidos o

nucleobases, una vitamina, un cofactor, una hormona, un antioxidante o un colorante, o el producto puede ser biomasa de algas. El alga mutante se puede cultivar fototróficamente y se puede cultivar en un estanque o canal de conducción. También se proporciona un producto preparado por un mutante de algas como se describe en el presente documento. También se incluye en este documento una biomasa de algas que comprende un alga mutante desregulada en aclimatación a poca luz, tal como cualquiera de los mutantes de LIHLA proporcionados en este documento.

La descripción describe moléculas de ácido nucleico aisladas. Por ejemplo, una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido que incluye una secuencia de aminoácidos que tiene al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75% o al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 9 o SEQ ID NO: 10. El ácido nucleico puede ser un ADNc, por ejemplo. El polipéptido codificado por la secuencia de nucleótidos puede tener al menos 30%, al menos 35%, al menos 40%, al menos 45% o al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con un polipéptido codificado por un gen natural en el que la secuencia de nucleótidos es diferente de la secuencia de un gen natural. Adicionalmente, la secuencia de nucleótidos puede codificar un polipéptido que tiene una secuencia de aminoácidos alterada con respecto a la secuencia de aminoácidos codificada por el gen de origen natural con el que tiene similitud de secuencia (por ejemplo, SEQ ID NO: 4; SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO 12, SEQ ID NO 13, SEQ ID NO 14, SEQ ID NO 15, SEQ ID NO 16, SEQ ID NO 17 o SEQ ID NO 18). La molécula de ácido nucleico aislada puede incluir una mutación en la secuencia de nucleótidos que puede resultar, por ejemplo, en una sustitución, adición o delección de un aminoácido. En algunos ejemplos, la molécula de ácido nucleico codifica un polipéptido truncado (por ejemplo, truncado en el extremo N o truncado en el extremo C) o internamente. En algunos ejemplos, la molécula de ácido nucleico puede codificar un polipéptido que tiene al menos 30%, al menos 35%, al menos 40%, al menos 45%, o al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con SEQ ID NO: 4 o SEQ ID NO: 8. Adicionalmente, el polipéptido codificado puede incluir un dominio de dedo de cinc TAZ y, en algunos ejemplos, puede incluir un dominio de dedo de cinc TAZ mutado. Además, la molécula de ácido nucleico puede comprender una secuencia de nucleótidos que tiene al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, o SEQ ID NO: 40, o una parte de la misma, en donde la secuencia de nucleótidos difiere de la de un gen de origen natural.

Una molécula de ácido nucleico aislada, como se proporciona en este documento, puede comprender una secuencia de nucleótidos que codifica un polipéptido que tiene al menos 30%, al menos 35%, al menos 40%, al menos 45%, o al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con SEQ ID NO: 4; SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 o SEQ ID NO: 18 y/o una secuencia de nucleótidos que tiene al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94 %, o al menos 95% de identidad con SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, o SEQ ID NO: 40, o una porción de la misma o un complemento de al menos una porción de la misma, operativamente unida a una secuencia de expresión heteróloga. Alternativamente o además, una molécula de ácido nucleico aislada comprende un vector que incluye una secuencia de ácido nucleico que codifica un polipéptido que tiene al menos 30%, al menos 35%, al menos 40%, al menos 45% o al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con SEQ ID NO: 4; SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 o SEQ ID NO: 18 y/o una secuencia de nucleótidos que tiene al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, o SEQ ID NO: 40, o una porción de la misma o un complemento de al menos una porción de la misma.

La descripción describe moléculas de ácido nucleico aisladas que comprenden una secuencia de nucleótidos que codifica un polipéptido que incluye una secuencia de aminoácidos que tiene al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, o al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 22 o SEQ ID NO: 23. El ácido nucleico puede ser un ADNc, por ejemplo. Además, el polipéptido codificado por la secuencia de nucleótidos puede tener al menos un 30%, al menos un 35%, al menos un 40%, al menos un 45%, o al menos un 50%, al menos un 55%, al menos un 60%, al menos un 65%, al menos un 70%, al menos un 75%, al menos un 80% o al menos un 85%, al menos un 90%, o al menos un 95% de identidad de secuencia con un polipéptido codificado por un gen

natural en el que la secuencia de nucleótidos es diferente en secuencia de un gen natural. La secuencia de nucleótidos puede codificar un polipéptido que tiene una secuencia de aminoácidos alterada con respecto a la secuencia de aminoácidos codificada por el gen de origen natural con el que tiene similitud de secuencia (por ejemplo, SEQ ID NO: 6; SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, o SEQ ID NO: 32). La molécula de ácido nucleico aislada puede incluir una mutación en la secuencia de nucleótidos que puede resultar, por ejemplo, en una sustitución, adición o delección de un aminoácido. En algunos ejemplos, la molécula de ácido nucleico codifica un polipéptido truncado (por ejemplo, truncado en el extremo N o truncado en el extremo C) o internamente. En algunos ejemplos, la molécula de ácido nucleico puede codificar un polipéptido que tiene al menos 30%, al menos 35%, al menos 40%, al menos 45%, o al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75% o, al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con SEQ ID NO: 6 o SEQ ID NO: 21. Adicionalmente, el polipéptido codificado puede incluir un dominio de unión al ADN de tipo myb y, en algunos ejemplos, puede incluir un dominio de unión a ADN de tipo myb mutado. Además, la molécula de ácido nucleico puede comprender una secuencia de nucleótidos que tiene al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94% o, al menos, 95% de identidad con SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, o SEQ ID NO: 40, o una parte de la misma, en donde la secuencia de nucleótidos difiere de la de un gen de origen natural.

La descripción describe moléculas de ácido nucleico aisladas que comprenden una secuencia de nucleótidos que codifica un polipéptido que incluye una secuencia de aminoácidos que tiene al menos 70%, al menos 75% o al menos 80% o al menos 85%, al menos 90%, o al menos el 95% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 64. El ácido nucleico puede ser un ADNc, por ejemplo. Además, el polipéptido codificado por la secuencia de nucleótidos puede tener al menos un 65%, al menos un 70%, al menos un 75%, al menos un 80% o al menos un 85% o, al menos, un 90%, o, al menos, un 95% de secuencia identidad con un polipéptido codificado por un gen de origen natural en el que la secuencia de nucleótidos no es 100% idéntica con la secuencia de un gen natural. La secuencia de nucleótidos puede codificar un polipéptido que tiene una secuencia de aminoácidos alterada con respecto a la secuencia de aminoácidos codificada por el gen natural con el que tiene similitud de secuencia (por ejemplo, de SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78). La molécula de ácido nucleico aislada puede incluir una mutación en la secuencia de nucleótidos que puede resultar, por ejemplo, en una sustitución, adición o delección de un aminoácido. En algunos ejemplos, la molécula de ácido nucleico codifica un polipéptido truncado (por ejemplo, truncado en el extremo N o truncado en el extremo C) o internamente. En algunos ejemplos, la molécula de ácido nucleico puede codificar un polipéptido que tiene al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 90% o al menos 95% de identidad de secuencia con SEQ ID NO: 63 o SEQ ID NO: 66. Además, la molécula de ácido nucleico puede comprender una secuencia de nucleótidos que tiene al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87% , al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o, al menos, 95% de identidad con la SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78, o una parte de la misma, en donde la secuencia de nucleótidos difiere de la de un gen natural.

También se describen en este documento construcciones de ácidos nucleicos para recombinación homóloga (que incluye, pero no se limita a, construcciones de sustitución de genes y knock-out), modificación del genoma y atenuación del gen (por ejemplo, construcciones de ARNi, antisentido y ribozima), que incluyen al menos una parte de las secuencias de nucleótidos proporcionadas en este documento o sus complementos.

Una construcción de ácido nucleico para recombinación homóloga puede incluir una secuencia de ácido nucleico que incluye al menos una porción de un gen natural que codifica un polipéptido como se proporciona en este documento que regula directa o indirectamente la respuesta de aclimatación a la luz, por ejemplo, un gen LAR u homólogo del mismo como se proporciona en este documento. Alternativamente o además, una construcción para recombinación homóloga puede incluir una secuencia de ácido nucleico que incluye una secuencia que está posicionada en un genoma huésped adyacente a un gen natural que codifica un polipéptido como se proporciona en este documento que regula directa o indirectamente la respuesta de aclimatación a la luz. En algunos ejemplos, una construcción para recombinación homóloga incluye una secuencia de ácido nucleico que incluye al menos una porción de un gen natural que codifica un polipéptido como se proporciona en este documento, en el que el gen tiene al menos una sustitución, delección, inserción o adición de aminoácidos con respecto a un polipéptido de tipo silvestre. Por ejemplo, el gen, o una porción del mismo, puede incluir la inserción de un gen marcador seleccionable.

Una construcción de ácido nucleico para atenuación génica, por ejemplo, una construcción de ribozima, ARNi o antisentido puede incluir al menos quince, al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, o al menos sesenta nucleótidos que tienen al menos un 80% de identidad, tal como al menos 85%, al menos 90%, al menos 95%, o al menos 99% o complementariedad a al menos una porción de la secuencia de un gen de origen natural, tal como un gen que codifica un polipéptido que incluye una secuencia de aminoácidos que tiene al menos una secuencia de aminoácidos que tiene al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, o al menos 80% o al menos 85%, al menos 90%, al menos 95%,

al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 22, SEQ ID NO: 23, o SEQ ID NO: 64. Una construcción de ácido nucleico para atenuación génica, por ejemplo, una construcción de ribozima, ARNi o antisentido puede incluir al menos quince, al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, o al menos sesenta nucleótidos que tienen al menos un 80%, tal como al menos 95% o aproximadamente 100% de identidad o complementariedad con la secuencia de un gen natural, tal como un gen que codifica un polipéptido que tiene al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 90%, o al menos 95% de identidad de secuencia con SEQ ID NO: 4; SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 6; SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78. Por ejemplo, una construcción de ácido nucleico para la atenuación génica, por ejemplo, una construcción de ribozima, ARNi o antisentido puede incluir al menos quince, al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, o al menos sesenta nucleótidos que tienen al menos un 80% de identidad o complementariedad con la secuencia de un gen natural, tal como un gen que codifica un regulador de aclimatación de luz (por ejemplo, un gen LAR) como se proporciona en el presente documento. Por ejemplo, una construcción de ácido nucleico para la atenuación génica, por ejemplo, una construcción de ribozima, ARNi o antisentido puede incluir al menos quince, al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, o al menos sesenta nucleótidos que tienen al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad o complementariedad con la SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 5, SEQ ID NO: 20, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 62, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75 o SEQ ID NO: 77, o una parte del mismo.

Breve descripción de los dibujos

Figura 1 (A) Experimento de cambio de luz de *Nannochloropsis* tipo silvestre experimentando fisiologías de aclimatación de luz intensa (cuadrados) y luz limitada (triángulos). Los cultivos se hicieron crecer durante 4,6 días a $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ (luz intensa) antes de reducir las intensidades a $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ (poca luz, triángulos) para que uno de los cultivos muestre la diferencia en la fluorescencia media de la clorofila entre los tratamientos. (B) El día 8, se midieron medidas fotofisiológicas de qP en 20 irradiancias para ilustrar cómo las células cultivadas con luz intensa (cuadrados) tienen un qP más alto en comparación con las células cultivadas con luz limitada (triángulos).

La Figura 2 proporciona un mapa del vector pSG-5534 (SEQ ID NO: 1), uno de varios vectores usados para la mutagénesis de inserción de *Nannochloropsis gaditana*, que incluía un promotor TCTP (SEQ ID NO: 2) que conduce un gen de resistencia a blasticidina de *Apergillus* (SEQ ID NO: 59), codón optimizado para *Nannochloropsis*. El sitio PI-PspI se usó para la linealización antes de la transformación.

La Figura 3 muestra el aumento en la fluorescencia de clorofila por célula en un cultivo de tipo silvestre durante la aclimatación a baja irradiación junto con un potencial mutante LIHLA. Las barras representan días sucesivos después de la transferencia a poca luz. Mientras que el cultivo de tipo silvestre se aclimata al aumentar su contenido de clorofila aproximadamente 2,5 veces, el contenido de clorofila del mutante permanece muy bajo, mostrando solo un aumento modesto.

La Figura 4 representa los resultados del monitoreo de la fluorescencia de clorofila durante 5 días (columnas de izquierda a derecha: día 1, día 2, día 3, día 4, día 5) durante el cultivo de aclimatación con poca luz de la cepa LIHLA GE-4574 y cuatro réplicas biológicas de controles de tipo silvestre (WT-3730). La fluorescencia media de la clorofila por célula se determinó mediante citometría de flujo.

La Figura 5 proporciona un gráfico del coeficiente de inactivación fotoquímico (qP) para 12 irradiancias diferentes para cultivos limitados en la luz de la cepa GE-4574 (diamantes) y réplicas biológicas del control de tipo silvestre WT-3730 (cuadrados y triángulos).

La Figura 6 proporciona un resumen de las propiedades fotofisiológicas de los mutantes de LIHLA obtenidos mediante la selección de mutantes con baja concentración de clorofila después de una aclimatación con poca luz. La tabla proporciona valores máximos de ETR_{PSII} (presentados como diferencia en veces del tipo silvestre), P_{max} por célula y P_{max} por miligramo de clorofila, y la cantidad de clorofila por célula para cada mutante. Las tres columnas más a la derecha proporcionan valores de P_{max} de tipo silvestre por célula y valores de P_{max} por miligramo de clorofila, y la cantidad de clorofila por célula a partir de la misma aclimatación con poca luz.

La Figura 7 proporciona una curva NPQ para un mutante LIHLA que tiene una mutación en el gen LAR1 (diamantes abiertos) en comparación con células de tipo silvestre (cuadrados negros) cuando tanto el mutante LIHLA como las células de tipo silvestre se aclimatan a poca luz.

- La figura 8 a) proporciona un gráfico de ETR_{PSII} para la cepa mutante LAR1 LIHLA GE-5440 (triángulos), la cepa mutante LAR2 LIHLA GE-5404 (cuadrados) y la cepa silvestre (WT-3730, diamantes) en un rango de irradiancias ; b) es un gráfico que representa NPQ en un intervalo de irradiancias para el mutante LAR1 LIHLA 5440 (triángulos), mutante LAR2 LIHLA 5404 (cuadrados) y la cepa de tipo silvestre WT-3730 (diamantes); c) es un gráfico que muestra qP en respuesta a la irradiancia para el mutante LAR1 LIHLA 5440 (triángulos), mutante LAR2 LIHLA 5404 (cuadrados) y la cepa de tipo silvestre WT-3730 (diamantes); d) muestra las curvas P/I para mutantes LAR1 aclimatados a la luz intensa (triángulos abiertos) o con poca luz (diamantes negros) así como para células de tipo silvestre aclimatadas a poca luz (círculos negros) y células silvestres aclimatadas a poca luz (cuadrados negros)
- La Figura 9 es un gráfico que representa las velocidades de transporte de electrones del fotosistema II (ETR_{PSII}) en función de la irradiancia para el mutante LIHLA LAR1 GE-5440 (diamantes abiertos) y WT-3730 de tipo silvestre (cuadrados negros) en un intervalo de intensidades de luz.
- La Figura 10 es un gráfico que representa las velocidades de transporte de electrones del fotosistema I en función de la irradiancia para el mutante LIHLA LAR1 GE-5440 (diamantes abiertos) y el tipo silvestre WT-3730 (cuadrados negros) en un intervalo de intensidades de luz.
- La Figura 11 muestra la penetración de la luz a varias profundidades (centímetros desde el fondo) en dos días diferentes después de la inoculación con cultivos tipo silvestre (diamantes [día 1] y triángulos [día 7]) y mutante LAR1 GE-4574 (cuadrados [día 1] y Xs [día 7]). La densidad celular en el día 1 fue de 2×10^6 células/ml. La densidad celular el día 7 fue de $4,5 \times 10^7$ células/ml.
- La Figura 12 representa gráficamente la cantidad de luz que alcanza varias profundidades de 400 L de minipondios operados en un invernadero. Los cuadrados son intensidades de luz medidas a partir de un cultivo del mutante LIHLA 4906 (mutado en el gen LAR2), y los diamantes son intensidades de luz medidas a partir de un cultivo de *Nannochloropsis* de tipo silvestre (WT-3730). El eje y muestra la profundidad (distancia desde la superficie) en centímetros.
- La Figura 13 es un mapa del locus genético que codifica el gen LAR1 en *Nannochloropsis gaditana* que muestra los sitios de las lesiones genéticas encontradas en varios mutantes de LIHLA. El dominio del dedo de cinc Pfam PF02135 TAZ (región azul, marcado como "Pfam 02135") está en la porción C-terminal de la proteína.
- La Figura 14 es un mapa del locus genético que codifica el gen LAR2 en *Nannochloropsis gaditana* que muestra los sitios de lesiones genéticas encontradas en varios mutantes de LIHLA. El dominio de unión a ADN de pfam PF00249 tipo myb (azul, marcado como "Pfam 00249") está en la parte N-terminal de la proteína.
- La Figura 15 es un gráfico de los valores de expresión de LAR1 y LAR2 a partir del análisis de ARN-seq utilizando bibliotecas de ARN para varias condiciones abióticas, así como durante los cambios de luz que demuestran la similitud altamente significativa de sus perfiles de expresión.
- La Figura 16 proporciona un alineamiento de secuencia de la secuencia de proteína LAR1 de *N. gaditana* con la secuencia de proteína LAR1 de *N. oceanica* ("No-LAR1").
- La Figura 17 proporciona un árbol filogenético bioinformático que muestra las relaciones entre homólogos de LAR1.
- La Figura 18 proporciona un alineamiento de secuencia de la secuencia de proteína LAR2 de *N. gaditana* con la secuencia de proteína LAR2 de *N. oceanica* ("No-LAR2").
- La Figura 19 representa gráficamente la expresión en veces normalizada con respecto a los niveles de expresión de tipo silvestre de varios genes medidos por qRT-PCR en mutantes LIHLA GE-4906 (mutado en el gen LAR2), GE-5409 (mutado en el gen LAR1) y GE-5440 (mutación insercional en el gen LAR1). La expresión en veces normalizada para estos mismos genes a partir de datos transcriptómicos para GE-4574 (mutado en el gen LAR1) se proporciona como referencia.
- La Figura 20 a) es un diagrama que muestra la aclimatación de las células de *Nannochloropsis* (en experimentos paralelos, tanto de tipo silvestre como de un mutante LIHLA LAR1) a la luz intensa ($500 \mu\text{mol fotones m}^{-2} \text{ s}^{-1}$), después de lo cual (en el tiempo 0 en el eje x) las células se transfirieron a poca luz durante dos días, mientras que las células de control se mantuvieron bajo condiciones de luz intensa durante dos días, b) proporciona el contenido de clorofila de las células durante el cambio de luz: círculos, células de tipo silvestre mantenidas con luz intensa después del tiempo 0; cuadrados, células de tipo silvestre transferidas desde luz intensa a poca luz en el tiempo 0, donde permanecieron durante dos días; triángulos abiertos, células mutantes LAR1 mantenidas con luz intensa después del tiempo 0; diamantes, células mutantes LAR1 transferidas de luz intensa a poca luz en el tiempo 0, donde permanecieron durante dos días, c) proporciona las curvas P/I para las células mutantes y de tipo silvestre dos días después del cambio de luz (tiempo 0): círculos, células de tipo silvestre mantenidas con luz intensa después del tiempo 0; cuadrados, células de tipo silvestre transferidas desde luz intensa a poca luz en el tiempo 0, donde permanecieron durante dos días; triángulos abiertos, células mutantes LAR1 mantenidas con luz intensa después del tiempo 0; diamantes, células mutantes LAR1 transferidas de luz intensa a poca luz en el tiempo 0, donde permanecieron durante dos días.

La Figura 21 proporciona un diagrama de los resultados del análisis de los niveles de transcripciones (cada transcripción representada por un punto) en el mutante LAR1 con respecto al tipo silvestre en condiciones de poca luz, con abundancia de transcripción relativa en el mutante reflejada en la posición a lo largo del eje y. Las mismas transcripciones se ubican a lo largo del eje x para reflejar su abundancia relativa en células de tipo silvestre aclimatadas a la luz escasa en comparación con sus niveles en células aclimatadas con luz intensa. Los genes se dividen en TRAC en función de si la expresión diferencial en el tipo mutante frente al natural sigue un patrón similar o diferente en el tipo silvestre aclimatado a la luz intensa frente al tipo silvestre aclimatado a la luz escasa. Las posiciones a lo largo del eje horizontal difieren para reflejar el efecto del entorno de luz sobre la expresión de las transcripciones individuales; las posiciones a lo largo del eje y varían en relación con el nivel de expresión en el mutante con respecto al tipo silvestre. Las líneas punteadas corresponden a los valores \log_2 de 1 ó -1, de modo que los puntos fuera de estos límites representan los genes que tienen un mayor aumento o disminución de 2 veces en la abundancia.

La Figura 22 proporciona gráficas de puntos lado a lado de transcripciones "TRAC I" que tienen el mismo patrón de regulación en el mutante LAR1 con respecto al tipo silvestre en condiciones de poca luz (diagrama izquierdo) que en células de tipo silvestre transferidas a la luz intensa frente a sus bajos niveles de luz (diagrama de la derecha). Una línea vertical punteada en cada trazo marca la posición a lo largo del eje x donde el cambio de \log_2 veces es cero, donde cuando no hay diferencia en el nivel de expresión, indica que la expresión del gen en el mutante es el doble o mayor que los niveles en el tipo silvestre.

La Figura 23 proporciona gráficas de puntos lado a lado de transcripciones "TRAC I" que tienen el mismo patrón de regulación en el mutante LAR2 con respecto al tipo silvestre en condiciones de poca luz (diagrama izquierdo) que en células naturales transferidas a niveles de luz intensa frente a su poca luz (diagrama de la derecha). Una línea vertical punteada en cada trazo marca la posición a lo largo del eje x, donde el cambio de \log_2 veces es cero, donde cuando no hay diferencia en el nivel de expresión indica que la expresión del gen en el mutante es el doble o mayor que los niveles en el tipo silvestre.

La Figura 24 proporciona gráficos de la evolución del oxígeno y el contenido total de clorofila por célula de *Nannochloropsis oceanica* tipo silvestre (WE5473) y cuatro cepas inactivadas de LIHLA: GE6054 y GE6055 son genes inactivados del gen LAR1, y GE6052 y GE6053 son genes inactivados del gen LAR2.

La Figura 25 representa la extinción fotoquímica (qP) en *Nannochloropsis oceanica* de tipo silvestre (WE5473) (diamantes sólidos), a) representa dos cepas de inactivación de LIHLA LAR1, GE6054 y GE6055; y b) representa dos cepas inactivadas LAR2, GE6052 y GE6053.

La Figura 26 es un gráfico que representa la correlación en niveles de expresión desregulados de transcritos en mutantes LAR1 con los niveles de expresión en mutantes LAR2.

La Figura 27 es un mapa de una construcción de ARNi utilizada para inactivar el gen LAR1 en *N. gaditana*.

La Figura 28 representa los niveles de expresión del gen LAR1 en aislados de *Nannochloropsis gaditana* (1-7) transformados con construcciones de ARNi que se dirigen al gen LAR1, con dos aislados de tipo silvestre (WT) mostrados como controles.

La Figura 29 representa las densidades celulares de tipo silvestre (WT-3730) (diamantes sólidos) y cinco mutantes de LIHLA en cultivos que crecieron durante 14 días en cultivos de laboratorio a escala que simulaban las condiciones del estanque.

La Figura 30 es un gráfico de barras que proporciona la productividad promedio por día medida por la acumulación total de carbono orgánico por un mutante LIHLA y la cepa WT-3730 natural de *Nannochloropsis* cultivada en cultivos de laboratorio a escala simulando condiciones de estanque, en las que los cultivos se diluyeron diariamente en un 15%. Se muestran los resultados de dos experimentos independientes.

La Figura 31 proporciona gráficos de evolución de oxígeno y contenido total de clorofila por célula de *Nannochloropsis oceanica* tipo silvestre (WE5473), mutante LIHLA GE-5440 y tres cepas mutantes rescatadas: GE5948, GE5950 y GE5951 todas sobreexpresando la proteína LAR1 en el fondo del mutante GE-5440.

La Figura 32 proporciona gráficos que representan la inactivación fotoquímica (qP), fluorescencia mínima (F_0) y fluorescencia máxima (F_m) en *Nannochloropsis oceanica* tipo silvestre (WE5473), mutante LIHLA GE-5440 y tres cepas mutantes rescatadas: GE5948, GE5950 y GE5951, todos sobreexpresan la proteína LAR1 en el fondo del mutante GE-5440.

La Figura 33 proporciona un diagrama de una región del genoma de *Nannochloropsis gaditana* que muestra donde el vector mutagenizante en un mutante LAR3 es insertado entre los genes 7250 y 7251.

La Figura 34 es un gráfico de barras que representa los niveles de expresión determinados por qRT-PCR del gen 7251 y tres genes de LHC en relación con los niveles de tipo silvestre. El eje y es con escala \log_2 .

La Figura 35 es un gráfico de barras que representa los niveles de fluorescencia de clorofila del mutante LAR3 GE-5489 transformado con varios genes considerados posibles loci de la mutación LAR3.

La figura 36 a) proporciona un mapa del locus LAR3 y b) una construcción diseñada para la recombinación homóloga en el locus LAR3 que incluye el gen "blast" (SEQ ID NO: 59) dirigido por el promotor TCTP de *Nannochloropsis* (SEQ ID NO: 2) como un marcador seleccionable.

La Figura 37 proporciona parámetros fotofisiológicos de la cepa 5473 de tipo silvestre de *Nannochloropsis oceanica* y mutantes modificados genéticamente de LAR1 (cepa 6054 inactivada), LAR2 (cepa 6053 inactivada) y LAR3 (cepa 6038 inactivada) en el fondo de *Nannochloropsis oceanica*. A) proporciona la clorofila por célula de tipo silvestre y mutantes LAR1, LAR2 y LAR3 modificados por ingeniería genética. B) proporciona P_{max} por mg de clorofila para la cepa de tipo silvestre y mutantes de LAR1 y LAR3 modificados por ingeniería genética. C) proporciona valores promedio de E_k para mutantes de tipo silvestre y LAR1, LAR2 y LAR3 modificados por ingeniería genética. D) proporciona ETR_{PSII} en un rango de intensidades de luz para mutantes de tipo silvestre y LAR1, LAR2 y LAR3 de tipo silvestre. E) proporciona valores de qP en un intervalo de intensidades de luz para mutantes de tipo silvestre y LAR1, LAR2 y LAR3 modificados por ingeniería genética.

La Figura 38 es un alineamiento de secuencia del dominio conservado del gen LAR3 con las regiones homólogas de otros genes heterokontes.

La Figura 39 proporciona gráficas de puntos lado a lado de transcritos "TRAC 1" que tienen el mismo patrón de regulación en el mutante LAR3 con respecto al tipo silvestre en condiciones de poca luz (diagrama izquierdo) que en células de tipo silvestre transferidas a la luz intensa frente a sus niveles de poca luz (diagrama de la derecha). Una línea vertical punteada en cada trazo marca la posición a lo largo del eje x donde el cambio \log_2 veces es cero, donde cuando no hay diferencia en el nivel de expresión indica que la expresión del gen en el mutante es el doble o mayor que los niveles en el tipo silvestre.

La Figura 40 es un gráfico que proporciona datos de expresión de experimentos de ARN-seq usando ARN aislado de células de tipo silvestre ("WE-3730"), mutante LAR2 (GE-5404) y LAR3 (células GE-5489), demostrando la expresión reducida característica de al menos dieciocho LHC en los mutantes LAR con respecto a las células de tipo silvestre aclimatadas a poca luz.

La Figura 41 proporciona gráficas de puntos lado a lado relativos a los niveles de expresión de transcripciones de LHC en células aclimatadas de poca luz naturales versus, en gráficos que proceden de izquierda a derecha, un mutante LAR1, un mutante LAR2 y un mutante LAR3, en el que los mutantes LAR también tienen aclimatamiento a poca luz. El gráfico de la derecha representa los niveles de expresión de los transcritos de LHC en células aclimatadas de poca luz frente a las células de tipo silvestre aclimatadas a la luz intensa. Una línea vertical discontinua en cada trazo marca la posición a lo largo del eje x donde el cambio de \log_2 veces es cero, donde no hay diferencia en el nivel de expresión. Las células de tipo silvestre aclimatadas a poca luz demuestran una expresión aumentada, en relación con las células de tipo silvestre de LHCs aclimatadas a poca luz (gráfico más a la derecha). Este patrón general se observa en los tres mutantes de LAR, donde los niveles de expresión de LHC de tipo silvestre aclimatado a poca luz para casi todos los LHC exceden los niveles de expresión en mutantes LAR aclimatados a poca luz. Los LHC de *nannochloropsis* ejemplificados son, procediendo de arriba hacia abajo: 2788 (SEQ ID NO: 79); 7958 (SEQ ID NO: 80); 810 (SEQ ID NO: 81); 7677 (SEQ ID NO: 82); 6329 (SEQ ID NO: 83); 9833 (SEQ ID NO: 84); 6477 (SEQ ID NO: 85); 9115 (SEQ ID NO: 86); 3454 (SEQ ID NO: 87); 790 (SEQ ID NO: 88); 1993 (SEQ ID NO: 89); 1373 (SEQ ID NO: 90); 4967 (SEQ ID NO: 91); 6755 (SEQ ID NO: 92); 7521 (SEQ ID NO: 93); 554 (SEQ ID NO: 94); 4250 (SEQ ID NO: 95); 171 (SEQ ID NO: 96); 4422 (SEQ ID NO: 97); 5134 (SEQ ID NO: 98); 2787 (SEQ ID NO: 99); 4249 (SEQ ID NO: 100); 3431 (SEQ ID NO: 101); y 7831 (SEQ ID NO: 102).

Descripción detallada

Definiciones

A menos que se defina lo contrario, todos los términos técnicos y científicos usados en la presente memoria tienen el mismo significado que entiende comúnmente cualquier experto habitual en la técnica a la que pertenece esta invención. En caso de conflicto, la presente aplicación, que incluye las definiciones, prevalecerá. A menos que el contexto requiera lo contrario, los términos singulares incluirán los plurales y los términos en plural incluirán el singular.

Aunque puedan usarse métodos y materiales similares o equivalentes a los descritos en este documento en la práctica o en el ensayo de la presente invención, los métodos y materiales adecuados se describen a continuación. Los materiales, métodos y ejemplos son solo ilustrativos y no pretenden ser limitantes. Otras características y ventajas de la invención serán evidentes a partir de la descripción detallada y de las reivindicaciones.

Para facilitar la comprensión de la presente invención, se definen a continuación varios términos y frases.

Tal como se usa en la presente descripción y reivindicaciones, las formas singulares "un", "uno/a" y "el/la" incluyen formas plurales, a menos que el contexto indique claramente lo contrario.

Como se usa en el presente documento, los términos "alrededor de" o "aproximadamente", cuando se hace referencia a cualquier valor numérico, pretenden indicar un valor de más o menos 10% del valor indicado. Por ejemplo, "alrededor de 50 grados C" (o "aproximadamente 50 grados C") abarca un rango de temperaturas de 45 grados C a 55 grados C, inclusive. De forma similar, "aproximadamente 100 mM" (o "aproximadamente 100 mM") abarca un intervalo de concentraciones de 90 mM a 110 mM, inclusive. Todos los intervalos proporcionados dentro de la solicitud incluyen los valores de los extremos superior e inferior del intervalo.

El término "y/o", como se usa en una frase tal como "A y/o B" en la presente memoria, pretende incluir "A y B", "A o B", "A" y "B".

El término "gen" se usa ampliamente para referirse a cualquier segmento de una molécula de ácido nucleico (típicamente ADN, pero opcionalmente ARN) que codifica un polipéptido o ARN expresado. Por lo tanto, los genes incluyen secuencias que codifican ARN expresado (que puede incluir secuencias codificantes de polipéptidos o, por ejemplo, ARN funcionales, tales como ARN ribosómicos, ARNt, ARN antisentido, microARN, ARN de horquilla corta, ribozimas, etc.). Los genes pueden comprender adicionalmente secuencias reguladoras requeridas o que afectan a su expresión, así como secuencias asociadas con la secuencia que codifica la proteína o el ARN en su estado natural, tales como, por ejemplo, secuencias de intrón, secuencias no traducidas 5' o 3', etc. En algunos ejemplos, un gen solo puede referirse a una porción que codifica una proteína de una molécula de ADN o ARN, que puede o no incluir intrones. Un gen es preferiblemente de más de 50 nucleótidos de longitud, más preferiblemente más de 100 nucleótidos de longitud, y puede tener, por ejemplo, entre 50 nucleótidos y 500.000 nucleótidos de longitud, tal como entre 100 nucleótidos y 100.000 nucleótidos de longitud o entre aproximadamente 200 nucleótidos y aproximadamente 50.000 nucleótidos de longitud, o aproximadamente 200 nucleótidos y aproximadamente 20.000 nucleótidos de longitud. Los genes se pueden obtener a partir de una variedad de fuentes, incluida la clonación a partir de una fuente de interés o la síntesis a partir de una información de secuencia conocida o predicha.

La expresión "ácido nucleico" o "molécula de ácido nucleico" se refiere a un segmento de ADN o ARN (por ejemplo, ARNm) y también incluye ácidos nucleicos que tienen estructuras modificadas (p. ej., ácidos nucleicos peptídicos, ácidos nucleicos bloqueados) o nucleobases modificadas o de origen no natural. Las moléculas de ácido nucleico pueden ser de doble cadena o monocatenarias; un ácido nucleico monocatenario que comprende un gen o una porción del mismo puede ser una cadena codificante (sentido) o una cadena no codificante (antisentido).

Una molécula de ácido nucleico puede "derivarse de" una fuente indicada, que incluye el aislamiento (en todo o en parte) de un segmento de ácido nucleico de una fuente indicada. Una molécula de ácido nucleico también puede derivarse de una fuente indicada mediante, por ejemplo, clonación directa, amplificación por PCR o síntesis artificial a partir de la fuente de polinucleótidos indicada o basada en una secuencia asociada con la fuente de polinucleótidos indicada. Los genes o moléculas de ácido nucleico derivadas de una fuente o especie particular también incluyen genes o moléculas de ácido nucleico que tienen modificaciones de secuencia con respecto a las moléculas de ácido nucleico fuente. Por ejemplo, un gen o molécula de ácido nucleico derivada de una fuente (p. ej., un gen de referencia particular) puede incluir una o más mutaciones con respecto al gen fuente o molécula de ácido nucleico que son no pretendidas o que se introducen deliberadamente, y si una o más mutaciones, incluyendo sustituciones, deleciones o inserciones, se introducen deliberadamente, las alteraciones de secuencia pueden introducirse mediante una mutación aleatoria o dirigida de células o ácidos nucleicos, mediante amplificación u otras técnicas de biología molecular, o mediante síntesis química, o cualquier combinación de las mismas. Un gen o molécula de ácido nucleico que se deriva de un gen o molécula de ácido nucleico de referencia que codifica un ARN o polipéptido funcional puede codificar un ARN o polipéptido funcional que tenga al menos 75%, al menos 80%, al menos 85% o, al menos 90%, o al menos 95% de identidad de secuencia con el ARN o polipéptido funcional referenciado o fuente, o con un fragmento funcional del mismo. Por ejemplo, un gen o molécula de ácido nucleico que se deriva de un gen o molécula de ácido nucleico referenciado que codifica un ARN o polipéptido funcional puede codificar un ARN o polipéptido funcional que tenga al menos 85%, al menos 90%, al menos 95%, al menos un 96%, al menos un 97%, al menos un 98%, o al menos un 99% de identidad de secuencia con el ARN o polipéptido funcional referenciado o fuente, o con un fragmento funcional del mismo.

Como se usa en este documento, un ácido nucleico "aislado" o proteína se elimina de su medio natural o el contexto en el que el ácido nucleico o proteína existe en la naturaleza. Por ejemplo, una proteína aislada o molécula de ácido nucleico se elimina de la célula u organismo con el que se asocia en su entorno de origen o natural. Un ácido nucleico o proteína aislada puede estar, en algunos casos, parcial o sustancialmente purificado, pero no se requiere un nivel particular de purificación para el aislamiento. Por lo tanto, por ejemplo, una molécula de ácido nucleico aislada puede ser una secuencia de ácido nucleico que ha sido extirpada del cromosoma, genoma o episoma en el que se integra en la naturaleza.

Una molécula de ácido nucleico "purificada" o una secuencia de nucleótidos, o una secuencia de proteína o polipéptido, está sustancialmente libre de material celular y componentes celulares. La molécula o proteína de ácido nucleico purificada puede estar libre de agentes químicos más allá del tampón o el disolvente, por ejemplo. "Sustancialmente libre" no pretende significar que otros componentes más allá de las nuevas moléculas de ácido nucleico sean indetectables.

Los términos "natural" y "silvestre" se refieren a una forma que se encuentra en la naturaleza. Por ejemplo, una

molécula de ácido nucleico, una secuencia de nucleótidos o una proteína de origen natural o de tipo silvestre puede estar presente y aislada de una fuente natural, y no estar modificada intencionalmente por manipulación humana.

Como se usa en la presente memoria, "atenuado" significa reducido en cantidad, grado, intensidad o resistencia. La expresión gen atenuado puede referirse a una cantidad y/o velocidad de transcripción significativamente reducida del gen en cuestión, o de traducción, plegamiento o ensamblaje de la proteína codificada. Como ejemplos no limitantes, un gen atenuado puede ser un gen mutado o alterado (por ejemplo, un gen alterado por delección parcial o total, truncamiento, desplazamiento de marco o mutación de inserción) o que tiene una expresión disminuida debido a la alteración de secuencias reguladoras de genes.

"Molécula de ácido nucleico exógena" o "gen exógeno" se refiere a una molécula o gen de ácido nucleico que se ha introducido ("transformado") en una célula. Una célula transformada se puede denominar célula recombinante, en la que se pueden introducir gen(es) exógeno(s) adicional(es). Un descendiente de una célula transformada con una molécula de ácido nucleico también se denomina "transformada" si ha heredado la molécula de ácido nucleico exógena. El gen exógeno puede ser de una especie diferente (y, por lo tanto, "heteróloga"), o de la misma especie (y, por lo tanto, "homóloga"), con relación a la célula que se transforma. Una molécula, gen o proteína de ácido nucleico "endógena" es una molécula, gen o proteína de ácido nucleico nativo tal como se produce en el huésped, o como es producido naturalmente por éste.

El término "nativo" se usa en este documento para referirse a secuencias de ácidos nucleicos o secuencias de aminoácidos ya que se producen de forma natural en el huésped. La expresión "no nativo" se usa en el presente documento para referirse a secuencias de ácidos nucleicos o secuencias de aminoácidos que no se producen naturalmente en el huésped. Una secuencia de ácido nucleico o una secuencia de aminoácidos que se ha eliminado de una célula, que se ha sometido a manipulación en el laboratorio y se ha introducido o reintroducido en una célula huésped, se considera "no nativa". Los genes sintéticos o parcialmente sintéticos introducidos en una célula huésped son "no nativos". Los genes no nativos incluyen además genes endógenos al microorganismo huésped ligados operativamente a una o más secuencias reguladoras heterólogas que se han recombinado en el genoma del huésped.

Una molécula de ácido nucleico "recombinante" o "modificada" es una molécula de ácido nucleico que se ha alterado mediante manipulación humana. Como ejemplos no limitantes, una molécula de ácido nucleico recombinante incluye cualquier molécula de ácido nucleico que: 1) ha sido sintetizada o modificada total o parcialmente in vitro, por ejemplo, usando técnicas químicas o enzimáticas (por ejemplo, mediante el uso de síntesis química de ácidos nucleicos, o mediante el uso de enzimas para la replicación, polimerización, digestión (exonucleolítica o endonucleolítica), ligadura, transcripción inversa, transcripción, modificación de bases (incluyendo, por ejemplo, metilación), integración o recombinación (incluyendo recombinación homóloga y específica de sitio) de moléculas de ácido nucleico); 2) incluye secuencias de nucleótidos conjuntas que no están combinadas en la naturaleza, 3) se ha diseñado usando técnicas de clonación molecular tales que carece de uno o más nucleótidos con respecto a la secuencia de moléculas de ácido nucleico natural, y/o 4) ha sido manipulado usando técnicas de clonación molecular tales que tiene uno o más cambios de secuencia o reordenamientos con respecto a la secuencia de ácido nucleico natural. Como ejemplos no limitantes, un ADNc es una molécula de ADN recombinante, como cualquier molécula de ácido nucleico que se ha generado por reacción(es) de polimerasa in vitro, o a la que se han unido enlazantes, o que se ha integrado en un vector, tal como un vector de clonación o vector de expresión.

La expresión "proteína recombinante", como se usa en el presente documento, se refiere a una proteína producida por ingeniería genética.

Cuando se aplica a organismos, el término recombinante, modificado o modificado genéticamente se refiere a organismos que han sido manipulados mediante la introducción de una secuencia de ácido nucleico recombinante heteróloga o exógena en el organismo, e incluye genes inactivados, mutaciones dirigidas y reemplazo de genes, reemplazo del promotor, eliminación o inserción, así como la introducción de transgenes o genes sintéticos en el organismo. Los organismos recombinantes o modificados genéticamente también pueden ser organismos en los que se han introducido construcciones para el gen "knock down". Dichos constructos incluyen, pero sin limitación, construcciones de ARNi, microARN, ARNhc, ARNip, antisentido y ribozima. También se incluyen organismos cuyos genomas han sido alterados por la actividad de meganucleasas o nucleasas con dedos de cinc. Una molécula de ácido nucleico exógena o recombinante se puede integrar en el genoma del organismo recombinante/modificado genéticamente o en otros casos no se integra en el genoma del organismo recombinante/modificado genéticamente. Como se usa en este documento, "microorganismo recombinante" o "célula hospedadora recombinante" incluye una progenie o derivados de los microorganismos recombinantes. Debido a que ciertas modificaciones pueden ocurrir en las generaciones sucesivas debido a la mutación o influencias ambientales, tal progenie o derivados pueden no ser, de hecho, idénticos a la célula parental, pero aún están incluidos dentro del alcance del término como se usa en el presente documento.

El término "promotor" se refiere a una secuencia de ácido nucleico capaz de unirse a la ARN polimerasa en una célula e iniciar la transcripción de una secuencia codificante corriente abajo (dirección 3'). Un promotor incluye la cantidad mínima de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Un promotor puede incluir un sitio de iniciación de la transcripción así como dominios de unión a proteínas

(secuencias consenso) responsables de la unión de la ARN polimerasa. Los promotores eucarióticos a menudo, pero no siempre, contienen cajas "TATA" y cajas "CAT". Los promotores procariotas pueden contener -10 y -35 secuencias consenso promotoras procarióticas. En la técnica se conoce un gran número de promotores, que incluyen promotores constitutivos, inducibles y reprimibles, de una variedad de fuentes diferentes. Las fuentes representativas incluyen, por ejemplo, tipos de células de algas, virus, mamíferos, insectos, plantas, levaduras y bacterias, y los promotores adecuados de estas fuentes están fácilmente disponibles o pueden prepararse sintéticamente, basándose en secuencias públicamente disponibles en línea o, por ejemplo, de los depósitos tales como la ATCC, así como otras fuentes comerciales o individuales. Los promotores pueden ser unidireccionales (iniciar la transcripción en una dirección) o bidireccionales (iniciar la transcripción en cualquier dirección). Un promotor puede ser un promotor constitutivo, un promotor reprimible o un promotor inducible.

El término "heterólogo", cuando se usa en referencia a un polinucleótido, gen, ácido nucleico, polipéptido o enzima se refiere a un polinucleótido, gen, ácido nucleico, polipéptido o enzima que proviene de una fuente o se deriva de una fuente distinta de la especie de organismo huésped. Por el contrario, se usa en este documento un polinucleótido, gen, ácido nucleico, polipéptido o enzima "homólogo" para denotar un polinucleótido, gen, ácido nucleico, polipéptido o enzima que se deriva de la especie del organismo hospedador. Cuando se hace referencia a una secuencia reguladora génica o a una secuencia auxiliar de ácido nucleico utilizada para mantener o manipular una secuencia génica (por ejemplo, un promotor, una región 5' no traducida, una región 3' no traducida, una secuencia de adición de poli A, una secuencia intrónica, un sitio de empalme, un sitio de unión de ribosoma, secuencia de entrada de ribosoma interna, región de homología de genoma, sitio de recombinación, etc.), "heterólogo" significa que la secuencia reguladora o secuencia auxiliar no está asociada de forma natural con el gen con el que la secuencia de ácido nucleico reguladora o auxiliar se yuxtapone en un constructo, genoma, cromosoma o episoma. Por lo tanto, un promotor unido operativamente a un gen al que no está unido operativamente en su estado natural (es decir, en el genoma de un organismo no genéticamente modificado) se denomina en este documento "promotor heterólogo", aunque el promotor puede derivar de la misma especie (o, en algunos casos, el mismo organismo) como el gen al que está vinculado.

Como se usa en el presente documento, el término "proteína" o "polipéptido" pretende abarcar un "polipéptido" singular así como varios "polipéptidos", y se refiere a una molécula compuesta de monómeros (aminoácidos) unidos linealmente por enlaces amida (también conocidos como enlaces peptídicos). El término "polipéptido" se refiere a cualquier cadena o cadenas de dos o más aminoácidos, y no se refiere a una longitud específica del producto. Por lo tanto, los péptidos, dipéptidos, tripéptidos, oligopéptidos, "proteína", "cadena de aminoácidos" o cualquier otro término utilizado para referirse a una cadena o cadenas de dos o más aminoácidos se incluyen dentro de la definición de "polipéptido" y el término "polipéptido" puede usarse en lugar de, o de manera intercambiable, con cualquiera de estos términos.

Los números de acceso genético y proteico, comúnmente proporcionados en este documento entre paréntesis después del nombre de un gen o especie, son identificadores únicos para un registro de secuencia públicamente disponible en el sitio web del Centro Nacional de Información Biotecnológica (NCBI) (ncbi.nlm.nih.gov) mantenido por los Institutos Nacionales de Salud de los Estados Unidos. El número de identificación de secuencia "identificador GenInfo" (GI) es específico de una secuencia de nucleótidos o aminoácidos. Si una secuencia cambia de alguna manera, se asigna un nuevo número de GI. Se encuentra disponible una herramienta de historial de revisiones de secuencia para rastrear los diversos números de GI, números de versión y fechas de actualización para las secuencias que aparecen en un registro de GenBank específico. La búsqueda y obtención de secuencias de ácidos nucleicos o genes o secuencias de proteínas basadas en números de acceso y números GI es bien conocida en las técnicas de, por ejemplo, biología celular, bioquímica, biología molecular y genética molecular.

Como se usa en este documento, las expresiones "identidad porcentual" u "homología", con respecto a las secuencias de ácidos nucleicos o polipéptidos, se definen como el porcentaje de residuos de nucleótidos o aminoácidos en la secuencia candidata que son idénticos a los polipéptidos conocidos, después de alinear las secuencias para el máximo porcentaje de identidad e introducción de huecos, si es necesario, para lograr el máximo porcentaje de homología. Las inserciones o deleciones N-terminales o C-terminales no se interpretará que afectan a la homología, y las deleciones y/o inserciones internas en la secuencia polipeptídica de menos de aproximadamente 30, menos de aproximadamente 20 o menos de aproximadamente 10 residuos de aminoácidos no se interpretará como que afectan a la homología. La homología o identidad a nivel de secuencia de nucleótidos o aminoácidos se puede determinar mediante el análisis BLAST (herramienta de búsqueda de alineación local básica) utilizando el algoritmo empleado por los programas blastp, blastn, blastx, tblastn y tblastx (Altschul (1997), *Nucleic Acids Res* 25, 3389 - 3402, y Karlin (1990), *Proc. Natl. Acad. Sci. USA* 87, 2264-2268), que están diseñados para la búsqueda de similitud de secuencia. El enfoque utilizado por el programa BLAST es primero considerar segmentos similares, con y sin espacios, entre una secuencia problema y una secuencia de la base de datos, luego evaluar la significación estadística de todas las coincidencias que se identifican, y finalmente resumir solo aquellas coincidencias que satisfacen un umbral de significación preseleccionado. Para una discusión de cuestiones básicas en la búsqueda de similitud de bases de datos de secuencias, véase Altschul (1994), *Nature Genetics* 6, 119-129. Los parámetros de búsqueda para el histograma, las descripciones, alineaciones, espera (es decir, el umbral de significación estadística para informar de coincidencias con las secuencias de las bases de datos), corte, matriz y filtro (baja complejidad) pueden tener la configuración predeterminada. La matriz de puntuación predeterminada utilizada por blastp, blastx, tblastn y tblastx es la matriz BLOSUM62 (Henikoff (1992), *Proc. Natl. Acad. Sci. USA* 89, 10915-10919),

recomendada para secuencias problema de más de 85 de longitud (bases de nucleótidos o aminoácidos).

Para blastn, diseñado para comparar secuencias de nucleótidos, la matriz de puntuación se establece por las relaciones de M (es decir, el puntaje de recompensa para un par de residuos coincidentes) a N (es decir, la puntuación de penalización para residuos no coincidentes), donde los valores predeterminados para M y N pueden ser +5 y -4, respectivamente. Se pueden ajustar cuatro parámetros de blastn de la siguiente manera: Q = 10 (penalización de creación de hueco); R = 10 (penalización por extensión de hueco); wink = l (genera aciertos de palabra en cada posición wink a lo largo de la consulta); y gapw = 16 (establece el ancho de la ventana dentro de la cual se generan las alineaciones de huecos). Los ajustes de parámetros de Blastp equivalentes para la comparación de secuencias de aminoácidos pueden ser: Q = 9; R = 2; wink = l; y gapw = 32. Una comparación Bestfit entre secuencias, disponible en el paquete GCG versión 10.0, puede usar los parámetros de ADN GAP = 50 (penalización de creación de hueco) y LEN = 3 (penalización de extensión de hueco), y las configuraciones equivalentes en las comparaciones de proteínas pueden ser GAP = 8 y LEN = 2.

Por lo tanto, cuando se hace referencia a las secuencias de polipéptido o ácido nucleico de la presente invención, se incluyen identidades de secuencia de al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 70%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o aproximadamente 100% de identidad de secuencia con el polipéptido de longitud completa o secuencia de ácido nucleico, o fragmentos de los mismos que comprenden una secuencia consecutiva de al menos 50, al menos 75, al menos 100, al menos 125, al menos 150 o más residuos de aminoácidos de la proteína completa; variantes de tales secuencias, por ejemplo, en las que se ha insertado al menos un residuo de aminoácido N- y/o C-terminal en, y/o dentro de, la(s) secuencia(s) descrita(s) que contiene(n) la inserción y la sustitución. Las variantes contempladas pueden incluir adicional o alternativamente aquellas que contienen mutaciones predeterminadas mediante, p. ej., recombinación homóloga o mutagénesis dirigida a sitio o PCR, y los polipéptidos o ácidos nucleicos correspondientes de otras especies, que incluyen, pero no se limitan a, los descritos en este documento, los alelos u otras variantes naturales de la familia de polipéptidos o ácidos nucleicos que contienen una inserción y sustitución; y/o derivados en los que el polipéptido se ha modificado covalentemente por medios de sustitución, químicos, enzimáticos u otros apropiados con un resto distinto de un aminoácido de origen natural que contiene la inserción y la sustitución (por ejemplo, un resto detectable, tal como una enzima).

Como se usa en el presente documento, la frase "sustitución conservativa de aminoácidos" o "mutación conservativa" se refiere a la sustitución de un aminoácido por otro aminoácido con una propiedad común. Una forma funcional de definir propiedades comunes entre aminoácidos individuales es analizar las frecuencias normalizadas de cambios de aminoácidos entre las proteínas correspondientes de organismos homólogos (Schulz (1979) Principles of Protein Structure, Springer-Verlag). De acuerdo con tales análisis, se pueden definir grupos de aminoácidos donde los aminoácidos dentro de un grupo se intercambian preferentemente entre sí y, por lo tanto, se parecen mucho en su impacto sobre la estructura proteica general (Schulz (1979) Principles of Protein Structure, Springer-Verlag). Los ejemplos de grupos de aminoácidos definidos de esta manera pueden incluir: un "grupo cargado/polar" que incluye Glu, Asp, Asn, Gln, Lys, Arg y His; un "grupo aromático o cíclico" que incluye Pro, Phe, Tyr y Trp; y un "grupo alifático" que incluye Gly, Ala, Val, Leu, Ile, Met, Ser, Thr y Cys. Dentro de cada grupo, los subgrupos también se pueden identificar. Por ejemplo, el grupo de aminoácidos cargados/polares se puede subdividir en subgrupos que incluyen: el "subgrupo con carga positiva" que comprende Lys, Arg y His; el "subgrupo cargado negativamente" que comprende Glu y Asp; y el "subgrupo polar" que comprende Asn y Gln. En otro ejemplo, el grupo aromático o cíclico se puede subdividir en subgrupos que incluyen: el "subgrupo de anillo de nitrógeno" que comprende Pro, His y Trp; y el "subgrupo de fenilo" que comprende Phe y Tyr. En otro ejemplo adicional, el grupo alifático se puede subdividir en subgrupos que incluyen: el "subgrupo alifático no polar grande" que comprende Val, Leu y Ile; el "subgrupo alifático ligeramente polar" que comprende Met, Ser, Thr y Cys; y el "subgrupo de residuos pequeños" que comprende Gly y Ala. Los ejemplos de mutaciones conservativas incluyen sustituciones de aminoácidos dentro de los subgrupos anteriores, tales como, aunque sin limitación: Lys para Arg o viceversa, de manera que se puede mantener una carga positiva; Glu para Asp o viceversa, de manera que se pueda mantener una carga negativa; Ser para Thr o viceversa, de modo que se pueda mantener un -OH libre; y Gln para Asn o viceversa, de modo que se pueda mantener un -NH₂ libre. Una "variante conservativa" es un polipéptido que incluye uno o más aminoácidos que se han sustituido para reemplazar uno o más aminoácidos del polipéptido de referencia (por ejemplo, un polipéptido cuya secuencia se describe en una publicación o base de datos de secuencias, o cuya secuencia ha sido determinada por secuenciación de ácidos nucleicos) con un aminoácido que tiene propiedades comunes, por ejemplo, que pertenece al mismo grupo o subgrupo de aminoácidos como se delineó anteriormente.

Como se usa en el presente documento, "expresión" incluye la expresión de un gen al menos en el nivel de producción de ARN, y un "producto de expresión" incluye el producto resultante, por ejemplo, un polipéptido o ARN funcional (p. ej., un ARN ribosomal, un ARNt, un ARN antisentido, un micro ARN, un ARNhc, una ribozima, etc.), de un gen expresado. La expresión "expresión aumentada" incluye una alteración en la expresión génica para facilitar la producción de ARNm aumentada y/o la expresión de polipéptido aumentada. La "producción incrementada" incluye un aumento en la cantidad de expresión del polipéptido, en el nivel de la actividad enzimática de un polipéptido, o una combinación de ambos, en comparación con la producción nativa o la actividad enzimática del polipéptido.

Algunos aspectos de la presente invención incluyen la delección, silenciamiento, inactivación o disminución incompleta, sustancial o completa de la expresión de secuencias de polinucleótidos particulares. Los genes pueden estar parcial, sustancialmente o completamente eliminados, silenciados, inactivados, o su expresión puede estar regulada negativamente para afectar la actividad realizada por el polipéptido que codifican, tal como la actividad de una enzima. Los genes pueden deleccionarse parcial, sustancialmente o completamente, silenciarse, inactivarse o regularse negativamente mediante la inserción de secuencias de ácido nucleico que interrumpen la función y/o expresión del gen (p. ej., Inserción viral, mutagénesis por transposón, ingeniería de meganucleasas, recombinación homóloga, u otros métodos conocidos en la técnica). Los términos "eliminar", "eliminación" y "inactivación" se pueden usar indistintamente con los términos "delección", "delección parcial", "delección sustancial" o "delección completa". En ciertas realizaciones, puede modificarse por ingeniería genética un microorganismo de interés mediante recombinación homóloga dirigida al sitio para eliminar un gen particular de interés. En otras realizaciones más, los constructos de ARNi o ADN antisentido (ADNas) pueden usarse para silenciar, inactivar o regular parcialmente, sustancialmente o completamente, un gen particular de interés.

Se puede entender que estas inserciones, delecciones u otras modificaciones de ciertas moléculas de ácido nucleico o secuencias de polinucleótidos particulares abarcan "modificación(es) genética(s)" o "transformación(es)" de manera que las cepas resultantes de los microorganismos o células huésped pueden entenderse que están "modificados genéticamente", "diseñado genéticamente" o "transformados".

Tal como se usa en la presente memoria, "regulación positiva" o "sobre-regulación" incluye un aumento en la expresión de un gen o molécula de ácido nucleico de interés o la actividad de una enzima, por ejemplo, un aumento en la expresión génica o actividad enzimática en comparación con la expresión o actividad en un gen o enzima por lo demás idéntico que no se ha regulado positivamente.

Como se usa en el presente documento, "regulado por disminución" o "regulación negativa" incluye una disminución en la expresión de un gen o molécula de ácido nucleico de interés o la actividad de una enzima, p. ej., una disminución en la expresión génica o actividad enzimática, en comparación con la expresión o actividad en un gen o enzima por lo demás idéntica, que no se ha regulado negativamente.

Como se usa en el presente documento, "mutante" se refiere a un organismo que tiene una mutación en un gen que ha surgido espontáneamente o es el resultado de una mutagénesis clásica, por ejemplo, usando irradiación gamma, UV o mutágenos químicos. "Mutante", como se usa en el presente documento, también se refiere a una célula recombinante que tiene una estructura alterada o expresión de un gen como resultado de la ingeniería genética que puede incluir, como ejemplos no limitantes, la sobreexpresión, incluida la expresión de un gen en diferentes tiempos, biológicos, o regulación ambiental y/o en un grado diferente al que se produce naturalmente y/o la expresión de un gen que no se expresa naturalmente en la célula recombinante; recombinación homóloga, que incluye knock-outs y knock-ins (por ejemplo, reemplazo génico con genes que codifican polipéptidos que tienen mayor o menor actividad que el polipéptido de tipo silvestre, y/o polipéptidos dominantes negativos); atenuación génica a través de ARNi, ARN antisentido, o ribozimas, o similares; y la ingeniería del genoma utilizando meganucleasas, tecnologías TALEN y/o CRISPR, y similares.

El término "Pfam" se refiere a una gran colección de dominios proteicos y familias de proteínas mantenidas por el Consorcio Pfam y disponibles en varios sitios web patrocinados a nivel mundial, que incluyen: pfam.sanger.ac.uk/ (Welcome Trust, Sanger Institute); pfam.sbc.su.se/ (Stockholm Bioinformatics Center); pfam.janelia.org/ (Janelia Farm, Instituto Médico Howard Hughes); pfam.jouy.inra.fr/ (Institut national de la Recherche Agronomique); y pfam.cccb.re.kr. El último lanzamiento de Pfam es Pfam 26.0 (noviembre de 2011) basado en la versión de la base de datos de proteínas UniProt 15.6, un compuesto de Swiss-Prot versión 57.6 y TrEMBL versión 40.6. Los dominios y familias de Pfam se identifican utilizando alineamientos de secuencias múltiples y modelos de Markov ocultos (HMM). Las asignaciones de familia o dominio de Pfam-A son asignaciones de alta calidad generadas por un alineamiento de semillas curadas utilizando miembros representativos de una familia de proteínas y modelos de perfil de Markov ocultos basados en el alineamiento de semillas. (A menos que se especifique lo contrario, las coincidencias de una proteína problema con un dominio o familia Pfam son coincidencias Pfam-A). Todas las secuencias identificadas que pertenecen a la familia se usan entonces para generar automáticamente una alineación completa para la familia (Sonnhammer (1998) *Nucleic Acids Research* 26, 320-322; Bateman (2000) *Nucleic Acids Research* 26, 263-266; Bateman (2004) *Nucleic Acids Research* 32, Database Issue, D138-D141, Finn (2006) *Nucleic Acids Research Database Issue* 34, D247-251, Finn (2010) *Nucleic Acids Research Database Issue* 38, D211-222). Al acceder a la base de datos de Pfam, por ejemplo, utilizando cualquiera de los sitios web de referencia anteriores, las secuencias de proteínas pueden consultarse contra los HMM utilizando el software de búsqueda de homología HMMER (por ejemplo, HMMER2, HMMER3, o una versión superior, hmmer.janelia.org/). Las coincidencias significativas que identifican a una proteína consultada como perteneciente a una familia pfam (o que tienen un dominio Pfam particular) son aquellas en las que la puntuación de bit es mayor o igual que el umbral de recopilación para el dominio Pfam. Los valores de expectativa (valores e) también se pueden usar como un criterio para la inclusión de una proteína consultada en un Pfam o para determinar si una proteína consultada tiene un dominio Pfam particular, donde valores bajos de e (mucho menos que 1,0, por ejemplo menos de 0,1, o menor o igual a 0,01) representan bajas probabilidades de que una coincidencia se deba al azar.

Cuando se hace referencia a un organismo fotosintético, tal como una alga, la expresión "aclimatado a poca luz"

significa que tiene un aumento de clorofila y de las propiedades fotosintéticas del organismo fotosintético después de haber estado expuesto a una baja intensidad de luz durante un período de tiempo suficiente para que se establezcan los cambios en la clorofila y las propiedades fotosintéticas en condiciones de poca luz. Poca luz puede ser, por ejemplo, menos de $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ y preferiblemente alrededor de $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o menos o $50 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o menos, y el período de tiempo para la aclimatación puede ser de al menos aproximadamente cuatro horas, al menos aproximadamente seis horas, al menos aproximadamente ocho horas, o al menos aproximadamente doce horas, al menos 24 horas o al menos 48 horas.

Un "ADNc" es una molécula de ADN que comprende al menos una parte de la secuencia de nucleótidos de una molécula de ARNm, con la excepción de que la molécula de ADN sustituye a la nucleobase timina, o T, en lugar de uridina, o U, que aparece en la secuencia de ARNm. Un ADNc puede ser de doble cadena o monocatenario y puede ser, por ejemplo, el complemento de la secuencia de ARNm. En ejemplos preferidos, un ADNc no incluye una o más secuencias de intrón que se producen en el gen de origen natural al que corresponde el ADNc (es decir, el gen tal como aparece en el genoma de un organismo). Por ejemplo, un ADNc puede tener secuencias desde aguas arriba de un intrón de un gen que se produce naturalmente yuxtapuesto a secuencias aguas abajo del intrón del gen de origen natural, donde las secuencias aguas arriba y aguas abajo no se yuxtaponen en una molécula de ADN en la naturaleza (es decir, las secuencias no están yuxtapuestas en el gen natural). Un ADNc puede producirse mediante transcripción inversa de moléculas de ARNm, o puede sintetizarse, por ejemplo, mediante síntesis química y/o usando una o más enzimas de restricción, una o más ligasas, una o más polimerasas (que incluyen, entre otras, polimerasas tolerantes a altas temperaturas que pueden usarse en reacciones en cadena de la polimerasa (PCR), una o más recombinasas, etc., basadas en el conocimiento de la secuencia de ADNc, donde el conocimiento de la secuencia de ADNc puede basarse opcionalmente en la identificación de regiones de codificación a partir de secuencias del genoma o compiladas a partir de los cDNA de las secuencias múltiples parciales.

Un mutante de algas "desregulado en aclimatación con poca luz" (o "Bloqueado en Aclimatación de luz intensa" o mutante LIHLA) es un mutante que no exhibe los cambios en el fenotipo y la expresión génica que son característicos de una célula de alga de tipo silvestre aclimatada a poca luz, incluyendo: un aumento sustancial en la clorofila y un aumento sustancial en la expresión de la mayoría de los genes de proteína compleja de recolección de luz (LHCP). Un mutante de algas desregulado en aclimatación con poca luz, cuando se aclimata a la poca luz, tiene disminuida la expresión con respecto a las células de tipo silvestre aclimatadas a poca luz, de genes múltiples (por ejemplo, al menos diez, al menos veinte, al menos treinta, al menos cuarenta o al menos cincuenta genes) que se regulan positivamente durante la aclimatación con poca luz de células de tipo silvestre. Además, un mutante de algas desregulado en aclimatación con poca luz tiene aumentada la expresión de los genes con respecto a las células silvestres aclimatadas a poca luz (por ejemplo, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, o al menos diez genes) que se regulan negativamente durante la aclimatación con poca luz de las células de tipo silvestre. Además, como se describe en la presente memoria, un mutante de algas desregulado en aclimatación con poca luz puede tener propiedades fotosintéticas que son significativamente diferentes de las propiedades fotosintéticas de las células de tipo silvestre cuando las células mutantes y de tipo silvestre se aclimatan a poca luz.

"Propiedades fotosintéticas", "propiedades fotosintéticas", "propiedades fotofisiológicas" o "parámetros fotofisiológicos" incluyen, sin limitación, la tasa fotosintética máxima, P_{max} (calculada por célula o por mg de base de clorofila), la intensidad a la que se satura la fotosíntesis, E_k , medido por la evolución de oxígeno, y α ("alfa") la pendiente inicial de la curva de fotosíntesis (evolución de oxígeno) versus intensidad de irradiancia (P/I). Las propiedades fotosintéticas adicionales incluyen diversos parámetros que se pueden medir usando la detección de fluorescencia, que incluyen, por ejemplo, la eficiencia fotosintética, F_v/F_m ; el rendimiento cuántico fotosintético del fotosistema II (PSII), (Φ_{PSII}), (extinción fotoquímica, o la proporción de centros PSII abiertos, qP ; inactivación no fotoquímica, NPQ; velocidad de transporte de electrones PSII, ETR_{PSII} ; velocidad de transporte de electrones PSI, ETR_{PSI} ; tamaño de sección transversal de PSI, y el tamaño de corte transversal de PSII. La enumeración en este documento no es exhaustiva, y los términos no excluyen otros parámetros que midan varios aspectos de la fotosíntesis.

La referencia a las propiedades que son "sustancialmente iguales" pretende significar que las propiedades están dentro del 25%, y preferiblemente dentro del 20%, del valor de referencia.

Mutantes LIHLA

Se proporcionan en este documento mutantes de algas que se desregulan en aclimatación en condiciones de poca luz. Un mutante de algas desregulado en aclimatación en condiciones de poca luz puede ser una microalga eucariota, por ejemplo, de una especie de microalgas eucarióticas marinas o de agua dulce. La transferencia de algas silvestres de luz intensa a poca luz resulta particularmente en un aumento en la cantidad de clorofila por célula (Chl/célula), junto con niveles más altos de las proteínas del complejo de recolección de luz (LHC). En los mutantes de algas descritos en este documento, la transferencia de luz intensa a poca luz no produce un aumento sustancial de la clorofila, y los niveles de ARNm para casi todos los transcritos de proteína LHC permanecen en niveles similares a los de las células silvestres en condiciones de estado aclimatado a alta iluminación. Se dice que tales mutantes están desregulados en condiciones de aclimatación a poca luz, y se denominan en el presente documento mutantes con aclimatación a la luz intensa en compartimentos o "LIHLA". "Poca luz" o "condiciones de poca luz" se

refiere a la intensidad de la luz en las longitudes de onda de radiación fotosintéticamente activa (PAR) (de aproximadamente 400 a aproximadamente 700 nm), y puede variar según la especie de alga pero, en general, se puede considerar que es menos de $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, preferiblemente menos de $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, y más preferiblemente $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o menos, o $50 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o menos. "Luz intensa" o las "condiciones de luz intensa" también pueden variar según la especie de alga pero, en general, es al menos $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, preferiblemente al menos $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, y más preferiblemente $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o superior, y puede ser $600 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o superior. Los mutantes tienen una tasa fotosintética máxima (P_{max}) por célula comparable a (por ejemplo, dentro del 70% de tipo silvestre, dentro del 80% o tipo silvestre, o dentro del 90% o 95% de tipo silvestre) o más alta que la P_{max} de células tipo silvestre, al tiempo que permite una mayor penetración de la luz en un cultivo de algas de células mutantes. Un mutante LIHLA, como se proporciona en este documento, puede alcanzar una densidad celular más alta en cultivo en comparación con una cepa de tipo silvestre.

Las propiedades de un mutante de algas LIHLA, que incluyen, sin limitación, propiedades fotosintéticas, perfiles de expresión génica, contenido de clorofila, propiedades de crecimiento y características de cultivo, como se menciona en esta solicitud, se comparan con las mismas propiedades de un organismo de tipo silvestre de la misma especie que el mutante LIHLA, preferiblemente la cepa progenitora del mutante LIHLA. Las propiedades de un mutante LIHLA que tiene un gen LAR alterado, atenuado, o directamente o indirectamente manipulado genéticamente que da como resultado una estructura o expresión alterada del gen LAR también se puede comparar con las mismas propiedades de una célula de control que no tiene un gen LAR alterado, atenuado, o manipulado genéticamente de otra manera, directa o indirectamente, que da como resultado una estructura o expresión alterada del gen LAR (independientemente de si la célula es de tipo silvestre). Una célula de control es sustancialmente idéntica a la del mutante LIHLA excepto que no tiene un gen LAR alterado, atenuado, manipulado genéticamente, directamente o indirectamente, que da como resultado una estructura o expresión alterada del gen LAR. Por ejemplo, una célula de control puede ser una célula recombinante o una célula mutada en un gen distinto del gen LAR cuyos efectos se están evaluando, etc.

Como se demuestra en la presente memoria, un mutante LIHLA, es decir, un mutante fotosintético de algas desregulado en aclimatación a poca luz, exhibe una reducción en el contenido de clorofila en condiciones de poca luz con respecto a un alga silvestre o de control de la misma cepa en las mismas condiciones de luz. Los mutantes se caracterizan por una cantidad reducida de clorofila por célula y pueden exhibir, por ejemplo, una reducción del 20% o más en la clorofila, una reducción del 25% o mayor en la clorofila, una reducción del 30% o mayor en la clorofila, una reducción del 35% o mayor en la clorofila, una reducción del 40% o mayor en la clorofila, una reducción del 45% o mayor en la clorofila, una reducción del 50% o mayor en la clorofila, una reducción del 55% o mayor en la clorofila, una reducción del 60% o mayor en la clorofila, una reducción del 65% o mayor en la clorofila, una reducción del 70% o mayor en la clorofila, una reducción del 75% o mayor en la clorofila, o una reducción del 80% o mayor en la clorofila. Preferiblemente, un mutante LIHLA muestra una reducción en la clorofila de al menos aproximadamente 40%, al menos aproximadamente 45%, o al menos aproximadamente 50% por célula en comparación con una célula de tipo silvestre que se desarrolla en condiciones de poca luz. La reducción en la clorofila es preferiblemente una reducción en la clorofila total. En algunos ejemplos, la clorofila b no se reduce selectivamente en el mutante. En ejemplos particulares, un mutante LIHLA está en una cepa de algas que no tiene clorofila b de forma natural, por ejemplo, un mutante LIHLA puede ser una especie de algas diatomeas o eustigmatofitas.

Un mutante LIHLA también exhibe una extinción fotoquímica más alta, qP (véase, por ejemplo, Maxwell y Johnson (2000) J. Exper. Botany 51: 659-668), en absoluto irradiaciones fisiológicamente relevantes por encima de aproximadamente $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por encima de aproximadamente $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por encima de aproximadamente $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por encima de aproximadamente $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por encima de aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por encima de aproximadamente $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o por encima de aproximadamente $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ con respecto a un alga silvestre o de control de la misma cepa cuando ambos mutantes, un tipo silvestre, están aclimatados a poca luz. Una cepa LIHLA también puede exhibir una eficacia fotoquímica máxima más alta, Fv/Fm, en comparación con una cepa de tipo silvestre aclimatada a poca luz. Además, una cepa LIHLA puede tener un valor más alto para ΦPSII , la eficiencia operativa del fotosistema II, que el valor para ΦPSII de una cepa silvestre.

Además del bajo contenido en clorofila y una mayor qP con respecto a la cepa de tipo silvestre o de control, el inicio de la inactivación no fotoquímica, o NPQ, de un mutante LIHLA puede ocurrir a una intensidad de luz más alta en comparación con la intensidad de luz del inicio NPQ en una cepa tipo silvestre o de control. Asimismo, además de exhibir el inicio de NPQ bajo una intensidad de luz más alta, un mutante LIHLA exhibe menos NPQ que las células de tipo silvestre o de control en todas las intensidades de luz fisiológicamente relevantes mayores que aproximadamente $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, tal como intensidades de luz superiores a aproximadamente $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, superiores a aproximadamente $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, superiores a aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, superiores a $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o superiores a $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por ejemplo, puede exhibir un NPQ más bajo a intensidades de luz de aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ a $2000 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o de aproximadamente $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ a $2000 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

Además, un mutante LIHLA tiene una tasa fotosintética máxima más alta en una base por clorofila que una célula silvestre o de control. Por ejemplo, un mutante LIHLA puede tener aproximadamente 1,5 veces o más, entre aproximadamente 1,5 veces y aproximadamente cinco veces, entre aproximadamente 1,5 veces y aproximadamente

4 veces, entre aproximadamente 2 veces y aproximadamente 4 veces, entre aproximadamente 2 veces y aproximadamente 3,5 veces, o entre aproximadamente 2 veces y aproximadamente 3 veces la P_{max} de una célula de control o tipo silvestre. Además, un mutante LIHLA puede tener al menos aproximadamente un 70%, al menos aproximadamente un 75%, o al menos aproximadamente un 80% de la velocidad fotosintética máxima por célula como alga de tipo silvestre de la misma cepa (por ejemplo, la cepa progenitora) o una cepa de control. La tasa fotosintética se puede medir como la tasa de evolución del oxígeno, que se mide en un rango de intensidades de luz, para saturar la fotosíntesis, para obtener la tasa máxima, o P_{max} . Un mutante de LIHLA puede tener, por ejemplo, sustancialmente la misma tasa fotosintética máxima que una célula de control o de tipo silvestre o una tasa de fotosíntesis máxima más alta que una célula de control o de tipo silvestre. Por "sustancialmente la misma tasa fotosintética máxima" se entiende que la tasa fotosintética máxima es al menos aproximadamente un 80% o al menos aproximadamente un 85% de la tasa fotosintética máxima de tipo silvestre, y preferiblemente al menos aproximadamente un 90% de la tasa de tipo silvestre o a al menos aproximadamente un 95% de la tasa de tipo silvestre. Además, el mutante fotosintético de algas puede tener una mayor irradiancia de saturación para la fotosíntesis (E_k) que un alga tipo silvestre de la misma cepa o una cepa de control y, en algunos ejemplos, también puede mostrar una disminución en la pendiente inicial de la curva de irradiancia fotosintética (α) con respecto a células de tipo silvestre o células de control.

En algunos ejemplos, un mutante LIHLA tiene al menos un 20%, 25%, 30%, 35%, 40%, 45% o 50% de reducción en la clorofila con respecto a una célula de tipo silvestre, exhibe una mayor inactivación fotoquímica (qP) respecto a todas las intensidades de luz fisiológicamente relevantes mayores que $400 \mu E \cdot m^{-2} \cdot s^{-1}$ con respecto a una célula de tipo silvestre; tiene una tasa mayor de fotosíntesis máxima (P_{max}) en una base por clorofila (por ejemplo, al menos 1,5 veces la P_{max} de las células de tipo silvestre, por ejemplo, al menos 2 veces la P_{max} de las células de tipo silvestre), con al menos un 75% de la P_{max} de células de tipo silvestre por célula; experimenta la saturación de la fotosíntesis a niveles de irradiación más altos (mayor E_k) que el tipo silvestre; exhibe una aparición retrasada de NPQ en respuesta a una intensidad de luz creciente en comparación con las células de tipo silvestre; y tiene niveles más bajos de NPQ respecto a todas las irradiancias mayores que aproximadamente $500 \mu E \cdot m^{-2} \cdot s^{-1}$ que las células de tipo silvestre, cuando tanto los mutantes LIHLA como las células de tipo silvestre se aclimatan a poca luz. En algunos ejemplos, un mutante LIHLA tiene al menos un 40%, 45% o 50% de reducción en la clorofila con respecto a una célula de tipo silvestre, exhibe un aumento de la inactivación fotoquímica (qP) respecto a todas las intensidades de luz fisiológicamente relevantes mayores que $200 \mu E \cdot m^{-2} \cdot s^{-1}$ con respecto a una célula de tipo silvestre; tiene una mayor tasa fotosintética máxima (P_{max}) respecto a una base por clorofila (por ejemplo, al menos 2 veces la P_{max} de células de tipo silvestre, por ejemplo, entre aproximadamente dos veces y aproximadamente tres veces la P_{max} de células de tipo silvestre), con al menos un 80% de P_{max} de células de tipo silvestre por célula; experimenta la saturación de la fotosíntesis a niveles de irradiación más altos (mayor E_k) que el tipo silvestre; exhibe una aparición retrasada de NPQ en respuesta a una intensidad de luz creciente en comparación con las células de tipo silvestre; y tiene niveles más bajos de NPQ que las células de tipo silvestre en todas las irradiancias mayores de aproximadamente $200 \mu E \cdot m^{-2} \cdot s^{-1}$ o mayores de aproximadamente $100 \mu E \cdot m^{-2} \cdot s^{-1}$, cuando tanto los mutantes LIHLA como las células de tipo silvestre se aclimatan a poca luz. En algunos ejemplos, un mutante LIHLA tiene una mutación en un gen que codifica un regulador de la respuesta de aclimatación a la luz y las características anteriores están en comparación con una célula control que no tiene una mutación en un gen que codifica un regulador de aclimatación de luz.

Adicionalmente, un mutante de LIHLA puede tener una actividad PSII (tasa de transporte de electrones a través de PSII, ETR_{PSII}) sustancialmente equivalente o mayor que la de las células control o de tipo silvestre. Por ejemplo, la ETR_{PSII} de un mutante LIHLA puede estar entre aproximadamente 1 y aproximadamente 4 veces la ETR_{PSII} de una célula silvestre o de control, por ejemplo, entre aproximadamente 1,5 veces y aproximadamente 3,5 veces, o entre aproximadamente 1 y aproximadamente 2 veces la ETR_{PSII} de una célula de tipo silvestre o control, entre aproximadamente 1,5 veces y aproximadamente 3,5 veces la ETR_{PSII} de una célula silvestre o de control, o más de 3 veces la ETR_{PSII} de una célula silvestre o de control, por ejemplo, entre aproximadamente 3 y aproximadamente 4 veces. Alternativamente o además, en diversos ejemplos, un mutante de LIHLA puede tener actividad de PSI (tasa de transporte de electrones a través de PSI, ETR_{PSI}) sustancialmente equivalente a la de una cepa tipo silvestre o de control.

Un cultivo de un mutante LIHLA de algas, como se proporciona en este documento, puede permitir una mayor penetración de luz en el cultivo que un alga silvestre o de control cuando el alga silvestre o de control se cultiva de la misma manera y la penetración de la luz en el cultivo se mide a la misma densidad de cultivo que la densidad celular de la cepa mutante LIHLA. Además, un mutante LIHLA de algas puede alcanzar densidades celulares más altas en un cultivo líquido durante un período de, por ejemplo, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, al menos diez, al menos once, al menos doce, al menos trece, al menos catorce, al menos quince, entre quince y veinte, al menos veinte, entre veinte y treinta, o al menos treinta días para un cultivo de al menos 0,5 litros, por ejemplo, al menos 1 litro, al menos 2 litros, al menos 5 litros, al menos 10 litros, al menos 20 litros, al menos 50 litros o al menos 100 litros.

Además de las propiedades anteriores, un mutante LIHLA de algas puede desregularse transcripcionalmente en global en condiciones de poca luz; es decir, un mutante LIHLA puede desregularse en la expresión de múltiples genes que están regulados diferencialmente en la cepa silvestre durante la aclimatación a poca luz. Por ejemplo, un mutante LIHLA puede exhibir una expresión desregulada de al menos diez, al menos veinte, al menos treinta, al

menos cuarenta, al menos cincuenta, o al menos 100 genes que se expresan diferencialmente en respuesta a una intensidad de luz, por ejemplo, que se expresan diferencialmente en células de tipo silvestre durante la aclimatación a poca luz. Los genes que se desrregulan en un mutante LIHLA de algas con poca luz incluyen, pero no se limitan a, genes que codifican proteínas del complejo de recolección de luz (LHC). Por ejemplo, bajo condiciones de poca luz, con respecto a una célula de tipo silvestre, un mutante LIHLA puede tener una menor expresión de al menos diez, al menos doce, al menos quince, o al menos veinte genes sensibles a la luz, incluyendo al menos cinco, al menos diez, o al menos doce, genes de la proteína LHC, y pueden tener una mayor expresión con respecto a una célula de tipo silvestre de al menos dos, al menos tres, al menos cuatro, al menos cinco, o al menos seis genes que no codifican proteínas complejas de recolección de luz. Un mutante LIHLA puede exhibir desrregulación de al menos veinte, al menos treinta, al menos cuarenta, al menos cincuenta, al menos sesenta, al menos setenta, al menos ochenta, al menos noventa, o al menos 100 genes donde la diferencia en el nivel de transcripción de los genes es al menos un \log_2 de 1. Además, al menos cinco, al menos diez, al menos doce, al menos catorce, al menos dieciséis, al menos dieciocho, al menos veinte, o al menos veintidós, genes que codifican proteínas LHC pueden estar regulados negativamente en el mutante con respecto al tipo silvestre cuando ambos son cultivados con poca luz por al menos un \log_2 de 1. Además, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, o al menos diez genes pueden estar regulados positivamente en un mutante LIHLA con respecto a una célula de tipo silvestre cuando ambos son cultivados con poca luz por al menos un \log_2 de 1.

Un mutante LIHLA de algas puede ser, por ejemplo, una microalga tal como, pero no limitada a, una especie de un género seleccionado del grupo que consiste en Achnanthes, Amphiprora, Amphora, Ankistrodesmus, Asteromonas, Boekelovia, Bolidomonas, Borodinella, Botrydium, Botryococcus, Bracteococcus, Chaetoceros, Carteria, Chlamydomonas, Chlorococcum, Chlorogonium, Chlorella, Chroomonas, Chrysothrix, Cricosphaera, Cricosphaera, Cryptothecodinium, Cryptomonas, Cyclotella, Dunaliella, Ellipsoidon, Emiliana, Eremosphaera, Ernodesmus, Euglena, Eustigmatos, Franceia, Fragilaria, Fragilaropsis, Gloeothamnion, Haematococcus, Halocafeteria, Hantzschia, Heterosigma, Hymenomonas, Isochrysis, Lepocinclis, Micractinium, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Neochloris, Nephrochloris, Nephroselmis, Nitzschia, Ochromonas, Oedogonium, Oocystis, Ostreococcus, Parachlorella, Parietochloris, Pascheria, Pavlova, Pelagomonas, Phœodactylum, Phagus, Picochlorum, Platymonas, Pleurochrysis, Pleurococcus, Prototheca, Pseudochlorella, Pseudoneochloris, Pseudostaurastrum, Pyramimonas, Pyrobotrys, Scenedesmus, Schizochlamydeella, Skeletonema, Spyrogyra, Stichococcus, Tetrachlorella, Tetraselmis, Thalassiosira, Tribonema, Vaucheria, Viridiella, Vischeria y Volvox. En algunos ejemplos, un mutante LIHLA es una especie que no tiene naturalmente clorofila b, tal como una diatomea o eustigmatofita. En algunos ejemplos, un mutante LIHLA es eustigmatofito tal como Ellipsoidon, Eustigmatos, Monodus, Nannochloropsis o Vischeria o una diatomea tal como, pero no limitado a, una especie de Amphora, Chaetoceros, Cyclotella, Fragilaropsis, Navicula, Nitzschia, Pavlova, Phaeodactylum, o Thalassiosira.

Un mutante fotosintético de algas que se desrregula en aclimatación a poca luz (es decir, un mutante LIHLA) puede ser un mutante generado por cualquier método factible, que incluye, pero no se limita a, irradiación UV, irradiación gamma o mutagénesis química. Los métodos para generar mutantes de cepas microbianas son bien conocidos.

Un mutante LIHLA de algas, como se proporciona en este documento, también puede ser un mutante de algas modificado genéticamente en el que uno o más genes, tales como, por ejemplo, el gen LAR1, LAR2 o LAR3 u homólogos del mismo, como se describe en la presente memoria, se han dirigido por recombinación homóloga para la eliminación o el reemplazo génico (por ejemplo, con una forma mutada del gen que puede codificar un polipéptido que tiene actividad reducida con respecto al polipéptido de tipo silvestre). Se incluyen aspectos de la ingeniería de un microorganismo en el que la introducción, adición, integración o incorporación de ciertas moléculas de ácido nucleico o secuencias de polinucleótidos particulares en microorganismos o células huésped afecta a la expresión de un gen en el microorganismo. Por ejemplo, un microorganismo de interés puede diseñarse mediante recombinación homóloga dirigida al sitio para insertar un gen particular de interés con o sin una secuencia de control de la expresión tal como un promotor en un locus genómico particular, o insertar un promotor en un locus genético del microorganismo huésped para afectar a la expresión de un gen particular o conjunto de genes en el locus.

Por ejemplo, el gen inactivado o reemplazo por recombinación homóloga puede ser por transformación de un fragmento de ácido nucleico (p. ej., ADN) que incluye una secuencia homóloga a la región del genoma que se va a alterar, donde la secuencia homóloga se interrumpe por una secuencia extraña, típicamente un gen marcador seleccionable que permite la selección para la construcción integrada. Las secuencias flanqueantes homólogas del genoma en cualquier lado de la secuencia extraña o de la secuencia del gen mutado pueden ser, por ejemplo, al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750 o al menos 2.000 nucleótidos de longitud. Se puede proporcionar un constructo genético inactivado o gen "knock in" en el cual una secuencia extraña está flanqueada por secuencias génicas diana en un vector que puede opcionalmente linealizarse, por ejemplo, fuera de la región que se someterá a recombinación homóloga, o se puede proporcionar como un fragmento lineal que no está en el contexto de un vector, por ejemplo, la construcción knock-out o knock-in puede ser un fragmento aislado o sintetizado, que incluya pero no se limite a un producto de PCR. En algunos casos, se puede usar un sistema marcador dividido para generar inactivados génicos por recombinación homóloga, donde pueden introducirse dos fragmentos de ADN que pueden regenerar un marcador seleccionable e interrumpir el locus genético de interés a través de tres eventos cruzados (Jeong et al. 2007) FEMS Microbiol Lett 273: 157 - 163).

Alternativamente o además, se puede generar un mutante LIHLA expresando un gen que codifica un regulador, en el que el gen ha sido mutado para codificar un regulador mutante negativo dominante. Por ejemplo, un gen regulador puede mutarse en una región que se une a un ácido nucleico o active una o más proteínas en una vía reguladora, de modo que la proteína pueda interactuar con (por ejemplo, unirse a) uno o más componentes de una ruta pero donde la interacción no dé como resultado la activación de proteínas adicionales en la ruta. En tales casos, el gen puede integrarse en el cromosoma del organismo huésped que puede ser o no el locus del gen de tipo silvestre, o puede introducirse en la célula como parte de una molécula de ácido nucleico episomal.

Alternativamente o además, un mutante de LIHLA genéticamente modificado puede modificarse para incluir una construcción para atenuar la expresión génica reduciendo la cantidad, estabilidad o traducibilidad del ARNm de un gen que codifica una proteína que regula la aclimatación a baja intensidad de luz. Por ejemplo, un alga se puede transformar con una construcción de ARN antisentido, RNAi o ribozima que se dirige a un ARNm de un regulador aclimatado a la luz usando métodos conocidos en la técnica. Por ejemplo, una construcción de ARN antisentido que incluye la totalidad o una parte de la región transcrita de un gen puede introducirse en una microalga para disminuir la expresión génica (Shroda et al. (1999) *The Plant Cell* 11: 1165-78; Ngiam et al. (2000) *Appl. Environ. Microbiol.* 66: 775-782; Ohnuma et al. (2009) *Protoplasma* 236: 107 - 112; Lavaud et al. (2012) *PLoS One* 7: e36806). Alternativamente o además, una construcción de ARNi (por ejemplo, una construcción que codifica un ARN de horquilla corta) dirigida a un gen regulador puede introducirse en un alga para reducir la expresión del regulador (véase, por ejemplo, Cerruti et al. (2011) *Eukaryotic Cell* (2011) 10: 1164 - 1172; Shroda et al. (2006) *Curr. Genet.* 49:69- 84). Otras estrategias de ingeniería genética para generar mutantes de LIHLA incluyen TALEN o ingeniería de genoma de nucleasa de dedo de cinc (Perez-Pinera et al. (2012) *Curr. Opin. Chem. Biol.* 16: 268 - 277) o tecnología CRISPR (por ejemplo, DiCarlo et al. (2013) *Nucl Acids Res* 41: doi: 10.1093/nar/gtk135).

Para la expresión antisentido, una secuencia de ácido nucleico del gen regulador de interés está operativamente unida a un promotor de modo que la cadena antisentido del ARN se transcribirá. La secuencia de ácido nucleico puede ser solo una porción del gen regulador de interés, o puede ser el gen completo de interés. La secuencia de nucleótidos puede ser, por ejemplo, de aproximadamente 30 nucleótidos a aproximadamente 3 kilobases o mayor, por ejemplo, de 30-50 nucleótidos de longitud, de 50 a 100 nucleótidos de longitud, de 100 a 500 nucleótidos de longitud, de 500 nucleótidos a 1 kb de longitud, de 1 kb a 2 kb de longitud, o de 2 a 5 kb. Por ejemplo, una secuencia antisentido puede ser de aproximadamente 100 nucleótidos a aproximadamente 1 kb de longitud. La construcción puede transformarse en algas usando cualquier método factible, que incluye cualquiera de los descritos en este documento.

Las construcciones de ARN catalítico (ribozimas) pueden diseñarse para formar pares de bases con un ARNm que codifica un gen tal como se proporciona en este documento para escindir el ARNm diana. En algunos ejemplos, las secuencias de ribozimas pueden integrarse dentro de una construcción de ARN antisentido para mediar en la escisión de la diana. Se pueden considerar varios tipos de ribozimas, su diseño y uso se conocen en la técnica y se describen, por ejemplo, en Haseloff et al. (1988) *Nature* 334: 585-591.

Además de la literatura científica anterior, el uso de construcciones de ARNi se describe en los documentos US2005/0166289 y WO 2013/016267, por ejemplo. Un ARN bicatenario con homología con el gen diana se suministra a la célula o se produce en la célula mediante la expresión de una construcción de ARNi, por ejemplo, una construcción de horquilla corta de ARNi (sh). La construcción puede incluir una secuencia que sea idéntica al gen diana, o al menos un 70%, 80%, 90%, 95% o entre 95% y 100% idéntica a una secuencia del gen diana. La construcción puede tener al menos 20, al menos 30, al menos 40, al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, o al menos 1 kb de secuencia homóloga respecto al gen diana. Los vectores de expresión se pueden diseñar usando promotores seleccionados para la expresión continua o inducible de una construcción de ARNi, tal como una construcción que produce un ARNhc.

Un mutante LIHLA en algunos ejemplos puede generarse a través de la dirección de un gen que codifica un regulador de aclimatación a baja intensidad de luz. Un regulador puede ser cualquier proteína que afecte directa o indirectamente a la aclimatación a la luz y puede ser, como ejemplos no limitantes, un factor de transcripción, un activador transcripcional, una proteína alostérica, una quinasa, una fosfatasa, una acetilasa, una deacetilasa, una metilasa, una desmetilasa, una nucleótido ciclasa o una fosfodiesterasa. Preferiblemente, un regulador afecta directa o indirectamente a la expresión de genes múltiples implicados en la aclimatación a la luz, que incluyen genes de proteínas LHC, así como genes que no codifican proteínas LHC.

Por ejemplo, un mutante LIHLA puede mutarse en un gen que codifica la proteína LAR1 ("Regulador de la Aclimatación de Luz 1") (anteriormente denominado Regulador-59) de *Nannochloropsis gaditana* (SEQ ID NO: 4) o cualquiera de sus ortólogos, tales como la proteína LAR1 de *Nannochloropsis oceanica* ("No-LAR1"; SEQ ID NO: 8), o un homólogo de la proteína LAR1 o No-LAR1 que tiene al menos un 30% de identidad con la SEQ ID NO: 4 o SEQ ID NO: 8 en cualquier especie de alga. Por ejemplo, un mutante de LIHLA puede mutarse en un gen que codifica el polipéptido de SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, o SEQ ID NO: 18, o puede mutarse en un gen natural que codifica un polipéptido que tiene al menos 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos

86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente un 100% de identidad de secuencia con SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, o SEQ ID NO: 18, donde el polipéptido incluye una secuencia de aminoácidos que tiene al menos un 40%, al menos 45%, o al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente un 100% de identidad de secuencia con SEQ ID NO: 9 o SEQ ID NO: 10.

La proteína LAR1 de *Nannochloropsis gaditana* (SEQ ID NO: 4) y la proteína No-LAR1 de *Nannochloropsis oceanica* (SEQ ID NO: 8) incluyen cada una un dominio de dedo de cinc TAZ (dominio pFam PF02135), un dominio comúnmente encontrado en activadores transcripcionales, que comprende los aminoácidos 554-632 de SEQ ID NO: 4 y 631-711 de SEQ ID NO: 8. Las secuencias de aminoácidos que comprenden un dominio de dedo de cinc TAZ más las secuencias conservadas en cualquier lado del dominio de dedo de cinc TAZ (referido en este documento como un dominio de dedo de cinc TAZ extendido, véase el Ejemplo 8), se proporcionan en este documento como SEQ ID NO: 9 para la proteína LAR1 de *N. gaditana* y SEQ ID NO: 10 para la proteína LAR1 de *N. oceanica* (No-LAR1).

Como se demuestra en el Ejemplo 12, la proteína LAR1 de *N. gaditana* (SEQ ID NO: 4) y la proteína No-LAR1 de *N. oceanica* (SEQ ID NO: 8) son homólogos funcionales u ortólogos en diferentes especies. Como se detalla en el Ejemplo 8, las secuencias de aminoácidos de la proteína LAR1 de *N. gaditana* (SEQ ID NO: 4) y la proteína No-LAR1 de *N. oceanica* (SEQ ID NO: 8) son aproximadamente idénticas en un 49,7%, mientras que el dominio de dedo de cinc TAZ extendido de la proteína LAR1 de *N. gaditana* (SEQ ID NO: 9) y el dominio de dedo de cinc TAZ extendido de la proteína No-LAR1 de *N. oceanica* (SEQ ID NO: 10) de estos ortólogos tienen aproximadamente un 81,8% de identidad de secuencia de aminoácidos. Análisis adicionales de genes relacionados en bases de datos públicas y privadas encontraron una agrupación filogenética distinta de genes que codificaban proteínas que tenían dominios de dedo de cinc TAZ con al menos un 40% de identidad con la SEQ ID NO: 9 (Tabla 3). Estos genes codifican supuestos ortólogos de la proteína LAR1.

Por lo tanto, en varios ejemplos, el gen que está mutado en un mutante LIHLA puede ser un gen que codifica un polipéptido que tiene al menos un 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 4 o SEQ ID NO: 8, o al menos un 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 o SEQ ID NO: 18. El polipéptido que codifica el gen regulador puede incluir un dominio de dedo de cinc TAZ y puede reclutar a pFam PF02135, por ejemplo, con una puntuación de bit mayor que el límite de recopilación (19,0) y un valor de E inferior a $1,00E-2$ o inferior a $1,00E-10$. Además, el polipéptido codificado que es al menos un 30% idéntico a SEQ ID NO: 4 o SEQ ID NO: 8, o es al menos un 80% o al menos un 85% idéntico a SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 o SEQ ID NO: 18 puede incluir una secuencia de aminoácidos que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 9 o SEQ ID NO: 10.

En ejemplos particulares, un mutante LIHLA puede mutarse en un gen que codifica un polipéptido que incluye una secuencia de aminoácidos que codifica un dominio de dedo de cinc TAZ que tiene al menos un 40%, al menos 64%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, o al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% con la secuencia de aminoácidos de la SEQ ID NO: 9 o SEQ ID NO: 10, donde el polipéptido incluye un dominio de dedo de cinc TAZ. Por ejemplo, un mutante LIHLA puede mutarse en un gen que codifica un polipéptido que tiene al menos un 50% de identidad con SEQ ID NO: 4 o SEQ ID NO: 8, y el polipéptido puede incluir una secuencia de aminoácidos que codifica un dominio de dedo de cinc, en el que la secuencia de aminoácidos tiene al menos un 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95% de identidad con la secuencia de aminoácidos de SEQ ID NO: 9 o SEQ ID NO: 10.

La descripción también describe mutantes LIHLA que están mutados en genes que comprenden una secuencia de nucleótidos que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos

- 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 3 o SEQ ID NO: 7, en donde el gen codifica un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o, al menos 99% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 9 o SEQ ID NO: 10. Además, el polipéptido codificado por el gen puede reclutar a pfam PF02135. Además, el polipéptido codificado por el gen puede tener al menos un 40%, al menos 45%, al menos 50%, al menos 55%, que tiene al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 4 o SEQ ID NO: 8.
- 15 La descripción describe mutantes de LIHLA que están mutados en genes que comprenden una secuencia de nucleótidos que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, o SEQ ID NO: 40, o una parte de la misma. El gen que está mutado en el mutante LIHLA puede codificar, en un alga de tipo silvestre, un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 9 o SEQ ID NO: 10.
- 25 En ejemplos adicionales, un mutante LIHLA puede mutarse en un gen que codifica la proteína LAR2 ("Regulador de Aclimatación de Luz 2") (anteriormente denominado Regulador-216) de *Nannochloropsis gaditana* (SEQ ID NO: 6) o cualquiera de sus ortólogos, tal como la proteína LAR2 de *Nannochloropsis oceanica* (No-LAR2; SEQ ID NO: 21), o un homólogo de la proteína LAR2 o la proteína No-LAR2 de cualquier especie de alga que tiene al menos un 50% de identidad con la SEQ ID NO: 6 o SEQ ID NO: 21.
- 30 La proteína LAR2 de *Nannochloropsis gaditana* (SEQ ID NO: 6) y la proteína No-LAR2 de *Nannochloropsis oceanica* (SEQ ID NO: 21) incluyen cada una un dominio de unión a ADN tipo myb (dominio pFam PF00249), un dominio comúnmente encontrado en reguladores transcripcionales, que comprenden los aminoácidos 101 - 144 de SEQ ID NO: 6, y los aminoácidos 66 - 109 de SEQ ID NO: 21. Una versión extendida de este dominio en *N. gaditana* (SEQ ID NO: 22) es 81% idéntica al dominio de unión a ADN de tipo myb extendido de *N. oceanica* (SEQ ID NO: 23).
- 35 Como se muestra en el Ejemplo 12, la proteína LAR2 de *N. gaditana* (SEQ ID NO: 6) y la proteína No-LAR2 de *N. oceanica* (SEQ ID NO: 21) son homólogos funcionales en diferentes especies u "ortólogos". Como se detalla en el Ejemplo 8, las secuencias de aminoácidos de la proteína LAR2 de *N. gaditana* (SEQ ID NO: 6) y la proteína No-LAR2 de *N. oceanica* (SEQ ID NO: 21) son aproximadamente 69% idénticas, mientras que los dominios de unión a ADN de tipo myb extendidos de la proteína LAR2 de *N. gaditana* (SEQ ID NO: 22) y la proteína N.-LAR2 de *N. oceanica* (SEQ ID NO: 23) son aproximadamente idénticos en un 81%. Análisis adicionales de genes relacionados tanto en bases de datos públicas como privadas encontraron una agrupación filogenética distinta de genes que codifican proteínas que tienen dominios de unión al ADN de tipo myb con al menos un 80% de identidad con la SEQ ID NO: 22 (Tabla 4). Estos genes codifican supuestos ortólogos de la proteína LAR2.
- 45 Por lo tanto, en varios ejemplos, el gen que está mutado en un mutante LIHLA puede ser un gen que codifica un polipéptido que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 6 o SEQ ID NO: 21, o al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, o SEQ ID NO: 32. El polipéptido codificado por el gen puede incluir un dominio de unión a ADN tipo myb y puede reclutar para pfam PF00249, por ejemplo, con una puntuación de bit más alta que el límite de recopilación (24,4) y un valor de E inferior a 1,00E-2 o inferior a 1,00E-10. Además, el polipéptido codificado que es al menos un 50% idéntico con SEQ ID NO: 6 o SEQ ID NO: 21, o al menos un 80% o al menos 85% idéntico con SEQ ID NO: 6, SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, o SEQ ID NO: 32 puede incluir una secuencia que tiene al menos un 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 22 o SEQ ID NO: 23.

En ejemplos particulares, un mutante LIHLA puede mutarse en un gen que codifica un polipéptido que incluye un dominio de unión al ADN de tipo myb, en el que la secuencia de aminoácidos tiene al menos un 65%, al menos 70%, al menos 75%, al menos 80%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95% de identidad con la secuencia de aminoácidos de SEQ ID NO: 22 o SEQ ID NO: 23. Por ejemplo, un mutante LIHLA puede mutarse en un gen que codifica un polipéptido que tiene al menos un 65% de identidad con SEQ ID NO: 6 o SEQ ID NO: 21, y el polipéptido puede incluir una secuencia de aminoácidos que codifica un dominio de unión a ADN tipo myb, en el que la secuencia de aminoácidos tiene al menos un 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95% de identidad con la secuencia de aminoácidos de SEQ ID NO: 22 o SEQ ID NO: 23.

La descripción describe mutantes de LIHLA que están mutados en genes que comprenden una secuencia de nucleótidos que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, por ejemplo, al menos 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad de secuencia con SEQ ID NO: 64, en donde el gen codifica un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 63 o SEQ ID NO: 66.

Además, la descripción describe mutantes de LIHLA que están mutados en genes que comprenden una secuencia de nucleótidos que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 62, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, o SEQ ID NO: 77, o una parte del mismo. El gen que está mutado en el mutante LIHLA puede codificar, en un alga de tipo silvestre, un polipéptido que incluya una secuencia de aminoácidos que tenga al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 64.

En ejemplos adicionales, un mutante LIHLA puede mutarse en un gen que codifica la proteína LAR3 ("Regulador de Aclimatación de Luz 3") de *Nannochloropsis gaditana* (SEQ ID NO: 63) o cualquiera de sus ortólogos, tal como la proteína LAR3 de *Nannochloropsis oceanica* (No-LAR2; SEQ ID NO: 66), o un homólogo de la proteína LAR3 o la proteína No-LAR2 de cualquier especie de alga que tenga al menos un 50% de identidad con la SEQ ID NO: 63 o SEQ ID NO: 66.

La proteína LAR3 de *Nannochloropsis gaditana* (SEQ ID NO: 63) y la proteína No-LAR3 de *Nannochloropsis oceanica* (SEQ ID NO: 66) incluyen cada una un dominio conservado de aproximadamente 100 aminoácidos (SEQ ID NO: 64), que comprende los aminoácidos 302 - 404 de SEQ ID NO: 63, y los aminoácidos 303 - 405 de SEQ ID NO: 66.

Como se demuestra en el Ejemplo 16, la proteína LAR3 de *N. gaditana* (SEQ ID NO: 63) y la proteína No-LAR3 de *N. oceanica* (SEQ ID NO: 66) son homólogos funcionales en diferentes especies, u "ortólogos". Como se detalla en el Ejemplo 17, las secuencias de aminoácidos de la proteína LAR3 de *N. gaditana* (SEQ ID NO: 63) y la proteína No-LAR3 de *N. oceanica* (SEQ ID NO: 66) son aproximadamente un 56% idénticas, mientras que el dominio conservado de la proteína LAR3 de *N. gaditana* (SEQ ID NO: 64) y la proteína No-LAR2 de *N. oceanica* (aminoácidos de SEQ ID NO: 23) son aproximadamente idénticas en un 96%.

Por lo tanto, en varios ejemplos, el gen que está mutado en un mutante LIHLA puede ser un gen que codifica un polipéptido que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 63 o SEQ ID NO: 66, o al menos un 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente el 100% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78. Además, el polipéptido codificado que es al menos un 50% idéntico con SEQ ID NO: 63 o SEQ ID NO: 66, o al menos 80% o al menos 85% idéntico a SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78 pueden incluir una secuencia que tenga al menos un 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 63 o la SEQ

ID NO: 66.

En ejemplos particulares, un mutante LIHLA puede mutarse en un gen que codifica un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 65%, al menos 70%, al menos 75%, al menos 80%, al menos 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95% de identidad con la SEQ ID NO: 64. Por ejemplo, un mutante de LIHLA puede mutarse en un gen que codifica un polipéptido que tiene al menos un 65% de identidad con la SEQ ID NO: 63 o la SEQ ID NO: 66.

La descripción describe mutantes de LIHLA que están mutados en genes que comprenden secuencias de nucleótidos que tienen al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, por ejemplo, al menos 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad de secuencia con SEQ ID NO: 62 o SEQ ID NO: 65, en el que el gen codifica un polipéptido que tiene una secuencia de aminoácidos que tiene al menos un 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la SEQ ID NO: 64. Además, el polipéptido codificado por el gen puede tener al menos un 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 63 o SEQ ID NO: 66.

La descripción describe mutantes de LIHLA que están mutados en genes que comprenden una secuencia de nucleótidos que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con SEQ ID NO: 62, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, o SEQ ID NO: 77, o una parte del mismo. El gen que está mutado en el mutante LIHLA puede codificar, en un alga de tipo silvestre, un polipéptido que incluya una secuencia de aminoácidos que tiene al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, o al menos 95% de identidad con la SEQ ID NO: 64.

Métodos de aislamiento de mutantes de LIHLA

La descripción describe un método para aislar mutantes de algas que están desregulados en aclimatación a poca luz. Los métodos incluyen: mutagenizar una población de algas; seleccionar la población mutagenizada de algas para una baja fluorescencia de clorofila; seleccionar los mutantes que retengan una baja fluorescencia de clorofila cuando los mutantes se mantengan en condiciones de poca luz; y seleccionar dichos mutantes seleccionados mediante fluorometría para identificar mutantes de algas de fluorescencia de clorofila baja estables a poca luz que tienen coeficientes fotoquímicos (qP) que son más altos que los coeficientes qP de las algas de tipo silvestre. Por ejemplo, los mutantes pueden cribarse para tener coeficientes qP más altos que las algas de tipo silvestre a intensidades de luz mayores que $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayores que $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayores que $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayores que $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o mayores que $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayores que $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por ejemplo, o mayores que $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

La mutagénesis puede ser por cualquier método, por ejemplo, mutagénesis insercional, mutagénesis química o por irradiación con radiación gamma o ultravioleta. Los métodos para generar mutantes de cepas microbianas son bien conocidos. Por ejemplo, irradiación gamma, irradiación UV y tratamiento con cualquiera de un gran número de mutágenos químicos posibles (p. ej., 5-bromo desoxiuridina, etil metano sulfonato (EMS), metil metano sulfonato (MMS), dietilsulfato (DES), nitrosoguanidina (NTG), compuestos de ICR, etc.) o el tratamiento con compuestos como los antibióticos enediyna que causan la rotura cromosómica (p. ej., bleomicina, adriamicina, neocarzinostatina) son métodos que se han empleado para la mutagénesis de algas, hongos y quitridios (véase, por ejemplo, la patente de Estados Unidos 8.232.090, la Solicitud de Patente de Estados Unidos 20120088831, la Solicitud de Patente de Estados Unidos 20100285557, la Solicitud de Patente de Estados Unidos 20120258498). En la técnica se conoce una gran cantidad de mutágenos químicos que incluyen, pero no se limitan a, agentes intercalantes, agentes alquilantes, agentes desaminadores, análogos de base. Los agentes intercalantes incluyen, como ejemplos no limitantes, los derivados de acridina o los derivados de fenantridina tales como el bromuro de etidio (también conocido como bromuro de 2,7-diamino-10-etil-6-fenilfenantridinio o bromuro de 3,8-diamino-5-etil-6-fenilfenantridinio). Los ejemplos no limitantes de agentes alquilantes incluyen derivados de nitrosoguanidina (por ejemplo, N-metil-N'-nitro-nitrosoguanidina), metanosulfonato de etilo (EMS), etil etanosulfonato, dietilsulfato (DES), metanosulfonato de metilo (MMS), ácido nitroso o HNO_2 , y las mostazas nitrogenadas o compuestos ICR. Los ejemplos no limitantes de análogos de bases que pueden usarse como mutágenos incluyen el compuesto 5-bromouracilo (también conocido como desoxinucleósido-5-bromodesoxiuridina), 5-bromo desoxiuridina y 2-aminopurina.

La mutagénesis puede incluir adicional o alternativamente la introducción de moléculas de ácido nucleico exógeno

en la célula microbiana de interés, como se ejemplifica en este documento. Por ejemplo, una molécula de ácido nucleico exógena introducida en la célula puede integrarse en un locus genético mediante integración aleatoria o dirigida, afectando a la expresión de los genes en los que se inserta el ADN extraño o genes que están próximos al ADN extraño en el genoma (p. ej. patente de los Estados Unidos 7.019.122; Patente de los Estados Unidos 8.216.844). Típicamente, la molécula de ácido nucleico introducida incluye un gen marcador seleccionable para la selección de transformantes que han integrado la construcción de la molécula de ácido nucleico exógeno. La molécula de ácido nucleico exógeno puede incluir un elemento transponible o un componente del mismo, tal como, por ejemplo, repeticiones invertidas que pueden reconocerse por una transposasa y/o un gen que codifica una transposasa, o la molécula de ácido nucleico exógeno puede basarse al menos en parte en un virus, tal como un virus integrador.

Para la mutagénesis de inserción aleatoria, una construcción incluye preferiblemente un marcador seleccionable que puede usarse para seleccionar transformantes que tienen una construcción integrada y, opcionalmente, también puede servir como marcador de segregación y marcador molecular para el aislamiento y la identificación de un gen interrumpido por el gen marcador seleccionable integrado. Alternativamente, se puede dirigir un locus genético específico, como se ilustra en el Ejemplo 12 de este documento. El locus genético puede codificar un regulador de aclimatación a la luz, tal como, pero no limitado a, LAR1 o un homólogo del mismo, o LAR2 o un homólogo del mismo. La construcción para la disrupción génica puede incluir, por ejemplo, un gen marcador seleccionable flanqueado por secuencias del locus genético de interés, por ejemplo, al menos una porción del gen que codifica un regulador y, opcionalmente, secuencias genómicas adicionales que rodean al gen. Dichas secuencias flanqueantes pueden comprender, por ejemplo, al menos 50 nucleótidos, al menos 100 nucleótidos, al menos 500 nucleótidos, o al menos 1 kilobase de la secuencia genómica.

Alternativamente o además, una molécula de ácido nucleico que codifica una variante de un regulador de aclimatación a la luz (por ejemplo, LAR1 o un homólogo de LAR1, LAR2 o un homólogo de LAR2, u otro regulador) puede introducirse en una célula de alga para generar un mutante LIHLA. El gen que codifica una variante puede dirigirse al locus del gen correspondiente para efectuar una sustitución génica, o puede transformarse en la célula para integrarse aleatoriamente o dirigirse a otro locus, o puede proporcionarse en un episoma. Como ejemplo no limitante, el gen puede codificar una variante que se trunca, se elimina internamente o incluye uno o más cambios de aminoácidos. En algunos ejemplos, la variante actúa como un negativo dominante para inhibir, en todo o en parte, la vía reguladora en la que participa el regulador de aclimatación a poca luz.

En otros ejemplos adicionales, se puede introducir una molécula de ácido nucleico que codifica un regulador de aclimatación a la poca luz, una construcción de ribozima, ARNi o antisentido para generar un mutante de LIHLA. Por ejemplo, una construcción que incluye secuencias que son complementarias ("antisentido") con respecto a una cadena codificante del gen regulador de la aclimatación a la luz puede incluir secuencias antisentido correspondientes a al menos una porción de un gen de alga nativo que codifica una proteína LAR1 o un homólogo de la misma, una proteína LAR2 o un homólogo de la misma u otro regulador de aclimatación de la luz.

Las algas y las bacterias fotosintéticas se pueden transformar mediante cualquier método adecuado, que incluye, como ejemplos no limitantes, la captación natural de ADN (Chung et al. (1998) *FEMS Microbiol Lett.* 164: 353-361; Frigaard et al. (2004) *Methods Mol. Biol.* 274: 325-40; Zang et al. (2007) *J. Microbiol.* 45: 241-245), conjugación (Wolk et al. (1984) *Proc. Natl. Acad. Sci. USA* 81, 1561-1565), transducción, transformación de perlas de vidrio (Kindle et al. (1989) *J. Cell Biol.* 109: 2589-601; Feng et al. (2009) *Mol. Biol. Rep.* 36: 1433-9; Pat. U.S. N° 5.661.017), transformación de triquitas de carburo de silicio (Dunahay et al. (1997) *Methods Mol. Biol.* (1997) 62: 503-9), biolística (Dawson et al. (1997) *Curr. Microbiol.* 35: 356-62; Hallmann et al. (1997) *Proc. Natl. Acad. USA* 94: 7469 - 7474; Jakobiak et al. (2004) *Protist* 155: 381 - 93; Tan et al. (2005) *J. Microbiol.* 43: 361-365; Steinbrenner et al. (2006) *Appl Environ. Microbiol.* 72: 7477-7484; Kroth (2007) *Methods Mol. Biol.* 390: 257-267; Pat. U.S. N° 5.661.017), electroporación (Kjaerulff et al. (1994) *Photosynth. Res.* 41: 277-283; Iwai et al. (2004) *Plant Cell Physiol.* 45: 171-5; Ravindran et al. (2006) *J. Microbiol. Methods* 66: 174 - 6; Sun et al. (2006) *Gene* 377: 140 - 149; Wang et al. (2007) *Appl. Microbiol. Biotechnol.* 76: 651-657; Chaurasia et al. (2008) *J. Microbiol. Methods* 73: 133 - 141; Ludwig et al. (2008) *Appl. Microbiol. Biotechnol.* 78: 729-35), transformación mediada por láser o incubación con ADN en presencia o después del pretratamiento con cualquiera de los dendrímeros de poli(amidoamina) (Pasupathy et al. (2008) *Biotechnol. J.* 3: 1078 - 82), polietilenglicol (Ohnuma et al. (2008) *Plant Cell Physiol.* 49: 117-120), lípidos catiónicos (Muradawa et al. (2008) *J. Biosci. Bioeng.* 105: 77-80), dextrano, fosfato de calcio o cloruro de calcio (Mendez-Alvarez et al. (1994) *J. Bacteriol.* 176: 7395-7397), opcionalmente después del tratamiento de las células con enzimas degradantes de la pared celular (Perrone et al. (1998) *Mol. Biol. Cell* 9: 3351-3365). La transformación mediada por *Agrobacterium* también se puede realizar en células de algas, por ejemplo, después de eliminar o dañar la pared de las células del alga (por ejemplo, el documento WO 2000/62601, Kumar et al. (2004) *Plant Sci.* 166: 731 - 738). Los métodos biolísticos son particularmente exitosos para la transformación de los cloroplastos de especies de plantas y algas eucarióticas (véase, por ejemplo, Ramesh et al. (2004) *Methods Mol. Biol.* 274: 355-307; Doestch et al. (2001) *Curr. Genet.* 39: 49-60; Pat. U.S. No. 7.294.506; documentos WO 2003/091413; WO 2005/005643; y WO 2007/133558).

Preferiblemente, la población que se ha sometido a un procedimiento de mutagénesis, ya sea por medios físicos o químicos, o por ingeniería genética, se criba mediante clasificación celular activada por fluorescencia (FACS). El procedimiento FACS puede usar una entrada, por la cual se excluyen las células que tienen fluorescencia por

encima de un cierto valor, o el procedimiento FACS puede tomar, por ejemplo, el 0,1, 0,2, 0,3, 0,4, 0,5, 1,0, 1,5, 2,0, 2,5, 3,0, 4,0, 5,0, o el 10% más bajo de células, en términos de intensidad de fluorescencia de clorofila por célula. Alternativamente, la entrada puede ser el nivel de fluorescencia de la clorofila de las células de tipo silvestre cuando se aclimatan a la luz intensa. Una exploración FACS puede ir seguida de un cribado aislado de una sola colonia para un contenido reducido de clorofila, que puede ser mediante inspección visual, por ejemplo, usando luz visible para observar una coloración pálida, o mediante fluorescencia, por ejemplo, con la ayuda de una cámara. Alternativamente o además, el nivel de clorofila de aislados clonales se puede medir y comparar cuantitativamente. Los niveles de clorofila se evalúan (bioquímicamente, visualmente y/o por fluorescencia) después de una aclimatación con poca luz para asegurar un fenotipo de aclimatación desregulada con poca luz.

El coeficiente de inactivación fotoquímica qP se puede medir por fluorometría (por ejemplo, fluorometría Dual PAM). La evaluación de qP se realiza preferiblemente después de que los aislados se aclimaten al crecimiento a poca luz, por ejemplo, a una luz de menos de o aproximadamente $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, menos de o aproximadamente $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, menos de o aproximadamente $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o menos de o aproximadamente $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o menos de o aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, menos de o aproximadamente $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, menos de o aproximadamente $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o menos de o aproximadamente $50 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. Preferiblemente, los aislados se criban para un fenotipo de clorofila bajo estable después de uno a cinco días de aclimatación a poca luz antes de medir qP. Los aislados que tienen un qP mayor que el de las células de tipo silvestre aclimatadas a poca luz, al menos a irradiaciones superiores a aproximadamente $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o superiores a aproximadamente $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, superiores a aproximadamente $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, superiores a aproximadamente $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, y preferiblemente a irradiaciones superiores a aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ se identifican como mutantes de LIHLA putativos.

Los aislados que tienen un fenotipo de clorofila baja estable en condiciones de poca luz, por ejemplo, al menos una reducción del 20% en clorofila, preferiblemente al menos una reducción del 30% en clorofila, y más preferiblemente al menos un 40%, al menos un 45% o al menos una reducción del 50% en la clorofila con respecto a las células de tipo silvestre en condiciones de poca luz, y que presentan un qP mayor a irradiancias mayores de $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, por ejemplo, mayores que $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, se analizan en busca de la tasa fotosintética midiendo la evolución del oxígeno.

La evolución del oxígeno se mide en un rango de irradiancias. La irradiancia a la que se satura la fotosíntesis (E_k) también puede calcularse, y los mutantes pueden seleccionarse para tener un E_k mayor que las células de tipo silvestre aclimatadas a poca luz. Los aislados que muestran una P_{max} (la tasa máxima de fotosíntesis) mayor que la P_{max} de la cepa de tipo silvestre en base a la clorofila, y sustancialmente la misma P_{max} que la cepa de tipo silvestre por célula, se seleccionan como mutantes LIHLA. Un mutante LIHLA puede tener una P_{max} por célula que es al menos un 70% de la P_{max} natural, al menos un 75% de la P_{max} natural, al menos un 80% de la P_{max} natural, al menos un 85% de la P_{max} natural, o al menos un 90% o al menos un 95% de la P_{max} natural, y puede tener una P_{max} por célula que sea más alta que la P_{max} por célula de una cepa de tipo silvestre comparable.

Los mutantes pueden evaluarse adicionalmente para determinar las diferencias en la inactivación no fotoquímica (NPQ) con respecto a las células de tipo silvestre. Preferiblemente, un mutante seleccionado inicia la NPQ a una mayor irradiancia que la cepa de tipo silvestre, y preferiblemente, la NPQ es más baja en todas las irradiancias mayores de aproximadamente $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $450 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que alrededor de $400 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $350 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que aproximadamente $300 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, mayor que $150 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, o mayor que $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ en un mutante seleccionado con respecto a la cepa de tipo silvestre.

En algunos ejemplos, las células mutagenizadas se criban para determinar la disminución en la cantidad de clorofila por célula, un aumento en P_{max} por mg de clorofila sin sustancialmente reducción en P_{max} por célula, mayor qP y E_k que las células de tipo silvestre, junto con una disminución en alfa, la pendiente de la curva de irradiancia fotosintética, y el inicio de NPQ a una irradiancia más alta que en las células de tipo silvestre. Además, las células pueden cribarse por su mayor penetración de la luz en el cultivo.

Además, las tasas de transporte de electrones para el fotosistema II y/o PSI pueden medirse por fluorometría para identificar mutantes que no están sustancialmente alterados en la función del fotosistema.

Los mutantes también pueden analizarse para determinar la expresión génica, por ejemplo, después de la aclimatación a poca luz, y los perfiles de expresión génica pueden compararse con los perfiles de expresión génica de las células silvestres aclimatadas a luz intensa y poca luz para identificar los mutantes con perfiles de expresión en condiciones de poca luz que demuestren una diferencia con respecto al perfil de expresión de las células de tipo silvestre con poca luz, similar a la diferencia en los perfiles de expresión de las células de tipo silvestre aclimatadas a la luz intensa en comparación con las células de tipo silvestre aclimatadas a poca luz.

Moléculas de ácido nucleico

La descripción describe moléculas de ácido nucleico aisladas que codifican reguladores que funcionan en la aclimatación de células fotosintéticas a la intensidad de la luz, que incluyen las secuencias de complemento de tales

5 secuencias de ácido nucleico. Las moléculas de ácido nucleico proporcionadas en este documento pueden usarse, por ejemplo, para generar construcciones de dirección de genes como se describe en este documento, y para construcciones de ARNi, y ribozimas, así como para construcciones de expresión. Las moléculas de ácido nucleico también pueden codificar reguladores variantes que actúan como proteínas negativas dominantes que pueden producirse en una célula de alga para producir un fenotipo mutante LIHLA, y también pueden usarse en estrategias para obtener genes adicionales que codifican polipéptidos que desempeñan un papel en la aclimatación a la luz.

10 En algunos ejemplos, una molécula de ácido nucleico aislada como se proporciona en este documento comprende una secuencia de ácido nucleico que codifica un polipéptido que comprende una secuencia de aminoácidos que codifica un dominio de dedo de cinc TAZ extendido que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, o al menos 80% de identidad con SEQ ID NO: 9 o SEQ ID NO: 10. La molécula de ácido nucleico puede codificar un polipéptido que tiene una mutación, con respecto a un gen de tipo silvestre, por ejemplo, un gen puede codificar un polipéptido que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de cualquiera de SEQ ID NO: 4; SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 o SEQ ID NO: 18, en el que el polipéptido tiene al menos una mutación con respecto a un gen de tipo silvestre. La mutación puede estar opcionalmente en un dominio de dedo de cinc TAZ (por ejemplo, en una secuencia que tenga al menos un 50%, al menos 65%, al menos 70%, al menos 75% o al menos 80% de identidad con la SEQ ID NO: 9 o SEQ ID NO: 10). El polipéptido codificado por el gen puede reclutar a pfam PF02135 con una puntuación de bit al menos tan alta como el punto de corte de reunión para pfam PF02135 (por ejemplo, 19,0) cuando se consulta frente a la base de datos Pfam. Además, el polipéptido codificado por el gen puede tener al menos un 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, que tiene al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 4 o SEQ ID NO: 8. La molécula de ácido nucleico puede codificar un polipéptido que tenga una mutación, con respecto a un gen de tipo silvestre, en un dominio de dedo de cinc TAZ (por ejemplo, en una secuencia que tenga al menos un 50%, al menos 65%, al menos 70%, al menos 75%, o al menos 80% de identidad con la SEQ ID NO: 9 o la SEQ ID NO: 10). Alternativamente o además, la molécula de ácido nucleico puede codificar un polipéptido truncado, desplazado en marco o suprimido internamente.

35 La descripción describe, en diversos ejemplos, moléculas de ácido nucleico que codifican polipéptidos que tienen al menos un 70%, al menos un 75%, al menos un 80% o al menos un 85%, por ejemplo al menos un 86%, al menos un 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, o SEQ ID NO: 18, en donde los polipéptidos incluyen un dominio de dedo de cinc TAZ extendido que tiene una secuencia de aminoácidos con al menos un 40% de identidad con SEQ ID NO: 9 o SEQ ID NO: 10. Los polipéptidos pueden tener, por ejemplo, al menos un 85%, al menos un 90%, o al menos un 95% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, o SEQ ID NO: 18, en la que los polipéptidos incluyen un dominio de dedo de cinc TAZ extendido que tiene una secuencia de aminoácidos con al menos un 40% de identidad con la SEQ ID NO: 9 o SEQ ID NO: 10. Las moléculas de ácido nucleico en varios ejemplos son ADNc, no tienen la secuencia de un gen natural y/o son construcciones para recombinación homóloga o atenuación génica.

50 La descripción describe moléculas de ácido nucleico aisladas que comprenden secuencias de nucleótidos que tienen al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, o SEQ ID NO: 40, así como moléculas de ácido nucleico que comprenden secuencias de nucleótidos complementarias a cualquiera de ellas, donde la secuencia de nucleótidos preferiblemente no es idéntica a la secuencia de nucleótidos del gen de origen natural. También se incluyen moléculas de ácido nucleico que comprenden secuencias de nucleótidos que tienen al menos un 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con al menos una porción de un gen de origen natural, en el que la molécula de ácido nucleico es una construcción para recombinación homóloga o atenuación génica (por ejemplo, una construcción para la expresión de ARNi, antisentido o ribozima), y en la que el gen de origen natural codifica un polipéptido que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, por ejemplo al

menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de cualquiera de SEQ ID NO: 4; SEQ ID NO: 8, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 o SEQ ID NO: 18. El gen de origen natural que está dirigido por la construcción antisentido, ARNi o ribozima puede, en algunos ejemplos, tener al menos un 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39 o SEQ ID NO: 40.

La descripción describe un ácido nucleico proporcionado en este documento que codifica un polipéptido que tiene al menos un 50% de identidad con la SEQ ID NO: 4 o SEQ ID NO: 8, en el que el polipéptido incluye un dominio de dedo de cinc TAZ extendido que tiene al menos un 80% de identidad con SEQ ID NO: 9 o SEQ ID NO: 10. Por ejemplo, un ácido nucleico como se proporciona en este documento puede codificar un polipéptido que tiene al menos un 85% de identidad con SEQ ID NO: 4 o SEQ ID NO: 8, donde el polipéptido incluye un dominio de dedo de cinc TAZ extendido que tiene una secuencia de aminoácidos con al menos un 85% de identidad con SEQ ID NO: 9 o SEQ ID NO: 10.

En otros ejemplos, una molécula de ácido nucleico aislada como se proporciona en este documento comprende una secuencia de ácido nucleico que codifica un polipéptido que comprende una secuencia de aminoácidos que codifica un dominio de unión a ADN de tipo myb extendido que tiene al menos un 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad con SEQ ID NO: 22 o SEQ ID NO: 23. La molécula de ácido nucleico puede codificar un polipéptido que tiene una mutación, con respecto a un gen de tipo silvestre, por ejemplo, un gen que codifica un polipéptido que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de cualquiera de SEQ ID NO: 6, SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, o SEQ ID NO: 32, en donde el polipéptido tiene al menos una mutación con respecto a un gen de tipo silvestre. La mutación puede estar opcionalmente en un dominio de unión al ADN de tipo myb (por ejemplo, en una secuencia que tenga al menos un 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad con SEQ ID NO: 22 o SEQ ID NO: 23). El polipéptido codificado por el gen puede reclutar a pfam PF00249 con una puntuación de bit al menos tan alta como el punto de corte de reunión para pfam PF00249 (p. ej., 24,4) cuando se consulta contra la base de datos Pfam. Además, el polipéptido codificado por el gen puede tener al menos un 40%, al menos 45%, al menos 50%, al menos 55%, que tiene al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97% o, al menos 98%, al menos 99%, o aproximadamente 100% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 6 o SEQ ID NO: 21. La molécula de ácido nucleico puede codificar un polipéptido que tiene una mutación, con respecto a un gen de tipo silvestre, en un dominio de unión a ADN de tipo myb (por ejemplo, en una secuencia que tiene al menos un 50%, al menos 65%, al menos 70%, al menos 75%, o al menos 80% de identidad con SEQ ID NO: 22 o SEQ ID NO: 23). Alternativamente o además, la molécula de ácido nucleico puede codificar un polipéptido truncado, modificado en el marco o eliminado internamente.

La descripción describe, en diversos ejemplos, moléculas de ácido nucleico que codifican polipéptidos que tienen al menos un 70%, al menos un 75%, al menos un 80%, o al menos un 85%, por ejemplo, al menos un 86%, al menos un 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 6, SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31 o SEQ ID NO: 32, en la que los polipéptidos incluyen un dominio de dedo de cinc TAZ extendido que tiene una secuencia de aminoácidos con al menos un 40% de identidad con SEQ ID NO: 22 o SEQ ID NO: 23. Los polipéptidos pueden tener, por ejemplo, al menos un 85%, al menos 90%, o al menos un 95% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 6, SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, o SEQ ID NO: 32, en la que los polipéptidos incluyen un dominio de dedo de cinc TAZ extendido que tiene una secuencia de aminoácidos con al menos un 40% de identidad con SEQ ID NO: 22 o SEQ ID NO: 23. Las moléculas de ácido nucleico en varios ejemplos son ADNc, no tienen la secuencia de un gen natural y/o son construcciones para la recombinación homóloga o atenuación génica.

La descripción describe además moléculas de ácido nucleico aisladas que comprenden secuencias de nucleótidos

que tienen al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5, SEQ ID NO: 20, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46 o SEQ ID NO: 47, así como moléculas de ácido nucleico que comprenden secuencias de nucleótidos complementarias a cualquiera de ellas, donde la secuencia de nucleótidos no es idéntica a la secuencia de nucleótidos del gen natural. También se incluyen moléculas de ácido nucleico que comprenden secuencias de nucleótidos que tienen al menos un 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con al menos una porción de un gen de origen natural, en el que la molécula de ácido nucleico es una construcción para recombinación homóloga o atenuación génica (por ejemplo, una construcción para la expresión de ARNi, antisentido o ribozima), y en la que el gen natural codifica un polipéptido que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de cualquiera de SEQ ID NO: 6; SEQ ID NO: 21, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, o SEQ ID NO: 32. El gen de origen natural que está dirigido por la construcción antisentido, ARNi o ribozima puede tener al menos un 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 5, SEQ ID NO: 20, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46 o SEQ ID NO: 47.

La descripción describe un ácido nucleico proporcionado en este documento que codifica un polipéptido que tiene al menos un 65% de identidad con la SEQ ID NO: 6 o SEQ ID NO: 21, en el que el polipéptido incluye un dominio de unión a ADN de tipo myb extendido que tiene al menos un 85% de identidad con SEQ ID NO: 22 o SEQ ID NO: 23. Por ejemplo, un ácido nucleico, como se proporciona en este documento, puede codificar un polipéptido que tiene al menos un 85% de identidad con SEQ ID NO: 6 o SEQ ID NO: 21, donde el polipéptido incluye un dominio de unión a ADN de tipo myb extendido que tiene una secuencia de aminoácidos con al menos un 95% de identidad con la SEQ ID NO: 22 o la SEQ ID NO: 23.

En otros ejemplos, una molécula de ácido nucleico aislada como se proporciona en este documento comprende una secuencia de ácido nucleico que codifica un polipéptido que comprende un dominio que codifica la secuencia de aminoácidos que tiene al menos un 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad con SEQ ID NO: 64. La molécula de ácido nucleico puede codificar un polipéptido que tiene una mutación, con respecto a un gen de tipo silvestre, por ejemplo, un gen que codifica un polipéptido que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de cualquiera de SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78, en donde el polipéptido tiene al menos una mutación con respecto a un gen de tipo silvestre. La mutación puede estar opcionalmente en un dominio conservado del polipéptido (por ejemplo, en una secuencia que tiene al menos un 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad con SEQ ID NO: 64). Además, el polipéptido codificado por el gen puede tener al menos un 40%, al menos 45%, al menos 50%, al menos 55%, que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o aproximadamente el 100% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 63 o la SEQ ID NO: 66. La molécula de ácido nucleico puede codificar un polipéptido que tiene una mutación con respecto a un gen de tipo silvestre en un dominio conservado del polipéptido (por ejemplo, en una secuencia que tiene al menos un 50%, al menos 65%, al menos 70%, al menos 75%, o al menos un 80% de identidad con SEQ ID NO: 64). Alternativamente o además, la molécula de ácido nucleico puede codificar un polipéptido truncado, modificado en el marco o eliminado internamente.

La descripción describe, en diversos ejemplos, moléculas de ácido nucleico que codifican polipéptidos que tienen al menos un 70%, al menos un 75%, al menos un 80% o al menos un 85%, por ejemplo al menos un 86%, al menos un 87%, al menos un 88%, al menos un 89%, al menos un 90%, al menos un 91%, al menos un 92%, al menos un 93%, al menos un 94%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, o al menos un 99% de identidad de secuencia con la

secuencia de aminoácidos de SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78, en el que los polipéptidos incluyen un dominio que tiene una secuencia de aminoácidos con al menos un 40% de identidad con SEQ ID NO: 64. Los polipéptidos pueden tener, por ejemplo, al menos un 85%, al menos 90%, o al menos 95% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, o SEQ ID NO: 78, en el que los polipéptidos incluyen un dominio conservado que tiene una secuencia de aminoácidos con al menos un 40% de identidad con la SEQ ID NO: 64. Las moléculas de ácido nucleico en varios ejemplos son ADNc, no tienen la secuencia de un gen natural y/o son construcciones para recombinación homóloga o atenuación génica.

La descripción describe moléculas de ácido nucleico aisladas que comprenden secuencias de nucleótidos que tienen al menos un 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de nucleótidos de la SEQ ID NO: 62, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75, o SEQ ID NO: 77, así como moléculas de ácido nucleico que comprenden secuencias de nucleótidos complementarias a cualquiera de las mismas, donde la secuencia de nucleótidos no es idéntica a la secuencia de nucleótidos del gen de origen natural. También se incluyen moléculas de ácido nucleico que comprenden secuencias de nucleótidos que tienen al menos un 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con al menos una porción de un gen de origen natural, en la que la molécula de ácido nucleico es una construcción para recombinación homóloga o atenuación génica (por ejemplo, una construcción para la expresión de ARNi, antisentido o ribozima), y en la que el gen de origen natural codifica un polipéptido que tiene al menos un 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80% o al menos 85%, por ejemplo al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de aminoácidos de cualquiera de SEQ ID NO: 63, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76 o SEQ ID NO: 78. El gen de origen natural que está dirigido por la construcción antisentido, ARNi o ribozima puede tener al menos un 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, o al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 62, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 71, SEQ ID NO: 73, SEQ ID NO: 75 o SEQ ID NO: 77.

La descripción describe un ácido nucleico proporcionado en este documento que codifica un polipéptido que tiene al menos un 65% de identidad con la SEQ ID NO: 63 o SEQ ID NO: 66, en el que el polipéptido incluye un dominio que tiene al menos un 85% de identidad con SEQ ID NO: 64. Por ejemplo, un ácido nucleico como se proporciona en este documento puede codificar un polipéptido que tiene al menos un 85% de identidad con SEQ ID NO: 63 o SEQ ID NO: 66, donde el polipéptido incluye un dominio que tiene una secuencia de aminoácidos con al menos un 95% de identidad con SEQ ID NO: 64.

La descripción describe variaciones de las secuencias de nucleótidos descritas en este documento, tales como las que codifican fragmentos funcionales o variantes de los polipéptidos como se describen en este documento. Tales variantes pueden ser de origen natural, o de origen no natural, tales como las inducidas por diversos mutágenos y procesos mutagénicos. Las variaciones pretendidas incluyen, pero no se limitan a, la adición, delección y sustitución de uno o más nucleótidos que pueden dar como resultado cambios conservadores o no conservadores de aminoácidos, incluyendo adiciones y delecciones. También se contempla la optimización de codones de secuencias de nucleótidos que codifican polipéptidos para la expresión en una célula huésped de interés.

La descripción describe constructos que comprenden una secuencia de ácido nucleico como se proporciona en el presente documento que puede incluir además una o más secuencias que regulan o median la transcripción, traducción o integración de secuencias de nucleótidos en un genoma huésped. Por ejemplo, la invención proporciona construcciones de expresión que comprenden uno o más "elementos de control de la expresión" o secuencias que regulan la transcripción de la expresión de un gen unido operativamente, o la traducción del ARN transcrito. Por ejemplo, un elemento de control de la expresión puede ser un promotor que se puede unir operativamente a un gen de interés o secuencia codificante de shRNA o antisentido en una construcción de expresión o "casete de expresión". Diversos promotores de algas se describen en la publicación de solicitud de patente estadounidense 2013/0023035, en la solicitud de patente estadounidense 13/486.930 presentada el 1 de junio de 2012, en la solicitud US 13/693.585 presentada el 4 de diciembre de 2012 y en la solicitud estadounidense 13/915.522 presentada el 11 de junio de 2013. Un promotor utilizado en una construcción puede, en algunos casos, ser regulable, por ejemplo, inducible.

Un promotor inducible puede responder, por ejemplo, a la intensidad de la luz o la temperatura alta o baja, y/o puede responder a compuestos específicos. El promotor inducible puede ser, por ejemplo, un promotor sensible a la

hormona (por ejemplo, un promotor que responde a ecdisona, tal como se describe en la Patente de Estados Unidos N° 6.379.945), un promotor metalotioniano (por ejemplo, la Patente de Estados Unidos N° 6.410.828), un promotor relacionado con la patogénesis (PR) que puede responder a un producto químico tal como, por ejemplo, ácido salicílico, etileno, tiamina y/o BTH (Patente de los Estados Unidos N° 5.689.044), o similares, o alguna combinación de los mismos. Un promotor inducible también puede responder a la luz o a la oscuridad (Patente de Estados Unidos N° 5.750.385, Patente de Estados Unidos N° 5.639.952, Patente de Estados Unidos N° 8.314.228), metales (Eukaryotic Cell 2: 995 - 1002 (2003)) o temperatura (Patente de Estados Unidos N° 5.447.858; Abe et al. Plant Cell Physiol. 49: 625-632 (2008); Shroda et al. Plant J. 21: 121 - 131 (2000)). Los ejemplos anteriores no son limitantes en cuanto a los tipos de promotores o promotores específicos que pueden usarse. La secuencia promotora puede ser de cualquier organismo, siempre que sea funcional en el organismo huésped. Los promotores inducibles se pueden formar fusionando una o más partes o dominios de un promotor inducible conocido a al menos una porción de un promotor diferente que puede operar en la célula huésped, p. ej. conferir inducibilidad a un promotor que opera en la especie huésped.

Quando la construcción de ácido nucleico no contiene un promotor en enlace operable con la secuencia de ácido nucleico que codifica el gen de interés (por ejemplo, un gen de deshidrogenasa) la secuencia de ácido nucleico se puede transformar en las células de manera que se una operativamente a un promotor endógeno mediante, por ejemplo, recombinación homóloga, integración específica de sitio, y/o integración del vector. En algunos casos, las secuencias huésped genómicas incluidas en una construcción de ácido nucleico para mediar la recombinación homóloga en el genoma del huésped pueden incluir secuencias reguladoras de genes, por ejemplo, una secuencia promotora, que puede regular la expresión de un gen o secuencia antisentido o ARNi de la construcción de ácido nucleico. En dichos ejemplos, el o los transgenes del constructo pueden unirse operativamente a un promotor que sea endógeno al microorganismo huésped. El(los) promotor(es) endógeno(s) puede(n) ser regulable(s), por ejemplo, inducible(s).

Las construcciones para recombinación homóloga en un genoma de algas (por ejemplo, para alteración o reemplazo génico de un gen regulador) pueden incluir una secuencia de nucleótidos de un gen regulador, tal como cualquiera proporcionado en este documento, o secuencias del genoma de algas que están adyacentes al gen regulador en el organismo huésped. Por ejemplo, una construcción para recombinación homóloga puede incluir al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de un gen regulador que incluye un dominio de dedo de cinc TAZ, tal como cualquiera descrito en este documento, y/o ADN genómico adyacente al mismo. Por ejemplo, las secuencias para mediar la recombinación homóloga en una construcción pueden incluir una o más secuencias de nucleótidos de, o adyacentes a, un gen de algas natural que codifica una proteína de dominio de dedo de cinc TAZ, en donde la proteína de dominio de dedo de cinc TAZ comprende una secuencia de aminoácidos que tiene al menos un 40%, por ejemplo, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, o al menos 99% de identidad con SEQ ID NO: 9 o SEQ ID NO: 10. La construcción puede incluir al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de SEQ ID NO: 3 o SEQ ID NO: 7 y/o una región adyacente del genoma de *Nannochloropsis*.

Alternativamente, las secuencias para mediar la recombinación homóloga en una construcción pueden incluir una o más secuencias de nucleótidos de, o adyacentes a, un gen de algas natural que codifica una proteína de dominio de unión a ADN de tipo myb, en el que la proteína de dominio de unión a ADN de tipo myb comprende una secuencia de aminoácidos que tiene al menos un 80%, por ejemplo, al menos 85%, al menos 90%, al menos 95% de identidad, o al menos un 99% de SEQ ID NO: 22 o SEQ ID NO: 23. Por ejemplo, una construcción para recombinación homóloga puede incluir al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de un gen regulador que incluye un dominio de unión al ADN de tipo myb, tal como cualquiera descrito en este documento, y/o ADN genómico adyacente al mismo. Por ejemplo, las secuencias para mediar la recombinación homóloga en una construcción pueden incluir una o más secuencias de nucleótidos de, o adyacentes a, un gen de algas natural que codifica una proteína de dominio de unión a ADN de tipo myb, en el que la proteína de dominio de unión a ADN de tipo myb comprende una secuencia de aminoácidos que tiene al menos un 40%, por ejemplo, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, o al menos 99% de identidad con SEQ ID NO: 22 o SEQ ID NO: 23. La construcción puede incluir al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de SEQ ID NO: 5 o SEQ ID NO: 20 y/o una región adyacente del genoma de *Nannochloropsis*.

Además, alternativamente, las secuencias para mediar en la recombinación homóloga en una construcción pueden incluir una o más secuencias de nucleótidos de, o adyacentes a, un gen de algas o heterokonte natural que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 80%, por ejemplo, al menos 85%, al menos 90%, al menos 95% de identidad, o al menos un 99% de SEQ ID NO: 64. Por ejemplo, una construcción para recombinación homóloga puede incluir al menos 50, al menos 100, al menos 200, al menos 300,

al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de un gen que incluye un dominio conservado, como cualquiera de los descritos en este documento, y/o ADN genómico adyacente al mismo. Por ejemplo, las secuencias para mediar la recombinación homóloga en una construcción pueden incluir una o más secuencias de nucleótidos de, o adyacentes a, un gen de algas o heterokonte de origen natural que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 40%, por ejemplo, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, o al menos un 99% de identidad con SEQ ID NO: 64. La construcción puede incluir al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de SEQ ID NO: 62 o SEQ ID NO: 65 y/o una región adyacente del genoma de *Nannochloropsis*.

También se proporcionan constructos para expresar el ARN (ARNi) antisentido o interferente o ribozimas para generar mutantes de LIHLA. Dichos constructos pueden incluir una o más secuencias que son complementarias, o antisentido, con respecto a las secuencias de ácido nucleico proporcionadas en la presente memoria que codifican polipéptidos reguladores. Por ejemplo, en este documento se proporcionan constructos de moléculas de ácido nucleico para la expresión de ARN antisentido, ARNh_c, microARN o una ribozima que comprende una secuencia de nucleótidos complementaria a al menos una porción de un gen de algas de origen natural que codifica una proteína de dominio de dedo de cinc TAZ, donde la proteína de dominio de dedo de cinc TAZ comprende una secuencia de aminoácidos que tiene al menos un 40%, por ejemplo, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, o al menos 99% de identidad con respecto a SEQ ID NO: 9 o SEQ ID NO: 10. La construcción puede incluir una secuencia complementaria de al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de SEQ ID NO: 3 o SEQ ID NO: 7 y/o una región no codificante de un ARNm que comprende SEQ ID NO: 3 o SEQ ID NO: 7.

Construcciones adicionales para expresar ARN antisentido o interferente (ARNi) o ribozimas para generar mutantes de LIHLA pueden incluir una o más secuencias que son complementarias, o antisentido, con respecto a las secuencias de ácido nucleico proporcionadas en este documento que codifican polipéptidos reguladores. Por ejemplo, en este documento se proporcionan constructos de moléculas de ácido nucleico para la expresión de ARN antisentido, ARNh_c, microARN o una ribozima que comprende una secuencia de nucleótidos complementaria a al menos una porción de un gen de algas de origen natural que codifica una proteína de dominio de unión a ADN de tipo myb, donde la proteína de dominio de unión a ADN de tipo myb comprende una secuencia de aminoácidos que tiene al menos 40%, por ejemplo, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, o al menos 99% de identidad con respecto a SEQ ID NO: 22 o SEQ ID NO: 23. La construcción puede incluir una secuencia complementaria de al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de SEQ ID NO: 5 o SEQ ID NO: 20 y/o una región no codificante de un ARNm que comprende SEQ ID NO: 5 o SEQ ID NO: 20.

Construcciones adicionales para expresar ARN antisentido o interferente (ARNi) o ribozimas para generar mutantes de LIHLA pueden incluir una o más secuencias que son complementarias, o antisentido, con respecto a las secuencias de ácido nucleico proporcionadas en este documento que codifican polipéptidos reguladores. Por ejemplo, en este documento se proporcionan construcciones de moléculas de ácido nucleico para la expresión de ARN antisentido, ARNh_c, microARN o una ribozima que comprende una secuencia de nucleótidos complementaria a al menos una porción de un gen de algas de origen natural que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 40%, por ejemplo, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, o al menos 99% de identidad con respecto a la SEQ ID NO: 64. La construcción puede incluir una secuencia complementaria de al menos 50, al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1.000, al menos 1.200, al menos 1.500, al menos 1.750, o al menos 2.000 nucleótidos de SEQ ID NO: 62 o SEQ ID NO: 65 y/o una región no codificante de un ARNm que comprende SEQ ID NO: 62 o SEQ ID NO: 65.

Métodos de producción de productos de algas

También se proporcionan en la presente memoria métodos para producir productos de algas cultivando algas que tienen una mayor eficacia fotosintética, tales como los mutantes de LIHLA descritos en este documento. Los métodos incluyen cultivar un mutante fotosintético de algas desregulado en aclimatación a poca luz en un medio adecuado para proporcionar un cultivo de algas y recuperar biomasa o al menos un producto del cultivo. El cultivo de algas es preferiblemente un cultivo fotoautótrofo, y el medio de cultivo preferiblemente no incluye una cantidad sustancial de carbono reducido, es decir, el cultivo no incluye carbono reducido en una forma o en un nivel que las algas puedan usar para crecer.

Las algas se pueden cultivar en cualquier recipiente adecuado, incluidos frascos o biorreactores, donde las algas

pueden estar expuestas a luz artificial o natural. El cultivo que comprende algas mutantes que están desreguladas en su respuesta a poca luz puede cultivarse en un ciclo de luz/oscuridad que puede ser, por ejemplo, un ciclo de luz/oscuridad natural o programado, y como ejemplos ilustrativos, puede proporcionar doce horas de luz a doce horas de oscuridad, catorce horas de luz a diez horas de oscuridad, dieciséis horas de luz a ocho horas de oscuridad, etc.

Cultivar se refiere al fomento intencional del crecimiento (por ejemplo, aumentos en el tamaño celular, contenido celular y/o actividad celular) y/o propagación (por ejemplo, aumentos en el número de células a través de la mitosis) de una o más células mediante el uso de condiciones controladas y/o seleccionadas. La combinación de tanto el crecimiento como la propagación puede denominarse proliferación. Como se demuestra en los ejemplos, los mutantes proporcionados en la presente que muestran una adaptación desregulada a baja intensidad de luz pueden alcanzar una mayor densidad celular del cultivo a lo largo del tiempo, por ejemplo, durante un período de una semana o más, con respecto a un cultivo de células de alga silvestre de la misma cepa que no están desreguladas en la aclimatación a poca luz. Por ejemplo, un mutante LIHLA puede cultivarse durante al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, al menos diez, al menos once, al menos doce, al menos trece, al menos catorce o al menos quince días, o al menos uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve o diez semanas o más.

Los ejemplos no limitantes de condiciones seleccionadas y/o controladas que se pueden usar para cultivar el microorganismo recombinante pueden incluir el uso de un medio definido (con características conocidas tales como pH, fuerza iónica y/o fuente de carbono), temperatura especificada, tensión de oxígeno, niveles de dióxido de carbono, crecimiento en un biorreactor o similares, o combinaciones de los mismos. En algunas realizaciones, el microorganismo o la célula hospedadora pueden crecer de manera mixotrófica, usando luz y una fuente reducida de carbono. Alternativamente, el microorganismo o la célula huésped pueden cultivarse fototróficamente. Al crecer fototróficamente, la cepa de algas puede usar de forma ventajosa la luz como fuente de energía. Se puede usar una fuente de carbono inorgánico, tal como CO₂ o bicarbonato, para la síntesis de biomoléculas por el microorganismo. "Carbono inorgánico", como se usa en el presente documento, incluye compuestos o moléculas que contienen carbono que no pueden ser utilizados como fuente de energía sostenible por un organismo. Típicamente, el "carbono inorgánico" puede estar en forma de CO₂ (dióxido de carbono), ácido carbónico, sales de bicarbonato, sales de carbonato, sales de hidrogenocarbonato o similares, o combinaciones de los mismos, que no pueden oxidarse más para obtener energía sostenible ni usarse como una fuente de poder reductor por parte de los organismos. Un microorganismo cultivado fotoautótroficamente puede cultivarse en un medio de cultivo en el cual el carbono inorgánico es sustancialmente la única fuente de carbono. Por ejemplo, en un cultivo en el que el carbono inorgánico es sustancialmente la única fuente de carbono, cualquier molécula de carbono orgánico (reducido) o compuesto de carbono orgánico que pueda proporcionarse en el medio de cultivo no puede ser absorbido y/o metabolizado por la célula para energía y/o no está presente en una cantidad suficiente para proporcionar energía sostenible para el crecimiento y la proliferación del cultivo celular.

Los microorganismos y las células hospedadoras que pueden ser útiles de acuerdo con los métodos de la presente invención se pueden encontrar en diversas ubicaciones y entornos en todo el mundo. El medio de crecimiento particular para la propagación y generación óptimas de lípidos y/u otros productos puede variar y puede optimizarse para promover el crecimiento, la propagación o la producción de un producto tal como un lípido, proteína, pigmento, antioxidante, etc. En algunos casos, ciertas cepas de microorganismos pueden ser incapaces de crecer en un medio de crecimiento particular debido a la presencia de algún componente inhibitorio o la ausencia de algún requerimiento nutricional esencial de la cepa particular del microorganismo o la célula huésped.

Los medios de crecimiento sólidos y líquidos están generalmente disponibles a partir de una amplia variedad de fuentes, como lo son las instrucciones para la preparación de medios particulares adecuados para una amplia variedad de cepas de microorganismos. Por ejemplo, varios medios de agua dulce y agua salada pueden incluir los descritos en Barsanti (2005) *Algae: Anatomy, Biochemistry & Biotechnology*, CRC Press para medios y métodos para cultivar algas. Las recetas de medios algales también se pueden encontrar en los sitios web de varias colecciones de cultivos de algas, que incluyen, como ejemplos no limitantes, la Colección de Cultivo de Algas UTEX (www.sbs.utexas.edu/utex/media.aspx); Colección de Cultivos de Algas y Protozoos (www.ccap.ac.uk); y Katedra Botaniky (botany.natur.cuni.cz/algo/caup-media.html).

Los métodos de cultivo pueden incluir opcionalmente inducir la expresión de uno o más genes para la producción de un producto, tal como, sin limitación, una proteína que participa en la producción de un lípido, una o más proteínas, antioxidantes o pigmentos, y/o que regula una ruta metabólica en el microorganismo. La expresión inductora puede incluir añadir un nutriente o compuesto al cultivo, eliminar uno o más componentes del medio de cultivo, aumentar o disminuir la luz y/o la temperatura, y/u otras manipulaciones que promuevan la expresión del gen de interés. Dichas manipulaciones pueden depender en gran medida de la naturaleza del promotor (heterólogo) operativamente unido al gen de interés.

La descripción describe microorganismos desregulados en aclimatación a baja intensidad de luz que pueden cultivarse en un "fotobiorreactor" equipado con una fuente de luz artificial, y/o tener una o más paredes que son lo suficientemente transparentes para iluminar, incluida la luz solar, para permitir, facilitar y/o mantener el crecimiento y la proliferación de microorganismos aceptables. Para la producción de productos de ácido graso o triglicéridos, los

microorganismos fotosintéticos o las células huésped pueden cultivarse adicional o alternativamente en matraces agitados, tubos de ensayo, viales, placas de microtitulación, placas de Petri o similares, o combinaciones de los mismos.

5 Adicionalmente o alternativamente, los microorganismos fotosintéticos recombinantes o las células hospedadoras pueden cultivarse en estanques, canales, recipientes de crecimiento basados en el mar, zanjas, conductos, conducciones o similares, o combinaciones de los mismos. En tales sistemas, la temperatura puede no estar regulada, o pueden emplearse diversos métodos o dispositivos de calentamiento o enfriamiento. Al igual que con los biorreactores estándar, una fuente de carbono inorgánico (tal como, pero sin limitarse a, CO₂, bicarbonato, sales de carbonato y similares), incluidos, entre otros, aire, aire enriquecido en CO₂, gases de combustión, o similares, o combinaciones de los mismos, se pueden suministrar al cultivo. Cuando se suministran gases de combustión y/u otras fuentes de compuestos inorgánicos que pueden contener CO además de CO₂, puede ser necesario pretratar dichas fuentes de manera que el nivel de CO introducido en el (foto)biorreactor no constituya un peligro y/o dosis letal con respecto al crecimiento, proliferación y/o supervivencia de los microorganismos.

15 Los mutantes LIHLA de algas pueden incluir uno o más genes no nativos que codifican un polipéptido para la producción de un producto, tal como, pero limitado a, un lípido, un colorante o pigmento, un antioxidante, una vitamina, un nucleótido, un ácido nucleico, un aminoácido, una hormona, una citocina, un péptido, una proteína o un polímero. Por ejemplo, el polipéptido codificado puede ser una enzima, regulador metabólico, cofactor, proteína transportadora o transportador.

20 Los métodos incluyen cultivar un mutante LIHLA que incluye al menos un gen no nativo que codifica un polipéptido que participa en la producción de un producto, para producir biomasa o al menos un producto de algas. Productos, tales como lípidos y proteínas, pueden recuperarse del cultivo por medios de recuperación conocidos por los expertos en la técnica, tales como por extracción de cultivo completo, por ejemplo, usando disolventes orgánicos. En algunos casos, la recuperación de los productos de ácidos grasos se puede mejorar mediante la homogeneización de las células. Por ejemplo, lípidos tales como ácidos grasos, derivados de ácidos grasos y/o triglicéridos se pueden aislar de las algas mediante la extracción de las algas con un disolvente a temperatura y/o presión elevadas, como se describe en la solicitud de patente de los Estados Unidos, comúnmente asignada, US 2016/0137951 titulada "Solvent Extraction of Products form Algae", presentada el 29 de febrero de 2012.

25 La biomasa puede cosecharse, por ejemplo, mediante centrifugación o filtración. La biomasa puede secarse y/o congelarse. Se pueden aislar productos adicionales a partir de biomasa, tales como, por ejemplo, lípidos o una o más proteínas.

30 La descripción describe una biomasa de algas que comprende biomasa de un mutante LIHLA de algas, tal como cualquiera descrito en este documento, por ejemplo, un mutante LIHLA de algas que incluye una mutación en un gen que codifica un polipéptido que tiene al menos un 40% de identidad con SEQ ID NO: 4 o SEQ ID NO: 8, un gen que codifica un polipéptido que tiene al menos 50% de identidad con SEQ ID NO: 6 o SEQ ID NO: 21, un gen que codifica un polipéptido que tiene al menos 50% de identidad con SEQ ID NO: 63 o SEQ ID NO: 66. Se incluye además un producto de algas producido por un mutante LIHLA, tal como cualquiera de los descritos en este documento, que incluye un mutante LIHLA de algas que incluye una mutación en un gen que codifica un polipéptido que tiene al menos un 40% de identidad con SEQ ID NO: 4 o SEQ ID NO: 8, un gen que codifica un polipéptido que tiene al menos 50% de identidad con SEQ ID NO: 6 o SEQ ID NO: 21, o un gen que codifica un polipéptido que tiene al menos un 50% de identidad con la SEQ ID NO: 64.

Ejemplos

Ejemplo 1. Determinación de métricas útiles para el estado de aclimatación con iluminación intensa.

45 Una cepa silvestre de *Nannochloropsis gaditana*, WT-3730, que es un aislado subcultivado de la cepa CCMP1894 de *N. gaditana*, obtenida del Centro Nacional Provasoli-Guillard de Algas marinas y Microbiota (NCMA, Maine, EE.UU.), anteriormente Colección de Cultivos del Fitoplancton Marino (CCMP) se usó como el fondo natural para todos los experimentos y la manipulación genética. Las células WT-3730 de *N. gaditana* fueron aclimatadas a intensidades de luz intensa ($500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) y poca luz ($100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) para evaluar las características del estado de aclimatación a la luz intensa que podría ser explotado durante la detección de derivados de LIHLA. Como se muestra en la Figura 1A, se cultivaron dos cultivos de tipo silvestre bajo una alta intensidad de luz durante 4 días para lograr una alta aclimatación a la luz. La fluorescencia de la clorofila se controló como una métrica para el proceso de aclimatación de luz intensa, que se caracteriza por una disminución de la fluorescencia asociada con una disminución en la recolección de luz de las proteínas de unión a la clorofila.

50 Una vez que la fluorescencia de la clorofila había alcanzado un nivel mínimo (aproximadamente 4,6 días), se retuvo un cultivo a la luz intensa mientras que el otro se transfirió a baja intensidad de luz. Los cultivos se dejaron crecer durante varios días después del cambio de luz y se siguió controlando la fluorescencia de la clorofila. Para el cultivo con poca luz cambiada, la fluorescencia de la clorofila aumentó a lo largo del período de crecimiento (representado por los triángulos en la Figura 1A), lo que indica una baja respuesta de la aclimatación a la luz, con un aumento del tamaño de la antena. El cultivo que se mantuvo a la luz intensa mostró un pequeño aumento con el tiempo en la

fluorescencia de la clorofila (representada por los cuadrados en la Figura 1A) que refleja un pequeño aumento en la recolección de luz asociada con un auto-sombreado a medida que las células crecían a densidades mayores. Para el Día 8, la fluorescencia de la clorofila era aproximadamente 3 veces mayor en las células que crecían en condiciones de poca luz en comparación con la fluorescencia de clorofila de las células mantenidas con luz intensa. En este punto, se controlaron varios parámetros de fluorescencia modulada por amplitud de pulso (PAM) para revelar las diferencias fisiológicas asociadas con el estado aclimatado con luz intensa. Se observaron diferencias en varios parámetros de fluorescencia de PAM; sin embargo, el parámetro qP, que refleja la presión de excitación de PSII, demostró ser un indicador particularmente fiable de la aclimatación de luz intensa en relación con el estado de baja iluminación aclimatada. La Figura 1B muestra que las células cambiadas a la luz intensa, representadas por los cuadrados, tienen mayor qP a todas las irradiancias mayores que aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. Por lo tanto, se eligió el parámetro qP como un parámetro de selección secundario inicial para comparar cultivos aclimatados con poca luz de tipo silvestre y supuestas algas mutantes de LIHLA.

Ejemplo 2. Mutagénesis por inserción de *Nannochloropsis gaditana*.

En experimentos separados, varias construcciones que incluían un gen de resistencia a blasticidina o bleomicina, codón optimizado para la expresión de *Nannochloropsis* y clonado corriente abajo de un promotor de *Nannochloropsis* (véase, por ejemplo, la Solicitud de Patente de Estados Unidos N° 13/915.522, presentada el 11 de junio de 2013 y Solicitud de Patente de Estados Unidos N° 13/486.930, presentada el 1 de junio de 2012) fueron transformadas en células de *Nannochloropsis gaditana*. Se proporciona un ejemplo de una de tales construcciones en la Figura 2, y la secuencia de una de tales construcciones se proporciona como SEQ ID NO: 1. La construcción incluye el gen de la resistencia a la blasticidina de *Aspergillus* optimizado para el codón para la expresión de *Nannochloropsis*, unido operativamente a un promotor de TCTP de *Nannochloropsis* (SEQ ID NO: 2). Para la transformación, las células de *Nannochloropsis gaditana* se cultivaron en medio PM064 y se recogieron a una concentración entre $1-3 \times 10^7$ células/ml. Las células se centrifugaron a $2500 \times g$ durante 10 minutos a 25°C para sedimentar las células. Después, las células se resuspendieron en una solución estéril de sorbitol 385 mM y se centrifugaron de nuevo, luego se lavaron dos veces más en sorbitol para eliminar todas las trazas del medio. El sedimento celular se resuspendió en sorbitol hasta una concentración final de 1×10^{10} células/ml. El ADN plasmídico linealizado de la construcción se alícuotó en tubos de micro fuga a una concentración entre 0,5-5 μg de ADN, y se mezclaron 100 ml de la mezcla de células con el ADN. La mezcla se transfirió a cubetas de electroporación enfriada con una distancia de separación de 2 mm. El electroporador se ajustó a 50 μF de capacitancia, 500 ohmios de resistencia y 2,2 kV de voltaje. Después de la electroporación, las muestras se resuspendieron en 1 ml de sorbitol y se incubaron en hielo durante unos minutos. Las células se transfirieron a tubos cónicos de 15 ml que contenían 10 ml de medio reciente y se dejaron recuperar durante la noche a la luz tenue ($\sim 5 \mu\text{mol fotones m}^{-2}\text{sec}^{-1}$). Al día siguiente, las células se sembraron en placas PM024 que contenían zeocina 5 $\mu\text{g/ml}$ o 100 $\mu\text{g/ml}$ de blasticidina a una concentración entre $5-7 \times 10^8$ células/ml. Las placas fueron incubadas bajo luz constante ($\sim 80 \mu\text{mol fotones m}^{-2}\text{sec}^{-1}$) hasta que aparecieron las colonias (alrededor de 2-3 semanas).

En un experimento separado, las células de *Nannochloropsis gaditana* se trataron con UV para inducir mutaciones y se sometieron a la selección de LIHLA. Para la mutagénesis por UV, las células se cultivaron hasta la mitad de la fase logarítmica y luego se diluyeron a 1×10^6 células/ml con medio de crecimiento PM064. Las suspensiones celulares se transfirieron mediante una pipeta a una placa de Petri de 100 mm y se colocaron dentro de un Stratalinker 2400 con la tapa de la placa retirada. La irradiación UV se llevó a cabo con 10.000, 25.000 y 50.000 $\mu\text{J/cm}^2$. Después de la irradiación, las suspensiones celulares se pipetearon en un matraz de agitación envuelto en papel de aluminio para evitar la exposición a la luz durante veinticuatro horas después de la irradiación.

El medio PM024 incluye: 35 ppt de Instant Ocean Sales, 10X de solución de enriquecimiento de agua marina F/2 de Guillard (50X stock de Sigma-Aldrich, St. Louis, MO, cat. No. G0154; concentraciones finales de componentes en los medios: nitrato de sodio 8,825 mM; Fosfato de sodio monobásico 0,32 mM; 0,205 μM de biotina; 0,420 μM de cloruro de cobalto $\cdot 6\text{H}_2\text{O}$; 0,400 μM de sulfato cúprico $\cdot 5\text{H}_2\text{O}$; EDTA disódico 0,11713 mM $\cdot 2\text{H}_2\text{O}$; 9,095 μM de cloruro de manganeso $\cdot 4\text{H}_2\text{O}$; Molibdato de sodio 0,248 $\mu\text{M} \cdot 2\text{H}_2\text{O}$; 2,965 μM de tiamina $\cdot \text{HCl}$; 0,037 μM de Vitamina B₁₂; 0,765 μM de sulfato de cinc $\cdot 7\text{H}_2\text{O}$.)

El medio PM064 incluye: 35 ppm de Instant Ocean Salts, 5X de solución de enriquecimiento de agua marina F/2 de Guillard (50X stock de Sigma-Aldrich, St. Louis, MO, cat. No. G0154; concentraciones finales de componentes en los medios: nitrato de sodio 4,413 mM; Fosfato de sodio monobásico 0,16 mM; 0,103 μM de biotina; 0,240 μM de cloruro de cobalto $\cdot 6\text{H}_2\text{O}$; 0,200 μM de sulfato cúprico $\cdot 5\text{H}_2\text{O}$; EDTA disódico 0,0585 mM $\cdot 2\text{H}_2\text{O}$; 4,54 μM de cloruro de manganeso $\cdot 4\text{H}_2\text{O}$; 0,124 μM de molibdato sódico $\cdot 2\text{H}_2\text{O}$; 1,48 μM de tiamina $\cdot \text{HCl}$; 0,0185 μM de Vitamina B₁₂; 0,382 μM de sulfato de cinc $\cdot 7\text{H}_2\text{O}$).

Ejemplo 3. Selecciones y evaluación fisiológica del mutante LIHLA GE-4574.

Inicialmente, se cribaron bibliotecas mutantes marcadas de inserción al azar de *Nannochloropsis* transformados para una pigmentación reducida usando técnicas de citometría de flujo. Las colonias resistentes a antibióticos que aparecen en las placas de transformación se resuspendieron en cultivo líquido y se aclimataron a bajas intensidades de luz y se clasificaron por citometría de flujo usando un citómetro de flujo BD FACSAria II (BD Biosciences, San Jose, CA) de manera que se seleccionaron células con baja fluorescencia de clorofila. En general, se seleccionó

aproximadamente del 0,5 al 2% de la población total de células con la menor fluorescencia de clorofila de este primer procedimiento de selección. En algunos casos, los transformantes aclimatados con poca luz se seleccionaron mediante FACS, en el que la clasificación se ajustó al nivel de fluorescencia de clorofila de un cultivo aclimatado con luz intensa natural. Se llevó a cabo un cribado primario adicional de colonias putativas de LIHLA aisladas mediante citometría de flujo mediante la selección de fenotipos visuales verdes pálidos o amarillos en placas. Con el fin de explorar las supuestas colonias de LIHLA a partir de otros mutantes de pigmento reducido y falsos positivos, las colonias seleccionadas se sometieron a una selección de cultivo secundario de rendimiento medio para aclimatar los aislados a condiciones de poca luz antes de las mediciones foto-fisiológicas. La fluorescencia de la clorofila se controló durante la aclimatación a poca luz para seleccionar las colonias que conservaban la fluorescencia clorofílica reducida característica del estado aclimatado a la luz intensa. La Figura 3 muestra un ejemplo de un mutante LIHLA y un cultivo de tipo silvestre que se someten a aclimatación de intensidad de luz intensa a poca, donde se monitorizó la fluorescencia de la clorofila diariamente. En estas condiciones, el cultivo de tipo silvestre aumentó su clorofila por célula durante diez días, mientras que en el mutante LIHLA, la clorofila aumentó solo ligeramente. Este rasgo, de retener sustancialmente el mismo bajo contenido de clorofila de una célula aclimatada a la luz intensa incluso cuando estaba aclimatada a condiciones de poca luz, es una característica de identificación de los mutantes de LIHLA. En las mismas condiciones de transferencia de luz intensa a poca luz y posterior aclimatación con poca luz, las células de tipo silvestre característicamente aumentan drásticamente su contenido de clorofila por célula.

Las líneas celulares que retuvieron la fluorescencia de clorofila reducida a una densidad de cultivo más alta (que promueve una respuesta de aclimatación a poca luz en el tipo silvestre) se analizaron después a través de medidas fotofisiológicas más avanzadas después de la aclimatación a poca luz.

Ejemplo 4. Evaluación fisiológica de los clones que tenían menos clorofila en condiciones de poca luz

Los parámetros fotofisiológicos PSII basados en fluorescencia se usaron para identificar cepas con rendimiento cuántico de PSII máximo similar o incrementado (F_v/F_m) y un coeficiente de inactivación fotoquímico (qP) mayor que la cepa de tipo silvestre, determinada en 12 niveles de irradiación. F_v/F_m se midió usando un fluorómetro Dual PAM (Walz, Effeltrich, Alemania). Una alícuota de 3 ml de células con una densidad celular de 1×10^8 células por ml (aproximadamente 5 mg de clorofila por ml) se adaptó a la oscuridad durante cinco minutos, después de lo cual se usó un rayo de medición de baja intensidad para obtener F_0 . Las células se expusieron luego a luz saturante para cerrar todos los centros de reacción, y luego se utilizó un segundo haz de medición de baja intensidad para obtener F_m (Maxwell y Johnson, (2000) J. Exper. Bot. 51: 659-668). La inactivación fotoquímica, o qP, es una medida de la proporción de centros de PSII abiertos, y también se midió usando el fluorómetro Dual PAM (Walz, Effeltrich, Alemania).

El contenido de clorofila de los mutantes se determinó extrayendo células congeladas con 80% de acetona tamponada con Na_2CO_3 , centrifugando y analizando el sobrenadante mediante HPLC (Lafarde et al. (2000) Appl Environ Microbiol 66: 64-72).

La inactivación no fotoquímico, o NPQ, también se midió usando un fluorómetro Dual PAM (Walz, Effeltrich, Alemania) en un rango de intensidades de luz. Los mutantes LIHLA exhibieron el inicio de NPQ a una intensidad de luz más alta que la requerida para el inicio de NPQ en células de tipo silvestre. Además, los mutantes LIHLA exhibieron un NPQ más alto en todas las intensidades de luz más altas que $100 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$.

La evolución del oxígeno se midió usando un electrodo de oxígeno de tipo Clark. Una alícuota de células que contenía $5 \mu\text{g}$ de clorofila por ml, o 10^8 células, se transfirió a la cámara del electrodo de oxígeno que se iluminó con una lámpara a $1500 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$. También se añadió bicarbonato de sodio (5 mM) a la cámara para asegurar que las células no tuvieran un límite de carbono. Las células de algas se expusieron a una intensidad de luz creciente mientras que la concentración de oxígeno se midió continuamente. La concentración de oxígeno se trazó en función de la intensidad de la luz para proporcionar una curva de irradiancia de fotosíntesis (P/I) que demostró la saturación lumínica de la fotosíntesis en las cepas, donde la velocidad de saturación luminosa de la evolución de oxígeno se denomina P_{max} . El valor de P_{max} se calculó sobre una base por miligramo de clorofila y por célula. E_k , la irradiancia de saturación para la fotosíntesis, también se calculó a partir de la evolución del oxígeno vs. la curva de intensidad de la luz (curva P/I) (Talling J. (1957) New Phytologist 56: 29-50. Se observó que los mutantes LIHLA tenían un α menor, la pendiente inicial de la curva P/I.

Ejemplo 5. Caracterización Funcional del Mutante LIHLA GE-4574

Se descubrió que uno de los aislados seleccionados para una caracterización adicional, la cepa GE-4574, tenía una reducción de aproximadamente 45% en la fluorescencia media de clorofila por célula en comparación con el tipo silvestre después de una aclimatación con poca luz (véase, por ejemplo, la Figura 4). Además del contenido reducido de clorofila, esta cepa también mostró una mayor F_v/F_m (Tabla 1) y mayor qP (Figura 5) a todas las intensidades de luz mayores de $200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$ que se probaron. El valor de qP se midió esencialmente de acuerdo con Maxwell y Johnson, (2000) J. Exper. Bot. 51: 659-668.

Tabla 1: Rendimiento cuántico máximo de PSII para la cepa GE-4574 y duplicados biológicos de *Nannochloropsis* de control de tipo silvestre (WT-3730).

Cepa	F_v/F_m
WT-3730	0,74
WT-3730	0,74
GE-4574	0,81

Es digno de mención que el valor de qP del mutante GEH-4574 LIHLA es significativamente mayor que el qP del tipo silvestre, en un rango de niveles de irradiación (Figura 5). Por ejemplo, a una irradiancia de $\sim 650 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$, la cepa GE-4574 tiene un valor de qP 1,7 veces mayor que el tipo silvestre, y a $\sim 1000 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$ e incluso $\sim 2200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$, el qP de la cepa GE-4574 es más del doble que el observado en la cepa de tipo silvestre.

Se implementaron otros filtros foto-fisiológicos para verificar el mantenimiento de una fotosíntesis equilibrada, consistente con un fenotipo LIHLA, en oposición a las deficiencias fotosintéticas que podrían reducir la productividad. Las mediciones de la máxima evolución de oxígeno por célula (P_{max}) y el fotosistema I (PSI) se realizaron en cultivos aclimatados a poca luz. La P_{max} para GE-4574 se redujo en aproximadamente un 10% por célula con respecto al tipo silvestre pero fue 2,5 veces mayor que el valor de tipo silvestre cuando se normalizó a la concentración de clorofila (Tabla 2). Los parámetros PSI se midieron controlando la diferencia entre las señales de transmitancia de 875 nm y 830 nm. Esto proporciona la proporción de centros de PSI oxidados a reducidos, y puede convertirse en una tasa de transporte de electrones (ETR) para PSI. Los parámetros de PSI para GE-4574 no indicaron deficiencias que pudieran perjudicar el crecimiento o la capacidad de realizar la fotosíntesis. Se encontró que las velocidades de transporte de electrones a través de PSI (ETR_{PSI}) y PSII (ETR_{PSII}) en mutantes de LIHLA eran sustancialmente equivalentes o mayores que las ETR_{PSI} y ETR_{PSII} de las células de tipo silvestre.

Tabla 2: Evolución máxima de oxígeno por célula y concentración de clorofila para el mutante LIHLA GE-4574 y duplicados biológicos del control tipo silvestre de Nannochloropsis (WT-3730).

ID de cepa	$P_{\text{max}}/\text{Célula}$ ($\mu\text{mol O}_2 \text{ horas}^{-1} \text{ célula}^{-1}$)	$P_{\text{max}}/[\text{clorofila}]$ ($\mu\text{mol O}_2 \text{ horas}^{-1} \text{ mg}^{-1}$)	Clorofila total (pg/célula)
WT-3730	0,18	151,1	0,12
WT-3730	0,19	161,0	0,12
GE-4574	0,17	404,9	0,04

Se aislaron originalmente un total de 35 mutantes usando el filtro para la clorofila reducida, el aumento de qP sobre irradiancias mayores que $200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$, un aumento en P_{max} por clorofila con P_{max} per células de al menos 70% de células de tipo silvestre, mayor E_k que las células de tipo silvestre, y la aparición de NPQ a una mayor irradiancia de la requerida para las células tipo silvestre, con NPQ reducido sobre irradiancias mayores de $200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$. Muchos de los mutantes que se aislaron tenían las siguientes características en comparación con las células de tipo silvestre bajo condiciones aclimatadas a poca luz: contenido de clorofila reducido de aproximadamente 30% a aproximadamente 75% de los niveles de tipo silvestre (típicamente, aproximadamente de dos a tres veces de disminución en la clorofila por célula); un aumento de aproximadamente dos a tres veces en P_{max} por clorofila, con menos de aproximadamente 25%, y típicamente no más de aproximadamente un 15% de disminución en P_{max} por célula; aproximadamente la misma o mayor velocidad de transporte de electrones fotosintéticos (por ejemplo, ETR_{PSII} típicamente al menos 1,5 veces o al menos 2 veces la velocidad de tipo silvestre); aumento de qP en todas las irradiancias mayores de $200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$; mayor E_k con respecto a las células de tipo silvestre; α disminuída (pendiente inicial de la curva de irradiancia de la fotosíntesis); inicio de NPQ a mayor irradiancia; disminución de NPQ en todas las irradiancias mayores de $200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$, y típicamente a todas las irradiancias mayores de $100 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$. La Figura 6 proporciona una lista de mutantes que tienen estas características junto con su contenido de clorofila, ETR_{PSII} y P_{max} . Los mutantes en la tabla se caracterizan por el locus genético que se encontró que era responsable del fenotipo mutante (LAR1, LAR2 o LAR3), como se describe a continuación en los ejemplos 7 y 10.

La Figura 7 es un gráfico de NPQ frente a la intensidad de la luz para uno de los mutantes de LIHLA (diamantes abiertos), que posteriormente se determinó que era un mutante LAR1, y la cepa progenitora de tipo silvestre (cuadrados negros). El mutante LIHLA demuestra el inicio tardío de NPQ con respecto a la irradiancia, y demuestra una NPQ drásticamente menor que el tipo silvestre en todas las irradiancias mayores de $200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$.

La figura 8a proporciona un gráfico del ETR_{PSII} para un mutante LAR1 LIHLA (triángulos), un mutante LAR2 LIHLA (cuadrados) y la cepa de tipo silvestre (diamantes) sobre todas las irradiancias mayores de $200 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$. La figura 8b muestra que NPQ se reduce drásticamente para los mutantes LIHLA en un rango de irradiancias

mayores de $100 \mu\text{mol fotones m}^{-2} \text{ s}^{-1}$ comparado con el tipo silvestre (mutante LAR1, triángulos, mutante LAR2, cuadrados, tipo silvestre, diamantes). La figura 8c es un gráfico que muestra un qP más alto en respuesta a la irradiancia para irradiancias mayores de $200 \mu\text{mol fotones m}^{-2} \text{ s}^{-1}$ para un mutante LAR1 LIHLA (triángulos), un mutante LAR2 LIHLA (cuadrados) y la cepa tipo silvestre (diamantes). La figura 8d proporciona curvas P/I para mutantes LAR1 aclimatados a la luz intensa (triángulos abiertos) o a poca luz (diamantes oscuros), así como células silvestres aclimatadas a la luz intensa (círculos oscuros) y aclimatadas a poca luz (cuadrados oscuros), con el mutante LIHLA que demuestra mayor E_k y mayor P_{max} por clorofila que las células de tipo silvestre a intensidades tanto altas como bajas de luz.

La Figura 9 y la Figura 10 muestran resultados ejemplares de pruebas de transporte de electrones a través del Fotosistema II (ETR_{PSII}) y el Fotosistema I (ETR_{PSI}) respectivamente en un mutante LIHLA LAR1. El mutante LIHLA muestra una ETR_{PSII} más alta en una amplia gama de irradiancias y no muestra un deterioro discernible de ETR_{PSI} .

Ejemplo 6. Penetración de la luz en estanques de cultivo de mutantes LIHLA

Los cultivos de cepas mutantes también se cribaron para una mejor penetración de la luz en el cultivo con respecto a la penetración de la luz en cultivos de cepas de tipo silvestre. También se evaluó la penetración de la luz en un sistema de crecimiento abierto utilizado para cultivar *Nannochloropsis* de tipo silvestre, así como el mutante GE-4574. Se cultivaron cultivos de *Nannochloropsis gaditana* de tipo silvestre (WT-3730) y mutantes GE-4574 en estanques de invernadero de 400 L en medio de crecimiento repleto de nitrógeno durante siete días en un ciclo de luz de 14 horas/10 horas de oscuridad. Se usaron células de tipo silvestre de *Nannochloropsis* y GE-4574 mutante para inocular estanques a una densidad celular de 2×10^6 células/ml. La penetración de la luz en el estanque se midió usando un sensor cuántico esférico de 4 Pi el día 1 así como el día 7 (proporcionando densidades celulares de 2×10^6 y $4,5 \times 10^7$ células/ml, respectivamente). La Figura 11 muestra que el primer día de cultivo, inmediatamente después de la inoculación, la diferencia en la penetración de la luz en los cultivos era sustancial, experimentando el cultivo del mutante GE-4574 niveles de luz mucho más altos, con respecto al cultivo de tipo silvestre, en todas las profundidades medidas. Por ejemplo, la cantidad de luz a aproximadamente 20 cm desde el fondo del estanque (es decir, aproximadamente 10 cm por debajo de la superficie) del cultivo de tipo silvestre es equivalente a la que se encuentra aproximadamente a 10 cm del fondo (es decir, a aproximadamente 20 cm debajo de la superficie del estanque) del cultivo mutante.

Cuando los cultivos alcanzaron una mayor densidad (día 7), la diferencia en la penetración de la luz en los cultivos de tipo silvestre y mutante fue menos pronunciada, pero aún significativa, a profundidades superiores a 20 centímetros desde el fondo del estanque (es decir, dentro de aproximadamente 10 cm de la superficie del estanque). Por ejemplo, aproximadamente $500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ de luz estaba disponible a 28 cm del fondo del estanque que contenía el cultivo mutante GE-4574, mientras que a la misma profundidad solo alrededor de $250 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ de luz estaba disponible en el estanque que contenía el cultivo de tipo silvestre. Por lo tanto, el mutante LIHLA permite una mayor penetración de la luz en el estanque que la cepa de tipo silvestre cultivada en condiciones idénticas, lo que permite que la fotosíntesis se produzca a mayores distancias de la superficie del estanque y a mayores densidades celulares.

En un experimento separado, se cultivaron cultivos de tipo silvestre y GE4906 (un mutante LAR2) en minipondios de invernadero de 400 L (5 estanques por cepa) con iluminación suplementaria mediante dos lámparas superiores en un ciclo de luz de 14 horas/10 horas de oscuridad. Las algas se cultivaron en medio PM074 (10X F/2) con nitrato adicional 4 mM añadido a los cultivos el día 7. Inicialmente, cada uno de los estanques se diluyó diariamente durante dos días para permitir que la cepa de tipo silvestre se aclimatara en condiciones de iluminación intensa. Un conjunto de cultivos de simulador de lámpara se dejó crecer más allá del inicio de la limitación de la luz; a los estanques restantes se les permitió hacer una asíntota y luego se diluyeron diariamente para mantener las cepas dentro de su rango óptimo de productividad. El perfil de profundidad para los cultivos de tipo silvestre y mutante se registró a una densidad de biomasa permanente de 300 mg/l, que correspondía a una OD_{730} de 0,75.

Los datos que proporcionan la irradiancia (medida usando un sensor cuántico esférico de 4 Pi) a varias profundidades del estanque se proporcionan en la Figura 12 y demuestran que, en cuanto al mutante LIHLA mutado en LAR1 GE4574 (Figura 11), el mutante LIHLA GE4906 mutado en LAR2 permite una mayor penetración de luz en el cultivo, experimentando el cultivo mutante GE4906 niveles de luz mucho más altos, con respecto al cultivo de tipo silvestre, en todas las profundidades medidas.

Ejemplo 7. Genotipado de mutantes LIHLA LAR1 y LAR2

Se aislaron múltiples mutantes que tenían el fenotipo LIHLA de 1) al menos una reducción del 40% en clorofila por célula con respecto al tipo silvestre según se midió por fluorescencia y espectroscopia, 2) mayor qP con respecto al tipo silvestre sobre irradiancias fisiológicamente relevantes (p. ej., mayor que aproximadamente $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ o mayor que aproximadamente $100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$), 3) equivalente (una diferencia del 20% o menos, y preferiblemente una diferencia de aproximadamente 10% o menos) o mayor P_{max} por célula en comparación con el tipo silvestre y al menos un P_{max} dos veces más alto por unidad de clorofila en comparación con una célula de tipo silvestre, 4) mayor E_k con respecto al tipo silvestre, y 5) aparición de NPQ a irradiancias más altas con respecto al tipo silvestre, con un NPQ más bajo a toda irradiancia más alta que alrededor de $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. De todos los mutantes que se

confirmaron mediante análisis fotofisiológicos que tenían los fenotipos LIHLA basados en estos parámetros, todos menos uno se mapearon por secuenciación de ADN para mostrar lesiones en uno de los dos genes. Trece mutaciones distintas se mapearon en un único gen anotado en una secuencia de genoma de *Nannochloropsis gaditana* patentada. Se encontró que el gen identificado, al que se hace referencia en este documento como "Regulador de Aclimatación a la luz 1" o LAR1 (SEQ ID NO: 3) (anteriormente denominado Regulador 59), codifica un polipéptido (SEQ ID NO: 4) que recluta para pfam PF02135 o el pfam "dedo de cinc TAZ", y por lo tanto se determinó que codifica un regulador de la transcripción del dedo de cinc TAZ putativo (Ponting et al. (1996) Trends Biochem. Sci. 21: 11-13). Las mutaciones encontradas en el gen LAR1 incluyeron inserciones marcadas, inserciones sin marcador y deleciones cortas que daban como resultado mutaciones de cambio de marco, así como una eliminación corta que creaba tanto una mutación del sentido erróneo como del punto de deleción. La Figura 13 muestra la posición y el tipo de mutaciones confirmadas para 13 mutantes de LIHLA asociados a LAR1. La confirmación de la posición de la lesión genética fue a través de una combinación de técnicas de caminata cromosómica, PCR y secuenciación de ADN, incluida la secuenciación del genoma utilizando un secuenciador MiSeq interno (Illumina, San Diego, CA).

Para la re-secuenciación del genoma de la cepa LIHLA, se usaron ADN genómicos completos de mutantes de *Nannochloropsis gaditana* para la preparación de la biblioteca de ADN de Nextera de acuerdo con el protocolo recomendado (Illumina Inc., San Diego, CA). Las bibliotecas generadas se secuenciaron por secuenciación de pares en un Instrumento Illumina MiSeq.

Además de las mutaciones encontradas en el gen LAR1, se identificaron cinco mutaciones distintas en un gen denominado "Regulador de Aclimatación a la luz 2" o LAR2 (SEQ ID NO: 5) (anteriormente denominado Regulador 216), entre los mutantes LIHLA aislados. Este locus se identificó en la secuencia patentada del genoma de *Nannochloropsis gaditana* como que codificaba de un factor de transcripción tipo myb. Se descubrió que el polipéptido codificado por este gen (SEQ ID NO: 6) reclutaba a pfam PF00249 ("dominio de unión a ADN de tipo myb"). Las lesiones genéticas LAR2 también se identificaron mediante una combinación de abordajes de caminata cromosómica y secuenciación; las posiciones y la naturaleza de las lesiones genéticas LAR2 se muestran en la Figura 14.

Se analizaron los conjuntos de datos transcriptómicos de *Nannochloropsis* (n = 113 bibliotecas de RNA-seq) para predecir genes con una función probablemente similar a LAR1 mediante el análisis de coexpresión. El análisis de coexpresión se realizó de dos maneras: 1) calculando las medidas de distancia por pares entre LAR1 y todos los otros genes y 2) realizando la inferencia de la red de genes en todos los genes de *Nannochloropsis gaditana* usando el algoritmo GENIE3 (Huynh-Thu et al. 2010) PLOS One 5 (9) e12776.doi: 10). Según todas las métricas, LAR2 fue clasificado como el gen más similar a LAR1, y el inverso también fue cierto. Trazar los valores de expresión de LAR1 y LAR2 para todas las condiciones abióticas (por ejemplo, temperatura, limitación de nutrientes, intensidad de luz, etc.) demuestra la similitud significativa de sus perfiles de expresión lo que sugiere una coregulación potencial de estos factores (Figura 15). Los resultados del análisis de inferencia de red también clasificaron a LAR2 como el gen superior con un patrón de expresión dependiente relativo a LAR1. Esto indica (como lo respalda su fenotipo altamente similar de los mutantes) que las proteínas codificadas por los genes LAR1 y LAR2 están en la misma ruta, y sugiere que pueden interactuar entre sí. Este análisis puede continuarse además para identificar otros miembros potenciales de la red de regulación de fotoaclimatación. Los mutantes LAR, como se proporcionan en este documento, pueden proporcionar un fondo útil para una manipulación genética adicional.

Ejemplo 8. Homólogos LAR1 y LAR2

Existe una similitud significativa entre el gen LAR1 de *N. gaditana* WT-3730 y un gen identificado en *Nannochloropsis oceanica* (cepa ID WT5473) (No-LAR1, SEQ ID NO: 7, que codifica el polipéptido de SEQ ID NO: 8) sobre el dominio del dedo de cinc anotado PF02135 TAZ (supuesto Dedo de Cinc Adaptador de la Transcripción) cerca del extremo C terminal de acuerdo con una alineación a pares de las secuencias de proteína de longitud completa (Figura 16). El dominio del dedo de cinc PFam PF02135 TAZ del polipéptido LAR1 de *N. gaditana* WT-3730, aminoácidos 554-632, se proporciona como SEQ ID NO: 9. El dominio del dedo de cinc PFam PF02135 TAZ del polipéptido WT5473 No-LAR1, aminoácidos 630-706, se proporciona como SEQ ID NO: 10. (La validación funcional del ortólogo LAR1 de *N. oceanica* No-LAR1 se demuestra en el Ejemplo 13, a continuación). Para descubrir potenciales ortólogos adicionales de LAR1, se realizaron búsquedas en 66 ensamblajes eucarióticos internos y disponibles públicamente (principalmente algas y plantas) para las secuencias de proteínas anotadas con el dominio PF02135 (según la puntuación de hmmscan en I-Evalue <0,01). Se extrajeron trescientas sesenta y cuatro secuencias de proteínas que contenían el dominio TAZ, y se alinearon usando Muscle (Edgar (2004) Nucl. Acids Res. 32: 1792-1797). Se construyó un árbol filogenético de máxima probabilidad (con 100 réplicas de arranque) usando Mega (Kumar et al. (2008) Briefings in Bioinformatics 9: 299-306). Una rama particular que contenía la proteína WT-3730 LAR1 agrupó el mismo conjunto de proteínas 87/100 veces y, por lo tanto, representa una agrupación probable de ortólogos LAR1 en otros estramenopilas y diatomeas (Figura 17). Cada proteína parece ser de una biblioteca de secuencias única (que corresponde a un conjunto único de ADNc o ADNg, cepa o especie) excepto en el caso de dos proteínas de *Ectocarpus silicosus* (conjunto del genoma de Ghent).

Utilizando la secuencia de dominio LAR1 TAZ (SEQ ID NO: 9) como una consulta a las secuencias de longitud completa de proteínas con dominios TAZ pronosticados, los hits superiores corresponden al subárbol definido del

árbol filogenético del dedo de cinc TAZ. Parece haber una diferencia sustancial en el valor E y el puntaje de bits para las coincidencias de similitud blastp fuera de la rama putativa de ortólogos TAZ. A pesar de la identidad de secuencia tan baja como 42% sobre el dominio TAZ, el análisis sugiere que las proteínas destacadas en la tabla son ortólogos probables en otras especies (Tabla 3).

Tabla 3. Homólogos de LAR1

Biblioteca_ID	% Identidad de dominio pfam	% ID de proteína entera	Longitud alineamiento	desajustes	aperturas de hueco	inicio q	final q.	inicios	finals	Valor E	puntuación bits
SEQ ID NO: 4 WT-3730 <i>Nannochloropsis gaditana</i>	100,0		79	0	0	1	79	554	632	1,00E-46	175
SEQ ID NO: 8 WT5473 <i>Nannochloropsis oceanica</i>	81,8	49,7	77	14	0	1	77	630	706	1,00E-37	145
SEQ ID NO: 11 CCMP1779 <i>Nannochloropsis oceanica</i>	81,8	49,5	77	14	0	1	77	555	631	2,00E-37	145
SEQ ID NO: 12 <i>Ectocarpus silicosus</i>	62,0		79	30	0	1	79	581	659	8,00E-26	106
SEQ ID NO: 13 <i>Thalassiosira pseudonana</i>	56,1		82	33	1	1	79	942	1023	1,00E-23	99,4
SEQ ID NO: 14 <i>Phaeodactylum tricornutum</i>	53,2		79	37	0	1	79	1011	1089	1,00E-20	89,4
SEQ ID NO: 15 WT0293 <i>Cyclotella</i> sp.	55,0		80	33	1	1	77	744	823	2,00E-20	88,6
SEQ ID NO: 16 <i>Ectocarpus silicosus</i>	55,4		74	33	0	1	74	444	517	7,00E-20	86,7
SEQ ID NO: 17 WT 0229 <i>Navicula</i> species	51,9		79	38	0	1	79	521	599	2,00E-19	85,5
SEQ ID NO: 18 <i>Aureococcus anopagefferens</i>	42,1		76	41	1	2	77	137	209	1,00E-14	69,3

- También hay una considerable identidad/similitud de secuencia entre los ortólogos *N. gaditana* (WT-3730) y *N. oceanica* (WT-5473) en el extremo N de LAR1, incluyendo un tramo perfectamente conservado de treinta aminoácidos: AKQQQLLKDSLTLADLKLHHEFERFQQATA (SEQ ID NO: 19). Utilizando la secuencia N-terminal (primeros 300 aminoácidos) de LAR1 como consulta en una búsqueda DELTA-Blast (Tiempo de Búsqueda Mejorado de Dominio Acelerado BLAST) sobre la base de datos de proteínas NCBI no redundante, se identificó una coincidencia significativa con una proteína XP_001952448.1, anotada como "PREDECIDA: nucleosoma-subunidad del factor de remodelación NURF301-tipo isoforma 1 [*Acyrtosiphon pisum*]". La región de similitud de secuencia en XP 001952448.1 es adyacente al dominio de PHD anotado. Se cree que NURF301 media en el reconocimiento de la histina H3 lisina a través del dominio PHD.
- Un enfoque similar se llevó a cabo para identificar supuestos ortólogos de LAR2 en otras especies. El alineamiento de LAR2 de *N. gaditana* (WT-3730) con un ortólogo identificado de *N. oceanica* (WT-5473) (No-LAR2, secuencia de gen proporcionada como SEQ ID NO: 20; polipéptido codificado proporcionado como SEQ ID NO: 21) muestra una identidad completa sobre el dominio de unión a ADN de tipo Myb PF00249 predicho (Figura 18). Una versión extendida del dominio pfam PF00249 tipo myb del polipéptido LAR2 de *N. gaditana* WT-3730, aminoácidos 71-351, se proporciona como SEQ ID NO: 22. Se proporciona una versión extendida del dominio de pfam PF00249 tipo myb del polipéptido No-LAR2 de *N. oceanica* WT-5473, aminoácidos 36-325, como SEQ ID NO: 23. (La validación funcional del ortólogo LAR2 de *N. oceanica* se muestra en los Ejemplos 12 y 13). Una característica notable de este tipo particular de proteínas de unión a ADN tipo Myb es un motivo de aminoácido "SHAQKY" (SEQ ID NO: 24) al final del dominio similar a Myb. Para LAR2, el motivo es ligeramente diferente: "THAQKY" (SEQ ID NO: 25). Por lo tanto, se realizaron búsquedas en 66 ensamblajes eucarióticos internos y públicos (principalmente algas y plantas) de secuencias de proteínas anotadas con el dominio PF00249 (anotado por hmmscan en I-Evalue <0,01) que contiene un motivo de seis aminoácidos THAQKY (SEQ ID NO: 25) o SHAQKY (SEQ ID NO: 24). Se identificaron un total de 1.409 proteínas con el dominio PF00249 y el motivo de seis aminoácidos.
- Se utilizó tanto el dominio anotado PFAM como una versión extendida del dominio PFAM como consultas para la comparación blastp contra el conjunto de 1.409 proteínas de unión al ADN tipo Myb. Se utilizó una versión extendida del dominio PF00249 porque el modelo PF00249 es relativamente corto (48 aminoácidos) y parece haber una similitud de secuencia extendida entre los ortólogos WT-3730 y WT-5473 en una región más grande. Sin embargo, usando ambas versiones de secuencia como consulta, se obtuvieron los mismos resultados principales que se muestran en la Tabla 4.
- Los homólogos putativos en otras especies (diatomeas y estramenopiles) se identifican en un nivel superior de identidad de secuencia (> 80%) sobre el dominio de tipo Myb en comparación con las homologías observadas para el dominio TAZ. Los emparejamientos en otras cepas y especies con un puntaje de similitud blastp mayor por valor E al dominio similar a Myb LAR2 en comparación con otras proteínas con dominios similares dentro de WT-3730 se consideran supuestos ortólogos como se destaca en la Tabla 4.

35

Tabla 3. Homólogos de LAR1

Objeto	% ID del dominio pfam PF00249	% ID de proteína entera	Longitud de alineamiento	desajustes	aperturas de hueco	inicio q	final q.	inicios	finals	Valor E	puntuación bits
SEQ ID NO: 6 WT-3730 Nannochloropsis gaditana	100,0		44	0	0	1	44	101	144	1,00E-24	101
SEQ ID NO: 21 WT5473 Nannochloropsis oceanica	100,0	69	44	0	0	1	44	66	109	1,00E-23	99,4
SEQ ID NO: 26 CCMP1779 Nannochloropsis oceanica	100,0		44	0	0	1	44	66	109	1,00E-23	99,4
SEQ ID NO: 27 Ectocarpus silicosus	86,4		44	6	0	1	44	7	50	3,00E-20	88,2
SEQ ID NO: 28 WT0229 Cyclotella sp.	86,4		44	6	0	1	44	72	115	4,00E-20	87,8
SEQ ID NO: 30 Phaeodactylum tricornutum	86,4		44	6	0	1	44	118	161	4,00E-20	87,8
SEQ ID NO: 31 Thalassiosira pseudonana	86,4		44	6	0	1	44	419	462	6,00E-20	87,4
SEQ ID NO: 32 Ectocarpus silicosus	81,8		44	8	0	1	44	32	75	1,00E-19	85,9

Los datos transcriptómicos que demuestran una fuerte coexpresión de los genes LAR1 y LAR2 en WT-3730 (Figura 15) sugieren una probable interacción o función relacionada de las proteínas codificadas. Entre las especies perfiladas, se predicen probables ortólogos para LAR1 (dominio de Taz) y LAR2 (dominio tipo Myb) en *Ectocarpus silicosus*, *Navicula* (cepa WT-0229), *Thalassiosira pseudonana* y *Phaeodactylum tricornutum*. Bajo la hipótesis de que las proteínas LAR1 y LAR2 interactúan, el análisis filogenético TAZ sugiere que podría haber ortólogos para la proteína LAR2 tipo Myb también en *Cyclotella* (cepa WT-0293) y *Aureococcus anophagefferens*.

Ejemplo 9. Transcriptómica de tipo silvestre y mutante LIHLA de *Nannochloropsis* aclimatado a luz intensa y poca luz

Con el fin de evaluar la respuesta transcripcional de todo el genoma de *Nannochloropsis gaditana* de tipo silvestre durante la aclimatación a poca y luz intensa, se llevaron a cabo experimentos de desplazamiento de luz transitoria. Las células se aclimataron por separado a la irradiación baja y alta cultivando a la intensidad de luz indicada durante 5 días con dilución de cultivo el día 3 para garantizar una exposición a la luz constante de las células, con todas las demás variables (por ejemplo, temperatura) estrechamente controladas y luego se transfirieron matraces experimentales de alta a baja (HL) y baja a alta (LH) irradiancia, mientras que los matraces de control se mantuvieron bajo las condiciones de irradiación previamente aclimatadas (poca o luz intensa).

Se probó previamente un rango de altas intensidades de luz para determinar el nivel apropiado de alta irradiación de luz para obtener un estado aclimatado de luz intensa sostenido en WT-3730 dentro del rango de densidad celular de 96 horas de crecimiento logístico y para probar la capacidad de esta cepa para adaptarse a una alta irradiancia. Se seleccionó una intensidad de luz de $500 \mu\text{mol fotones} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ PAR porque 1) esta era la irradiación máxima más alta a la que las células de tipo silvestre podían cultivarse sin aglutinación inducida por estrés a la densidad de células de partida deseada de 2×10^6 células/ml, manteniendo aún la aclimatación a la luz intensa a la densidad celular final después de 96 h de crecimiento logístico; y 2) las tasas más altas de evolución de oxígeno máximo por unidad de clorofila (P_{max}), y 3) la mayor diferencia en la cantidad de clorofila por célula (Chl/célula) (de los $50 \mu\text{mol fotones} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ control PAR de poca luz) se determinaron a esta intensidad. Tanto P_{max} como Chl/célula son indicadores ampliamente aceptados de aclimatación fotosintética a los cambios en la intensidad de la luz. En las células de tipo silvestre, se indujo un aumento de aproximadamente 2 veces en P_{max} , mientras que la cantidad de clorofila por célula (Chl/célula) disminuyó 2-3 veces durante en el transcurso de 48 horas después de pasar de poca a la luz intensa. Estos cambios se reprodujeron consistentemente cuando las células de *Nannochloropsis* se cambiaron de poca a la luz intensa.

Los experimentos de cambio de poca a la luz intensa y de luz intensa a poca de escala transcriptómica de WT-3730 se repitieron 3 veces para generar triplicados biológicos para cuatro puntos de tiempo durante la aclimatación de luz intensa a poca luz. Los cultivos se aclimataron durante 48 horas antes del cambio de luz. Los cultivos se hicieron crecer en volúmenes de 100 ml comenzando a aproximadamente 2×10^6 células/ml y se hicieron crecer a aproximadamente 1×10^7 células/ml en el momento del cambio. Las células se cultivaron axénicamente en matraces de cultivo de tejidos de 100 cm de Corning de perfil bajo (Parte nº 3816), se sellaron con tapones de goma previamente tratados en autoclave penetrados por tubos de PTFE rojo 1/16 "ID x 1/8" OD (Cole-Parmer parte Nº. EW-96130 -02) y se mezclaron mediante burbujeo con mezclas de CO_2 :aire al 1% filtradas con $0,2 \mu\text{M}$ a una velocidad de 15 mL/min (+/- 3 mL/min). Para cada experimento, se sedimentaron 40 ml de muestras de cultivo y se congelaron inmediatamente en nitrógeno líquido en 4 puntos de tiempo (0h (T_0), 4 h, 24 h y 48 h). La reproducibilidad de una respuesta deseada a las condiciones de luz intensa y poca luz se validó de nuevo en este experimento: se mejoró la evolución de O_2 en los matraces adaptados a la luz intensa y se observó una disminución de 2-3 veces en Chl/célula a las 24 y 48 horas después del cambio de luz intensa. Estos cambios fueron completamente reversibles durante el cambio de luz intensa a poca.

El contenido de clorofila, la tasa fotosintética (P_{max}) y los parámetros de fluorescencia de la clorofila Dual PAM (por ejemplo, qP) también se controlaron para mostrar que tuvo lugar una aclimatación fisiológicamente exitosa. El ARN se extrajo de muestras de sacrificio extraídas en diversos momentos durante el cambio de luz y se sometió a secuenciación profunda de todo el genoma utilizando HiSeq.

El ARN se extrajo de muestras adaptadas de poca a la luz intensa recogidas a las 0, 4, 24 y 48 h después del cambio de luz de todos los experimentos. La calidad final del ARN se determinó mediante análisis Agilent Bioanalyzer 2100. Todas las muestras tenían números de integridad de ARN mayores que 7, con la mayoría entre 8 y 9. Al menos $10 \mu\text{g}$ de ARN de cada muestra se envió a Ambry Genetics (San Diego, CA) para la secuenciación del transcrito. Además de la secuenciación del ARN poliA, 16 de las 24 muestras también fueron tratadas con agotamiento de ARNr RiboZero™ (Kit de hoja de planta) del ARNr para la secuenciación total del ARN. La secuenciación de ARN total tratada con RiboZero permitió la cuantificación de transcritos codificados por cloroplastos no capturados por la secuenciación de poliA del ARN. El análisis de Ribo-Zero tratado frente a muestras purificadas de ARN poliA reveló patrones similares de transcripciones de genes codificados en el núcleo, aunque las muestras tratadas con RiboZero permitieron un análisis adicional de la transcripción de genes mitocondriales y cloroplastos.

Además, el mutante LIHLA GE-4574 se aclimató a poca luz usando las mismas condiciones que se usaron para la aclimatación de poca luz del tipo salvaje. El ARN se extrajo tanto de células de tipo silvestre (WT-3730) como de

células mutantes de LIHLA GE-4574 en diversos momentos durante la aclimatación de luz y se sometió a un perfil transcriptómico.

Las muestras de ARN se agotaron de ARNr por dos métodos independientes. Las muestras se dividieron en dos partes alícuotas y se purificaron con polyA o se trataron usando el kit magnético RiboZero™ (Hoja de Planta) después de lo cual se fragmentaron y secuenciaron por Ambry Genetics (Aliso Viejo, CA). El ARNm se secuenció usando la secuenciación por síntesis (Illumina HiSeq) para generar lecturas de extremo de 100 pares usando el procedimiento de ARNm Seq (descrito en Mortazavi et al. (2008) Nature Methods 5: 621-628. Las lecturas mapeables se alinearon con la secuencia del genoma de referencia de *N. gaditana* usando TopHat (tophat.cbcb.umd.edu/). Los niveles de expresión se calcularon para cada gen anotado normalizado para la longitud del gen y el número total de lecturas mapeables por muestra utilizando el componente Cuffdiff del software Cufflinks (cufflinks.cbcb.umd.edu). Los niveles de expresión en unidades de fragmentos por kilobase por millón (FPKM) se mostraron para cada gen en cada muestra usando parámetros estándar. FPKM es una medida de los niveles transcripcionales relativos que se normaliza para las diferencias en la longitud del transcrito.

El análisis global de las transcripciones con diferencias significativas (FDR menor o igual a 0,05) en sus niveles de expresión entre el mutante LIHLA GE-4574 y la cepa progenitora de tipo silvestre WT-3730 bajo las mismas condiciones de poca luz, mostró el patrón de expresión diferencial de estos genes. El paquete de software edgeR se usó para probar genes para la expresión diferencial entre las dos cepas, véase Robinson et al. (2009) Bioinformatics 26: 139-140. De los 340 genes expresados que se analizaron, la abundancia de transcripciones del mutante LIHLA GE-4574 cultivado con poca luz en relación con su abundancia en células de tipo silvestre bajo condiciones de poca luz, parecían increíblemente similares al WT durante el cambio de poca a la luz intensa (LH) y opuesto al WT durante el cambio de luz intensa a poca (HL). Es decir, en base al análisis transcriptómico, la mayoría de los genes regulados por la luz que se encontraron que tenían niveles de transcripción alterados en cultivos de tipo silvestre aclimatados a poca luz con respecto a sus niveles de transcripción con luz intensa, se encontraron desregulados en el mutante LIHLA; es decir, no tenían, en células mutantes aclimatadas a poca luz, la abundancia de transcripción observada en células de tipo silvestre aclimatadas a poca luz. Esta desregulación a gran escala de los genes que son regulados diferencialmente por la luz, en un mutante con un perfil transcripcional de luz intensa bloqueado, sugiere enérgicamente que la proteína LAR1 es un regulador global positivo de la respuesta de aclimatación a poca luz en *Nannochloropsis*. El perfil transcripcional de los genes anotados como polipéptidos de recolección de luz bajo condiciones de cambio de luz fue particularmente llamativo, con casi la totalidad de los transcritos de LHCP reducidos en abundancia en el mutante bajo crecimiento a poca luz con respecto a los niveles de tipo silvestre y reducidos de forma similar en el tipo silvestre bajo crecimiento a la luz intensa pero presente en mayor abundancia en el tipo silvestre bajo crecimiento a poca luz.

Ejemplo 10. qRT-PCR de genes regulados durante la aclimatación a poca luz

Uno de los mutantes LIHLA aislados (GE5440) portaba una única inserción en el gen LAR1 ("Regulador de la Aclimatación a la luz - dominio del dedo de cinc") (SEQ ID NO: 3) y tenía solo una mutación puntual conservativa adicional en una región intergénica de otra región del genoma (según lo determinado por la secuenciación del genoma (MiSeq)). Con el fin de averiguar si este mutante también se desregulaba en la respuesta transcriptómica a poca luz, se realizó RT-PCR en una selección de transcripciones génicas (proporcionadas en la Tabla 5) que eran representativas de la respuesta transcripcional a la luz en el tipo silvestre. Un mutante LAR1 adicional con buenas propiedades de crecimiento, GE5409, y un mutante LAR2, GE4906, también se incluyeron en este análisis.

Tabla 5. Genes regulados diferencialmente en células mutantes LIHLA en relación con el tipo silvestre

Gen ID	Gen	Abundancia relativa (arriba/abajo)	Cambio veces Log2 en abundancia
7003	Proteína desconocida	↑	3,1
2936	Proteína pronosticada	↑	1,9
4948	Enzima metabólica	↑	1,7
5289	Proteína con dominio de unión a peptidoglicano	↑	1,3
4070	Subunidad de enzima	↑	1,1
2286	Proteína relacionada con la fotosíntesis	↓	-1,4
1517	Proteína relacionada con la fotosíntesis	↓	-1,5
1584	Proteína relacionada con la fotosíntesis	↓	-1,6

2281	Proteína pronosticada	↓	-1,7
1091	Proteína relacionada con la fotosíntesis	↓	-1,8
9309	Proteína del complejo de cosecha de luz 3	↓	-2,1
8780	Proteína del complejo de cosecha de luz 2	↓	-2,4
2073	Proteína del complejo de cosecha de luz (VCP1)	↓	-2,5
0480	Proteína del complejo de cosecha de luz 3	↓	-2,6
8185	Proteína del complejo de cosecha de luz 3	↓	-2,6
6440	Proteína del complejo de cosecha de luz 3	↓	-2,7
0651	Proteína inducible por luz temprana HV60	↓	-2,7
0881	Proteína del complejo de cosecha de luz 2	↓	-2,8
3534	Proteína del complejo de cosecha de luz 2	↓	-2,9
7454	Proteína del complejo de cosecha de luz (VCP2)	↓	-3,1
2796	Proteína del complejo de cosecha de luz 10	↓	-3,1
4746	Proteína del complejo de cosecha de luz 2	↓	-3,6
0194	Proteína del complejo de cosecha de luz 5	↓	-3,9
0499	Proteína del complejo de cosecha de luz 5	↓	-4,4

5 Como se proporciona en un gráfico de resumen en la Figura 19, todos los mutantes probados demostraron la desregulación, con respecto a la cepa de tipo silvestre, de los genes que se expresaban diferencialmente cuando las células de tipo silvestre se cambiaban de luz intensa a poca luz. La respuesta transcripcional global fue similar para todos los mutantes examinados; es decir, todos los genes ensayados que estaban regulados positivamente en relación con el tipo silvestre en el conjunto de datos transcriptómicos del mutante LAR1 GE4574 original (Ejemplo 9) se regularon positivamente en estas cepas mutantes adicionales que se perfilaron. Además, todos los genes regulados negativamente en relación con el tipo silvestre en el conjunto de datos transcriptómicos del mutante LAR1 GE4574 original también se regularon negativamente, y se regularon negativamente en una proporción similar, en estas cepas mutantes.

10 Ejemplo 11. Transcriptómica para el análisis de genes regulados por LARI

15 En otros experimentos de transcriptómica, el mutante LIHLA LAR1 GE5440 se cultivó con luz intensa ($500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) antes de cambiar a poca luz ($50 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) y cultivar durante dos días más, o, como control, manteniendo las células aclimatadas a la luz intensa durante dos días adicionales. Las células de *N. gaditana* de tipo silvestre se sometieron exactamente a la misma pauta: se aclimataron a la luz intensa antes de cambiar a poca luz y se cultivaron durante dos días, o se mantuvieron continuamente con luz intensa (diagramado en la Figura 20A). La Figura 20B muestra la cantidad de clorofila por célula a lo largo del tiempo de estos experimentos de cambio de luz, donde las células silvestres aclimatadas a alta luminosidad (cuadrados) aumentaron su contenido de clorofila aproximadamente dos veces durante los dos días siguientes a un cambio de luz intensa a poca luz, pero disminuyeron su clorofila levemente cuando, en lugar de cambiar a poca luz, se mantuvieron bajo luz intensa durante los dos días adicionales (círculos). Por el contrario, el mutante LAR1 aumentó su clorofila solo ligeramente durante los dos días siguientes a un cambio de luz intensa a poca (diamantes), dando como resultado un nivel de clorofila que era esencialmente el mismo que el nivel de clorofila de las células silvestres mantenidas con luz intensa, demostrando claramente el fenotipo "Bloqueado en Aclimatación a la luz intensa". Las células mutantes de LAR1 control que permanecieron con luz intensa durante el experimento (triángulos) mantuvieron su bajo nivel de clorofila, similar al tipo silvestre. Bajo estas condiciones, la curva de irradiancia para la fotosíntesis (Figura 20C) mostró que la evolución del oxígeno fotosintético a irradiancias más altas que $200 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ fue el más bajo en células de tipo silvestre que se cambiaron de luz intensa a poca (cuadrados). La P_{max} por clorofila fue algo más alta en células de tipo silvestre que se mantuvieron bajo luz intensa (círculos). Sorprendentemente, el mutante LAR1 LIHLA demostró un P_{max} mayor y E_k que el tipo silvestre, independientemente de si se mantuvieron con luz intensa (triángulos) o se cambiaron a poca luz (diamantes). P_{max} por clorofila fue aproximadamente el doble del valor para el mutante LIHLA encontrado en células de tipo silvestre, que tenía aproximadamente el doble de contenido de clorofila, indicando que el P_{max} por célula era aproximadamente equivalente al P_{max} de las células de tipo silvestre.

El ARN se extrajo en los puntos temporales 0, 4 h, 24 h y 48 h, donde el punto temporal 0 h fue el tiempo en que las células se desplazaron de luz intensa a poca luz (véase la Figura 20A) y se analizó como se proporcionó en el Ejemplo 9. Se utilizó RNA-seq para analizar la respuesta transcripcional global bajo condiciones de luz intensa en estado estacionario ($500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) o condiciones de cambio de luz intensa ($500 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) a poca luz ($50 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) para el tipo silvestre y el mutante LAR1.

El \log_2 de veces el valor de cambio de la respuesta de tipo silvestre al cambio a poca luz (es decir, el nivel de transcripción a poca luz (24 horas después del cambio) en comparación con el nivel de expresión a la luz intensa (pre-cambio) para cada gen se trazó con respecto al eje horizontal relativo al \log_2 de veces el valor de cambio del mismo gen para el mutante LA1 aclimatado a poca luz frente al tipo silvestre aclimatado a poca luz en el eje vertical. La agrupación resultante de los puntos graficados, que representan los genes expresados diferencialmente en el punto de tiempo de 24 horas (24 horas después de pasar de luz intensa a poca), reveló categorías discretas de ascendencia reguladora de la transcripción (TRAC) que fueron numeradas de -I+ a IV+ y de I- a IV- (Figura 21). Los TRAC hacen posible la organización de la regulación transcripcional del mutante LAR1 en relación con la respuesta de tipo silvestre a la luz posible: TRAC 0 son genes de categoría nula que no variaron significativamente en las condiciones de respuesta en el límite de significación (índice de descubrimiento falso, FDR, de 0,05 o menos) seleccionado en los puntos de tiempo probados. La expresión del gen TRAC I está desacoplada de la variable abiótica (intensidad de la luz) por la variable genética probada (mutación LAR). Muchas proteínas fotosintéticas están en esta categoría, junto con muchas desconocidas. Los genes TRAC II están regulados por la intensidad de la luz, y también regulados diferencialmente, pero en mayor grado, en el mutante. Los genes TRAC III son aquellos que se encuentran expresados diferencialmente entre el mutante y el tipo silvestre cuando se aclimatan a las mismas condiciones ambientales (poca luz), aunque no se expresan significativamente diferencialmente en el tipo silvestre en respuesta al cambio de luz. Los genes TRAC IV no mostraron expresión diferencial estadísticamente significativa entre el mutante y el tipo silvestre cuando se aclimataron a las mismas condiciones ambientales (poca luz); sin embargo, se ha encontrado que se regulan diferencialmente en respuesta al cambio de luz.

Los genes TRAC I incluyen una proporción significativa de genes fotosintéticos sensibles a la luz, incluidos los LHC, VCP, genes de biogénesis de clorofila, genes de ensamblaje de tilacoides, así como muchos genes que codifican proteínas desconocidas. Dichos genes, y otros identificados mediante este análisis, se pueden usar para la ingeniería adicional de cepas, por ejemplo, para alterar su expresión para modular el fenotipo LIHLA. La sorprendente similitud de la expresión de los genes TRAC I con los genes regulados por la intensidad de la luz en células de tipo silvestre se muestra en los gráficos lado a lado de la Figura 22 en los que los genes que están regulados positivamente en el tipo silvestre con luz intensa también se regulan positivamente en el mutante LAR1 LIHLA con respecto al tipo silvestre en aclimatación a poca luz, y los genes que están regulados negativamente en el mutante en condiciones de poca luz en comparación con el tipo silvestre también se regulan negativamente en células de tipo silvestre aclimatadas a la luz intensa.

La Figura 22 representa, en gráficos de gráficos de puntos uno al lado del otro, 127 genes TRAC I individuales representados por puntos ordenados verticalmente. En el gráfico de la derecha, cada punto que representa un gen se posiciona horizontalmente de acuerdo con su nivel de expresión en células silvestres aclimatadas a la luz 24 horas en relación con su nivel de expresión en células aclimatadas a la luz intensa, con genes con mayor expresión relativa en células aclimatadas a poca luz (es decir, el mayor grado de "sobreexpresión" en células aclimatadas a poca luz frente a células aclimatadas a la luz intensa) representados por puntos situados más hacia la derecha (\log_2 valores positivos) y genes que tienen niveles más bajos de expresión en células silvestres aclimatadas a condiciones de poca luz frente a las aclimatadas a la luz intensa representadas por puntos situados a la izquierda (\log_2 valores negativos). El eje x se calibra en una escala logarítmica 2 (\log_2), tal que "0" en el eje x indica que no hay cambios en los niveles de expresión basados en la aclimatación a la luz, "1" en el eje x representa una duplicación en el nivel de una transcripción, y "-1" en el eje x representa una reducción a la mitad en el nivel de una transcripción expresada en células aclimatadas a 24 horas de poca luz en relación con las células aclimatadas a la luz intensa. Por ejemplo, se sabe que los genes LHC se sobreexpresan bajo aclimatación a poca luz con respecto a sus niveles de expresión en células aclimatadas a la luz intensa; casi todos los genes caracterizados por el dominio pfam como LHC están representados por puntos encontrados en el extremo superior de la escala mostrada en la Figura 22, y su posición horizontal está a la derecha (lado positivo) de la línea punteada vertical a lo largo del eje x. El gráfico adyacente de la izquierda está alineado con el gráfico de la derecha, de modo que los puntos que representan el mismo gen se encuentran en la misma posición vertical en el gráfico de la izquierda que en el gráfico de la derecha, pero en el gráfico de la izquierda los puntos se posicionan horizontalmente de acuerdo con su nivel de expresión relativa en células de tipo silvestre aclimatadas a poca luz con respecto a las células mutantes LIHLA LAR1 (cuando ambas están aclimatadas a poca luz). El gráfico muestra que para un gran número de genes, los niveles de expresión relativa en el tipo silvestre frente al mutante LIHLA LAR1 son sorprendentemente similares a los niveles de expresión relativos en células tipo silvestre aclimatadas a poca luz frente a una aclimatación a la luz intensa, siendo regulados positivamente los genes regulados positivamente en las células tipo silvestre aclimatadas a poca luz en las células de tipo silvestre en comparación con las células mutantes LAR1 LIHLA, y siendo regulados negativamente los genes regulados negativamente en células de tipo silvestre aclimatadas a poca luz en las células de tipo silvestre con respecto a las células mutantes LAR1 LIHLA. Por lo tanto, el patrón de expresión génica en el mutante LAR1 LIHLA demuestra que los mutantes LIHLA se desregulan globalmente en condiciones de aclimatación de poca luz, ya que bajo condiciones de aclimatación de poca luz su patrón de expresión génica se

asemeja mucho al de las células de tipo silvestre aclimatadas a la luz intensa.

Ejemplo 12. Transcriptómica de un mutante LAR2

Los análisis transcriptómicos de un mutante LAR2 se realizaron de una manera similar pero simplificada como se describe para LAR1 en el Ejemplo 9. Los cultivos triplicados de WT-3730 y mutante LAR2 GE-5404 se llevaron todos a aclimatación a poca luz como se representa en la Figura 3. Las células se sedimentaron por centrifugación (4000 x g durante 5 minutos) y los sedimentos celulares se resuspendieron en tampón de lisis (Tris 50 mM, pH 8,0, EDTA 20 mM, pH 8,0, NaCl 300 mM) después de lo cual las resuspensiones se transfirieron a un tubo de microcentrifuga de 2 ml que contenía aproximadamente 0,5 ml de perlas de zirconio de 200 μm . Las células se lisaron mecánicamente batiendo perlas durante 3 minutos y luego los residuos se sedimentaron durante 2 minutos a 11,8 x g. A los sobrenadantes que contenían ARN se añadió SDS al ~1,5% p/v y se incubó durante 15 minutos a 50°C. Después de la incubación, se extrajo el ARN con un volumen igual de Fenol ácido: CHCl_3 seguido de una segunda extracción con 24 ml de 1-bromo-3-cloropropano. El ARN purificado se precipitó con una concentración final de LiCl 2,5 M con incubación durante la noche a 20°C. El sedimento de ARN se lavó una vez con etanol al 80%, se secó y luego se resuspendió en agua. El ARN se trató con ADNsa sin ADN siguiendo el protocolo del fabricante (Life Technologies, Carlsbad, CA). La calidad final del ARN se determinó mediante análisis de Agilent Bioanalyzer 2100 de acuerdo con las instrucciones del fabricante (Agilent Technologies, Santa Clara, CA). El ARNm se enriqueció usando selección de poliA y luego se usó para generar bibliotecas Illumina TruSeq Stranded mRNA LT (siguiendo las instrucciones del fabricante, Illumina Inc., San Diego, CA). Se utilizó RNA-seq para analizar la respuesta transcripcional global bajo una condición aclimatada de estado estacionario a poca luz ($50 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) para el tipo silvestre y el mutante LAR2. La Figura 23 es otra representación gráfica de puntos (como se describe en el Ejemplo 11, Figura 22), en este caso de la similitud de patrones de expresión génica en un mutante LAR2 LIHLA (GE-5404 que contiene una inserción del vector en la región promotora del gen LAR2, ver la Figura 14) en comparación con el tipo silvestre. La Figura 23 representa, en gráficos de lado a lado para ser vistos juntos, 186 genes individuales representados por puntos ordenados verticalmente. Los genes representados son genes TRAC I a partir del análisis transcriptómico de células tipo silvestre con aclimatación de poca luz vs. luz intensa y células de tipo silvestre vs. mutantes GE5404 LAR2 bajo poca luz, es decir, están desreguladas en el mutante LAR2 LIHLA bajo aclimatación a poca luz. En el gráfico de la derecha, cada punto que representa un gen se coloca horizontalmente de acuerdo con su nivel de expresión en células de tipo silvestre aclimatadas a poca luz en relación con su nivel de expresión en células aclimatadas a la luz intensa, con los genes con la expresión relativa más alta en células aclimatadas a poca luz (es decir, el mayor grado de "sobrexpresión" en células aclimatadas a poca luz frente a células aclimatadas a la luz intensa) representados por puntos posicionados a la derecha (\log_2 valores positivos) y con los genes con niveles de expresión más bajos en células de tipo silvestre aclimatadas a poca luz vs. luz intensa representados por puntos situados a la izquierda (\log_2 valores negativos). El gráfico del lado izquierdo está alineado con el gráfico de la derecha, de modo que los puntos que representan el mismo gen se encuentran en la misma posición vertical en el gráfico izquierdo que en el gráfico derecho, pero en el gráfico izquierdo los puntos se colocan horizontalmente de acuerdo a su nivel de expresión relativa en células de tipo silvestre frente a células mutantes LIHLA LAR2 (cuando ambas están aclimatadas a poca luz). El gráfico muestra que para una gran cantidad de genes, los niveles de expresión relativa en el mutante de tipo silvestre frente a LIHLA LAR2 son sorprendentemente similares a los niveles de expresión relativa en células tipo silvestre aclimatadas a poca luz vs. aclimatadas a la luz intensa, con los genes regulados positivamente en células tipo silvestre aclimatadas a poca luz siendo regulados positivamente en las células de tipo silvestre en comparación con las células mutantes LAR2 LIHLA, y los genes regulados negativamente en células de tipo silvestre aclimatadas a poca luz estando regulados negativamente en las células de tipo silvestre con respecto a las células mutantes LAR2 LIHLA. Por lo tanto, el patrón de expresión génica en el mutante LAR2 LIHLA muestra que los mutantes LIHLA están globalmente desregulados en aclimatación a poca luz, ya que bajo condiciones de poca luz su patrón de expresión génica se asemeja al de las células de tipo silvestre aclimatadas a la luz intensa.

Ejemplo 13. Inactivación de ortólogos LAR1 y LAR2 en *Nannochloropsis oceanica*.

Para demostrar aún más que las mutaciones en LAR1 y LAR2 son responsables del fenotipo LIHLA, y también para determinar la funcionalidad de genes homólogos de otras especies de algas, se identificaron ortólogos de estos reguladores en *Nannochloropsis oceanica*, una especie en la que se ha demostrado la recombinación homóloga. Las construcciones que incluían el ortólogo LAR1 de *N. oceanica* (gen No-LAR1, SEQ ID NO: 7) interrumpido por un gen de resistencia a blasticidina (SEQ ID NO: 48) y ortólogo LAR2 de *N. oceanica* (gen No-LAR2, SEQ ID NO: 20) interrumpido por un gen de resistencia a blasticidina (SEQ ID NO: 49) se linealizaron y se transformaron independientemente en la cepa W-5473 de *N. oceanica* esencialmente de acuerdo con el procedimiento proporcionado en el Ejemplo 2. Las colonias resistentes a antibióticos se aislaron y cribaron por PCR para la integración de la construcción inactivada mediante recombinación homóloga. Los aislamientos inactivados recombinantes se cribaron según su fluorescencia en clorofila baja.

La figura 24B muestra la clorofila total por célula de la cepa de tipo silvestre WE-5473 aclimatada a poca luz de *Nannochloropsis oceanica*, que se usó para generar los inactivados, así como la clorofila por célula de mutantes inactivados LAR1 aclimatados a poca luz (GE-6054 y GE-6055) y mutantes inactivados LAR2 (GE6052 y GE6053). En todas las cepas inactivadas, consistentes con el fenotipo LIHLA, la cantidad de clorofila por célula se reduce al menos en un 50%. La Figura 24 también proporciona gráficos que muestran las tasas de evolución de oxígeno para las cepas de tipo silvestre y de inactivación. En una base por célula, las tasas de evolución de oxígeno de los

mutantes inactivados son algo inferiores a las del tipo silvestre (Figura 24A), aunque las tasas de evolución de oxígeno por clorofila de los mutantes inactivados son significativamente más altas que las tasas de evolución del oxígeno por clorofila del tipo silvestre (Figura 24C). La Figura 25 muestra los resultados de los estudios de fluorescencia para determinar qP de dos cepas inactivadas del gen LAR1 y dos cepas inactivadas del gen LAR2. Los parámetros de fluorescencia de la cepa silvestre de *N. oceanica* WE-5473 se proporcionan en los mismos gráficos para su comparación. Para ambos inactivados, consistente con el fenotipo LIHLA visto en los mutantes de *N. gaditana*, qP es significativamente mayor en los inactivados en todas las irradiancias sobre 100 $\mu\text{mol fotones m}^{-2} \text{s}^{-1}$ que en la cepa progenitora de tipo silvestre. Por lo tanto, el homólogo de LAR1 en *N. oceanica*, que muestra aproximadamente el 49,5% de identidad con el polipéptido LAR1 de *N. gaditana* y aproximadamente el 82% de identidad sobre el dominio de TAZ extendido (Tabla 3), demostró ser un homólogo funcional u ortólogo de LAR1 de *N. gaditana*. La homología de LAR2 en *N. oceanica*, que demuestra aproximadamente el 69% de identidad con el polipéptido LAR1 de *N. gaditana* y el 100% de identidad sobre el dominio de unión de ADN tipo myb extendido (Tabla 4), se demostró que era un homólogo funcional, u ortólogo, de LAR2 de *N. gaditana*.

Los inactivados LAR1 y LAR2 en *N. oceanica* se usaron en experimentos adicionales de transcriptómica para analizar los patrones de expresión génica en estos mutantes de aclimatación de luz globalmente desregulados. Los genes que se expresaron diferencialmente en el mutante LAR1 inactivado bajo aclimatación a poca luz con respecto al *N. oceanica* tipo silvestre bajo aclimatación a poca luz fueron sorprendentemente similares a los genes que se expresaron diferencialmente en el mutante LAR2 inactivado bajo aclimatación a poca luz con respecto al tipo silvestre de *N. oceanica* bajo aclimatación a poca luz. Esto se representa en la Figura 28, donde el valor para el \log_2 cambio en veces de LAR1 en el nivel de expresión (mutante vs. tipo salvaje) para cada gen se representa frente al \log_2 cambio en veces de LAR2 en el nivel de expresión (mutante vs. tipo salvaje) para el mismo gen. Se puede ver que el trazado de la expresión diferencial para LAR1 vs. LAR2 muestra que casi todos los puntos caen en una línea diagonal directa de 1:1, lo que indica que estos genes inactivados producen efectos transcripcionales globales casi idénticos.

Además de los inactivados de *N. oceanica*, se generaron líneas knock-down de *N. gaditana* que exhibían las características de LIHLA del mutante LAR1 (ver Figura 6) usando la construcción de RNAi mostrada en la Figura 27. La construcción de ARNi dirigida al gen LAR1 (SEQ ID NO: 50) se introdujo en *N. gaditana*, dando como resultado niveles variables de transcripción de LAR1 en diferentes líneas de inactivación (Figura 28). La Tabla 6 proporciona datos que muestran que la caída de LAR1-ARNi exhibía características fisiológicas de otros mutantes LAR1, que incluyen más del doble de P_{max} por clorofila sin disminución de P_{max} por célula, ETR_{PSII} máximo superior y E_k mayor.

Tabla 6. Foto-fisiología de la línea inactivada de ARNi LAR1 de *N. gaditana*

	Células/ml	Chl mg/L prom.	pmol O_2 /hora/célula	$\mu\text{mol O}_2$ /hora/mg chl	ETR_{PSII} max	Cambio en veces de ETR_{PSII} max	E_k (μE)
WT-3730	4,84 E07	15,717	0,022	68,72	45,3	1,00	190,4
LAR1-ARNi	1,05 E08	15,314	0,023	154,76	93,5	2,06	353,9

Ejemplo 14. Cultivos mutantes de LIHLA en el sistema de cultivo a escala reducida

Para las pruebas de laboratorio de los cultivos de LIHLA, se colocó una caja de luz junto a una serie de matraces de cultivo que incluían 500 ml de medio PM074 (10x f/2) para replicar un día soleado en un invernadero con un ciclo de 12 horas de luz/12 horas de oscuridad con una irradiancia máxima de $\sim 1800 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$. Se alimentó continuamente una mezcla de 1% de CO_2 . Los cultivos se diluyeron a una densidad celular de 2×10^7 células ml^{-1} durante 3 días para una aclimatación en estado estacionario. El día 7, se añadió un nitrato 4 mM adicional. Las células se dejaron crecer durante 12 días y los recuentos celulares se registraron diariamente.

Se descubrió que todos los mutantes LIHLA de *Nannochloropsis gaditana* GE-5492, GE-4908 y GE-5494 (gen LAR1 alterado), GE5491 (gen LAR2 alterado) y GE5489 (gen LAR3 alterado, descrito a continuación en el ejemplo 16) alcanzaron un mayor densidad celular que la cepa de tipo silvestre WT-3730, como se representa gráficamente en la Figura 29.

En un experimento separado, se cultivaron la cepa silvestre WT-3730 y el mutante LIHLA LAR1 NE-5282 en una serie de matraces de cultivo en el mismo sistema a escala reducida que incluía 500 ml de medio PM074 (10x f/2) para replicar un día de sol en un invernadero (14 horas de luz a 10 horas de oscuridad) con una irradiación máxima de $\sim 1800 \mu\text{mol fotones m}^{-2} \text{s}^{-1}$. Se alimentó una mezcla de CO_2 al 1% a los cultivos continuamente. Los cultivos se diluyeron diariamente eliminando 15% en volumen del cultivo y reemplazándolo con 15% a una densidad celular de 2^7 durante 3 días para un volumen en estado estacionario del cultivo y reemplazándolo con medio recién preparado. Las células se dejaron crecer durante 10 días y el carbono orgánico total se midió diariamente diluyendo 2 ml de cultivo celular hasta un volumen total de 20 ml con agua DI. Se inyectaron tres inyecciones por medición en un analizador TOC-Vcsh de Shimadzu para la determinación del carbono total (TC) y el carbono inorgánico total (TIC). El horno de combustión se ajustó a 720°C, y se determinó el TOC restando TIC de TC. El rango de calibración de 4

puntos fue de 2 ppm a 200 ppm correspondiente a 20-2000 ppm para cultivos no diluidos con un coeficiente de correlación de $r^2 > 0,999$. El experimento fue repetido una vez. La Figura 30 muestra que el mutante LIHLA tenía una acumulación de biomasa diaria promedio ligeramente más alta. Aunque el aumento fue pequeño por día, dichos aumentos pueden acumularse para proporcionar ventajas en cultivos a largo plazo.

5 Ejemplo 15. Complementación de mutantes LIHLA

Con el fin de confirmar adicionalmente que las lesiones en el gen LAR1 y el gen LAR2 son responsables del fenotipo LIHLA, el mutante de *N. gaditana* GE5440, que tiene una inserción de construcción de vector en el gen LAR1 (Figura 13), se transformó con una construcción de la expresión en el que el gen LAR1 (SEQ ID NO: 3) se unió operativamente a un promotor TCTP de *Nannochloropsis* (SEQ ID NO: 2). La construcción se transformó en el mutante de *N. gaditana* GE5440 esencialmente de acuerdo con los métodos proporcionados en el Ejemplo 2. El análisis de qRT-PCR se realizó para confirmar que los genes LAR1 intactos transformados se expresaban en las células. Las cepas LAR1 complementadas, cepas del mutante LAR1 parental y células "abuelas" de tipo silvestre se cultivaron bajo niveles de irradiación moderadamente bajos ($100 \mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ durante 10 días y se diluyeron diariamente a una densidad celular de 1×10^8 células ml^{-1}). Durante este crecimiento semicontinuo, a poca luz se aclimata el control WT, se monitoreó la autofluorescencia de clorofila por célula hasta que fue alta y estable, lo que indica que se había logrado una aclimatación en condiciones de poca luz en estado estacionario. Entonces las células se adaptaron a la oscuridad y se controlaron mediante fluorescencia, como se describió anteriormente.

Como se muestra en la Figura 31, los mutantes complementados GE-5948, GE-5950 y GE-5951 han aumentado la clorofila por célula en comparación con el mutante GE-5440, aunque el nivel de clorofila natural por célula no se ha restaurado completamente en los mutantes complementados. De forma similar, los mutantes rescatados (GE5948, GE5950, GE5951) tienen tasas de evolución de O_2 por célula con las tasas de aproximación de tipo silvestre, y las tasas de evolución de O_2 por célula son muy similares a las tasas de tipo silvestre. La Figura 32 muestra que los mutantes complementados han restaurado en gran medida los valores de F_0 y F_m , así como los valores de qP que se aproximan a los valores de tipo silvestre, proporcionando una confirmación adicional de que LAR1 es un regulador de la aclimatación a la luz y que las mutaciones en el gen LAR1 son responsables del Bloqueo en el fenotipo de Aclimatación a la luz intensa.

Ejemplo 16. Genotipado de un mutante LAR3

Entre las treinta y cinco cepas de LIHLA aisladas, todas excepto una cepa, GE-5489, tenían una mutación en el locus LAR1 o LAR2. La re-secuenciación de los loci LAR1 y LAR2 en la cepa GE-5489 indicó que los genes LAR1 y LAR2 estaban intactos en GE-5489, indicando que un locus genético diferente era responsable del fenotipo LIHLA en esta cepa.

La secuenciación del genoma completo de GE-5489 identificó la ubicación genómica del vector insertado, así como varias deleciones de pares de bases únicas y polimorfismos de un solo nucleótido en otras partes del genoma. Se encontró que GE-5489 tenía el vector introducido insertado en una región intergénica ~ 200 pb aguas abajo del locus 7250 y ~ 300 pb cadena arriba del locus 7251 (Figura 33). Además, la resecuenciación del genoma identificó, con alta confianza, varias mutaciones adicionales presentes en esta cepa en relación con la cepa parental WT-3730, incluidas las deleciones del desplazamiento del marco de un solo par de bases en proteínas conservadas predichas E2-352 y E2-2665.

Para determinar la identidad del gen cuya interrupción era responsable del fenotipo LIHLA, se investigó la posibilidad de que el sitio de inserción del vector podría haber afectado a la expresión del supuesto gen de LPAAT putativo 7251. Se realizó un análisis de PCR de transcripción inversa cuantitativa (qRT-PCR) de GE-5489 para determinar si la abundancia de transcripción del gen 7251 difería de la observada en la cepa de tipo silvestre WT-3730. Además del gen E1-7251, se incluyeron tres genes de LHC en este experimento qRT-PCR, que en otros mutantes de LIHLA se ha demostrado que carecen de la regulación positiva característica en condiciones de poca luz observadas en cepas de tipo silvestre.

Para qRT-PCR, el ARN se aisló a partir de células WT-3730 de tipo natural de *N. gaditana* aclimatadas a poca luz ($50 \mu\text{E}$) y el mutante LAR3 GE-5489 como se describe en el Ejemplo 12. El ARN tratado con DNasa se transcribió de forma inversa usando Bio-Rad iScript de acuerdo con los protocolos del fabricante (Bio-Rad, Hercules, CA). El ADNc resultante se utilizó luego como plantilla en las reacciones de qPCR usando Bio-Rad SsoAdvanced, nuevamente de acuerdo con los protocolos del fabricante.

Tabla 6. Cebadores para qRT-PCR de genes expresados diferencialmente

Gen diana	Cebador	Secuencia
5307251	5-19	GTGAAACCAGCACTCAATCTCTC (SEQ ID NO: 51)
5307251	5-20	AGTTCGAATATCCTGCAATCGT (SEQ ID NO: 52)

5310499	F3	GGGAGGCTGAGATCAAGCAC (SEQ ID NO: 53)
5310499	F4	CAGGAGCCCTACCGTCATGT (SEQ ID NO: 54)
5310194	F5	TACTTGCAGGAGGCCGAGAT (SEQ ID NO: 55)
5310194	F6	AGGGAGCTAACAGCGTGGAC (SEQ ID NO: 56)
5312796	G1	TGCCCCGAAGAAGCTAGATG (SEQ ID NO: 57)
5312796	G2	CACGACCCAGCCTAGGAAAC (SEQ ID NO: 58)

El análisis demostró que los tres transcritos de proteína LHC ensayados por qRT-PCR estaban regulados negativamente de forma similar con respecto al tipo silvestre, como se observó en las otras cepas de LIHLA investigadas. El gráfico que se muestra en la Figura 34 proporciona una escala logarítmica 2 en el eje y, y las barras representan la expresión en veces de los genes evaluados en el mutante LAR3 GE-5489 con respecto al tipo silvestre WT-3730. La Figura 34 demuestra que el supuesto gen LPAAT E1-7251 se expresa en el mutante LAR3 GE-5489 a un nivel esencialmente idéntico al tipo silvestre, mientras que cada uno de los genes de LHC ensayados 10499, 10194 y 12796 se expresan a niveles varias veces más bajos en el mutante LAR3 GE-5489 con respecto a la cepa de tipo silvestre WT-3730, como se esperaba para los mutantes de LIHLA, cuando tanto el mutante LAR3 como la cepa de tipo silvestre están aclimatados a poca luz. Experimentos adicionales tampoco lograron identificar los cambios consistentes en la abundancia del transcrito para los transcritos sentido o antisentido del gen E1-7250 que se encuentra cadena arriba del sitio de inserción del vector o un supuesto ARN no codificante que flanquea el sitio de inserción del vector. Un análisis similar de los niveles de transcripción para el gen E1-7250 que está cadena arriba del sitio de inserción del vector tampoco pudo mostrar ninguna diferencia en la expresión de ese locus en relación con el tipo silvestre.

Aunque el análisis de transcripción no proporcionó indicaciones claras del locus responsable, dos enfoques adicionales confirmaron la identidad del gen LAR3 responsable del fenotipo LIHLA en GE-5489. En un enfoque, se llevaron a cabo experimentos de complementación para recuperar el fenotipo de tipo silvestre en la cepa GE-5489. En un segundo enfoque, se crearon inactivados dirigidos de varios loci en la especie relacionada *Nannochloropsis oceanica*. En ambos casos, se identificó una proteína conservada pronosticada (E2-352 en *N. gaditana* y ortólogo E1-8005 en *N. oceanica*) como el locus responsable del fenotipo LIHLA (Tabla 6).

Tabla 6. Validación genética del gen E2-352 como locus LAR3

Gen diana	ID Gen	Complementación, mutante <i>N. gaditana</i> GE-5489	Inactivación (KO) ortólogo <i>N. oceanica</i>
Proteína PHD-dedo	E1-7250	No se complementó	Fluorescencia de clorofila del tipo silvestre
LPAAT supuesto	E1-7251	No se complementó	KO sin éxito
Supuesto ARN no codificante		No se complementó	Fluorescencia de clorofila del tipo silvestre
Proteína "1" predicha conservada	E2-352	Complementación parcial	Fenotipo LIHLA
Proteína "2" predicha conservada	E2-2665	No se puede clonar	No se puede clonar

Los experimentos de complementación en *N. gaditana* dieron como resultado una complementación parcial para recuperar cerca de los niveles de clorofila por célula silvestre a través de la expresión heteróloga del marco de lectura abierto E2-352 ("proteína 1 predicha conservada") en GE5489. En estos experimentos, los vectores linealizados que incluían el gen que se iba a analizar junto con el promotor nativo y las secuencias terminadoras se transformaron en el mutante LIHLA GE-5489 esencialmente según el procedimiento proporcionado en el Ejemplo 2. Los transformantes se sembraron en placas que contenían higromicina y después de 2-3 semanas, las colonias se recogieron, se sembraron y luego se cultivaron en cultivo líquido para una aclimatación a poca luz antes de la selección usando un citómetro de flujo BD Accuri C6 (BD Biosciences, San Jose, CA) para determinar la fluorescencia media de la clorofila del cultivo en una base por célula. Las células de tipo silvestre (cepa de *N. gaditana* WT-3730), que contienen un vector vacío, también fueron aclimatadas a poca luz para usarse como control de los niveles de fluorescencia de clorofila por célula. La Figura 35 proporciona un gráfico que muestra la

fluorescencia media de clorofila por célula de cepas transformadas con el supuesto ARN no codificante, el supuesto gen LPAAT (E1-7251, la proteína PHD-dedo (E1-7250), un vector vacío como control y la "proteína 1 predicha conservada" (gen E2-352). La Figura 35 muestra claramente que la complementación con una proteína pronosticada conservada intacta (E2-352) da como resultado la recuperación de los niveles casi silvestres de clorofila media por célula.

Los inactivados dirigidos de genes ortólogos en una especie relacionada confirmaron estos resultados (Tabla 5). Las construcciones inactivadas para el locus LAR3 de *N. oceanica* incluyeron un gen de resistencia a la blasticidina (SEQ ID NO: 59) flanqueado en ambos lados por secuencias que variaban de 1-2 kb del locus que se estaba probando. La Figura 36 proporciona un diagrama de la construcción inactivada LAR3 que muestra las secuencias del locus LAR3 de *N. oceanica* (flanco izquierdo, SEQ ID NO: 60) en cualquier lado del gen de resistencia a la blasticidina (SEQ ID NO: 61). El fragmento de inactivación lineal se generó por PCR y se transformó en *N. oceanica* esencialmente de acuerdo con el protocolo en el Ejemplo 2.

El análisis fisiológico detallado de la cepa LAR3 de *N. Oceanica* inactivada junto con las cepas inactivadas LAR1 y LAR2 de *N. oceanica* muestra claramente que el inactivado dirigido de este gen produce una fisiología similar a la identificada en las otras cepas LIHLA (Figura 37). Se ha demostrado que las cepas inactivadas LAR1 (6054), LAR2 (6053) y LAR3 (6538) genéticamente modificadas tienen un contenido de clorofila significativamente reducido por célula (Figura 37A) y una Ek mayor (Figura 37C) con respecto a la cepa de *N. oceanica* de tipo silvestre 5473. Además, las velocidades de transporte de electrones del fotosistema II (ETR_{PSII}) también se incrementan a intensidades de luz superiores a 100 o superiores a 200 para los inactivados LAR1, LAR2 y LAR3 (Figura 37D), como lo es qP (Figura 37E). La Figura 37B muestra los niveles más altos de evolución de oxígeno (P_{max}) por clorofila para los inactivados LAR1 y LAR3.

Ejemplo 17. Homólogos LAR3

E2-352 (el gen LAR3, SEQ ID NO: 62) codifica una proteína predicha (SEQ ID NO: 63) de 841 aminoácidos y un peso molecular de 86,23 kDa. La alineación de secuencias múltiples (MUSCLE) de las secuencias traducidas de LAR3 y los 5 éxitos de homología principales de otros heterokontes/estramenopiles muestra una secuencia de ~100 aa que está altamente conservada (Figura 38). Esta sección corresponde a los residuos 302-405 de LAR3 (SEQ ID NO: 64). Esta región conservada puede usarse como un gancho para identificar los supuestos homólogos en otras especies tales como *Cydotella* y *Navicula*. La identidad de secuencia en esta región es de aproximadamente el 80% para las mejores coincidencias en estas cepas. En el valor $E < 1$, no hay coincidencias con los dominios PFAM-A conocidos en esta región altamente conservada. Existe una coincidencia significativa con un dominio PFAM-B (Pfam-B 10967) sobre los residuos 299 a 414 de LAR3 en el valor $E < 2,1E-12$, pero esta alineación automatizada de PFAM-B no está caracterizada.

El dominio conservado de la proteína LAR3 de *N. gaditana* (SEQ ID NO: 64) tiene una identidad del 96% con los aminoácidos 303 a 405 del ortólogo LAR3 de *Nannochloropsis oceanica* (SEQ ID NO: 66, codificado por SEQ ID NO: 65). En toda la proteína LAR3, el ortólogo LAR3 de *N. gaditana* y *N. oceanica* son 56% idénticos en el nivel de aminoácidos. La equivalencia funcional de estas proteínas se demuestra mediante los experimentos de inactivado realizados en *N. oceanica* en el Ejemplo 16, anterior, que dieron como resultado el fenotipo LIHLA.

El dominio conservado de la proteína LAR3 de *N. gaditana* (SEQ ID NO: 64) tiene un 62% de identidad con los aminoácidos 138 a 236 de un "producto de proteína no identificado" de *Phytophthora ramorum* (SEQ ID NO: 68, codificado por SEQ ID NO: 67); 86% de identidad con los aminoácidos 263 a 364 de una "supuesta proteína no caracterizada" de *Ectocarpus siliculosus* (SEQ ID NO: 70, codificada por SEQ ID NO: 69); 86% de identidad con los aminoácidos 591 a 692 de una "proteína hipotética" de *Aureococcus anophagefferens* (SEQ ID NO: 72, codificada por SEQ ID NO: 71); 76% de identidad con los aminoácidos 116 a 217 de una "proteína predicha" de *Thalassiosira pseudonana* (SEQ ID NO: 74, codificada por SEQ ID NO: 73); 87% de identidad con los aminoácidos 55 a 156 de una "proteína predicha" de *Phaeodactylum tricornutum* (SEQ ID NO: 76, codificada por SEQ ID NO: 75); y una identidad del 78% con los aminoácidos 176 a 263 de una "proteína no caracterizada" de *Thalassiosira oceanica* (SEQ ID NO: 78, codificada por SEQ ID NO: 77). Estos polipéptidos son supuestos ortólogos de los polipéptidos LAR3 de *Nannochloropsis*.

Ejemplo 18. Transcriptómica de LAR3

El perfil transcriptómico del mutante LAR3 GE-5489 se realizó exactamente como para el mutante LAR2 descrito en el Ejemplo 12. La Figura 39 es el mismo tipo de representación gráfica del gráfico de puntos de la similitud de los patrones de expresión génica en un mutante LIHLA en comparación con el tipo silvestre como se proporciona para un mutante LAR1 LIHLA en la Figura 22 y el mutante LAR2 en la Figura 23. La Figura 39 representa, en gráficos de lado a lado para ser vistos juntos, 327 genes individuales representados por puntos ordenados verticalmente. Los genes representados son genes TRAC I a partir del análisis transcriptómico de células silvestres con aclimatación a poca luz vs. luz intensa y células mutantes de tipo silvestre vs. GE5489 LAR3 bajo poca luz, es decir, están desreguladas en el mutante LAR3 LIHLA bajo aclimatación a poca luz. En el gráfico de la derecha, cada punto que representa un gen se coloca horizontalmente de acuerdo con su nivel de expresión en células de tipo silvestre aclimatadas a poca luz en relación con su nivel de expresión en células aclimatadas a la luz intensa, con genes con

la expresión relativa más alta en células aclimatadas a poca luz (es decir, el mayor grado de "sobreexpresión" en células aclimatadas a poca luz frente a células aclimatadas a la luz intensa) representadas por puntos situados a la derecha (\log_2 valores positivos) y genes que tienen niveles de expresión más bajos en células de tipo silvestre aclimatadas a poca luz vs. luz intensa representadas por puntos situados a la izquierda (\log_2 valores negativos). El gráfico del lado izquierdo está alineado con el gráfico de la derecha, de modo que los puntos que representan el mismo gen se encuentran en la misma posición vertical en el gráfico izquierdo que en el gráfico derecho, pero en el gráfico izquierdo los puntos se colocan horizontalmente de acuerdo a su nivel de expresión relativa en células de tipo silvestre frente a células mutantes LIHLA LAR3 (cuando ambas están aclimatadas a poca luz). El gráfico muestra que, para un gran número de genes, los niveles de expresión relativa en el tipo silvestre frente al mutante de LIHLA LAR3 son notablemente similares a los niveles de expresión relativa en células aclimatadas a poca luz vs. aclimatadas a la luz intensa, con los genes regulados positivamente en las células de tipo silvestre aclimatadas a poca luz estando regulados positivamente en las células de tipo silvestre en comparación con las células mutantes LAR3 LIHLA, y los genes regulados negativamente en células de tipo silvestre aclimatadas a poca luz estando regulados negativamente en las células de tipo silvestre con respecto a las células mutantes LAR3 LIHLA. Por lo tanto, el patrón de expresión génica en el mutante LAR3 LIHLA demuestra que los mutantes LIHLA se desregulan globalmente en aclimatación a poca luz, ya que bajo condiciones de aclimatación a poca luz su patrón de expresión génica se asemeja al de las células de tipo silvestre aclimatadas a la luz intensa. Además de todas las características fenotípicas y fisiológicas de una cepa LIHLA, GE-5489 muestra una desregulación global similar de los genes de la fotosíntesis y específicamente aquellos involucrados en la aclimatación a poca luz. Incluso en condiciones de aclimatación a poca luz, la gran mayoría de las proteínas del complejo de recolección de luz (LHC) están significativamente reguladas negativamente en esta cepa (Figura 40). Esta desregulación global de los genes relacionados con la fotosíntesis también es evidente entre todos los genes que se encuentran en los estramenopiles fotosintéticos, pero no en los estramenopiles no fotosintéticos, los genes "fotocorte".

Se ha demostrado que tres mutantes LIHLA independientes que tienen mutaciones en diferentes genes (LAR1, LAR2 y LAR3) se desregulan globalmente en la aclimatación a poca luz. Se demuestra que más de 100 genes que están regulados hacia arriba o hacia abajo por células de tipo silvestre en aclimatación a poca luz no están regulados, es decir, no están regulados de manera similar, por mutantes LAR1, LAR2 y LAR3 aclimatados a poca luz. Entre los genes desregulados están los genes que codifican las proteínas de unión a la clorofila de recuperación de luz o LHC. Las secuencias de LHC analizadas en este experimento se identificaron en el conjunto del genoma interno de *Nannochloropsis gaditana* mediante la inclusión del dominio pfam PF00504 (proteína de unión a clorofila a-b) en las secuencias polipeptídicas codificadas. Las secuencias de ácido nucleico se proporcionan como SEQ ID NO: 79 - SEQ ID NO: 102. La Figura 41 proporciona gráficos de puntos alineados para mutantes LAR1, LAR2 y LAR3, así como células de tipo silvestre, en las que los genes analizados codifican LHC o proteínas relacionadas. De arriba a abajo, los genes de *Nannochloropsis* analizados son, desde la parte superior a la parte inferior de la gráfica, la SEQ ID NO: 79, anotada como que codifica una proteína de unión a clorofila a/c de fucoxantina; SEQ ID NO: 80, también anotada como que codifica una proteína de unión a clorofila a/c de fucoxantina; SEQ ID NO: 81, anotada como que codifica una proteína compleja de recolección de luz 5 (precursor); SEQ ID NO: 82, también anotada como que codifica una proteína compleja de recolección de luz 5 (precursor); SEQ ID NO: 83, anotada como que codifica una proteína 10 del complejo de recolección de luz (precursor); SEQ ID NO: 84, también anotada como que codifica una proteína 10 del complejo de recolección de luz (precursor); SEQ ID NO: 85, anotada como que codifica una proteína Beta-Ig-H3/fasciclina; SEQ ID NO: 86, anotada como que codifica una proteína inducible por la luz temprana de masa molecular baja HV60; SEQ ID NO: 87, anotada como que codifica una proteína 2 de complejo de recolección de luz; SEQ ID NO: 88, anotada como que codifica la proteína 3 del complejo de recolección de luz; SEQ ID NO: 89, anotada como que codifica una proteína 3 del complejo de recolección de luz; SEQ ID NO: 90, anotada como que codifica una proteína 2 de complejo de recolección de luz; SEQ ID NO: 91, anotada como que codifica una proteína de unión a clorofila a/b del complejo I de recolección de luz; SEQ ID NO: 92, anotada como que codifica una proteína 2 del complejo de recolección de luz; SEQ ID NO: 93, también anotada como que codifica una proteína 2 del complejo de recolección de luz; SEQ ID NO: 94, anotada como que codifica una proteína del complejo de recolección de luz; SEQ ID NO: 95, también anotada como que codifica una proteína 2 de complejo de recolección de luz; SEQ ID NO: 96, anotada como que codifica una proteína 3 del complejo de recolección de luz; SEQ ID NO: 97, anotada como que codifica una proteína 2 del complejo de recolección de luz; SEQ ID NO: 98, anotada como que codifica una proteína 2 del complejo de recolección de luz; SEQ ID NO: 99, anotada como que codifica una proteína de unión de fucoxantina clorofila a/c; SEQ ID NO: 100, anotada como que codifica una proteína 3 del complejo de recolección de luz; SEQ ID NO: 101, anotada como que codifica una proteína del complejo de recolección de luz; y SEQ ID NO: 102, anotada como que codifica un precursor de beta-Ig-H3/fasciclina.

El gráfico de la izquierda muestra a lo largo del eje x la relación de expresión de cada uno de los 22 genes LHC en el tipo silvestre aclimatado a poca luz en comparación con las células LAR1 aclimatadas a poca luz, el siguiente gráfico muestra la expresión de tipo silvestre aclimatada a poca luz de cada gen LHC como una relación de su expresión en células LAR2 aclimatadas a poca luz, el tercer gráfico muestra la expresión de tipo silvestre aclimatada a poca luz de cada gen LHC como una relación de su expresión en células LAR3 aclimatadas a poca luz, y el gráfico más a la derecha muestra la expresión de tipo silvestre aclimatada a poca luz de cada gen de LHC como una relación de su expresión en células de tipo silvestre aclimatadas a la luz intensa. Los gráficos muestran que la expresión de al menos 20 genes LHC en células silvestres aclimatadas a poca luz difiere de la expresión del LHC en células

mutantes LIHLA aclimatadas a poca luz de una manera muy similar a la diferencia entre células tipo silvestre aclimatadas a poca luz y aclimatadas a la luz intensa. Es decir, los mutantes LIHLA se asemejan a las células de tipo silvestre aclimatadas a la luz intensa en su patrón de expresión de LHC, a pesar de que están aclimatadas a poca luz.

5 Listado de secuencias

```

<110> SYNTHETIC GENOMICS, INC.
      BAILEY, Shaun
      MCCARREN, Jay
      LIEBERMAN, Soyan Leung
      MEUSER, Jonathan E.
      ROMANO, Anna E.
      YEE, Daniel
      SORIAGA, Leah
      BROWN, Robert C.
      WEISSMAN, Joseph C.
      PRINCE, Roger C.
      NIELSEN, Robert D.
      SCHWARTZ, Ariel S.

<120> MUTANTES DE ALGAS QUE TIENEN UN FENOTIPO ACLIMATADO A LA LUZ INTENSA EN COMPARTIMENTOS

<130> SGI1700-4WO

<150> US 61/733,956
<151> 06-12-2012

<150> US 61/810,216
<151> 04-09-2013

<150> US 61/869,590
<151> 23-08-2013

<150> US 61/881,342
<151> 23-09-2013

<160> 102

<170> PatentIn version 3.5

<210> 1
<211> 4118
<212> ADN
<213> Secuencia artificial

<220>
<223> Constructo pSG-05534 para mutagénesis de inserción

<400> 1
gcgccaata cgaaaacgc ctctccocgc gcgttgccg attcattaat gcagctggca      60
cgacaggtt cccgactgga aagcggcag tgagcgcaac gcaattaatg tgagttagct      120
cactcattag gcaccccagg ctttacactt tatgcttcog gctogtatgt tgtgtggaat      180
tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg ccaagcttgc      240
atgcctgcag gtgcactota gagcgtgcag gtgtacagat tgaaggaaac aatggagata      300
tctttggcag ttgaaaacg tgttogaatc atgctcttct actctccaac tgagaogaaa      360
tttatagcgc catctcgtt ctgactacca ggcttaggaa ggcctcatca caagctggat      420
cggttcgaat taagcaggca ctgaagccaa gtttgcgaaga cagccacott ttaattctct      480
caaacactt tctcaattca gcccgtaaa tatgccgatt cacagcggcc aagatagagg      540

```

ggaggttagc aagaatggtg cgatccctcc ccagtcggtg cctcgcacac aacctaggac 600
 ttcacctttc catggaaaat tgagaagtga atattggttt tcttacggca taccagatga 660
 aatcatgacc cctaaacatg aagagctgca ggcaaaacac ctgctctgga cgagcacgat 720
 gaaatctoga gaaccocgocg tacttcagtt gatcccgcat gatgacggcc gccattgaaa 780
 taagccacct cactttattc tagcaccgat ttccaccggt gtgagggccg aacgaggaca 840
 atttcgtgcg aaacaagcac gaacgcgcac acgattagta ggacagacga gcagatcgat 900
 ggcattgccc acggtctcgc gttctcggcg accaggacaa cggagcagag ggaggcctgc 960
 cgagttccga ggggcatttt agtcccaaaa ttgtgttgac acgtgaacaa gtggcttgaa 1020
 aagaggaagg aatgcctggg gtttcccttc gagagcggga actcgtttgt gcgtcatcct 1080
 agctacccat ggtccctttg tgggggagc tgtttcgtcc taccgaatgt gtggcgctcc 1140
 atgcatcttc tgcctcccaa accaccaaca tgagcacgcg aaggaaggag aaaaaagtgg 1200
 ccgcaacggt ctcttctcat atttattgtc tcatcaciaa cataggtaca taatacaaca 1260
 atcatggatc ccgggtacc gagctcgatg gccaagcctt tgtccaaga ggaatccacg 1320
 ctgatcgaac gtgcaactgc gaccatcaac agcataccta ttagcgagga ctactcgggtg 1380
 gccagtgcag ccctctcgtc cgacggtcgg atctttaccg gcgtgaatgt atatcatttc 1440
 accggagggc catgcgcgga gctcgtggtc ctcggaacgg ccgctgcggc tgctgcogga 1500
 aatctgacgt gcatagtggc catcgggaac gaaaaccgcg gcattctgtc tccgtgcggg 1560
 cgatgtcggc aggtgctgct tgacttgac cgggggatca aggcaattgt caaagattcc 1620
 gatgggcagc ccacagcggg tggcatcagg gagttgcttc cctctggcta cgtctgggag 1680
 ggttgaatc cactggcogt cgttttaciaa cgtcgtgact gggaaaacc tggcgttacc 1740
 caacttaac gccttgacgc acatcccctt ttcgcccagc ccataatacc cataatagct 1800
 gtttgccact ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca acagttgocg 1860
 agcctgaatg gcgaatggcg cctgatgcgg tattttctcc ttaocgatct gtgcggtatt 1920
 tcacaccgca tatggtgcac tctcagtaca atctgctctg atgcccata gttaagccag 1980
 ccccgacacc cgccaacacc cgctgacgcg ccctgacggg cttgtctgct ccggcatcc 2040
 gcttacagac aagctgtgac cgtctccggg agctgcatgt gtcagagggt ttcaccgtca 2100
 tcaccgaaac gcgcgagacg aaagggcctc gtgatacgc tatttttata ggttaatgtc 2160
 atgataataa tggtttctta gacgtcaggt ggcacttttc ggggaaatgt gcgcggaacc 2220
 cctatttgtt tatttttcta aatacattca aatatgtatc cgtcctatgag acaataacc 2280
 tgataaatgc ttcaataata ttgaaaaagg aagagtatga gtattcaaca tttccgtgtc 2340
 gcccttattc ccttttttgc ggcattttgc cttcctggtt ttgctcacc agaaaocgtg 2400
 gtgaaagtaa aagatgctga agatcagttg ggtgcacgag tgggttacat cgaactggat 2460

ES 2 682 279 T3

ctcaacagcg gtaagatcct tgagagtttt cgccccgaag aacgttttcc aatgatgagc 2520
 acttttaaaag ttctgctatg tggcgcggtta ttatcccgta ttgacgccgg gcaagagcaa 2580
 ctcggtcgcc gcatacacta ttctcagaat gacttggttg agtactcacc agtcacagaa 2640
 aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgcat aacctagagt 2700
 gataaacactg cggccaactt acttctgaca acgatcggag gaccgaagga gctaaccgct 2760
 tttttgcaca acatggggga tcatgtaact cgccttgatc gttgggaacc ggagctgaat 2820
 gaagccatac caaacgacga gcgtgacacc acgatgcctg tagcaatggc aacaacgttg 2880
 cgcaaactat taactggcga actacttact ctagcttccc ggcaacaatt aatagactgg 2940
 atggaggcgg ataaagttgc aggaccactt ctgcgctcgg cccttcggc tggctggttt 3000
 attgctgata aatctggagc cggtgagcgt gggctcgcg gtatcattgc agcactggg 3060
 ccagatggta agccctccc tatcgtagtt atctacacga cggggagtca ggcaactatg 3120
 gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca ttggtaactg 3180
 tcagaccaag tttactcata tatactttag attgatttaa aacttcattt ttaatttaa 3240
 aggatctagg tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt 3300
 tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt 3360
 tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt 3420
 ttgccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag 3480
 ataccaaata ctgttcttct agttagccg tagttaggcc accacttcaa gaactctgta 3540
 gcaccgccta catacctcgc tctgctaate ctgttaccag tggctgctgc cagtggcgat 3600
 aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggctc 3660
 ggctgaacgg ggggttcgtg cacacagccc agcttgagc gaacgacct caccgaactg 3720
 agatacctac agcgtgagct atgagaaagc gccacgcttc ccgaaggag aaaggcggac 3780
 aggtatccgg taagcggcag ggtcggaaca ggagagcgca cgaggagct tccaggggga 3840
 aacgcctggt atctttatag tcctgtcggg tttcgccacc tctgacttga gcgtcgattt 3900
 ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggcttttta 3960
 cggttcctgg ccttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat 4020
 tctgtggata accgtattac cgcctttgag tgagctgata ccgctcggc cagccgaacg 4080
 accgagcgca gcgagtcagt gagcggaggaa gcggaaga 4118

<210> 2
 <211> 1002
 <212> ADN
 <213> Secuencia artificial

<220>

<223> Promotor TCTP

<400> 2

```

cgtgcagggtg tacagattga aggaacaat ggagatatct ttggcagttg aaaaccgtgt      60
tcgaatcatg cttttctact ctccaactga gacgaaattt atagcgccat ctcgcttctg      120
actaccaggc ttaggaaggc ctcatcacia gctggatcgg ttcgaattaa gcaggcactg      180
aagccaagct tgcaagacag ccacctttta attctctcaa aacactttct caattcagcc      240
cggtaaatat gccgattcac agcggccaag atagagggga ggtagcaag aatggtgcga      300
tccctcccca gtcggtgcct cgacacaaac ctaggacttc acctttccat ggaaaattga      360
gaagtgaata ttggttttct tacggcatat cagatgaaat catgaccctt aaacatgaag      420
agctgcaggc aaaacacctg ctctggacga gcacgatgaa atctogagaa cccgccgtac      480
ttcagttgat cccgcatgat gacggccgcc attgaaataa gccacctcac tttattctag      540
caccgatttc caccgttggt agggccgaac gaggacaatt tcgtgcgaaa caagcacgaa      600
cgcgcacacg attagtagga cagacgagca gatcgatggc atcgggcacg gtctcgcgtt      660
ctcggcgacc aggacaacgg agcagagggga ggcctgccga gttccgaggg gcattttagt      720
ccaaaattgt gttgacacgt gaacaagtgg cttgaaaaga ggaaggaaat gcctgggttt      780
cccttcgaga gcgggaactc gcttgtgcgt catcctagct acccatggtc cctttgtggg      840
ggaggtggtt tcgtcctacc gaatgtgtgg cgctccatgc atcttctgcc tcccaaacca      900
ccaacatgag cacgcgaagg aaggagaaaa aagtggccgc aacgttctct tctcatattt      960
attgtctcat cacaacata ggtacataat acaacaatca tg                               1002
    
```

<210> 3

<211> 1995

<212> ADN

<213> Secuencia artificial

<220>

<223> Gen que codifica LAR1

<400> 3

```

atgagcaaca tccttgccgc aggggactac gacatgcaca gaagtagcga tgtggaactt      60
aagcaagagg cgtcagccaa catgaaatca aacgcccagg ttggcctgca tccgtcacag      120
aatcagcagc ttctgcaaca acaagtgcag cagccgcagg gacaggagga gagggggccg      180
aagacagcga cgccaccatg tctatcagag ggaagatata cctcatttct ggctccgctg      240
aatcactga cctccccctg ggcgtcttcg gtgttcgagg cggacgcgaa gcagcagcaa      300
ctcttgaaag attccctcac cgcagacctt aaactgctct tgcaacgagtt tgaacgcttc      360
cagcaagcga cagcattagt gtcgagagag ggctcgaaaag aggtggaggc aatggagcgg      420
gcggcgaaaag tggaattctt cctaggctac atcgaaaagg tgcttcagga acttgccggc      480
    
```

ES 2 682 279 T3

gccgacgcac cgaagctcca ggaattagag gttoggatca agaccagcct ccttccattg 540
 aaggggcaag tggatgaaca gcttgcatct tccctgctct cgtcctccgc cctcgggtggt 600
 ctgcagcatg agccttctct ctccgcatcc atcccgctgc cttcctcttc tcttctctcc 660
 tcatgcagca cccacaccac tcccccatc tccccgatat caggggagaa gatgacagtc 720
 caggatgatg gccggagggg gaggacgcat ccgacggccg ccgogctcat gccctccgtg 780
 cgagtccaac gcctcgacag ctctccagt ggcgccacca cctgctcgga gaactccgag 840
 gaggggcgcg gacagcttga cgatatggag tgcctgagcc tgctcatgga ggaagacggc 900
 cagggactgg ggaggccaca ggacaggacg gggggcggga gggagtgggg cctgggcccc 960
 gacgaggacg tgacggacgc ggccagcctg gtgtcggagg aaagcagcaa cgtcttcgct 1020
 ccctcgcccc gggaggcgct ggagatgctg gagacgatca gcagggggct ggggtctggg 1080
 caaaagaggc tggccttgtc ggagggcagc acgggcagcc tgggcttggc ctacagcgtcc 1140
 ttcacctcgg ttgagagcat gcagcagctc aagagggcgc gcagcaccgt gattcccagc 1200
 atcagcagca atagcatcag caccaccacc accagcagcg tgagcatcgg ccatggaagc 1260
 agcgaggttt cgaaccctc catgtccgtt gcgagcgcgg agaagcctca ggtgcgacag 1320
 gtogagtatc agtgcggggt ctgtcgggaa tcgtacagcg ctgccgcctc cctgaacccc 1380
 tgggtggcgc tggagaagca ggagtgccca caatgcaaga agctgcagat tctcggatc 1440
 gacattaacc tgccggccaa caccatggac tatcaccocg ccctgctcgc ggaggagggg 1500
 gacgacgacg acgaggaaga ggggctgggc ctgggagggg gcggattggt ccctggggaa 1560
 gagtttcggg ggatggggaa gggaggggag gactcggcgg tcggcggcgg gataggcgtc 1620
 ccggaggagg aggacgggca ggccttctcc ccggcgcaag cctcgcagat attggagtgt 1680
 atgtcgacg cccggacctg tcccggccac caccactcgg aggccaccg cgcggtgtgt 1740
 acctccacca agtacctgat gctgcatgtg agggactgcg acgggaagac cctggacggg 1800
 gaggcctgtg gcttctctg gtgccgcctc tgcaagcacc tgctgggtca cctggtgcgt 1860
 tgttatgagt cggagcagtg cagcatttgt cgtccccaga agcgtgagcc gtgagaggaa 1920
 gcggtcgcgt gcaagagggg gaccagcggg gcggtgaggg agggtgtgta tcgggcaactg 1980
 acgagcttgt gctaa 1995

<210> 4
 <211> 664
 <212> PRT
 <213> *Nannochloropsis gaditana*
 <400> 4

Met Ser Asn Ile Leu Ala Ala Gly Asp Tyr Asp Met His Arg Ser Ser
 1 5 10 15

ES 2 682 279 T3

Asp Val Glu Leu Lys Gln Glu Ala Ser Ala Asn Met Lys Ser Asn Ala
 20 25 30

Gln Val Gly Leu His Pro Ser Gln Asn Gln Gln Leu Leu Gln Gln Gln
 35 40 45

Val Gln Gln Pro Gln Gly Gln Glu Glu Arg Gly Pro Lys Thr Ala Thr
 50 55 60

Pro Pro Cys Leu Ser Glu Gly Arg Tyr Thr Ser Phe Leu Ala Pro Leu
 65 70 75 80

Lys Ser Leu Thr Ser Pro Val Ala Ser Ser Val Phe Glu Ala Asp Ala
 85 90 95

Lys Gln Gln Gln Leu Leu Lys Asp Ser Leu Thr Ala Asp Leu Lys Leu
 100 105 110

Leu Leu His Glu Phe Glu Arg Phe Gln Gln Ala Thr Ala Leu Val Ser
 115 120 125

Arg Glu Gly Ser Lys Glu Val Glu Ala Met Glu Arg Ala Ala Lys Val
 130 135 140

Glu Phe Phe Leu Gly Tyr Ile Gly Lys Val Leu Gln Glu Leu Ala Gly
 145 150 155 160

Ala Asp Ala Pro Lys Leu Gln Glu Leu Glu Val Arg Ile Lys Thr Ser
 165 170 175

Leu Leu Pro Leu Lys Gly Gln Val Val Asn Lys Leu Ala Ser Ser Leu
 180 185 190

Leu Ser Ser Ser Ala Leu Gly Gly Leu Gln His Glu Pro Ser Ser Ser
 195 200 205

Ala Ser Ile Pro Ser Pro Ser Ser Ser Pro Ser Ser Ser Cys Ser Thr
 210 215 220

His Thr Thr Pro Pro Ile Ser Pro Val Ser Gly Glu Lys Met Thr Val
 225 230 235 240

Gln Asp Asp Gly Arg Arg Glu Arg Thr His Pro Thr Ala Ala Ala Leu
 245 250 255

Met Pro Ser Val Arg Val Gln Arg Leu Asp Ser Ser Ser Ser Gly Ala

ES 2 682 279 T3

Gly Ser Gly Leu Val Pro Gly Glu Glu Phe Arg Gly Met Gly Lys Gly
515 520 525

Gly Glu Asp Ser Ala Val Gly Gly Gly Ile Gly Val Pro Glu Glu Glu
530 535 540

Asp Gly Gln Ala Phe Ser Pro Ala Gln Ala Ser Gln Ile Leu Glu Leu
545 550 555 560

Met Ser His Ala Arg Thr Cys Pro Gly His His His Ser Glu Ala His
565 570 575

Arg Ala Val Cys Thr Ser Thr Lys Tyr Leu Met Leu His Val Arg Asp
580 585 590

Cys Asp Gly Lys Thr Leu Asp Gly Glu Ala Cys Gly Phe Ser Trp Cys
595 600 605

Arg Pro Cys Lys His Leu Leu Gly His Leu Val Arg Cys Tyr Glu Ser
610 615 620

Glu Gln Cys Ser Ile Cys Arg Pro Gln Lys Arg Glu Pro Cys Glu Glu
625 630 635 640

Ala Val Ala Cys Lys Arg Glu Thr Ser Gly Ala Val Arg Glu Gly Val
645 650 655

Tyr Arg Ala Leu Thr Ser Leu Cys
660

<210> 5
<211> 1821
<212> ADN
<213> Secuencia artificial

<220>
<223> Gen que codifica LAR2

<400> 5
atgggaatcg gctcagaggg cgtcgctgag gtagggagcg gcacaagcag cgcaccgtgt 60
ggcagcaact cagcaagccc aatgatgcct gtcgtgactg cgagtcagaa cggggctgag 120
agcgtagcag gctctatgct acctgtaagt tgcgtggcct ccgctgcgat cggccctcaa 180
acgtcctcga cgacagtggg agtcgctgcc agcggccctg cagcgtcccc aacgacctcg 240
aagctgcaca aaggcacgtc ctctatgacc caactcggcg aaaccggtcg agagaacacg 300
ggtcgatgga cttgcgagga acatgttttg ttcttaaaag gcttggagat gcacgggaaa 360
ggctggaaaa aaattgcaa attgattaa acccggaccg ttgtccagat cgggaccac 420

ES 2 682 279 T3

gcgcagaaat atttccaaaa attggcaaaa gcaaagaaaa atgggcacca tggggacatg 480
 cttggaatgg aaggcacacg attcggggga aagcgtgtga agttcactgg gaaacggcga 540
 gggttggtct acgggagcta cttggtcggg gcgaggcga cttccgcagc catctctcct 600
 gcottgcagt cctacatgcc cgggtccttg gccgggcgtg aggaagggga agcgcgtgctg 660
 gataaggagg aggacgccgc catcgagaaa gggctttatc gatttctgtc tcccgtcgtgta 720
 ttggacgctg cggcgagtaa cctggacgcg accgcgccgg aagttttacc tccgagcaca 780
 ccgggaacag gagtgcacgc caacggggta gtgggggcag acggcgagac gacggaggag 840
 gatgggagca gcggcgggga caacgtggac gtcagtgaga cgatcgatga cgctgactcc 900
 tcatcggggc agcccctgcc gcgcttgccc cgcgtcacga atgacatgta tgagcgtatg 960
 agtgtgccca cctggtttat gaagggcggg gatatacgaag agctcctggc cgatgctgcc 1020
 gcgatcgact ggcgcgagga ctccggggga gacgcggtga aggctgaaga gcgcggggcg 1080
 agcattttga acgccaatat tgagagttcc gaccaagctc agagccacag gaatggcgag 1140
 aaagtggccg ttccgaatag ggtcgcagcc gtgaaagtgg caggaaacat ttccgcagac 1200
 agttcttaca taacatcggg tgccagccaa tgcgtcaacc accctcctac gaccaacact 1260
 atcgatggca agcgcagcaa catgaattcg gggcagtcgc tgccaaccgc gaatggacgg 1320
 aacggtgcag cctcaggacg atgcgtcaca gggcgtggac agcagcaaaa gaaaaagcag 1380
 ccaaaaacac aagaatcggg gaaccacggc aagcagcaat tagtgaacaa gcatagcgcc 1440
 cctacggggag acatgttcca gtcaaaggca ggcgctggca cgatcgtcct tcaaggatgat 1500
 atcaactgct ttgcatcttt ggaccgcac catggtgagc tgaaggagga gcattcccac 1560
 caagagctac gtttgaaga gttgcagcaa agcggcgtca acgatgacac gttcgcacac 1620
 atggatttct tggcgaacga cgaggcggc gtggaccatc acgcaggcca ccttcacacc 1680
 atttctacgc acgatgacgt gcatagtcat gccgaccatg acgttcatat gcgcaactat 1740
 gacgtctttt ggaaagatac ggtacctgac ggcgaccatg gcttgctgct cgatgatttt 1800
 gacggcggaa tcgaattttg a 1821

<210> 6
 <211> 606
 <212> PRT
 <213> Nannochloropsis gaditana
 <400> 6

Met Gly Ile Gly Ser Glu Ala Val Ala Glu Val Gly Ser Gly Thr Ser
 1 5 10 15

Ser Ala Pro Cys Gly Ser Asn Ser Ala Ser Pro Met Met Pro Val Val
 20 25 30

ES 2 682 279 T3

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

ES 2 682 279 T3

Val Asp Val Ser Glu Thr Ile Asp Asp Ala Asp Ser Ser Ser Gly Glu
 290 295 300

Pro Leu Pro Arg Leu Ala Arg Val Thr Asn Asp Met Tyr Glu Arg Cys
 305 310 315 320

Ser Val Pro Thr Trp Phe Met Lys Gly Gly Asp Ile Glu Glu Leu Leu
 325 330 335

Ala Asp Ala Ala Ala Ile Asp Trp Arg Glu Asp Ser Gly Gly Asp Ala
 340 345 350

Val Lys Ala Glu Glu Arg Gly Ala Ser Ile Leu Asn Ala Asn Ile Glu
 355 360 365

Ser Ser Asp Gln Ala Gln Ser His Arg Asn Gly Glu Lys Val Ala Val
 370 375 380

Pro Asn Arg Val Ala Ala Val Lys Val Ala Gly Asn Ile Ser Ala Asp
 385 390 395 400

Ser Ser Tyr Ile Thr Ser Gly Ala Ser Gln Cys Val Asn His Pro Pro
 405 410 415

Thr Thr Asn Thr Ile Asp Gly Lys Arg Ser Asn Met Asn Ser Gly Gln
 420 425 430

Ser Leu Pro Thr Ala Asn Gly Arg Asn Gly Ala Ala Ser Gly Arg Cys
 435 440 445

Val Thr Gly Arg Gly Gln Gln Gln Lys Lys Lys Gln Pro Lys Thr Gln
 450 455 460

Glu Ser Gly Asn His Gly Lys Gln Gln Leu Val Asn Lys His Ser Ala
 465 470 475 480

Pro Thr Gly Asp Met Phe Gln Ser Lys Ala Gly Ala Gly Thr Ile Val
 485 490 495

Leu Gln Gly Asp Ile Asn Cys Phe Ala Ser Leu Asp Pro His His Val
 500 505 510

Glu Leu Lys Glu Glu His Ser His His Glu Leu Arg Leu Glu Glu Leu
 515 520 525

Gln Gln Ser Gly Val Asn Asp Asp Thr Phe Ala His Met Asp Phe Leu

ES 2 682 279 T3

530		535		540	
Ala Asn Asp Glu Ala Pro Val Asp His His Ala Gly His Leu His Thr					
545		550		555	560
Ile Ser Thr His Asp Asp Val His Ser His Ala Asp His Asp Val His					
		565		570	575
Met Arg Asn Tyr Asp Val Phe Trp Lys Asp Thr Val Pro Asp Gly Asp					
		580		585	590
His Gly Leu Leu Leu Asp Asp Phe Asp Gly Gly Ile Glu Phe					
		595		600	605

<210> 7
 <211> 2241
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> ortólogo LAR1

<400> 7

atgtcctccc ccaagaacat tctggccccc gcaagcttgt ccctaacaac ttacaataaa	60
cccagccacg atctggcag cccaaagacg caacaccatc accatggcct gcatcatcag	120
caccagcaca agcagcagta ccagcaccaa cagcaactgc agcacgcca tgttctcggg	180
ggcaaatcag tagccggttc caacaagatc cttcctttca cctcatctat ggatgaagtt	240
aaatatgccg ctggcggggt gatcaagccc gggagcgaag gcctgggcag catgctctct	300
acacccttga cgcctcggc ttctcttatt gcatcgtcgg tggtagaggc cgatgcaaag	360
caacaacaac tgctgaagga ttccctcacc gccgacctca aattgctcct gcaogaattc	420
gaacgcttcc agcaggccac tgcagcagca gccggaacag gaggcgtggg cgaggaggag	480
gcagctgagc gatgcagcaa agtggagttc tttttgggat acattgagcg agttctccac	540
gatttggcag gggctgacgc atcgaaactg caggatctcg aggtgcggat caagacaagt	600
ctactaccac tcaaggcca ggtggttagc cagcttgcgg cgcaaaataa taactcccct	660
cctcccata aggagcagca gtcactctgg ttccatcctt cttccacctg ctctccctc	720
tcctctctt ctccgtctc cagcgtgcac acaactcctc ctggatctcc cttggcaagg	780
gaggagacgg tgatgagcac ctacggtccc tttaccact ccgcgctgc agcagcagac	840
gcagtcttcc tctctcttc ctctgctgct tctcggtga tgccgcctgt gagtttccgg	900
agggatcaga gtgacataag tgggattacg togtgctogt catcatogtc gtcatgcggg	960
ggggaaggag gtgcggggca tatggaagat ttagatttgg agtgttttag cttgataatg	1020
gacgaggctg ccgctactgc tccttttact actgcaggta atggtgggaa ggatctgccg	1080

ES 2 682 279 T3

```

gccaatggct catgtgttga tgatgacatg acggatgctg gctctctaac ttcagaggaa      1140
agcagcgtgt ttgtgtcctc cccccgggag tctcctctt cgctcagcac tgtagtacg      1200
ggcttgacc caacaagcag caacagtga aataagcgga ccttgcoctt agaatttccc      1260
agcagcagca gcagcagcag cagcagcagt ctttctctgg ccacagcgtc ctcogtatct      1320
acagacactc tgcagcagcc gctcaagcgc gccgcagcg tcctcctacc caccagcagc      1380
agcagcagca gcagttgtgc tacgcatgca cctcctgcc tcgcctcctc ctccaactcc      1440
tctcttctt ctttctctc cctttctct tctctctgg tggcgacagt ggcagcagca      1500
gatgtcagca agcctctct togtcaggtt gagtatcagt gcggtgcttg gcocgacacc      1560
tacaccgctg cctcctcct caatccctgg tgggcctcg agcgacagga gtgtcccaag      1620
tgcaagaagg ttcaagttcc acgcattgac ataaacctgc ctgccaacac catggaatac      1680
caccggctt tgctcgcgga ggaaggcgt gatgacgatg atgatgaggt gggaggcggg      1740
agggaggag ggatgatgat gatgccgggg ggaggggatg gacatggaca tttggaggaa      1800
agagaggagg gagagacgag tgagaaagg agcggtgaa gtagcgtact agaggaggac      1860
gaggaagcag tgttgagccc tatgcaagct tcacagctgt tgagtttgtt ggagcatgcg      1920
cggacatgcc cgggcaatca tgctgccgag aaacaccagg ctgtgtgtac gagtgccaaag      1980
tatctcatgt tgcattgtg ggattgtgat ggcaggacgt tagatgggga ggcttgtgga      2040
ttctcgtgt gcaggccttg caaacacttg ctcgggcacc tggccggtg ttacgaagcc      2100
gagaagtgtc agatctgctg tttctcacac caagaagagg aggagaaagt ggaaaagaag      2160
gtgatgatga gtgtggagga gatgattgag gagaagagg tgagggtcga cacctacagg      2220
agcttgacga gcttgagctg a                                          2241

```

```

<210> 8
<211> 746
<212> PRT
<213> Nannochloropsis oceanica

```

```
<400> 8
```

```
Met Ser Ser Pro Lys Asn Ile Leu Ala Pro Ala Ser Leu Ser Leu Asn
1          5          10          15
```

```
Asn Tyr Asn Lys Pro Ser His Asp Leu Gly Ser Pro Lys Thr Gln His
          20          25          30
```

```
His His His Gly Leu His His Gln His Gln His Lys Gln Gln Tyr Gln
          35          40          45
```

```
His Gln Gln Gln Leu Gln His Ala His Val Leu Gly Gly Lys Ser Val
          50          55          60
```

ES 2 682 279 T3

Ala Gly Ser Asn Lys Ile Leu Pro Phe Thr Ser Ser Met Asp Glu Val
65 70 75 80

Lys Tyr Ala Ala Gly Gly Leu Ile Lys Pro Gly Thr Gln Gly Leu Gly
85 90 95

Ser Met Leu Ser Thr Pro Leu Thr Pro Ser Ala Ser Leu Ile Ala Ser
100 105 110

Ser Val Val Glu Ala Asp Ala Lys Gln Gln Gln Leu Leu Lys Asp Ser
115 120 125

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

Gln Ala Thr Ala Ala Ala Ala Gly Thr Gly Gly Val Gly Glu Glu Glu
145 150 155 160

Ala Ala Glu Arg Ser Thr Lys Val Glu Phe Phe Leu Gly Tyr Ile Glu
165 170 175

Arg Val Leu His Asp Leu Ala Gly Ala Asp Ala Ser Lys Leu Gln Asp
180 185 190

Leu Glu Val Arg Ile Lys Thr Ser Leu Leu Pro Leu Lys Gly Gln Val
195 200 205

Val Ser Gln Leu Ala Ala Gln Asn Asn Asn Ser Pro Pro Pro His Lys
210 215 220

Glu Gln Gln Ser Ser Trp Phe His Pro Ser Ser Thr Cys Ser Ser Leu
225 230 235 240

Ser Ser Ser Ser Ser Val Ser Ser Val His Thr Thr Pro Pro Gly Ser
245 250 255

Pro Leu Ala Arg Glu Glu Thr Val Met Ser Thr Tyr Gly Pro Phe Thr
260 265 270

His Ser Arg Val Ala Ala Ala Asp Ala Val Phe Leu Ser Ser Ser Ser
275 280 285

Ala Ala Ser Arg Leu Met Pro Pro Val Ser Phe Arg Arg Asp Gln Ser
290 295 300

Asp Ile Ser Gly Ile Thr Ser Cys Ser Ser Ser Ser Ser Ser Cys Gly

ES 2 682 279 T3

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

ES 2 682 279 T3

His Pro Ala Leu Leu Ala Glu Glu Gly Asp Asp Asp Asp Asp Asp Glu
 565 570 575

Val Gly Gly Gly Arg Glu Gly Gly Met Met Met Met Pro Gly Gly Gly
 580 585 590

Asp Gly His Gly His Leu Glu Glu Arg Glu Glu Gly Glu Thr Ser Glu
 595 600 605

Lys Gly Ser Gly Gly Ser Ser Val Leu Glu Glu Asp Glu Glu Ala Val
 610 615 620

Leu Ser Pro Met Gln Ala Ser Gln Leu Leu Ser Leu Leu Glu His Ala
 625 630 635 640

Arg Thr Cys Pro Gly Asn His Ala Ala Glu Lys His Gln Ala Val Cys
 645 650 655

Thr Ser Ala Lys Tyr Leu Met Leu His Val Arg Asp Cys Asp Gly Arg
 660 665 670

Thr Leu Asp Gly Glu Ala Cys Gly Phe Ser Trp Cys Arg Pro Cys Lys
 675 680 685

His Leu Leu Gly His Leu Val Arg Cys Tyr Glu Ala Glu Lys Cys Gln
 690 695 700

Ile Cys Cys Phe Ser His Gln Glu Glu Glu Glu Lys Val Glu Lys Lys
 705 710 715 720

Val Met Met Ser Val Glu Glu Met Ile Glu Glu Lys Gly Met Arg Val
 725 730 735

Asp Thr Tyr Arg Ser Leu Thr Ser Leu Ser
 740 745

<210> 9
 <211> 77
 <212> PRT
 <213> Nannochloropsis gaditana

<400> 9

Ala Ser Gln Ile Leu Glu Leu Met Ser His Ala Arg Thr Cys Pro Gly
 1 5 10 15

His His His Ser Glu Ala His Arg Ala Val Cys Thr Ser Thr Lys Tyr
 20 25 30

Leu Met Leu His Val Arg Asp Cys Asp Gly Lys Thr Leu Asp Gly Glu
 35 40 45

Ala Cys Gly Phe Ser Trp Cys Arg Pro Cys Lys His Leu Leu Gly His
 50 55 60

Leu Val Arg Cys Tyr Glu Ser Glu Gln Cys Ser Ile Cys
 65 70 75

<210> 10
 <211> 77
 <212> PRT
 <213> *Nannochloropsis oceanica*
 <400> 10

Ala Ser Gln Leu Leu Ser Leu Leu Glu His Ala Arg Thr Cys Pro Gly
 1 5 10 15

Asn His Ala Ala Glu Lys His Gln Ala Val Cys Thr Ser Ala Lys Tyr
 20 25 30

Leu Met Leu His Val Arg Asp Cys Asp Gly Arg Thr Leu Asp Gly Glu
 35 40 45

Ala Cys Gly Phe Ser Trp Cys Arg Pro Cys Lys His Leu Leu Gly His
 50 55 60

Leu Val Arg Cys Tyr Glu Ala Glu Lys Cys Gln Ile Cys
 65 70 75

<210> 11
 <211> 671
 <212> PRT
 <213> *Nannochloropsis* CCMP 1179
 <400> 11

Met Asp Glu Val Lys Tyr Ala Ala Gly Gly Leu Ile Lys Pro Gly Thr
 1 5 10 15

Gln Gly Leu Gly Ser Met Leu Ser Thr Pro Leu Thr Pro Ser Ala Ser
 20 25 30

Leu Ile Ala Ser Ser Val Val Glu Ala Asp Ala Lys Gln Gln Gln Leu
 35 40 45

Leu Lys Asp Ser Leu Thr Ala Asp Leu Lys Leu Leu Leu His Glu Phe
 50 55 60

Glu Arg Phe Gln Gln Ala Thr Ala Ala Ala Ala Gly Thr Gly Gly Val

ES 2 682 279 T3

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

ES 2 682 279 T3

Thr Val Ser Thr Gly Leu Asp Pro Thr Ser Ser Asn Ser Gly Asn Lys
 325 330 335

Arg Thr Leu Pro Leu Glu Phe Pro Ser Ser Ser Ser Ser Ser Ser Ser
 340 345 350

Ser Ser Ser Leu Ser Leu Ala Thr Ala Ser Ser Val Ser Thr Asp Thr
 355 360 365

Leu Gln Gln Pro Leu Lys Arg Ala Arg Ser Val Ile Leu Pro Thr Ser
 370 375 380

Ser Ser Ser Ser Ser Ser Cys Ala Thr His Ala Pro Pro Cys Leu Ala
 385 390 395 400

Ser Ser Ser Thr Ser Ser Ser Ser Ser Phe Ser Ser Pro Phe Ser Ser
 405 410 415

Ser Ser Val Ala Thr Val Ala Ala Ala Asp Val Ser Lys Pro Leu Leu
 420 425 430

Arg Gln Val Glu Tyr Gln Cys Gly Ala Cys Ala Asp Thr Tyr Thr Ala
 435 440 445

Ala Ser Ser Leu Asn Pro Trp Trp Ala Leu Glu Arg Gln Glu Cys Pro
 450 455 460

Lys Cys Lys Lys Val Gln Val Pro Arg Ile Asp Ile Asn Leu Pro Ala
 465 470 475 480

Asn Thr Met Glu Tyr His Pro Ala Leu Leu Ala Glu Glu Gly Asp Asp
 485 490 495

Asp Asp Asp Asp Glu Val Gly Gly Gly Arg Glu Gly Gly Met Met Met
 500 505 510

Met Pro Gly Gly Gly Asp Gly His Gly His Val Glu Glu Arg Glu Glu
 515 520 525

Gly Glu Thr Ser Glu Lys Gly Ser Gly Gly Ser Ser Val Leu Glu Glu
 530 535 540

Asp Glu Glu Ala Val Leu Ser Pro Met Gln Ala Ser Gln Leu Leu Ser
 545 550 555 560

Leu Leu Glu His Ala Arg Thr Cys Pro Gly Asn His Ala Ala Glu Lys
 565 570 575

His Gln Ala Val Cys Thr Ser Ala Lys Tyr Leu Met Leu His Val Arg
 580 585 590

Asp Cys Asp Gly Arg Thr Leu Asp Gly Glu Ala Cys Gly Phe Ser Trp
 595 600 605

Cys Arg Pro Cys Lys His Leu Leu Gly His Leu Val Arg Cys Tyr Glu
 610 615 620

Ala Glu Lys Cys Gln Ile Cys Cys Phe Ser His Gln Glu Glu Glu Glu
 625 630 635 640

Lys Val Glu Lys Lys Val Met Met Ser Val Glu Glu Met Ile Glu Glu
 645 650 655

Lys Gly Met Arg Val Asp Thr Tyr Arg Ser Leu Thr Ser Leu Ser
 660 665 670

<210> 12
 <211> 787
 <212> PRT
 <213> Ectocarpus siliculosus

<400> 12

Met Leu Ala Gln Leu Ala Gln Gln Gln Gln Gln Ser Ser Pro Ser Thr
 1 5 10 15

Ala Ala Thr Ala Thr Pro Ser Thr Glu Thr Ile Gly Thr Asp Ala Gly
 20 25 30

Gly Asn Thr Pro Pro Pro Pro Ser Gly Ser Leu Phe Leu Gly Gly Val
 35 40 45

Gly Gly Gly Lys Asp Gly Ala Thr Ala Val Ser Asp Gly Trp Val Val
 50 55 60

Gly Ser Pro Ser Pro Thr Leu Pro Pro Ser Ser Ser Asn Glu Ser Gly
 65 70 75 80

Gly Gly Gly Gly Gly Gly Gly Arg Val Asn Ile Ser Thr Cys Asp Ser
 85 90 95

Gly Gly Gln Leu Ser Gly Thr Thr Ser Ser Ala Ser Leu Pro Arg Ala
 100 105 110

Cys Ser Ser Leu Ser Phe Ala Asp Leu Gly Asp Leu Cys Gly Asp Gln
 115 120 125

Glu Gly Gly Asp Val Phe Leu Val Gly Gly Glu Asp Cys Asp Asp Asp
 130 135 140

Ala His Gly Ser Gly Ile Gly Leu Asp Leu Tyr Asp Leu Gly Gln Gly
 145 150 155 160

Gly Lys Asp Pro Phe Leu Pro Glu Gly Gln Gln Arg Ala Gly Arg Trp
 165 170 175

Ser Ser Ile Ser Thr Arg Ser Ser Ser Ser Thr Ala Ser Thr Asp Gly
 180 185 190

Glu Glu Leu Ser Asp Asp Ser Leu Ser Val Asp Ala Asp Cys Ser Gly
 195 200 205

Ser Ser Ser Thr Pro Cys Ser Pro Ser Ala Val Ser Ser Gly Glu Pro
 210 215 220

Ser Leu Leu Ala Ala Ala Ala Ala Ala Thr Arg Ala Ala Lys Ala Ile
 225 230 235 240

Lys Thr Glu Ala Arg Pro Asp Gly Ser Ala Gly Leu Glu Ala Gly Ala
 245 250 255

Val Gln Gln Gly Val Ala Ala Gly Gly Ala Ser Ala Thr Lys Ser Gly
 260 265 270

Ile Asp Gln Glu Met Gly Glu Leu Ser Glu Leu Phe Ala Pro Asp Ala
 275 280 285

Phe Leu Met Ser Thr Val Leu Asp Thr Glu Met Glu Ala Gly Gly Gly
 290 295 300

Gly Ala Ala Arg Ala Gly Ala Ser Gly Ile Glu Val Ser Ile Glu Pro
 305 310 315 320

Thr Ala Thr Glu Pro Ser Lys Ala Ser Gln His Ala Ala Ala Val
 325 330 335

Gly Pro Arg Pro Thr Thr Ala Ala Gly Ala Ser Ala Val Ala Val Thr
 340 345 350

Ala Pro Ala Ala Val Val Lys Met Glu Phe Leu Pro Ala Ala Gly Ala
 355 360 365

Ser Ala Pro Ser Pro Ala Ser Pro Pro Ala Pro Thr Ala Ala Ala Ala
 370 375 380

ES 2 682 279 T3

Ala Ser Val Leu Pro Ala Pro Ala Ala Val Ala Pro Lys Gln Glu Cys
 385 390 395 400

Arg Cys Gly Gln Ala Ala Cys Pro Ser Leu Ile Thr Ala Ala Arg Lys
 405 410 415

Arg Pro Val Ala Glu Leu Ser Pro Ser Leu Ser Gly Ala Asp Gly Pro
 420 425 430

Ala Pro Leu Ser Met Val Thr Gly Leu Pro Phe His Gln Ser Ser Arg
 435 440 445

Gln Arg Lys Arg Gln Arg Ser Leu Ala Pro Val Leu Ser Ala Pro Arg
 450 455 460

Thr Val Ser Tyr Glu Cys Ser Leu Cys Lys Glu Ser Tyr Pro Ser Glu
 465 470 475 480

Ile Ala Ser Asn Pro Trp Trp Ser Leu Phe Leu His Glu Cys Pro Arg
 485 490 495

Cys His Arg Met Gln Ile Pro Arg Val Asp Ala Thr Ser Ala Ala Val
 500 505 510

Ser Val Asp Tyr Ile His Ala Val Cys Ala Glu Glu Gly Glu Gly Cys
 515 520 525

Asp Ser Asp Gly Tyr Gly Ser Glu Ser Cys Ser Asp Ser Asp Asp Asp
 530 535 540

Val Thr Asp Asp Gly Arg Glu Arg Glu Gly Ile Ala Ala Phe Asp Thr
 545 550 555 560

Asp Ile Ile Ala Gly Asp Ser Gln Ala Gly Cys Lys Glu Gly Arg Leu
 565 570 575

Ser Thr Phe Gln Ala Ser Arg Leu Leu Val Leu Met Ser His Ala Arg
 580 585 590

Thr Cys Pro Gly His His Ala Asn Pro Lys His Ala Glu Val Cys Arg
 595 600 605

Ser Thr Lys Phe Leu Met Leu His Met Arg Asp Cys Thr Gly His Thr
 610 615 620

Ala Asn Gly Asp Pro Cys Glu His Arg Trp Cys Arg Pro Cys Lys Ser

ES 2 682 279 T3

```

625                630                635                640

Leu Leu Ser His Leu Val Arg Cys Pro Asp Pro Asn Thr Cys Arg Ile
        645                650                655

Cys Thr Pro Leu Asp Leu Pro Gly Pro Leu Arg Gln Leu Arg Asp Leu
        660                665                670

Asn Val Ala Gln Ala Arg His Ala Ser Ala Ala Ala Ala Ala Ala Ala
        675                680                685

Ala Ala Thr Thr Ser Thr Thr Ala Pro Ser Ser Ala Ala Val Ala Val
        690                695                700

Pro Gly Val Gly Gly Val Pro Ser Ser Leu Ala Ala Ala Ala Ala Pro
705                710                715                720

Ala Ala Ala Ala Arg Val Pro Ser Met Pro Ala Pro Ala Thr Ala Val
        725                730                735

Asp Ala Ala Thr Val Lys Ser Glu Glu Met Cys Arg Pro Ala Gly Ala
        740                745                750

Ala Thr Ala Met Gly Leu Ala Val Ser Ala Thr Thr Thr Thr Ser Leu
        755                760                765

Arg Gln Gln Leu Gln Gln Arg Val Val Gly Gly Ala Gly Gly Val Val
770                775                780

Thr Thr Arg
785

<210> 13
<211> 1135
<212> PRT
<213> Thalassiosira pseudonana

<400> 13

Met Asp Gly Ser Ser Asp Asp Asn Arg Lys Arg Pro His Pro Ala Ser
1                5                10                15

Asn Ala Asn Asp Val Ala Ser Ser Glu Val Ala Ala Gly Asp Asn Lys
        20                25                30

Arg Val Lys Ile Glu Asp Val Gly Gln Ala Ala Ala Ala Val Glu Thr
        35                40                45

Gln Asp Val Pro Gln Thr Gln Ala Ser Val Asp Ala Val Ala Ser Ser

```

ES 2 682 279 T3

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

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

ES 2 682 279 T3

Gly Ala Asp Ala Ala Lys Ala Val Val Val Ala Glu Glu Glu Arg Lys
 565 570 575

Arg Leu Lys Tyr Leu Glu Glu Ser Ala Ala Arg Val Ala Gly Val Ala
 580 585 590

Pro Ala Gly Gly Gly Ala Leu Asp Arg Lys Pro Ala Ser Arg Pro Thr
 595 600 605

Ala Ile Glu Ala Ala Ala Pro Lys Pro Pro Glu Gly Pro Ala Thr Met
 610 615 620

Ala Ala Arg Ala Arg Ala Ile Ala Leu Ala Ala Thr Asn Asn Asn Thr
 625 630 635 640

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

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

Pro Leu Ala Ala Ala Thr Ala Ala Asn Gln Met Ala His Leu Gly Met
 675 680 685

Lys Pro Val Lys Pro Lys Lys Pro His Leu Ala Pro Asp Phe Asn Asp
 690 695 700

Pro Ala Leu Thr Ala Thr Gln Gln Asn Glu Leu Arg Leu Lys Glu Ala
 705 710 715 720

Arg Trp Arg Gln Arg Lys Arg Arg Lys Glu Arg Arg Arg Lys Arg Ser
 725 730 735

Gly Val Val Val Asp His Ala Ala Met Val Thr Gln His Ala Gln Ser
 740 745 750

Met Gln Glu Pro Asn Val His Ala Ser Ser Ser Asn Asp Ala Val Pro
 755 760 765

Ala Gln Pro Met Val Leu Arg Val Asn Lys Asn Gly Ala Tyr Gly Pro
 770 775 780

Arg Thr Val Glu Tyr Val Cys Ala Val Cys Asn Glu Gly Tyr Val Ser
 785 790 795 800

Thr Cys Glu Met Asn Pro Trp Trp Ala Leu Ile Asn His Glu Cys Pro
 805 810 815

Lys Cys Gly Lys Asn Gln Ile Pro Arg Leu Asp Ile Ser Ala Pro Asn
 820 825 830

Asn Val Ile Glu Tyr His Pro Ala Leu Leu Val Gln Glu Asp Gly Lys
 835 840 845

Pro Val Ser Ala Pro Val Ser Asn Gly Gly Asp Ser Ser Ser Val Gln
 850 855 860

Met Gln Tyr Leu Pro Arg His Leu Ala Lys Lys Ser Ser Leu Ser Asp
 865 870 875 880

Ser Glu Val Ser Gln Thr Asp Glu Ser Asp Gly Glu Gly Gly Thr Glu
 885 890 895

Glu Tyr Phe Asp Glu Ser Ser Asp Asp Glu Glu Ser Val Asn Lys Asn
 900 905 910

Ala Leu Asp Ser Phe Ala Lys Glu Glu Arg Ala Glu Arg Glu Asp Tyr
 915 920 925

Gly Phe Glu Phe Lys Gly Glu Thr Leu Ser Asp Asp Gln Ala Lys Arg
 930 935 940

Leu Leu Ile Leu Ile Glu His Ala Ser Ile Cys Pro Gly Arg His Arg
 945 950 955 960

Ser Ala Lys His Arg Asn Val Cys His Ser Thr Lys Tyr Leu Met Leu
 965 970 975

His Val Arg Asp Cys Pro Gly Leu Leu Ser Asn Gly Asp Val Cys Pro
 980 985 990

Phe Pro Trp Cys Arg Lys Thr Lys His Leu Leu Tyr His Leu Val Ser
 995 1000 1005

Cys Glu Lys Ser Asn Asp Gly Lys Glu Cys Gly Ile Cys Cys Pro
 1010 1015 1020

Lys Asn Leu Ser Ser Asn Leu Ser Glu Leu Val Gly Leu Asn Lys
 1025 1030 1035

His Arg Arg Lys Gln Phe Val Asp Arg Thr Lys Ala Ile Val Ala
 1040 1045 1050

Ala Ala Lys Arg Gln Gln Leu Ala Ala Ala Arg Ala Lys Ala Val

ES 2 682 279 T3

1055 1060 1065

Ala Pro Arg Ala Ala Val Gln His Gln Tyr Arg Gly Pro Val Val
 1070 1075 1080

Arg Lys Gly Pro Ile Pro Ala Ala Thr Thr Tyr Ala Ala Pro Pro
 1085 1090 1095

Pro Ser Ala Ser Thr Val Ser Tyr Ala Thr Thr Ser Arg Gly Thr
 1100 1105 1110

His Met Pro Ser Thr Asn Asn Pro Ile Ile Gln Ser Pro Pro Asp
 1115 1120 1125

Ala Lys Thr Ser Asn Gln Phe
 1130 1135

<210> 14
 <211> 1148
 <212> PRT
 <213> *Phaeodactylum tricornutum*

<400> 14

Val Gly Arg Leu Ala Leu Thr Leu Arg Arg Arg Leu Ala Gly Ile Ser
 1 5 10 15

Ala Ala Glu Pro Leu Ser Ser Val Thr Thr Val Val Phe Gly Leu Ala
 20 25 30

Lys Thr Ala Gln Gly Asp Ala Arg Ser Phe Leu Phe Trp Ala Gly Lys
 35 40 45

Ala Ala Glu Ile Val Lys Lys Ala Pro Cys Asp Thr Glu Asn Glu Glu
 50 55 60

Thr Thr Val Gly Ala Lys Thr Glu Arg Ile Ala Gln His Arg Thr Asp
 65 70 75 80

Arg Arg Gln Arg His Val Leu Met Val Thr Met Ala Ile Pro Thr Asn
 85 90 95

Val Cys Leu Leu Leu Ser Glu Ile Phe Arg Phe Leu Met Met Glu Met
 100 105 110

Glu Glu His Asp Leu Ala Gln Gln Lys Asp Pro Gly Lys Val Ser Ser
 115 120 125

Cys Ser Asp Ser Lys Gly Glu Gly His Ala Leu Glu Val Gln Leu Val

Ser Val Ser Ser Ser Asp Ser Leu Arg Asp Leu Ser Ala His Arg Pro
 385 390 395 400

Gln His Pro Gln Asn Thr Thr Arg Leu Pro Val Ser Ser Ser Thr Thr
 405 410 415

Thr Val Thr Ser Gly Ser Asn Ser Pro Leu Ser Ala Gly Pro Val Ser
 420 425 430

Ala Gln Ala Pro Pro Ser Pro Leu Leu Pro Leu Lys Ala Thr Lys Met
 435 440 445

Ser His Leu Arg Gln Lys Tyr Met Gln Glu Leu Glu Tyr Met Leu Cys
 450 455 460

Glu Phe Gln Lys Leu Glu Arg Gln Leu Leu Gly Ala Lys Ala Thr Thr
 465 470 475 480

Ala Glu Ser Ala Gly Ser Arg Glu Arg Arg Glu Lys Leu His Ser Phe
 485 490 495

Ile Thr His Leu Ser Asp Thr Ile Gln Asn Ile Gln Thr Gly Cys Gln
 500 505 510

Leu Glu Ser Glu Gly Lys Ser Thr Val Gly Glu Ala Ser Lys Gln Asp
 515 520 525

Ile Ala Gln Glu Ala Ala Leu Ala Asp Leu Thr Cys Glu Lys Gly Glu
 530 535 540

Glu Glu Asn Val Gln Lys Leu Glu Glu His Ile Leu Ala Asn Leu Leu
 545 550 555 560

Pro Val Lys Val Arg Leu Lys Lys Gln Leu Ala Ala Gln Gln Gly Ala
 565 570 575

Lys His Asn Pro Ala Gly Met Pro Val Ala Gln Arg Gly Leu Val Ala
 580 585 590

Pro Ser Glu Gly Gly Lys Gly Thr Phe Ala Ala Ala Ala Glu Glu Arg
 595 600 605

Arg Lys Gln Leu Ala Asp Ala Ala Ala Ala Ala Gln Gly Phe Asp His
 610 615 620

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

Lys Pro Leu Gln Gly Asn Gly Ser Ser Leu Thr Arg Asn Leu His Gly
 645 650 655
 Ser Thr Leu Gly Ser Ala Ile Lys Val Gly Thr Asp Lys Ser Lys Ile
 660 665 670
 Leu Phe Ala Gly Leu Ala Ile Gly Ser Ser Gln Val Lys Ser Ser Val
 675 680 685
 Asn Ala Ala Ser Ser Val His Gln Leu Val Ile Lys Asp Pro Ala Leu
 690 695 700
 Leu Glu Leu Ala Arg Gln Gln Ser Ala Ser Lys Gln Gln Glu Asp Leu
 705 710 715 720
 Pro Pro Gln Thr Gln Gln Glu Asp Ser Pro Thr Gln Ser Lys Pro Asn
 725 730 735
 Ser Leu Leu Pro Pro Ser Ser Ser Glu Pro Asn Asp Ser Pro Glu Asp
 740 745 750
 Thr Asn Arg Lys Ala Ile Ser Leu Lys Val Ser Pro Ala Val Ala Ser
 755 760 765
 Ala Ala Ala Leu Ala Ala Ser Glu Gln Pro Asp Ala Val Leu Ser Lys
 770 775 780
 Ala Pro Pro Ser Arg Leu Asp Asp Val Asp Ala Thr Tyr Pro Asp Met
 785 790 795 800
 Pro Ser Ala Ala Leu Thr Asp Glu Glu Arg Arg Thr Leu Arg Arg Leu
 805 810 815
 Lys Arg Arg Lys Lys Arg Arg Lys Arg Lys Ala Glu Ala Thr Pro Val
 820 825 830
 Thr Ala Ala Ala Thr Ala Ala Pro Val Ile Asn Arg His His Lys Pro
 835 840 845
 Thr Thr Lys Lys Arg Gly Pro Arg Thr Val Glu Tyr Met Cys Ala Leu
 850 855 860
 Cys Asn Glu Val Tyr Asn Ser Thr Cys Asp Tyr Asn Pro Trp Trp Ala
 865 870 875 880
 Leu Ala Gln His Asp Cys Pro Lys Cys Arg Lys Asn Gln Ile Pro Arg
 885 890 895

Val Asp Ile Ser Ala Pro Ala Asn Thr Ile Glu Tyr His Pro Ala Leu
 900 905 910

Leu Ala His Ala Asp Glu Asn Gly Gly Ser Thr Pro Thr Pro Pro Ala
 915 920 925

Ala Ile Val Lys Pro Val Thr Thr Val Ser Ala Pro Val Thr Ser Val
 930 935 940

Pro Lys Cys Gly Asn Asp Ser Asp Ser Phe Gly Ser Asp Leu Ser Asp
 945 950 955 960

Asp Asp Leu Asp Gly Leu Leu Ser Asp Thr Asp Ser Glu Gly Ser Gly
 965 970 975

Glu Ile Gly Met Glu Arg Ile Asp Ala Leu Ser Pro Ala Glu Gln Ala
 980 985 990

Glu Asn Glu Tyr Phe Gly Val Glu Tyr Lys Gly Pro Lys Leu Lys Asp
 995 1000 1005

Ser Glu Ala Ala Arg Leu Leu Ile Leu Met Gly His Ala Ser Thr
 1010 1015 1020

Cys Pro Cys Lys His Gln Ser Ile Lys His Arg Glu Thr Cys Arg
 1025 1030 1035

Asn Thr Lys Trp Met Met Leu His Val Arg Asp Cys Pro Gly Thr
 1040 1045 1050

Thr Ser Ser Phe Asp Val Cys Pro Phe Pro Trp Cys Arg Lys Val
 1055 1060 1065

Lys His Leu Leu Tyr His Leu Val Ser Cys Arg Asp Ala Lys His
 1070 1075 1080

Cys Glu Ile Cys Ser Pro Thr Lys Leu Asn Gln Asn Met Ile Leu
 1085 1090 1095

Leu Lys Gly Leu Asn Gln His Arg Phe Met Gln Tyr Arg Glu Arg
 1100 1105 1110

Leu Ile Gly Arg Gly Lys Ala Leu Thr Lys Val Ser Asn Ser Ala
 1115 1120 1125

Pro Lys Asn Thr Pro Ala Gln Ala Gln His Lys Thr Phe Ile Asp

1130 1135 1140

Val Ser Gln Met Leu
1145

<210> 15
<211> 948
<212> PRT
<213> *Cyclotella cryptica*

<400> 15

Met Leu Val Glu Phe Lys Lys Leu Glu Arg Gln Leu Leu Gly Ala Pro
1 5 10 15

Ile His Gln Gln Gln Asn Pro Pro Lys Ala Glu Pro Lys Gly Ser Arg
20 25 30

Glu Arg Arg Glu Lys Leu His Gly Phe Ile Leu His Leu Glu Asp Thr
35 40 45

Ile Arg Gln Val Glu Glu Gly Cys Ala Leu Glu Lys Arg Glu Glu Gly
50 55 60

Ser Ile Asp Gln Val Val His Gly Ala Ser Gly Ser Leu Ala Asn Asn
65 70 75 80

Gln Gln Glu Glu Val Glu Glu Glu Glu Lys Lys Ser Ser Asp Ala Ser
85 90 95

Leu Gln Gln Pro Gln His Thr Ala Gln Thr Asn Thr Ala Ser Asn Leu
100 105 110

Asn Asn Leu Asn Lys Thr Thr Ser Asn Met Glu Gly Pro Pro Lys Lys
115 120 125

Phe Thr Ala Ala Glu Ala Ala Leu Ser Ser Leu Pro Pro Glu Lys Glu
130 135 140

Arg Glu Glu Ser Val Gln Arg Leu Glu Glu His Ile Leu Ala Asn Leu
145 150 155 160

Leu Pro Val Lys Val Arg Leu Thr Lys Gln Leu Ala Ala Gln Lys Gly
165 170 175

Ala Thr Arg Asn Pro Val Thr Ala Pro Val Arg Ala Gly Ala Ala Asn
180 185 190

Thr Val Ala Gly Gly Thr Ile Ala Glu Ala Val Glu Ala Lys Arg Arg

ES 2 682 279 T3

195					200					205					
Ala	Gln	Glu	Glu	Glu	Leu	Leu	Lys	Lys	Gln	Leu	Gln	Gln	Arg	Gln	Ile
210						215					220				
Thr	Thr	Ser	Gln	Tyr	Gly	Lys	Pro	Ile	Gly	Gly	Ala	Gly	Ser	Ser	Leu
225					230					235					240
Thr	Ala	Arg	Leu	His	Gly	Gly	Val	Leu	Gly	Ser	Asn	Ala	Pro	Ala	Ala
				245					250					255	
Gly	Ala	Ser	Gly	Ser	Glu	Thr	Thr	Lys	Arg	Pro	Ile	Leu	Tyr	Ala	Gly
			260					265					270		
Val	Ala	Pro	Gly	Ser	Ser	Gln	Val	Pro	Ser	Thr	Ile	Lys	Thr	Val	Ser
		275					280					285			
Gly	Ala	His	Pro	Gly	Leu	Ile	Gly	Lys	Asp	Ala	Thr	Lys	Ala	Val	Ala
	290					295					300				
Leu	Ala	Glu	Glu	Glu	Arg	Arg	Arg	Leu	Lys	Asn	Leu	Glu	Glu	Asn	Ala
305					310					315					320
Thr	Arg	Val	Ala	Leu	Gly	Val	Ala	Ala	Lys	Pro	Ser	Pro	Ala	Ala	Ser
				325					330					335	
Ala	Leu	Asp	Pro	Arg	Lys	Pro	Ala	Ala	Leu	Pro	Thr	Leu	Asn	Ala	Ser
			340					345					350		
Ala	Leu	Pro	Lys	Gln	Pro	Glu	Gly	Pro	Ala	Thr	Leu	Ala	Ala	Arg	Ala
		355					360					365			
Arg	Ala	Val	Ala	Leu	Ala	Ala	Ala	Ser	Ser	Gly	Gly	Ala	Thr	Ser	Lys
	370					375					380				
Ile	Ser	Arg	Pro	Asn	Gln	Gln	Phe	Pro	Thr	Arg	Ser	Leu	Gln	Gln	His
385					390					395					400
His	Val	Lys	Lys	Gly	Pro	Pro	Val	Met	Ala	Ala	Pro	Ala	Ala	Thr	Met
				405					410					415	
Ala	Ala	Pro	Ala	Ala	Thr	Ile	Ala	Ala	Pro	Arg	Ala	Pro	His	Ile	Asn
			420					425					430		
Gln	Tyr	Thr	Lys	Pro	Ser	Ala	Ala	Val	Pro	Tyr	His	Ser	Val	Pro	Thr
		435					440					445			

ES 2 682 279 T3

Pro Pro Met Thr Ala Lys Ser Lys Lys Pro His Ile Ala Pro Asn Phe
 450 455 460

Asn Asp Pro Ser Leu Thr Pro Glu Gln Arg Phe Glu Leu Arg Leu Lys
 465 470 475 480

Glu Ala Arg Trp Arg Gln Arg Lys Arg Arg Arg Glu Arg Arg Arg Lys
 485 490 495

Arg Leu Glu Gly Tyr Leu His Ala Ala Gly Ala Tyr His Ile Val Pro
 500 505 510

Leu Ala Val Gln Gln Gln Pro Ser Leu Ser Leu Ser Ser Gln Pro His
 515 520 525

Leu Gln Glu Thr Gln Pro Glu Arg Val Ala Ser Thr Val Val Thr Pro
 530 535 540

Asn Ala Pro Ala Pro Pro Pro Pro Val Asn Thr Pro Pro Pro Pro Val
 545 550 555 560

Thr Ser Ala Ser Arg Pro Lys Lys Asn Gly Ala Tyr Gly Pro Arg Thr
 565 570 575

Val Glu Tyr Val Cys Ala Val Cys Asn Glu Thr Tyr Ile Ser Thr Cys
 580 585 590

Glu Phe Asn Pro Trp Trp Ala Leu Thr Ser His Asp Cys Pro Lys Cys
 595 600 605

Gly Lys Pro Gln Ile Pro Lys Leu Asp Ile Ser Thr Pro Ala Asn Glu
 610 615 620

Ile Asp Tyr His Pro Ala Leu Leu Ser Gln Glu Asp Asn Ala Lys Pro
 625 630 635 640

Gln Ser Ser Ser Val Ser Ala Ser Ala Asn Ala Ser Asn Pro Ala Val
 645 650 655

Ala Ala Pro Gln Val Ala Gln Pro Val Gln Tyr Met Pro Lys Pro Pro
 660 665 670

Ala His Met Lys Lys Asn Phe Leu Leu Ser Asp Ser Glu Val Ser Leu
 675 680 685

Thr Asp Glu Ser Asp Gly Glu Gly Gly Gly Gly Lys Tyr Asp Glu Ser
 690 695 700

Ser Glu Glu Glu Asp Thr Ser Tyr Asp Asn Asp Met Asp Ser Val Thr
705 710 715 720

Arg Glu Glu Arg Val Glu Lys Glu Glu Phe Gly Phe Asp Tyr Lys Gly
725 730 735

Glu Val Leu Ser Glu Asp Gln Ala Arg Arg Leu Leu Val Leu Ile Glu
740 745 750

His Ala Ser Ile Cys Pro Gly Arg His Arg Ser Ala Lys His Arg Asn
755 760 765

Val Cys His Ser Thr Lys Tyr Met Met Leu His Val Arg Asp Cys Cys
770 775 780

Gly Leu Leu Ser Asn Gly Asp Val Cys Pro Phe Pro Trp Cys Arg Lys
785 790 795 800

Thr Lys His Leu Leu Tyr His Leu Val Thr Cys Thr Lys Asn Asp Asp
805 810 815

Gly Ser Lys Cys Ser Ile Cys Cys Pro Glu Asn Leu Ser Ser Asn Leu
820 825 830

Met Asp Leu Val Gly Leu Asn Ser Tyr Arg Arg Lys Ile Phe Val Glu
835 840 845

Arg Ala Lys Ala Val Ala Ala Ala Ala Ala Ala Thr Arg His Gln Met
850 855 860

Ala Ile Ala Lys Ala Lys Ala Ala Ala Gln Ser Ser Thr Gln Pro His
865 870 875 880

Val Leu Thr Ser Ser Gln Ile Leu Thr Ala Pro Ser Thr Asn His Asn
885 890 895

Tyr Glu Ser Gln Thr Lys Pro Phe Ala Thr Ala Ser Thr His Val Ala
900 905 910

Thr His Ala Thr Gln Pro Leu Gly Asn Ala Arg Arg Gly Ser Ser Val
915 920 925

Gln Asp Ala Thr Ile Ala Ser Asn Asn Thr Gly Pro His Leu Asp Ser
930 935 940

Glu Val Val Leu
945

ES 2 682 279 T3

<210> 16
 <211> 549
 <212> PRT
 <213> Ectocarpus siliculosus

 <400> 16
 Met Val Met Ala Ala Thr Ala Asp Gly Gly Val Cys Glu Ile Lys Met
 1 5 10 15

 Asp Asp Ser Gly Arg Leu Thr His Ile Trp Trp Gln Thr Arg Glu Gln
 20 25 30

 Ala Arg Gln Arg Ser Ser Asp Asp Asp Ala Pro Thr Lys Gly Gly Val
 35 40 45

 Ala Leu Gly Gly Val Gln Ala Ser Pro Asp Ser Glu Val Ser Arg Arg
 50 55 60

 Leu Asp Leu Leu Met Asp Lys Ser Phe Gly Asp Leu Gln Phe Leu Ser
 65 70 75 80

 Asp Glu Phe Ala Lys Leu Glu Val Val Val Ala Pro Ala Val Gln Glu
 85 90 95

 Arg Asp Asp Ser Gly Lys Ala Gly Thr Asp Ser Lys Leu Gly Arg Leu
 100 105 110

 Arg Phe Phe Thr Thr His Val Arg Arg Thr Met Ala Arg Met Arg Asp
 115 120 125

 Ala Arg Ser Gly Arg Asp Pro Met Ser Met Ser Gln Leu Ala Leu Leu
 130 135 140

 Glu Glu His Ile Ala Thr Ser Ile Ala Pro Gly Gly Gly Gly Gly Val
 145 150 155 160

 Ser Ser Ser Arg Ser Ser Ser Val Ser Ser Ser Met Gly Gly Arg Ala
 165 170 175

 Gly Ala Glu Glu Leu Val Gly Ala Gly Arg Glu Glu Glu Glu Glu Trp
 180 185 190

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

 Val Ser Leu Gly Leu Val Ser Pro Ala Pro Ala Ser Val Pro Asp Arg
 210 215 220

ES 2 682 279 T3

Gly Ala Asp Gly Gly Ile Tyr Asp Pro His Ala Cys Ser Ala Arg Leu
 225 230 235 240

Thr Pro Ala Ser Ser Ser Ser Ser Leu Ser Arg Leu Cys Trp Gly Ser
 245 250 255

Gly Gly Gly Thr Gly Gly Ser Ala Gly Gly Gly Gly Arg Asp Val Ala
 260 265 270

Arg Gly Gly Arg Pro His Arg His Thr Arg Arg Asp His Val Asp Met
 275 280 285

Leu Ser Gln Leu Glu Ala Glu Gly Val Phe Ala Ala Asp Asp Ser Tyr
 290 295 300

Ser Pro Cys Gly Gly Gly Gly Gly Ser Arg Gly Gly Gly Gly Met Gly
 305 310 315 320

Gly Val Ala Phe Ser Pro Glu Pro Arg Glu Val Arg Tyr Gln Cys Gly
 325 330 335

Ala Cys Ala Ala Ser Tyr Ala Ala Thr Val Ser Gly Asn Pro Trp Trp
 340 345 350

Leu Leu Val Arg Gln Glu Cys Pro Ile Cys His Lys Met Gln Ile Pro
 355 360 365

Arg Val Asp Ile Leu Asn Pro Thr Asn Asn Val Glu Ser His Ile Ala
 370 375 380

Phe Leu Thr Glu Asn Ala Ser Asp Gly Asp Gly Ser Cys Met Asp Trp
 385 390 395 400

Asp Gly Glu Thr Ser Asp Glu Asn Ser Gly Asp Glu Tyr Ser Gly Asp
 405 410 415

Glu Arg Gln Gly Leu Ser Ala Gly Gly Ser Met Ser Gly Gly Asp Gly
 420 425 430

Ser Gly Leu Gly Pro Thr Leu Asp Ser Asp Gln Ala Ala Lys Leu Leu
 435 440 445

Val Leu Met Cys His Ala Arg His Cys Pro Gly Asn His Arg Ser Ala
 450 455 460

Arg Leu Ala Glu Val Cys Arg Ser Val Lys Phe Leu Met Leu His Leu

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

ES 2 682 279 T3

Asn Ser Thr Cys Asp Tyr Asn Pro Trp Trp Ala Leu Thr Gln Glu Glu
 385 390 395 400

Cys Pro Lys Cys Gln Lys Thr Gln Ile Pro Arg Ile Asp Ile Gly Ala
 405 410 415

Pro Ala Asn Ala Ile Glu Tyr His Pro Ala Leu Leu Ala His Ala Asp
 420 425 430

Glu Ser Ala Gly Ala Ala Glu Pro Ser Ala Val Leu Glu Pro Gln Asp
 435 440 445

Leu Pro Val Pro Ser Thr Thr Gly Asp Asp Met Glu Tyr Ser Asp Val
 450 455 460

Asp Asp Ser Asp Leu Ser Asp Glu Asp Gly Leu Leu Ser Asp Ala Ser
 465 470 475 480

Leu Asp Leu Asp Ser Asp Asp Ser Glu Ile Ala Asp Ser Glu Asn Met
 485 490 495

Ser Pro Ala Glu Gln Ala Glu Ser Glu Lys Phe Gly Ala Glu Tyr Asp
 500 505 510

Gly Pro Lys Phe Ser Asp Ala Glu Ala Ala Arg Leu Leu Asn Leu Met
 515 520 525

Leu His Ala Ser Thr Cys Pro Cys Arg His Lys Ser Ser Glu His Tyr
 530 535 540

Asp Val Cys Arg Ser Val Lys Trp Met Met Leu His Val Arg Asp Cys
 545 550 555 560

Pro Gly Thr Thr Ser Thr Phe Asp Val Cys Pro Phe Pro Trp Cys Arg
 565 570 575

Lys Ala Lys His Leu Leu Tyr His Leu Leu Ser Cys Glu Asn Pro Gln
 580 585 590

Ser Cys Pro Ile Cys Ser Pro Val His Leu Asn Ala Ser Met Lys Ser
 595 600 605

Leu Arg Gly Leu Asn His Tyr Arg Leu Lys Lys Gln Gln Gln Cys Val
 610 615 620

Ile Gly Ala Ser Gln Ser Pro Gly Lys Pro Ser Ala Ala Arg Gly Gly
 625 630 635 640

ES 2 682 279 T3

His Ser Ser Pro Lys Lys Ser Asn Glu Ser Thr Gln Asp Glu Val Asp
 645 650 655
 Thr Cys Arg Asp Thr Leu Asp Ala Phe Val Val Gly Thr Lys Glu Pro
 660 665 670
 Thr Glu Ile Ser Ser Ser Ser Val Pro Ser Asn Ser Thr Ala Ala Val
 675 680 685
 Asp Gln Pro Ala His Glu Asp Glu Val Lys Thr Gln Leu Tyr Glu Phe
 690 695 700
 Val Asp Gln Trp Glu Asp Pro Glu His Pro Ala Ala Leu Lys Asp Asp
 705 710 715 720
 Gly Leu Asn Asn Ser Gln Leu Thr Val Gly His Cys Asp Asn Asp Val
 725 730 735
 Leu Ile Lys Gln Glu Asp Asp Asn Ser Gln
 740 745
 <210> 18
 <211> 327
 <212> PRT
 <213> *Aureococcus anophagefferens*
 <400> 18
 Met Arg Arg Ala Ala Ala Glu Ala Lys Ala Met Glu Pro Leu Arg Ala
 1 5 10 15
 Leu Ala Phe Ala Gln Ser Val Lys Tyr Tyr Cys Ser Val Cys Glu His
 20 25 30
 Ala Tyr Glu Thr Thr Ser Asp Ala Asn Pro Leu Trp Thr Leu Ala Arg
 35 40 45
 His Ser Cys Pro Gln Cys Gly Ala Leu Gln Tyr Pro Glu Ile Glu Ile
 50 55 60
 Asp Asp Ile Ser Leu Pro Thr Ala Glu Ala Arg Ala Pro Asp Asp Ala
 65 70 75 80
 Ser Thr Leu Ser His Ala Gln Gly Asp Gly Ala Arg Ser Pro Pro Arg
 85 90 95
 Pro Ala Ala Leu Gly Arg Ala Ala Ala Ala Asp Asp Gly Glu Val Leu
 100 105 110

ES 2 682 279 T3

Leu Arg Ala Asp Ala Pro Arg Lys Arg Ser Ala Pro Asp Asp Asp Ala
 115 120 125
 Ser Thr Met Ser His Ala Gln Gly Ala Ala Leu Leu Glu Leu Phe Asp
 130 135 140
 His Val Arg Ser Cys Pro Gly Arg His Gln Ser Ala Ala His Ala Arg
 145 150 155 160
 Val Cys Ala Gly Ala Lys Phe Val Met Leu His Ala Arg Asp Cys Asp
 165 170 175
 Ala Ala Pro Gly Thr Cys Gly Val Glu Trp Cys Gly Ala Val Lys Gly
 180 185 190
 Leu Leu Ser Arg Val Val Cys Gly Gln Gln Gly Asp Lys Cys Val Val
 195 200 205
 Cys Ala Glu Pro Ser Glu Ala Met Asp Val Gly Asp Ala Ser Pro Pro
 210 215 220
 Thr Val Ser Pro Asp Ala Met Asp Val Asp Asp Ala Ser Pro Pro Arg
 225 230 235 240
 Ala Ser Ser Asp Ala Pro Met Ala Asp Ala Pro Pro Ala Glu Pro Arg
 245 250 255
 Val Gly Ala Ala Ala Pro Ser Pro Pro Gln Ser Thr Asp Arg Ala Phe
 260 265 270
 Glu Ala Pro Thr Ala Val Pro Ser Pro Val Arg Ala Ser Pro Ser Arg
 275 280 285
 Gly Gly Ala Pro Ser Ser Arg Arg Leu Ala Asp Ile Thr Met Ser Glu
 290 295 300
 Asn Val Glu Met Asp Phe Pro Arg Asp Phe Cys Phe Ala Asp Glu Leu
 305 310 315 320
 Arg Arg Arg Gly Asp Ala Ala
 325

<210> 19
 <211> 30
 <212> PRT
 <213> Nannochloropsis gaditana
 <400> 19

ES 2 682 279 T3

Ala Lys Gln Gln Gln Leu Leu Lys Asp Ser Leu Thr Ala Asp Leu Lys
 1 5 10 15

Leu Leu Leu His Glu Phe Glu Arg Phe Gln Gln Ala Thr Ala
 20 25 30

<210> 20
 <211> 1797
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> *Nannochloropsis oceanica*

<400> 20
 atgcctatgg tcaccctctc ccaggatgca actactaccg cggccggcag catgatgctt 60
 cccctccttc cctccatccc tgcttccgct accagcgctt cttttaccgc tccctcagcg 120
 tctccacga ccaccaaacc tcccaaaggt acttctctca tgactcagct tggagagaca 180
 gggcgagaga atactggccg atggacttgt gaagagcatg tgctgtttct taaaggccta 240
 gaaatgcacg gcaagggttg gaagaaaatc gcaaagctaa tcaagaccgg aacgggtggtc 300
 caaatccgca cgcacgcgca aaagtacttc cagaaattgg caaaagccaa gaagaacggg 360
 caccatggtg atatgctcgg aatggaaggc tcacactttg ggggaaaacg tgtcaaattt 420
 accggaaaagc gacgtgggct tgtctataat tcgtatttag taggtgccga ggccacctct 480
 ggggctatct cccggcggtt gcagacgttt atgccggcga acttggggat ggaggggcag 540
 cgtgtaggcc ttatgacgga taaggaggag gatgcagcaa tcgagaaggg actttatcgt 600
 ttctctccc ccgtagtgtt ggatcccgcc acgcgtaatc tggacgcctc cgctcctgag 660
 atcttgccct taccaccag cactccagcg atgggcgtgc accataccag tagcagaggt 720
 agcagcagag gggggttggg tggagagaca acgggagagg aggacggcgg aagcgattcg 780
 attgtaatgg gggatggagg gagcgatcaa gatgcagagt cgtcgttggg cgagcccttg 840
 ccaacttttg cgcgggtgac accggagatg tacacacggt gtggagttcc ggaatggttt 900
 aagaaagggg gggacattga cgaattgctc attgatgcag cgggactcga ttggagaagt 960
 gactcgggtg gggacgcacg gaaggtggtg gatcaaggga caagtatattt gaatgogaat 1020
 attaatggtg gtaattgtgc gactgtggcc ccggcagtgg tgcggaaggg ctgtggttagc 1080
 aacactaata agatgaactc agcggcgcct gtgctgaaca tgacaggggt ggctggggct 1140
 ggagggcttt cagggtgga gggcaagggt agcgacacta gcgaaggtag cagcagcaac 1200
 ggcagcagca agaacatggc tttgacggca aatgogtcgg cgggatgtgg acaagggagc 1260
 tggggagtgc gggggggggc acagaaaaag cagcagcagc agcaacatga ggtaccaacg 1320
 cagcagcagc aggtaccaac gcagcagcag caggttcacg gcattcacgt gaaggaggaa 1380

```

gggatggagc tctgagagt catggcggat agagggactg ttcacgggtca cgtccatgag      1440
gaggatggct ttgcccgtt tgacccccat catgtcgagc tgaaggagga gcaactccac      1500
catgatttgt tgctggaaga gctgccccac gacagcaacc acgatgaagc cctggccccat      1560
attgtgttct cagtgaatgg agagtggat cttcattcct tgccgagggg tgcaggaggg      1620
gggggagcgc atgtgcatgc gccgccggtt gtggtggggg ggagacagca tcaactatcat      1680
cataatgaca atgatatcca tttgcatgag taagacgagt atttgaaga ggaagggggc      1740
gatggcgggc atgggttgtt gttgttgag gatttggatg ggggaatcga gttttag      1797

```

```

<210> 21
<211> 598
<212> PRT
<213> Nannochloropsis oceanica

```

```
<400> 21
```

```
Met Pro Met Val Thr Leu Ser Gln Asp Ala Thr Thr Thr Ala Ala Gly
1                               5                               10                               15
```

```
Ser Met Met Leu Pro Leu Leu Pro Ser Ile Pro Ala Ser Ala Thr Ser
                20                               25                               30
```

```
Ala Ser Phe Thr Ala Pro Ser Ala Ser Ser Thr Thr Thr Lys Ser Pro
                35                               40                               45
```

```
Lys Gly Thr Ser Ser Met Thr Gln Leu Gly Glu Thr Gly Arg Glu Asn
                50                               55                               60
```

```
Thr Gly Arg Trp Thr Cys Glu Glu His Val Leu Phe Leu Lys Gly Leu
65                               70                               75                               80
```

```
Glu Met His Gly Lys Gly Trp Lys Lys Ile Ala Lys Leu Ile Lys Thr
                85                               90                               95
```

```
Arg Thr Val Val Gln Ile Arg Thr His Ala Gln Lys Tyr Phe Gln Lys
                100                              105                              110
```

```
Leu Ala Lys Ala Lys Lys Asn Gly His His Gly Asp Met Leu Gly Met
                115                               120                               125
```

```
Glu Gly Ser His Phe Gly Gly Lys Arg Val Lys Phe Thr Gly Lys Arg
                130                               135                               140
```

```
Arg Gly Leu Val Tyr Asn Ser Tyr Leu Val Gly Ala Glu Ala Thr Ser
145                               150                               155                               160
```

ES 2 682 279 T3

Ala Ala Ile Ser Pro Ala Leu Gln Thr Phe Met Pro Ala Asn Leu Gly
 165 170 175

Met Glu Gly Glu Arg Val Gly Leu Met Thr Asp Lys Glu Glu Asp Ala
 180 185 190

Ala Ile Glu Lys Gly Leu Tyr Arg Phe Leu Ser Pro Val Val Leu Asp
 195 200 205

Pro Ala Thr Arg Asn Leu Asp Ala Ser Ala Pro Glu Ile Leu Pro Leu
 210 215 220

Pro Pro Ser Thr Pro Ala Met Gly Val His His Thr Ser Ser Arg Gly
 225 230 235 240

Ser Ser Arg Gly Gly Leu Asp Gly Glu Thr Thr Gly Glu Glu Asp Gly
 245 250 255

Gly Ser Asp Ser Ile Val Met Gly Asp Gly Gly Ser Asp Gln Asp Ala
 260 265 270

Glu Ser Ser Leu Gly Glu Pro Leu Pro Thr Leu Ala Arg Val Thr Pro
 275 280 285

Glu Met Tyr Thr Arg Cys Gly Val Pro Glu Trp Phe Lys Lys Gly Gly
 290 295 300

Asp Ile Asp Glu Leu Leu Ile Asp Ala Ala Gly Leu Asp Trp Arg Ser
 305 310 315 320

Asp Ser Gly Gly Asp Ala Arg Lys Val Val Asp Gln Gly Thr Ser Ile
 325 330 335

Leu Asn Ala Asn Ile Asn Gly Ser Asn Cys Ala Thr Val Ala Pro Ala
 340 345 350

Val Val Arg Lys Gly Cys Gly Ser Asn Thr Asn Lys Met Asn Ser Ala
 355 360 365

Ala Pro Val Leu Asn Met Thr Gly Leu Ala Gly Ala Gly Gly Leu Ser
 370 375 380

Gly Trp Lys Gly Lys Gly Ser Asp Thr Ser Glu Gly Ser Ser Ser Asn
 385 390 395 400

Gly Ser Ser Lys Asn Met Ala Leu Thr Ala Asn Ala Ser Ala Gly Cys
 405 410 415

Gly Gln Gly Ser Trp Gly Val Arg Gly Gly Ala Gln Lys Lys Gln Gln
 420 425 430

Gln Gln Gln His Glu Val Pro Thr Gln Gln Gln Gln Val Pro Thr Gln
 435 440 445

Gln Gln Gln Val His Gly Ile His Val Lys Glu Glu Gly Met Glu Leu
 450 455 460

Leu Arg Val Met Ala Asp Arg Gly Thr Val His Gly His Val His Glu
 465 470 475 480

Glu Asp Gly Phe Ala Ala Phe Asp Pro His His Val Glu Leu Lys Glu
 485 490 495

Glu His Ser His His Asp Leu Leu Leu Glu Glu Leu Pro His Asp Ser
 500 505 510

Asn His Asp Asp Ala Leu Ala His Ile Val Phe Ser Val Asn Gly Glu
 515 520 525

Ser Asp Leu His Ser Leu Pro Arg Gly Ala Gly Gly Gly Ala His
 530 535 540

Val His Ala Pro Pro Val Val Val Gly Gly Arg Gln His His Tyr His
 545 550 555 560

His Asn Asp Asn Asp Ile His Leu His Ala Tyr Asp Ala Tyr Leu Glu
 565 570 575

Glu Glu Gly Ala Asp Gly Gly His Gly Leu Leu Leu Leu Glu Asp Leu
 580 585 590

Asp Gly Gly Ile Glu Phe
 595

- <210> 22
- <211> 281
- <212> PRT
- <213> *Nannochloropsis oceanica*

<400> 22

Ser Ala Pro Ala Ala Ser Pro Thr Thr Ser Lys Leu His Lys Gly Thr
 1 5 10 15

Ser Ser Met Thr Gln Leu Gly Glu Thr Gly Arg Glu Asn Thr Gly Arg
 20 25 30

ES 2 682 279 T3

Trp Thr Cys Glu Glu His Val Leu Phe Leu Lys Gly Leu Glu Met His
 35 40 45

Gly Lys Gly Trp Lys Lys Ile Ala Lys Leu Ile Lys Thr Arg Thr Val
 50 55 60

Val Gln Ile Arg Thr His Ala Gln Lys Tyr Phe Gln Lys Leu Ala Lys
 65 70 75 80

Ala Lys Lys Asn Gly His His Gly Asp Met Leu Gly Met Glu Gly Thr
 85 90 95

Arg Phe Gly Gly Lys Arg Val Lys Phe Thr Gly Lys Arg Arg Gly Leu
 100 105 110

Val Tyr Gly Ser Tyr Leu Val Gly Ala Glu Ala Thr Ser Ala Ala Ile
 115 120 125

Ser Pro Ala Leu Gln Ser Tyr Met Pro Gly Ser Trp Ala Gly Arg Glu
 130 135 140

Glu Gly Glu Ala Leu Ser Asp Lys Glu Glu Asp Ala Ala Ile Glu Lys
 145 150 155 160

Gly Leu Tyr Arg Phe Leu Ser Pro Val Val Leu Asp Ala Ala Ala Ser
 165 170 175

Asn Leu Asp Ala Thr Ala Pro Glu Val Leu Pro Pro Ser Thr Pro Gly
 180 185 190

Thr Gly Val His Ala Asn Gly Val Val Gly Ala Asp Gly Glu Thr Thr
 195 200 205

Glu Glu Asp Gly Ser Ser Gly Gly Asp Asn Val Asp Val Ser Glu Thr
 210 215 220

Ile Asp Asp Ala Asp Ser Ser Ser Gly Glu Pro Leu Pro Arg Leu Ala
 225 230 235 240

Arg Val Thr Asn Asp Met Tyr Glu Arg Cys Ser Val Pro Thr Trp Phe
 245 250 255

Met Lys Gly Gly Asp Ile Glu Glu Leu Leu Ala Asp Ala Ala Ala Ile
 260 265 270

Asp Trp Arg Glu Asp Ser Gly Gly Asp
 275 280

<210> 23
 <211> 290
 <212> PRT
 <213> *Nannochloropsis oceanica*
 <400> 23
 Thr Ala Pro Ser Ala Ser Ser Thr Thr Thr Lys Ser Pro Lys Gly Thr
 1 5 10 15
 Ser Ser Met Thr Gln Leu Gly Glu Thr Gly Arg Glu Asn Thr Gly Arg
 20 25 30
 Trp Thr Cys Glu Glu His Val Leu Phe Leu Lys Gly Leu Glu Met His
 35 40 45
 Gly Lys Gly Trp Lys Lys Ile Ala Lys Leu Ile Lys Thr Arg Thr Val
 50 55 60
 Val Gln Ile Arg Thr His Ala Gln Lys Tyr Phe Gln Lys Leu Ala Lys
 65 70 75 80
 Ala Lys Lys Asn Gly His His Gly Asp Met Leu Gly Met Glu Gly Ser
 85 90 95
 His Phe Gly Gly Lys Arg Val Lys Phe Thr Gly Lys Arg Arg Gly Leu
 100 105 110
 Val Tyr Asn Ser Tyr Leu Val Gly Ala Glu Ala Thr Ser Ala Ala Ile
 115 120 125
 Ser Pro Ala Leu Gln Thr Phe Met Pro Ala Asn Leu Gly Met Glu Gly
 130 135 140
 Glu Arg Val Gly Leu Met Thr Asp Lys Glu Glu Asp Ala Ala Ile Glu
 145 150 155 160
 Lys Gly Leu Tyr Arg Phe Leu Ser Pro Val Val Leu Asp Pro Ala Thr
 165 170 175
 Arg Asn Leu Asp Ala Ser Ala Pro Glu Ile Leu Pro Leu Pro Pro Ser
 180 185 190
 Thr Pro Ala Met Gly Val His His Thr Ser Ser Arg Gly Ser Ser Arg
 195 200 205
 Gly Gly Leu Asp Gly Glu Thr Thr Gly Glu Glu Asp Gly Gly Ser Asp
 210 215 220

ES 2 682 279 T3

Ser Ile Val Met Gly Asp Gly Gly Ser Asp Gln Asp Ala Glu Ser Ser
 225 230 235 240

Leu Gly Glu Pro Leu Pro Thr Leu Ala Arg Val Thr Pro Glu Met Tyr
 245 250 255

Thr Arg Cys Gly Val Pro Glu Trp Phe Lys Lys Gly Gly Asp Ile Asp
 260 265 270

Glu Leu Leu Ile Asp Ala Ala Gly Leu Asp Trp Arg Ser Asp Ser Gly
 275 280 285

Gly Asp
 290

<210> 24
 <211> 6
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Proteína

<400> 24

Ser His Ala Gln Lys Tyr
 1 5

<210> 25
 <211> 6
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Proteína

<400> 25

Thr His Ala Gln Lys Tyr
 1 5

<210> 26
 <211> 598
 <212> PRT
 <213> Nannochloropsis oceanica CCMP1779

<400> 26

Met Pro Met Val Thr Leu Ser Gln Asp Ala Thr Thr Thr Ala Ala Gly
 1 5 10 15

Ser Met Met Leu Pro Leu Leu Pro Ser Ile Pro Ala Ser Ala Thr Ser
 20 25 30

ES 2 682 279 T3

Ala Ser Phe Thr Ala Pro Ser Ala Ser Ser Thr Thr Thr Lys Ser Pro
 35 40 45

Lys Gly Thr Ser Ser Met Thr Gln Leu Gly Glu Thr Gly Arg Glu Asn
 50 55 60

Thr Gly Arg Trp Thr Cys Glu Glu His Val Leu Phe Leu Lys Gly Leu
 65 70 75 80

Glu Met His Gly Lys Gly Trp Lys Lys Ile Ala Lys Leu Ile Lys Thr
 85 90 95

Arg Thr Val Val Gln Ile Arg Thr His Ala Gln Lys Tyr Phe Gln Lys
 100 105 110

Leu Ala Lys Ala Lys Lys Asn Gly His His Gly Asp Met Leu Gly Met
 115 120 125

Glu Gly Ser His Phe Gly Gly Lys Arg Val Lys Phe Thr Gly Lys Arg
 130 135 140

Arg Gly Leu Val Tyr Asn Ser Tyr Leu Val Gly Ala Glu Ala Thr Ser
 145 150 155 160

Ala Ala Ile Ser Pro Ala Leu Gln Thr Phe Met Pro Ala Asn Leu Gly
 165 170 175

Met Glu Gly Glu Arg Val Gly Leu Met Thr Asp Lys Glu Glu Asp Ala
 180 185 190

Ala Ile Glu Lys Gly Leu Tyr Arg Phe Leu Ser Pro Val Val Leu Asp
 195 200 205

Pro Ala Thr Arg Asn Leu Asp Ala Ser Ala Pro Glu Ile Leu Pro Leu
 210 215 220

Pro Pro Ser Thr Pro Ala Met Gly Val His His Thr Ser Ser Arg Gly
 225 230 235 240

Ser Ser Arg Gly Gly Leu Asp Gly Glu Thr Thr Gly Glu Glu Asp Gly
 245 250 255

Gly Ser Asp Ser Ile Val Met Gly Asp Gly Gly Ser Asp Gln Asp Ala
 260 265 270

Glu Ser Ser Leu Gly Glu Pro Leu Pro Thr Leu Ala Arg Val Thr Pro

ES 2 682 279 T3

Ser Asp Leu His Ser Leu Pro Arg Gly Ala Gly Gly Gly Ala His
530 535 540

Val His Ala Pro Pro Val Val Val Gly Gly Arg Gln His His Tyr His
545 550 555 560

His Asn Asp Asn Asp Ile His Leu His Ala Tyr Asp Ala Tyr Leu Glu
565 570 575

Glu Glu Gly Ala Asp Gly Gly His Gly Leu Leu Leu Leu Glu Asp Leu
580 585 590

Asp Gly Gly Ile Glu Phe
595

<210> 27
<211> 676
<212> PRT
<213> Ectocarpus siliculosus

<400> 27

Met Ser Ser Gln Leu Gln Val Asp Asp Gly Ala Ala Ser Ser Ser Asp
1 5 10 15

Asn Leu Asp Asn Glu Gly Val Cys Gln Pro Gly Asn Glu Asn Thr Gly
20 25 30

Arg Trp Thr Ser Asp Glu His Arg Leu Phe Leu Arg Gly Leu Glu Leu
35 40 45

His Gly Lys Gly Trp Lys Gln Ile Ala Thr Leu Ile Gln Thr Arg Thr
50 55 60

Val Val Gln Ile Arg Thr His Ala Gln Lys Tyr Phe Gln Lys Leu Ser
65 70 75 80

Lys Ala Gln Ala Ser Gly Thr Ser His Leu Asp Pro Ala Thr Leu Met
85 90 95

Ser Thr Met Asp Ala Gly Lys Pro Arg Pro Ala Ser Val Ser Arg Asn
100 105 110

Leu Arg Ser Ser Thr Met Ala Asn Ser Pro Ala Glu Ser Glu Gly Arg
115 120 125

Leu Met Ser Leu Arg Lys Arg Arg Asn Arg Gly Arg His Pro Arg His
130 135 140

ES 2 682 279 T3

Asp Asp Asp Gln Asp Tyr Asp Thr Asn Ser Glu Glu Ser Tyr Asp Tyr
 145 150 155 160

Ala Arg Ser Ser Ala Thr Thr Arg Thr Arg Arg Arg Arg Ser Val
 165 170 175

Ser Ser Gly Ser Ser Gly Gly Gly Gly Gly Ser Ser Glu Ser Glu
 180 185 190

Glu Glu Asp Gly Glu Gly Gly Gly Arg Gly His Arg Gly Val Tyr Arg
 195 200 205

Glu Thr Ala Leu Gly Ala Thr Thr Thr Thr Thr Met Met Met Met Val
 210 215 220

Arg Gln Ala Glu Glu Ala Pro Ala Phe Thr Val Gly Gly Gly Asn Ser
 225 230 235 240

Gly Ala His Gly Ala Glu Phe Asp Glu Glu Glu Ala Glu Glu Glu Leu
 245 250 255

Asp Glu Ser Val Asp Val Glu Gly His Asn Ser Asn Asn Lys Asn Val
 260 265 270

Gly Ile Ser Ser Ser Leu Leu His Gly Gly Thr Ala Gly Thr Trp Ser
 275 280 285

Lys Arg Arg Pro Ala Lys Ser His Arg Pro Ser Pro Thr Lys Ala Ser
 290 295 300

Lys Ala Ala Ala Ile Ala Ala Ala Gly Gly Ala Ala Asp Ala Arg Ala
 305 310 315 320

Leu Ala Trp Glu Ala Ala Ala Ala Lys Ala Ala Ala Gly Glu Glu Gly
 325 330 335

Glu Glu Ala Ala Ala Asp Gly Val Val Gly Gly Gly Ser Gly Thr Lys
 340 345 350

Arg His Arg Ser Glu Ser Val Ser Ser Ser Ser Asn Asp Ala Ser Leu
 355 360 365

Gly Lys Thr Ile Lys Ser Leu Lys Arg Thr Gly Gly Ser Ser Gln Thr
 370 375 380

Thr Arg Ile Ser Pro Thr Ser Val Ala Asp Val Asn Ser Phe Met Ser
 385 390 395 400

ES 2 682 279 T3

Phe Pro Val Pro Gln Thr Glu Arg Ser Asp Met Ala Met His His Leu
 405 410 415
 Pro Gln Gln Leu Ala Pro Ser Trp Cys Thr Lys Pro Gln Asp Thr Trp
 420 425 430
 Met Ala Gly Ala Gly Leu Asp Met Gly Gly Leu Glu Ala Asp Asp Ala
 435 440 445
 Gly Pro Phe Arg Trp Phe Ile Asp Glu Arg Ser Leu Ser Ala Pro Gly
 450 455 460
 Ala Phe Phe Gln Pro Pro Val Glu Thr Trp Ser Gly Ser Asp Thr Thr
 465 470 475 480
 Asp Val Ser Thr Ser Ala Gly Ser Asp His Gln His His Arg Thr Val
 485 490 495
 Val Pro Glu His Leu Pro Ser Asn Thr Asn Lys Val Asp Met Ser Pro
 500 505 510
 Gly Gly Val Leu Thr Ser Asp Asn Thr His Asn Gln Glu Asp Val Ser
 515 520 525
 Ile Ala Lys His Arg Gly Met Met Glu Pro Ile Val Pro Ile Ser Asp
 530 535 540
 Ala Ala Met Ala Ser Glu Val Ser Arg His Val Glu Ala Asp Ser Ala
 545 550 555 560
 Thr Ala Cys Ala Ala Ala Gly Cys Ala Gly Gly Ala Ser Ala Ser Ala
 565 570 575
 Gly Gly Ala Gly Asp Gly Gly His Ser Leu Asp Ile Thr Ser Gly Leu
 580 585 590
 Trp Leu Asp His Pro Gln Glu Ala His Glu Ala Val Ala Ser Pro Leu
 595 600 605
 Met Tyr Gly Leu Gly Leu Ser Ala Pro Ala Asn Thr Gly Gly Ala Gly
 610 615 620
 Gly Gly Gly Gly Ala Gly Gly Leu Gly Gly Gly Leu Ser Leu Pro Ser
 625 630 635 640
 Leu Asp Glu Glu Glu Val Leu Gly Phe Leu Ala Cys Ala Glu Glu Asn
 645 650 655

Ser Arg Gln Ala Ala Ala Ala Ala Ala Gly Gly Asp Ala Asp Glu Asn
 660 665 670

Asp Leu Leu Val
 675

<210> 28
 <211> 453
 <212> PRT
 <213> Navicula sp. WT0229

<400> 28

Met Lys Val Glu Ala Thr His Gln His Ala Ser Gly Ala Asp His Ser
 1 5 10 15

Ser Thr Leu Asp Ala Pro Ala Pro Val Leu Ser His Ser Gly Pro Ala
 20 25 30

Ala Ser Asn Ala Ser Asn Asn Ala Pro Lys Ser Lys Lys Lys Lys Gly
 35 40 45

Gln Pro Ala Ile Thr Ile Ala Ala Ala Pro Ser Gln Ala Val Ala Pro
 50 55 60

Leu Ala Pro Gly Glu Asn Thr Gly Arg Trp Thr Ala Glu Glu His Arg
 65 70 75 80

Leu Phe Leu Gln Gly Leu Glu Gln His Gly Lys Gly Trp Lys Lys Ile
 85 90 95

Ala Ser Leu Ile Lys Ser Arg Thr Val Val Gln Ile Arg Thr His Ala
 100 105 110

Gln Lys Tyr Phe Gln Lys Leu Ala Lys Ala Arg Gln Asn Gly Glu Glu
 115 120 125

Gly Asp Ile Thr Met Glu Gly Arg Gly Gly Thr Ala Ser Ile Thr Ser
 130 135 140

Ser Thr Thr Ala Thr Ala Ala Leu Thr Asn Lys Arg Arg Arg His Ile
 145 150 155 160

Thr Gly Thr Lys Arg Lys Val Ile Gln Ser Ile Val Ala Ser Ala Gln
 165 170 175

Arg Gln Ala Lys Lys Ala His Leu Pro Glu Ile Gly Asp Ala Lys Lys
 180 185 190

Ser Pro Val Val Pro Gly Val Ala Pro Ala Leu Ala Tyr Tyr Val Thr
 195 200 205

Pro Ser Gln Ser Ser Ser Val Ser Ala Gly Ser Asp Val Phe Thr Glu
 210 215 220

Gly Asn Leu Ser Gly Pro Val Leu Glu Asp Ser Leu Phe Arg Phe Leu
 225 230 235 240

Thr Pro Val Pro Ile Ala Ser Asp Asp Val Asn Glu Val Ala Arg Gln
 245 250 255

Ala Gly Ala Asn Pro Ile Thr Leu Pro Ser Ser Asn Ser His Ala Leu
 260 265 270

Ser Ser Val Pro Gly Ser Ser Pro Thr Gly Val Gln Glu Leu Ser Ile
 275 280 285

Tyr Pro Ser Trp Thr Asp Ala Lys Asp Pro Pro Ser Trp Tyr Ala Lys
 290 295 300

Gly Ala Asp Val Asp Ala Leu Leu Asp Val Ser Asp Thr Leu Asp Trp
 305 310 315 320

Leu Ala Asp Thr Gly Asp Leu Asp Glu Glu Tyr Gln Pro Gln Ser Asn
 325 330 335

Asp Ile Asp Thr Cys Ser Phe Gly Gly Gln Gly Glu His His Glu Met
 340 345 350

Gly Ile Ser Asn Val His Asn Asn Thr Ser Val Ser Ser Leu Thr His
 355 360 365

Val Asp Pro Asn Met Val Ser Val Val Pro Pro Leu Pro Ser Leu Phe
 370 375 380

Glu Gly Asn Pro Asp Val Val Glu Ala Glu Leu Thr Val Gly Lys Asp
 385 390 395 400

Val Asn Val Ala Ile Gly Asn Ala Ser Leu Leu Ile Ala Pro Ser Asp
 405 410 415

Pro Asn Ala Glu Thr Leu Gln Val Phe Asp Ser Pro Met Glu Glu Gln
 420 425 430

Glu Phe Val Ser Thr Leu Leu Glu Thr Thr Ala Glu Ser Ser Asp Asn

ES 2 682 279 T3

195		200		205											
Arg	Lys	Ala	Ile	Gln	Ser	Val	Val	Ala	Ser	Ala	Gln	Arg	Gln	Gly	Lys
	210						215						220		
Lys	Leu	Ala	Ala	Ala	Lys	Thr	Asn	Pro	Thr	Arg	His	His	Pro	Leu	Pro
	225						230						235		240
Pro	Pro	Leu	Pro	Thr	Val	Ala	Pro	Ala	Leu	Ala	His	Tyr	Thr	Leu	Pro
					245										255
Ser	Thr	Ala	Met	Met	Ala	Lys	Asn	Gly	Thr	Ala	Val	Lys	Glu	Glu	Tyr
					260										270
Val	Ser	Pro	Thr	Asn	Leu	Ser	Gly	Pro	Ala	Leu	Glu	Asp	Ser	Leu	Phe
															285
Arg	Phe	Leu	Thr	Pro	Leu	Pro	Val	Ser	Glu	Pro	Pro	Leu	Asn	Glu	Val
															300
Ala	Arg	Gln	Ala	Gly	Ala	Asn	Pro	Ile	Ser	Leu	Pro	Thr	Asp	Asn	Pro
															320
Ser	Ser	Ile	Pro	Thr	Val	Gly	Ala	Gly	Glu	Ile	Ser	Pro	Thr	Gly	Val
															335
Ser	Asp	Leu	Met	Leu	Tyr	Pro	Ser	Trp	Thr	Asp	Ser	Lys	Glu	Pro	Pro
															350
Ser	Trp	Tyr	Ser	Lys	Gly	Ala	Asp	Ile	Asp	Ala	Leu	Leu	Asp	Met	Gly
															365
Asp	Ser	Leu	Asp	Trp	Leu	Asp	Asp	Thr	Gly	Asp	Leu	Asn	Glu	Ser	Tyr
															380
Val	Pro	Pro	Val	Val	Asp	Thr	Ala	Met	Ala	Ala	Pro	Glu	Pro	His	Thr
															400
Thr	Phe	His	Arg	Tyr	Ser	Asp	Leu	Gly	His	Ser	Lys	Gly	Leu	His	Ser
															415
Thr	Ser	Val	Thr	Ser	Leu	Pro	His	Val	Asp	Ser	Asn	Ala	Asn	Val	Glu
															430
Ser	Val	Val	Pro	Pro	Leu	Pro	Ser	Ile	Phe	Asp	Gly	Ala	Pro	Asp	Ser
															445

Gly Glu His Leu Glu Thr Thr Glu Gly Met Val Pro Ser Asn Ser Thr
 450 455 460

Ser His Leu Ala Asp Glu Ile Asp Asp Ser Glu Gly Ile His Glu His
 465 470 475 480

Leu Gln Val Phe Asp Ser Pro Leu Glu Glu Asn Asp Phe Val Ser Ala
 485 490 495

Ile Leu Glu Glu Asp Thr Ile Asp Val Thr Ala Ala Leu Ala Ala Ser
 500 505 510

<210> 30
 <211> 1633
 <212> PRT
 <213> *Phaeodactylum tricornutum*

<400> 30

Met Thr Thr Ala Pro Leu Thr Ala Ala Gln Ser Ser Pro Leu Thr Ile
 1 5 10 15

Ala Asn Thr Thr Thr Thr Thr Ile Asn Ser Lys Pro Leu Val Asp Ser
 20 25 30

Leu Thr Asn Thr Val Pro Glu Thr Val Gln Ser Ser Gly Arg Asp Gly
 35 40 45

Asn Ser Ser Asn Ser Glu Thr Asn Lys Ser Asn Ser Asn Arg Thr Pro
 50 55 60

Thr Lys Phe Ala Ala Leu Ser Glu Val Pro His Arg Arg Ser Ala Leu
 65 70 75 80

Val Ser Thr Ala His Ser Val Thr Ser Lys Ser Lys Gln Asn Ser Thr
 85 90 95

Pro Pro Val Asp Met Ala Thr Ala Ser Gly Ser Ala Ser Gln Gly Ser
 100 105 110

His Gly Glu Asn Thr Gly Arg Trp Thr Ala Glu Glu His Arg Leu Phe
 115 120 125

Leu Gln Gly Leu Glu Gln His Gly Lys Gly Trp Lys Lys Ile Ala Ser
 130 135 140

Leu Ile Lys Ser Arg Thr Val Val Gln Ile Arg Thr His Ala Gln Lys
 145 150 155 160

ES 2 682 279 T3

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

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

ES 2 682 279 T3

Asp Ser Asp Asp Asn Asn Asn Asn Asn Asn Asn Ser His Gln Lys Pro
 675 680 685
 Ser Ala Asp Leu Phe Gly Ser Gln Pro Asp Ala Leu Ala Phe Val Pro
 690 695 700
 Gln Gln Arg Pro Gly Glu Val Val Arg His Gly Pro Val Phe Asp Ala
 705 710 715 720
 Pro Val Arg Gln Leu Leu Phe Leu Pro Asp Ser His Phe Leu Ala Ile
 725 730 735
 Ala Thr Glu Ala Gly Leu Ala Val Val Ser Thr Asp Thr Asp Ser Gly
 740 745 750
 Ile Gly Gly Gly Ser Leu Asp Thr Asn His Asn Glu Asn Val Asn His
 755 760 765
 His Asn Val Lys Tyr Leu His Arg Glu Ala Gln Thr Ala His Asp Glu
 770 775 780
 Ser Gly Ile Arg Gly Leu Ala Leu Trp Gln Ala Lys Asp Cys Arg Ile
 785 790 795 800
 Leu Ser Ser Leu Ala Met Asp Gly Arg Leu Cys His Trp Asp Val Ser
 805 810 815
 Ala Pro Thr Pro Thr Leu Trp Lys Leu Leu His Arg Glu Thr Val Pro
 820 825 830
 Thr Val Thr Lys Pro Asp Leu Gly Glu Met Leu Gly Ala Asp Ala Trp
 835 840 845
 Asp Arg Ser Thr Ile Pro Val Ala His Ser His Glu Ser Ile Leu Phe
 850 855 860
 Leu Pro Gly Glu Thr Tyr Val Gln Ala Arg Arg Tyr Arg Asn His Thr
 865 870 875 880
 Trp Glu Leu Leu Gln Ser Pro Thr Gly Ala Thr Asn Thr Thr Asp Lys
 885 890 895
 Val Gln Gly His Ile Glu Ala Ile Val Ala Met Ala Pro Ala Pro Asn
 900 905 910
 Pro Arg Asp Pro Tyr Leu Val Thr Ser Gly Arg Asp Gly Arg Val Val

ES 2 682 279 T3

Ser Val Thr Phe Leu Arg Gly Gln Ala Gly Ile Asn Arg Ser Thr
 1160 1165 1170

Ile Asp Ile His Phe Thr Asp Ser Ala Phe Arg Arg Pro Val Ser
 1175 1180 1185

Phe Thr Asp Asn Met Gly Phe Ile Leu Gly Ser Leu Gly Glu Asp
 1190 1195 1200

Gly Gly Ile Phe Ala Thr Asp Leu Ala Glu Asp Glu Asp Ile Asp
 1205 1210 1215

Glu Glu Asp Asp Asp Met Asp Gly Leu Asn Val Ser Ala Ala Thr
 1220 1225 1230

Lys Ala Ala Val Lys Arg Ser Arg Lys Gly Pro Ser Asn Lys Pro
 1235 1240 1245

Thr Gly Ser Ser Ile Tyr Phe His Arg Phe Glu Thr Phe Gly Ser
 1250 1255 1260

Leu Arg Asp Lys Asp Trp Tyr Leu Thr Leu Pro Asp Gly Glu Arg
 1265 1270 1275

Ala Leu Gly Cys Ala Ser Gly Glu Gly Trp Ala Ala Val Val Thr
 1280 1285 1290

Ser Arg Arg Phe Leu Arg Leu Phe Ser Ser Gly Gly Asn Gln Gly
 1295 1300 1305

Glu Val Leu Trp Leu Asn Gly His Pro Val Thr Met Ala Gly Arg
 1310 1315 1320

Gly Arg Phe Val Ala Val Val Tyr His Glu Ser Thr Pro Leu Pro
 1325 1330 1335

Asp Gly Thr Gln Lys Leu Gly Tyr Leu Val Leu Asp Ala Met Ala
 1340 1345 1350

Asn Arg Val Val Ala Lys Gly Pro Val Ser Cys Ile Ser Gly Ala
 1355 1360 1365

Ser Thr Leu Ser Trp Leu Gly Phe Ser Asn Asp Gly Ser Leu Leu
 1370 1375 1380

Ala Met Asp Ser Asp Gly Met Leu Ser Met Leu Val Cys Ala Ser
 1385 1390 1395

ES 2 682 279 T3

Ser Leu Asp Ala Glu Gly Pro Thr Glu Lys His Trp Glu Trp Met
 1400 1405 1410

Pro Met Leu Asp Thr Val Gly Leu Arg Lys Ser Arg Asp Asp Ser
 1415 1420 1425

Phe Trp Pro Val Thr Val Tyr Asp Gly Lys Leu Val Cys Val Pro
 1430 1435 1440

Leu Lys Gly Gly Met Lys His Pro Asp Ala Val Arg Arg Pro Val
 1445 1450 1455

Thr Ala Ala Leu Gly Phe Arg Leu Pro Leu Ala Arg Gly Pro Leu
 1460 1465 1470

Thr Lys Thr His Thr Leu Glu Glu Leu Ala Val Arg Ala Ala Ile
 1475 1480 1485

Ala Leu Gly Gln Lys Lys Ala Ile His Glu Ile Ser Arg Glu Gly
 1490 1495 1500

Asp Glu Asp Asp Glu Asp Phe Glu Lys Glu Tyr Arg Ser Leu Ser
 1505 1510 1515

Ala Gln Val Asp Lys Val Thr Leu Lys Met Phe Ala Ala Ile Ala
 1520 1525 1530

Glu Ala Gly Lys Leu Glu Arg Ala Leu Asp Leu Val Glu Arg Leu
 1535 1540 1545

His Leu Glu Lys Ser Tyr Asp Leu Ala Met Thr Ile Gly Asp Arg
 1550 1555 1560

His Arg Lys Leu Val Asp Leu Ile Glu Glu Ala Lys Asp Arg Lys
 1565 1570 1575

Phe Gly Asp Pro Gly Ser His Gln Ala Glu Phe Thr Thr Lys Ala
 1580 1585 1590

Glu Ser Pro Asn Tyr Gln Arg Pro Arg Ile Ser Pro Asp Ser Ala
 1595 1600 1605

Gly Ala Lys Arg Ser Leu Asp Asp Glu Asp Glu Asp Val Arg Ser
 1610 1615 1620

Arg Leu Val Arg Arg Lys Pro Thr Phe Ala
 1625 1630

ES 2 682 279 T3

```

<210> 31
<211> 991
<212> PRT
<213> Thalassiosira pseudonana

<400> 31

Met Ser Thr Thr Leu Ser Ala Val Gln Phe His Phe Gln Glu Thr Thr
1          5          10          15

Thr Thr Ala Ala Ala Asn Glu Asn Gly Ser Gly Asp Glu Lys Ile Ile
          20          25          30

Ser Thr Ala Ala Thr Thr Thr Thr Thr Thr Thr Glu Gly Ser Leu Ala
          35          40          45

Ala Glu Glu Gly Ala Ile Gln Ser Pro Arg Lys Ile Ser Glu Gly Asp
          50          55          60

Tyr Thr Leu Gly Ala Leu Leu Thr Ser Ser Gly Gly Ala Ser Val Ser
65          70          75          80

Arg Val Val Thr Lys Glu Ser Leu Leu Gly Asn Asn Asn Asn Gly Ser
          85          90          95

Glu Gln Thr Thr Ala Thr Gly Pro Lys Ala Phe Ser Val Pro Ser Pro
          100          105          110

Leu Asp His Asn Asp Gly Gly Ser Phe Ser Phe Ser Ala Pro Glu Thr
          115          120          125

Thr Phe Gln Ser Ala Pro Leu Asn Ser Gly Gly Ser Ser Ser Ala Ser
          130          135          140

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

Lys Ile Asp Phe Ser Val Pro Cys Pro Leu Asn Ser Phe His Gln Val
          165          170          175

Asp Tyr Ser Thr Thr Asn Asp Gly Lys Thr Thr Pro Val Glu Glu Glu
          180          185          190

Thr Leu Tyr Leu Pro Pro Glu Asp Ser Asn Asn Asn Asn Ala Val Val
          195          200          205

Ser Ala Tyr Asn Ser Pro Ile Pro Pro Leu Phe Ser Ser Ser Ser Ala
          210          215          220

```

ES 2 682 279 T3

Pro Ile Thr Ile Asn Asp Pro Leu Ile Ser Leu Arg Ser Ser Ser Pro
 225 230 235 240

Leu Leu Tyr Ala Ala Ser Phe Asn Thr Pro Gln Glu Glu Asn Thr Ser
 245 250 255

Trp Ala Ala Val Gly Gly Ala Thr Ser Ile Val Ser Thr Gly Leu Ser
 260 265 270

Val Gln Val Pro Val Glu Val Ala Ala Val Ser Thr Ala Ala Val Val
 275 280 285

Ala Ser Asn Ser Ala Val Val Val Thr Pro Pro Thr Thr Leu Glu Glu
 290 295 300

Asp Gly Ser Ser Ile Ser Ile Ala Lys Lys Lys Lys Ser Lys Ala Pro
 305 310 315

Lys Thr Gln Lys Ser Arg Thr Lys Ser Lys Gln Pro Thr Ala Thr Ser
 325 330 335

Gly Val Val Ala Pro Ile Ala Arg His Ser Ala Thr Thr Asn Thr Thr
 340 345 350

Met Asp Arg His Asn Met Gly Ser Ala Pro Ser Pro Leu Gly Thr Thr
 355 360 365

Ala Ser Thr Met Ser Ile Ser Ile Pro His His Thr Tyr Gln His Gln
 370 375 380

Gln Gly Ala Val Asp Glu Pro Glu Thr Pro Ser Ser Ser Gly Ser Pro
 385 390 395 400

Ser Gly Asn Thr Val Ser Thr Leu Asn Pro Asn Thr Asn Ala Val Glu
 405 410 415

Asn Thr Gly Arg Trp Thr Ala Glu Glu His Arg Leu Phe Leu Gln Gly
 420 425 430

Leu Glu Gln His Gly Lys Gly Trp Lys Lys Ile Ala Gly Leu Ile Lys
 435 440 445

Ser Arg Thr Val Val Gln Ile Arg Thr His Ala Gln Lys Tyr Phe Gln
 450 455 460

Lys Leu Ala Lys Ala Arg Ala Gly Asp Gly Ser Gly Ile Pro Met Ile

ES 2 682 279 T3

Ala Pro Asn Gln Val Lys Met Pro Val Ala His Ala His Tyr Leu Gln
725 730 735

Gln Gln Ala Ala Gln Asn Pro Ser Pro Thr Gly Val Thr Glu Met Thr
740 745 750

Phe Pro Ser Trp Val Asp Pro Asn Asn Pro Pro Gln Trp Tyr Asn Glu
755 760 765

Gly Gly Asp Ile Asp Asn Leu Leu Asp Glu Ala Glu Ala Leu Asp Trp
770 775 780

Leu Thr Asp Thr Gly Asp Ile Asn Glu Thr Tyr Pro Pro Ala Val Ala
785 790 795 800

Ala Thr Gln Ala Ala Val Ala Asn Asp Pro Ser Asp Tyr Glu Pro Thr
805 810 815

Pro Val Asn Asn Tyr Ser Tyr His Val Asp Asn Ser Gly Ala Thr Thr
820 825 830

Pro Ser Leu Ala Ser Val Asp Pro Thr Ala Val Lys Met Glu Asp His
835 840 845

Thr Met Val His Asp Pro Thr Gly Met Met His Pro Ser Ala Asp Ser
850 855 860

Leu Ser Phe Leu Val Asp Pro Pro Glu Glu Gly Val Thr Ser Ala Glu
865 870 875 880

Leu Pro Ala Phe Leu Asn Asp Asp Ser Pro Gln His Gly Val Ala Ser
885 890 895

Met Pro Phe Ser Ser Ala Thr Glu Ala Val Glu Ser His Thr Phe Glu
900 905 910

Asn Lys Ile Asn Thr Ala Val Ser Asp Gly Asn Leu Met Gly Phe Pro
915 920 925

Asp Leu Asp Met Gly Asp Glu Gln Ala Phe Val Ser Ala Leu Leu Asp
930 935 940

Asn Ser Gly Gln Ser Thr Leu Ser Phe Pro Lys Leu Asn Ser Glu Leu
945 950 955 960

His Ile Gly Ser Leu Ser Gly Ile Gly Leu Ser Gly Val Gly Val Ser
965 970 975

Ser Asp Ala Leu Gly Glu Pro Leu Glu Asp Asp His Leu Asp Asp
 980 985 990

<210> 32
 <211> 408
 <212> PRT
 <213> Ectocarpus siliculosus

<400> 32

Met Thr Pro Lys Asn Thr Gly Arg Trp Thr Tyr Asp Glu His Arg Leu
 1 5 10 15

Phe Leu Arg Gly Leu Glu Leu His Gly Lys Gly Trp Lys Lys Ile Ala
 20 25 30

Ser Leu Ile Lys Thr Arg Thr Val Val Gln Ile Arg Thr His Ala Gln
 35 40 45

Lys Tyr Phe Gln Lys Ile Ala Lys Ala Lys Gln Asn Gly Glu His Gly
 50 55 60

Asp Val Ala Met Asp Ser Lys Gly His Gly Ser Arg Arg Lys Ser Arg
 65 70 75 80

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

Pro Ser Leu Gln Pro Tyr Val Ala Gly Ala Gly Ala Val Glu Thr Gly
 100 105 110

Leu Tyr Arg Phe Leu Ser Pro Ile Thr Ile Gln Asp Leu Glu Arg Pro
 115 120 125

Pro Ala Ala Ala Gly Gln Gly Gly Thr Ala Gly Gly Val Gly Gly Gly
 130 135 140

Ala Ala Gly Leu Gly Gly Asn Gly Gly Ala Ser Leu Gln Ser Gln Leu
 145 150 155 160

Ala Gly His Ala Thr Ser Ser Ala Gly Ala Ser Pro Thr Thr Thr
 165 170 175

Ala Gly Ala Ala Met Ala Ala Ala Ala Ala Val Leu Pro Pro Asn Trp
 180 185 190

Tyr Arg Ala Gly Arg Gly Val Asp Asp Leu Leu Thr Glu Ala Glu Gly
 195 200 205

ES 2 682 279 T3

Leu Asp Trp Leu Ala Asp Ser Gly Gly Ala Val Ala Val Thr Ser Val
 210 215 220

Ser Ala Val Asn Pro Ala Ala Ala Ala Ala Ala Ser Pro Ala Pro Ala
 225 230 235 240

Ala Val Ala Gly Ala His Ala Gly Val Ala Arg Val Ala Ser Ser Ser
 245 250 255

Glu Leu Ser Ser Gln Pro Leu Leu Lys Lys Val Arg Ser Thr Thr Val
 260 265 270

Asp Ser Ser Ala Pro Leu Ala Thr Ala Thr Thr Met Asn Thr Ser Ser
 275 280 285

Thr Ala Ala Ala Ala Met Gly Gly Ile Ala Arg Pro Ser Ala Ser Ala
 290 295 300

Ala Ala Met Thr Pro Ser Ser Trp Ser Leu Pro Pro Arg Pro Val
 305 310 315 320

Ala Cys Pro Pro Pro Ala Ala Pro Pro Thr Ala Gly Val Ala Val Asp
 325 330 335

Glu Leu Thr Pro Ser Gly Leu Gln Gly Ser Asp Ala Glu Glu Gly His
 340 345 350

Leu Ser Gly Leu Ser Asp Leu Glu Ser Leu Met Glu Asp Gln Glu Leu
 355 360 365

Pro Gly Ser Gln Pro Phe Gly Pro Val Ser Thr Ala Ala Glu Val Ala
 370 375 380

Gly Ala Gly Gly Asp Gly Met Asp Asp Gly Met Val Asp Glu Asp Ala
 385 390 395 400

Phe Val Asn Ala Leu Leu Asp Ala
 405

- <210> 33
- <211> 2016
- <212> ADN
- <213> Nannochloropsis CCMP 1179

<400> 33
 atggatgaag ttaaatatgc cgctggcggg ctgatcaagc cgggacgca aggcctgggc 60

agcatgctct ctacaccctt gacgcctctg gcttctctta ttgcatcgtc ggtggtagag 120

gccgatgcaa agcaacaaca actgctgaag gattccctca ccgccgacct caaattgctc	180
ctgcacgaat tcgaacgctt ccagcaggcc actgcagcag cagcgggaac aggaggcgtg	240
ggcgaggagg aggcagctga gcgatcgacg aaagtggagt tctttttggg atacattgag	300
cgagttctcc acgatttggc aggggctgac gcacgaaac tgcaggatct cgaggtgcgg	360
atcaagacaa gtctactacc actcaagggc caggtgggta gccagcttgc ggcgcaaat	420
aataactccc ctctcccca taaggagcag cagtcacttt ggttccatcc ttcttccacc	480
tgctctccc tctctctctc ttcttccgtc tccagcgtgc acacaactcc tcttggatct	540
cccttgcaaa gggaggagac ggtgatgagc acctacggtc cctttaccca ctcccgcgtc	600
gcagcagcag acgcagcttt tctctctctc tctctctctc ctctctggct gatgcgcct	660
gtgagtttcc ggagggatca gaggacata agtgggatta cgtcgtgctc gtcactcctg	720
tcgtcatgcg ggggggaagg aggtgcgggg catatggacg atttagattt ggagtgtttt	780
agcttgataa tggacgaggc tgcgctact gctcctttta ctactgcagg taatgggtggg	840
aaggatctgc cggccaatgg ctcatgtgtt gatgatgaca tgacggatgc tggctctcta	900
acttcagagg aaagcagcgt gtttgtgtcc tcccccggg agtctctctc ttgcctcagc	960
actgttagta cgggcttgga cccaacaagc agcaacagtg gaaataagcg gaccttgccc	1020
ttagaatttc ccagcagcag cagcagcagc agcagcagca gcagtctttc tctggccaca	1080
gcgtctctcg tatctacaga cactctgcag cagccgctca agcgcgcccc cagcgtcctc	1140
ctaccacca gcagcagcag cagcagcagt tgtgtctacgc atgcacctcc ctgcctggcc	1200
tctctctcca cctctctctc ttcttctttc tcttctctct tctctctctc ttgcgtggcg	1260
acagtggcag cagcagatgt cagcaagcct ctcttctctc aggttgagta tcagtgcggt	1320
gcttgcccg acacctacac cgtgcctcc tccctcaatc cctgggtgggc cctcgagcga	1380
caggagtgtc ccaagtgcaa gaaggttcaa gttccacgca ttgacattaa cctgcctgcc	1440
aacaccatgg aataccaccc ggctttgctc gcggaggaag gcgatgatga cgatgatgat	1500
gaggtgggag gcgggagga gggagggatg atgatgatgc cggggggagg ggtggacat	1560
ggacatgtgg aggaaagaga ggaggagag acgagtgaga aaggagcgg tggaagtagc	1620
gtactagagg aggacgagga agcagtggtg agccctatgc aagcttcaca gctgttgagt	1680
ttgttgagc atgcgcggac atgcccgggc aatcatgctg ccgagaaaca ccaggctgtg	1740
tgtaocagtg ccaagtatct catgttgcct gtgagggatt gtgatggcag gacgttagat	1800
ggggaggctt gtggattctc gtggtgcagg ccttgcaaac acttgctcgg gcacctggtc	1860
cggtgttacg aagccgagaa gtgtcagatc tgctgtttct cacaccaaga agaggaggag	1920
aaagtggaaa agaaggtgat gatgagtgtg gaggagatga ttgaggagaa aggaatgagg	1980
gtcgacacct acaggagctt gacgagcttg agctga	2016

<210> 34
 <211> 2364
 <212> ADN
 <213> *Ectocarpus siliculosus*

<400> 34

atgctggcgc aactggcgcgca gcagcagcag cagtccctccc cctccaccgc cgcaacggct	60
acgcogtcta ccgaaaccat cggcaccgac gccggcggca acacgcctcc cccccccagc	120
ggaagcctct tcctcggggg cgtcggcggc ggcaaggacg gcgccaccgc ggtcagcgc	180
ggctgggtgg ttgggtcgcc ctcccctact ctccccccgt cgtcctccaa cgagagcggc	240
ggcggcggcg gggcggcggg ccgcgtgaac atttccacgt gcgattccgg cgggcagttg	300
tctgggacga cgtcttcggc gtcgctgcg agagcgtgct ctccgctgtc ctccgcgac	360
ctgggcgacc tgtgcgggga tcaggaggga ggagacgttt ttctagtgg gggggaggac	420
tgcgacgacg atgccaccg gtcggggatc ggtcttgacc tgtacgacct gggacagggg	480
ggcaaagacc cttttcttc ccgagggcag caacgcgcgc gtcggtggag ctccatctcc	540
actaggtcgt cgtcttcgac ggcattccac gacggggaag aactctcggc cgattcgtc	600
tccgtcgatg cggactgcag cggcagcagc agcacgcgcg gcagcccgtc tgcggtcagc	660
tccggcgagc cgtccttgct ggccggcggc gcggcggcaa cgagagcgc caaagcaatc	720
aagaccgagg cccggcccga cggcagcgc ggtctagaag cggcgcggg acagcagggg	780
gttgcgcgtg ggggcgcttc tgccaccaag agcggcatcg accaggagat gggagagctg	840
tggagctgt tcgctccga tgccctcctg atgagtaccg tcctggacac cgagatggag	900
gccggcggcg gcggtgcggc gcgagcagga gcgtcgggca tcgaggtttc gatcagccc	960
accgccactg agcccagcaa ggcctcgcag catgctgctg ctgcggtcgg cccataggcct	1020
acgacggcag caggggctag tgctgtcgcc gtcactgcgc cggccgcggg ggtcaagatg	1080
gagttcctgc ctgcgcggg agcgtccgct ccttctccgg catcaccacc cgctccact	1140
gctgcgcgtg ccgcctcgt tttgccggca ccggcagcgg tagctccgaa acaggagtgc	1200
cgctgcggac aggcggcgtg tccgtcgtg atcacggccg cccgaaagcg accggtggcg	1260
gagctctcgc ctccgttgtc gggggcggac ggccccgcgc cgtgtccat ggtcacgggc	1320
ctgcccttc accagtcctc gcggcagcgg aagcggcagc ggagcctcgc cccggtgctc	1380
tgggcgcgc gcaccgtgag ctaogagtgc tcgctctgca aggagagcta cccgtccgag	1440
atcgcgtcca acccgtggtg gtcgctgttc ctgcacgagt gccacggtg ccaccggatg	1500
cagatacctc gagtcgacgc gacgagcgcg gccgtgagcg tggactacat ccaocccgtg	1560
tgccgcggag aaggggaggg ctgcgacagc gacgggtacg gcagcagtc gtgctccgac	1620
agcgcgacg acgttaccga tgacggcagg gagaggggaag ggatagccgc ctccgacacg	1680

gacatcatog cgggggactc gcaggccggc tgcaaggagg gccgcctgtc gacottccag	1740
gggtcccgcc tgctggtgct catgtcccac gccaggacgt gccccgggca ccaagccaac	1800
cccaagcaag cagaggtctg ccggtogacc aagttoctca tgctgcacat gcgogactgc	1860
acgggtcaca ccgctaacgg ggatccgtgc gagcacccgt ggtgccggcc gtgcaagagc	1920
ctcctgagcc acctcgtccg gtgccgggac ccgaacacct gccggatctg cacgcogctc	1980
gacctaccgg gcccgtcgog acagctgagg gatcttaacg tcgcccaggc gcggcatgcc	2040
tccgccgggg ccgccggcgc cgctgccgct actacttcta ctaccgctcc ttctttctgt	2100
gcogtggggg tgcccggogt gggcgggtgt cctcttcgc ttgogggggc ggcggogcgg	2160
gctgctgctg ccagggtgcc gtogatgccg gctcctgoga cggcggtaga cgcogctacc	2220
gtgaagtctg aggagatgtg cggccccgcc ggggccgcta cggcgatggg gctggcgggt	2280
tctgcgacga cgacgacgtc gctgoggcaa cagctgcagc agcgggtggt tgggggtgog	2340
ggaggggtag tcacgactcg ttga	2364

<210> 35
 <211> 3408
 <212> ADN
 <213> *Thalassiosira pseudonana*

<400> 35	
atggacgggt catccgacga caaccgtaag cggccgcacc ctgcatccaa tgccaacgat	60
gtggcttctg cagaggttgc tgccggagat aacaaacgag taaaaatcga agatgtagga	120
caggcagcag cggcgggtgga gacacaagat gtgccacaga ctccaggctc ggtggatgct	180
gttgettcat ctattgtacc gacgacgcaa ccgcagcctc tgacacagga agctgttcca	240
gcagccgtat cgtccacaac aactaccgct gttgtatcct ctccgacaca aatatctgcg	300
ggggaagcgg cggctctgac tgcttcgaaa actgcaacag ctaagccagc agcagacggg	360
gacttaatgc ctgctgcctt cctcacatcg tcgtcgtcaa ctcaagctgt tgctccagca	420
actcaagctc agagtgtgac accagcagtt gtccaaagta caactcagca tgttcctgct	480
gctagcacat cagcacagcc tccagcaaaa aaagaggtca agccagcatc aacaccatcc	540
accaccgtog tccaaccgca acgaccacaa aaacaacgag caccatctcc acccccaccc	600
ctaaaagcac ttactttcca ccacctccac aaaaaatatg gccccgagtt agactacatg	660
ctcgttgaat ttcgtaaact ggaacgacaa ctgttgggtg ctccattca tgctgccgct	720
gctgcatctg ctgctcccaa agcaggagga gcagcagctg ccgaggtgaa gccaaaggtg	780
gaaccgaagg ggtcgagaga acgaagggag aagttgcatg ggtttatatt gcatttggag	840
gatacgatac ggcaggtaga ggaggggtgt gcogttgaac ggagtgaaag gaatttgaag	900
tgcgagaaca atggtagtgg cggcgggaac ttgaagagtg aagagtgcgc tgctagtcat	960

catcaacaac aaccgaaaca acaaccgttt gaggaggaga agaagtcac tcatgcttca 1020
ttgcagccgc agcccaatac agcatcagct gcttcacatg ccgcaacaat aatatcaact 1080
ccaaacaaca acaactggcga agcaccocaa ttcacagcag cagacgcac actttoccaa 1140
ctccccccag aaaaagaacg ggaagaatca gtccagcgcc tcgaagagca catccttgcc 1200
aacctccttc cagtaaagat acgtctgacg agacaattgg cagcacaana gggagctaca 1260
aagaatccaa tcacagctcc actacgtgct ggctcgggtg cgcacggggg cgttcaaaag 1320
gcaggcgtga gtattgcaga ggctgtggag gcaaaaagaa aggcacaaga agagagggtg 1380
ttgcaacaac aattggtgca aaagtcttct gttcctgtga gcaaaagacat tccttctcag 1440
tttgggaaac cgattgggtc tggcagttca tcgttgacag ctcggttgca tggaggagt 1500
ctgggtgctt cggggtctgg agctgctgct gctagtctct ccaatgctgc gtcagggtct 1560
gcagcgagtg gtgcagcaac tccatcgaaa cgcagcaattc tgtatgcagg tgtagctcct 1620
ggatcgacac aggttccatc gtcagtccat gcagtctctg gagttcatcc tggaaatggtc 1680
ggagctgatg ctgccaaggc ggtagttgtg gcagaggagg aacgaaagcg actcaagtac 1740
ttggaagaga gtgcagctcg tgtagcgggc gtggctcctg cgggaggagg agccttggac 1800
agaaagcctg catctcgtcc aacagcaatt gaagcggcag cgcgaaacc gccagaaggt 1860
ccagcgacaa tggccgcacg tgctcgggca attgcattgg ctgcgacaaa caacaatact 1920
gggccttctg cagcgagtaa agtatcccgt ccaaatcagc agatacctgg tatcactgcc 1980
aaacagctac aacaacaaca tttgaagaag gtgtctcccc tggcagctgc aacggctgcc 2040
aatcaaatgg cacatttagg gatgaaacca gtcaaaccaa agaaaccaca tcttgcacca 2100
gatttcaacg atccggcatt gacagccaca caacaaaatg agttaagcct caaagaggca 2160
cgttggaggc aacggaagcg tcgaaaggaa cgtcgtcgta aacggtcggg cgttgttgtt 2220
gatcatgagg cgatgggttac tcagcatgct cagtcaatgc aagagcccaa tgttcaagca 2280
tcttctcca atgacgcagt tccggcgcag ccaatggttt tgagggtgaa caagaatgga 2340
gcgtatggac ccgcactgt ggaatacgtg tgcgctgtct gtaacgagg atatgtttct 2400
acttgcgaga tgaatccttg gtgggcctta atcaatcatg aatgtccaaa gtgtgaaaag 2460
aatcagattc cagcctcga tatctctgca ccaacaacg taatagaata ccatcctgca 2520
ttgctggttc aagaagatgg caaacagtt tcagcgcagc tgtccaacgg tggcgactct 2580
tcaagcgtgc aatgcaata tctcccacgg catcttgcta aaaagtctc tttgtctgat 2640
tcggaagtaa gtcaaacga cgcagcgcac ggagaagggtg gtaccgaaga atactttgat 2700
gaaagcagcg atgacgagga atcagtaaac aagaatgcgc tcgattcctt tgctaaagag 2760
gagagagcgg aacgcgagga ctatggcttt gagttcaaag gagagacatt aagtgatgat 2820

caagcaaagc ggttgcttat attgattgag catgcatcca tttgtcctgg aagacaccga 2880
tctgcaaagc acagaaatgt ctgtcacagc acaaaatact tgatgcttca cgtgogagat 2940
tgtcccggct tgctatccaa cggagacgtc tgtccatttc cttggtgccg aaagacaaaa 3000
caccttcttt accatcttgt ctcatgogag aaaagtaacg acggtaaaga gtgoggtatt 3060
tgttgcccaa agaatctatc ttccaacctt tcggagttag ttggtcttaa caaacaccgt 3120
cggaaagcaat ttgtggatcg gacaaaggcg attgtggcag ctgcgaaacg tcagcaactt 3180
gctgcagcca gagcaaaagc cgtogctccc agagctgcog tccaacatca gtatogtgg 3240
ccggtggtac gaaagggctc gattccagct gcaacaactt atgctgctcc acctccatca 3300
gcatctacgg tttcttatgc aactacttct cgcggaaccc atatgccatc gacgaacaat 3360
ccgatcatcc aatctcctcc agatgcaaaa acatccaacc aattttga 3408

<210> 36
<211> 3447
<212> ADN
<213> *Phaeodactylum tricornicum*

<400> 36
gtcggteget tggcgttgac tttgogccgc agattggccg gtatttctgc agcggaaaccg 60
ctgtogtcag tcacaactgt cgttttcggg ttggcgaaaa cagcccaggg cgtgogcga 120
tccttcttt tctgggcagc aaaggccgcg gaaattgtca aaaaggctcc atgtgacacc 180
gagaacgaag aaacaactgt cggtgccaaa acagaaagaa tcgcgcaaca ccgtaccgat 240
cgacgacaac ggcattgtct catggtaacg atggcaattc ctacgaatgt ttgottgott 300
ctctcagaaa tattccggtt cttaatgatg gagatggagg agcacgattt ggcgcaacaa 360
aaagaccccg gtaaagtctc gagctgttct gacagtaagg gagaaggaca cgcgctcgaa 420
gtgcaactcg tttacgtttc ttccgggtg gggttccgtg tttcttggcg tagtgogteg 480
tcgttgttgt cctcacgggg gcgggggtat ccaatcttgt cggatggttg tggtgcccgt 540
tcgcagtcca caaacgaggc agtttccagg actaccgaca gttogaacca cactttgcta 600
cagaattoga atctctcgca gcagcctccc ctccctcttc tgcttctgc cactgattct 660
ttgoggcac cgcgaaaccc tctctacagc aggaatgat cacacgacac gaatagcgt 720
ataggcgttt ccgacccgcg tacgcatacg agccgtgcca tgagcagtca cacgtogctg 780
caatactoga gttccggagg catogogaac atctctacca caaccgatcc tccacacaaa 840
cgactcaagt tggaccatgc catgagccac acatogctcg gcaaccatc cttgagctat 900
cacgattttg ccgcacatta cgacagtcgc agtaccttac aactagtag caccatggat 960
ctaggcgttt tgcggaaaga agattccttg ggcattgatg gcaaggacgg cgaogacgag 1020
gacgacgaaa atgatcagaa cgaccogata tctccacag ctgtacgaca agcgaocggtc 1080

caacctactg	ctcttccgaa	tgaaagtgcg	aaaccacac	acccactac	agcgaacgta	1140
gccaccacaa	attcogtttc	gtcctcogac	agtctgcgcg	atctatcogc	acaccgtcca	1200
caacatccac	agaatactac	togtcttccc	gtttcttctg	caacgactac	ggtaacatcg	1260
ggttogaatt	ctcogctctc	tgoggggocg	gtatcagccc	aagctcctcc	ctogcctctg	1320
ttacctctca	aggctaccaa	aatgtcacac	ctccgccaaa	aatacatgca	agaactagag	1380
tacatgctgt	gtgagttcca	aaagctggaa	ogtcagctac	taggtgccaa	ggcagcgaca	1440
gccgaatccg	ctggcagccg	cgaacgtcga	gaaaaactgc	attcgttcat	caocgacctg	1500
agcgatacga	tccagaacat	acagaccgga	tgtcagctag	agtcggaggg	aaaatcaacc	1560
gtcggagaag	cttccaagca	agatatagcc	caggaggccg	cgctggcaga	tttgacgtgc	1620
gaaaaggggg	aagaggaaaa	cgtgcaaaaag	ctggaagagc	acattctagc	caatctgttg	1680
cccgtaaaag	tccggctcaa	gaaacaactg	gogggcccagc	aaggtgccaa	gcataaccocg	1740
gcggggatgc	cggttgcgca	aaggggacta	gtggcaccga	gcgaaggtgg	taaaggcacg	1800
tttgcggcag	cggccgaaga	gcgcagaaaag	caattggcgg	acgcggccgc	cgcggcacaa	1860
ggcttccgatc	atacacacgt	accggcgga	ccggttcctc	cagaccagac	acaatttggg	1920
aaaccactac	aaggaaaacgg	ctcctcgttg	acgcgaaatt	tgcatggatc	cactttggga	1980
tccgcgatta	aagtgggaac	ggataagtcc	aaaattttgt	tcgctggttt	ggcgatcgga	2040
tcgtcgcaag	taaagtctgc	ggtcaacgca	gcttcgtcgg	tacatcagct	cgtaattaag	2100
gatcccgctt	tgttgaggtt	ggctcgccaa	cagagcgcgt	caaaacaaca	agaggacctt	2160
ccaccgcaaa	cacaacaaga	agactctcca	acgcaaagca	aacccaattc	gctgctgcct	2220
ccttctcctg	ccgagccgaa	tgactctcca	gaggatacaa	accgtaaggc	tatatcacta	2280
aaagtttcgc	ctgctggttc	ttctgcagca	gctttggccg	cgtctgagca	accagacgca	2340
gtcttgtcaa	aggetccacc	aagcagatta	gatgatgttg	atgccaccta	ccccgacatg	2400
ccatcggcag	ctttaaccga	tgaagaacgg	cgaaccctcc	gtcgtctcaa	acgccgaaaa	2460
aagagacgaa	aacgcaaggc	cgaagcaact	ccagtcacgg	cagcggccac	ggcagcacca	2520
gtgatcaatc	gccatcacia	gccgacgaca	aaaaaacggg	gacctcggac	ggtggaatac	2580
atgtgtgctt	tgtgtaacga	agctctacaat	ctctacctgtg	attataatcc	ttggtgggct	2640
ctggtcaac	atgattgtcc	aaaatgtcga	aaaaatcaga	taccgcgggt	agatattagc	2700
gcacctgcca	atacgatoga	atatcatcog	gcggtgctag	ctcacgcaga	cgaaaatggc	2760
ggtagtactc	cgacaccgcc	tcagcaata	gtgaagccag	tcacaactgt	gtcggctcct	2820
gtcactagtg	tgccaaaatg	tggtaatgat	tcogattcgt	tcggatctga	cttgtcagac	2880
gatgatcttg	acggcctggt	gtcagacact	gactcggagg	gctcgggaga	aatagggtatg	2940
gaaagaatag	atgcgctatc	gcctgcggaa	caagcagaga	atgaatattt	tgggggtgaa	3000

tacaaggggc caaaattgaa agacagtgaa gctgctcggc tactgattct catggggcat 3060
 gogtgcacct gtccttgcaa gcatcaatcg atcaaacatc gtgaaacctg cagaaatacg 3120
 aatggatga tgttgcatgt tcgggattgt ccaggaacta catcttcggt tgatgtctgc 3180
 ccatttccat ggtgcccga agtcaagcat ttgttgatc atcttgctc gtgtcggat 3240
 gccaaact gtgagatctg ctaccgacc aagctcaacc aaaatatgat cctgttaaag 3300
 gggttgaatc agcaccgctt catgcaatat agggagcggc tgatcggccg tggaaaggcg 3360
 ttgacaaagg tgtcaaatag tgcgccgaaa aatactccag ctccaggcca gcacaaaact 3420
 ttcacgacg tttcgcaaat gctgtaa 3447

<210> 37
 <211> 2847
 <212> ADN
 <213> *Cyclotella cryptica*

<400> 37
 atgctcgtcg aattcaaaaa actogaacgt caacttctcg gggcacctat ccacagcaa 60
 caaaaccctc cgaaggcga accaaagga agccgtgaac gcagggaaaa gttgcacggc 120
 ttcattttgc atctcgagga tacaatacgc caggttgaag agggatgtgc tttggaaaag 180
 ogtgaggagg gatcgataga ccaagtgtt catggtgctt cgggatcact ggogaataac 240
 caacaggaag aagtggaaga ggaagagaag aatcatccg atgcgtcact tcaacaaccc 300
 caacacacgg cacaaacca caccgccagc aatctcaaca atctcaataa aaccacgagt 360
 aatatggaag gtccacogaa aaagttcaca gcgcgcgaag ccgccctctc ctccctcccg 420
 cccgaaaaag aacgcgaaga atctgtccaa cgactcgaag aacacatact cgccaaccta 480
 ttgccagtca aagttogatt gaccaaaca ctgcgcgcac aaaaagggtc gacacgtaat 540
 cctgtcacag cccccgtccg agctggcgc gcaatacag tggcaggggg gaccatcgt 600
 gaagcggctg aagctaaacg aagggtcag gaggaggaat tgttgaagaa acaactgcaa 660
 caaaggcaga ttacgactag tcaatatggt aaaccattg gtggggcggg atcttcggtg 720
 actgctagat tgcattgagg tgttcttgg tccaatgcc ctgctgcagg tgcaagtgga 780
 totgagacga cgaagcgtcc tatactctac ggggggtgg cgctggatc gtctcaagtt 840
 ccgtcaacta tcaaaaccgt atcgggagct catcccggc taattgaaa agatgccacc 900
 aaagctgtgg ctttgccga ggaagaacga agacgcttga aaaatcttga agagaatgcc 960
 acccgtgtcg cattgggagt tgcgcgaaa ccactccag ctgcacagc attggatccc 1020
 agaaaaccag cggcgtgcc cacgttgaat gcatcagcat tgccaaagca gccggaagg 1080
 cccgccacgc ttgcggccc ggcacgagcc gtggcattgg cggctgccag cagcggaggc 1140
 gcgacgtcga aaatttcgag acccaatcaa caattccga ccaggtcatt gcagcaacat 1200

catgtgaaga aaggcoctcc tgtgatggct gctccagctg cgacgatggc tgctccagct	1260
gcgacgattg ctgctcccag agctcctcat atcaatcagt ataccaagcc atccgoggcc	1320
gtgccttatac attctgtccc gacccctccc atgacggcaa aatcgaagaa accccacata	1380
gogcccaatt tcaacgaccc atctctcact cccgagcaac gattcgagtt acgtttaaag	1440
gaagcacggg ggaggcagcg taaacgacgg agagaacggc gacgcaaacg tttggagggg	1500
tatttacatg ctgctggagc gtatcatata gttcccttag ccgtccaaca gcaaccatct	1560
ctgtctctgt cgtcgcaacc gcatttgcaa gaaacacagc cggaacgagt tgcattccact	1620
gtcgttacc ctaatgctcc agcgctccg cctccagtaa aactccgcc accaccagtc	1680
acctcggcct cgcgacccaa aaagaatgga gcttacgggc ctcgaaacagt agagtatgtc	1740
tgtgcgggtt gtaatgagac atacatatca acttgtgaat tcaacccttg gtgggactg	1800
acaagtcaag attgtcccaa atgoggcaaa ccgcaaatcc caaagcttga catttctacc	1860
cctgcaaatg aaatcgacta tcatcctgct ctattgagtc aagaagataa cgctaaacct	1920
cagagttctt ctgtgtggc ttggcgaat gcaagcaacc ctgcagtagc tgccccacag	1980
gttgcaaac cagtccaata catgcccaag ccgcccgtc acatgaagaa gaacttcttg	2040
ctctcagatt cogaagtcag tttgacagac gagagtgatg gcgaaggtgg cgggtgaaag	2100
taagacgaga gcagcgaaga agaagacacg agttacgaca atgacatgga ttcggtgacg	2160
cgcaagaac ggttagagaa ggaagagttt ggatttgact acaagggaga ggtattgagt	2220
gaagaccaag cgaggagggt gttggtgctg atcgaacacg cctccatttg ccttgaaga	2280
caccgatcag caaaacatcg caacgtatgc catagcacia aatacatgat gcttcatgtc	2340
agagattgct ggggtctatt atccaacggg gatgtgtgcc cattcccttg gtgocgaaaa	2400
acaaaacacc ttctctatca ccttgcact tgcacaaaga acgacgacgg cagcaaatgt	2460
tcaatatgct gtccggaaaa cctttcgtca aatctcatgg atttggttg cttgaattca	2520
tatcgccgga agatTTTTGT agagcgagcg aaggctgtag cggctgcagc cggggcgaca	2580
cgcatcaaa tggcgatagc aaaagcaaaa gctgccgcgc aatcatcaac acaaccgcat	2640
gtcttaacaa gctctcagat tctgactgca ccttcaacta accataatta cgagtcgcag	2700
acaaaacctt tcgctactgc atcgacccat gttgcaactc atgctacaca accattagga	2760
aatgctagga gggctcttc cgttcaagat gctacgatag catcgaacaa tacaggaccc	2820
caocttgatt cggagttgt tttatag	2847

<210> 38
 <211> 1650
 <212> ADN
 <213> *Ectocarpus silicosis*

ES 2 682 279 T3

<400> 38
 atggtcatgg ccgcgacggc cgacggcggg gtgtgcgaga tcaagatgga tgacagcggc 60
 cgctgacgc acatctggtg gcaaacgaga gaggcggcgc gacagcgctc gagcgatgac 120
 gatgctccga cgaagggtgg cgttgcggtg ggcggtgtgc aagccagccc cgactcggag 180
 gtgtcaagac ggctggatct actcatggac aagtccttcg gagatcttca gttcctctcg 240
 gacgaatttg cgaagctgga ggtcgtggtg gcgccggcag tgcaggagcg ggacgacagc 300
 ggcaaggcgg gcacagactc gaagctgggg aggctccgct tcttcaccac acacgtccgt 360
 cggacgatgg cccggatgcg ggacgcgagg agcggggcgcg accccatgtc catgtcccag 420
 ctggctctgc tagaggagca catcgcgacg tccattgctc cggggggggg gggaggggta 480
 tccagttcgc ggagctcgtc ggtgtcgagc agcatgggcg gtagagccgg cgcggaggag 540
 ctggtcggag ccgggcggga ggaggaggag gagtggtcgg tgggcctgga gctgcacgac 600
 gaggcgttcg gcagcggcgg gagcgtttct ctgggcctgg tgtcgccgc cccggcgtcg 660
 gtccccgaca gaggagcggg cggaggtatc tacgaccccc acgcctgctc cgcgcgcttg 720
 acgcctgcaa gtagcagttc ttccctctcg cgcctctgct ggggcagcgg cggcggaacc 780
 ggcggtagcg cgggcggcgg ggaagagat gttgcccggg gggggaggcc acaccgtcac 840
 acgcgggcag atcacgtgga catgctgagc cagctcgagg cggagggggg gttcgcccg 900
 gacgacagct actcgccctg cggcgggtgt ggccgcagcc gcggaggcgg tggcatgggc 960
 ggggtcgcgt tttcgccaga gccccgggag gtgaggtacc agtgccgggc gtgcgcggcg 1020
 agctacgagg cgacggtgtc gggaaacca tgggtggctgc tggtgaggca ggagtcccc 1080
 atatgccaca agatgcaaat acctcgcgtg gacatcttga atcccaccaa caacgtcgag 1140
 agtcacatcg cgtttctgac ggagaatgct tctgacggcg atggcagctg catggactgg 1200
 gacggggaaa cgagcgatga aaactccggc gacgagtact cgggggacga gcggcaaggg 1260
 ttgtccgcgg gaggaagcat gtcgggcggc gacggcagcg ggcttggacc tacctggac 1320
 tccgaccagg ccgccaagct gctggtgctt atgtgccacg cccggcactg cccggggaat 1380
 caccgctccg ctaggctggc agaggtgtgc cgaagcgtca agtttctgat gctgcacttg 1440
 cgagactgcg acgggaagac caggaacgga gacccttgtc cgatgccgtg gtgcgagcct 1500
 tgcattgtcc tctgcatca cctcatccag tgtccggagt cgacgggatg caagggggcac 1560
 gagcacaaga acaggggtga tctgcctagc caaacaccgt caggtttcgg acgcctgaga 1620
 acgcttcggt ttgcggcgt tttttcttga 1650

 <210> 39
 <211> 2241
 <212> ADN
 <213> Navicula sp. WT0229

<400> 39

atggaagaac	cgctggtggt	gactgtgaaa	cgagaaaaca	taaatgtatc	accogaacca	60
ctcaagaatg	atggggtcgt	cagtcacgat	gaggctactt	cgacgacggc	agattccctc	120
gtctccgaag	ccgctccact	caaagcgacc	acctttcgac	acctgaatct	aaagtatttg	180
goggaactcg	agtacatgct	ctgtgaattc	cagaaactcg	aacgacaatt	gttgggagct	240
cggaacttgc	aagcgagcga	gtcggatggg	tcgagagagc	gaagagaaaa	gctacattcc	300
tttattctgc	atctagaaga	tacgattcag	cagattcacg	caggatgcga	aacagaaggg	360
aatcaacgg	agactaccgg	attggtcaag	ccatcgaacg	aaaagaaaga	ggaagaagcc	420
gtacagaagc	tgaagaaca	catcctggcc	aacttgttgc	cggtcaaagt	cagactaaca	480
aaacaactcg	cggcgcagca	gggcgccaaa	cacaaccggg	ctgccatgcc	ggtgcgcggg	540
gtggtctcgg	aatcgtccaa	agaaacggat	acttcccaat	ttgaaagcc	gttagagggc	600
ggcgggtcca	gcttgacaca	aaaactccac	ggaagaacat	tgggcgcgga	agggagggca	660
cacggacacg	gcgtgggtac	cgtccacagc	aaacgggccc	aagccaaggt	tttatatgct	720
ggaatggcga	ttggtagtga	taaacatcaa	atgagaagtt	cactgagtgc	ggcgagttcc	780
gogcatcgac	tgttgttgca	aggaaaggac	gtcacggaca	agagtcgcac	goggccgcga	840
gaagaaaaaa	tgatcgagag	aagaggcgcg	acaaacaaag	acggtggtat	tcctccaagt	900
gatgacctac	caccacagac	cgacgcatcc	tttgcctcatg	ccggcagacc	ggacgccgac	960
tcggtcaact	cattgctatc	tgccgagcga	cgacgcttgc	agagaaaacg	acgccacaaa	1020
cgaaaacgaa	ttcttttgca	agatccgcaa	cagcaagccg	ctcttgcaaa	aagaaaaaag	1080
ggaaccagca	gcagtaagaa	acggggaccg	cgaaatgtgg	aatacatgtg	tgctctttgc	1140
aatgaagtgt	ataactogac	atgcgattac	aatccctggt	gggcgctgac	gcaggaagaa	1200
tgtcccaagt	gtcaaaaaac	tcagattccc	cgaatcgaca	ttggtgcccc	cgccaatgcc	1260
atcgaatacc	atcccgcctt	gctagcacac	goggacgaat	ctgccggcgc	agcagaacca	1320
agtgctgtcc	ttgaaccaca	agacttgctt	gtcccttcga	caacaggaga	tgatatggaa	1380
tatagcgacg	tggacgactc	ggatttatca	gatgaagatg	gactgctaag	tgatgcgagt	1440
cttgacttgg	attcggacga	ctccgaaatt	gcogattcgg	agaatatgtc	tcctgccgag	1500
caggccgagt	cggaaaagtt	tggtgctgaa	tacgacggtc	ccaaattttc	cgacgcggaa	1560
goggctcgtc	ttttaaat	gatgctacac	gcttcgacat	gtccttgacg	acacaagtca	1620
tcggaacatt	acgacgtgtg	tcggagtgtg	aagtggatga	tgttacatgt	tcgagactgc	1680
cctggaacaa	cgtcaacttt	tgacgtttgt	ccctttccct	ggtgtagaaa	ggcaaagcac	1740
ctgctctacc	atcttttatc	ttgcgaaaac	ccccaaagtt	gtcccatttg	ctcacctgta	1800
cacttgaatg	cgagcatgaa	atcgcttcgt	gggctgaatc	actatagact	gaaaaagcaa	1860

ES 2 682 279 T3

caacagtgtg taataggtgc aagtcagtca ccagggaaac catccgcagc gaggggcggc 1920
cattcatctc cgaaaaagtc caacgagtc actcaggatg aagtggacac atgcagagac 1980
acgctagacg cgtttgtagt tggaaacgaag gagccgacgg agatttcgtc atcgtcggta 2040
ccctcaaaact ctacagccgc cgtogatcag ccagctcatg aagatgaagt gaaaacccaa 2100
ttgtatgaat ttgtagatca atgggaagat cccgaacatc ctgccgctct gaaagatgat 2160
ggattgaaca attcacagtt gacggttggt cattgcgaca atgatgtatt gataaagcag 2220
gaggacgaca attcgcagta a 2241

<210> 40
<211> 984
<212> ADN
<213> *Aureococcus anophagefferens*

<400> 40
atgcggcgcg cggcggcgga ggcgaaagcg atggaaccgc tgcgtgcgct ggcgttcgct 60
cagtcggtca agtactactg ctccgtgtgc gagcacgcct acgagacgac gtcggacgcg 120
aaccggttat ggacgctcgc gcgccacagc tgccccagc gggcgcgct ccagtaccg 180
gagatcgaga tcgacgacat ctccgtgccg acggccgagg cgcgcgcgcc cgacgacgcg 240
tcgacgctgt cccatgccc a gggcgacggc gcgcgctcgc cgcgcggcc cgcggcgctc 300
ggccgcgcgg ccgcgcgca cgacggcgag gtgctcctcc gcgcgcgacg gccgcggaag 360
cgcagcgcgc ccgacgacga cgcgtcgacg atgtcccacg cccagggcgc ggcgctcctc 420
gagctcttcg accacgtgcg gagctgcccc ggccgccacc agtccgcggc ccacgcgcgc 480
gtctgcgcgg gcgccaagtt cgtcatgctc cacgcgcgcg actgcgacgc cgcgccgggc 540
acctgcggcg tcgagtggcg cggcgcggtc aagggcctcc tctcgcggt cgtctgcggc 600
cagcagggcg acaagtgcgt cgtctgcgcg gagccgtccg aggcgatgga cgtgggcgac 660
gcgtcgcgc ccacggtgtc gccggacgcc atggacgtgg acgacgcgct gccgccagc 720
gcgtcgtcgg acgcgccgat gccgcgcgcg ccgcggcgag agccgcggt cgggtgcggcc 780
gcgccgtcgc cgcgccagtc gacggaccgg gcgttcgagg cgcgccaggc cgtgccgtcg 840
ccggtccgcg cgtcgcgcgc gcggggcggc gcgcgctcgt cgcggcgct cgcggacatc 900
acgatgtcgg agaacgtcga gatggacttc ccgcgtgact tctgcttcgc cgacgagctg 960
cggcggcggg gcgacgcgc gtga 984

<210> 41
<211> 1797
<212> ADN
<213> *Nannochloropsis oceanica* CCMP1779

<400> 41
atgcctatgg tcaccctctc ccaggatgca actactaccg cggccggcag catgatgctt 60

ES 2 682 279 T3

cccctccttc cctccatccc tgettccgct accagcgcct cttttaccgc tccttcagcg 120
 tcctccacga ccaccaaadc tccc aaaggc acttctccca tgactcagct tggagagaca 180
 gggcgagaga atactggccg atggacttgt gaagagcatg tgctgtttct taaaggccta 240
 gaaatgcacg gcaagggtt gaagaaaac gcaaagctaa tcaagaccg aacggtggtc 300
 caaatccgca cgcacgcca aaagtacttc cagaaattg caaaagccaa gaagaacggg 360
 caccatggcg atatgctcgg aatggaaggc tcacactttg ggggaaaacg tgtcaaattt 420
 accggaaagc gacgtgggct tgtctataat tcgtatttag taggtgccga ggccacctct 480
 ggggctatct ccccgcggtt gcagacgttt atgcgggoga acttggggat ggaggggcgag 540
 cgtgtaggcc ttatgacgga taaggaggag gatgcagcaa tcgagaaggg actttatcgt 600
 ttctctccc ccgtagtgct ggatcccgcc acgcgtaac tggacgcctc cgtcctgag 660
 atcttgccct taccaccag cactccagcg atggcggtgc accataccag tagcagaggt 720
 agcagcagag gggggttga tggagagaca acgggagagg aggacggcgg aagcattcgg 780
 attgtaatgg gggatggagg gagcgatcaa gatgcagagt cgtcgttggg cgagcccttg 840
 ccaactttgg cgcgggtgac accggagatg tacacacggt gtggagtcc ggaatggttt 900
 aagaaagggg gggacattga cgaattgctc attgatgcag cggactcga ttggagaagt 960
 gactcgggtg gggacgcacg gaaggtggcg gatcaagga caagtatttt gaatgcgaat 1020
 attaatggtg gtaattgtgc gactgtggcc ccggcagtgg tgcggaaggg ctgtggtagc 1080
 aacactaata agatgaactc agcggcgcct gtgctgaaca tgacagggtt ggtggggct 1140
 ggagggtctt cagggtgga gggcaagggt agcgacacta gcgaaggtag cagcagcaac 1200
 ggcagcagca agaacatggc tttgacggca aatgcgtcgg cgggatgtgg acaagggagc 1260
 tggggagtgc gggggggggc acagaaaaag cagcagcagc agcaacatga ggtaccaacg 1320
 cagcagcagc aggtaccaac gcagcagcag caggttcacg gcattcacgt gaaggaggaa 1380
 gggatggagc tcctgagagt catggcggat agagggactg ttcacggcca cgtccatgag 1440
 gaggatggct ttgcggcgtt tgacccccat catgtcagc tgaaggagga gcactcccac 1500
 catgatttgt tgctggaaga gctgccccac gacagcaacc acgatgacgc cctggcccat 1560
 attgtgttct cagtgaatgg agagtccgat cttcattcct tgccgagggg tgcaggaggg 1620
 gggggagcgc atgtgcatgc gcgcgggtt gtgggtgggg ggagacagca tcactatcat 1680
 cataatgaca atgatatcca tttgcatcgc tacgacgcgt atttgaaga ggaaggggag 1740
 gatggcgggc atgggttgtt gttgttgag gatttggat ggggaatcga gtttttag 1797

<210> 42
 <211> 2031
 <212> ADN

<213> *Ectocarpus silicosis*

<400> 42

atgtcgtctc aactccaggt agacgacggc gcagccagca gcagcgataa cttggacaac	60
gaggggtgtgt gccagccggg caacgaaaac acagggcggt ggacttogga cgagcaccgg	120
ctgtttttga ggggcctcga gcttcacggc aaggggtgga agcagatcgc caccctcctc	180
cagacgagga ccgtcgttca aattcgcact catgccaga agtatttcca gaagctgtcc	240
aaggcgcagg ctagtggcac gtcacacctg gaccccgcaa cctcatgag caccatggac	300
gctggaaagc ctgcctccgc ctctgtgtca cggaaacttg gcagcagcac aatggcgaac	360
agcccgggcg agtcggaagg gaggctcatg agcctgagga agcgcgcgaa togtggccgc	420
cacccgagac acgacgatga ccaagactac gacaccaaca gcgaggagag ctacgactac	480
gcccgcagca gcgccaccac gcgcacccgc cgacgcaggc ggagcgtcag cagcggcagc	540
agcggcggcg gcggcggcg atcagcgaag agcgaagagg aggacgggga aggtggtggc	600
aggggtcacc gcggtgtgta ccgcgagacg gcgctggggc cgacgacgac gacgacgatg	660
atgatgatgg tgaggcaggc agaagaagcg cctgcgttta cgtcgggtgg cggcaacagt	720
ggtgcgcacg gcgcggagtt tgatgaggag gaggcggagg aggagttgga cgagagcgtg	780
gatgtcagag gacacaacag caacaacaaa aaogtcggta tttcgtcgtc gctcctgcac	840
ggcggcaccg ccggtacgtg gagcaagcgg cggccggcca aaagccaccg gccttccccg	900
acgaaggcgt ccaaggccgc cgcattgcc gccgcggcg gcgcggccga cgcgccgcgc	960
ctcgcgtggg aggcgcggc ggcaaggct gccgcgggag aggagggcga ggaggctgct	1020
gctgatggtg tagtggggcg tgggtccggg acgaagcgtc accggagcga gagcgttagc	1080
agcagcagca acgacgcgtc gctcggcaag accatcaaga gcctcaagcg cacgggcggg	1140
tcgtctcaga ccacacggat tagtcctact tcagtggcgg atgtcaacag cttcatgtcg	1200
ttcccggtgc cgcaaacgga gcgctcggac atggcgatgc accacctacc gcagcagttg	1260
gccccctcct ggtgcaccaa gcccagggac acctggatgg ccggggcagg gctggacatg	1320
gggggtctcg aggcggacga cgcggggccc ttccgggtgt tcatcgacga gcgttcgctt	1380
tcgcaccag gcgcgttttt ccagcccgcg gtggagacct ggagcggttc cgacacgacc	1440
gacgtctcca cgtccgcggg cagcgaccac cagcaccacc ggaccgtcgt acccgagcat	1500
cttcctcca acacgaacaa ggtcgcacatg tccccggggg gcgttttgac tagcgacaac	1560
acgcataatc aagaggacgt ttccatcgcc aagcaccggg gcatgatgga gccaatcgtc	1620
cccatttcg acgcggcaat gccagcag gtcagccgtc acgtcagggc cgacagcgc	1680
accgcctgtg ctgcgcggg ctgcgcggc gggcgaagcg cgagcgcgg cgggcgcggc	1740
gacggcggcc actcgcctcga catcacctcg ggcctctggc tcgaccacc gcaggaggcc	1800

cacgaggccg tcgcgtcgcc cctgatgtac ggcctcggcc tgtecgctcc cgccaacacc 1860
 gggggcgccg gcggcggggg cggggcgggg gggctcggcg ggggcctgtc tctgccttcc 1920
 ctcgacgagg aggaggtcct ggggttcttg gcctgcgcgg aggagaacag caggcaggcc 1980
 gccgcgcg cgccggggg cgaocgggat gagaacgatt tgctagtctg a 2031

<210> 43
 <211> 1362
 <212> ADN
 <213> Navicula sp. WT0229

<400> 43
 atgaaagtog aggcaactca tcaacacgcg tggggggctg atcattcttc aacgctggac 60
 gccccgcctc cagttctctc tcattctggt ccagcggcgt ctaatgcatc gaataatgca 120
 cccaaatcga aaaagaaaa aggccaaccg gcgataacga ttgctgctgc ccctagccaa 180
 gctgttgctc ctttggtctc gggggagaat acaggtcggg ggactgctga agaacatcgc 240
 ttgttcttgc aaggcctcga acagcacggg aaaggatgga aaaagatcgc ttctctgata 300
 aaatcacgaa cagtggtgca gatcogaacc catgctcaaa aatacttcca gaagttggca 360
 aaggcacgcc agaatggaga ggaaggtgat attacaatgg aagggcgcgg aggaactgct 420
 tccatcactt cgagcacgac tgcaacagct gcgctgacaa acaagcgtcg togccatctc 480
 acaggaacaa agcgcgaagg gatacaatcc attgtagcat cagcgcgaacg gcaagcaaag 540
 aaagcgcatt tgcctgaaat cggagatgca aagaagatc cggttgttcc aggggttgca 600
 cccgcttttag cgtactatgt cacgccaagt cagtcgtcct cagtaagcgc tggttccgac 660
 gtgtttacgg aaggaaatct atctggcccc gttctogaag actctttgtt caggtttttg 720
 accccagttc cgattgcctc tgatgatgtg aatgaggtag cacgccaggc gggcgcgaac 780
 ccgattactt tgccctcttc gaatagtcac gcattatcat ctgttccagg aagctcgcct 840
 actggagtcc aggagttgtc aatttaccba tcgtggactg acgcgaaaga tcccccttct 900
 tggtaaccca agggagcaga tgtggatgcc ctctggatg tatcggacac tcttgattgg 960
 ttagcagata cgggtgatct tgacgaagaa taccaaccgc agtcaaatga catcgacact 1020
 tgttcattcg gtggccaagg cgagcaccat gagatgggaa tatccaatgt gcacaacaac 1080
 actagcgtga gttcgcttac tcacgtcgac ccaaatatgg tctcggttgt cctcctcta 1140
 ccctctctct ttgaaggtaa tctgatgta gtagaagcgg agctaactgt agggaaagat 1200
 gtaaacgtcg ccatcggaaa cgcctccctt cttatcgcgc caagcgatcc taatgcggag 1260
 acccttcagg tgttgatag ccccatggaa gagcaagagt tcgtgtctac cttactggaa 1320
 acaactgcog aaagtagcga caacctggcc gtcttaagtt ag 1362

<210> 44

```

<211> 1539
<212> ADN
<213> Phaeodactylum tricornutum

<400> 44
atgaccacgg cgccactcac tgcagctcaa ttttctccct tgactattgc caatacagcg      60
acaacgacaa tcaatagcaa gcctctcgtg gattctctta ccaatacagt tccagagaca      120
gtccaattct ctggacgaga cggtaattcc tcaaacagtg agaccaacaa aagcaacagc      180
aacaggactc ccaactaaatt tcttgcgttg tccgaagtgt cgcacgacg ctccggcatcg      240
gtgtctacgg ctcaactccgt cacgtcgaaa tccaacacaga attctactcc accggtagac      300
atggcaacgg cgtccggatc agccagtcaa ggctcgacag gcgaaaacac gggacgctgg      360
accgcggaag aacaccgctt gttcttacag gggttggaac agcatggcaa gggatggaag      420
aaaatcgcgt cgctcatcaa gtccgcaacc gtctgacaga ttcggacgca cgcccagaag      480
tactttcaga aattggccaa ggctcgccaa aatggggaag aaggcgatgt cgccatggaa      540
ggtcgcggtg gcgtggcttc cattacctcc gtctcgacaa ctgctgtttt acccaagcga      600
cgtcgccaga caaccggaac aaaacgcaag gccattcaat ccgtcgtggc ttcogcccag      660
cggcaaggca agaaacttgc cgccgcaaaag acgaatccta ctcgacacca tccttgcgg      720
ccgcccctac caacggtcgc ccccgcactc ggcattaca ctctcccag tactgcatg      780
atggccaaaa acggcacccg agtgaaggaa gaatacgtct cgcccaccaa tctttcagga      840
ccggccctag aagattcatt attccgcttc ttaaccccgc ttccggtatc ggaaccaccg      900
ctcaacgaag tagctcgtca agcgggtgcc aaccccattt ctctcccac cgacaaccca      960
agctctattc caacgggtgg tgcaggagaa atctcgccca cgggagtttc ggatttgatg     1020
ctttaccctc cgtggacaga ctcaaaagag ccaccttctt ggtacagcaa gggcgcgac     1080
attgacgcat tgctcgatat gggggattcg ttggactggt tggacgacac gggggatttg     1140
aacgagtcat atgtaccacc cgtcgtggac acagcaatgg ccgctccaga accgcacacg     1200
acctttcaca ggtactccga tctgggacat tcaaagggac ttcacagtac cagtgtgacg     1260
tctctgccac atgtcgattc caacgcaaat gtggaatccg ttgtgccgcc acttcctcc     1320
atattcgatg gagccccoga ctogggagag catcttgaga ccacggaagg gatggtacct     1380
tccaacagta cttctcactt ggccgatgaa atcgacgaca gtgaaggcat acacgaacac     1440
ctacaagtat ttgacagtcc tttggaggag aacgacttcc tatcggccat cctogaagaa     1500
gacacgattg atgttacagc agctctagca gcgagctaa                               1539

<210> 45
<211> 4902
<212> ADN
<213> Phaeodactylum tricornutum

```

<400> 45
atgaccacgg ogccgctcac tgcagctcaa tcttctcoct tgactatagc caatacgaag 60
acaacgacaa tcaatagcaa gcctctcgtg gattctctta ccaatactgt tccagagaca 120
gtccaatcct ctggacgaga cggtaattcc tcaaacagtg agaccaacaa aagcaacagc 180
aacaggactc ccactaaatt tgctgcgttg tccgaagtgc cgcacgacg ctccggcattg 240
gtgtctacgg ctactccgt cagctcgaaa tccaaacaga attctactcc accggtagac 300
atggcaacgg cgtccggatc agccagtcaa ggctcgcaag gcgaaaacac gggacgctgg 360
accgcggaag aacaccgctt gttcttacag gggttggaac agcatggcaa gggatggaag 420
aaaatcggct cgtcatcaa gtgcggaacc gtgtacaga ttccgaagca cgcacagaag 480
tactttcaga aattggccaa ggctcgccaa aatggggaag aaggcagatg cgcacagaa 540
ggctcggtg gcgtggcttc cattacctcc gtctcgacaa ctgctgtttt acccaagcga 600
cgtcgccaga caaccggaac aaaacgcaag gccattcaat ccgctcgtggc ttcgcccag 660
cggcaaggca agaaacttgc cgcgcgcaag acgaatccta ctgacacca tcccttgccg 720
ccgcccctac caacggctgc ccccgcactc ggcattaca ctctcccag tactgogatg 780
atggccaaaa acggcaccgc agtgaaggaa gaattcgtct cgcaccacaa tctttcagga 840
ccgcccctag aagattcatt attccgcttc ttaaccccgg ttcggatc ggaaccaccg 900
ctcaacgaag tagctcgtca agccgggtgc aaccgattt ctctcccac cgacaacca 960
agctctattc caacgggtggg tgcaggagaa atctcgcca cgggagttc agatttgatg 1020
ctttaccctc cgtggacaga ctcaaaagag ccaccttctt ggtacagcaa gggcgcgac 1080
attgacgcat tgctcgatat gggggattcg ttggactggt tggacgacac gggggatttg 1140
aacgagtcac atgtaccacc cgtcgtggac acagcaatgg ccgctccaga accgcacacg 1200
acctttcaca ggtactccga tctgggacat tcaaaggac ttcacagtac cagtgtgacg 1260
tctctgccac atgtcgattc caacgcaaat gtggaatccg ttgtgccgcc acttccctcc 1320
atattcgatg gagccccga ctccggagag catcttgaga ccacggaagg gatgaaagcc 1380
caagcatcgg atctctcgca caaacacacg cactcgttta cggcctgtcc gcttcgactc 1440
accatgggca cgatagtatc tacagtgggt cggagcgcag agacagattt tttcgtcacc 1500
gagtctctac agttcctact gtcttggaaac gagggaccaa gacggcaaac gcggcggatc 1560
catgatccaa cactccttga cgggaattgac agtaacgaga atattctoga cagcagatc 1620
gccccatca caatgccaac gaccaattcg ccagtcgaaa gccgaacggt gtctggacg 1680
gatctcgggt tggacacgga cgaagatcac ccgcctcgtt tgttgcgtgt gcgggacgac 1740
gtgattgcct acggcggcga cgaaggcaca ctggtaacct tacctttcgt cacaacgagc 1800
aacgccccaa atgccgacac ggggtccacg cgacccttgg ccgtgcgtcg ctttgatgaa 1860

gacgccatac gtgccgtcgc agtctcggac gacggaaccc gcggtgccgt cggaacggat 1920
agcgggtgcca ctcttttcta ccgttaocgag ttggatggac acgtagtaga cgcaoctgga 1980
aaggggctcg tctcccgaca cggatttgtc acgcacgaca gtgacgacaa caacaacaac 2040
aacaacaact cccaccagaa accctcggca gacttgtttg gatcgcagcc cgacgcctc 2100
gcctttgtcc cacagcaacg tcccggggaa gtcgtccgtc acggaccctg ctttgacgt 2160
cctgtaocgc aactcctctt tctcccgcac tcgcatttcc tcgccattgc caocggaagcc 2220
ggattggccg ttgtttccac cgataocgac agcggcattg gtggtggcag tctggacact 2280
aaccacaacg agaacgtcaa ccaccacaac gtcaaatacc tccaccggga agcccaaacg 2340
gcacacgacg aatccggcat acgoggactc gcctctggc aagcaaagga ctgtcgtata 2400
ctctctcac tcgccatgga cggcgtctc tgtcactggg atgtctctgc tcccactccc 2460
acactctgga aactactgca ccgogagaca gtaccgacgg ttaccaagcc cgacctgggc 2520
gaaatgctcg gtgccgatgc ctgggatcga tccaccatcc ccgtcgcca tcccacgaa 2580
agcatactct ttttgcccgg agaaacctac gtacaggcgc gtcgctaccg caaccacacc 2640
tgggaactcc tacagtcccc tacoggggccc accaatacta ccgacaaaagt acagggacac 2700
attgaagcca ttgtcgcctt ggcctcggca cccaacctc gagatccgta cctcgtcacc 2760
agtggacogc acggaocgag cgtcctctgg aactacagt actctcatca cgacaacaac 2820
aacaacgaca acaatccaaa cgacaatggt gacgggcaca ttgtctttca aaaacaaatc 2880
ctccagacgg attccgcccc aactcatttg ttgtggacac tggaccaacc gacgcaaacg 2940
gaacgtctcg acatggtgac ccgocagcga cactggacta ctctggtagg acgocaccag 3000
attgctccgg cctgtccaac cactgcagtg acccaagaga tctccctccc acaccgcaa 3060
tcagccgatt ccgtgocggg aaaagagaag gaacatgacg cagactcggga cgacagcgtt 3120
gatgactttt ctctgaacaa accttcacaa caccaaaaga atccgtttgt ggacgacgag 3180
gocggagcag acaacgatga cgatacgtc gatacggcct ccgctggaaa actggagacg 3240
acctcaccaa cggacaagcg ccctccaat cttaacagca gcgctctcga agaaccacc 3300
aatgatctag acgacgactc catcgggtgac gatgacgact ccttcacaa cattccgact 3360
ctcaccacgc gacattcoga ttogatocag tggcctgaac cacaaccocg ctttggctct 3420
tcttcacat ccgttgaatt gactcgcgc tttttgtgct ggaatcacat tgggtccggt 3480
acgtttcttc gaggacaggc cggcatcaac ccgacgacga tcgacattca ctttacggac 3540
tcggcatttc gtcggcccgt ttccttcacc gataatatgg gcttcattct ggggtccctg 3600
ggggaagacg gcggaatatt ccgcaocgac ttggoggaag acgaggatat tgatgaggag 3660
gacgacgata tggacggctt gaacgtgtcg gctgctacca aggcgcocgt caaacgttcg 3720
cgcaagggtc ctctgaacaa accgaccggg tcgagcattt actttcatcg ctctgaaacg 3780

ttcggatcct tacgcgacaa ggattggtac ttgacgctcc cagatgggga gggggctttg 3840
 ggggtgtcgt ccggtgaagg atgggcccgc gtcgtaacga gtcgccgttt cttgcggctc 3900
 ttttcttcgg gcggcaatca aggagaggtg ctttggtga acggccacc cgtcaccatg 3960
 gctggacggg gacggttcgt cgcggtgta tatcacgaa gtacaccgtt accagatgga 4020
 acacaaaaac tcgataactt ggtgttgat gcgatggcga atcgcgtagt tgccaagggg 4080
 ccagtgtcat gtattagcgg tgcacgact ctttcatggt tggggttcag caatgatgga 4140
 tctctgctgg ccatggattc ggatggtatg ctgtcaatgt tggtttgcc atcatccttg 4200
 gatgcccagg gaccgacgga aaaacactgg gaatggatgc caatgctgga cacggtggg 4260
 ttacgtaa atccgggacga ttccttctgg ccggtcacag tttatgacgg aaagttggtg 4320
 tgcgtcccgc tcaagggtgg gatgaagcat cctgatgccc tgcgccgtcc cgtcacggcc 4380
 gctctcggct ttcgtcttcc cctggcccgg ggtcctttga ccaagacgca cacggttgaa 4440
 gagcttgccg tgcgcgccgc gattgcgcta gggcagaaaa aggcaattca cgagattagc 4500
 cgggaaggcg acgaggacga cgaggacttt gaaaaagaat accgttcctt ttccgcccac 4560
 gtggacaagg tcacgctaaa aatgtttgca gcaatcggcg aagccggtaa attggagcgc 4620
 gctttggatt tgggtggagcg tttgcatttg gaaaagagct acgacttggc catgacgatt 4680
 ggcgaccggc accgcaaact tgtcgatttg atcgaagagg ccaaggatcg caagtttggg 4740
 gatccgggat cgcaccaagc cgagtttacg accaaagcgg aatccccgaa ctatcaacgc 4800
 ccccgcatct ctccagattc ggcggtggca aaacgcagtc ttgacgatga ggaagaggac 4860
 gtccgaagcc gtctcgtgcg tcgcaaacca acgtttgcct aa 4902

<210> 46
 <211> 2976
 <212> ADN
 <213> *Phaeodactylum tricornerum*

<400> 46
 atgtcaacaa cactctccgc cgttcaattt cacttccaag aaacgacaac aacagcagca 60
 gccaacgaga acggcagcgg tgacgagaag ataatatcaa cggcagcaac gacgacgacg 120
 acgacgacgg agggatcatt agcagcggag gaaggagcaa tccagtctcc tagaaagatt 180
 agcgagggag actacacact cggcgccttc ctgaccagca gggcggtgct ctcgctctcc 240
 agagtgggtga cgaaggaatc tctcctcggc aacaacaaca acggcagcga acaaacacc 300
 gctactggcc ccaaagcctt ctcgctcctt tctcccctcg atcacaacga cgggggatcg 360
 tttctgttct ctgctccoga gaccaccttc cagtcggccc cgctcaacag cgggtggtagt 420
 agcagcgcga gtcgctgtgg cgaacaagaa gtcgatcaag cgtttgctgc cgagatcgca 480
 aagattgact tttcgggtgc ttgtcctttg aactcgtttc atcaagtoga ttacagtact 540

accaacgacg	gaaagacaac	tccggtagaa	gaagagacac	totaccttcc	tctggaagat	600
tccaacaaca	acaacgcagt	agtatcagcc	tacaactctc	ccatcccccc	actcttttca	660
tcttcatcag	ctccaattac	aatcaacgat	ccactcattt	cattaagatc	ttccagtcca	720
ttactctaog	ctgcttcttt	caatacacca	caagaggaaa	acaccagttg	ggctgctggt	780
ggtggtgcta	catccattgt	ttcaacgggt	ctttcggttc	aagttccggt	agaagtagcc	840
gcagtgtcca	cggcagcagt	tgttgcaagt	aatagtgctg	ttggtgtaac	tccaccaact	900
actttggagg	aggacggtag	tagtattagt	attgccaaga	agaaaaagag	taaggctccc	960
aaaacacaaa	agtcaaggac	caagtctaag	caaccaacag	caacgtcggg	agttgtagca	1020
ccaatcgcaa	ggcactcggc	aactacgaat	actacaatgg	atcgacacaa	catgggatct	1080
gcaccatccc	cattgggaac	aacagccagc	acaatgtcca	tctccattcc	ccaccacacc	1140
taccaacacc	aacaaggagc	ggtagacgaa	ccogaaacac	cctcctcctc	cggctctccc	1200
tccggcaaca	ccgtatccac	cctcaacccc	aacaccaacg	ccgtcgaaaa	caccggccgt	1260
tggacagcag	aagaacatcg	tctcttctc	caagggttgg	aacaacacgg	gaaaggatgg	1320
aaaaagattg	ccggtcttat	taaatcacgt	actgtggttc	agattcgtac	tcatgctcaa	1380
aagtactttc	aaaagttagc	gaaggctcgt	gcgggggatg	ggagtgggat	tccaatgatt	1440
ggtggtggtg	cgggggagga	tagtccggag	ttgggtcctc	aagcagcagt	ggcggtgagt	1500
aataccatga	atagtagtgg	gggaggaaaa	gcaggagggg	ggttgcatat	gttgccctct	1560
gogaatacgg	tgacgatgcg	tactgtgaat	caacagcatc	atggaaatca	cggtaacatg	1620
tccttgggga	cggattctgt	gagtatggcc	gcggcagctg	gtcagattga	cattgcgtcg	1680
ggagtttcca	caagcagtgg	gggagcttct	tcgaggaata	ccaccaccac	ttccggggga	1740
ttgaagcgtc	gtaccaatgc	caagtcagcc	ggaggggggg	gagggacgaa	acgtcgtgcc	1800
attggatcag	tggtgaggag	tgcatgagg	gaggggagga	atgtcaagcg	ccaaaagatt	1860
gocgaggcta	ggcgcaatgg	aggtgctaca	gctgcgacaa	gtggaggaga	tggaggggtg	1920
cccaatccac	tcccagccat	ttccaacatt	ttggatccgt	atgtcccatc	ggtggcatcg	1980
gctgcggctg	ttgctgggac	gaagaaggga	cggggaagac	agcagattgt	gcagacggcg	2040
actcatggtt	ctttgcccc	ggctgctttg	gaggatgcag	tgtttctgtc	cctaacacca	2100
gcaccaggag	ctccactctc	ccatccttcc	gcctcgtcac	aaccaatcag	tgatcccctt	2160
gcacccaacc	aggtcaagat	gcccgtagct	caogctcact	acctccaaca	gcaggcagct	2220
caaaacccaa	gccccacggg	tgtgaccgag	atgacctttc	cttcgtgggt	ggatcccaac	2280
aatcctcctc	agtggataaa	cgaagggggg	gacattgaca	atctcctoga	cgaagcggag	2340
gctttagatt	ggcttacgga	tacgggagac	atcaacgaga	cgtatccacc	agcggttgct	2400

gccactcaag ctgccgtggc gaatgatccc tctgattatg agccaacacc cgtcaataac 2460
tattettacc acgttgacaa ctcaggtgcc accactccat cactggcacc tctogatecc 2520
accgcogtca aaatggaaga tcacaccatg gttcacgacc ctaccggaat gatgcatcca 2580
tctgctgact cgctgtcctt cttgggtgat cctcccagag agggagttac ctcagctgaa 2640
ttgccagcgt tccttaacga tgactcccc cagcacgggg tgcctccat gccctttct 2700
tggcgactg aagcagtgga gtctcact tttgaaaaca aaattaacac tgcogttct 2760
gacggaaatc tcattgggtt cccgatcta gacatggag acgaacaggc atttgtctcg 2820
gcgctattgg acaactctgg gcagagcact ttgtcgttc ccaagttgaa ttcagagttg 2880
cacatcggga gtctgtcagg gattggattg agtggcggtg gtgttagtag cgatgccttg 2940
ggagagcctt tggaggatga tcacctgat gattga 2976

<210> 47
<211> 1227
<212> ADN
<213> *Ectocarpus siliculosus*

<400> 47
atgacgcga agaacacggg gogatggacg taogacgagc acaggctctt cctgcggggg 60
ctggagctgc acggcaaggg ttggaagaag atogcgtcgc tcatacaagac gaggacgggtg 120
gtgcagatcc ggaccacgc gcagaagtac ttccagaaga tcgccaaggc caagcagaac 180
ggggagcacg gggacgtggc gatggactcg aaggggcacg ggtctcgaag aaaaagcaga 240
gggcgagaa gaatggaaga cctgctgagg ggtcgacgg cagtggcccc ttcctgcag 300
ccctacgtgg cgggggcagg ggcggtggag accgggctct accgcttctt gtccccgata 360
acaatacagg acctggagcg gccaccagct gcggcgggac agggggggcac agcaggcggg 420
gtggcgggg gggcggtcg actcggcggg aacggcgggg cttecgcttca gtcccagctg 480
gccgggcacg caacgtcgtc ggcaggagcg tcgccgacaa cgacaacggc gggagcggca 540
atggcgggcg cagcggcggc ccttcctcgg aactggtacc gagcaggccg tggcgtggac 600
gacctgctca ccgaggctga ggggctagac tggctggccg acagcggggg tgcggtggcc 660
gtcacgtcgg tgtcggcgg taaccggcc gccgcgcgc cgcctcccc cgcctccgcc 720
gccgttgccc gtgcccaacg cggcgtcgcc agggctcgcg cgtcgtcggg gctgtcagac 780
cagcctttgc tcaagaaggc cgggtccacc accgtggact cgtcggcccc gctggccacg 840
ggcagcaga tgaacacctc ctgaccgct gccgctgcca tggcgggcat cgcgcggccc 900
tcggccagcg ctgcccgat gacgcctgct tcttggctgt tgcgcgcgc cggccgggtg 960
gcgtgccgc cgcctgcgc acctccaacc gctggagtg ctgtggacga gctgaccccc 1020
tccgcttgc agggctcggg tgccgaggag gggcacctt cgggcttgtc ggacctggaa 1080

ES 2 682 279 T3

agcctcatgg aggaccaaga gctcccgggc tctcagccgt ttggaccagt gtccaccgct 1140
gccaagtgg ccggcgccgg cggtgatggc atggacgacg gcatggttga cgaagacgca 1200
ttcgtcaacg cgcttctaga cgctag 1227

<210> 48
<211> 5414
<212> ADN
<213> Secuencia artificial

<220>
<223> Construcción inactivada de LAR1

<400> 48
caacgtcttt ctctcggtaa cgtcggcgcg aggacagggt gtgcccaacg cctgcaggat 60
gccaaccatc tggatgtcgt gacgtcgccc tctctcgaag agtgcctcgt ttccgtagtc 120
gcggttgcgc acagtatatt tcacagcact tcaccaacca cgctagaaaa atgggtgcgc 180
ctggtctttc tatcagtgtc caaagaaagc acctgtttc gtctttgaac gtgtatcatg 240
caagatactt gatcgaggtc gcgatagaga cggccatctg ggtccgggtt cttgcgccat 300
tcggaaccgc aaaaccgtaa ttatgagggg agcaagtggc tctctccccg gcgggcttgt 360
tggacggact gctttgatct tcgtaaagcg cccattttc gcacgcgctt tctacttget 420
agccttctta agtcaagggg gagtgcaggc gaocgcgoga ggctgtgagt gcaggggctc 480
ggacatcatc aagcagcaac ttccgagcaa caaggtcaag atagaaacag aaggcgagca 540
gctgaagaag ggtatatagt gcttttgtgg ctatgtgcac ccatccaagc tttgctaacc 600
ctgtgccaaag caacctcgt cacaagctct tctcaaatgc ccttacttta tttcccacac 660
aggagcgcgg gggcgtcttg ttttgcctcc atttcttggg aagcgaatca ggaaccaaag 720
gtacacgaga acgctctcca gtacacagga gacctaattt gtccgattta ggcaaacctc 780
tcttttggca cgtgggcacc atcttgtgct cgtgtacgaa ggtgttcccg tagtttcaga 840
gtgtgtactt gtggtcgtgt cgtgtgaatt gcccgctctc tcccacctc taagaaacaa 900
gatgcctcgt tctgcagtac gggcattggt tcacttcttc cttegggtgc gagactcgac 960
ggacaacagt cacacacaca caacacacac agtgtttttc actgcttttt cattgaattc 1020
tgcggtgttt tagtgcggtg gcgggtgtct caagggctca ggtgactggc gcttctccac 1080
cccctacact ttttgaacga gaacgaggag tgctgcgtct tcagacaaac caagcaccat 1140
gtcctcccc aagaacattc tggccccgc aagcttgtcc ctaaacaatt acaataaacc 1200
cagccacgat ctggcagcc caaagacgca acaccatcac catggcctgc atcatcagca 1260
ccagcacaag cagcagtacc agcaccaaca gcaactgcag cagcccatg ttctcgggtg 1320
caaatcagta gccggttcca acaagatcct tcctttcacc tcatctatgg atgaagttaa 1380
atatgccgct ggccggctga tcaagcccgg gacgcaaggc ctgggcagca tgctctctac 1440

acccttgaag coctoggtt ctcttattgc atogtgggtg gtagaggccg atgcaaagca 1500
 acaacaactg ctgaaggatt ccctcaccgc cgacctcaaa ttgctcctgc acgaattcga 1560
 acgcttccag caggccactg cagcagcagc gggaacagga ggctgggcg aggaggaggc 1620
 agctgagcga tcgacgaaag tggagttctt tttgggatac attgagcgag ttctccacga 1680
 tttggcaggg gctgacgcat cgaaactgca ggatctcgag gtgcggatca agacaagtct 1740
 actaccactc aagggccagg tggtagcca gcttgcggcg caaaataata actccctcc 1800
 tccccataag gagcagcagt catcttggtt ccatecttct tccacctgct cctccctctc 1860
 ctctcttct tccgtctcca gctgacac aactctctct ggatctcctt tggcaagggg 1920
 ggagacggtg atgagcacct acggtccctt taccactcc cgcgtcgag cacgtgcagg 1980
 tgtacagatt gaaggaaaca atggagatat ctttggcagt tgaaaaccgt gttogaatca 2040
 tgcttttcta ctctccaact gagaagaaat ttatagcgcc atgtcgctt tgaactaccag 2100
 gcttaggaag gcctcatcac aagctggatc gtttogaatt aagcaggcac tgaagccaag 2160
 cttgcaagac agccacctt taattcctc aaaacacttt ctcaattcag cccggtaaatt 2220
 atgccgattc acagcggcca agatagaggg gaggttagca agaattgtgc gatccctccc 2280
 cagtcgttgc ctgcacaca acctaggcct tcaccttcc atggaaaatt gagaagtga 2340
 tattggtttt cttacggcat atcagatgaa atcatgacct ctaaaccatga agagctgcag 2400
 gcaaaacacc tgctctggac gagcagatg aaatctcgag aaccgcctg acttcagttg 2460
 atcccgcatg atgacggccg ccattgaaat aagccacctc actttattct agcaccgatt 2520
 tccaccgttg tgagggccga acgaggacaa tttcgtcgca aacaagcacg aacacgcaca 2580
 cgattagtag tacagacgag cagatcgatg gcatgoggca cggctctcg tctctggcga 2640
 ccaggacaac ggagcagagg gaggcctgcc gagttccgag gggcatttta gtccaaaatt 2700
 gtgttgacac gtgaacaagt ggcttgaaaa gaggaaggaa atgcctgggt ttcccttcca 2760
 gagcgggaac tcgcttgtgc gtcacctag ctaccatgg tcccttgtg ggggaggctg 2820
 tttcgtcta ccgaatgtgt ggcgtccat gcatcttctg cctcccaaac caccaacatg 2880
 agcacgogaa ggaaggagaa aaaagtggcc gcaacgttct cttctcatat ttattgtctc 2940
 atcacaaca taggtacata atacaacaat catggatccc cgggtaccga gctcgatggc 3000
 caagccttta tcccaagagg aatccacgct gatogaacgt gcaactgoga ccatcaacag 3060
 catacctatt agcagaggact actcgggtgc cagtgcagcc ctctcgtccg acggtcggat 3120
 ctttaccggc gtgaatgtat atcatttcac cggagggcca tgcgcggagc tcgtggtcct 3180
 cggaacggcc gctcgggtg ctgcgggaaa tctgacgtgc atagtggcca togggaacga 3240
 aaaccgggc attctgtctc cgtgcgggag atgtcggcag gtgctgctt acttgcaccc 3300

gggatcaag gcaattgtca aagattccga tgggcagccc acagcgggtg gcatcagggg	3360
gttgcttccc tctggctacg tctgggaggg ttgacagcag cagttgtgct acgcatgcac	3420
ctccctgcct cgctcctcc tccacctcct cctctcttc tttctctcc cctttctctt	3480
cctcttcggt ggcgacagtg gcagcagcag atgtcagcaa gcctctcctt cgtcaggttg	3540
agtatcagtg cgggtccttg gcgacacct acaccgctgc ctctccctc aatccctggt	3600
ggccctcga gcgacaggag tgtcccaagt gcaagaaggt tcaagttcca cgcattgaca	3660
taaacctgcc tgccaacacc atggaatacc acccggcttt gctcgcggag gaaggcagtg	3720
atgacgatga tgatgaggtg ggaggcggga gggaggaggt gatgatgatg atgccggggg	3780
gaggggatgg acatggacat ttggaggaaa gagaggaggg agagacgagt gagaaagggg	3840
gcggtggaag tagcgtacta gaggaggacg aggaagcagt gttgagccct atgcaagctt	3900
cacagctggt gaggttgtg gagcatgocg ggacatgccc gggcaatcat gctgccgaga	3960
aacaccaggc tgtgtgtacg agtgccaagt atctcatggt gcatgtgagg gattgtgatg	4020
gcaggacggt agatggggag gcttgtgat tctcgtggtg caggccttg caaacacttg	4080
tgggacacct ggtccggtgt tacgaagcog agaagtgtca gatctgctgt ttctcacacc	4140
aagaagagga ggagaaagtg gaaaagaagg tgatgatgag tgtggaggag atgattgagg	4200
agaaaggaat gagggctgac acctacagga gcttgacgag cttgagctga gacaaggtgt	4260
gatgagaaga ggaggggaag aaggagggga catatctgta tggccgcatg tattgtggtg	4320
agaaatgtgc gtgtcttgat gtgatattt taaatgtctg agtgagtga tgtccgagtg	4380
agtgaatgtc cgagtgagtg aatgtccgcg cttgtgtgta tcggtaggta cgagcgaggg	4440
aaaagagtga tctaattgtc ctgcctgtcc atcagtctga tgacaactga gaaaaaacat	4500
cgattgggag agtgaaagta gaaatgatgt atgaaaccaa ccaggagtga cggcaacgat	4560
gctggcgctg catgtgcaaa cgagcgcagc attaatcga cgagtacatg gacaagaaga	4620
caogtgtatg agagagagga aatgagatag agagagggag agagaatgag agagtaaaag	4680
atgatttatt tatttatatc gaaaatctat tgcccgccca tctatgattt agatctatta	4740
cccacctatc tacatagcga tcgagtgcg aataaataag gcatctacat atatgataaa	4800
ttctacgct tgattogtgc tttogattct atatgccatt ggtagataga gaccaatagt	4860
cagcggcagc atcagcgcga gaacgatgaa gcctgtcttt ctgatgtacg agtagtgtga	4920
aggaagagat caaagagggg atgacgtatc aggacaagta ggaagaggcg ttcgtttggg	4980
aacgggaaat atgacgaatg agcatgcacg tggcaactcat cgcgtgaagc agaggggtct	5040
gggggaaacc ctggaagcaa gaagacgacg acaagaacga tgatcatgcc tatcaataca	5100
ctgagaagaa atacacaagt aacagatgta gttccatagg acaacaaaa cagaatcaga	5160
ggtcgatatg aaaacccaaa aaaaaacca ataagattca atttagaac acccaaaaat	5220

ES 2 682 279 T3

cagcgagcag aagttagcac gtaccaoggt aacaacagag aagagcaatc gacgaccata 5280
 actaaaaata gaaaacaaaa aataaaaaag atagagaaaa cccacgccag tccgcccccc 5340
 agaaattoga gggctaoggt gcoctgacga atacagcgag ttttcottga gaagggtaat 5400
 aacacacaag aaac 5414

<210> 49
 <211> 3415
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Construcción inactivada de LAR2

<400> 49
 ctgtatgctc gtcgtcgatc cagttaaagc gcgatgctta ccagaccccc tcatctagta 60
 aattgtgctt ggatctgagc catccccccg agtacattgc ttgaccogt cgctgtttaa 120
 tctgacagat tgctataata gcaccaatga acaatcattt ttattttcac acgcaatctc 180
 acaacatcct ctctctcccc tcaaaagcag gatgcaacta ctaccgoggc cggcagcatg 240
 atgcttcccc tcttccctc catccctgct tccgctacca gcgctcttt taccgctccc 300
 tcagcgtcct ccacgaccac caaatctccc aaaggtactt cctccatgac tcagcttggg 360
 gagacagggc gagagaatac tggcogatgg acttgtgaag agcatgtgct gtttcttaaa 420
 ggctagaaa tgcaocggcaa gggttggaag aaaatcgcaa agctaataca gacccgaacg 480
 gtggtccaaa tccgcaogca cgcgcaaaag tacttccaga aattggcaaa agccaagaag 540
 aacgggcacc atggtgatat gctcggatg gaaggctcac actttggggg aaaacgtgct 600
 aaatttaccg gaaagcgagc tgggcttgtc tataattogt atttagtagg tgccgaggcc 660
 acctctgagg ctatctcccc ggcgttgagc acgtttatgc cggcgaactt ggggatggag 720
 ggcgagcgtg taggccttat gacggataag gaggaggatg cagcaatcga gaagggactt 780
 tatcgtttcc tctccccogt agtgctggat cccgccacgc gtaatctgga cgcctccgct 840
 cctgagatct tgcccttacc acccagcact ccagcgatgg gcgtgcacca taccagtagc 900
 agaggtagca gcagaggggg gttggatgga gagacaacgg gagaggagga cggcgggaagc 960
 gattcgattg taatggggga tggagggagc gatcaagatg cagagtcgtg caggtgtaca 1020
 gattgaagga aacaatggag atatctttgg cagttgaaaa ccgtgttoga atcatgcttt 1080
 tctactctcc aactgagagc aaatttatag cgcctatgog cttctgacta ccagtcttag 1140
 gaaggcctca tcacaagctg gatcggttcg aattaagcag gcaactgaag caagcttgca 1200
 agacagccac cttttaatto cctcaaaaaca cttctcaat tcagccoggt aaatatgcoo 1260
 attcacagcg gccaaagatag aggggaggtt agcaagaatg ttgcgatccc tcccagtcg 1320

ttgcctcgca	cacaacctag	gccttcacct	ttccatggaa	aattgagaag	tgaatattgg	1380
ttttcttaag	gcatatcaga	tgaaatcatg	accocctaac	atgaagagct	gcaggcaaaa	1440
cacctgctct	ggacgagcac	gatgaaatct	cgagaaccog	ccgtacttca	gttgatcccg	1500
catgatgaag	gocgccattg	aaataagcca	cctcacttta	ttctagcacc	gatttccacc	1560
gttgtgaggg	ccgaacgagg	acaatttcgt	gcgaacaag	cacgaacacg	cacacgatta	1620
gtagtacaga	cgagcagatc	gatggcatgc	ggcacggtct	cgogttctcg	gcgaccagga	1680
caacggagca	gagggaggcc	tgccgagttc	cgaggggcat	tttagtccaa	aattgtgttg	1740
acacgtgaac	aagtggcttg	aaaagaggaa	ggaaatgcct	gggtttccct	tcgagagcgg	1800
gaactcgctt	gtgcgtcatc	ctagctaccc	atggctccct	tgtgggggag	gctgtttcgt	1860
cctaccgaat	gtgtggcgct	ccatgcatct	tctgcctccc	aaaccaccaa	catgagcacg	1920
cgaaggaagg	agaaaaaagt	ggcgcacaag	ttctcttctc	atatttattg	tctcatcaca	1980
aacatagta	cataatacaa	caatcatgga	tccccgggta	ccgagctcga	tggccaagcc	2040
tttatcccaa	gaggaatcca	cgctgatcga	acgtgcaact	gcgaccatca	acagcatacc	2100
tattagcgag	gactactcgg	tggccagtgc	agccctctcg	tcgacgggtc	ggatctttac	2160
ggcgtgaa	gtatatcatt	tcaccggagg	gccatgcgcg	gagctcgtgg	tcctcggaac	2220
ggccgctcgc	gctgctgccc	gaaatctgac	gtgcatagtg	gccatogggg	acgaaaaccg	2280
cgccattctg	tctccgtcgc	ggcgatgtcg	gcaggtgctg	cttgacttgc	accoggggat	2340
caaggcaatt	gtcaaagatt	ccgatgggca	gcccacagcg	gttggcatca	gggagttgct	2400
tcctctggc	taogtctggg	agggttgagt	ggtagcaaca	ctaataagat	gaactcagcg	2460
gcgcctgtgc	tgaacatgac	agggttgctt	ggggctggag	ggctttcagg	gtggaagggc	2520
aagggtagcg	acactagcga	aggtagcagc	agcaacggca	gcagcaagaa	catggctttg	2580
acggcaaatg	cgtcggcggg	atgtggacaa	gggagctggg	gagtgcgggg	gggggcacag	2640
aaaaagcagc	agcagcagca	acatgaggtg	ccaacgcagc	agcagcaggt	accaacgcag	2700
cagcagcagc	ttcacggcat	tcacgtgaag	gaggaagggg	tggagctcct	gagagtcag	2760
gcggatagag	ggactgttca	cggtcacgtc	catgaggagg	atggctttgc	ggcgtttgac	2820
ccccatcatg	tcgagctgaa	ggaggagcac	tcccaccatg	atttgttgc	ggaagagctg	2880
ccccacgaca	gcaaccacga	tgacgccctg	gcccattattg	tgttctcagt	gaatggagag	2940
tcggatcttc	attccttgcc	gaggggtgca	ggaggggggg	gagcgcagtg	gcatgcgccg	3000
ccggttctgg	tgggggggag	acagcatcac	tatcatcata	atgacaatga	tatccatttg	3060
catgcgtaag	acgcgtattt	ggaagaggaa	ggggcggatg	gcgggcatgg	gttgttgttg	3120
ttggaggatt	tggatggggg	aatcgagttt	tagagatgga	gagggagaag	ggaagtgtgc	3180
atgtgtaaga	aaagaaagga	ggaaattcaa	ggaagagctg	gaagggagag	aaaaagatgc	3240

ES 2 682 279 T3

ggctataatg gtgttcatga aggcacattg ccacaaaaat aaacgaacgg ggagaagagg 3300
 cgtttggaaa gaggagagag atgtgaaagg acctgaggta gagtgttggg tgaattacac 3360
 taaacgccaa ccttgtgtcc ttgtgtcggc ogtgcaogaa cgatccagac gaagg 3415

 <210> 50
 <211> 4553
 <212> ADN
 <213> Secuencia artificial

 <220>
 <223> Construcción de ARNi de LAR1

 <400> 50
 cagggtgcc ctcccagggt caatacagct cttaaagggc cgcaagcttg cggccaacat 60
 gtcactcaga ggtacgtcgg tcgcaactgc gtgcacttcg tggccgagga gcaggactaa 120
 agatcttcta gagtgggggc ggcggccgc ttgagcaga catgataaga tacattgatg 180
 agtttgaca aaccacaact agaatgcagt gaaaaaatg ctttatttgt gaaatttgtg 240
 atgctattgc tttatttcta accattataa gctgcaataa acaagttaac aacaacaatt 300
 gcattcattt tatgtttcag gttcaggggg aggtgtggga ggttttttaa agcaagtaaa 360
 acctctaaa atgtggtaaa atcgataagg atctgaacga tggagcggag aatgggcgga 420
 actgggcgga gttagggggc ggatggggcg agttaggggc gggactatgg ttgctgacta 480
 attgagatgc atgctttgca tacttctgcc tgctggggag cctggggact ttccacacct 540
 ggttctgac taattgagat gcatgctttg catacttctg cctgctgggg agcctgggga 600
 cttccacac cctaactgac acacattcca cagccatag cacgtgaagg gogaattcgt 660
 ttaaacctgc aggactagtc gtcatgataa taatggttc ttagacgtca ggtggcactt 720
 ttggggaaa tgtgocgga acccctattt gtttattttt ctaaatacat tcaaataatg 780
 atccgctcat gagacaataa cctgataaa tgcttcaata atattgaaa aggaagagta 840
 tgagtattca acatttcctg gtgcocctta ttccctttt tgggcattt tgccttctg 900
 tttttgctca ccagaaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac 960
 gagtgggta catogaactg gatctcaaca gcggtaagat ccttgagagt tttgccccg 1020
 aagaacgttt tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc 1080
 gtattgacgc cgggcaagag caactcggtc gcgcataca ctattctcag aatgacttgg 1140
 ttgagtactc accagtcaca gaaaagcacc ttacggatgg catgacagta agagaattat 1200
 gcagtctgc cataaccatg agtgataaca ctgcccga cttacttctg acaacgatcg 1260
 gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta actgccttg 1320
 atcgttggga accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc 1380

ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt 1440
 cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct 1500
 cggcccttcc ggctggctgg tttattgctg ataaatctgg agccgggtgag cgtgggtctc 1560
 gcggtatcat tgcagcactg gggccagatg gtaagccctc cegtatcgta gttatctaca 1620
 cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct 1680
 cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt 1740
 taaaacttca tttttaattt aaaaggatct aggtgaagat cctttttgat aatctcatga 1800
 ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca 1860
 aaggatcttc ttgagatcct tttttctgac gcgtaatctg ctgcttgcaa acaaaaaaac 1920
 cacgctacc agcgggtggt tgtttgccgg atcaagagct accaactctt tttccgaagg 1980
 taactggctt cagcagagcg cagataccaa atactgttct tctagtgtag cogtagttag 2040
 gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac 2100
 cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt 2160
 tacgggataa ggccagcggc tcgggctgaa oggggggttc gtgcacacag ccagcttgg 2220
 agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcggccagc 2280
 ttccogaagg gagaaaggcg gacaggtatc oggtaagcgg cagggctgga acaggagagc 2340
 gcacgagggg gcttccaggg ggaacgcct ggtatcttta tagtctgtc gggtttcgcc 2400
 acctctgact tgagcgtcga tttttgtgat gctcgtcagg gggcgggagc ctatggaaaa 2460
 acggcagcaa cgcggccttt ttaocgttcc tggccttttg ctggcctttt gctctacgta 2520
 tctctcaggt cgactctaga gaggattgat ttccgagtca aagattcgta tagctgcaaa 2580
 tgcactcatt gatgagatac ctgaagagga ggatgacaaa cagtacttgg caagtgaatc 2640
 ttttcgggaa ttcatacggg attttgatgg actggatgca ctaatggagt cggcttcacg 2700
 gcogttcggg ggaggagaag atccaaggat gaatgcattg actttgttgg gagaggtgga 2760
 ggaatcattg aggatgtttg tcaagattgc gaagtgatga ttgcatgcat cgcgctgatg 2820
 aaagaaatgt gctcaatgtc gacgaaccat aggcatgat actggagaga atgaagacat 2880
 agtaacgcag gagacgtcaa ctgaatccaa aactcgatgg attccaata caaattatgt 2940
 aaacttatca ttctttatcc cagccaaggc agcctaaact cgtaccccca aagtagtggc 3000
 gccaacagcg cgacagagtg aggatcttca accctcccag acgtagccag agggaagcaa 3060
 ctccctgatg ccaaccgctg tgggctgccc atcggaatct ttgacaattg ccttgatccc 3120
 cgggtgcaag tcaagcagca cctgcogaca togcccgcac ggagacagaa tgccgoggtt 3180
 ttogttcccg atggccaacta tgcaogtcag atttcggca gcagccgag cggccgttcc 3240
 gaggaccacg agctccgccc atggccctcc ggtgaaatga tatacattca cgcgggtaaa 3300

gatccgaccg tcggacgaga gggctgcact ggcaccgag tagtcctcgc taataggtat 3360
gctgttgatg gtgcagttg cacgttcgat cagcgtggat tctcttggg acaaaggett 3420
ggccatggca tgattgttg attatgtacc tatgtttgtg atgagacaat aaatatgaga 3480
agagaacggt ggggccactt ttttctcctt ccttcgcgtg ctcatgttg tggtttggga 3540
ggcagaagat gcatggagcg ccacacattc ggtaggacga aacagcctcc cccacaaagg 3600
gaccatgggt agctaggatg acgcacaagc gagttccgc tctogaaggg aaaccaggc 3660
attccttcc tctttcaag ccactgttc acgtgtcaac acaatttgg actaaaatgc 3720
ccctcgaac tcggcaggcc tccctctgct ccgttgctct ggtcgcgag aacgcgagac 3780
cgtgccgat gccatcgatc tgctcgtctg tactactaat cgtgtgcgtg ttcgtgcttg 3840
tttcgcacga aattgtctc gttcggccct cacaacggtg gaaatcgtg ctagaataaa 3900
gtgaggtggc ttatttcaat gggggcgc atcatgccc atcaactgaa gtacggcggg 3960
ttctcgagat tcatcgtgc tcgtccagag caggtgtttt gcctgcagct cttcatgttt 4020
aggggtcatg atttcatctg atatgcgta agaaaaccaa tattcactc tcaatttcc 4080
atgaaaagg gaagccctag gttgtgtgc aggcaacgac tggggagga tcgcaacatt 4140
cttgctaacc tccctctat cttggcgcgt gtgaatcgc atatttacc ggctgaattg 4200
agaaagtgt ttgaggaat taaaagggtg ctgtcttgca agcttggctt cagtgcctgc 4260
ttaattcgaa ccgatccagc ttgtgatgag gccttctaa gcctggtagt cagaagcgc 4320
atggcgcct aaatttcgct tcagttggag agtagaaaag catgattoga acacggtttt 4380
caactgcaa agatatctc attgttctc tcaatctgta cacctgcacg ggtaccgagc 4440
tcgcttccat tcaggtcag gtggcccgcc tccatgcacc ggcagcaac gggggaggc 4500
agacaaggta tagggcggc cctcataatc aaagatgagc cagccacgaa gct 4553

<210> 51
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Sintética

<400> 51
gtgaaaccag cactcaatct ctc

23

<210> 52
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Sintética

<400> 52		
agttcgaata tcttgaatc gt		22
<210> 53		
<211> 20		
<212> ADN		
<213> Secuencia artificial		
<220>		
<223> Sintética		
<400> 53		
gggaggctga gatcaagcac		20
<210> 54		
<211> 20		
<212> ADN		
<213> Secuencia artificial		
<220>		
<223> Sintética		
<400> 54		
caggagccct accgtcatgt		20
<210> 55		
<211> 20		
<212> ADN		
<213> Secuencia artificial		
<220>		
<223> Sintética		
<400> 55		
tacttgcagg aggccgagat		20
<210> 56		
<211> 20		
<212> ADN		
<213> Secuencia artificial		
<220>		
<223> Sintética		
<400> 56		
agggagctaa cagcgtggac		20
<210> 57		
<211> 20		
<212> ADN		
<213> Secuencia artificial		
<220>		
<223> Sintética		
<400> 57		
tgccccgaag aagctagatg		20

<210> 58
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 58
 cacgacccag cctaggaaac 20

<210> 59
 <211> 399
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Sintética
 <400> 59
 atggccaagc ctttgccca agaggaatcc acgctgatcg aacgtgcaac tgcgaccatc 60
 aacagcatac ctattagcga ggactactcg gtggccagtg cagccctctc gtcogacggt 120
 cggatcttta ccggcgtgaa tgtatatcat ttcaccggag ggccatgccc ggagctcgtg 180
 gtccctcgaa cggccgctgc ggctgctgcc ggaaatctga cgtgcatagt ggccatcggg 240
 aacgaaaacc gcggcattct gtctccgtgc gggcgatgtc ggcaggtgct gcttgacttg 300
 cacccgggga tcaagcaat tgtcaaagat tccgatgggc agcccacagc ggttggcatc 360
 agggagttgc ttccctctgg ctacgtctgg gagggttga 399

<210> 60
 <211> 1602
 <212> ADN
 <213> *Nannochloropsis gaditana*
 <400> 60
 gtgtggtgaa gcagggaggt gggagaaggt cctggagctc ttggcggaga tgcottcaat 60
 gggcgtgtca cccaacgtgg tggcatacac ggctgctgtc gctgctgtg gacgaggagg 120
 gcagccagag cgggctttgg ggttgctgcg ggaaatgaaa agcgagggga tcagtccgaa 180
 tgtacaatgc tacaatactt tgctttgggc actggctaaa ggtggggact ggcagcagtg 240
 tctggcgtaa ttggaggaga tgaaggaaga gggcaagacg gatgctcaca gttatcgtat 300
 tgtgatgggc gttgtaaaaa gtaagcaagg gcaagagcat gttatggagg ctcttcaggg 360
 ggatatggag gggctgagga ccgaggcggc gaatgacagc tttggatgcc ggtaaataat 420
 ttttaagatt acggaaagtg tagcttgtaa ttagcaattt gtgagtgtgc atagacctga 480
 tagaaaagat agaaaataaa gttgaacaaa gtatcaattg tgaattcaat tcaaggatat 540
 ataaatagaa tctcgcccag gtgtcaaagt aagttctctc tgctattccc cgtccatggt 600

ttogatctgt actttcctcc atctcccggc caccaaaaag ttagcacaac gtttacggaa 660
 ccagaagcca ataacagtat ttagggtaat catttttagac gtgctaagtt ccacataccg 720
 aaataaatgt atttacaaaa ggtttcactc gtgcgcttga ccagctcaca cgacccccgaa 780
 gcagaagtct tcttttcctt caaacccatgc accgcacagc attgctaaga gaggggagtgt 840
 cactctgttg cttctctcga tgcacgaaaa ggtgctcctc cgagggctcc caacttatct 900
 cagtggcagg gatgtctctc atcacatttc tgctcattcg acaacgtcat gagaagatct 960
 ggtgtctttc tgcggattgc gcctcgcaca acactattgt tcttgggagc gctctttcgg 1020
 cctgacacaa atgcccgctgc cttcaaggcg tgggtattht cctagtgtct ttcacgtaaa 1080
 gcgtgcgcac gtctggccac tatttggtac tcgactgtgc ctccatttgt cctttttatg 1140
 gtgatgtctt tttctttgtc cacttagata caccttggcg tgccgggtggc cagtgtttaa 1200
 gcctacccca agcaggcttc cttgctacca atagcttgcc ggtaaaaatg ctccatcgt 1260
 cttattcggc ctctccattc tghtaacggt ccctcccctc cctcggacag tcccggcctc 1320
 ccctctcgtc ggcaccatag cgaagggtca gtccaccacc agcgcgcgag tctcgtcccg 1380
 agcgcggccc gtggcttggc tcaattgctt ttggaagccg acagcaagca cggaaatgcg 1440
 ccgtagctgg ccacaaaatt cctcgcactaa cgacacctag ttagccaacc acgtcgcagt 1500
 ccattcacca cggctggaca gggcgagcta ggcgacaggt tctttcccgc acctggccag 1560
 accccaccgg gccaccagac gataatgggg atgcctagca ac 1602

<210> 61
 <211> 1134
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 61
 aattcaagtc ggggttgttg ccattgacgg atggtacgac attgaggaca tttttgagta 60
 agctgcttaa ttgtgatcct atgaggataa gcaaaaaatt tgtgggtcag aattgcatcg 120
 gaaagcaagt gtttcgtcgg cggcagcaag cagagatgga caggttgtcc cctgaagaga 180
 ttgaacggag tcgacaagaa cttgccgaac tagagagacg atttttggag cgcgtggtc 240
 agacaaatcg atgtaagaat tctgctgtgg ggacgaagca tgggaaagac tggcgtcgg 300
 gtcaggtggc gccggaggtg ttgcgacgac agcaagagta cctgggcccga ggtgggcagg 360
 tgggtgggtaa tcccctcctc cccccctgga tgctccctcc cacccgggat gcatcgatgg 420
 agggagggga gggcattcaa ggtggagatg gtcattcgaa ttctcgcga gagagggggg 480
 ggatggggat gaccggctgc ggtggtggtg gtagttcttt tcgtttgatg ggtggaggaa 540
 cgggagcagg gagcaagcag cagcagcagc agagccagca gcagcagcag ccgccttttt 600
 cggatttgcc ctgcgcacac cagttgcacc atcctcagtc acatctccat catccacgac 660

ES 2 682 279 T3

gtatcccctc ctcatcgtcg tccaccctag gagggctgaa cacgggtagt ctctcogcga 720
 tcgcgtcggc gaatatgtca cgcaccgatt ctcaggatag cetaacagcg ttaggtggca 780
 caatgtcggc agacgctcct ggacatttgg cggactttcc gogtgtgacc agcctaggag 840
 atcttgccgg actcaacatg tttacctcca gccctttccc ctccacggac aatctcctca 900
 gccttcacca gcagcaacac gcccaaagtg ctcaaattgt tccggccact gctgcttctc 960
 ttgcccgcga tagcaccacc agcttagcag cagcggcagc agcagcagca gtggctcogt 1020
 cgagcacacc gatcagtcgt agcaccagtg gcaacggtaa tggcgggagc agcagcagca 1080
 gcagcagcag cagcagcagc agcaatgcga atggtcttcc accactagcc aatg 1134

<210> 62
 <211> 2526
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 62
 atgatggaga tgccaaccaa ccgcagcggc cctgggacgg actcgggaga gaagtgccca 60
 aatcctccca cggttggtgc acttgtcacg actacgcttt cgcocaaaag tgaatctcct 120
 aagagagttg ctgttcccac gttaggagcg cccgtggcgg ctccaccggc gtcgcocgtg 180
 gcogtggcag caagggggca cggatcctcc ctgcagaatg ggctcccagc ccgcaacggg 240
 gacggtggca gctctttgac gccaaattct ggcccaagag aatttcttca agtggcgagt 300
 aaaccgcctg gcacggagat gcctgctcta gacgcgtcaa caagtggatt gatacataag 360
 ccggcgagta ggcagccgcg cacggctctc tcacagcacc acgatgcggt gatacagcga 420
 gagctgaatg cgatcattgg tcaaaagcgg gaacacaggg accgcagtga agagtogact 480
 gtgcttcttc cgtccacctc tccgatctcg gttcccgtag caagacacca agcagcgcga 540
 cgacccgggc caggccccgc ctgcgatgcc ctccccgect cgcocccgcg acaagcaggc 600
 tcaaacgccca gaggaagcgg gcagacggtc tcgaataggt cgcctccatc ccggaacggc 660
 caagccetta tgctcctac gcgtcactcg acaagggggg ccaccgcgcg agcggcggt 720
 ggtattagca agtccatcga agaggcctgc catgccgccg ccctagagga cggcgtggtg 780
 agctcgaagg ggggaggcgc gctgaatgct ggatacggga atgctaagca aggtagcgtg 840
 gttagcccgg gcgagacaag cgcgctgaca agcgtcaata aagctggcaa ggtcaagaac 900
 ggtctgcgac gagggaagtg gacgcggag gaagaggctt acgcgaaccg tttgattgtc 960
 gagttcaaat ctggattgtt accgctcact gatggaacaa cgtacggac atttctgtca 1020
 aaacttttga attgcgatcc catgogtacc tccaagaaat tcgtggggca aaattgtatc 1080
 gggagcaag tttttcgacg ccggcagcag gcggacttgg atgcctatc cacggaogag 1140
 attgagcgca gtcgacaaga gctagcggag ctggagcgcc gttttttaga gcgcgttgct 1200

ES 2 682 279 T3

caaacaacc gttgcaaaa ttcggttgca ggctccaagc acggtaaaga cagcgccgcg 1260
 gggcaagtgg ctctgaggt gccacgtcgg caacaggagt accttgggcc tgacggccgg 1320
 attggtggga acccogtct tccgccttgg atgctgccac cagctgcaca ggctgcgaac 1380
 gaaggggagc ctgcgtcagc ctcaacaacg ccgcgctcaa actgttcgag cggtatcaac 1440
 agcggcttcc ggggagccgc aggccccaat tcgaccaagg gaacaagcag gcatcatcag 1500
 aatattgcc ctgaacacac aggcgtact ctcatcaca tcagccacct tcatcagcct 1560
 tcgggccttc ggcgcgggcc tggcggatct gacactggga ttttgtcggc aactgcagcg 1620
 acaaacatgc ctgaaacgga ctcacaggac agtctgtcgg tgctagggcg caccatctcg 1680
 gcagatgcac tcggccacct cagcgacttc ccccatacca ccagtttggg tgaccttgc 1740
 ggctgaata tgttcacatc cagccctttt ccctccaccg acaatctttt gagcctgcac 1800
 cagcatcagc accaacagac ggtagcggcg gcaccagag gccatactac tctctcagcg 1860
 cataacgcaa caagcgtggc ggcagcggcg gcagcagcag ccgtcgcccc ttcgagcacg 1920
 tccatcagtc gcaaacacag ggacagtgcc ccccgggcc ccaacgtgag tggacacggg 1980
 tctcagctgc caatgagcac taccogtggc actgtggcaa cggtcacaca aatcctgtc 2040
 ataacagcaa ccgcgaggaa tgcogtggc gcaaacactg ctggatcggc tcatcaacg 2100
 acgacgaccc aaaagcggat tcogcgtgtg gattcggcca caggcttgtc ctccctgcgt 2160
 gtcagcagcg ggctgcctcg aaacacctcc gttgaggact ttctgtcgtc ggtggattac 2220
 ggagatattc ctgcaccaga caaggacctc ctctcgaagt gcgttttccc tcagaacagc 2280
 gctgcacgcg ctgccagac cgctgcctca cagcagcgtc ccgtggcggt cacaggcatt 2340
 gggagcctt cagtcactct cgctggcggt gtgaaggtgg aaggcggcgc acccagga 2400
 ccaatacccg ggggctcaca cgatgttgcg ggaggaaaac gtgatcgagg tgatacctta 2460
 cctacgatat cgacaggaag ttcagcgtg ggtgccaggt tgaagcaacc gaaaattgaa 2520
 agataa 2526

<210> 63
 <211> 841
 <212> PRT
 <213> *Nannochloropsis gaditana*

<400> 63

Met Met Glu Met Pro Thr Asn Arg Ser Gly Pro Gly Thr Asp Ser Gly
 1 5 10 15

Glu Lys Cys Pro Asn Pro Pro Thr Val Gly Ala Leu Val Thr Thr Thr
 20 25 30

ES 2 682 279 T3

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

ES 2 682 279 T3

Leu Thr Ser Ala Asn Lys Ala Gly Lys Val Lys Asn Gly Leu Arg Arg
 290 295 300

Gly Lys Trp Thr Pro Glu Glu Glu Ala Tyr Ala Asn Arg Leu Ile Val
 305 310 315 320

Glu Phe Lys Ser Gly Leu Leu Pro Leu Thr Asp Gly Thr Thr Leu Arg
 325 330 335

Thr Phe Leu Ser Lys Leu Leu Asn Cys Asp Pro Met Arg Ile Ser Lys
 340 345 350

Lys Phe Val Gly Gln Asn Cys Ile Gly Lys Gln Val Phe Arg Arg Arg
 355 360 365

Gln Gln Ala Asp Leu Asp Arg Leu Ser Thr Asp Glu Ile Glu Arg Ser
 370 375 380

Arg Gln Glu Leu Ala Glu Leu Glu Arg Arg Phe Leu Glu Arg Val Ala
 385 390 395 400

Gln Thr Asn Arg Cys Lys Asn Ser Val Ala Gly Ser Lys His Gly Lys
 405 410 415

Asp Ser Ala Ala Gly Gln Val Ala Pro Glu Val Pro Arg Arg Gln Gln
 420 425 430

Glu Tyr Leu Gly Pro Asp Gly Arg Ile Val Gly Asn Pro Val Leu Pro
 435 440 445

Pro Trp Met Leu Pro Pro Ala Ala Gln Ala Ala Asn Glu Gly Glu Pro
 450 455 460

Ala Ser Ala Ser Thr Thr Pro Arg Ser Asn Cys Ser Ser Gly Ile Asn
 465 470 475 480

Ser Gly Phe Arg Gly Ala Ala Gly Pro Asn Ser Thr Lys Gly Thr Ser
 485 490 495

Arg His His Gln Asn Ile Ala Pro Glu His Thr Gly Ala Thr Leu His
 500 505 510

His Ile Ser His Leu His Gln Pro Ser Gly Leu Arg Arg Gly Pro Gly
 515 520 525

Gly Ser Asp Thr Gly Ile Leu Ser Val Thr Ala Ala Thr Asn Met Pro
 530 535 540

Arg Thr Asp Ser Gln Asp Ser Leu Ser Val Leu Gly Arg Thr Ile Ser
 545 550 555 560
 Ala Asp Ala Leu Gly His Leu Thr Asp Phe Pro His Thr Thr Ser Leu
 565 570 575
 Gly Asp Leu Ala Gly Leu Asn Met Phe Thr Ser Ser Pro Phe Pro Ser
 580 585 590
 Thr Asp Asn Leu Leu Ser Leu His Gln His Gln His Gln Gln Thr Val
 595 600 605
 Ala Ala Ala Pro Arg Gly His Thr Thr Leu Ser Ala His Asn Ala Thr
 610 615 620
 Ser Val Ala Ala Ala Ala Ala Ala Ala Val Ala Pro Ser Ser Thr
 625 630 635 640
 Ser Ile Ser Arg Asn Thr Gly Asp Ser Ala Pro Arg Gly Pro Asn Val
 645 650 655
 Ser Gly His Gly Ser Gln Leu Pro Met Ser Thr Thr Arg Gly Thr Val
 660 665 670
 Ala Thr Val Thr Gln Asn Pro Val Ile Thr Ala Thr Ala Arg Asn Ala
 675 680 685
 Gly Gly Ala Asn Thr Ala Gly Ser Val Ser Ser Thr Thr Thr Gln
 690 695 700
 Lys Arg Ile Pro Arg Val Asp Ser Ala Thr Gly Leu Ser Ser Leu Arg
 705 710 715 720
 Val Met Ser Gly Leu Pro Arg Asn Thr Ser Val Glu Asp Phe Leu Ser
 725 730 735
 Leu Val Asp Tyr Gly Asp Ile Pro Ala Pro Asp Lys Asp Leu Leu Ser
 740 745 750
 Lys Cys Val Phe Pro Gln Asn Ser Ala Ala Arg Ala Ala Gln Thr Ala
 755 760 765
 Ala Ser Gln Gln Ala Ala Val Ala Phe Thr Gly Ile Gly Lys Pro Ser
 770 775 780
 Val Ile Leu Ala Gly Gly Val Lys Val Glu Gly Gly Ala Pro Thr Gly

ES 2 682 279 T3

gcgcactgg ggccccacc ggccctcgtc aacacggcca cagcgggggc agcaggttgt 120
 gcatcctcca cccttaatgg catcagcagc catggtaacg gcaacagcag cagcaacaca 180
 catcacaggt ccgccaagct gcacgatggg gcagcccgcg ttgccaatag gacgacgacc 240
 accactagca aaataaaaagt ggaggaacca accaacgggtg gtggaagcat ggaggccgca 300
 tcagcggcgg tcgatggcaa ggacgctctc attcagcatg aactcaatgc gctactcagc 360
 caaaagcgtg acgcccgtgc tgatggcctg agccaagatt cctcccctgc ctcatccact 420
 tcctccacc ccaacgaccg cccatccatg atgcctacga ccatgaacgc ccctctcttc 480
 aaaccttctc ctactgctgc tgcttctttg tcttcgtcac cagcagtgcc aaccaccgga 540
 ggaggagcag ggaaagggct tatgcaagtc gctgcctcca gaccctccc tacctcatcg 600
 togtcgtccg caacagcgcg gccccccctg gccgccacgc ggcattcagc gagaggcgca 660
 aogggcggcg cagcagcagc ggggatcagc aagtccattg aggaggcctg tcattgccct 720
 gccctagaag atggtgtgct tggatccaaa ggaggggggg gcgggggagc ggtaggagga 780
 ggtggtggag gaggggcagg acatggccat gccaaagcaag gtgggggagg aattggtaat 840
 ggaggcgtat cgtcgtcaac ggcagctagt aatggcagtg ggaaaccggg gaaggctaag 900
 aatggattga ggagagggaa atggacgccc gaggaggagg cgtatgcaaa tcgattaatc 960
 gtggaattca agtcgggggt gttgccattg acggatggta cgacattgag gacatttttg 1020
 agtaagctgc ttaattgtga tcctatgagg ataagcaaaa aatttgtggg tcagaattgc 1080
 atcgaaagc aagtgtttcg tcggggcag caagcagaga tggacagggt gtcocctgaa 1140
 gagattgaac ggagtcgaca agaacttccc gaactagaga gacgattttt ggagcgcgtg 1200
 gctcagacaa atcgatgtaa gaattctgct gtggggacga agcatgggaa agactcggcg 1260
 tcgggtcagg tggcgcggga ggtgttgcca cgacagcaag agtacctggg cccaggtggg 1320
 caggtggtgg gtaatcccct cctccccccc tggatgctcc ctcccaccgc ggatgcatcg 1380
 atggaggag gggagggcat tcaaggtgga gatggtcatt cgaattcctc gccagagagg 1440
 ggggggatgg ggatgaccgg tcgcggtggt ggtggtagt cttttcgttt gatgggtgga 1500
 ggaacgggag caggagcaaa gcagcagcag cagcagagcc agcagcagca gcagccgct 1560
 ttttcggatt tgccctgcgc acaccagttg caccatctc agtcacatct ccatcatcca 1620
 cgacgtatcc cctcctcctc gtgctcacc ctaggagggc tgaacaoggg tagtctctcc 1680
 gogatcgcgt cggcgaatat gtcacgcacc gattctcagg atagcctaac agcgttaggt 1740
 ggcacaatgt cggcagacgc tcttgacat ttggcggact ttccgcgtgt gaccagccta 1800
 ggagatcttg ccggactcaa catgtttacc tccagccctt tcccctccac ggacaatctc 1860
 ctacgccttc accagcagca acacgcccga agtgcctaaa ttgttcggc cactgctgct 1920

ES 2 682 279 T3

tctcttgccg cgcatagcac caccagctta gcagcagcgg cagcagcagc agcagtggct 198
 ccgtcgagca caccgatcag tcgtagcacc agtggcaacg gtaatggcgg gagcagcagc 204
 agcagcagca gcagcagcag cagcagcaat gcgaatggtc ttccaccact agccaatgcc 210
 agcagcaagg tgatagtgc tggggcgggt cggttggcga caggaaagag caacaacggc 216
 aacataacta atagtggagg gacgagcagt agcactaccg cagcaggagg acggcgtatg 222
 ccccggtgg actcgccac aggttatcc tccttacggg tgatgagcgg attacctcgg 228
 aatacatcgg tgaagactt tctatccttg gtcgattacg gtgacatccc ggcgccagac 234
 caagacctgc tctccaagtg cgtctttcct caagccatca aggcgcctgt gaccagcag 240
 tctgcctgg cctttaccgg aatoggaagg agttccggcgg tgccggctgg aggaggtggt 246
 ggaggagtgg ggggagattc ggatgatgtg gctcatacaa gaggaggagg aggaggagga 252
 ggannnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnag cagcagcagc agcagcagca 258
 acagcggctt tgtaacatc aggggtaaag gggactggtg ggaagagggg tcgggcagac 264
 ttccctcgg cggcagtggg gggacaaaag ccgaaacaag ccaaggtgga tgggtaa 269

<210> 66
 <211> 898
 <212> PRT
 <213> Nannochloropsis oceanica

<220>
 <221> características_misc
 <222> (842) .. (853)
 <223> Xaa puede ser cualquier aminoácido natural

<400> 66

Met Ser Ala Thr Pro Ser Gly Pro Ser Phe Lys Val Asp Thr Asn Lys
 1 5 10 15

Thr Ala Ala Gly Ala Ala Leu Gly Pro Pro Pro Ala Ser Val Asn Thr
 20 25 30

Ala Thr Ala Gly Ala Ala Gly Cys Ala Ser Ser Thr Leu Asn Gly Ile
 35 40 45

Ser Ser His Gly Asn Gly Asn Ser Ser Ser Asn Thr His His Arg Ser
 50 55 60

Ala Lys Leu His Asp Gly Ala Ala Arg Ile Ala Asn Thr Thr Thr Thr
 65 70 75 80

Thr Thr Ser Lys Ile Lys Val Glu Glu Pro Thr Asn Gly Gly Gly Ser
 85 90 95

Met Glu Ala Ala Ser Ala Ala Val Asp Gly Lys Asp Ala Leu Ile Gln
 100 105 110

His Glu Leu Asn Ala Leu Leu Ser Gln Lys Arg Asp Ala Ala Ala Asp
 115 120 125

Gly Arg Ser Gln Asp Ser Ser Pro Ala Ser Ser Thr Ser Ser Thr Pro
 130 135 140

Thr Thr Ala Pro Ser Met Met Pro Thr Thr Met Asn Ala Pro Leu Phe
 145 150 155 160

Lys Pro Ser Pro Thr Ala Ala Ala Ser Leu Ser Ser Ser Pro Ala Val
 165 170 175

Pro Thr Thr Gly Gly Gly Ala Gly Lys Gly Leu Met Gln Val Ala Ala
 180 185 190

Ser Arg Pro Leu Pro Thr Ser Ser Ser Ser Ser Ala Thr Ala Gln Ala
 195 200 205

Pro Leu Ala Ala Thr Arg His Ser Thr Arg Gly Ala Thr Ala Ala Ala
 210 215 220

Ala Ala Ala Gly Ile Ser Lys Ser Ile Glu Glu Ala Cys His Ala Ala
 225 230 235 240

Ala Leu Glu Asp Gly Val Leu Gly Ser Lys Gly Gly Gly Gly Gly Gly
 245 250 255

Ala Val Gly Gly Gly Gly Gly Gly Gly Ala Gly His Gly His Ala Lys
 260 265 270

Gln Gly Gly Gly Gly Ile Gly Asn Gly Gly Val Ser Ser Ser Thr Ala
 275 280 285

Ala Ser Asn Gly Ser Gly Lys Pro Gly Lys Ala Lys Asn Gly Leu Arg
 290 295 300

Arg Gly Lys Trp Thr Pro Glu Glu Glu Ala Tyr Ala Asn Arg Leu Ile
 305 310 315 320

Val Glu Phe Lys Ser Gly Leu Leu Pro Leu Thr Asp Gly Thr Thr Leu
 325 330 335

Arg Thr Phe Leu Ser Lys Leu Leu Asn Cys Asp Pro Met Arg Ile Ser
 340 345 350

Lys Lys Phe Val Gly Gln Asn Cys Ile Gly Lys Gln Val Phe Arg Arg
 355 360 365
 Arg Gln Gln Ala Glu Met Asp Arg Leu Ser Pro Glu Glu Ile Glu Arg
 370 375 380
 Ser Arg Gln Glu Leu Ala Glu Leu Glu Arg Arg Phe Leu Glu Arg Val
 385 390 395 400
 Ala Gln Thr Asn Arg Cys Lys Asn Ser Ala Val Gly Thr Lys His Gly
 405 410 415
 Lys Asp Ser Ala Ser Gly Gln Val Ala Pro Glu Val Leu Arg Arg Gln
 420 425 430
 Gln Glu Tyr Leu Gly Pro Gly Gly Gln Val Val Gly Asn Pro Leu Leu
 435 440 445
 Pro Pro Trp Met Leu Pro Pro Thr Ala Asp Ala Ser Met Glu Gly Gly
 450 455 460
 Glu Gly Ile Gln Gly Gly Asp Gly His Ser Asn Ser Ser Pro Glu Arg
 465 470 475 480
 Gly Gly Met Gly Met Thr Gly Arg Gly Gly Gly Ser Ser Phe Arg
 485 490 495
 Leu Met Gly Gly Gly Thr Gly Ala Gly Ser Lys Gln Gln Gln Gln Gln
 500 505 510
 Ser Gln Gln Gln Gln Gln Pro Pro Phe Ser Asp Leu Pro Cys Ala His
 515 520 525
 Gln Leu His His Pro Gln Ser His Leu His His Pro Arg Arg Ile Pro
 530 535 540
 Ser Ser Ser Ser Ser Thr Leu Gly Gly Leu Asn Thr Gly Ser Leu Ser
 545 550 555 560
 Ala Ile Ala Ser Ala Asn Met Ser Arg Thr Asp Ser Gln Asp Ser Leu
 565 570 575
 Thr Ala Leu Gly Gly Thr Met Ser Ala Asp Ala Leu Gly His Leu Ala
 580 585 590
 Asp Phe Pro Arg Val Thr Ser Leu Gly Asp Leu Ala Gly Leu Asn Met

ES 2 682 279 T3

	595					600						605					
Phe	Thr	Ser	Ser	Pro	Phe	Pro	Ser	Thr	Asp	Asn	Leu	Leu	Ser	Leu	His		
	610					615					620						
Gln	Gln	Gln	His	Ala	Gln	Ser	Ala	Gln	Ile	Val	Pro	Ala	Thr	Ala	Ala		
625					630					635						640	
Ser	Leu	Ala	Ala	His	Ser	Thr	Thr	Ser	Leu	Ala							
				645					650						655		
Ala	Ala	Val	Ala	Pro	Ser	Ser	Thr	Pro	Ile	Ser	Arg	Ser	Thr	Ser	Gly		
			660					665					670				
Asn	Gly	Asn	Gly	Gly	Ser												
		675					680						685				
Ser	Asn	Ala	Asn	Gly	Leu	Pro	Pro	Leu	Ala	Asn	Ala	Ser	Ser	Thr	Val		
	690					695						700					
Ile	Val	Ala	Gly	Gly	Gly	Arg	Leu	Ala	Thr	Gly	Lys	Ser	Asn	Asn	Gly		
705					710					715					720		
Asn	Ile	Thr	Asn	Ser	Gly	Gly	Thr	Ser	Ser	Ser	Thr	Thr	Ala	Ala	Gly		
				725					730					735			
Gly	Arg	Arg	Met	Pro	Arg	Val	Asp	Ser	Ala	Thr	Gly	Leu	Ser	Ser	Leu		
			740					745					750				
Arg	Val	Met	Ser	Gly	Leu	Pro	Arg	Asn	Thr	Ser	Val	Glu	Asp	Phe	Leu		
		755					760					765					
Ser	Leu	Val	Asp	Tyr	Gly	Asp	Ile	Pro	Ala	Pro	Asp	Gln	Asp	Leu	Leu		
	770					775					780						
Ser	Lys	Cys	Val	Phe	Pro	Gln	Ala	Ile	Lys	Ala	Pro	Val	Thr	Gln	Gln		
785					790					795					800		
Ser	Ala	Val	Ala	Phe	Thr	Gly	Ile	Gly	Arg	Ser	Ser	Ala	Val	Pro	Ala		
				805					810					815			
Gly	Gly	Gly	Gly	Gly	Gly	Val	Gly	Gly	Asp	Ser	Asp	Asp	Val	Ala	His		
			820					825					830				
Thr	Arg	Gly	Xaa														
		835					840						845				

ES 2 682 279 T3

Xaa Xaa Xaa Xaa Xaa Ala Ala Ala Ala Ala Ala Ala Thr Ala Ala Leu
 850 855 860

Leu Thr Ser Gly Val Lys Gly Thr Gly Gly Lys Arg Gly Arg Ala Asp
 865 870 875 880

Leu Pro Ser Ala Ala Val Val Gly Gln Lys Pro Lys Gln Ala Lys Val
 885 890 895

Asp Gly

- <210> 67
- <211> 2343
- <212> ADN
- <213> *Phytophthora ramorum*

<400> 67
 atgggaacgg gagaagcgt gcacattcct cgcgcggacc cgagctccat gagcgcgtac 60
 ctgcctccgc agcettacaa ctcgtcgggg gtgcgaccgg acggaccccc gtacctgccg 120
 cccgggtcgt gtctcgtgtg ctacaacccc caggtggagc tgctgctgga gccgtgccac 180
 caccagttcc acgcgtcctg tatcgagcgc tggctcagca aggacaaggt ctgccccacc 240
 tgctggaagc ccatccaggc gccgcgccgc ctcgtcccgc agcagtagc acagccgcaa 300
 cagcaggagc gccggcgata ccccgagac agcaaagcgc acgtgttgga gcccgggag 360
 ccgcccaccc cagccgccag cgcggagccc agcgcgcagc caccgcggc tatgoggaag 420
 ggcaagtgga cggccgagga gagcgcgtac tgcgaccggg tgatcgagga gttcaagaag 480
 ggaaacctgc cgctggctga gggcacgacg ctgcgcacgt ttctgagcaa actgctgaac 540
 tgogacccca tgcgcatctc gaagaagtac acgggcgacc agtgcatcgg caagatcatt 600
 ttccgcggga gagaggacga cgtgtccaag gacgacatgg agagcatccg caaggacctg 660
 gccgagctcg agaagacgta cctggagagg gagcagtaca accagcggcg gcgcgagaag 720
 cggctcaggt cggagctctc gagggacaag agtcgctttg cggccaccag gtccattgga 780
 tacgcagcgg caggcaattc ggcgccgatg cggccgcccc accctcagca gcagcaaaa 840
 cagggagggt acccgcaggc tgcagtccag cccatgacaa agcaggagcc gcgacctgga 900
 ccgggtcctg tgcagcagcc caactatgga gctccaggtc gaggtggtat gcctgtccag 960
 cctcctctac accccccaca gcagagcaat aacgtgccct cgcacctatt taatggcgac 1020
 cacagcaatg tatcgatgct tggcattgct ggcgcgcaga cacaggggtca aggtcaggtt 1080
 ccgaaccagg acaacaagcc ggctagcagc tccgctatgg acggcgggga cgggtttccg 1140
 cgogtgcgt ccatcgacag cttttcttgt ctgttccttc gcgtggccag tattgaaaac 1200
 ttccagcagc cgacgtcctc tggttttggc ggcataaact ctgtcgggtc gtaccogtgc 1260

acggagactc aaaacacgat gaccagcagt ggattogacg cccagccttc cggtttgccc 1320
 aaaccattgt cgattggcga aggccttaac gcctacttcc ctcgatcca gtcgctagaa 1380
 cagctttcaa acctactgca agaccacgga ccgaacagtc cacgtggagg agcaacatcg 1440
 tcgtogactc agaacagtcg ggataccatg gagaattcaa gcgacagcaa acaacagatt 1500
 aaggacaatt ttacatccgg aggtcatcgg aggaggctgg gcgagaatgg tattaaggaa 1560
 gagcagccgc gggaaagcaga tttcaacgca cagacagggc ccagcagcag caataacagc 1620
 agcaacagtt ctacgagcag cgacagcagt gccgtttcca cttegatgag cgtgacgacc 1680
 ggttccacta gcaactggact gacaaagcgc ctacgcccc gtcacaatgc atcggcctct 1740
 tcaggttcag gcaaccagat ccagattccc aagccgctga acaagatgcc gcgaagctcg 1800
 tcgggcatct tcccgcgcgt gccgtccatg gacaagatgc ctcgtgtgcc ttcgctcgac 1860
 aagatgcctc gcggtgcac tctogacaag ctgcgcgcga tcccgtctat ggacaagctc 1920
 cacacggtgg gcggcggcga gcaacgcac cccaggggtgc cgtccatgga caagatggcg 1980
 cgtgtgccga gctcggacat gctgtctcgc tttggctcca gcgaccacct tagcagcttc 2040
 ccgtcgttct ctaacctaa cagttgtct tcaagcgcct cgtacgacaa gctgagctcg 2100
 ctgggtgggt ttaagtggg cttcccgcgc aattcgtcca ttgaggacat tttgtctctc 2160
 gtggcctcgt ccgagtcgac tggctctccg tcaaattggc cgacgttgca gctgagtgca 2220
 ctggccgcag ttgcgggcga ggagtcctcc catattgcca acgagcggaa gcggcggctg 2280
 gaaagctccc agcaggacgc atcgttgcca tcagataaca agaagagcaa attatcggcg 2340
 tga 2343

<210> 68
 <211> 780
 <212> PRT
 <213> *Phytophthora ramorum*

<400> 68

Met Gly Thr Gly Glu Ala Leu His Ile Pro Arg Ala Asp Pro Ser Ser
 1 5 10 15

Met Ser Ala Tyr Leu Pro Pro Gln Pro Tyr Asn Ser Ser Gly Val Arg
 20 25 30

Pro Asp Gly Pro Pro Tyr Leu Pro Pro Gly Ser Cys Leu Val Cys Tyr
 35 40 45

Asn Pro Gln Val Asp Val Leu Leu Glu Pro Cys His His Gln Phe His
 50 55 60

Ala Ser Cys Ile Glu Arg Trp Leu Ser Lys Asp Lys Val Cys Pro Thr
65 70 75 80

Cys Trp Thr Pro Ile Gln Ala Pro Arg Arg Leu Val Pro Gln Gln Tyr
85 90 95

Ala Gln Pro Gln Gln Gln Asp Gly Gly Gly Tyr Pro Ala Asp Ser Lys
100 105 110

Ala His Val Leu Glu Ala Ala Glu Pro Pro Thr Pro Ala Ala Ser Ala
115 120 125

Asp Ala Ser Ala Gln Pro Pro Ala Ala Met Arg Lys Gly Lys Trp Thr
130 135 140

Ala Glu Glu Ser Ala Tyr Cys Asp Arg Leu Ile Glu Glu Phe Lys Lys
145 150 155 160

Gly Asn Leu Pro Leu Ala Glu Gly Thr Thr Leu Arg Thr Phe Leu Ser
165 170 175

Lys Leu Leu Asn Cys Asp Pro Met Arg Ile Ser Lys Lys Tyr Thr Gly
180 185 190

Asp Gln Cys Ile Gly Lys Ile Ile Phe Arg Arg Arg Glu Asp Asp Val
195 200 205

Ser Lys Asp Asp Met Glu Ser Ile Arg Lys Asp Leu Ala Glu Leu Glu
210 215 220

Lys Thr Tyr Leu Glu Arg Glu Gln Tyr Asn Gln Arg Arg Arg Glu Lys
225 230 235 240

Arg Leu Glu Ser Glu Leu Ser Arg Asp Lys Ser Arg Phe Ala Ala Thr
245 250 255

Arg Ser Ile Gly Tyr Ala Ala Ala Gly Asn Ser Ala Pro Met Arg Pro
260 265 270

Pro His Pro Gln Gln Gln Gln Gln Gln Gly Gly Tyr Pro Gln Ala Ala
275 280 285

Val Gln Pro Met Thr Lys Gln Glu Pro Arg Pro Gly Pro Gly Pro Val
290 295 300

Gln Gln Pro Asn Tyr Gly Ala Pro Gly Arg Gly Gly Met Pro Val Gln
305 310 315 320

Pro Pro Leu His Pro Pro Gln Gln Ser Asn Asn Val Pro Ser His Leu
 325 330 335
 Phe Asn Gly Asp His Ser Asn Val Ser Met Leu Gly Ile Ala Gly Ala
 340 345 350
 Gln Thr Gln Gly Gln Gly Gln Val Pro Asn Gln Asp Asn Lys Pro Ala
 355 360 365
 Ser Ser Ser Ala Met Asp Gly Gly Asp Gly Phe Pro Arg Val Ser Ser
 370 375 380
 Ile Asp Ser Phe Ser Cys Leu Phe Pro Arg Val Ala Ser Ile Glu Asn
 385 390 395 400
 Phe Gln His Ala Thr Ser Ser Gly Phe Gly Gly Met Asn Ser Val Gly
 405 410 415
 Ser Tyr Pro Ser Thr Glu Thr Gln Asn Thr Met Thr Ser Ser Gly Phe
 420 425 430
 Asp Ala Gln Pro Ser Gly Leu Pro Lys Pro Leu Ser Ile Gly Glu Gly
 435 440 445
 Leu Asn Ala Tyr Phe Pro Arg Ile Gln Ser Leu Glu Gln Leu Ser Asn
 450 455 460
 Leu Leu Gln Asp His Gly Pro Asn Ser Pro Arg Gly Gly Ala Thr Ser
 465 470 475 480
 Ser Ser Thr Gln Asn Ser Arg Asp Thr Met Glu Asn Ser Ser Asp Ser
 485 490 495
 Lys Gln Gln Ile Lys Asp Asn Phe Thr Ser Gly Gly His Arg Arg Arg
 500 505 510
 Leu Gly Glu Asn Gly Ile Lys Glu Glu Gln Pro Arg Glu Ala Asp Phe
 515 520 525
 Asn Ala Gln Thr Gly Pro Ser Ser Ser Asn Asn Ser Ser Asn Ser Ser
 530 535 540
 Ser Ser Ser Asp Ser Ser Ala Val Ser Thr Ser Met Ser Val Thr Thr
 545 550 555 560
 Gly Ser Thr Ser Thr Gly Leu Thr Lys Arg Leu Ser Pro Ser His Asn
 565 570 575

Ala Ser Ala Ser Ser Gly Ser Gly Asn Gln Ile Gln Ile Pro Lys Pro
 580 585 590

Leu Asn Lys Met Pro Arg Ser Ser Ser Gly Ile Phe Pro Arg Val Pro
 595 600 605

Ser Met Asp Lys Met Pro Arg Val Pro Ser Leu Asp Lys Met Pro Arg
 610 615 620

Val Ala Ser Leu Asp Lys Leu Pro Arg Ile Pro Ser Met Asp Lys Leu
 625 630 635 640

His Thr Val Gly Gly Gly Glu Gln Arg Ile Pro Arg Val Pro Ser Met
 645 650 655

Asp Lys Met Ala Arg Val Pro Ser Ser Asp Met Leu Ser Arg Phe Gly
 660 665 670

Ser Ser Asp His Leu Ser Ser Phe Pro Ser Phe Ser Asn Leu Ser Thr
 675 680 685

Leu Ser Ser Ser Ala Ser Tyr Asp Lys Leu Ser Ser Leu Gly Gly Phe
 690 695 700

Lys Ser Gly Phe Pro Arg Asn Ser Ser Ile Glu Asp Ile Leu Ser Leu
 705 710 715 720

Val Ala Ser Ser Glu Ser Thr Gly Leu Pro Ser Asn Gly Ser Thr Leu
 725 730 735

Gln Leu Ser Ala Leu Ala Ala Val Ala Gly Glu Glu Ser Ser His Ile
 740 745 750

Ala Asn Glu Arg Lys Arg Arg Leu Glu Ser Ser Gln Gln Asp Ala Ser
 755 760 765

Leu Ala Ser Asp Asn Lys Lys Ser Lys Leu Ser Ala
 770 775 780

<210> 69
 <211> 2799
 <212> ADN
 <213> Ectocarpus siliculosus

<400> 69
 atgcagcctg ggatttcccc cagagttaag cggggagtgg cggtgggcagc ggtggcgggg 60
 gggggggggg ggagagagac gtgcaggagc actagcggta cgattgcgat cgaggcggct 120

ggacagagcgg ctgcgggtgc tgatgggggtg attgcgagggc cagcagttctc tcaacaacaa 180
 ccgcagctgc agcgcgcagc caccaaggcc cccgtgccgg tagcggcggc tccgaccccg 240
 gccagggccg ccgcctacgc gaaagcgggtg gcgacggcgc agcagcaaca gggggcccac 300
 gccgcgcgca cggcggcggc ggcggcggcg agggtagcga gatcgaacgc gatcgccact 360
 tcgtacacaa gcaacgccgc cgcgcgccga gcagcagcag cagcggcggc agccgcggcc 420
 gccccctcgg ctgcgggtgc gtgcggggcg acgcatcggc aacagcagca ccccggtgtc 480
 gtcgtcgccg cccggcccac cgtggtggcc cggggccaca gcaacggcca catctacaac 540
 gcggcccagc ggcagcagca gcagcagcac gcggccgctt ttggggcgat gggggcagcg 600
 gcggcggcgg cggcggcggg gggtcaccat caccaccatc acttgcaagc gtttcagcag 660
 ggcgggggca gtgggttctc ggggctgcag gggggcgcta gcgcgatggg ggcggcggcg 720
 gcagcggcgg cagcggcagc ggcggcggcg gcggcgggtc aggcgaaagc gaagaagccg 780
 aacgggttga ggagagggaa gtggacgtcc gaggaggagg actacgcgaa ccgccttacc 840
 caggagttca agagcgggtt gttgcgcgtg acggacggaa cgacgctcag aacgttctcg 900
 agcaagctgc tcaactgcga ccccatgcgc atctccaaga agttcgtggg tagcaactgc 960
 atcgggaagc aggtgttccg caggaggcag gcgcacatgg accggttaca gccgcgccac 1020
 atcgagagga gccgctctca gcttgcaagc ctagagagga ggttcttggg gcgggtagcc 1080
 atgacgaacc gatgcaagc cggcggggag ataggctcga aagaaggcgg cagcagcttc 1140
 gacctcctgc gcccgaaatc gccacctccg gctaccacgc cgtggatgct gccgcgcgct 1200
 cccccctcgg ttactctcgc ggggacgctg tcgcccgcgg cggcgcagca gagggcgaac 1260
 gtggcggggg ctctcgcaga tgaagccaca gggggcgtgc tcgctgagac aacagcggta 1320
 gcagcgggag cggcagggca ccagcgaat agcaccacac cgcacacctc accaccatca 1380
 tcggcagcag caacagcacc agggtcggcg gaggtagcgc catcgcctgc atcagcagcg 1440
 ttgcgcgtca gttccgtttt gggcgcgagc catgcggggc caggggggagt cgcgaaagca 1500
 ccgaacgggc tgggtccttg tggcccacc gaaaaggagt ctcaacacag gccctcacg 1560
 tcgcagcagc aaccgaatgc tcaaggtacg agcaccactg ctctggcccc tgccccggt 1620
 accccagctg agataaacga tggtggggtc aaggcggggc cggcgcggg tatcgctaac 1680
 ggcaaccacc tggcagcaac aacgtcagcg gcgtcagcag tggcagcaac accgcctgct 1740
 tcgtcagggc ccggtgtatc ggctgctgcg acaagggcta atgggtgcgc tgcacgacc 1800
 agtcgcggctg cggccacgcc ggcgtcgttc ctagacgcgg gaggcgcgct gccgtcgttg 1860
 cccgggactg gggacgggag ggtagcttcc gcgatcccc gaagtaccgc tgcagttggt 1920
 cccaagcagg aactcacggc cagtgggaca ggcttgggtg tccgcacatc gaggtcgggg 1980

gcaagcaacg gcttcaacac cagctccggc ctgggtggaa gggcatcga cggaccaca 2040
 tctctcgagg cctctcogct gctggagctg ccacacatcc aaggtatgac caaccttgct 2100
 gtgctggggc tgtggcctgg cgcctttccc ccttcgggat cttcgacagg gacaacgact 2160
 gggactcctc tcaaggogat gaacggcggg ttgaagagaa cgccttcatg ggcgaggatc 2220
 agctcgtacg aacacctgca gagcctcgat gaaggggctc cctgtatcc gacgacaaac 2280
 ggcaacggag caggcagtgt gagcagcaac atcgcgagca gtagcagcag caacagccat 2340
 tccgcatctg gactaggagg aggcaattct agcggtgcca gggcactag caccatgacg 2400
 tcatcatcat cgccatcgcc atcctcagcc gcgcacctt cgccttcat gccgagaaac 2460
 acctcggteg aggacttct cagccttgtc gagtccggag acattccgcc tccagagggt 2520
 gccgcgggt tcacggtgcc catgtggatc tacagcggag actcgagcaa caccacgaac 2580
 ggcaacccca acggacacgc tctaattggg agttogaatg gtgcttcaac tctgctgcc 2640
 agcgagcgtc cagcggggca ctctaaaccg tctcaaaaat ctaagaacaa gggctcgggt 2700
 agcggttctg ctgcgggggc ttctggaatg tcatcactcg agaagtccaa gtcgaaacgg 2760
 ctgaagtoga cgcaggctaa ggcgagggcc aagagttga 2799

<210> 70
 <211> 932
 <212> PRT
 <213> *Ectocarpus siliculosus*

<400> 70

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

ES 2 682 279 T3

Ala Arg Ser Asn Ala Ile Ala Thr Ser Tyr Thr Ser Asn Ala Ala Ala
115 120 125

Ala Pro Ser Ala
130 135 140

Ala Gly Ala Ser Pro Ala Thr His Arg Gln Gln Gln His Pro Gly Val
145 150 155 160

Val Val Ala Ala Arg Pro Thr Val Val Ala Pro Gly His Ser Asn Gly
165 170 175

His Ile Tyr Asn Ala Ala Gln Arg Gln Gln Gln Gln His Ala Ala
180 185 190

Ala Phe Gly Ala Met Gly Ala Ala Ala Ala Ala Ala Ala Gly Gly
195 200 205

His His His His His His Leu Gln Gln Phe Gln Gln Gly Gly Gly Ser
210 215 220

Gly Phe Leu Gly Leu Gln Gly Gly Ala Ser Ala Met Gly Ala Ala Ala
225 230 235 240

Ala Val Gln Ala Lys
245 250 255

Ala Lys Lys Pro Asn Gly Leu Arg Arg Gly Lys Trp Thr Ser Glu Glu
260 265 270

Glu Asp Tyr Ala Asn Arg Leu Ile Gln Glu Phe Lys Ser Gly Leu Leu
275 280 285

Pro Leu Thr Asp Gly Thr Thr Leu Arg Thr Phe Leu Ser Lys Leu Leu
290 295 300

Asn Cys Asp Pro Met Arg Ile Ser Lys Lys Phe Val Gly Ser Asn Cys
305 310 315 320

Ile Gly Lys Gln Val Phe Arg Arg Arg Gln Ala Asp Met Asp Arg Leu
325 330 335

Gln Pro Ala Asp Ile Glu Arg Ser Arg Ser Gln Leu Ala Asp Leu Glu
340 345 350

Arg Arg Phe Leu Glu Arg Val Ala Met Thr Asn Arg Cys Lys Thr Gly
355 360 365

Gly Glu Ile Gly Ser Lys Glu Gly Gly Ser Ser Phe Asp Leu Leu Arg
 370 375 380

Pro Asn Gln Pro Pro Pro Ala Thr Gln Pro Trp Met Leu Pro Pro Ser
 385 390 395 400

Pro Pro Ala Val Thr Leu Ala Gly Thr Leu Ser Pro Ala Ala Ala Thr
 405 410 415

Gln Arg Ala Asn Val Ala Gly Ala Pro Ala Asp Glu Ala Thr Gly Gly
 420 425 430

Val Leu Ala Glu Thr Thr Ala Val Ala Ala Glu Ala Ala Gly His Gln
 435 440 445

Arg Asn Ser Thr Thr Pro Pro Pro Arg Pro Pro Ser Ser Ala Ala Ala
 450 455 460

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

Leu Pro Ser Ser Ser Val Leu Gly Ala Ser His Ala Gly Pro Gly Gly
 485 490 495

Val Ala Lys Ala Pro Asn Gly Leu Val Leu Gly Gly Pro Pro Glu Lys
 500 505 510

Glu Ser Gln His Arg Pro Leu Thr Ser Gln Gln Gln Pro Asn Ala Gln
 515 520 525

Gly Thr Ser Thr Thr Ala Leu Ala Pro Ala Pro Ala Thr Pro Ala Glu
 530 535 540

Ile Asn Asp Val Gly Val Lys Ala Gly Ser Ala Ala Gly Ile Ala Asn
 545 550 555 560

Gly Asn His Leu Ala Ala Thr Thr Ser Ala Ala Ser Ala Val Ala Ala
 565 570 575

Thr Pro Pro Ala Ser Ser Gly Ser Gly Val Ser Ala Ala Ala Thr Arg
 580 585 590

Ala Asn Gly Ala Pro Ala Ser Thr Ser Ala Ala Ala Ala Thr Pro Ala
 595 600 605

Ser Phe Leu Asp Ala Gly Gly Ala Leu Pro Ser Leu Pro Gly Thr Gly

ES 2 682 279 T3

610						615										620
Asp	Gly	Arg	Val	Ala	Ser	Ala	Ile	Pro	Gly	Ser	Thr	Ala	Ala	Val	Val	
625					630					635					640	
Pro	Lys	Gln	Glu	Leu	Thr	Ala	Ser	Gly	Thr	Gly	Leu	Gly	Val	Arg	Thr	
				645					650					655		
Ser	Arg	Ser	Gly	Ala	Ser	Asn	Gly	Phe	Asn	Thr	Ser	Ser	Gly	Leu	Gly	
			660					665					670			
Gly	Ser	Gly	Ile	Asp	Arg	Thr	Thr	Ser	Leu	Glu	Ala	Leu	Ser	Leu	Leu	
		675					680					685				
Glu	Leu	Pro	His	Ile	Gln	Gly	Met	Thr	Asn	Leu	Ala	Val	Leu	Gly	Leu	
	690					695					700					
Trp	Pro	Gly	Ala	Phe	Pro	Pro	Ser	Gly	Ser	Ser	Thr	Gly	Thr	Thr	Thr	
705					710					715					720	
Gly	Thr	Pro	Leu	Lys	Ala	Met	Asn	Gly	Gly	Leu	Lys	Arg	Thr	Pro	Ser	
				725					730					735		
Trp	Ala	Arg	Ile	Ser	Ser	Tyr	Glu	His	Leu	Gln	Ser	Leu	Asp	Glu	Gly	
			740					745					750			
Ala	Pro	Leu	Tyr	Pro	Thr	Thr	Asn	Gly	Asn	Gly	Ala	Gly	Ser	Val	Ser	
		755					760					765				
Ser	Asn	Ile	Ala	Ser	Ser	Ser	Ser	Ser	Asn	Ser	His	Ser	Ala	Ser	Gly	
	770						775				780					
Leu	Gly	Gly	Gly	Asn	Ser	Ser	Gly	Gly	Ser	Gly	Thr	Ser	Thr	Met	Thr	
785				790						795					800	
Ser	Ser	Ser	Ser	Pro	Ser	Pro	Ser	Ser	Ala	Ala	Ala	Pro	Ser	Pro	Phe	
				805					810					815		
Met	Pro	Arg	Asn	Thr	Ser	Val	Glu	Asp	Phe	Leu	Ser	Leu	Val	Glu	Ser	
			820					825					830			
Gly	Asp	Ile	Pro	Pro	Pro	Glu	Gly	Ala	Ala	Gly	Phe	Thr	Val	Pro	Met	
		835					840					845				
Trp	Ile	Tyr	Ser	Gly	Asp	Ser	Ser	Asn	Thr	Thr	Asn	Gly	Asn	Pro	Asn	
	850					855					860					

Gly His Ala Pro Asn Gly Ser Ser Asn Gly Ala Ser Thr Pro Ala Ala
 865 870 875 880

Ser Glu Arg Pro Ala Gly His Ser Lys Pro Ser Ser Lys Ser Lys Asn
 885 890 895

Lys Gly Ser Gly Ser Gly Ser Ala Arg Gly Ala Ser Gly Met Ser Ser
 900 905 910

Leu Glu Lys Ser Lys Ser Lys Arg Leu Lys Ser Thr Gln Ala Lys Ala
 915 920 925

Glu Ala Lys Ser
 930

<210> 71

<211> 3288

<212> ADN

<213> *Aureococcus anophagefferens*

<400> 71

atgaagcagc gcaaggcgcc ggcgaaagcc actgcgccgg atgcggaoga cgagcccaag 60
gatcgcacgt cgctctacgg ctcgctgctg cgccgctttc cgctggccat gaacgccgtc 120
caggccggcg cgctgagcgc ggcgtcccag ctgctgtcgc agcggctcaa gggcgcgcg 180
gcgctcgact tcgcgccggc gctgcggttc gcgctcatca gcgctttcgt ggtgacgccc 240
gtgagcaagg tcttcttttc catogtcggc aagttccggc tccggacgcc cgcgtcgtg 300
gcgctcgact tcttcgtcgg cgggcccgtc ctcaactgcg ccttcatcgc cgcgctccac 360
ggcctccagg gccaggacct ggccttcac ctggcgctgc tgcggtcgcg ggccttctgg 420
gtgacatagg tcttcggctc gaacaagggtg tggctgcccg cgaaggctgc catgtactcg 480
ctcgtgcccg cggagtaactg ggggctctgg tgctcctgcg tctccttogg ctggggcacc 540
gtcctcgcga cgatgcctc gaagaagaag aaggcgagt acacgaattg ggcgcggctc 600
ggcggcggtt cggcgcctt cgtcgtcgtt tgggtgcgcg gcgctacat aaattacggg 660
cggatgaacg agaagctgga gaagcagagc aagatcgtcc cgcctgtcca gtcggactc 720
gagaccccgc tgccgctcga cccgtcagcg caggccgtcg ggaccccgct gccgctcagc 780
ccgtcagcgc gggcgcgtcg cgcgcggggc ttgccggtgg cccaggtcgt cgcgcgata 840
ccggggccga agaccagct cgtcgtcgcg gcgtcccccg gcccgctcgg agtccatttc 900
aaggcgggga cggcggcggt cgcggagctt tccccgggt cgcagctcgc gggcgaggtc 960
gagatcggcg acgtcctcga gtcctcaac ggacgcccgg ctaccgcgcg gtcctcgcg 1020
gccaggcgcg gcacctcga ggagaacgac gacggcgaaa cgcgcgggac cctcgtcttc 1080
gcgcgcccg cccggaggtt cgcctccac gcggcgccgg gctccctcgg cgtcgtcttc 1140

ES 2 682 279 T3

gccccggca cgacgcggt cggacgcgt cccgaacgag gaagcgttct ggggcccggc 1200
 cgatcgcagc tcttcggctt cgtcgcagag ggcgacgcgc tcctctccgt gaacggccgc 1260
 ccggccgcgc ccgcgagccc ggcgagggg ggcgtcctcg acgaggagga cgacggccgc 1320
 cgcgcgcgac tcctcgtctt cgagcgcggc cccaagcagc gaaacgagct ctgcaccaac 1380
 gtcgagttcg tcatacgcgc gaagcccggg ccgctcggcg tcgtcttcaa gaaccggcca 1440
 atcttcggga acgggcccgc cccgcagcgt ctcccacacg ctcccgaatc gttctcgcga 1500
 aagagcatcg cgcgccggg cgaccaatcg ctgtttctcg gcgcgccgaa accgatccct 1560
 gtttctcgcg gcgaggaagc gaccatggcc gcgacggagg acgcgcgcgc cctgacgcag 1620
 aacgtcgcag ccttgaggc cgcggcccgc gcgccctgc ccgacgagtc gcagatcgag 1680
 gcgtccgcgg agaagtcgcc gccgcggagc ccgcggtcgc gcaaggcggc cggcgacaag 1740
 cccgcgcgca agccgcggg cgagaacggc ctgcggcgcg gcaagtggac cgtcagggag 1800
 gaggcctacg cgaaccggct gatccacgag ttcaagctcg gcctgctgcc gctcaccgac 1860
 ggcaacgcgc tcggacggt cctgtogaag ctgctcaact gcgacccat gcgcatctcg 1920
 aagaagttcg tgggctcga ctgcatcggc aagcaggtct tcgcgccgcg ccaggcggac 1980
 atggaccggc tcacgcccga cgacatcaag cggagccgct acgagctcgc ggaactggag 2040
 cgcgggttcc tcacgcggt cgcgcagagc caccggtccg ccaagtccgg cggcgcgggc 2100
 gccaaagggc tcaagggcgg cgacggcaag gccctggcg gcgggctcat gcaggcccag 2160
 cagcggccca tgctcgcgc gtggctcctg ccgcgcacg cggccgcggg cgcgccgacg 2220
 gtcgcgcgg gcgcggggc cgtcgcctc gccgcgcct acggcctgcc gtectacgcg 2280
 gcgcgcgcgc ccgcgcggc gccgcgcgc ccgcgcgcgc cgaagcccga cgcggacgcg 2340
 gcggccgcgg agaaggcga gcgcgcggc ctcgagggcc tgcacctgcc gtcgctccag 2400
 tcgacgcgt cgtccagtc gctggcctc acggcgcgc agccgtcctt cgcgtccctc 2460
 ggcgcgcct gccgctggc gccgaaactc gccagggct ggcagagctc gaactcgtc 2520
 gcgggcgacg cggccgcggc cggccggccg cgcgcgcga gcggcgacgc gggcctgagc 2580
 tcctggcgt ccttctcca cctgtcagc acggccgagc actcgcgcg gctccgcgcc 2640
 gcggcgcgc ccaaggcgc ggcgcgcgc ctgcacatcg cggccgctgc caagcccgcg 2700
 gaggcgcgc cgtcgcgcg gccgcgcgc ccgcgcgcgc ccgcgcgcgc ggcggccgc 2760
 gcgcgcgcgc cgcgcgcgc cgcgcgcgc gccccgcgc cggccgcgc cgcgcgcgc 2820
 gcgctcgcgt cccgcacgg ccgcacgcgc gagctgacgc ccgagcccga ggagccgacg 2880
 gcgaagcggc ggtcggagc ttcgcgaag cgcgagcggc cgcgccacgt gtcgagctcc 2940
 gacgaggact ccaacggcgg cgcgcgcgc gccgaggacc tcgaccgcgc gacgctcgcg 3000

```

cagcagcgca aggacgtgcg ccgcgagcgc aaccggcagc acgcgcgcggt ctgcgagag      3060
cgcaagcgcc agaagctcga gcacctccag gaggagaacg acgcgctccg cggccaggag      3120
gcccgcctga tggaccagcg cgaccgcctc aacgcgcgcc tctgcgcggt cgagtacgag      3180
aaccgcgcgc tccgcgctg gatccagaac cagcgggggc acggcgcgcc gccgcccggc      3240
ccggccgccc acgacgacga ggcgagccgg cccgtgcccg accactga                    3288

```

```

<210> 72
<211> 1095
<212> PRT
<213> Aureococcus anophagefferens

```

```
<400> 72
```

```
Met Lys Gln Arg Lys Ala Pro Ala Lys Ala Thr Ala Pro Asp Ala Asp
1                               5                               10                               15
```

```
Asp Glu Pro Lys Asp Arg Thr Ser Leu Tyr Gly Ser Leu Leu Arg Arg
                20                               25                               30
```

```
Phe Pro Leu Ala Met Asn Ala Val Gln Ala Gly Ala Leu Ser Ala Ala
35                               40                               45
```

```
Ser Gln Leu Leu Ser Gln Arg Leu Lys Gly Ala Ala Ala Leu Asp Phe
50                               55                               60
```

```
Ala Pro Ala Leu Arg Phe Ala Leu Ile Ser Ala Phe Val Val Thr Pro
65                               70                               75                               80
```

```
Val Ser Thr Val Phe Phe Ser Ile Val Gly Lys Phe Arg Leu Arg Thr
85                               90                               95
```

```
Pro Ala Ser Leu Ala Leu Asp Phe Phe Val Gly Gly Pro Phe Leu Asn
100                              105                              110
```

```
Cys Ala Phe Ile Ala Ala Leu His Gly Leu Gln Gly Gln Asp Leu Ala
115                              120                              125
```

```
Phe Ile Leu Gly Val Leu Arg Ser Arg Ala Phe Trp Val Asp Met Val
130                              135                              140
```

```
Leu Gly Ser Asn Lys Val Trp Leu Pro Ala Lys Val Ala Met Tyr Ser
145                              150                              155                              160
```

```
Leu Val Pro Pro Glu Tyr Trp Gly Leu Trp Cys Ser Cys Val Ser Phe
165                              170                              175
```

```
Gly Trp Gly Ile Val Leu Ala Thr Ile Ala Ser Lys Lys Lys Lys Ala
```


ES 2 682 279 T3

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

ES 2 682 279 T3

Gln Ser His Arg Ser Ala Lys Ser Gly Gly Ala Gly Ala Lys Gly Val
690 695 700

Lys Gly Gly Asp Gly Lys Ala Leu Gly Gly Gly Leu Met Gln Ala Gln
705 710 715 720

Gln Arg Pro Met Leu Ala Pro Trp Leu Leu Pro Pro His Ala Ala Ala
725 730 735

Gly Ala Pro Thr Val Arg Ala Gly Ala Gly Ala Val Ala Ile Ala Ala
740 745 750

Pro Tyr Gly Leu Pro Ser Tyr Ala Ala Pro Pro Pro Ala Ala Ala Pro
755 760 765

Pro Pro Pro Pro Pro Pro Lys Pro Asp Ala Asp Ala Ala Ala Glu
770 775 780

Lys Ala Lys Arg Ala Ala Leu Glu Gly Leu His Leu Pro Ser Leu Gln
785 790 795 800

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

Phe Ala Ser Leu Gly Ala Ala Trp Pro Ser Ala Pro Asn Leu Ala Gln
820 825 830

Gly Trp Gln Ser Ser Asn Ser Leu Ala Gly Asp Ala Ala Ala Ala Gly
835 840 845

Arg Pro Arg Ala Pro Ser Gly Asp Ala Gly Leu Ser Ser Trp Pro Ser
850 855 860

Phe Ser His Leu Val Thr Thr Ala Asp Asp Ser Pro Arg Leu Pro Pro
865 870 875 880

Ala Ala Pro Pro Lys Ala Ala Ala Pro Pro Leu Asp Ile Ala Ala Ala
885 890 895

Ala Lys Pro Ala Glu Ala Pro Pro Leu Ala Arg Pro Ala Pro Pro Ala
900 905 910

Pro Pro Ala Pro Pro Ala Ala Ala Ala Pro Pro Pro Pro Pro Ala
915 920 925

Ala Ala Ala Pro Pro Ala Ala Ala Ala Val Ala Ala Leu Ala Ser
930 935 940

ES 2 682 279 T3

Pro His Gly Arg Thr Arg Glu Leu Thr Pro Glu Pro Glu Glu Pro Thr
 945 950 955 960

Ala Lys Arg Arg Ser Glu Pro Ser Ala Lys Arg Glu Arg Pro Pro His
 965 970 975

Val Ser Ser Ser Asp Glu Asp Ser Asn Gly Gly Ala Gly Ala Ala Glu
 980 985 990

Asp Leu Asp Pro Thr Thr Leu Glu Gln Gln Arg Lys Asp Val Arg Arg
 995 1000 1005

Glu Arg Asn Arg Gln His Ala Arg Val Ser Arg Glu Arg Lys Arg
 1010 1015 1020

Gln Lys Leu Glu His Leu Gln Glu Glu Asn Asp Ala Leu Arg Arg
 1025 1030 1035

Gln Glu Ala Ala Leu Met Asp Gln Arg Asp Arg Ile Asn Ala Arg
 1040 1045 1050

Leu Leu Arg Val Glu Tyr Glu Asn Arg Ala Leu Arg Ala Trp Ile
 1055 1060 1065

Gln Asn His Ala Gly Asp Gly Ala Pro Pro Pro Pro Pro Ala Ala
 1070 1075 1080

Asp Asp Asp Glu Ala Ser Arg Pro Val Pro Asp His
 1085 1090 1095

<210> 73
 <211> 2013
 <212> ADN
 <213> *Thalassiosira pseudonana*

<400> 73
 atgcaagctg cccgatcggc agcgaaggca gatggagtag cccaaccaa taatgatgta 60
 gcaaaagcat ctgattctca agagaaggcg aatgcaaagt ctgatgaaga ggagaaaagt 120
 gaaacgatg atgccgtccc aatggagat gaaaaagatg ccgaagacgc agccgctccg 180
 gcttctgcc aaaacggatc ggctctcaac cagatgctga gtgccatcaa cggaccaccc 240
 aatcaagtga cttccaccaa cctttagga tcagtagatg atccgtctgc tctatctgaa 300
 tctcccgcg aacoggacag tgcggccgat cgtogtcggg ctccccttcg togtgggaaa 360
 tggactgagg aggaagaagc ctacgctagt cgtctcattc aagagttcaa agctgggttg 420
 ctccccctca ccgatggcac aactctccgt accttcttga gtaagctctt gaactgtgat 480

ES 2 682 279 T3

cctatgcgta tctccaagaa gtttgttggg agcaactgta tcggcaaaca agtctttcgt 540
 cgacggggag ctgatgttaa caacttgact ccagcacaaa tccagcaaac tcgtctcgaa 600
 ctatccgagt tggagaagag gttcttggat cgogtctcgc agaacaagaa atctggtggt 660
 tcccccaaga gtgaacggcc ggccagcaag cccagctctg gaaatgatgg tggctctctct 720
 ggcgtgtctg ggatgtcggg aataggcaac atgaacaagt ctgctgcggc ggctggctgt 780
 gcactgctcc aaggaaacaa gggcgggtga aatggagata gtggtcctac agggttgctt 840
 gctcagcttc aagccagtca accaggaatg tttgatgcga atactgcaat ggcttacaat 900
 tccaatgctt caggtggcgg gcaggcagtc atgggagtca acagtgcttc cattaacaac 960
 ctcatgcttc aaaccggcat gacggccgag caaatctccc aactcactca gacaaagggc 1020
 attaactcct cggcatccct tgctaatttg cttggaaaga agcgaagctt tgatggtctc 1080
 atgtccttgg actttcagag tatgcagagt attgataatt tggccaactt gattcagcaa 1140
 ggaataccta gtcaatctct ccacaagaat caaatgaaga atttcgactg gaacagtgg 1200
 gccggcgctc aaggaagcga cgcaggggct ccttcatctg gcactaaggg atctcttgag 1260
 aatcttgttc tcagtctatc tggaaacaac acccagcaga ttgataacaa taatgccact 1320
 gcttcaatga gcaatcaagc aaacgtaaac tatggcaatc ttcttcagag tatgcaaggg 1380
 aacgcacaaa atggaaacat caatgacctt cttcagagta tgcaccagtc tgctaacaac 1440
 aacaacaaca tgaatcagaa tcagaacttt ggaatctcc ttcaaggcat gggcaattca 1500
 ttcttcaga atcccatgat gggtaatgat tttttaacaa taatgaatgg tgggtggcga 1560
 gacttgtctc agcagaatat gatgcatttc aataatacat ttgcaatgca gcagaatccc 1620
 atgatggcgg ctgcaattgc gcaacaacag cttcttgcctc aagcgggagg aaaccccgcc 1680
 attgctaattg tcttagctca acaaggtatt atggcggcga tgacgaacat ggctaacaac 1740
 tttggcaaca actttggcaa tcaaaacaca aacgacttgc ttcagcagtt aattgccag 1800
 caacaaggag aaagtggata caataaccaa ctacaacagc agcaacatca taatcatcat 1860
 cagcaacatc agcagcagca gcagcagcag caacaacagg ggcaagcaca ggaggggaac 1920
 aagcgttaact tcgacaaaat gaagcaaggt ggagggattg accaaggtgg agacgatgga 1980
 caggctgcaa agaaacaaca atgcgacggt tga 2013

<210> 74
 <211> 670
 <212> PRT
 <213> *Thalassiosira pseudonana*

<400> 74

Met Gln Ala Ala Arg Ser Ala Ala Lys Ala Asp Gly Val Ala Gln Pro
 1 5 10 15

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

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

Asn Ile Met Asn Gly Gly Gly Gly Asp Leu Ser Gln Gln Asn Met Met
 515 520 525

His Phe Asn Asn Thr Phe Ala Met Gln Gln Asn Pro Met Met Ala Ala
 530 535 540

Ala Ile Ala Gln Gln Gln Leu Leu Ala Gln Ala Gly Gly Asn Pro Ala
 545 550 555 560

Ile Ala Asn Val Leu Ala Gln Gln Gly Ile Met Gly Gly Met Thr Asn
 565 570 575

Met Ala Asn Asn Phe Gly Asn Asn Phe Gly Asn Gln Asn Thr Asn Asp
 580 585 590

Leu Leu Gln Gln Leu Ile Ala Gln Gln Gln Gly Glu Ser Gly Tyr Asn
 595 600 605

Asn Gln Leu Gln Gln Gln Gln His His Asn His His Gln Gln His Gln
 610 615 620

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gly Gln Ala Gln Glu Gly Asn
 625 630 635 640

Lys Arg Asn Phe Asp Lys Met Lys Gln Gly Gly Gly Ile Asp Gln Gly
 645 650 655

Gly Asp Asp Gly Gln Ala Ala Lys Lys Gln Gln Cys Asp Val
 660 665 670

- <210> 75
- <211> 1902
- <212> PRT
- <213> *Phaeodactylum tricornutum*
- <400> 75

Ala Thr Gly Thr Cys Ala Ala Gly Thr Cys Thr Ala Thr Thr Gly Ala
 1 5 10 15

Thr Ala Cys Ala Thr Gly Cys Cys Gly Cys Cys Cys Ala Ala Ala Gly
 20 25 30

Cys Gly Cys Gly Gly Cys Gly Cys Ala Gly Gly Cys Cys Gly Ala Cys
 35 40 45

Gly Gly Thr Cys Thr Cys Thr Cys Thr Gly Thr Gly Cys Cys Thr Gly
 50 55 60

Gly Gly Cys Ala Thr Cys Gly Cys Thr Thr Gly Gly Ala Gly Gly Cys
 65 70 75 80
 Gly Gly Cys Gly Cys Ala Cys Thr Cys Cys Ala Thr Gly Gly Cys Gly
 85 90 95
 Gly Cys Gly Cys Cys Ala Cys Cys Gly Cys Ala Gly Cys Ala Gly Cys
 100 105 110
 Cys Thr Cys Ala Cys Cys Cys Thr Gly Cys Thr Cys Thr Gly Gly Ala
 115 120 125
 Thr Gly Thr Ala Ala Thr Thr Cys Cys Cys Ala Ala Cys Ala Ala Thr
 130 135 140
 Ala Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Cys Cys Cys Gly Cys
 145 150 155 160
 Cys Cys Cys Thr Thr Cys Gly Thr Cys Gly Ala Gly Gly Gly Ala Ala
 165 170 175
 Ala Thr Gly Gly Ala Cys Thr Cys Cys Ala Gly Ala Gly Gly Ala Gly
 180 185 190
 Gly Ala Ala Gly Cys Gly Thr Ala Cys Gly Cys Gly Ala Ala Thr Cys
 195 200 205
 Gly Cys Thr Thr Gly Ala Thr Thr Cys Thr Ala Gly Ala Ala Thr Thr
 210 215 220
 Cys Ala Ala Ala Thr Cys Thr Gly Gly Cys Cys Thr Thr Thr Thr Gly
 225 230 235 240
 Cys Cys Cys Cys Thr Gly Ala Cys Gly Gly Ala Thr Gly Gly Gly Ala
 245 250 255
 Cys Thr Ala Cys Ala Thr Thr Gly Cys Gly Thr Ala Cys Cys Thr Thr
 260 265 270
 Cys Thr Thr Gly Thr Cys Cys Ala Ala Ala Thr Thr Gly Cys Thr Cys
 275 280 285
 Ala Ala Cys Thr Gly Cys Gly Ala Thr Cys Cys Cys Ala Thr Gly Cys
 290 295 300
 Gly Thr Ala Thr Thr Thr Cys Cys Ala Ala Ala Ala Ala Gly Thr Thr
 305 310 315 320

Cys Gly Thr Gly Gly Gly Ala Ala Gly Cys Ala Ala Cys Thr Gly Cys
 325 330 335
 Ala Thr Cys Gly Gly Cys Ala Ala Gly Cys Ala Ala Gly Thr Cys Thr
 340 345 350
 Thr Cys Cys Gly Ala Ala Gly Ala Cys Gly Cys Ala Cys Gly Gly Cys
 355 360 365
 Ala Gly Ala Thr Cys Thr Cys Ala Ala Cys Ala Gly Gly Cys Thr Ala
 370 375 380
 Ala Cys Ala Cys Cys Gly Gly Ala Gly Cys Ala Gly Ala Thr Thr Cys
 385 390 395 400
 Ala Gly Cys Ala Ala Ala Gly Thr Cys Gly Cys Gly Cys Cys Gly Ala
 405 410 415
 Ala Cys Thr Gly Ala Gly Cys Gly Ala Ala Cys Thr Cys Gly Ala Ala
 420 425 430
 Cys Gly Cys Cys Gly Ala Thr Thr Thr Thr Thr Gly Gly Ala Gly Cys
 435 440 445
 Gly Thr Gly Thr Thr Gly Cys Gly Cys Ala Ala Ala Cys Cys Ala Ala
 450 455 460
 Thr Cys Gly Gly Gly Thr Cys Ala Ala Gly Thr Cys Gly Thr Cys Cys
 465 470 475 480
 Gly Gly Thr Gly Thr Thr Gly Gly Cys Gly Gly Gly Gly Cys Gly Gly
 485 490 495
 Cys Thr Thr Cys Cys Gly Cys Gly Gly Cys Thr Cys Cys Cys Gly Ala
 500 505 510
 Ala Gly Cys Thr Ala Thr Cys Ala Thr Cys Ala Thr Gly Gly Gly Gly
 515 520 525
 Ala Gly Ala Cys Cys Cys Ala Ala Ala Ala Thr Ala Gly Ala Ala Cys
 530 535 540
 Ala Cys Gly Ala Gly Ala Cys Ala Ala Thr Gly Gly Ala Gly Cys Cys
 545 550 555 560
 Thr Cys Cys Cys Ala Gly Thr Cys Cys Thr Cys Cys Gly Thr Gly Gly
 565 570 575

Thr Thr Gly Cys Ala Gly Cys Cys Thr Cys Cys Gly Thr Thr Thr Gly
 580 585 590

Gly Gly Thr Ala Cys Ala Ala Ala Cys Ala Ala Gly Gly Cys Gly Cys
 595 600 605

Gly Gly Gly Ala Gly Cys Gly Gly Cys Ala Thr Thr Cys Gly Cys Gly
 610 615 620

Gly Cys Thr Gly Cys Thr Ala Ala Thr Cys Thr Cys Thr Cys Cys Gly
 625 630 635 640

Gly Ala Thr Cys Cys Ala Ala Cys Ala Gly Thr Cys Gly Cys Gly Cys
 645 650 655

Cys Gly Cys Cys Gly Cys Cys Gly Cys Thr Gly Gly Ala Cys Gly Cys
 660 665 670

Gly Cys Cys Ala Thr Gly Cys Thr Cys Gly Cys Thr Gly Gly Thr Ala
 675 680 685

Cys Ala Ala Gly Thr Gly Gly Thr Cys Thr Cys Ala Ala Cys Gly Gly
 690 695 700

Ala Gly Ala Thr Cys Gly Cys Cys Gly Gly Gly Cys Gly Ala Gly Thr
 705 710 715 720

Gly Gly Gly Ala Thr Gly Ala Gly Thr Thr Cys Ala Cys Ala Gly Gly
 725 730 735

Ala Ala Cys Thr Thr Thr Thr Gly Gly Cys Cys Ala Thr Gly Gly Cys
 740 745 750

Cys Gly Ala Ala Thr Thr Thr Cys Ala Ala Cys Gly Thr Cys Ala Ala
 755 760 765

Gly Cys Ala Thr Cys Cys Cys Ala Ala Thr Cys Gly Ala Cys Cys Ala
 770 775 780

Thr Gly Ala Thr Gly Cys Ala Gly Ala Ala Thr Cys Cys Cys Thr Thr
 785 790 795 800

Thr Cys Ala Thr Cys Ala Ala Gly Gly Gly Thr Cys Gly Gly Cys Ala
 805 810 815

Ala Cys Cys Ala Ala Cys Thr Thr Gly Cys Thr Cys Gly Cys Thr Gly

			820						825						830		
Cys	Thr	Ala	Cys	Gly	Cys	Gly	Gly	Thr	Cys	Gly	Gly	Gly	Cys	Ala	Gly		
		835					840						845				
Thr	Ala	Gly	Cys	Ala	Ala	Cys	Gly	Gly	Cys	Cys	Thr	Thr	Thr	Cys	Thr		
	850					855					860						
Ala	Gly	Cys	Gly	Cys	Ala	Gly	Cys	Gly	Thr	Thr	Cys	Thr	Cys	Gly	Cys		
865					870					875					880		
Ala	Ala	Thr	Thr	Ala	Gly	Cys	Gly	Cys	Ala	Ala	Ala	Gly	Cys	Gly	Thr		
				885					890					895			
Gly	Thr	Cys	Gly	Gly	Cys	Thr	Gly	Cys	Cys	Cys	Gly	Ala	Cys	Thr	Thr		
			900					905					910				
Thr	Cys	Thr	Gly	Gly	Ala	Cys	Thr	Gly	Cys	Ala	Ala	Ala	Gly	Cys	Ala		
		915					920						925				
Ala	Cys	Ala	Ala	Thr	Gly	Cys	Ala	Thr	Cys	Cys	Ala	Thr	Gly	Ala	Ala		
	930					935					940						
Cys	Ala	Ala	Thr	Thr	Thr	Ala	Ala	Thr	Gly	Cys	Thr	Ala	Ala	Ala	Gly		
945					950					955					960		
Ala	Cys	Gly	Gly	Gly	Ala	Cys	Thr	Cys	Thr	Cys	Thr	Cys	Gly	Ala	Gly		
				965					970					975			
Ala	Ala	Cys	Ala	Gly	Thr	Thr	Gly	Ala	Cys	Thr	Cys	Ala	Ala	Cys	Thr		
			980					985					990				
Ala	Gly	Cys	Ala	Ala	Gly	Ala	Gly	Ala	Cys	Cys	Gly	Thr	Gly	Gly	Thr		
		995					1000						1005				
Cys	Thr	Cys	Thr	Cys	Gly	Thr	Cys	Ala	Gly	Ala	Thr	Thr	Cys	Cys			
1010						1015						1020					
Cys	Thr	Cys	Thr	Cys	Gly	Ala	Ala	Cys	Ala	Thr	Gly	Ala	Thr	Cys			
1025						1030					1035						
Cys	Ala	Gly	Cys	Gly	Ala	Cys	Ala	Gly	Ala	Ala	Cys	Thr	Cys	Gly			
1040						1045					1050						
Thr	Thr	Cys	Gly	Ala	Thr	Gly	Cys	Thr	Cys	Thr	Thr	Ala	Thr	Gly			
1055						1060					1065						

Thr Cys Gly Thr Thr Gly Gly Ala Thr Thr Thr Thr Cys Ala Ala
 1070 1075 1080
 Ala Gly Thr Cys Thr Thr Cys Ala Ala Thr Cys Cys Ala Thr Cys
 1085 1090 1095
 Gly Ala Thr Ala Ala Thr Cys Thr Thr Gly Cys Cys Ala Ala Cys
 1100 1105 1110
 Thr Thr Ala Ala Thr Thr Cys Ala Ala Ala Cys Ala Gly Gly Gly
 1115 1120 1125
 Ala Cys Cys Gly Cys Ala Gly Gly Ala Ala Gly Thr Cys Ala Thr
 1130 1135 1140
 Thr Cys Cys Ala Ala Thr Ala Thr Ala Cys Cys Gly Gly Ala Ala
 1145 1150 1155
 Thr Cys Gly Gly Gly Thr Ala Thr Gly Ala Ala Gly Ala Ala Cys
 1160 1165 1170
 Gly Cys Gly Gly Ala Cys Thr Thr Thr Gly Gly Cys Thr Ala Thr
 1175 1180 1185
 Thr Cys Thr Thr Cys Ala Cys Gly Ala Cys Ala Gly Ala Ala Cys
 1190 1195 1200
 Ala Thr Ala Ala Cys Gly Gly Gly Ala Gly Cys Thr Thr Cys Thr
 1205 1210 1215
 Thr Cys Thr Gly Gly Ala Gly Ala Cys Thr Thr Ala Thr Cys Gly
 1220 1225 1230
 Ala Ala Cys Gly Cys Cys Gly Cys Thr Cys Gly Thr Cys Gly Cys
 1235 1240 1245
 Thr Thr Ala Gly Cys Gly Ala Gly Cys Gly Cys Gly Gly Gly Ala
 1250 1255 1260
 Cys Gly Ala Ala Thr Gly Gly Ala Ala Ala Gly Thr Cys Thr Thr
 1265 1270 1275
 Cys Thr Gly Cys Ala Ala Thr Cys Ala Ala Thr Gly Thr Cys Cys
 1280 1285 1290
 Ala Ala Cys Ala Ala Thr Ala Ala Thr Thr Thr Thr Cys Ala Cys
 1295 1300 1305

Ala Ala Cys Ala Ala Thr Gly Gly Gly Ala Thr Cys Gly Gly Thr
 1310 1315 1320

Gly Gly Ala Ala Ala Cys Ala Ala Thr Gly Ala Thr Ala Cys Gly
 1325 1330 1335

Cys Cys Cys Gly Ala Thr Thr Cys Thr Ala Ala Cys Gly Thr Gly
 1340 1345 1350

Ala Ala Thr Cys Thr Thr Ala Ala Cys Ala Ala Cys Thr Thr Gly
 1355 1360 1365

Cys Thr Thr Cys Ala Ala Thr Cys Thr Ala Thr Gly Cys Ala Thr
 1370 1375 1380

Gly Gly Thr Gly Gly Thr Ala Gly Cys Ala Thr Gly Cys Thr Cys
 1385 1390 1395

Gly Gly Thr Ala Thr Gly Gly Gly Cys Gly Ala Cys Cys Gly Cys
 1400 1405 1410

Ala Gly Cys Ala Gly Cys Gly Cys Ala Gly Cys Ala Thr Cys Thr
 1415 1420 1425

Cys Thr Cys Cys Thr Thr Gly Gly Thr Thr Cys Cys Gly Gly Gly
 1430 1435 1440

Ala Ala Cys Gly Gly Ala Cys Cys Cys Ala Gly Thr Gly Cys Gly
 1445 1450 1455

Gly Thr Gly Ala Gly Thr Cys Thr Ala Gly Cys Ala Ala Ala Thr
 1460 1465 1470

Thr Thr Gly Cys Thr Cys Cys Gly Cys Cys Ala Ala Gly Ala Thr
 1475 1480 1485

Thr Cys Thr Thr Cys Cys Ala Cys Thr Gly Gly Thr Thr Thr Gly
 1490 1495 1500

Ala Cys Gly Gly Cys Thr Thr Thr Gly Cys Gly Thr Ala Thr Gly
 1505 1510 1515

Cys Ala Gly Gly Ala Cys Gly Gly Thr Thr Thr Ala Ala Ala Thr
 1520 1525 1530

Cys Ala Ala Cys Gly Cys Ala Ala Thr Thr Cys Gly Ala Gly Cys
 1535 1540 1545

Gly Thr Thr Gly Ala Thr Gly Ala Cys Thr Thr Thr Thr Thr Gly
 1550 1555 1560

Ala Gly Cys Thr Thr Gly Gly Thr Thr Gly Cys Gly Gly Cys Cys
 1565 1570 1575

Gly Gly Ala Gly Ala Thr Ala Thr Cys Cys Cys Gly Cys Ala Cys
 1580 1585 1590

Cys Ala Ala Gly Ala Thr Cys Cys Ala Thr Cys Ala Thr Thr Gly
 1595 1600 1605

Cys Thr Gly Ala Ala Thr Gly Thr Thr Cys Cys Ala Thr Thr Gly
 1610 1615 1620

Ala Thr Gly Cys Ala Thr Cys Ala Gly Cys Ala Ala Gly Gly Gly
 1625 1630 1635

Cys Cys Gly Cys Cys Ala Gly Gly Ala Thr Cys Ala Ala Gly Cys
 1640 1645 1650

Gly Ala Ala Gly Cys Thr Gly Cys Thr Gly Cys Cys Ala Ala Gly
 1655 1660 1665

Thr Thr Ala Ala Thr Gly Gly Cys Cys Cys Ala Gly Cys Ala Gly
 1670 1675 1680

Cys Ala Ala Ala Thr Gly Cys Thr Ala Cys Ala Ala Gly Cys Thr
 1685 1690 1695

Thr Cys Gly Gly Gly Gly Ala Ala Thr Thr Cys Thr Gly Cys Ala
 1700 1705 1710

Cys Thr Gly Gly Cys Gly Ala Ala Thr Gly Cys Gly Cys Thr Gly
 1715 1720 1725

Gly Cys Ala Thr Cys Thr Cys Gly Thr Thr Cys Thr Thr Thr Thr
 1730 1735 1740

Gly Gly Thr Ala Ala Thr Thr Thr Gly Cys Ala Ala Ala Ala Cys
 1745 1750 1755

Thr Cys Cys Cys Ala Thr Cys Ala Thr Ala Gly Cys Gly Gly Ala
 1760 1765 1770

Gly Gly Cys Ala Thr Gly Cys Ala Thr Thr Cys Gly Ala Cys Ala

ES 2 682 279 T3

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

Leu Ser Ser Asp Ser Leu Ser Asn Met Ile Gln Arg Gln Asn Ser Phe
 340 345 350
 Asp Ala Leu Met Ser Leu Asp Phe Gln Ser Leu Gln Ser Ile Asp Asn
 355 360 365
 Leu Ala Asn Leu Ile Gln Thr Gly Thr Ala Gly Ser His Ser Asn Ile
 370 375 380
 Pro Glu Ser Gly Met Lys Asn Ala Asp Phe Gly Tyr Ser Ser Arg Gln
 385 390 395 400
 Asn Ile Thr Gly Ala Ser Ser Gly Asp Leu Ser Asn Ala Ala Arg Arg
 405 410 415
 Leu Ala Ser Ala Gly Arg Met Glu Ser Leu Leu Gln Ser Met Ser Asn
 420 425 430
 Asn Asn Phe His Asn Asn Gly Ile Gly Gly Asn Asn Asp Thr Pro Asp
 435 440 445
 Ser Asn Val Asn Leu Asn Asn Leu Leu Gln Ser Met His Gly Gly Ser
 450 455 460
 Met Leu Gly Met Gly Asp Arg Ser Ser Ala Ala Ser Leu Leu Gly Ser
 465 470 475 480
 Gly Asn Gly Pro Ser Ala Val Ser Leu Ala Asn Leu Leu Arg Gln Asp
 485 490 495
 Ser Ser Thr Gly Leu Thr Ala Leu Arg Met Gln Asp Gly Leu Asn Gln
 500 505 510
 Arg Asn Ser Ser Val Asp Asp Phe Leu Ser Leu Val Ala Ala Gly Asp
 515 520 525
 Ile Pro His Gln Asp Pro Ser Leu Leu Asn Val Pro Leu Met His Gln
 530 535 540
 Gln Gly Pro Pro Gly Ser Ser Glu Ala Ala Ala Lys Leu Met Ala Gln
 545 550 555 560
 Gln Gln Met Leu Gln Ala Ser Gly Asn Ser Ala Leu Ala Asn Ala Leu
 565 570 575
 Ala Ser Arg Ser Phe Gly Asn Leu Gln Asn Ser His His Ser Gly Gly
 580 585 590

ES 2 682 279 T3

Met His Ser Thr Thr Ser Ala Ala Ala Leu Ala Met Ala Gln Ala Arg
 595 600 605

Ala Ala Ala Asn Gly Ser Lys Arg Asn Leu Asp Tyr Leu Ser Gly Gly
 610 615 620

Tyr Ile Gly Gln Gly Asp Ser Lys Arg
 625 630

<210> 77
 <211> 792
 <212> ADN
 <213> *Phaeodactylum tricornutum*

<400> 77
 atgctgccac ccccccggc gaacgctgca tctgtgagca aattgccaat gggactccg 60
 tctcogcttt cggttgacct aacaagctct tcagttgtca cagtaggcac aggagggacc 120
 aacaaggccg gctcgtgag gcaaacgcg acaccgtatc atcatcacac acggacgact 180
 tctgttaatc gatcagcgac tgtctcgtca agggaaatgg cctgcaaaa taatacagtt 240
 gcaactcggc tgacgaagta cccaacctcg atcagcaccg gggcgaacac ccatgctaag 300
 ggcacagcga ccaacatggc atcogccacg gcaactggga gcagtttaag atacgcagca 360
 agaccoccaa ctgttccatc cagcgggtgc gtagecgtcg catcttccct tgogcagtcg 420
 gcacagacgc cttcatcatt acatcgacct gtagegggct tgcgcacta cacaatcccg 480
 cccgacgctt catcgatctc taccaaaaag aaaggtggtc aacctctcag aagaggggaag 540
 tggacgaccg aagaggaggc atatgcggtc aggctaatac atgagtttaa atcaggctta 600
 ctcccgctca cggatggaac gactctcagc aactttctat cgaagctgtt aaattgcgac 660
 ccgatgagga tatcgaagaa atttgtgggc aacaattgta ttggaaagca agtctttcga 720
 cggaaagtgg cggacataaa cagtttgacc cctgcgcaga tatcccaaat aagggttgag 780
 ttgagcgagt ga 792

<210> 78
 <211> 263
 <212> PRT
 <213> *Thalassiosira oceanica*

<400> 78

Met Leu Pro Pro Pro Pro Ala Asn Ala Ala Ser Val Ser Lys Leu Pro
 1 5 10 15

Met Gly Thr Pro Ser Pro Leu Ser Val Asp Pro Thr Ser Ser Ser Val
 20 25 30

ES 2 682 279 T3

Val Thr Val Gly Thr Gly Gly Thr Asn Lys Ala Gly Ser Leu Arg Gln
 35 40 45

Asn Ala Thr Pro Tyr His His His Thr Arg Thr Thr Ser Val Asn Arg
 50 55 60

Ser Ala Thr Val Ser Ser Arg Glu Met Ala Leu Gln Asn Asn Thr Val
 65 70 75 80

Ala Thr Arg Leu Thr Lys Tyr Pro Thr Ser Ile Ser Thr Arg Ala Asn
 85 90 95

Thr His Ala Lys Ala Thr Ala Thr Asn Met Ala Ser Ala Thr Ala Leu
 100 105 110

Gly Ser Ser Leu Arg Tyr Ala Ala Arg Pro Ala Thr Val Pro Ser Ser
 115 120 125

Gly Val Val Ala Ala Ala Ser Ser Leu Ala Gln Ser Ala Gln Thr Pro
 130 135 140

Ser Ser Leu His Arg Pro Val Ala Gly Leu Ser His Tyr Thr Ile Pro
 145 150 155 160

Pro Asp Ala Ser Ser Ile Ser Thr Lys Lys Lys Gly Gly Gln Pro Leu
 165 170 175

Arg Arg Gly Lys Trp Thr Thr Glu Glu Glu Ala Tyr Ala Ala Arg Leu
 180 185 190

Ile His Glu Phe Lys Ser Gly Leu Leu Pro Leu Thr Asp Gly Thr Thr
 195 200 205

Leu Arg Asn Phe Leu Ser Lys Leu Leu Asn Cys Asp Pro Met Arg Ile
 210 215 220

Ser Lys Lys Phe Val Gly Asn Asn Cys Ile Gly Lys Gln Val Phe Arg
 225 230 235 240

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

Ile Arg Val Glu Leu Ser Glu
 260

<210> 79
 <211> 633
 <212> ADN

<213> *Nannochloropsis gaditana*

<400> 79

```

atgaagaccg ccgctctgct caccgtctcc tcctcatgg gcgcctccgc ctttgtggcc      60
cccgccccca agttcagccg caccgcgggt gttgcccgca tgcctctoga ggacgaggcc      120
ggcgtgaccg cccccctggg ctactgggac ccgcttggtc tctccgcoga tggatgatgc      180
gagaaattca accgttaccg cggcatcgag atcaagcag gccgagtggc catgcttgcc      240
atgctccaca cctggtgac cggcctcggc gtgaagctcc ccggccttgt ggctgccggc      300
gacggcatcc ccgcctccat gcccgggggc atcaacgcc aacacctcgg cgcttggggc      360
gcacagggat gggcgcaggc gctcctcttc tgctccgcc tcgaggtcct gccccccag      420
aaggaggaca agatccccgg ggatgtgcag ccgcacacct ctgccttcgc caagctcgag      480
gacaagaccg aggaggaggc gctcgcctac cagaacaagg agatcaaca cggccgcctg      540
gccatggttg cctggaccgg agccactgtg ggcgccctcc tcaccaacgg cgaggacccc      600
atcaccaccc ttctctccaa gctcggcaac taa                                     633
    
```

<210> 80

<211> 654

<212> ADN

<213> *Nannochloropsis gaditana*

<400> 80

```

atgcgttttt tcgaacaaaa tggaaggacc gccgctctgc tcaccgtctc ctccctcatg      60
ggagcctctg cctttgtggc ccccgcccc aagttcagcc gcaccgcggg tgttgccgc      120
atgtccttgc agggcgaggc cggcgtgacc gcccccttg gctactggga cccccctggc      180
ttctccgcgg atggtgatgt cgagaagttc aaccgttacc gcgccatoga gatcaagcac      240
ggccgagtgg ccatgcttgc catgctccac accctggtga ccggcctcgg cgtgaagctc      300
ccccgccttg tggctgccgg tgacggcatc ccgcctcca tgcgcgggg catcaacgcc      360
atcacctcgg gcgcttgggc cgcacagggg tggcgcagg tgctcctctt ctgctccgcc      420
ctcgaggtcc tggcccccca gaaggaggac aagatccccg gggatgtgca gcccgacacc      480
tctgccttgc ccaagttcga ggacaagacc gaggaggagg cactcgccta ccagaacaag      540
gagatcaaca acggccgcct ggccatggtt gcctggaccg gagccactgt gggcgccctc      600
ctaccaacg gcgaggaccc catcaccacc ctctggcca agctcggcaa ctaa                                     654
    
```

<210> 81

<211> 672

<212> ADN

<213> *Nannochloropsis gaditana*

<400> 81

```

atgcgctctg ttgccttcat cgtcggcacc acggcagcca tgtctactgc ttttgtcctc      60
    
```

ES 2 682 279 T3

ccctctgcgc ccaaggcacc tcagaccag ctggetgcga agagcaaggc cgtcccttc 120
 ctggaagccc ccaaggccct ggatggctcc ctccccggg acgttggctt cgaccccctg 180
 aacttgctgg acatcgattt tgacttcacc tacttgatgg tgectaccaa gtgggacgag 240
 totgcacgg ggctttoggc gctcaagtgg ttccgggagg ctgagatcaa gcacgggcgc 300
 tttgccatgc ttgccgtctt gggatgggtg gctgtggaca tgggtctgcg ccttcccgtg 360
 gccaaagtac cggggtacaa cgcogtgcag gccacgacg tgttctgtaa gagcggagac 420
 atgacggtag ggctcctggc catoggcttc ctggaggtgg tgatgggcgc aggcactac 480
 gaaatgagca aaggttctga toggggggcc ggcacttca gcttcgacct cttgggcctg 540
 ggcaaggacc ccgccaagta cgcacgctac caggtgtcgg aaatcaagaa cggggcctg 600
 gccatgcttg ctttcggcgg catcgctacg caggtgtgc tgaccaacgg aggctttcct 660
 tactaccagt aa 672

<210> 82
 <211> 816
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 82
 atgtgcggca tattatccat tgataccaag tatctgcgtt gtctgaatat gtgtgatta 60
 gatgtgaata gtgtcagtg ccttgtaaag gtagccttcc atgctccagg atcccgtcta 120
 ccatcggtac gtgctttcct gatctattct aatccgttcc tcacctgcgt ctataacctc 180
 acacagattt gttctctaac taaaattatg aagctcttgt ctttcgcttg cctcatcggc 240
 gccgctgcgc cctttgtgcc ccccatgccc gctacgactc gtgcccggcg cgtgtctcc 300
 atgatggctg agaagtccaa atccatccct tttttgcccc gcccgctgc gctggacggc 360
 actgccccgg gagacgtggg ctttgacctg gtcgggttca caagctggct gcccttggcg 420
 tacttgacgg aggcgagat caagcactgc cgcctcgcca tgttggaac cctcgggtgg 480
 attgtcgogg attttgtcca cctccctggg gatgtccaag ctgttagctc cctggccgcc 540
 cacgacgtgg ctgtaaagtc gggcgccctc gcccaaacc tcactctggac ctccatcgcg 600
 gaagcgatct ccgtcgttgc catctctcag atgctcgagg gctctggccg ccagcccggg 660
 gacttcaagt tcgatcccct gggctttgcc aaagacgagc agactctgaa gaagctgcag 720
 ctgaacgagc tcaagaatgg acgtttggcc atgctcgctc tctctggcat tgtcaccctaa 780
 gccgcctga ccggcactc cttcccgtac atgtaa 816

<210> 83
 <211> 660
 <212> ADN
 <213> *Nannochloropsis gaditana*

ES 2 682 279 T3

<400> 83
atgcgcacca ccgcttgctt gctcgcgggt ctgggcctca ccagagcttt tctggccccc 60
accctcccca ggcgccgctc gttcagccgc accgccgtga ccatgaagtt gggctcctct 120
gcacagggcc tagccggaag cgacatcgag ttccccgagt tcgaccccct gggattcacc 180
aacaacccaa agccccgagc attggactgg taccgtgccc ctgagctcaa gcaogggcgc 240
gttgccatgt tggcggccct gggtcagatc gtgcagcact tctacaccct ccccgacgcg 300
tccggcgtct tctccgctgg agaccggccc atcgaggccc tgaacaaggt ggtggcagag 360
cggcccttgg cagccattca gattggcctg gccatctttg cggcggaggg gctggggcag 420
ttcaaccagg ccaagcccgg gcaggccccc ggggatctgg gctgggaccc tctgaacttg 480
aagagcgacg accccgagat ctacgctaag gtccagaacc gcgagctcaa gaacggggcg 540
cttgcctatga tcgccatcgc cggcatgttg gtgcaggaga acttgacggg cctcggcgtc 600
atcgagcagt gggcacaagg cgacatcaac ccattcgggg acggacgtgg cttcttctag 660

<210> 84
<211> 603
<212> ADN
<213> *Nannochloropsis gaditana*

<400> 84
atgaagttcg cctccctcct gtctttgctt gggcccccct tcttgccctc tgetttcctt 60
gctccggccc ccaagactac tcgcccgcgc ggtgtggtct ccatggtcca atccaaagct 120
ctgcctttct tggctgcccc gaagaagcta gatggttccc tggcgggoga ctttggttcc 180
gaccccatgg ggatctcgga tcaggtggca aatttgaagt acgtacgcgc ggcogagctt 240
aaacacgggc gggcggccat gcttggttcc ctaggctggg tcgtgacgca attcgtccac 300
ctcccggggg agatctacgc ggaatccaac ccgctgaagg cgatcgccgc cgtgcccctc 360
atcagccaag tgcaaatctt tctcttcacg gctgcccacg agctggcgac cctggatagg 420
acgtacacgg cggacaagcc ttgggacctg ggcttcgacc ccatgaacat gagcaagggg 480
aagtccgagc aacagatgaa ggatctggag ctcaaggagc tcaagaatgg ccgcctggca 540
atgatcgcca tcatcggact catggcccag accgcctaca cgggcattcc ccttttctct 600
tga 603

<210> 85
<211> 1218
<212> ADN
<213> *Nannochloropsis gaditana*

<400> 85
atgttctgtg tcgaagcgtc cggggaccgg tccaccccgc atcgtcctcg ggtccttatt 60
tttctccaca accttcagat aaacttgtct atactctcta gtacaatgag tgtcctactt 120

ES 2 682 279 T3

tctctcttgg cgtgcgccgc cacggcctct gctttccttc tcccactcc taccacgcgt 180
 cctgcctcgg cgtcttttgc gtccactgtg ggcgccaagg acatcgtcga cacggccgtc 240
 gacgcaggat ccttcaagac gcttgccacc gcactcacgg ccgcccgcct cgtggagacg 300
 ctgaagggac ccggcccttt cacgggtgtc gcaccacccg acgaggcctt caacaaactc 360
 gaggcagcga ccctgaacgc tctcctggcc gacaaggaga agctcacctc cattcttact 420
 taccacgtcg tgtcaggccg cgtgcccgtc gcgacgtgc tgaagctgac cagcgccaag 480
 actgtgcaag ggggagacgt tgctgtcgtc gtctccagca ccggcgtcaa ggtcaacaac 540
 gcgaacgtgg tggcgacgga cgtggaggcc tcgaacggca tcatccacat catcgattcc 600
 gtcctcatcc ccggtgccgt ccgcccctcc gcggtcgggt ctaaggagggt ggacaagaag 660
 agcctggcct accgcgagggt gatggaggcg ggcgccacgg ctccccttgg cttctttgac 720
 cccatcgggc tcagcaccgg gaagacctt aaggaactga agaagtggcg cgagtccgag 780
 gtcaagcacg gtcgggttgc catgctggcc gtctcggcg tectctcca ggaggtttc 840
 gcgcctttct acaaccccga gaccggcagc agcgacctg gccccccat tttccacttc 900
 caggagctgg aggcctcaa ccctttctc ttcgttttcc tcattctggg catcgccatc 960
 gtcgagtcct tcaccatcag caaaggctgg gagtctccgg aggaaatgcg cgcggggcgc 1020
 aacacgatcg ccggtctccg agacacgtac gtggccggcg acttgaggt cgacccccctc 1080
 ggcttgctc ctacggcgga cgtggacggc ttcacaaacc agcgttccaa ggagctcaac 1140
 aacggtcgtc tggccatgat tgcttggcg ggcatggtc tccaagagct gatctcaac 1200
 tcaaaaatct tctcctaa 1218

<210> 86
 <211> 705
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 86
 atgaagacgg cagcgatttt ctccctgttg gctgttgccg ccaacgcctt tgtgctcct 60
 actcccaaag ctcccagac tgtcgtgaag gggactccct tcgataccgt tgacaaggcg 120
 gaaatggcgg tggagaacaa ggcctgtcc ggcatgagcg tctttgagcg cgcctatggcc 180
 gatttcaacg tccgctacct ccgctggcc gcgttgggt tgggcccag cgtgaaggcc 240
 gagcgtgga acggcggca cgcctatgtt ggggtgatcg ccctgttggc gacggggtac 300
 gccagggcg acgggctgct ccggagggg ggcatggac cgaaggagt gggcacgtgg 360
 gtcacgtca acacggacct tgccacggc gcagttacca ccatccctgc cgcggagcc 420
 gccatcgca togcccacct ccacctggg gcggtgggtg tctttgctgc ctacagcaag 480
 aacagcgtca aggacaggct gctcttggaa ccgggggaga aggacaggc tctgcccgc 540

ES 2 682 279 T3

ctcttccctg cectgcgccc cgggctgacc aaggacgccc agatcctcaa cgggcgcac 600
 gccatgctgg ggctcattgc ccttgctgcc gtogctgccc caacggggca ggacatcctc 660
 tccgtcattg accagggctc gggggggctt cttctcaagg cctga 705

<210> 87
 <211> 888
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 87
 atggcatgcc cccatggcct gagaatcgtt aactttttca ctggtgcccc tcaagagcct 60
 aactttttgt tcttaoqttt cctogctctt cctcccgaa aaactogtgc catgtcctcc 120
 gcttcattca caaatgctcc ttgcccttct ccttccctc agagttcacc taagaagatg 180
 accatgttcc gcaagctggc cctggccctt ctctgctcct ccgcggctgc cttccatgca 240
 ccgcaccccg taagccgcag ctcaagcagc ctccaggccc agaacaaacc cgcgcacccg 300
 cccgctgcag cccctgctgc cgctgctact gggggggcgg ctgctgctcc ccccccccc 360
 gogaagcccc aaccogtgtt ctogaagtgc gtgccgttcc tcttgaagcc caagaacttg 420
 gatgggatgg tggcgatgt gggcttgcac cctctgggcc tggtgagta cgtggacac 480
 ogatggctgc gagaggcga gctgaagaac ggccgcgtgc caatgcttgc cttcttgggc 540
 ttogtggctc aggaatttat ccgcctccc ggagacctgt actcggagcc caacggggctc 600
 aaagccttct tccaagctgg accccagccc ctcaattcaga tcttctctt ctgtgggttt 660
 ttggaattca cctccacaa ggggaagatg acgcaaatgg acatgtttgc ggaocgaaag 720
 cgcgagcccc ggaactttgg ctttgatccc ttgggtttcg gcaaggatcc ctccaaacgc 780
 gagcgactgg cactggcaga gctcaagaac ggacgtctgg ccatgatcgc gatcgggtgg 840
 cttatccacc atgcctcgt gacgggcat gccacgttta gcagctaa 888

<210> 88
 <211> 717
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 88
 atgaagttcc tgcagtcct tactgggtgc ctogccacag cctctgcttt cgtgcctcat 60
 gccacgctct ccgctggcgc acgaagcagc agcgtctga ggatgtcggc atcggacatg 120
 atcgggattg acgtcagac gggcggctgc ttogatcccc tggccttctc caaggatgaa 180
 caatccctgt acaagtaccg gcaagtgcag ctcaagcacg gccagatggc gatgctcgc 240
 tgctgggca cctggtgca gtctacacc caoctcccc acgacgtttt cagcaatccc 300
 cggcccctgg gagcaactcg acagctcctg tccgagcgtc ccttggcaat cctccagatc 360
 gtcatagcca tcactttgat cgaggtgacg agcgggaagc aggaccccca gctggcgc 420

ES 2 682 279 T3

ggtcagttgg gacgcttcgg tgaggccttc aagcccaggg gcgaggccga aatggccgcc 480
 gtccagctca aggagcttaa aaatgggcca cttgcgatgg taagacaagg atggagcttg 540
 ttgagattta gaatgctgtg ggaaggggtc gccatcatcg gacagtgggt gcaagagctg 600
 ttgacgggcc agggccocat tgagcagatc acctccgggc acatctcccc cttoggcgac 660
 ggacagggtc tttctgtaa gtgccagtta aacaaccacc aagtcctcat tggttga 717

<210> 89
 <211> 741
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 89
 atgatgcgcg cegttgtttt gctggggctg ctgtcctctg ccgcgccttt cattcctgcg 60
 actttcaaca caggcaagag ctcatagcc cgcaaaaact tgggtggtggc ctccgtttct 120
 gacttgatcg gagcggacac ggagacgggc gggatctggg atcctctggg cctgagccag 180
 aacgaggggt ccctccgtaa gtaccgcgag tgcgagctca agcacgggcg tgtagccatg 240
 gctgccatgc tcggtatctt tgtccaaggt ctctaccacc ttccggatcc cgtctttagc 300
 aaccctcgcc ctctggtgct cttgcagcag gtctacgagc aacgtcccga ggccatttgg 360
 cagatcatcg taggactggg cgtgatcgag ttcacggggg gcgacaaaa ggaggaccgt 420
 gccctggcg acttgaactt tggttcttct ttcacccca agagcgagaa ggagttcgag 480
 gaattgcagc tgaaggagct gaaaaacggg gcctggcga tgggtggcctc catgggcgcc 540
 ctgctccagg aatacctgac aggccagggg ccggtggagc aagtcttggc tggccacttg 600
 aatcccttgc gtgacggcca agggttcttc tgtaaaaacc cagccgactt cttggcaccg 660
 caaacgggag ggcccatggc gagaacaagc ttgatttttc atttcatogg gcacacggag 720
 aacacttccg gcgtagaata g 741

<210> 90
 <211> 402
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 90
 atgaagttct tggccctcgc caccatcgtg agcactgcct ctgccttctt ggtcctggt 60
 ccgatacaca ccgtaacca gggccgcatg tacatggcog accaggctcc tgtggatgag 120
 gaagtgcggg ccatgcctt gcogttcgtc gccaaagcgc agaacttgaa tggcgagctt 180
 gctggggatg taggctttga tcctttcaag ttctctgaca agggatgatg ggccaaattc 240
 cgcgtggctg agctgaagca tggacgtgtg gccatgcttg cggtagttgg tgtgctcgtg 300
 caagaacttt accagtggaa cgaaaacttt ccctccaaga acttctgga ggccctcaag 360

accgcccccg cectcggtct gctccagctc ttcttctcc tg 402

<210> 91
 <211> 663
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 91
 atgaagacag tatttgctct tgetgcctc ctccatgog cgctggttt tgtggtccg 60
 gtggccttcc ccacagcctc ctccgaggg aaagccgtgc gaggcgcac taccatgatg 120
 gcggaacgat ccaagtcctt tcccttctg atgaagccca agaacttga cgggtccatg 180
 ggggtgacg ttggcttoga ccccttggc ctgtcagaga tcaacgaagt cggcatcgac 240
 ctgtactggt tgcgggaggc ggaactcaag cactgccgcc tgggatgat ggccgaggcc 300
 ggcattctct tctgagggc tgtgggacc gcccccgtt tccctccac caagagccaa 360
 atggatgct tctggaccgt gtaccgag aagccttccc tctggggggc ggcttggta 420
 gctattgca ttttgaagt gatctcgggt gtgccacca cccagggccg tcaaacggg 480
 gaccgagcgc cggcgacta caactcgac cccctgggt ttggaaagga ccccgccaag 540
 ttcaaagact tgcagctcaa ggaagtgaag aacggccgcc ttgccatgat tctgcccga 600
 ggcattgctc ttcaaggtgt ctccactcac caaggtgat tgcaaaactt gaacggaac 660
 tag 663

<210> 92
 <211> 681
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 92
 atgtcacaca gctacttcc tctctcata cagagcacca tgcggtgoc tgetgcctc 60
 atgccacc cggccaccgc ctccgcttc ctgccccg cctccctcc ctctctcc 120
 tcttccgcc gcaactcaagg cgtgtctcc atggacatct cggcatogt gggctccgac 180
 gtggagggtc cgaattoga ccccttggc ctggccaaga acaaggacga ggaaacgctg 240
 ggctgtacc gcgcccga actcaagcac gggcgtgtg gcatgctggc ctccgtcggg 300
 tacctcgtcc aggggtgta ccacttccc gacgggccc tggaggcctc caagccatc 360
 gacgccctgc tcaaggtctc aagcgaacgc ccgtggccg ccgtgcagat cccatcgc 420
 atgccgccc tggagtgct cgtgcctcc atccagaagt acacggcccc gggagacctg 480
 accttgacc ccttggctt gaagcccga gacctgagg agctcggga gctgcagctg 540
 aaggagctga agaacggcg gttggccatg ttggccacgg cggcttggc cggccaggaa 600
 tacgtgactg gccagggcc agtggagcag ctgctctcg gccacatct cccctcgg 660
 gacggccaag gggcgttctg a 681

<210> 93
 <211> 915
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 93
 atgaagacct cccttgttac cttggcgtg gccacagtcc ccgccactgc ettcattgggc 60
 ggcttcatgg gcaagaactt cgcgcgcccc gctaaggtag cctccaccac caccaccacc 120
 atgttctctga acttcggctc caagaaggca ggcgcgcccc ccaagaccgc gcccaagacc 180
 gccctgcca agaaggcggg ggtctctatc cccaaaccgc cgggcaaagt ggccgccaag 240
 cctgcggcca agcccgcttc caagccggtg tccaaagtgg cgcccaagaa agctgcccc 300
 gcgcctaaga aagccgcccc cgcgcccaag aaagccgccc cgacaagggc cgcagccaag 360
 ctggccgccc ccctatccaa ccctgatttc gcggcggtt tgatcggctc ggatgtggag 420
 gcgatccgat tcgatccctg gaacctggcc agtgagcag accccgaagg cctcgcattg 480
 taccgatccg ccgagctcaa gcacggccgc atctgcatgc tcgccccctt ggtctctctc 540
 gtgcagtcct cctaccacct ccccgacgag gtcttcagca actccaaggg cctcgacgag 600
 ctcttcagg tctcagcggg gggcctcag gccatgccc agatcctcat cgcctcggc 660
 gccatcgagg ttgcggggct ggcagccagc gagggcaagg cccccggaga cttcggcttc 720
 gacccctga acctgaagcc caagacggag gaggtttca acgagctgca gctgaaggag 780
 atcaagaacg gacggctggc gatggtggga gtggcgggga tgctcattca ggagaccctg 840
 accaaccagg gcgtgctoga gcagatcaac agtggctact tgagcccttt caacgacggc 900
 cagggcgtgt tctag 915

<210> 94
 <211> 729
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 94
 atgctgttac tctctttcct tgctattatc ggcaactgccc ctgccttctg caagcccacc 60
 ctccctgctg ctgggtctcg cactcgtgcc ggcgctctcc gcatgaacct cgcgagatc 120
 gagctggagg cgggcaagac ctctccttcc cccgatgget tcgatcctct cggctctgctc 180
 aaggataagt cattcaagga gctcaagaag tggcgcgagg cggagctcaa gcaagccgc 240
 gttgccatgc ttgcgctcct gggcacggcc gtgcaagaga acttcacccc cctgtggggc 300
 ttcaacgaga aggagatgga tgggtccatc tttcacttcc aggagatcca aaactctac 360
 cccctttctt ggaccgccc ccttttcatc atcggcatca ttgaggctcg caccatctcc 420
 accggatggg atgagaacat ggccgatct tcccagatcg ccggcgtgaa ggaggactac 480

ES 2 682 279 T3

atctgcggga acctgggcct ggacccccctc aagatcatcg aaaatgacga cgaggaggct 540
 ttctgttctt accgcaacaa gtctcaccgg ttcttgcctt acctcatctt tcaggagctc 600
 aacaacggcc gtctggctat gatcgcagcc gctgggatca ccgtccagga gaagttcgtc 660
 accaacggcc tccccgagtt cgagttccac cgattcgccc tctcggaagt ctacaacttc 720
 ttcttctaa 729

<210> 95
 <211> 903
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 95
 atgaagaccg cagtcctctt cctgtcgggt ctggtcggcg cccaggcttt cctcaccccc 60
 aacgcccggg tgccagccgc aaccaagatg gaggccatca aggcggggaa gggagccaag 120
 gccggttaagg ccgctactgc caacaactac ctcaaggagg tggagcctgg ctcgggccca 180
 ctcgaaaagt ggtacgtccg ggaogaccog accctctcga ccgcccttc ctgggtcaca 240
 cggcccgaga tcggggaocg ctctctggtg ggggacttcg gtttogatcc ttttggcctt 300
 tccaagatct tcgatgtgaa ttggttgccg aatgocggagc tcaagcacgg acgattggcc 360
 atgttgcca cgctgggtat ggttacgcc gagctggtgc aggcocctgc tggattcgag 420
 ggcttcaagt ttgcgccga gttctcggag ctgaacgcc ttaaggcatt gagcgcctg 480
 cctaccttgg gtttgccgca gatcattctg gccatatctt tcgtggagggt ggcgacctt 540
 agcaaggtgt acaatgagcg cttcacgttc gaagacaacc tgacccoctt ggagcgcgaag 600
 aaagtctcc agggccgctt ctctgacctt tctggcgcg ccaagacaca agccaaagcg 660
 ggogtgaacc cgttcgggaa cgcogtggac gtgggttttc aggatccga gaagttcgtg 720
 cccggggacc tgggcttcga ccccgctcgc ttcacagaca acggcatcaa cccggactac 780
 gccctggccg agatcaagca cgcocgctt gccatgctt ggcagccgg catgcttatc 840
 caggaattcc tccaacccaa gggatatctg agccagaocg tcgagtgggc ccagagccag 900
 taa 903

<210> 96
 <211> 714
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 96
 atggatagaa agtggaatga gaatcggcc accaggatgc accacactcc tgtcttcccc 60
 tctccgctc gagtgacatc cacactgtca aatatgaaga ctgcatttgc ctctctagcc 120
 ctctggggcg tctccactgc cttcatgcc actgctcgc gcatgacccg aggcocgagt 180
 gtgcgcatgg ctgtgaacga gatgattggc tcggatggtg agaccaacgg tgtctttgac 240

ccoccttggtc ttgccaagga tgaggcootcc ctctaccggt cccgtttgat tgaactcaag 300
 caocggccggc ttgctatggt ggcttgctc ggtacccttg tgcaatcgtt ctaccatctg 360
 ccogacgagg tcttctccaa cccccgccc ctgcgcgcc tcgcacaagt gtattctgag 420
 cgtcccgtag ccttctggca gatcttctg gctatcggtg ccatcgagct taccatcggc 480
 aagcaggaca ccgccaaccg cgcctccggt gacgtcggct tcggcgtgc tttcatcccg 540
 gacgatggcg aagacttcgc ggctctccag ctcaaggagc ttaagaacgg ccgcttggt 600
 atgatggcca tcattggaca gttcgtgcag gagaagctga cgggccaggg accgatcgag 660
 cagctggttg aggggcactt ttctccctc ggtgatggac agggagcttt ctaa 714

<210> 97
 <211> 327
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 97
 atgaagtgca tcgtatttgc ggcctgctg gcgtctgcca ctgcttttat ggctcctgct 60
 ccgctggtt cgtctcgctc cgcgttgaga atgggactgg aaggccaggg gggctacgac 120
 gtcgagactg gcggcaaacc atgggacccg ctgggcttgc ccgggatctc ggagcgcgcaac 180
 aacttgggca ttaaccgcga catcaagtgg ctgcaggagt ccgagatcaa gcaocggacgg 240
 accgccatgc tggccttctc cggagtgatc gtgcccggat cccttggggg gtatgtgccc 300
 acctaccccc agctatggga agagtag 327

<210> 98
 <211> 879
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 98
 atggtgagca gcgctcgttc ttgccaccat cttaccctcg tccccctct ctctgacttt 60
 acacacgtcg taccatctca cattctctca acacaacatc ttctccacc agcttttttt 120
 tttctgctc gttcaagaat gagggtcatt gctcttctct ctctgcctc gatggcaagt 180
 gccttcctcg cccctacacc cttggcccgt cgtgccaccg gcgcgggtgcg catggccggg 240
 gatgacgacc tgtctactgc gcttctcttc gacaagcgac cgcaccaact ggacggcagc 300
 cttctgggg atgtggggtt tgaccggctc ggcttctoga acaaccctcc ccgtccttgg 360
 ctgatcggag gctcgggccc ctccctgaag tggtaaccgt aagcagagat tgttcacggc 420
 cgagtgcga tgcttgccgc cctcggatgg gtgttcccc acatctatca ccttccgggg 480
 aacgaggacg ttgggtgga tgcgtttgcg aacctgaacc ctttcggggc tttgaccacc 540
 gtcocggccg cgggcctctg gcagatgcc gccacgggtg gcgcgattga gcttttccgc 600

gtgaaccgtg tcatccgtgg agacaaggag gccgggtgatc tcggactcgg ccagggagag 660
 gggcgatgga accccttcgg attcaactac tcocaggagg agtatgccga gaagcagttg 720
 caggagatca agaacggacg tctggcgatg gtgggcatct tgggtcttct tttgcaggcc 780
 agtgtgactg ggaagggtat tgtgcagcag ctggggggct ctttcgacgt gcctgaagct 840
 gtctccaagg ctggctacta cttccctgac ggcgtctaa 879

<210> 99
 <211> 633
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 99
 atgaagaccg ccgctctgct cacogtctcc tccctcatgg gcgcctccgc ctttgtggcc 60
 cccgccccca agttcagccg caccgcgggt gttgcccgca tgccttcga ggacgaggcc 120
 ggggtgaccg cccccctggg ctactgggac ccgcttggtt tctccgcga tggatgatgc 180
 gagaaattca accgttaccg cgcctcagag atcaagcacg gccgagtggc catgcttgcc 240
 atgctccaca ccctggtgac cggcctcggc gtgaagctcc ccggccttgt ggctgcgggg 300
 gacggcatcc ccgctccat gcccggggc atcaacgcc tcaacctcgg cgttgggcc 360
 gcacagggat gggcgcaggt gctcctcttc tgctccgccc tcgaggtcct ggccccccag 420
 aaggaggaca agatccccgg ggatgtgcag ccgacacct ctgccttcgc caagctcag 480
 gacaagaccg aggaggaggc gctgcctac cagaacaagg agatcaaca ccgccgcctg 540
 gccatggttg cctggaccgg agccactgtg ggogccctcc tcaccaacgg cgaggacccc 600
 atcaccacc ttctctcaa gctcggcaac taa 633

<210> 100
 <211> 726
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 100
 atgcttgccc catgtccgac gggggggcga ccccgctgg acgtgcgctt gcagaacgtt 60
 tactccctcc tctagacaa tggaatccac tcccccttct tcaaaatgaa gtccgcgatg 120
 gtacttgctg gcagcttggc tgtggcctct gcattcgttc ccgcggcccc taagatgtct 180
 cgcactcgtg ggatgacacg catggcctg aacgacatcc tgggctcaga cgtggagact 240
 ggogcgctgt gggaccccc caatttttcc aaagatgagg gcagcctgta cogatatcgc 300
 gccgtggagc tcaagcacgg ccgattggcg atgcttgccg tgtaggcct gtgggtctct 360
 gagttctatc acccctcta cgacggaaag atctcaccgg gcatcaaggc catcggggaa 420
 ctgcggggcc cggcctggct gcagatcctt gccactattg gtgtgattga actgaccgtg 480
 ggaaagcagg acacggagaa taaggcaccg ggcgacctcg gcttcgggta caacttaac 540

ES 2 682 279 T3

cccttcaaga acgacccoga gaagttoGCC gagctgcaac tcaaagaact taaaaacggg 600
 cgcttggcca tgcttggcgc ggcaggaatt ctgctgcagg agagtatcac cggccagact 660
 tgcttgcagc agatgcgcgc gcagcatctc agcccgttg gtgaocggaca gggcttttc 720
 tgggga 726

<210> 101
 <211> 714
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 101
 atgattgccc agcctgctcg gctcttgctc tgcttggccc tcccggccct tgccaccgcc 60
 ttcttggccc gtgctcccac cacagtacct cacactcact cggtgaggaa agggcgtact 120
 gttgttatgg agtctatctc ggagttcctc aagtttgatg cgaactcgg cgagttggag 180
 gtgaccaaag cgggagtctc tgctccattc gggttcttcg atcccttgaa cttttacgga 240
 ggcaaaacca ttgcacagaa gaaaaagctg cgcgagtctg aactcaaca cggctgcgctt 300
 gctatgatgg cggtgctggg cattttcgtc caagaactct ggcacacctc tctggcaggt 360
 gatgatttcg aaaccgcgct cagccaatac tacggcgtga cggagtattat tcccgacttc 420
 ggcgoggggg cctctttgt cctttctctt tttgagttca aaagcatcgc tatgggatgg 480
 gatcccgtcg aggagactct caaagagcca aaaattgccg gtatgcgtga tgattatgtg 540
 ggggagact tacagtaoga tctctgggc ttctgtcccg acgatgaoga agcctttgtg 600
 cgcacgcgca ctaaggagct caacaacggg cgtatggcca tgatgcgcgc cattggcacc 660
 atcggacagg agctggccac gggcgtgcct gttgtcaagt cgctctttg gtaa 714

<210> 102
 <211> 1338
 <212> ADN
 <213> *Nannochloropsis gaditana*

<400> 102
 atgttgcgca ctcaatccat cgttacaat tttctactc tcatgcattt cgacacgcag 60
 gaatogaggg cacaatttgg tttctagcgc cgcggctca tcaaggagtc catcgctatg 120
 gtccagagccc ggtgtcggaa cctttcccc tgctgatttc accccacttg cctctttaag 180
 tggcogtcca ttatctgoot ttctacatc ctccgagaat tttacgttgt tacggccttt 240
 gtcccgccat ttccacggc ggagaccccc aggaaccttt tttgtccctc cgtgcaccc 300
 cttccacctc cgcgcgcoga gaccgacctc ctctctactc tggcccaaga cggacgcttc 360
 acacgcctcc tccacctttt caaagaggta gggctogaac ccacctgag tggccctggg 420
 ccgtggactg tgcttgcgcc cacggacgcg gctctggacg accaggtagc caacctcacc 480

ES 2 682 279 T3

ctcatcgtgg ctgcacggga gccaaagcgg cacctgacac cettgctgtc ctaccacatc	540
atcccaggga aagccatggg cctgggagac ctggtggccg ccggtgacgc gaggactcgg	600
gtcaagacgt tggagcgcag cggtttgtcg atcgtctct atggggcac gtcctcgtg	660
gacgaaggcc gtatcctgga gtggggcatg gaagcaggca acggagtcgt gcatgtgatg	720
gacagcctcc ttctccctgc cctcccgtcc gacttcgata ctttcgcgc ccgggagccc	780
gcctcccogg gccccgtcct gggggtgacc gctcccctcg gctgctggga tcctctgggc	840
ttgtggacgg gaaaggacgc cgcaacgcaa gcccgtttgc gtgacgcgga aatcaagcac	900
gcgcgcttgg ccatgctggc tgcaataggg atcctcacgc aggagctggt ggtaccact	960
catcogtaog cgttccagcc tgtattggat gcattgcgtg gcagcacaca catggctccc	1020
gacgtttcat ttctactacc ttgctcctcg attgcggcgt tggcagcgat cettgctccc	1080
ggagtcgcgt gggaaagtgtg gtcactgcga agagcctcga aggacaaggc gccgggcagg	1140
ccatccaagg aggaggcact caccagata ctgaaacgag aatccccoga ctttgaggca	1200
gagagggagg aagagcgtga aaaaatcgat gcggggaata gagaattgaa caacggtaga	1260
ctggctatga tagctgtcgt tggaatgcta gtgcaggaaa tcgtcacggg tcgaagcgtt	1320
ctgagtcctt ggatgtga	1338

REIVINDICACIONES

1. Un mutante de algas desregulado en aclimatación a poca luz, en donde el mutante de algas exhibe al menos una reducción del 20% en clorofila en condiciones de poca luz con respecto a un alga de tipo salvaje de la misma cepa y además en el que el mutante de algas:
 - 5 exhibe un mayor inactivación fotoquímica (qP) a todas las irradiancias fisiológicamente relevantes por encima de $400 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ con respecto a un alga de tipo silvestre de la misma cepa; exhibe un inicio de inactivación no fotoquímica (NPQ) a una intensidad de luz más alta que la intensidad de luz a la cual aparece NPQ en un alga de tipo salvaje de la misma cepa y exhibe un NPQ más bajo en todas las irradiancias fisiológicas por encima de $400 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ que es exhibida por un alga de tipo salvaje de la misma cepa;
 - 10 adicionalmente en el que el mutante está mutado en un gen que codifica un polipéptido que comprende un dominio de dedo de cinc pfam PF02135 TAZ, en el que el polipéptido comprende la proteína LAR1 (SEQ ID N°: 4), la proteína No-LAR1 (SEQ ID N°: 8) o un homólogo de la proteína LAR1 (SEQ ID NO: 4) o proteína No-LAR1 (SEQ ID NO: 8) que tiene al menos un 80% de identidad con SEQ ID NO: 4 o SEQ ID NO: 8, con la condición de que el alga no se obtenga exclusivamente por medio de un proceso esencialmente biológico.
 - 15 2. Un mutante de algas según la reivindicación 1, en donde el mutante tiene aproximadamente la misma tasa fotosintética máxima (P_{max}) que un alga de tipo silvestre de la misma cepa, o en el que el mutante tiene una irradiancia de saturación más alta para la fotosíntesis (EK) que un alga silvestre de la misma cepa.
 3. Un mutante de algas según la reivindicación 1, en el que un cultivo del mutante tiene una mayor penetración de la luz en el cultivo que un cultivo de un alga de tipo silvestre de la misma cepa.
 - 20 4. Un mutante de algas según la reivindicación 1, en donde el mutante muestra la desregulación de al menos veinte genes sensibles a la luz, opcionalmente en donde la abundancia de transcritos de los genes sensibles a la luz difiere en el mutante con respecto a una célula silvestre por al menos $\log_2 1,0$.
 5. Un mutante de algas según la reivindicación 1, en donde el mutante se ha generado por irradiación UV, irradiación gamma o mutagénesis química, o en el que el mutante es un mutante modificado genéticamente, opcionalmente en el que el mutante se ha modificado genéticamente mediante mutagénesis insercional.
 - 25 6. Un mutante de algas de acuerdo con la reivindicación 1, en el que el mutante exhibe una velocidad máxima de transporte de electrones del fotosistema II (ETR_{PSII}) más alta que un alga de tipo silvestre de la misma cepa.
 7. Un mutante de algas según la reivindicación 1, en el que el mutante pertenece a un género seleccionado del grupo que consiste en Achnanthes, Amphiprora, Amphora, Ankistrodesmus, Asteromonas, Boekelovia, Bolidomonas, Borodinella, Botrydium, Botryococcus, Bracteococcus, Chaetoceros, Carteria, Chlamydomonas, Chlorococcum, Chlorogonium, Chlorella, Chroomonas, Chrysothrix, Cricosphaera, Cryptocodinium, Cryptomonas, Cyclotella, Dunaliella, Ellipsoidon, Emiliana, Eremosphaera, Ernodesmus, Euglena, Eustigmatos, Franceia, Fragilaria, Gloeothamnion, Haematococcus, Halocafeteria, Heterosigma, Hymenomonas, Isochrysis, Lepocinlis, Micractinium, Monodus, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Neochloris, Nephrochloris, Nephroselmis, Nitzschia, Ochromonas, Oedogonium, Oocystis, Ostreococcus, Pavlova, Parachorella, Pelagomonas, Phaeodactylum, Phagus, Picochlorum, Platymonas, Pleurochrysis, Pleurococcus, Prototheca, Pseudochlorella, Pseudoneochloris, Pseudostaurastrum, Pyramimonas, Pyrobotrys, Scenedesmus, Skeletonema, Spyrogyra, Stichococcus, Tetraselmis, Thalassiosira, Tribonema, Vaucheria, Viridiella, Vischeria y Volvox.
 - 30 8. Una biomasa de algas que comprende un mutante de algas según la reivindicación 1.
 9. Un método para producir un producto de algas, que comprende cultivar un mutante de algas según la reivindicación 1 y aislar al menos un producto del cultivo, opcionalmente en el que el producto es biomasa de algas, o en el que el producto es un lípido, una proteína, un péptido, uno o más aminoácidos, uno o más nucleótidos, una vitamina, un cofactor, una hormona, un antioxidante o un pigmento o colorante.
 - 35 10. Un método según la reivindicación 9, en el que el alga se cultiva fototróficamente, opcionalmente en el que el alga se cultiva en un estanque o conducto.
 - 45 11. Un método según la reivindicación 9, en el que el producto es un lípido.
 12. Un método según la reivindicación 11, en el que el mutante de algas se modifica por ingeniería genética para incluir al menos un gen exógeno que codifica un polipéptido que participa en la producción de un lípido.

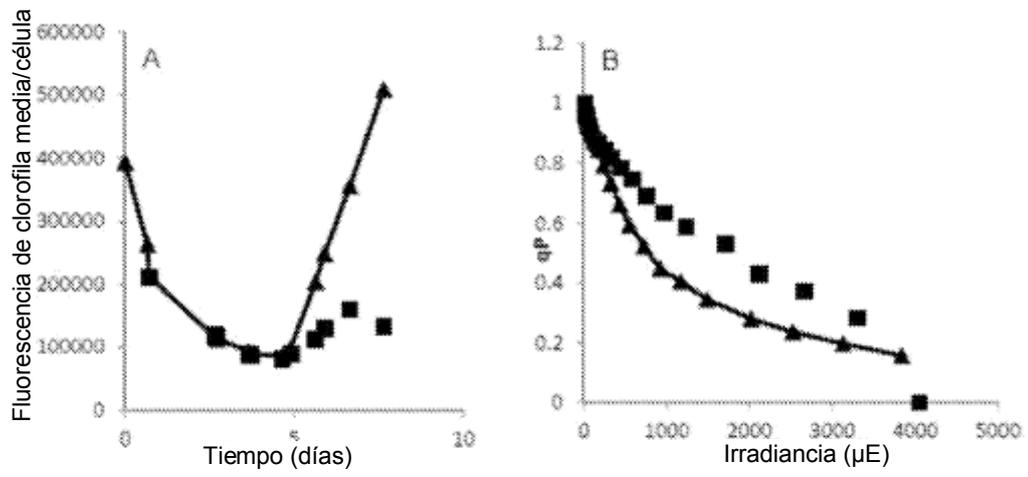


FIG. 1

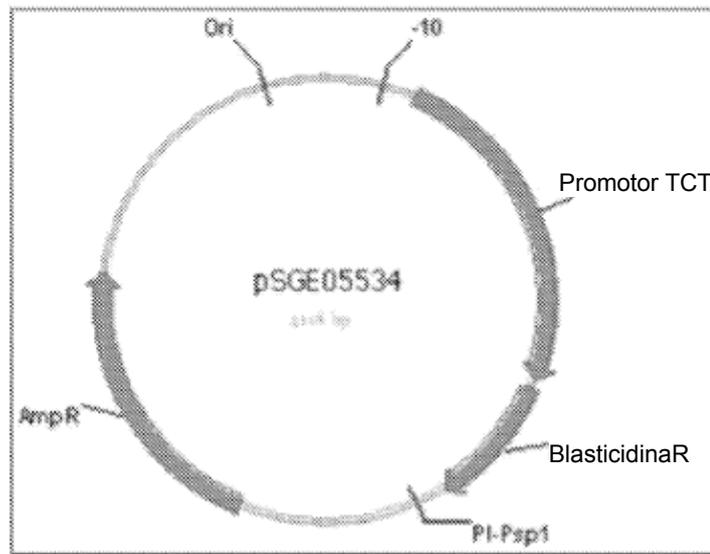


FIG. 2

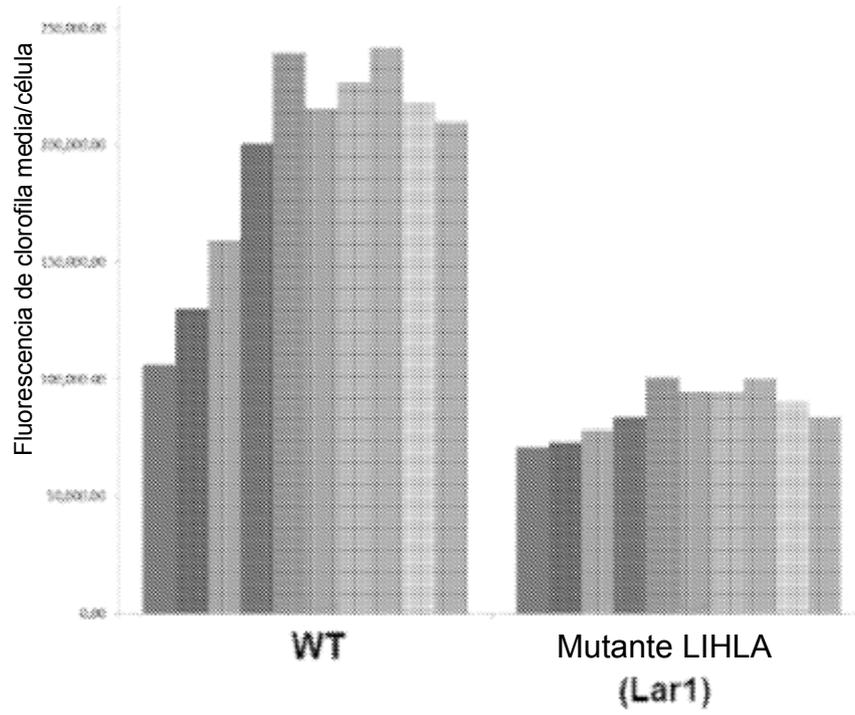


FIG. 3

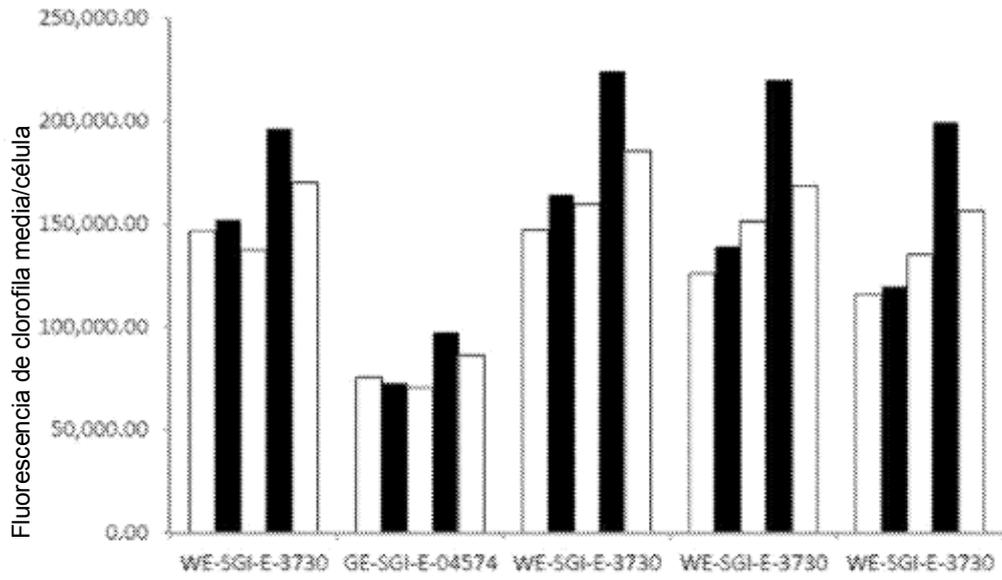


FIG. 4

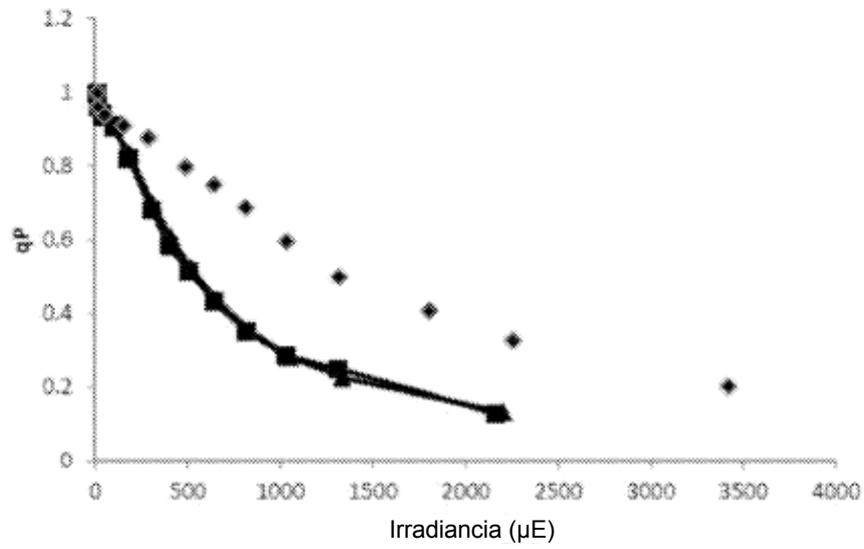


FIG. 5

ES 2 682 279 T3

LIMS ID	Genotipo (mutación detectada)	ETRmax (X veces mayor que WT)	pmol O ₂ /min /célula	μmol O ₂ /min /mg chl	Promedio total Chl (pg/célula)	WT pmol O ₂ /min /célula	WT μmol O ₂ /min/mg chl	Promedio total WT (pg/célula)
GE4574	LAR1	2,62	0,017	405	0,04	0,019	161	0,13
GE4904	LAR1	2,44	0,023	353	0,07	0,028	150	0,21
GE4908	LAR1	2,23	0,024	330	0,07	0,028	150	0,21
GE4911	LAR1	2,1	0,022	320	0,07	0,028	150	0,21
GE5099	LAR1	2,1	0,086	1654	0,05	0,086	534	0,16
GE5117	LAR1	2,6	0,094	1753	0,05	0,086	534	0,16
NE5282	LAR1	2,00	0,029	374	0,08	0,029	145	0,20
NE5284	LAR1	2,22	0,026	429	0,06	0,029	145	0,20
NE5287	LAR1	2,03	0,024	347	0,07	0,029	145	0,20
NE5292	LAR1	2,76	0,031	422	0,07	0,029	145	0,20
GE5406	LAR1	2,22	0,019	254	0,08	0,017	107	0,16
GE5409	LAR1	2,5	0,026	444	0,06	0,025	163	0,16
GE5438	LAR1	2,51	0,022	391	0,06	0,025	163	0,16
GE5439	LAR1	2,79	0,020	374	0,05	0,025	163	0,16
GE5440	LAR1	2,7	0,022	389	0,06	0,029	145	0,20
GE5441	LAR1	2,81	0,023	395	0,06	0,029	145	0,20
GE5492	LAR1	2,91	N/A	225	7,18 mg/L	N/A	101	19,34 mg/L
GE5494	LAR1	2,95	N/A	239	2,72 mg/L	N/A	86	11,75 mg/L
GE5838	LAR1	1,53	0,33	492	0,13	0,037	229	0,16
GE4906	LAR2	2,33	0,023	358	0,07	0,028	150	0,21
GE5098	LAR2	1,47	0,094	1341	0,07	0,086	534	0,16
GE5404	LAR2	2,28	0,025	244	0,10	0,017	107	0,16
GE5405	LAR2	1,88	0,026	220	0,12	0,017	107	0,16
GE5407	LAR2	2,1	0,026	232	0,11	0,017	107	0,16
GE5408	LAR2	2,18	0,035	294	0,12	0,025	163	0,16
GE5491	LAR2	2,47	N/A	237	7,8 mg/L	N/A	101	19,34 mg/L
GE5489	LAR3	1,88	N/A	203	8,58	N/A	101	19,34 mg/L

FIG 6

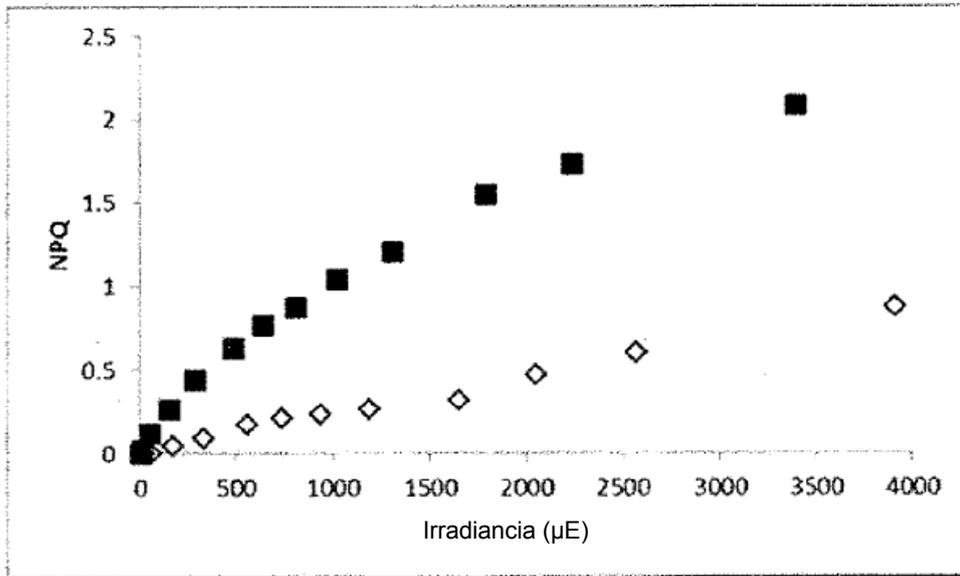


FIG. 7

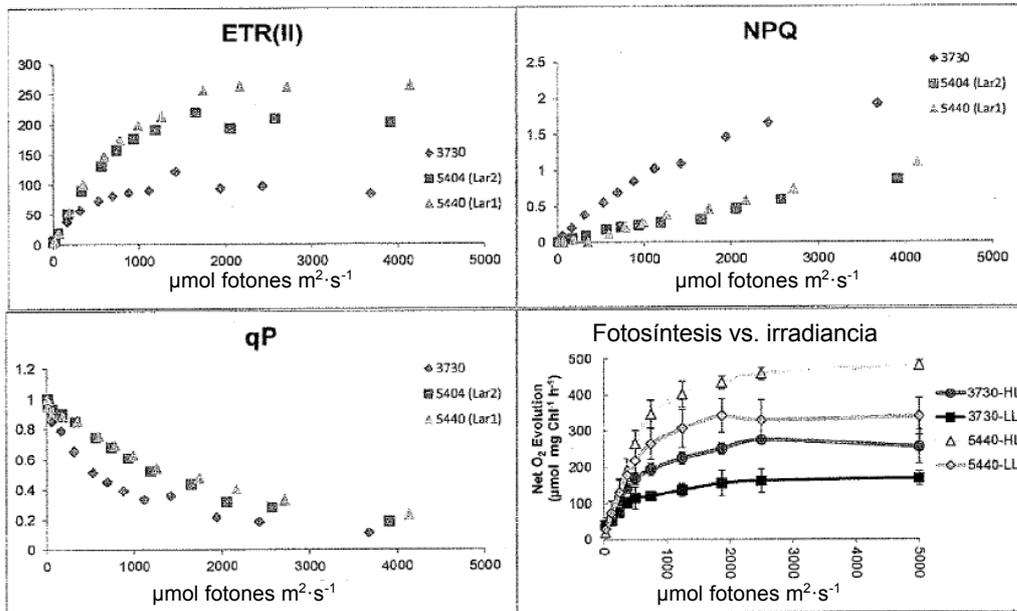


FIG. 8

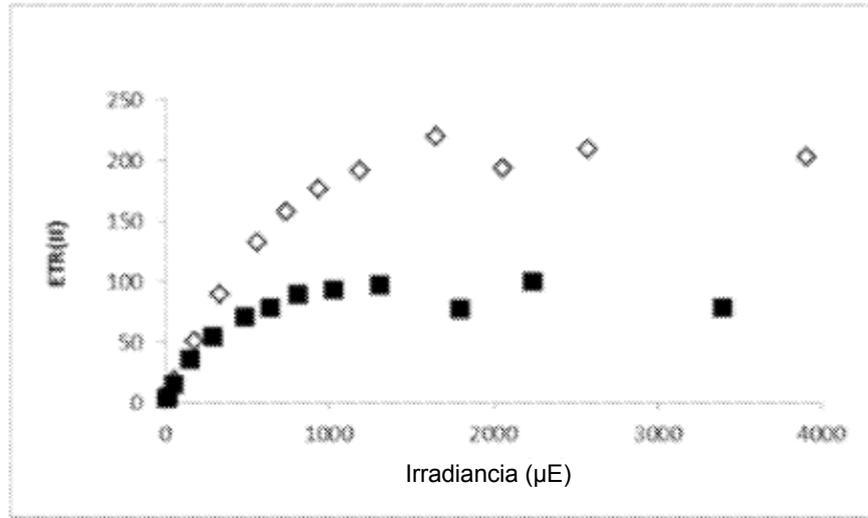


FIG. 9

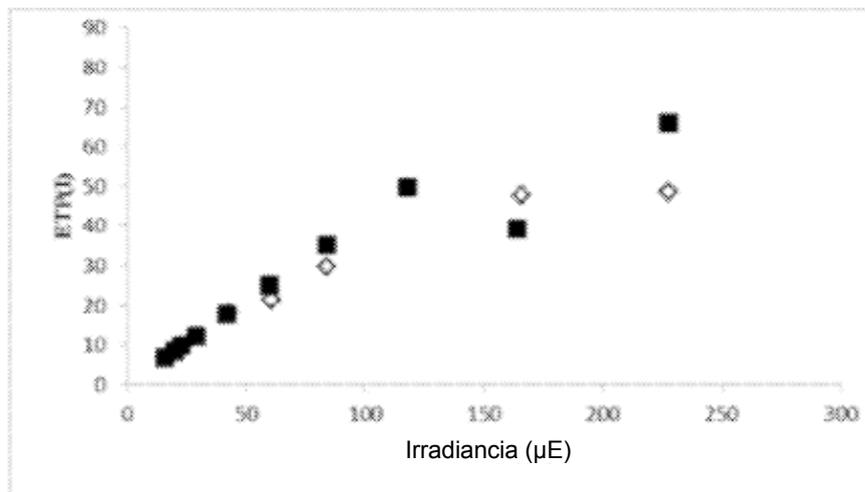


FIG. 10

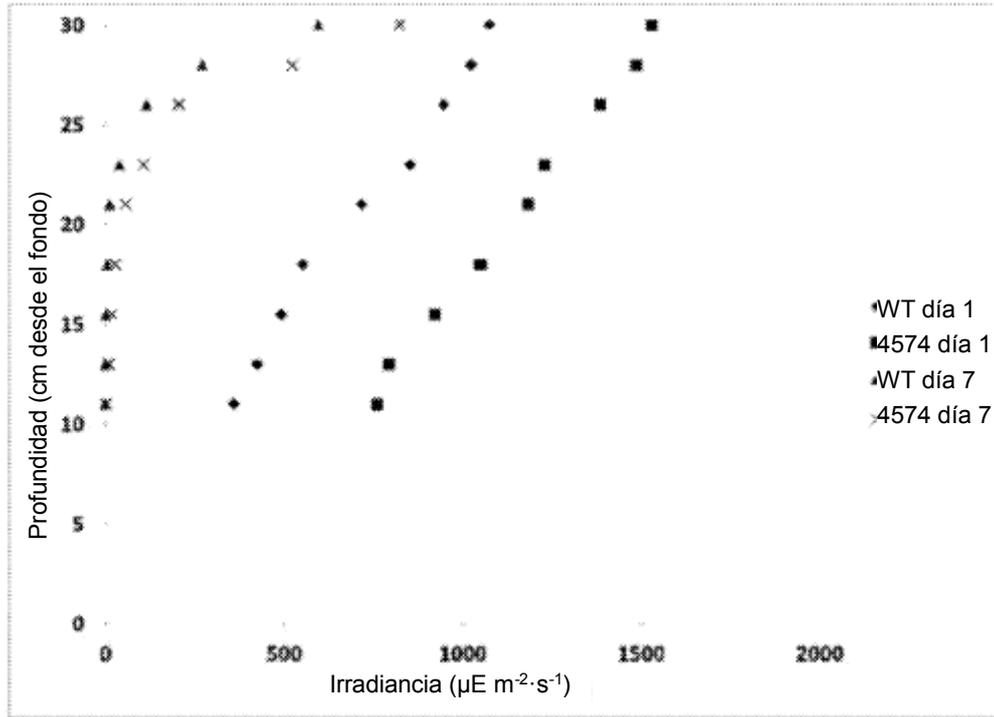


FIG. 11

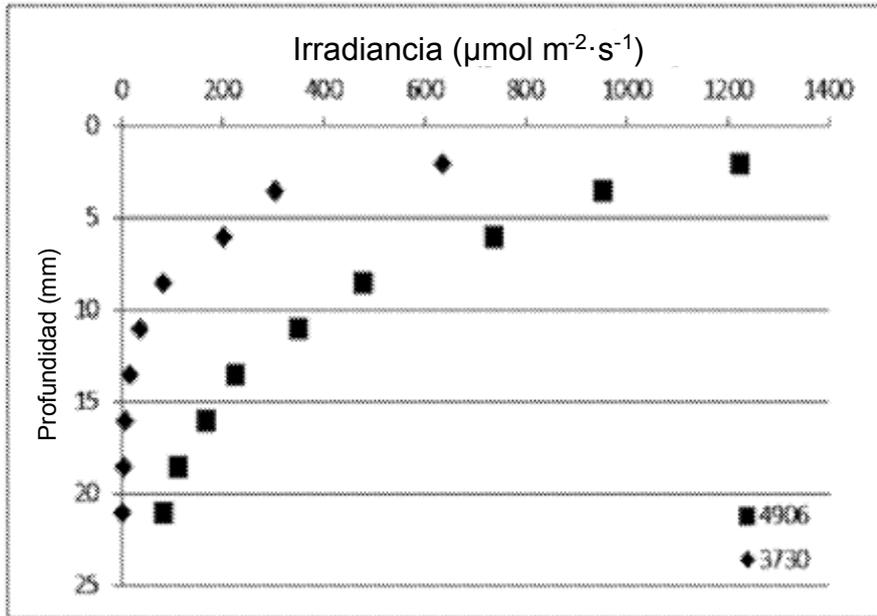


FIG. 12

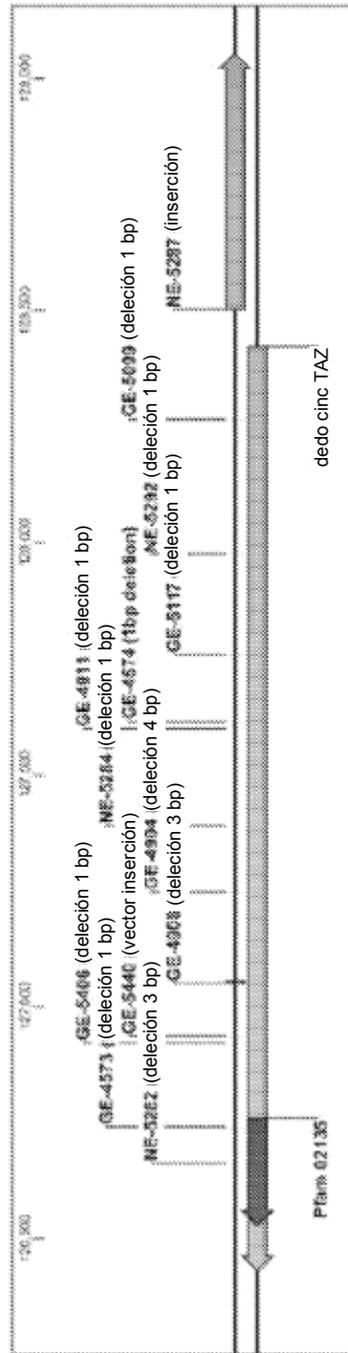


FIG. 13

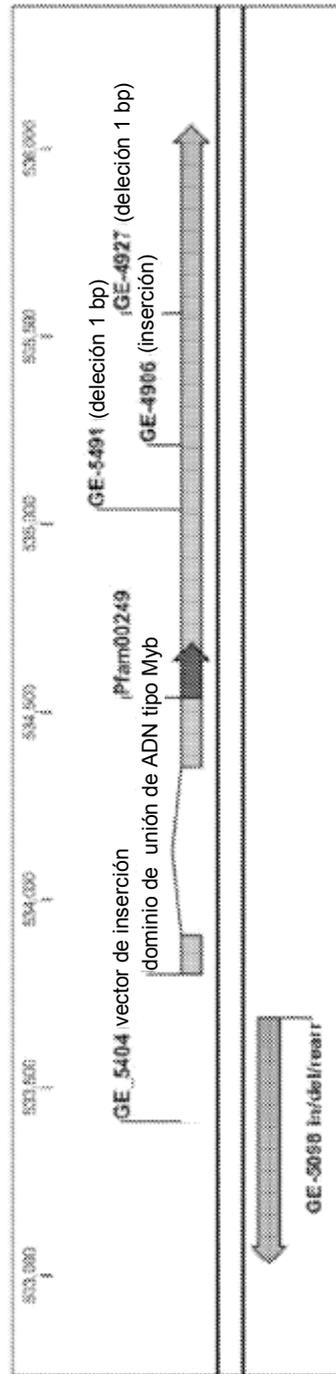


FIG. 14

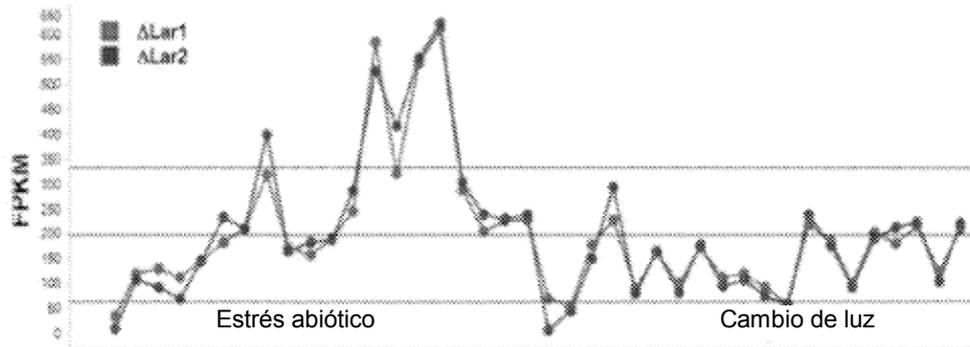
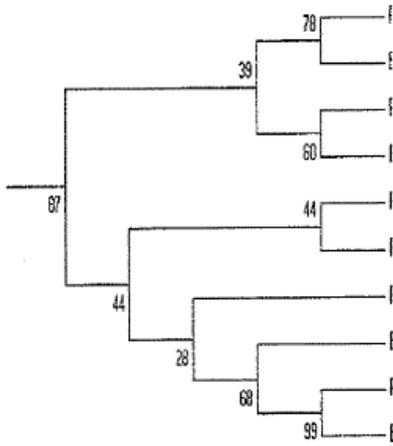


FIG. 15

Figura 17



- Dedo de cinc TAZ *Thalassiosira pseudonana* JGI v3 SGIv2
- 177 Dedo de cinc TAZ 0293 nuclear v1.3 (PASA+rastros intrón+rastros blastx+transcritos no alternativos)
- Dedo de cinc TAZ *Phaeodactylum tricornutum* JGI v2 SGIv2
- 154 Cadena A de proteína de unión CREB WT0229 *Navicula ADNc* ensamblaje CLC
- Cadena A de proteína de unión CREB *Ectocarpus siliculosus* Ghent
- Subunidad del complejo de reconocimiento de origen *Ectocarpus siliculosus* Ghent
- Dedo de cinc TAZ *Aureococcus anophagefferens* JGI v1
- 180 Dedo de cinc TAZ ve03730 anotación genoma nuclear v2012 05 29
- 50 Dedo de cinc TAZ *nannochloropsis oceanica* CCMP1779
- 184 Dedo de cinc TAZ ve05473 *Nannochloropsis Oceanica*, genoma nuclear con anotación v130301

ES 2 682 279 T3

```

EMRE1EUKT5308216 MGIGSEAVAEVGSGTSSAPCGSNSASPMMPVVTASQNAAESVAGSMLPVSCVASAAIAIQ 60
EMRE1EUKT5787835 -----MPMVTLSDATTTAAGSMM-----LPLLPS 25
                        **:** **:*: :.*****:      :. *.

EMRE1EUKT5308216 TSSTTVGVAAASAPAA$PTT$KLHKGT$SMTQLGETGRENTGRWTCEEHVLFKQLEMHGK 120
EMRE1EUKT5787835 IPASATSASFTAPSA$STTTK$PKGT$SMTQLGETGRENTGRWTCEEHVLFKQLEMHGK 85
                        .: :. . . . . : **:** **:* * *****

EMRE1EUKT5308216 GWKKIAKLIKTRTVVQIRTHAQKYFQKLAKAKKNHGHGDMLGMEGTRFPGGKRVKFTGKRR 180
EMRE1EUKT5787835 GWKKIAKLIKTRTVVQIRTHAQKYFQKLAKAKKNHGHGDMLGMEGSHFPGGKRVKFTGKRR 145
                        *****:*****

EMRE1EUKT5308216 GLVYGSYLVGAEATSAAISPALQSYMFGSWAGREEGEAL-$DKEEDAAIEKGLYRFLSPV 239
EMRE1EUKT5787835 GLVYNSYLVGAEATSAAISPALQTFMFPANLGMGERVGLMTDKEEDAAIEKGLYRFLSPV 205
                        ***.*****:***. . . * . * :*****

EMRE1EUKT5308216 VLDAAASNLDATAPEV--LPPSTPGTGVHANGVVA-----DGETT-EEDGSSGGDNVDV 291
EMRE1EUKT5787835 VLDPATRNLDASAPEILPLPSTPAMGVHHTSSRGSSRGGLDGETTGEEDGSSDSIVMGD 265
                        ***.*: ***:***: *****. *** .. *: ***** ***,*.. :.

EMRE1EUKT5308216 SETIDDADSSSGEPLRLARVTNDMYERC$VPTWFMKGGDIEELLADAAAIDWREDSGGD 351
EMRE1EUKT5787835 GGS$DQDAESSLGEPLPTLARVTPEMYTRCGVPEWPKKGGDIDELLIIDAAGLDWRS$SGGD 325
                        . : :**:* ***** ***** :** **.* ** *****:*** ***:***.*****

EMRE1EUKT5308216 AVKAEERGASILNANIESSDQAQSHRNQEKVAVPNRVAAVKVAGNISADSSYITSGASQC 411
EMRE1EUKT5787835 ARKVVDQGT$ILNANINGSNCATVAPAVVRKCGG$NTNKMNSAAPVLNMTGLAGAGGLSG 385
                        * *. :*:*****:*. * : . . . . :*. : :. :*. .

EMRE1EUKT5308216 VNHPTTNTIDCKR$NMNSGQSLPTANGRNGAASGR-----CVTGRGQQQKKKQPKTQES 466
EMRE1EUKT5787835 WKGR-GSDTSEGSS$SKN$MAL$TANASAGCGG$GWVGGGAQKQKQKQHEVPTQQQQ 444
                        : :*:*. ** .* : ***. *..* . : *****: *. *:.

EMRE1EUKT5308216 GNHGKQQLVNKHSAPTG-DMFQSKAGAGTIVLQGDII--NCFASLDPHHVELKEEHSHE 522
EMRE1EUKT5787835 VPTQQQVHGIVHKEEGMELLRVMADRGTV--HGHVHEEDGFAAFDPHHVELKEEHSHE 502
                        **: . * * : : : * . ** : * : : : **:*****:

EMRE1EUKT5308216 LRLEELQQSGVNDTFAHMDFLANDEAPVDH-----HAGHLHT---ISTHDDVHSHA 571
EMRE1EUKT5787835 LLEELPHDSNHDDALAHIVFSVNGESDLHSLPRGAGGGGAHVHAPPVVVGGRQHYYHHN 562
                        * **** .. :***:***: * .*. : :. . .*: :. :. :. : *

EMRE1EUKT5308216 DHDVHMRNYDVFWKDTVPDGDHG-LLLD$FDGGIEF 606
EMRE1EUKT5787835 DNDIHLHAYDAYLEEGADGGHGLLLELDLGGIEF 598
                        *:***: **.: :. .**.* ***:*:*****

```

FIG. 18

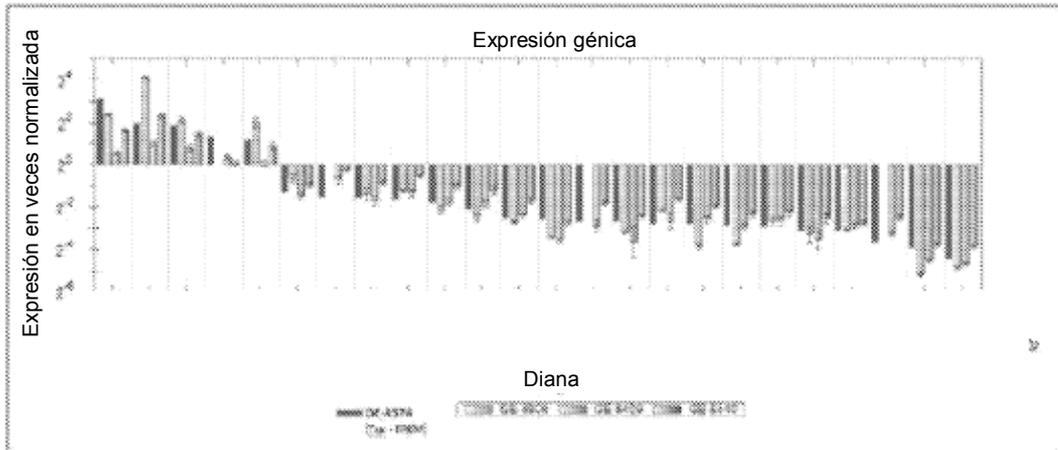


FIG. 19

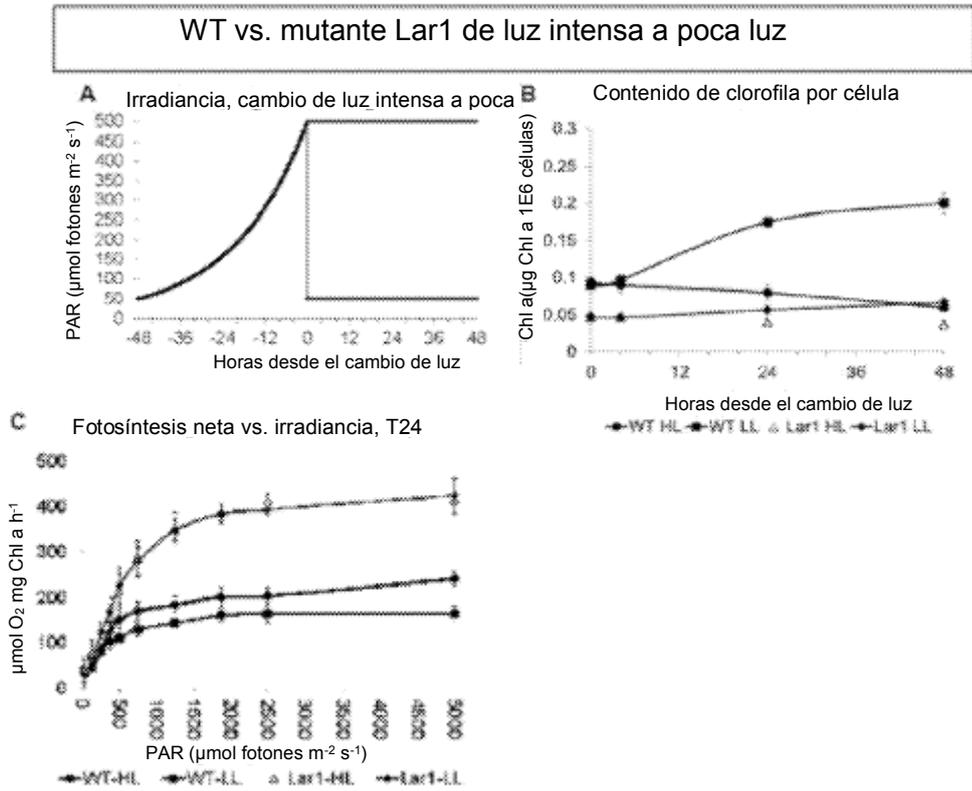


FIG. 20

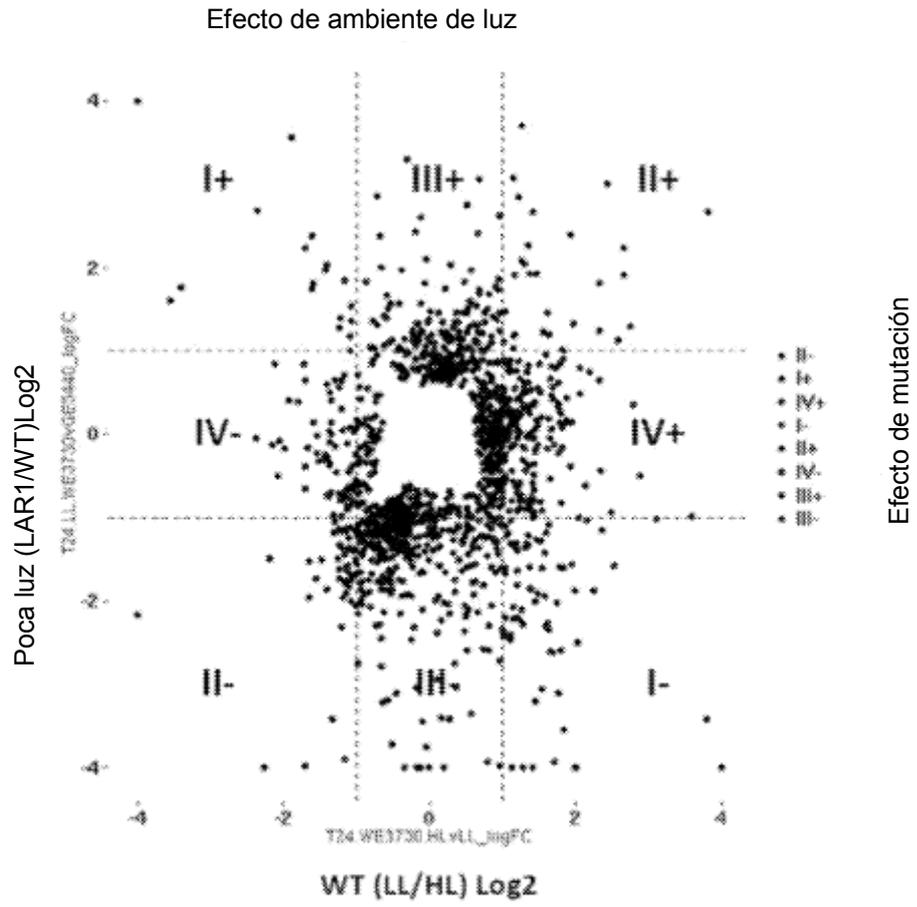


FIG. 21

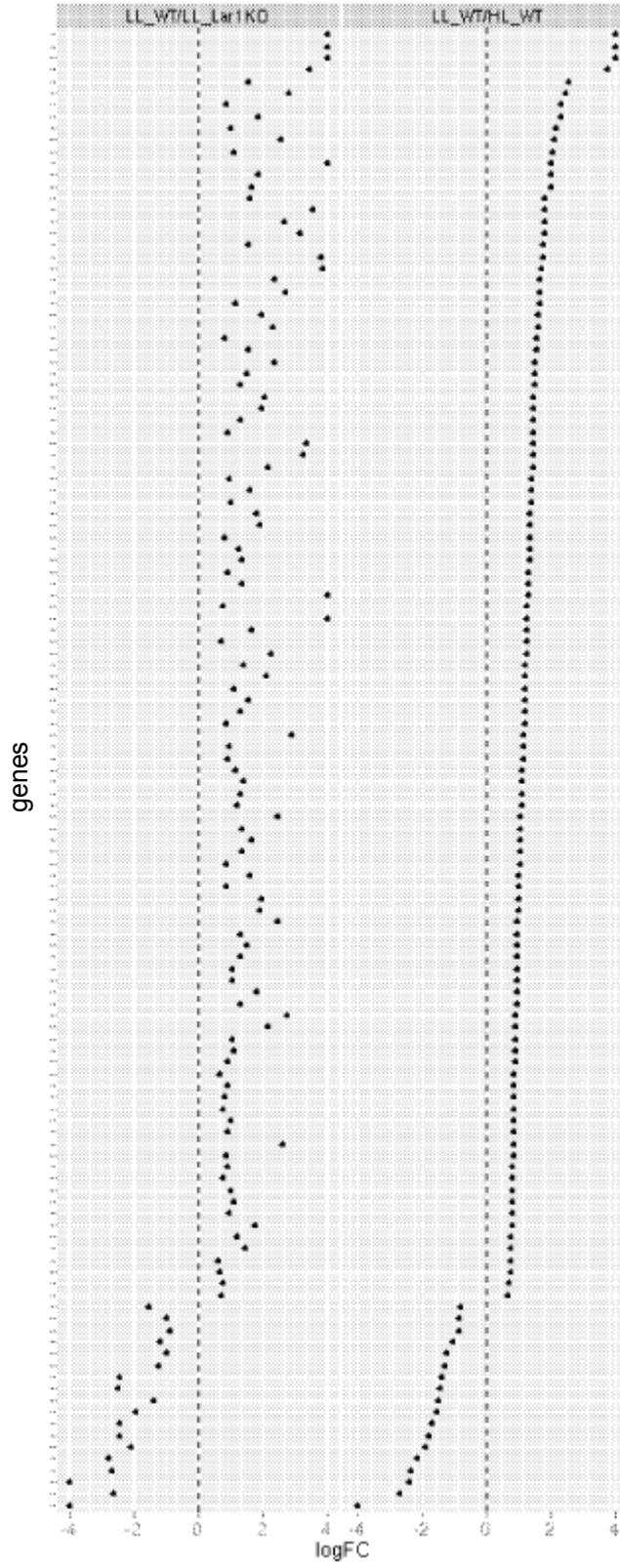


FIG. 22

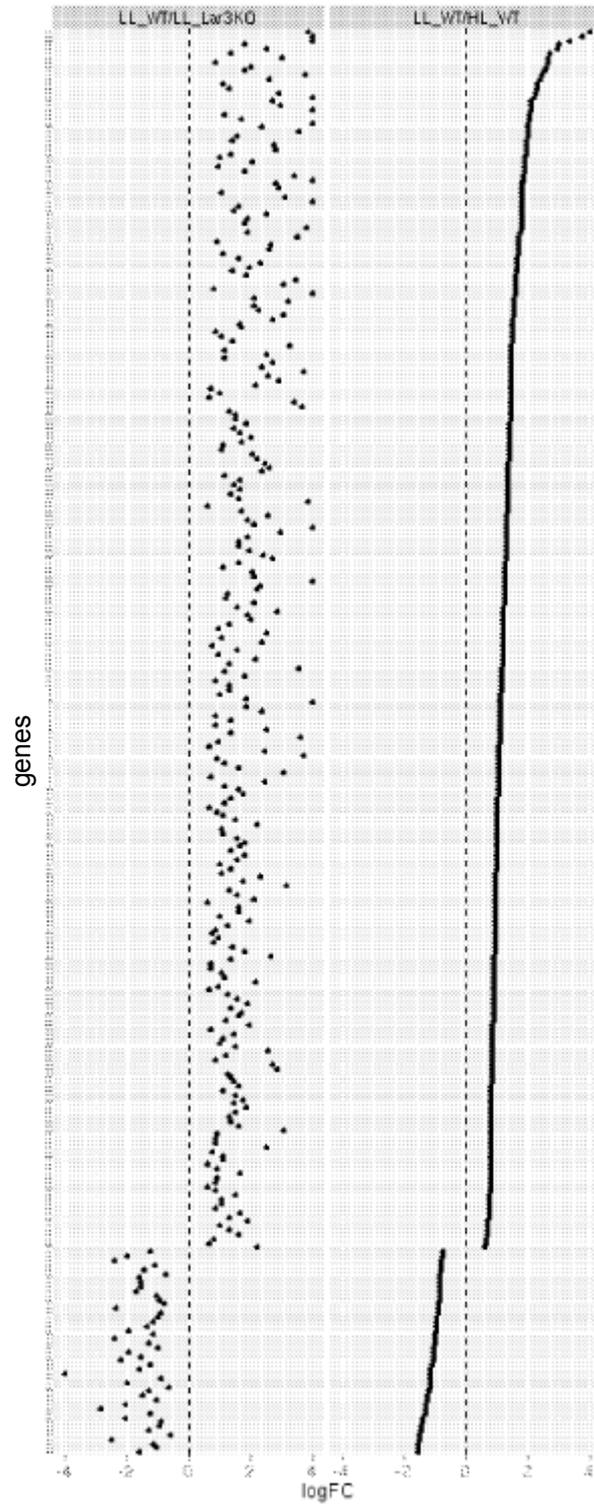


FIG. 23

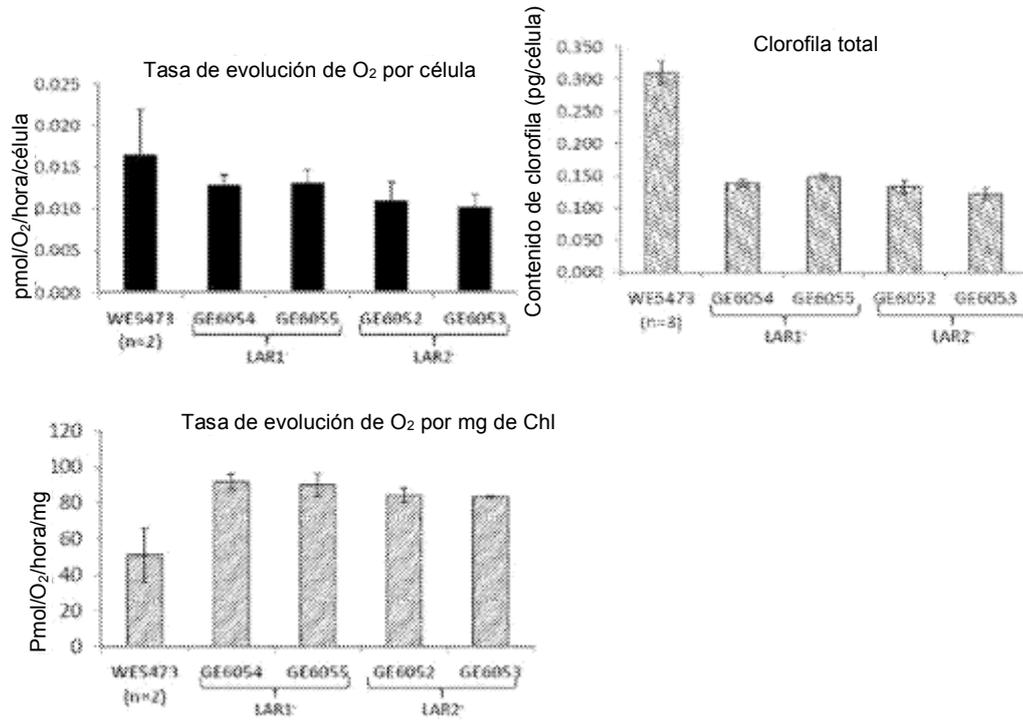


FIG. 24

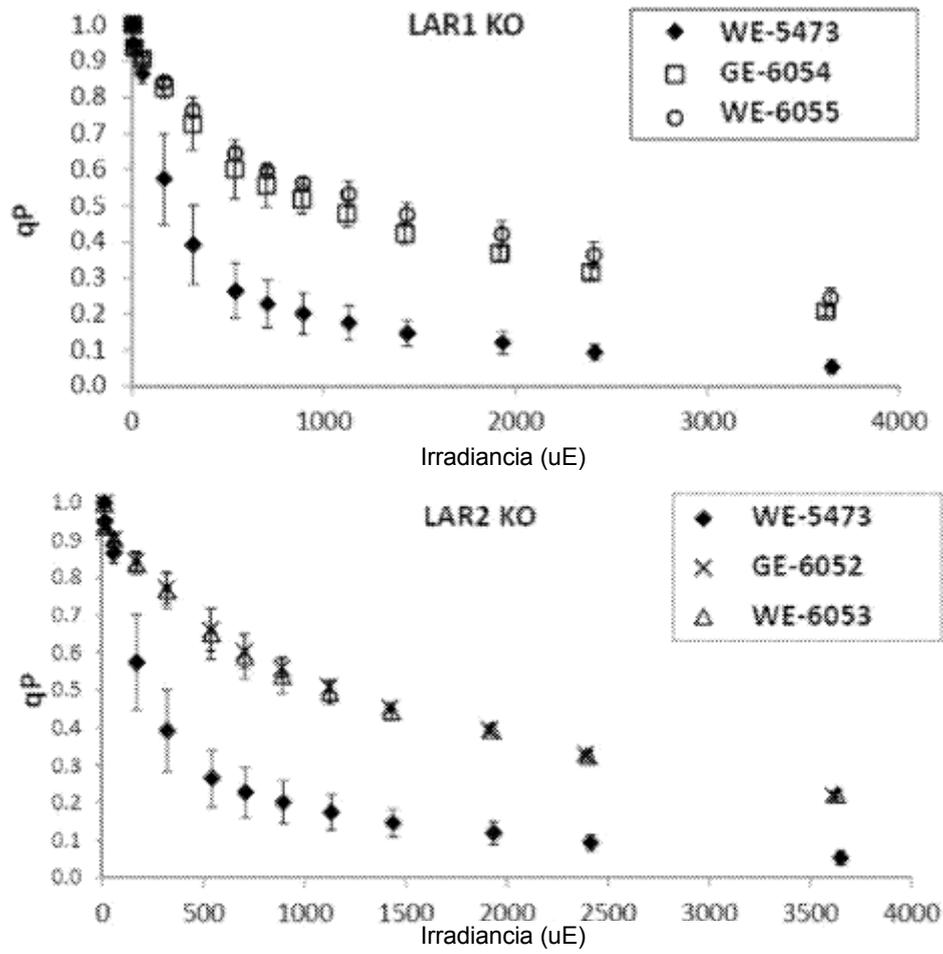


FIG. 25

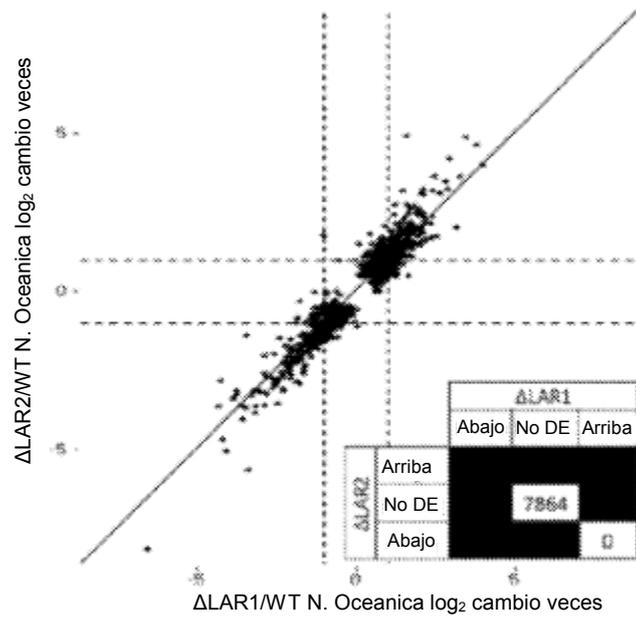


FIG. 26

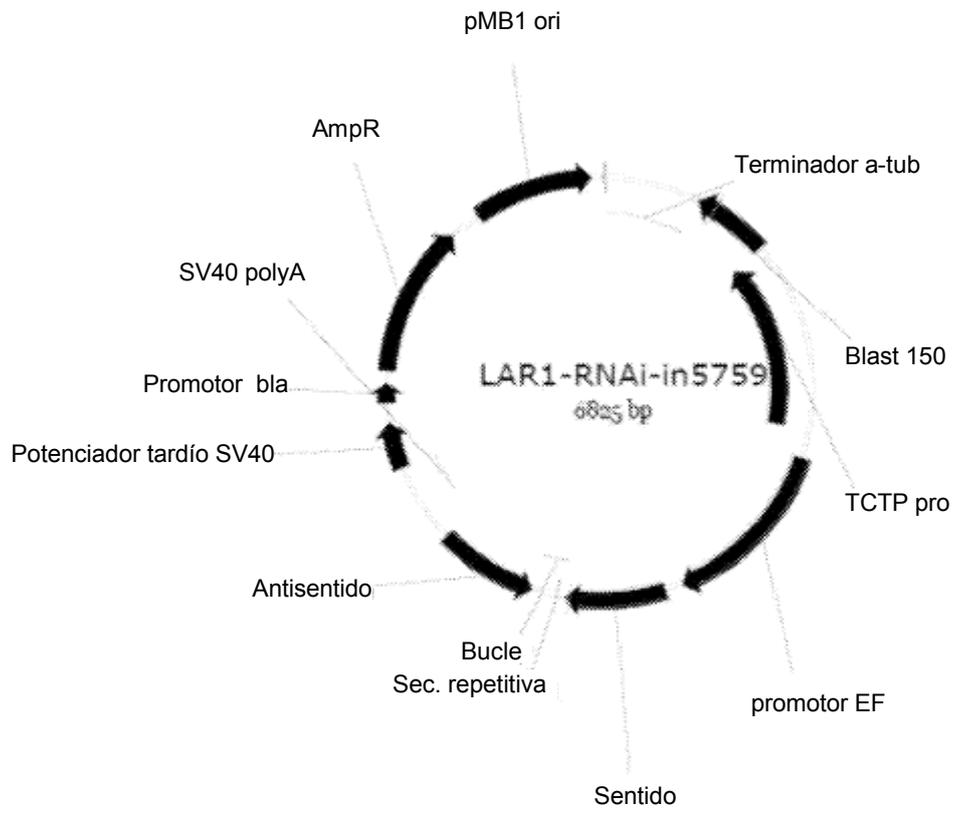


FIG. 27

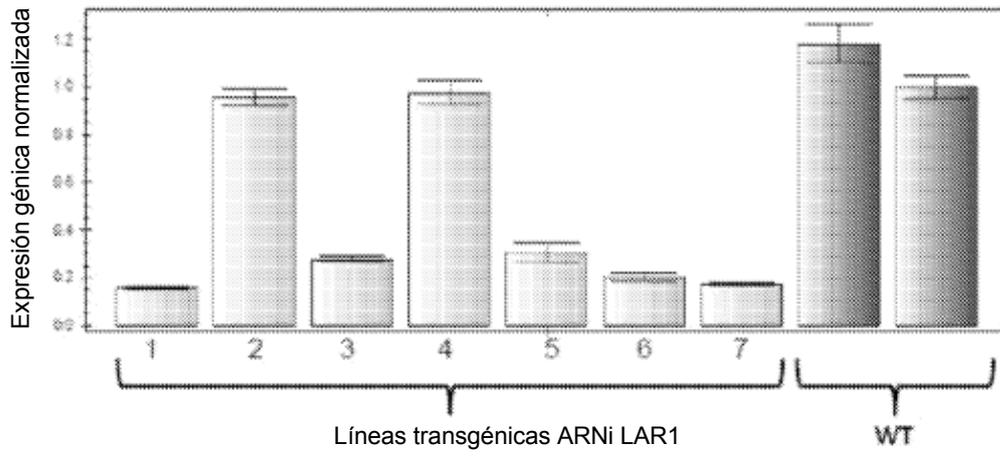


FIG. 28

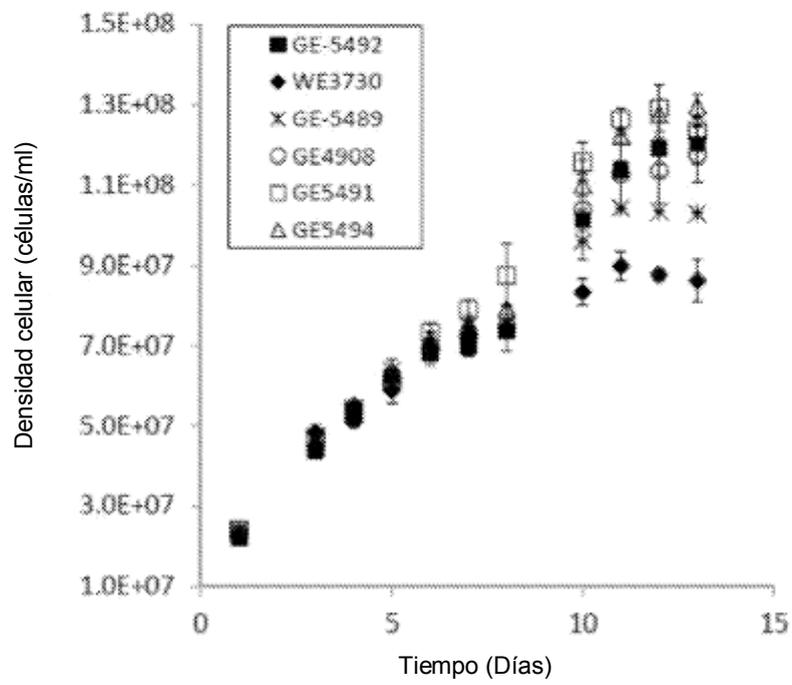


FIG. 29

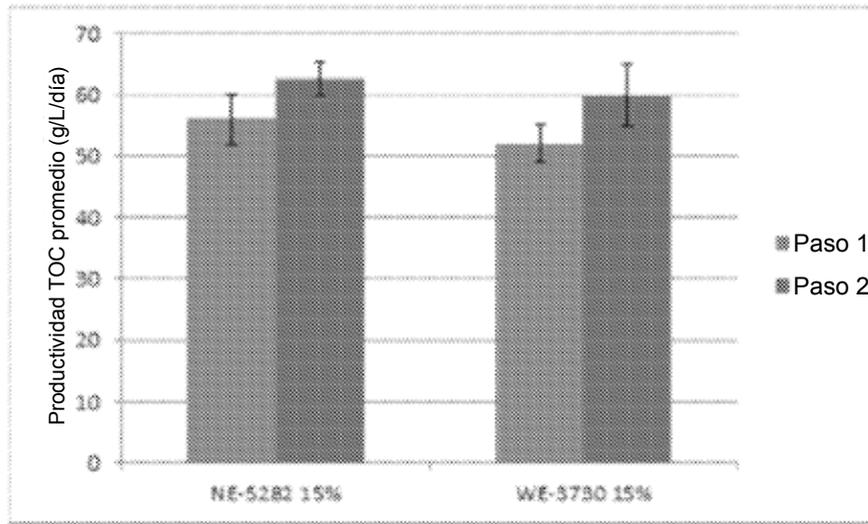


FIG. 30

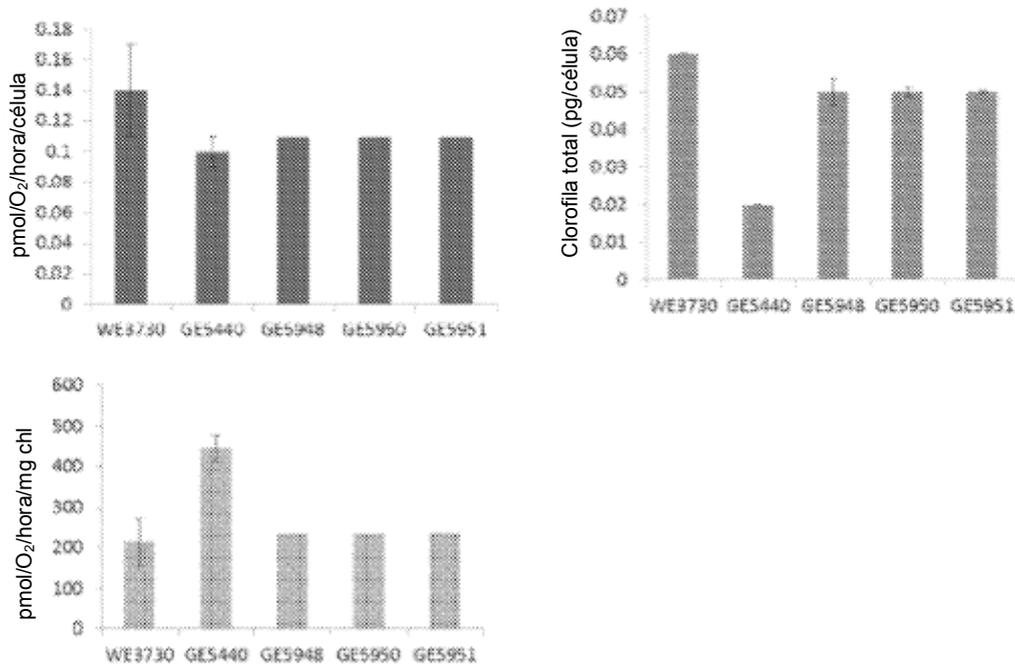


FIG. 31

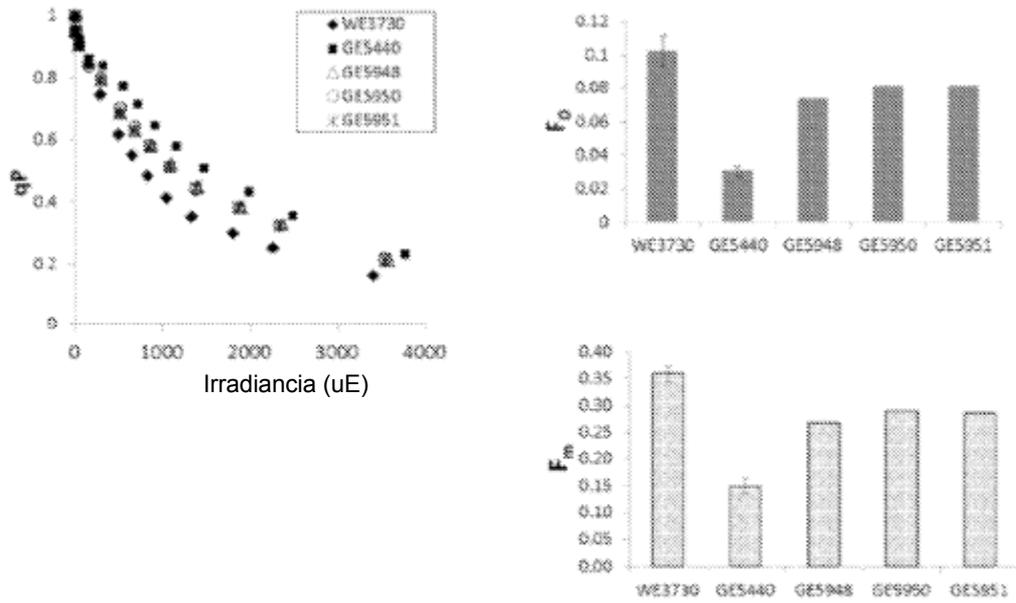


FIG. 32

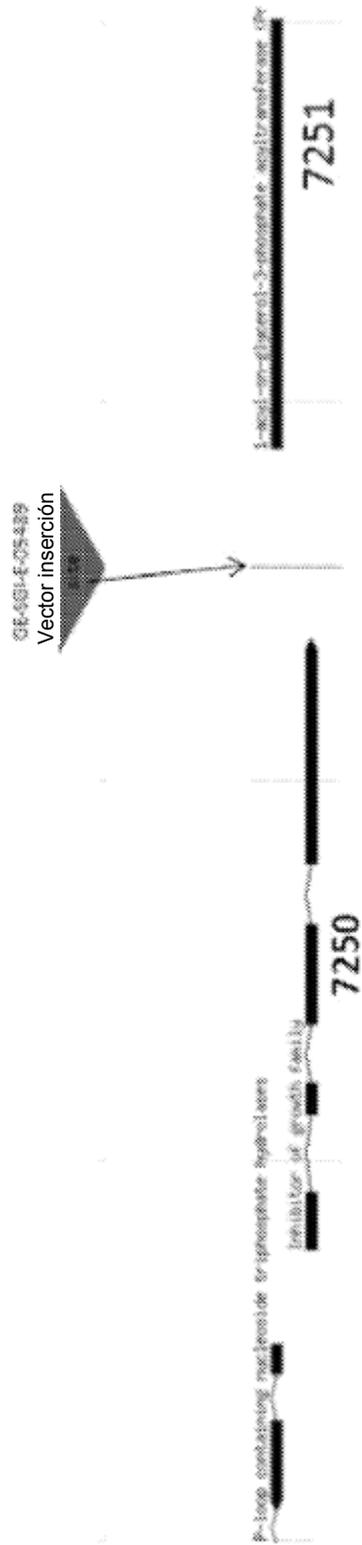


FIG. 33

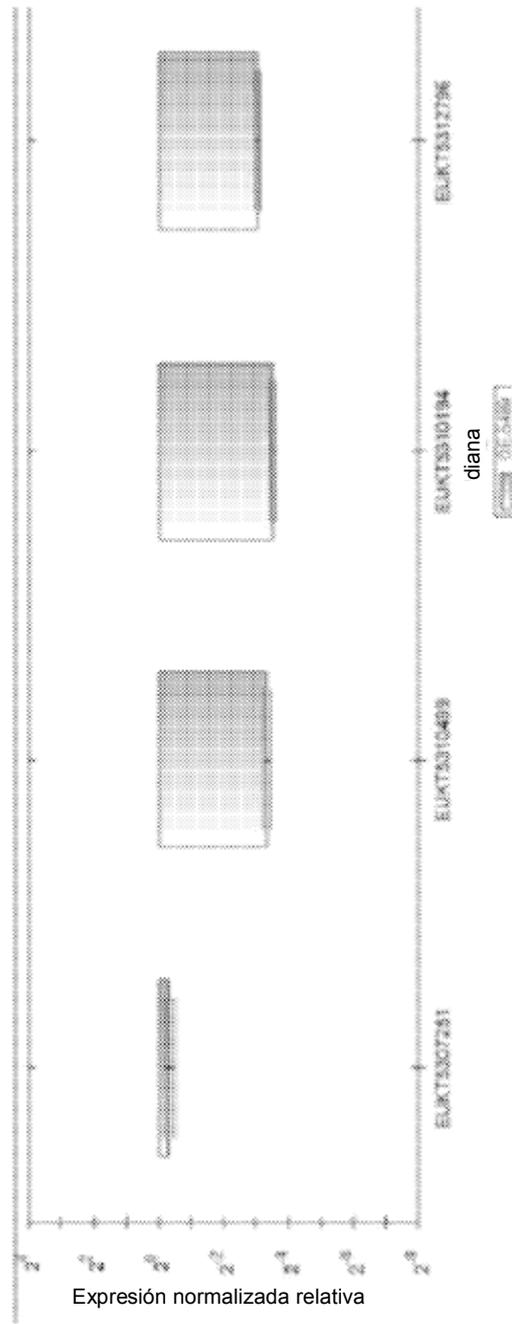


FIG. 34

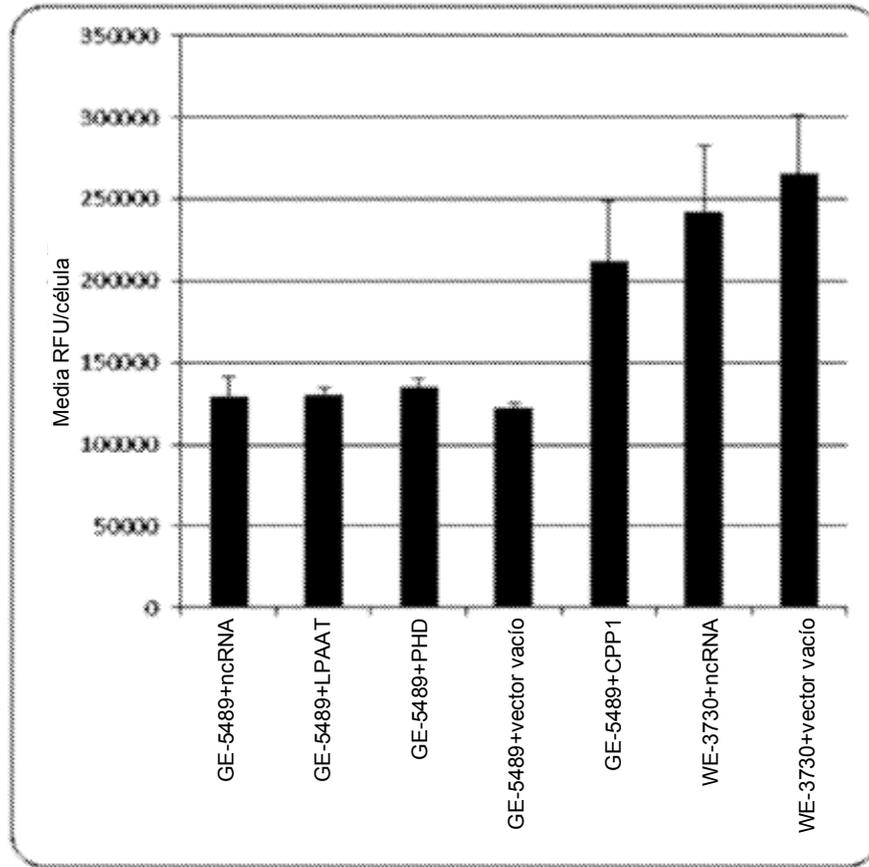


FIG. 35

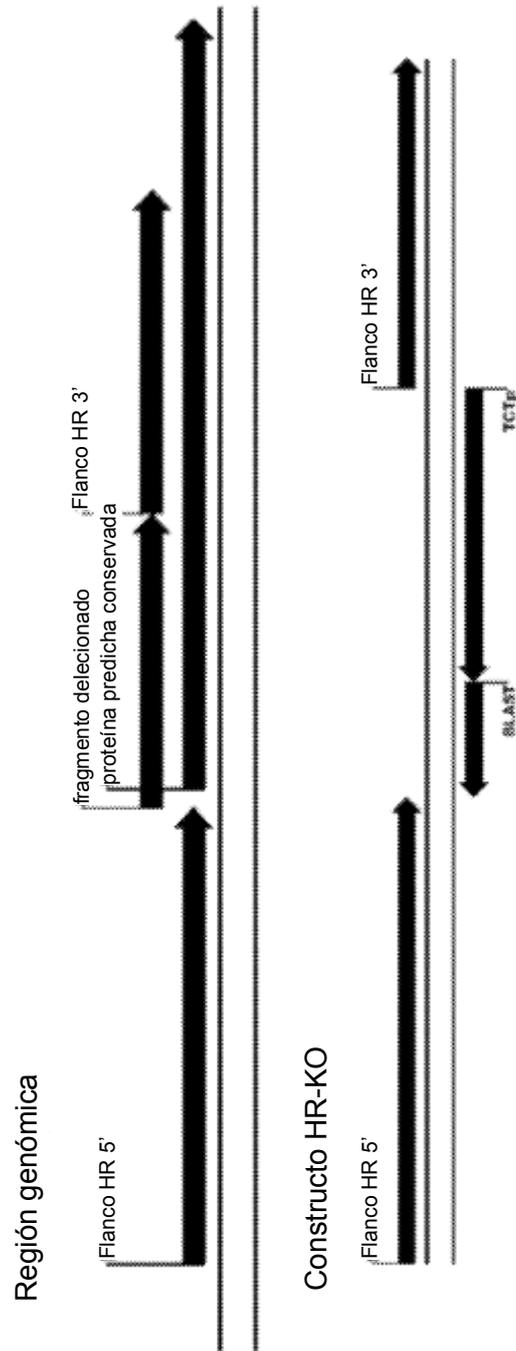


FIG. 36

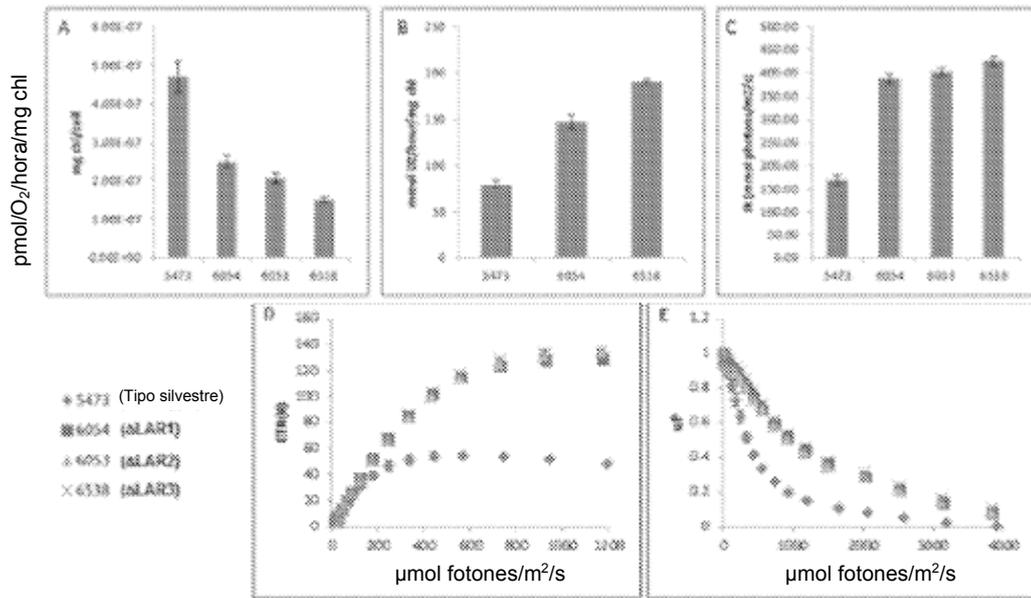


FIG. 37

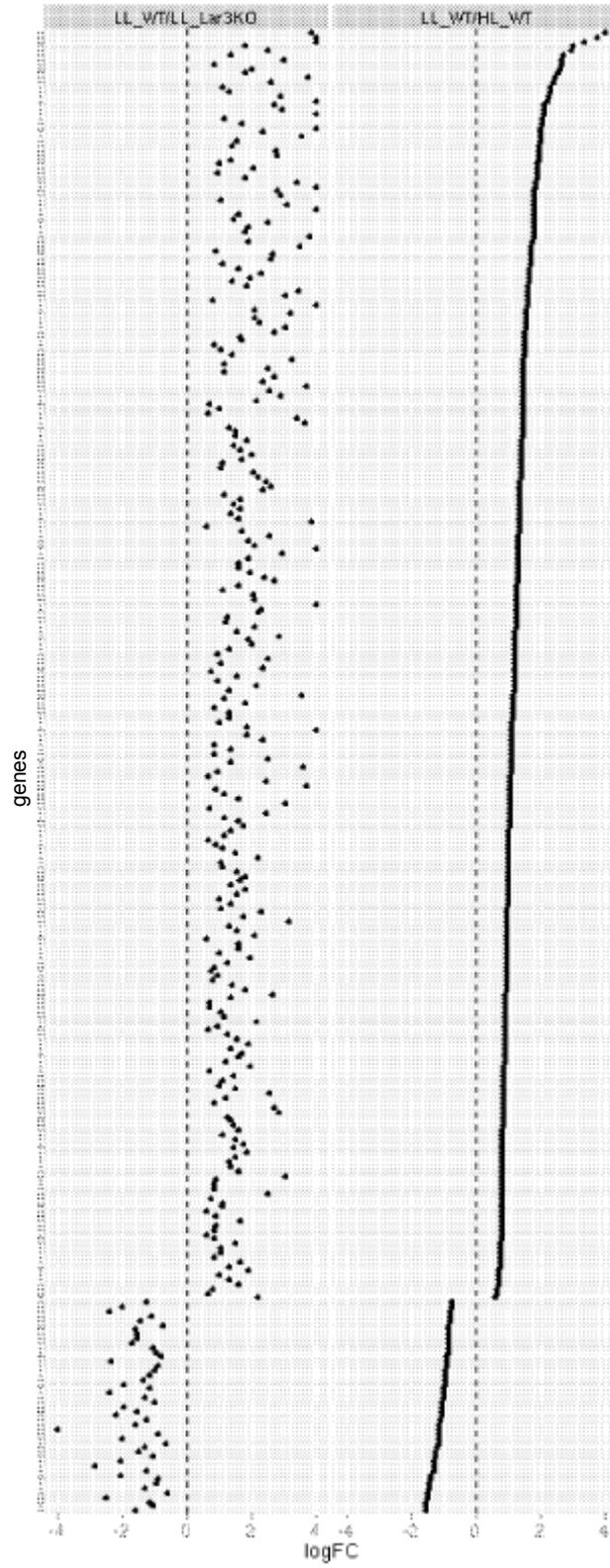


FIG. 39

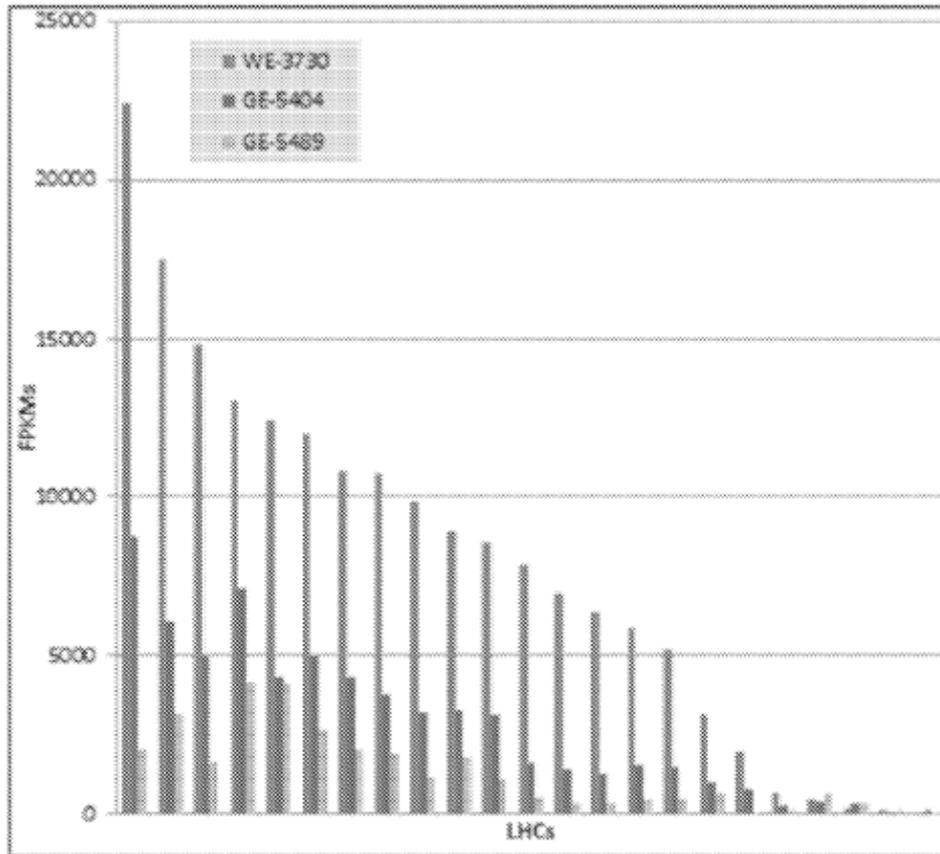


FIG. 40

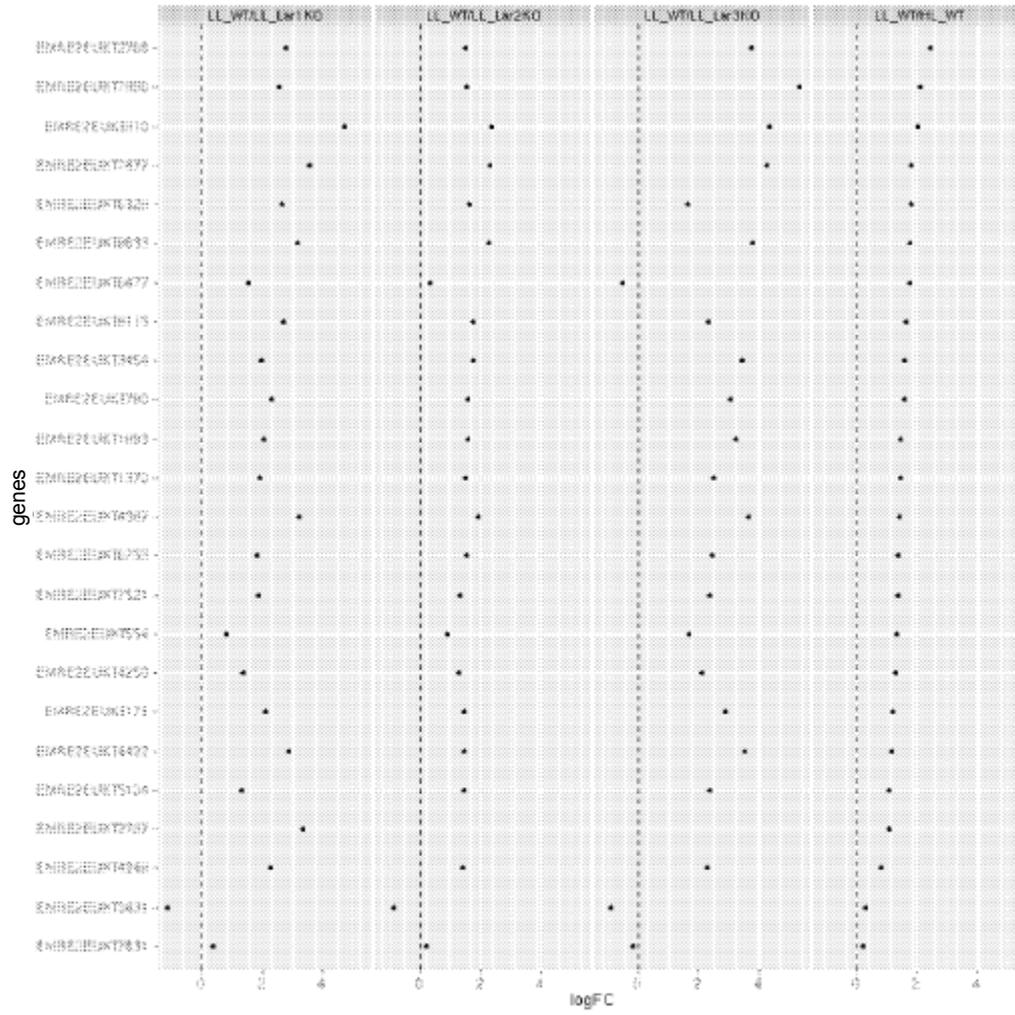


FIG. 41