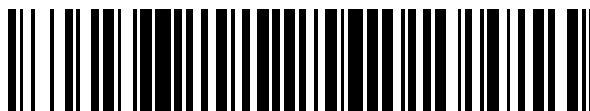


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 423 209**

51 Int. Cl.:

C12N 15/82 (2006.01)

C07K 14/405 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.11.2008** **E 08851200 (9)**

97 Fecha y número de publicación de la concesión europea: **08.05.2013** **EP 2225378**

54 Título: **Plantas que tienen un aumento de características relacionadas con el rendimiento y un método para elaboración de las mismas**

30 Prioridad:

22.11.2007 EP 07121362

26.11.2007 US 990132 P

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

18.09.2013

73 Titular/es:

CROPDESIGN N.V. (100.0%)

TECHNOLOGIEPARK 3

9052 GENT, BE

72 Inventor/es:

FRANKARD, VALERIE;

ALLEN, ANDY y

BOWLER, CHRIS

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 423 209 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

Plantas que tienen un aumento de características relacionadas con el rendimiento y un método para elaboración de las mismas.

La presente invención se refiere en general al campo de la biología molecular y se relaciona con un método para incrementar diferentes características de la planta relacionadas con el rendimiento, mediante el incremento de la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido transportador de amonio (AMT). La presente invención también se relaciona con plantas que tienen una mayor expresión de una secuencia de ácido nucleico que codifica un polipéptido AMT, en donde dichas plantas tienen más características relacionadas con el rendimiento con respecto a las plantas que sirven de control. La invención también provee construcciones útiles en los métodos de la invención.

La población mundial siempre en crecimiento y la oferta cada vez más escasa de tierras cultivables disponibles para la agricultura impulsan la investigación para lograr una mayor eficiencia de la agricultura. Los medios convencionales para la mejora hortícola y de los cultivos utilizan técnicas selectivas de fitomejoramiento para identificar las plantas que tienen rasgos deseables. Sin embargo, dichas técnicas selectivas de fitomejoramiento tiene varias desventajas, a saber, que estas técnicas son típicamente de trabajo intensivo y traen como resultado plantas que contienen a menudo componentes genéticos heterogéneos que no siempre resultan en la transmisión de la característica deseable a partir de las plantas progenitoras. Los avances en biología molecular le han permitido a la humanidad modificar el germoplasma de animales y de plantas. La modificación por ingeniería genética de las plantas supone el aislamiento y manipulación de material genético (típicamente en la forma de ADN o ARN) y la posterior introducción de ese material genético en una planta. Dicha tecnología tiene la capacidad de proporcionar cultivos o plantas que tienen diferentes características económicas, agronómicas u hortícolas mejoradas.

Una característica de particular interés económico es un mayor rendimiento. El rendimiento normalmente se define como la producción medible de valor económico de un cultivo. Esto puede ser definido en términos de cantidad y/o de calidad. El rendimiento depende directamente de varios factores, por ejemplo, de la cantidad y tamaño de los órganos, de la arquitectura de la planta (por ejemplo, la cantidad de ramas), de la producción de semillas, de la senectud de la hoja y más. El desarrollo de la raíz, la captación de nutrientes, la tolerancia al estrés y el vigor inicial también pueden ser factores importantes para determinar el rendimiento. Por lo tanto, la optimización de los factores mencionados puede contribuir al incremento del rendimiento de los cultivos.

El rendimiento de semillas es una característica particularmente importante, ya que las semillas de muchas plantas son importantes para la nutrición humana y animal. Cultivos tales como maíz, arroz, trigo, canola y soja dan cuenta de más de la mitad de la ingesta total de calorías por parte de la humanidad, ya sea a través del consumo directo de las semillas en si mismas o a través del consumo de productos alimenticios producidos con semillas procesadas. Son además una fuente de azúcares, aceites y muchas clases de metabolitos utilizados en procesos industriales. Las semillas contienen un embrión (la fuente de nuevos brotes y raíces) y un endospermo (la fuente de nutrientes para el desarrollo del embrión durante la germinación y durante el desarrollo temprano de las plántulas). El desarrollo de una semilla involucra muchos genes, y requiere de la transferencia de metabolitos a partir de las raíces, hojas y tallos dentro de la semilla en desarrollo. El endospermo, en particular, asimila los precursores metabólicos de carbohidratos, aceites y proteínas y los sintetiza en macromoléculas de almacenamiento para llenar el grano.

La biomasa de la planta es producida por cultivos forrajeros como la alfalfa, el maíz ensilado y el heno. Se han utilizado muchos procedimientos en los cultivos de granos. Los principales entre todos estos son los estimados del tamaño de la planta. El tamaño de la planta puede medirse en muchas formas dependiendo de la especie y de la etapa de desarrollo, pero incluye el peso seco total de la planta, el peso seco por encima del suelo, el peso fresco por encima del suelo, el área foliar, el volumen del tallo, la altura de la planta, el diámetro del rosetón, la longitud de las hojas, la longitud de la raíz, la masa de la raíz, el número de brotes y el número de hojas. Muchas especies mantienen una relación conservadora entre el tamaño de las diferentes partes de la planta en una etapa de desarrollo dada. Estas relaciones alométricas se utilizan para extrapolar a partir de una de estas medidas de tamaño con respecto a otra (por ejemplo, Tittone et al., 2005, Agric Ecosys & Environ 105: 213). El tamaño de la planta en una etapa de desarrollo temprano se correlacionará típicamente con el tamaño de la planta posterior en el desarrollo. Una planta más grande con un área foliar más grande puede absorber típicamente más luz y dióxido de carbono que una planta más pequeña y por lo tanto probablemente ganará un mayor peso durante el mismo periodo (Fasoula & Tollenaar, 2005, Maydica 50: 39). Esto se suma a la continuación potencial de la ventaja microambiental o genética que la planta tenía para obtener el tamaño más grande inicialmente. Existe un componente genético fuerte en el tamaño de la planta y la velocidad de crecimiento (por ejemplo, ter Steege et al., 2005, Plant Physiology 139: 1078), y de este modo para una gama de genotipos diversos, el tamaño de la planta bajo una condición ambiental probablemente se correlaciona con el tamaño bajo otra (Hittalmani et al., 2003, Theoretical Applied Genetics 107: 679). De esa forma se utiliza un ambiente estándar como determinante para los ambientes diversos y dinámicos encontrados en diferentes sitios y momentos por los cultivos en el campo.

Otra característica importante para muchos cultivos es el vigor inicial. La mejora del vigor inicial es un objetivo importante de los programas modernos de fitomejoramiento de arroz tanto en zonas de cultivo templadas como

tropicales de arroz. Raíces largas son importantes para el anclaje apropiado del suelo en arroz sembrado en agua. Cuando el arroz se siembra directamente en campos inundados, y donde las plantas deben emerger rápidamente a través del agua, los brotes más largos están asociados con el vigor. Donde se practica el cultivo por medio de sembradoras, mesocotilos y coleoptilos más largos son importantes para una buena aparición de las plántulas. La capacidad para modificar por ingeniería genética el vigor inicial en las plantas sería de gran importancia en la agricultura. Por ejemplo, un pobre vigor inicial ha sido una limitación para la introducción de híbridos de maíz (*Zea mays* L.) con base en el germoplasma del Cinturón de Maíz en el Atlántico Europeo.

El índice de cosecha, la relación del rendimiento de semillas con respecto al peso seco por encima del suelo, es relativamente estable bajo muchas condiciones ambientales y de este modo se puede obtener una muy buena correlación entre el tamaño de la planta y el rendimiento de granos (por ejemplo, Rebetzke et al., 2002, *Crop Science* 42: 739). Estos procesos están intrínsecamente enlazados porque la mayor parte de la biomasa del grano depende de la productividad fotosintética presente o almacenada por parte de las hojas y el tallo de la planta (Gardener et al., 1985, *Physiology of Crop Plants*. Iowa State University Press, páginas 68 - 73). Por lo tanto, la selección del tamaño de la planta, incluso en etapas tempranas de desarrollo, ha sido utilizada como un indicador del rendimiento potencial futuro (por ejemplo, Tiftonell et al., 2005, *Agric Ecosys & Environ* 105: 213). Cuando se analiza el impacto de las diferencias genéticas sobre tolerancia al estrés, la capacidad de estandarizar las propiedades del suelo, la temperatura, la disponibilidad de agua y de nutrientes y la intensidad de la luz es una ventaja intrínseca del invernadero o de los ambientes con cámaras de crecimiento de plantas en comparación con el campo. Sin embargo, las limitaciones artificiales en el rendimiento debido a una pobre polinización provocada por la ausencia de viento o de insectos, o insuficiente espacio para la maduración de la raíz o el desarrollo del dosel, pueden restringir el uso de estos ambientes controlados para analizar las diferencias de rendimiento. Por lo tanto, las mediciones del tamaño de la planta en el desarrollo temprano, bajo condiciones estándar en una cámara de crecimiento o invernadero, son prácticas estándar para proporcionar una indicación de las ventajas potenciales de rendimiento genético.

Otra característica importante es aquella de la tolerancia mejorada al estrés abiótico. El estrés abiótico es una de las causas principales de pérdida de cultivos en el mundo, reduciendo los rendimientos promedio para la mayoría de las plantas de cultivo en más del 50% (Wang et al. (2003) *Planta* 218: 1 - 14). Los estreses abióticos pueden ser provocados por sequía, salinidad, extremos de temperaturas, toxicidad química, exceso o deficiencia de nutrientes (macroelementos y/o microelementos), radiación y estrés oxidativo. La capacidad de incrementar la tolerancia de una planta al estrés abiótico sería una ventaja económica importante para los granjeros a nivel mundial y permitiría el cultivo de granos en condiciones adversas y en territorios donde el cultivo no puede ser posible de otro modo.

Por lo tanto, se puede incrementar el rendimiento del cultivo optimizando uno de los factores anteriormente mencionados.

Dependiendo del uso final, se puede favorecer a modificación de ciertas características de rendimiento con respecto a otras. Por ejemplo para aplicaciones tales como la producción de forraje o madera, o recursos de biocombustibles, puede ser deseable un incremento en las partes vegetativas de una planta, y para aplicaciones tales como la producción de harina, almidón, o de aceite, puede ser particularmente deseable un incremento en los parámetros de la semilla. Incluso entre los parámetros de la semilla, algunos pueden verse favorecidos sobre otros, dependiendo de la aplicación. Diferentes mecanismos pueden contribuir para incrementar el rendimiento de semilla, ya sea en la forma de mayor tamaño de la semilla o el incremento en el número de semillas.

Un enfoque para incrementar las características relacionadas con el rendimiento (rendimiento de semilla y/o de biomasa) en plantas puede ser a través de la modificación de los mecanismos inherentes al crecimiento de una planta, tal como el ciclo celular o diferentes rutas de señalización involucradas en el desarrollo de la planta o en mecanismos de defensa.

Ahora se ha encontrado que se pueden incrementar las diferentes características relacionadas con el rendimiento en las plantas con respecto a plantas que sirven de control, por medio del aumento de la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido transportador de amonio (AMT). El aumento de características relacionadas con el rendimiento comprende uno o más de: aumento del vigor inicial, aumento de la biomasa aérea, aumento de la biomasa de las raíces, aumento del rendimiento total de semillas por planta, aumento de la tasa de llenado de las semillas, aumento del número de semillas llenas, y mayor índice de cosecha.

Antecedentes

El amonio y el nitrato son fuentes primarias de nitrógeno para el crecimiento y desarrollo de las plantas. Las plantas requieren de transportadores para la adquisición tanto de amonio como de nitrato. Existen transportadores de amonio y de nitrato no sólo en las plantas, sino en casi todos los organismos. Los transportadores de amonio (AMT) existen por lo general en un genoma como familias de genes, por ejemplo, al menos: seis en *Arabidopsis thaliana*, ocho en *Chlamydomonas reinhardtii* (Gonzales-Ballester et al. (2004) *Plant Molec Biol*. 56: 863 - 878), catorce en el álamo (Couturier et al. (2007). *New Phytologist* 174: 137 - 150), seis de la diatomea *Phaeodactylum tricornutum* (Allen (2005) *J Phycology* 41).

Con base en el análisis filogenético, se identificaron tres subfamilias de transportadores de amonio (Loqué & von Wiren (2004) J. Exp. Bot 55 (401): 1293 - 1305):

- 5 1. la subfamilia de AMT, incluyendo los transportadores de tipo AMT1 de la planta, y transportadores de amonio cianobacterial;
2. la subfamilia de MEP, incluyendo los transportadores de tipo AMT2 de la planta, los transportadores de MEP de levadura, la AmtB de *E. coli*, y otros homólogos procariotas;
- 10 3. La subfamilia Rh, incluyendo sólo los antígenos de los grupos sanguíneos Rh humanos y de animales.

10 Todos los polipéptidos AMT son proteínas de membrana altamente hidrófobas con al menos 10, más comúnmente 11 hélices putativas que abarcan toda la membrana. Se ha mostrado en numerosos reportes que los polipéptidos AMT son capaces de incorporar de amonio en un amplio rango de concentración, aunque con diferentes afinidades de un organismo a otro. Dentro de ciertos organismos, tales como plantas, se identificaron transportadores de amonio de baja y alta afinidad (Gazzarini et al. (1999) Plant Cell 11: 937 - 47). Además de las propiedades de afinidad, se han identificado varios otros mecanismos de regulación para la captación de amonio, por ejemplo, a niveles transcripcional y post-transcripcional (Yuan et al (2007) Plant Phys. 143: 732 - 744).

20 La sobreexpresión de una secuencia de ácido nucleico de arroz que codifica una AMT1 se llevó a cabo en dos variedades cultivadas de arroz (Taipéi 309 y Jarrah), utilizando un promotor de ubiquitina del maíz para expresión constitutiva. La biomasa de brotes y raíces de líneas transgénicas disminuyeron durante la etapa vegetativa temprana y de plántulas en comparación con el tipo silvestre, sobre todo cuando se cultiva en condiciones de alta nutrición de amonio (Hoque et al. (2006) Functional Plant Biol 33: 153 - 163). Los autores concluyeron que la disminución de la biomasa de las plantas transgénicas en las primeras etapas de crecimiento podría haber sido causada por la acumulación de amonio en la raíces debido a la incapacidad para que coincida la asimilación de amonio con la mayor captación de amonio.

25 En la patente de los Estados Unidos No. 6.620.610, se describe una secuencia de ácido nucleico que codifica un polipéptido AMT1 de *Arabidopsis thaliana*, plásmidos que comprenden la secuencia de ácido nucleico que codifica un AMT1 para la expresión en levaduras y bacterias.

30 En la patente de los Estados Unidos No. 6.833.492 se describen secuencias de ácidos nucleicos que codifican un polipéptido AMT1 de soja, maíz, trigo y arroz. Se describe una secuencia de ácido nucleico que codifica un polipéptido AMT1 o un polipéptido AMT que tiene 90% de identidad de secuencia de aminoácidos con el polipéptido aislado AMT1 de soja. Se describen plantas y semillas que comprenden una secuencia recombinante de ácido nucleico que codifica dicha secuencia de polipéptido, así como métodos para producir tales plantas. En el documento WO 2006/076423 se describe una secuencia de ácido nucleico que codifica un polipéptido AMT1 de la diatomea *Cylindrotheca fusiformis*.

40 Sorprendentemente, se ha encontrado ahora que el aumento en la expresión de una secuencia de ácido nucleico que codifica un polipéptido ATM produce plantas que tienen más características relacionadas con el rendimiento en relación con las plantas de control.

45 De acuerdo con una forma de realización, se proporciona un método para aumentar las características relacionadas con el rendimiento en plantas con relación a las plantas de control, que comprende aumentar la expresión de una secuencia de ácido nucleico que codifica un polipéptido AMT como se define aquí, en una planta. El aumento de características relacionadas con el rendimiento comprende uno o más de: aumento del vigor inicial, aumento de biomasa aérea, aumento de biomasa de la raíz, aumento del rendimiento total de semillas por planta, aumento de la tasa de llenado de las semillas, aumento del número de semillas llenas, aumento del número de flores por panícula, y aumento del índice de cosecha.

Definiciones

55 Polipéptido(s) / Proteína(s)

Los términos "polipéptido" y "proteína" se utilizan indistintamente aquí y se refieren a aminoácidos en una forma polimérica de cualquier longitud, unidos por medio de enlaces peptídicos.

60 Polinucleótido(s) / Ácido(s) nucleico(s) / Secuencia(s) de ácido nucleico / Secuencia(s) de nucleótidos

Los términos "polinucleótido(s)", "secuencia(s) de ácido nucleico", "secuencia(s) de nucleótidos", "ácido(s) nucleico(s)" se utilizan indistintamente aquí y se refieren a nucleótidos, ya sea ribonucleótidos o desoxirribonucleótidos o una combinación de ambos, en una forma polimérica no ramificada de cualquier longitud.

Planta(s) de control

La elección de plantas control adecuadas es una parte rutinaria de la programación de un experimento y puede incluir las plantas correspondientes de tipo silvestre o las plantas correspondientes sin el gen de interés. La planta control es típicamente de la misma especie de planta o incluso de la misma variedad que la planta que va a ser evaluada. La planta control puede ser también un nulicigoto de la planta que va a ser evaluada. Una "planta de control" tal como se utiliza aquí se refiere no sólo a plantas enteras, sino también a partes de plantas, incluyendo semillas y partes de la semilla.

Homólogo(s)

Los "homólogos" de una proteína abarcan péptidos, oligopéptidos, polipéptidos, proteínas y enzimas que tienen sustituciones, supresiones y/o inserciones de aminoácidos con relación a la proteína no modificada en cuestión y que tienen actividad biológica y funcional similar a la de la proteína no modificada de la cual se derivan.

Una supresión se refiere a la remoción de uno o más aminoácidos de una proteína.

Una inserción se refiere a uno o más residuos de aminoácidos que se introducen en un sitio predeterminado en una proteína. Las inserciones pueden incluir fusiones en el terminal N y/o en el terminal C así como inserciones dentro de la secuencia de uno solo o de múltiples aminoácidos. En general, las inserciones dentro de la secuencia de aminoácidos serán más pequeñas que las fusiones en el terminal N o en el terminal C, del orden de aproximadamente 1 a 10 residuos. Los ejemplos de proteínas o péptidos de fusión del terminal N o C incluyen el dominio de enlazamiento o el dominio de activación de un activador transcripcional tal como el que se utiliza en el sistema de doble híbrido de levadura, las proteínas de recubrimiento de fago, la etiqueta de (histidina)₆, la etiqueta de glutatión S transferasa, la proteína A, la proteína de enlazamiento de maltosa, la dihidrofolato reductasa, el epítipo Tag-100, el epítipo c-myc, el epítipo FLAG®, lacZ, CMP (péptido de enlazamiento de calmodulina), el epítipo HA, el epítipo de proteína C y el epítipo VSV.

Una sustitución se refiere al reemplazo de aminoácidos de la proteína por otros aminoácidos que tienen propiedades similares (tales como hidrofobicidad, hidrofiliidad, antigenicidad, propensión a formar o romper estructuras helicoidales α o estructuras de láminas β similares). Las sustituciones de aminoácidos son típicamente de residuos individuales, pero pueden agruparse dependiendo de las restricciones funcionales colocadas sobre el polipéptido; las inserciones usualmente serán del orden de aproximadamente 1 a 10 residuos de aminoácidos. Las sustituciones de aminoácidos son preferiblemente sustituciones conservadoras de aminoácidos. Las tablas de sustitución conservadora son bien conocidas en la técnica (véase por ejemplo Creighton (1984) Proteins. W. H. Freeman and Company (Eds.) y la Tabla 1 siguiente).

Tabla 1: Ejemplos de sustituciones conservadoras de aminoácidos

Residuo	Sustituciones conservadoras	Residuo	Sustituciones conservadoras
Ala	Ser	Leu	Ile; Val
Arg	Lys	Lys	Arg; Gln
Asn	Gln; His	Met	Leu; Ile
Asp	Glu	Phe	Met; Leu; Tyr
Gln	Asn	Ser	Thr; Gly
Cys	Ser	Thr	Ser, Val
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp; Phe
His	Asn; Gln	Val	Ile; Leu
Ile	Leu, Val		

Las sustituciones, supresiones y/o inserciones de aminoácidos se pueden realizar fácilmente utilizando técnicas de síntesis de péptidos bien conocidas en la técnica, tales como la síntesis de péptidos en fase sólida y similares, o por medio de manipulación de ADN recombinante. Los métodos para la manipulación de las secuencias de ADN para producir las variantes de sustitución, inserción o supresión de una proteína son bien conocidos en la técnica. Por ejemplo, las técnicas para realizar mutaciones de sustitución en sitios predeterminados en el ADN son bien conocidas por aquellos ordinariamente capacitados en la técnica e incluyen la mutagénesis M13, mutagénesis *in vitro* del Gen T7 (USB, Cleveland, OH), mutagénesis dirigida al sitio QuickChange (Stratagene, San Diego, CA), mutagénesis dirigida al sitio mediada por PCR u otros protocolos de mutagénesis dirigida al sitio.

Derivados

Los "derivados" incluyen péptidos, oligopéptidos, polipéptidos los cuales pueden, en comparación con la secuencia de aminoácidos de la forma de la proteína de origen natural, tal como la proteína de interés, incluir sustituciones de aminoácidos con residuos de aminoácidos que no son de origen natural, o adiciones de residuos de aminoácidos

que no son de origen natural. Los “derivados” de una proteína también abarcan péptidos, oligopéptidos, polipéptidos los cuales incluyen residuos de aminoácidos de origen natural alterados (glicosilados, acilados, prenilados, fosforilados, miristoilados, sulfatados, etc.) o alterados de origen no natural comparado con la secuencia de aminoácidos de una forma de origen natural del polipéptido. Un derivado puede incluir también uno o más sustituyentes o adiciones que no son de aminoácidos comparado con la secuencia de aminoácidos de la cual se deriva, por ejemplo una molécula reportera u otro ligando, enlazado en forma covalente o no covalente a la secuencia de aminoácidos, tal como una molécula reportera que se enlaza para facilitar su detección, y residuos de aminoácidos de origen no natural con relación a la secuencia de aminoácidos de una proteína de origen natural.

10 Ortólogo(s) / Parálogo(s)

Los ortólogos y los parálogos abarcan conceptos evolucionarios utilizados para describir las relaciones ancestrales de los genes. Los parálogos son genes dentro de la misma especie que se han originado a través de la duplicación de un gen ancestral; los ortólogos son genes de diferentes organismos que se han originado a través de especiación, y se derivan también de un gen ancestral común.

Dominio

El término "dominio" se refiere a un conjunto de aminoácidos conservados en posiciones específicas a lo largo de una alineación de secuencias de proteínas relacionadas por evolución. Mientras que los aminoácidos en otras posiciones pueden variar entre homólogos, los aminoácidos que son altamente conservados en posiciones específicas indican aminoácidos que son probablemente esenciales en la estructura, estabilidad o función de una proteína. Identificados por su alto grado de conservación en secuencias alineadas de una familia de homólogos de proteína, se pueden utilizar como identificadores para determinar si cualquier polipéptido en cuestión pertenece a una familia de polipéptidos previamente identificada.

Motivo / Secuencia de consenso / Firma

El término "motivo" o "secuencia de consenso" o "firma" se refieren a una región conservada corta en la secuencia de proteínas relacionadas por evolución. Los motivos son frecuentemente partes de dominios altamente conservados, pero pueden incluir también solamente parte del dominio, o estar localizados fuera del dominio conservado (si todos los aminoácidos del motivo caen fuera de un dominio definido).

Hibridación

El término “hibridación” como se define aquí es un proceso en donde las secuencias nucleotídicas complementarias sustancialmente homólogas se aparean entre sí. El proceso de hibridación puede ocurrir completamente en solución, es decir ambas moléculas de ácido nucleico complementarias están en solución. El proceso de hibridación puede ocurrir también con una de las moléculas complementarias de ácido nucleico inmovilizadas a una matriz tal como perlas magnéticas, perlas de Sefarosa o cualquier otra resina. El proceso de hibridación puede ocurrir además con una de las moléculas complementarias de ácido nucleico inmovilizadas a un soporte sólido tal como una membrana de nitrocelulosa o de nailon o inmovilizada, por ejemplo, por medio de fotolitografía, por ejemplo a un soporte de vidrio síliceo (éste último conocido como arreglos de secuencias de ácido nucleico o microarreglos o como chips de secuencias de ácido nucleico). Con el propósito de permitir que se produzca la hibridación, las moléculas de ácido nucleico son generalmente desnaturalizadas térmica o químicamente para fundir una hebra bicatenaria en dos hebras individuales y/o para remover horquillas u otras estructuras secundarias de las moléculas de ácido nucleico monocatenario.

El término “rigurosidad” se refiere a las condiciones bajo las cuales tiene lugar una hibridación. La rigurosidad de la hibridación está influenciada por condiciones tales como la temperatura, la concentración salina, la fuerza iónica y la composición del amortiguador de hibridación. En general, las condiciones de baja rigurosidad se seleccionan para que sean aproximadamente 30°C menores al punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y pH definidos. Las condiciones de rigurosidad media se dan en los casos donde la temperatura está 20°C por debajo de la T_m , y las condiciones de alta rigurosidad se dan en los casos cuando la temperatura está 10°C por debajo de la T_m . Las condiciones de hibridación de alta rigurosidad se utilizan típicamente para aislar las secuencias de hibridación que tienen una gran similitud de secuencia con la secuencia de ácido nucleico objetivo. Sin embargo, las secuencias de ácido nucleico pueden desviarse en secuencia e incluso codificar un polipéptido sustancialmente idéntico, debido a la degeneración del código genético. Por lo tanto, algunas veces pueden ser necesarias condiciones de hibridación de rigurosidad media, para identificar tales moléculas de secuencia de ácido nucleico.

La T_m es la temperatura en condiciones de una fuerza iónica y pH definidos, a la cual 50% de la secuencia objetivo hibrida a una sonda perfectamente emparejada. La T_m depende de las condiciones de la solución y de la composición base y la longitud de la sonda. Por ejemplo, las secuencias más largas hibridan específicamente a temperaturas más elevadas. La velocidad máxima de hibridación se obtiene aproximadamente desde 16°C hasta 32°C por debajo de la T_m . La presencia de cationes monovalentes en la solución de hibridación reduce la repulsión electrostática entre las dos hebras de la secuencia de ácido nucleico promoviendo de este modo la formación del

híbrido; este efecto es visible para concentraciones de sodio de hasta 0,4 M (para concentraciones más elevadas, este efecto puede ignorarse). La formamida reduce la temperatura de fusión de los dúplex ADN-ADN y ADN-ARN con 0,6 hasta 0,7°C para cada por ciento de formamida, y la adición de 50% de formamida permite que la hibridación se lleve a cabo entre 30 a 45°C, si bien la velocidad de la hibridación se disminuirá. La falta de emparejamiento de los pares de bases reduce la velocidad de hibridación y la estabilidad térmica de los dúplex. En promedio y para sondas largas, la T_m disminuye aproximadamente 1°C por % de falta de emparejamiento de las bases. La T_m puede calcularse utilizando las siguientes ecuaciones, dependiendo de los tipos de híbridos:

1) híbridos de ADN - ADN (Meinkoth y Wahl, Anal. Biochem, 138: 267 - 284, 1984):

$$T_m = 81,5^\circ\text{C} + 16,6 \times \log_{10} [\text{Na}^+]^a + 0,41 \times \% [\text{G/C}]^b - 500 \times [\text{L}]^{-1} - 0,61 \times \% \text{ de formamida}$$

2) híbridos de ADN - ARN o ARN - ARN:

$$T_m = 79,8 + 18,5 (\log_{10} [\text{Na}^+]^a) + 0,58 (\% \text{ de G/C}^b) + 11,8 (\% \text{ de G/C}^b)^2 - 820/\text{L}^c$$

3) híbridos de oligo - ADN u oligo - ARN^d:

Para <20 nucleótidos: $T_m = 2 (I_n)$

Para 20 - 35 nucleótidos: $T_m = 22 + 1,46 (I_n)$

^a o para otro catión monovalente, pero sólo preciso en el rango de 0,01 - 0,4 M.

^b sólo preciso para % de GC en el rango de 30% a 75%.

^c L = longitud del dúplex en pares de bases.

^d oligo, oligonucleótido; I_n = longitud efectiva del cebador = $2 \times (\text{no. de G/C}) + (\text{no. de A/T})$.

El enlazamiento no específico se puede controlar utilizando cualquiera entre una cantidad de técnicas conocidas tales como, por ejemplo, bloqueo de la membrana con soluciones que contienen proteína, adiciones de ARN heterólogo, ADN, y SDS al amortiguador de hibridación, y tratamiento con ARNasa. Para las sondas no homólogas, se puede llevar a cabo una serie de hibridaciones variando uno de (i) disminución progresiva de la temperatura de apareamiento (por ejemplo de 68°C hasta 42°C) o (ii) disminución progresiva de la concentración de formamida (por ejemplo desde 50% hasta 0%). La persona ordinariamente capacitada conoce diferentes parámetros que pueden ser alterados durante la hibridación y los cuales o bien mantendrán o cambiarán las condiciones de rigurosidad.

Además de las condiciones de hibridación, la especificidad de la hibridación típicamente también depende de la función de los lavados posteriores a la hibridación. Para eliminar el medio resultante de la hibridación no específica, se lavan las muestras con soluciones salinas diluidas. Los factores críticos de tales lavados incluyen la fuerza iónica y la temperatura de la solución final de lavado: cuánto más baja la concentración de sal y más alta la temperatura de lavado, mayor la rigurosidad del lavado. Las condiciones de lavado se llevan a cabo típicamente en o por debajo de la rigurosidad de la hibridación. Una hibridación positiva produce una señal que es al menos dos veces aquella de la señal de fondo. En general, las condiciones rigurosas adecuadas para los ensayos de hibridación de la secuencia de ácido nucleico o los procedimientos de detección por amplificación génica son como se expusieron anteriormente. Las condiciones más o menos rigurosas también se pueden seleccionar. Una persona normalmente capacitada en la técnica conoce diferentes parámetros que pueden ser alterados durante el lavado y que o bien mantendrán o cambiarán las condiciones de rigurosidad.

Por ejemplo, las condiciones típicas de hibridación de alta rigurosidad para los híbridos de ADN de longitud superior a 50 nucleótidos abarcan la hibridación a 65°C en 1x SSC o a 42°C en 1x SSC y 50% de formamida, seguido por un lavado a 65°C en 0,3x SSC. Los ejemplos de condiciones de hibridación de rigurosidad media para los híbridos de ADN de longitud superior a 50 nucleótidos abarcan la hibridación a 50°C en 4x SSC o a 40°C en 6x SSC y 50% de formamida, seguido por un lavado a 50°C en 2x SSC. La longitud del híbrido es de la longitud anticipada para la hibridación del ácido nucleico. Cuando se hibridan moléculas de ácido nucleico de secuencia conocida, se puede determinar la longitud del híbrido por medio de la alineación de las secuencias y la identificación de las regiones conservadas descritas aquí. 1x SSC es NaCl 0,15 M y citrato sódico 15 mM; la solución de hibridación y las soluciones de lavado pueden incluir además 5 x del reactivo de Denhardt, 0,5 - 1,0% de SDS, 100 µg/ml de ADN de esperma de salmón fragmentado y desnaturalizado, 0,5% de pirofosfato de sodio.

Para los propósitos de definir el nivel de rigurosidad, puede hacerse referencia a Sambrook et al. (2001) Molecular Cloning: a laboratory manual, 3a Edición, Cold Spring Harbor Laboratory Press, CSH, New York o a Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989 y sus actualizaciones anuales).

Variante de Empalme

El término "variante de empalme" tal como se utiliza aquí abarca las variantes de una secuencia de ácido nucleico en las cuales se han suprimido, reemplazado, desplazado o agregado los intrones y/o exones seleccionados, o en las cuales se han acortado o alargado los intrones. Tales variantes serán aquellas en las cuales se retiene sustancialmente la actividad biológica de la proteína; esto se puede lograr reteniendo selectivamente segmentos

funcionales de la proteína. Tales variantes de empalme pueden hallarse en la naturaleza o pueden ser artificiales. Los métodos para predecir y aislar tales variantes de empalme son bien conocidos en la técnica (ver por ejemplo Foissac and Schiex (2005) BMC Bioinformatics 6: 25).

5 Variante alélica

- Los alelos o las variantes alélicas son formas alternativas de un gen dado, localizados en la misma posición cromosómica. Las variantes alélicas abarcan los Polimorfismos de un Solo Nucleótido (SNP por sus siglas en inglés), así como también los Polimorfismos de Inserción/Supresión pequeña (INDEL por sus siglas en inglés). El tamaño de los INDEL es usualmente inferior a 100 pb. Los SNP y los INDEL forman el conjunto más grande de variantes de secuencia en las cadenas polimórficas de origen natural de la mayoría de los organismos.

Transposición de genes / Evolución dirigida

- 15 La transposición de genes o la evolución dirigida consiste de iteraciones de la transposición del ADN seguida por la detección y/o selección apropiada para generar variantes de secuencias de ácido nucleico o porciones de las mismas que codifican las proteínas que tienen una actividad biológica modificada (Castle et al., (2004) Science 304(5674): 1151 - 4; patentes estadounidenses Nos. 5.811.238 y 6.395.547).

20 Elemento regulador / Secuencia de control / Promotor

- Los términos "elemento regulador", "secuencia de control" y "promotor" se utilizan todos en forma intercambiable en la presente descripción y se deben tomar en un contexto amplio para referirse a secuencias reguladoras de ácido nucleico capaces de efectuar la expresión de las secuencias a las cuales están ligadas. El término 'promotor' típicamente se refiere a una secuencia control de la secuencia de ácido nucleico localizada secuencia arriba del inicio de la transcripción de un gen y la cual está involucrada en el reconocimiento y enlazamiento de la ARN polimerasa y otras proteínas, dirigiendo de este modo la transcripción de un de ácido nucleico operativamente enlazado. Abarcados por los términos antes mencionados, se hallan las secuencias reguladoras de la transcripción que se derivan de un gen genómico eucariota clásico (que incluye la caja TATA la cual es necesaria para la iniciación precisa de la transcripción, con o sin una secuencia de la caja CCAAT) y elementos reguladores adicionales (es decir, secuencias activadoras secuencia arriba, reforzadoras y silenciadoras) los cuales alteran la expresión génica en respuesta a estímulos de desarrollo y/o externos, o en una forma específica del tejido. También se incluyen dentro del término una secuencia reguladora de la transcripción de un gen procariota clásico, en cuyo caso puede incluir una secuencia de caja -35 y/o secuencias reguladoras de la transcripción de caja -10. El término "elemento regulador" también abarca una molécula sintética de fusión o un derivado que confiere, activa o mejora la expresión de una molécula de la secuencia de ácido nucleico en una célula, tejido u órgano.

- Un "promotor de una planta" incluye elementos reguladores, los cuales median la expresión de un segmento de una secuencia de codificación en células de plantas. El "promotor de la planta" preferiblemente se origina a partir de una célula de la planta, por ejemplo, a partir de la planta la cual es transformada con la secuencia de ácido nucleico que va a ser expresada en el proceso de la invención y descrita aquí. Esto también aplica a otras señales reguladoras de la "planta", tales como terminadores de la "planta". Los promotores secuencia arriba de las secuencias nucleotídicas útiles en los métodos de la presente invención pueden ser modificados mediante una o más sustituciones, inserciones y/o supresiones de nucleótidos sin interferir con la funcionalidad o actividad ya sea de los promotores, el marco de lectura abierto (ORF por sus siglas en inglés) o la región reguladora 3' tal como terminadores u otras regiones reguladoras 3' las cuales se localizan lejos del ORF. Además es posible que la actividad de los promotores se incremente por la modificación de su secuencia, o que sean completamente reemplazados por promotores más activos, incluso promotores de organismos heterólogos. Para la expresión en plantas, la molécula de la secuencia de ácido nucleico debe, como se describió anteriormente, estar operativamente enlazada a un promotor adecuado, o contener un promotor adecuado, el cual expresa al gen en el momento adecuado y con el patrón de expresión espacial requerido.

- Para la identificación de promotores funcionalmente equivalentes, se pueden analizar la fuerza del promotor y/o el patrón de expresión de un promotor candidato por ejemplo enlazando operativamente el promotor a un gen reportero y ensayando el nivel de expresión y el patrón del gen reportero en diferentes tejidos de la planta. Los genes reporteros adecuados bien conocidos incluyen por ejemplo a la beta-glucuronidasa o a la beta-galactosidasa. La actividad del promotor se ensaya mediante la medición de la actividad enzimática de la beta-glucuronidasa o la beta-galactosidasa. La fuerza del promotor y/o el patrón de expresión pueden ser comparados luego con la de un promotor de referencia (tal como el utilizado en los métodos de la presente invención). En forma alternativa, la fuerza del promotor se puede ensayar cuantificando los niveles de ARNm o comparando los niveles de ARNm de la secuencia de ácido nucleico utilizada en los métodos de la presente invención, con niveles de ARNm de los genes de mantenimiento tales como el ARNr 18S, utilizando métodos conocidos en la técnica, tal como las transferencias tipo Northern con análisis densitométrico de autorradiogramas, PCR cuantitativa en tiempo real o RT-PCR (Heid et al, 1996, Genome Methods 6: 986 - 994). En general, por "promotor débil" se entiende un promotor que dirige la expresión de una secuencia de codificación a un nivel bajo. Por "nivel bajo" se entiende a niveles de aproximadamente 1/10.000 de transcritos hasta aproximadamente 1/100.000 de transcritos, hasta

aproximadamente 1/500.0000 de transcriptos por célula. Por el contrario, un "promotor fuerte" dirige la expresión de una secuencia de codificación a nivel alto, o aproximadamente a 1/10 de transcriptos hasta aproximadamente 1/100 de transcriptos hasta aproximadamente 1/1.000 de transcriptos por célula. En general, por "promotor de fuerza media" se entiende un promotor que dirige la expresión de una secuencia de codificación a un nivel que está en todos los casos por debajo de aquel obtenido bajo el control de un promotor 35S del CaMV.

Operativamente enlazado

El término "operativamente enlazado" tal como se utiliza aquí, se refiere a un enlace funcional entre la secuencia promotora y el gen de interés, de modo que la secuencia promotora pueda iniciar la transcripción del gen de interés.

Promotor constitutivo

Un "promotor constitutivo" se refiere a un promotor que es transcripcionalmente activo durante la mayoría de las fases, pero no necesariamente todas, de crecimiento y desarrollo y bajo la mayoría de las condiciones ambientales, en al menos una célula, tejido u órgano. La Tabla 2a a continuación muestra ejemplos de promotores constitutivos.

Tabla 2a: Ejemplos de promotores constitutivos de la planta

Fuente de Genes	Referencia
Actina	McElroy et al, Plant Cell, 2: 163 - 171, 1990
HMGB	WO 2004/070039
GOS2	de Pater et al, Plant J Nov; 2(6): 837 - 44, 1992, WO 2004/065596
Ubiquitina	Christensen et al, Plant Mol. Biol. 18: 675 - 689, 1992
Ciclofilina de arroz	Buchholz et al, Plant Mol Biol. 25(5): 837 - 43, 1994
Histona H3 de maíz	Lepetit et al, Mol. Gen. Genet. 231: 276 - 285, 1992
Histona H3 de alfalfa	Wu et al. Plant Mol. Biol. 11: 641 - 649, 1988
Actina 2	An et al, Plant J. 10(1); 107 - 121, 1996
Subunidad pequeña de Rubisco	Patente de los Estados Unidos No. 4.962.028
Ocs	Leisner (1988) Proc Natl Acad Sci USA 85(5): 2553
SAD1	Jain et al., Crop Science, 39 (6), 1999: 1696
SAD2	Jain et al., Crop Science, 39 (6), 1999: 1696
V-ATPasa	WO 01/14572
Proteínas de caja G	WO 94/12015

Promotor ubicuo

Un promotor ubicuo es activo sustancialmente en todos los tejidos o células de un organismo.

Promotor regulado por desarrollo

Un promotor regulado por desarrollo es activo durante ciertas etapas del desarrollo o en partes de la planta que experimentan cambios en el desarrollo.

Promotor inducible

Un promotor inducible ha inducido o incrementado la iniciación de la transcripción en respuesta a un estímulo químico (para una revisión véase Gatz, 1997, Annu. Rev. Plant Physiol. Plant Mol. Biol., 48: 89 - 108), ambiental o físico, o puede ser "inducible por estrés", es decir activado cuando una planta es expuesta a diferentes condiciones de estrés, o un promotor "inducible por patógenos", es decir activado cuando una planta es expuesta a diferentes patógenos.

Promotor específico de un órgano / Específico de un tejido

Un promotor específico de un órgano o específico de un tejido es uno que es capaz de iniciar Preferiblemente la transcripción en ciertos órganos o tejidos, tales como las hojas, las raíces, el tejido de la semilla, etc. Por ejemplo, un "promotor específico de la raíz" es un promotor que es transcripcionalmente activo predominantemente en las raíces de la planta, sustancialmente con la exclusión de cualquier otra parte de una planta, mientras que aún permite cualquier expresión defectuosa en estas otras partes de la planta. Los promotores capaces de iniciar la transcripción en ciertas células únicamente se denominan aquí como "específicos de la célula".

Ejemplos de promotores específicos de la raíz se enumeran en la Tabla 2b a continuación:

Tabla 2b: Ejemplos de promotores específicos de la raíz

Fuente de Genes	Referencia
RCc3 de arroz	Xu et al (1995) Plant Mol Biol 27(2): 237 - 48
Transportador de fosfato PHT1 de Arabidopsis	Kovama et al., 2005
Transportador de fosfato de Medicago	Xiao et al., 2006
Pyk10 de Arabidopsis	Nitz et al. (2001) Plant Sci 161 (2): 337 - 346
Genes RB7, RD2, RD5, RH12 específicos de la raíz del tabaco	Conkling et al. (1990) Plant Phys 93(3): 1203-1211
Lectina específica de la raíz de cebada	Lerner & Raikhel (1989) Plant Phys 91: 124 - 129
Proteína rica en hidroxiprolina específica de la raíz	Keller & Lamb (1989) Genes & Dev 3: 1639 - 1646
CDC27B/hobbit de Arabidopsis	Blilou et al. (2002) Genes & Dev 16: 2566 - 2575

Un promotor específico de semilla es transcripcionalmente activo predominantemente en el tejido de la semilla, pero no necesariamente exclusivamente en el tejido de la semilla (en casos de expresión defectuosa). El promotor específico de la semilla puede estar activo durante el desarrollo de la semilla y/o durante la germinación. Ejemplos de promotores específicos de semilla se muestran en la Tabla 2c a continuación. Otros ejemplos de promotores específicos de semilla se presentan en Qing Qu y Takaiwa (Plant Biotechnol. J. 2, 113 - 125, 2004), cuya divulgación se incorpora aquí por referencia como si se la hubiera expuesto en su totalidad.

Tabla 2c: Ejemplos de promotores específicos de la semillas

Fuente de Genes	Referencia
Gene específicos de la semilla	Simon et al., Plant Mol. Biol. 5: 191, 1985.
	Scofield et al., J. Biol. Chem. 262: 12202, 1987.
	Baszczyński et al., Plant Mol. Biol. 14: 633, 1990.
Albúmina de la nuez del Brasil	Pearson et al., Plant Mol. Biol. 18: 235 - 245, 1992.
Legumina	Ellis et al., Plant Mol. Biol. 10: 203 - 214, 1988.
glutelina (arroz)	Takaiwa et al., Mol. Gen. Genet. 208: 15 - 22, 1986;
	Takaiwa et al., FEBS Letts. 221: 43 - 47, 1987.
Zeina	Matzke et al Plant Mol Biol, 14(3): 323 - 32 1990
NapA	Stalberg et al, Planta 199: 515 - 519, 1996.
Glutenina-1 LMW y HMW de trigo	Mol Gen Genet 216: 81 - 90, 1989; NAR 17: 461 - 2, 1989
SPA de trigo	Albani et al, Plant Cell, 9: 171-184, 1997
α , β , γ -gliadinas de trigo	EMBO J. 3: 1409 - 15, 1984
Promotor ltr1 de cebada	Diaz et al. (1995) Mol Gen Genet 248(5): 592 - 8
Hordeína B1, C, D de cebada	Theor Appl Gen 98: 1253 - 62, 1999; Plant J 4: 343 - 55, 1993; Mol Gen Genet 250: 750 - 60, 1996
DOF de cebada	Mena et al, The Plant Journal, 116(1): 53 - 62, 1998
blz2	EP99106056.7
Promotor sintético	Vicente-Carbajosa et al., Plant J. 13: 629 - 640, 1998.
Prolamina NRP33 de arroz	Wu et al, Plant Cell Physiology 39(8) 885 - 889, 1998
α -globulina G1b-1 de arroz	Wu et al, Plant Cell Physiology 39(8) 885 - 889, 1998
OSH1 de arroz	Sato et al, Proc. Natl. Acad. Sci. USA, 93: 8117 - 8122, 1996
α -globulina REB/OHP-1 de arroz	Nakase et al. Plant Mol. Biol. 33: 513 - 522, 1997
ADP-glucosa pirofosforilasa de arroz	Trans Res 6: 157 - 68, 1997
Familia de genes ESR del maíz	Plant J 12: 235 - 46, 1997
α -kafirina del sorgo	DeRose et al., Plant Mol. Biol 32: 1029 - 35, 1996
KNOX	Postma-Haarsma et al, Plant Mol. Biol. 39: 257 - 71, 1999
Oleosina del arroz	Wu et al, J. Biochem. 123: 386, 1998
Oleosina de girasol	Cummins et al., Plant Mol. Biol. 19: 873 - 876, 1992
40S putativa del arroz, PRO0117	WO 2004/070039
Proteína ribosomal	
Alanina aminotransferasa del arroz, PRO0136	No publicada
Inhibidor ITR1 de tripsina (cebada), PRO0147	No publicada
WSI18 de arroz, PRO0151	WO 2004/070039
RAB21 de arroz, PRO0175	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039

(continuación)

Fuente de Genes	Referencia
α -amilasa (Amy32b)	Lanahan et al, Plant Cell 4: 203 - 211, 1992; Skriver et al, Proc Natl Acad Sci USA 88: 7266 - 7270, 1991
Gen tipo catepsina β	Cejudo et al, Plant Mol Biol 20: 849 - 856, 1992
Cebada Ltp2	Kalla et al., Plant J. 6: 849 - 60, 1994
Chi26	Leah et al., Plant J. 4: 579 - 89, 1994
Maíz B-Perú	Selinger et al., Genetics 149: 1125 - 38, 1998

Un promotor específico del tejido verde como se define aquí, es un promotor que es transcripcionalmente activo predominantemente en tejido verde, sustancialmente con la exclusión de cualquier otra de las partes de una planta, mientras que aún permite cualquier expresión defectuosa en estas otras partes de la planta.

Ejemplos de promotores específicos del tejido verde que pueden ser utilizados para llevar a cabo los métodos de la invención son presentados en la Tabla 2d a continuación.

Tabla 2d: Ejemplos de promotores específicos del tejido verde

Gen	Expresión	Referencia
Ortofosfato diquinasa del maíz	Específica de la hoja	Fukavama et al., 2001
Fosfoenolpiruvato carboxilasa del maíz	Específica de la hoja	Kausch et al., 2001
Fosfoenolpiruvato carboxilasa del arroz	Específica de la hoja	Liu et al., 2003
Pequeña subunidad de Rubisco del arroz	Específica de la hoja	Nomura et al., 2000
Beta expansina EXBP9 del arroz	Específica del brote	WO 2004/070039
Pequeña subunidad de Rubisco del guandú	Específica de la hoja	Panguluri et al., 2005
RBCS3A del guisante	Específica de la hoja	

Otro ejemplo de un promotor específico del tejido es un promotor específico del meristema, el cual es transcripcionalmente activo predominantemente en el tejido meristemático, sustancialmente con la exclusión de cualquier otra de las partes de una planta, mientras que aún permite cualquier expresión defectuosa en estas otras partes de la planta. Ejemplos de promotores específicos del meristema que pueden ser utilizados para llevar a cabo los métodos de la invención se presentan en la Tabla 2e a continuación.

Tabla 2e: Ejemplos de promotores específicos del meristema

Fuente del gen	Patrón de expresión	Referencia
OSH1 del arroz	Meristema apical del brote, desde la etapa globular embrionaria hasta la etapa de plántula	Sato et al. (1996) Proc. Natl. Acad. Sci. USA, 93: 8117 - 8122
Metalotioneína del arroz	Específico del meristema	BAD87835.1
WAK1 & WAK 2	Meristemas apicales de raíz y de brote, y en hojas y sépalos en expansión	Wagner & Kohorn (2001) Plant Cell 13(2): 303 - 318

Terminador

El término "terminador" abarca una secuencia de control que es una secuencia de ADN al extremo de una unidad de transcripción la cual señala el procesamiento 3' y la poliadenilación de un transcripto primario y la terminación de la transcripción. El terminador puede derivarse del gen natural, a partir de una variedad de otros genes de planta, o del ADN-T. El terminador que va a ser añadido puede derivarse de, por ejemplo, los genes de la nopalina sintasa o de la octopina sintasa, o en forma alternativa de otro gen de la planta, o con menor preferencia a partir de cualquier otro gen eucariota.

Modulación

El término "modulación" significa en relación con expresión o con expresión génica, un proceso en el cual el nivel de expresión es cambiado por dicha expresión génica en comparación con la planta de control. Preferiblemente el nivel de expresión se incrementa. La expresión original, no modulada puede ser de cualquier tipo de expresión de un ARN estructural (ARNr, ARNt) o de ARNm con posterior traducción. El término "que modula la actividad" significa cualquier cambio de la expresión de las secuencias de ácido nucleico de la invención o las proteínas codificadas, lo cual conduce a un mayor rendimiento y/o a un mayor crecimiento de las plantas.

Expresión incrementada / sobreexpresión

El término "expresión incrementada" o "sobreexpresión" tal como se utiliza aquí significa cualquier forma de expresión que sea adicional al nivel original de expresión de tipo silvestre.

Los métodos para incrementar la expresión de los genes o de los productos génicos están bien documentados en la técnica e incluyen, por ejemplo, la sobreexpresión conducida por promotores apropiados, el uso de incrementadores de la transcripción o incrementadores de la traducción. Las secuencias de ácido nucleico aisladas que sirven como elementos promotores o incrementadores pueden ser introducidas en una posición apropiada (típicamente secuencia arriba) de una forma no heteróloga de un polinucleótido para favorecer la expresión de una secuencia de ácido nucleico que codifica al polipéptido de interés. Por ejemplo, los promotores endógenos pueden ser alterados *in vivo* por medio de mutación, supresión, y/o sustitución (véase, Kmiec, patente de los Estados Unidos No. 5.565.350; Zarling et al., WO9322443), o se pueden introducir los promotores aislados en una célula de una planta en la orientación y distancia apropiados de un gen de la presente invención para controlar la expresión del gen.

Si se desea una expresión polipeptídica, en general es deseable incluir una región de poliadenilación en el extremo 3' de un región de codificación del polinucleótido. La región de poliadenilación puede derivarse del gen natural, a partir de una variedad de otros genes de la planta, o a partir del ADN-T. La secuencia del extremo de 3' que va a ser añadida puede derivarse, por ejemplo, de los genes de la nopalina sintasa o de la octopina sintasa, o en forma alternativa a partir de otro gen de la planta, o menos preferiblemente a partir de cualquier otro gen eucariota.

También se puede añadir una secuencia del intrón a la región no traducida (UTR) 5' o en la secuencia codificadora de la secuencia de codificación parcial para incrementar la cantidad del mensaje de maduración que se acumula en el citosol. La inclusión de un intrón que puede empalmarse en la unidad de transcripción tanto en los constructos de expresión de plantas como de animales ha demostrado que incrementa la expresión génica tanto en los niveles de ARNm como de proteína hasta 1000 veces (Buchman y Berg (1988) Mol. Cell Biol. 8: 4395 - 4405; Callis et al. (1987) Genes Dev. 1: 1183 - 1200). Tal incremento del intrón de la expresión génica es típicamente mayor cuando se localiza cerca del extremo 5' de la unidad de transcripción. El uso de los intrones del maíz, intrón Adh1-S 1, 2 y 6 y el intrón de Bronce-1 es conocido en la técnica. Para información general, véase: The Maize Handbook, Capítulo 116, Freeling and Walbot, Eds., Springer, N.Y. (1994).

Gen endógeno

La referencia aquí a un gen "endógeno" no sólo se refiere al gen en cuestión tal como se encuentra en una planta en su forma natural (es decir, sin que exista intervención humana alguna), sino que también se refiere al mismo gen (o un ácido nucleico/gen sustancialmente homólogo) en un forma aislada posteriormente (re)introducido en una planta (un transgén). Por ejemplo, una planta transgénica que contiene dicho transgén puede encontrar una reducción sustancial de la expresión del transgén y/o una reducción sustancial de la expresión del gen endógeno.

Expresión disminuida

La referencia que se hace aquí a una "expresión disminuida" o a una "reducción o eliminación sustancial" de la expresión se pretende que signifique una disminución en la expresión del gen endógeno y/o de los niveles de polipéptido y/o de la actividad del polipéptido con relación a las plantas de control. La reducción o la eliminación sustancial es en orden creciente de preferencia una reducción de al menos 10%, 20%, 30%, 40% o 50%, 60%, 70%, 80%, 85%, 90%, o 95%, 96%, 97%, 98%, 99% o más en comparación con las de las plantas de control.

Para la reducción o la eliminación sustancial de la expresión de un gen endógeno en una planta, se requiere de una longitud suficiente de nucleótidos sustancialmente contiguos de una secuencia de ácido nucleico. A fin de llevar a cabo el silenciamiento génico, este puede ser tan pequeño como de 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 o menos nucleótidos, alternativamente este puede ser tanto como el gen completo (incluyendo la UTR de 5' y/o 3', ya sea en parte o completa). El alargamiento de nucleótidos sustancialmente contiguos puede derivarse de la secuencia de ácido nucleico que codifica la proteína de interés (gen objetivo), o de cualquier secuencia de ácido nucleico capaz de codificar un ortólogo, un parálogo o un homólogo de la proteína de interés. Preferiblemente, el alargamiento de los nucleótidos sustancialmente contiguos es capaz de formar enlaces de hidrógeno con el gen objetivo (ya sea una cadena sentido o antisentido), más preferiblemente, el alargamiento de los nucleótidos sustancialmente contiguos tiene, en orden creciente de preferencia, una identidad de secuencia del 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100% con respecto al gen objetivo (ya sea una cadena sentido o antisentido). Una secuencia de ácido nucleico que codifica un polipéptido (funcional) no es un requerimiento para los diferentes métodos discutidos aquí para la reducción o eliminación sustancial de la expresión de un gen endógeno.

Esta reducción o eliminación sustancial de la expresión puede lograrse mediante el uso de herramientas y técnicas de rutina. Un método para la reducción o eliminación sustancial de la expresión del gen endógeno es mediante el silenciamiento mediado por ARN utilizando una repetición invertida de una secuencia de ácido nucleico o una parte de la misma (en este caso un alargamiento de nucleótidos sustancialmente contiguos derivados del gen de interés, o de cualquier secuencia de ácido nucleico capaz de codificar un ortólogo, parálogo u homólogo de la proteína de interés), preferiblemente capaz de formar una estructura de horquilla. Otro ejemplo de un método de silenciamiento del ARN involucra la introducción de secuencias de ácido nucleico o de partes de las mismas (en este caso un alargamiento de nucleótidos sustancialmente contiguos que derivan del gen de interés, o de cualquier secuencia de ácido nucleico capaz de codificar un ortólogo, parálogo u homólogo de la proteína de interés), en una orientación sentido en una planta. Otro ejemplo de un método de silenciamiento de ARN involucra el uso de secuencias de

ácido nucleico antisentido. El silenciamiento génico también puede lograrse mediante mutagénesis de inserción (por ejemplo, inserción de ADN-T o inserción de un transposón) o mediante estrategias como las descritas, entre otros, por Angeli y Baulcombe ((1999) Plant J 20(3): 357 - 62), (Amplicón VIGS WO 98/36083), o Baulcombe (WO 99/15682). Otros métodos, tal como el uso de anticuerpos dirigidos a un polipéptido endógeno para inhibir su función en la planta, o la interferencia en la ruta de señalización en la cual está involucrado un polipéptido, serán bien conocidos por aquellos normalmente capacitados en el arte. Se pueden utilizar microARN artificiales (miARNa) y/o naturales para desactivar la expresión génica y/o la traducción del ARNm. Los miARN endógenos son los ARN monocatenarios pequeños típicamente de 19 - 24 nucleótidos de largo. Los microARN artificiales (miARNa), que tienen típicamente 21 nucleótidos de longitud, pueden ser modificados por ingeniería genética específicamente para regular en forma negativa la expresión génica de genes individuales o múltiples de interés. Los determinantes de la selección objetivo del ARN micro de una planta son bien conocidos en la técnica. Los parámetros empíricos para el reconocimiento del objetivo han sido definidos y pueden ser utilizados para ayudar en el diseño de los ARNmia específicos (Schwab et al., (2005) Dev Cell 8(4): 517 - 27). Las herramientas convenientes para el diseño y generación de los ARNmia y sus precursores también se encuentran disponibles al público (Schwab et al., (2006) Plant Cell 18(5): 1121 - 33).

Para un desempeño óptimo, las técnicas de silenciamiento génico utilizadas para reducir la expresión en una planta de un gen endógeno requieren del uso de secuencias de ácido nucleico de plantas monocotiledóneas para la transformación de plantas monocotiledóneas, y de plantas dicotiledóneas para la transformación de plantas dicotiledóneas. Preferiblemente, se introduce una secuencia de ácido nucleico de cualquier especie de planta dada en esa misma especie. Por ejemplo, una secuencia de ácido nucleico de arroz se transforma en una planta de arroz. Sin embargo, no es un requerimiento absoluto que la secuencia de ácido nucleico que va a ser introducida se origine a partir de la misma especie de planta que la planta en la cual será introducida. Es suficiente que exista homología sustancial entre el gen objetivo endógeno y la secuencia de ácido nucleico que va a ser introducida.

Se describieron anteriormente ejemplos de diferentes métodos para la reducción o eliminación sustancial de la expresión en una planta de un gen endógeno. Un persona normalmente capacitada en la técnica podría adaptar fácilmente los métodos anteriormente mencionados para el silenciamiento a fin de lograr la reducción de la expresión de un gen endógeno en una planta completa o en partes de la misma a través, por ejemplo, del uso de un promotor apropiado.

Marcador seleccionable (gen) / Gen reportero

Los términos "marcador seleccionable", "gen marcador seleccionable" o "gen reportero" incluyen a cualquier gen que confiera un fenotipo a una célula en la cual se exprese para facilitar la identificación y/o la selección de las células que son transfectadas o transformadas con un constructo de una secuencia de ácido nucleico de la invención. Estos genes marcadores permiten la identificación de una transferencia exitosa de las moléculas de la secuencia de ácido nucleico mediante una serie de principios diferentes. Los marcadores adecuados se pueden seleccionar de los marcadores que confieren resistencia antibiótica o herbicida, que introducen una nueva característica metabólica o que permiten una selección visual. Los ejemplos de genes marcadores seleccionables incluyen a los genes que confieren resistencia a los antibióticos (tales como nptII que fosforilan la neomicina y la kanamicina, o hpt, que fosforila higromicina, o genes que confieran resistencia, por ejemplo, a la bleomicina, estreptomycin, tetraciclina, cloramfenicol, ampicilina, gentamicina, genética (G418), espectinomycin o blasticidina), a los herbicidas (por ejemplo bar el cual proporciona resistencia a Basta®; aroA o gox que proporcionan resistencia contra el glifosato, o los genes que confieren resistencia, por ejemplo, a la imidazolinona, fosfinotricina o sulfonilurea), o genes que proporcionen una característica metabólica (tal como manA que le permite a las plantas utilizar manosa como única fuente de carbono o xilosa isomerasa para la utilización de xilosa, o marcadores antinutritivos tales como la resistencia a la 2-desoxiglucosa). La expresión de los genes marcadores visuales resulta en la formación de color (por ejemplo β -glucuronidasa, GUS o β -galactosidasa con sus sustratos coloreados, por ejemplo X-Gal), luminiscencia (tal como el sistema de luciferina / luciferasa) o fluorescencia (Proteína Fluorescente Verde, GFP, y sus derivados). Esta lista representa sólo una pequeña cantidad de posibles marcadores. La persona ordinariamente capacitada en la técnica está familiarizada con tales marcadores. Se prefieren marcadores diferentes, dependiendo del organismo y del método de selección.

Se sabe que después de una integración estable o transitoria de secuencias de ácido nucleico en las células de una planta, sólo una minoría de las células incorpora el ADN foráneo y, si se desea, lo integra en su genoma, dependiendo del vector de expresión utilizado y de la técnica de transfección utilizada. Para identificar y seleccionar estos integrantes, se introduce usualmente un gen que codifica un marcador seleccionable (tal como aquellos descritos anteriormente) en las células huésped junto con el gen de interés. Estos marcadores pueden ser utilizados, por ejemplo, en mutantes en los cuales estos genes no son funcionales, por ejemplo, por medio de la supresión mediante métodos convencionales. Además, se pueden introducir las moléculas de la secuencia de ácido nucleico que codifican un marcador seleccionable en una célula huésped en el mismo vector que contiene la secuencia que codifica los polipéptidos de la invención o que se utilizan en los métodos de la invención, o también en un vector separado. Las células que han sido transfectadas en forma estable con la secuencia de ácido nucleico introducida pueden ser identificadas por ejemplo mediante selección (por ejemplo, las células que han integrado el marcador seleccionable sobreviven mientras que las otras células mueren).

Ya que los genes marcadores, particularmente los genes para resistencia a los antibióticos y herbicidas, no se requieren más o no son deseables en la célula huésped transgénica una vez que se han sido introducidos con éxito las secuencias de ácido nucleico, el proceso de acuerdo con la invención para introducir las secuencias de ácido nucleico emplea en forma conveniente técnicas que permiten la remoción o escisión de estos genes marcadores. Un de tales métodos es el que se conoce como cotransformación. El método de cotransformación emplea dos vectores en forma simultánea para la transformación, un vector que porta la secuencia de ácido nucleico de acuerdo con la invención y un segundo vector que porta el gen o los genes marcadores. Una gran proporción de transformantes recibe, o, en el caso de las plantas, incluye (hasta 40% o más de los transformantes), ambos vectores. En el caso de la transformación con *Agrobacterias*, los transformantes usualmente reciben solamente una parte del vector, es decir la secuencia flanqueada por el ADN-T, la cual usualmente representa el casete de expresión. Los genes marcadores pueden ser removidos posteriormente de la planta transformada por medio de cruzamientos. En otro método, los genes marcadores integrados en un transposón se utilizan para la transformación junto con la secuencia de ácido nucleico deseada (conocida como la tecnología Ac/Ds). Los transformantes se pueden cruzar con una fuente de transposasa o se transforman los transformantes con un constructo de la secuencia de ácido nucleico que confiere la expresión de una transposasa, en forma transitoria o estable. En algunos casos (aproximadamente el 10%), el transposón salta fuera del genoma de la célula huésped una vez a tenido lugar la transformación en forma exitosa y se pierde. En otro número de casos, el transposón salta a una ubicación diferente. En estos casos el gen marcador debe ser eliminado mediante la realización de cruzamientos. En microbiología, se desarrollaron técnicas que hacen posible, o facilitan, la detección de tales eventos. Un método ventajoso adicional se fundamenta en lo que se conoce como sistemas de recombinación, cuya ventaja es que se puede prescindir de la eliminación por cruzamiento. El sistema más conocido de este tipo es el que se conoce como sistema Cre/lox. Cre1 es una recombinasa que remueve las secuencias localizadas entre la secuencias de loxP. Si el gen marcador se integra entre las secuencias de loxP, es eliminado una vez que ha tenido lugar con éxito la transformación, por medio de la expresión de la recombinasa. Otros sistemas de recombinación son los sistemas HIN/HIX, FLP/FRT y REP/STB (Tribble et al., J. Biol. Chem., 275, 2000: 22255 - 22267; Velmurugan et al., J. Cell Biol., 149, 2000: 553 - 566). Es posible una integración específica del sitio en el genoma de la planta de las secuencias de ácido nucleico de acuerdo con la invención. Naturalmente, estos métodos se pueden aplicar también a microorganismos tales como levaduras, hongos o bacterias.

Transgénico / Transgén / Recombinante

Para los propósitos de la invención, "transgénico", "transgén" o "recombinante" significan con respecto, por ejemplo, a una secuencia de ácido nucleico, un casete de expresión, una construcción génica o un vector que comprende la secuencia de ácido nucleico o un organismo transformado con las secuencias de ácido nucleico, los casetes de expresión o los vectores de acuerdo con la invención, todos aquellos constructos logrados por métodos recombinantes en los cuales o bien

- (a) las secuencias de ácido nucleico que codifican las proteína útiles en los métodos de la invención, o
- (b) la(s) secuencia(s) genética(s) de control que está(n) operativamente enlazada(s) con la secuencia de ácido nucleico de acuerdo con la invención, por ejemplo un promotor, o
- (c) a) y b)

no están localizadas en su ambiente genético natural o han sido modificadas por métodos recombinantes, siendo posible que la modificación tomara la forma, por ejemplo, de una sustitución, adición, supresión, inversión o inserción de uno o más residuos nucleótidos. Se entiende que el ambiente genético natural significa el locus genómico o cromosómico natural en la planta original o la presencia en una biblioteca genómica. En el caso de una biblioteca genómica, el ambiente genético natural de la secuencia de ácido nucleico es preferiblemente retenida, al menos en parte. El ambiente flanquea la secuencia de ácido nucleico al menos en un costado y tiene una longitud de secuencia de al menos 50 pb, Preferiblemente de al menos 500 pb, especialmente preferiblemente de al menos 100 pb, lo más preferible al menos 5000 pb. Un casete de expresión de origen natural - por ejemplo la combinación de origen natural del promotor natural de las secuencias de ácido nucleico con la correspondiente secuencia de ácido nucleico que codifica un polipéptido útil en los métodos de la presente invención, como se definió anteriormente - convierte un casete de expresión transgénica cuando este casete de expresión es modificado por métodos sintéticos no naturales ("artificiales") tales como, por ejemplo, tratamiento mutagénico. Los métodos adecuados son descritos, por ejemplo, en la patente de los Estados Unidos No. 5.565.350 o en WO 00/15815.

Una planta transgénica para los propósitos de la invención se entiende entonces que significa, como se indicó anteriormente, que las secuencias de ácido nucleico utilizadas en el método de la invención no están en su locus natural en el genoma de dicha planta, siendo posible para las secuencias de ácido nucleico que sean expresadas en forma homóloga o heteróloga. Sin embargo, como se mencionó anteriormente, transgénico también significa que, mientras las secuencias de ácidos nucleicos de acuerdo con la invención o utilizadas en el método de la invención, estén en su posición natural en el genoma de una planta, la secuencia ha sido modificada y con respecto a la secuencia natural, y/o que las secuencias reguladoras de las secuencias naturales han sido modificadas. Transgénico Preferiblemente significa que la expresión de las secuencias de ácidos nucleicos de acuerdo con la invención en un locus no natural en el genoma, es decir, que tiene lugar una expresión homóloga o, Preferiblemente

heteróloga de las secuencias de ácido nucleico. Las plantas transgénicas preferidas se mencionan en la presente invención.

Transformación

El término "introducción" o "transformación" a los cuales se hace referencia aquí, abarcan la transferencia de un polinucleótido exógeno dentro de una célula huésped, independientemente del método utilizado para la transferencia. El tejido de la planta capaz de propagación clonal posterior, ya sea por medio de organogénesis o de embriogénesis, puede ser transformado con una construcción genética de la presente invención y se puede regenerar una planta entera a partir del mismo. El tejido particular escogido variará dependiendo de los sistemas de propagación clonal disponibles y más adecuados, para la especie particular que va a ser transformada. Los ejemplos de tejidos objetivo incluyen discos de hoja, polen, embriones, cotiledones, hipocotiledones, mega gametofitos, tejido del callo, tejido meristemático existente (por ejemplo, meristema apical, yemas axilares, y meristemas de la raíz), y tejido meristemático inducido (por ejemplo, meristema del cotiledón y meristema del hipocotiledón). El polinucleótido puede ser introducido en forma transitoria o estable en una célula huésped y puede mantenerse no integrado, por ejemplo, como un plásmido. En forma alternativa, puede integrarse al genoma huésped. La célula de la planta transformada resultante puede ser luego utilizada para regenerar una planta transformada en una forma conocida por las personas ordinariamente capacitadas en la técnica.

La transferencia de genes foráneos dentro del genoma de una planta se denomina transformación. La transformación de una especie de planta es hoy en día una técnica bastante rutinaria. En forma conveniente, se puede utilizar cualquiera de los diferentes métodos de transformación para introducir el gen de interés en una célula ancestral adecuada. Los métodos descritos para la transformación y regeneración de plantas a partir de tejidos de la planta o de células de la planta pueden ser utilizados para una transformación transitoria o estable. Los métodos de transformación incluyen el uso de liposomas, electroporación, productos químicos que aumentan la incorporación de ADN libre, inyección del ADN directamente a la planta, bombardeo de partículas, transformación que utiliza virus o polen y microproyección. Los métodos pueden ser seleccionados a partir del método de calcio/polietylén glicol para protoplastos (Krens, F. A. et al., (1982) *Nature* 296, 72 - 74; Negritiu I et al. (1987) *Plant Mol Biol* 8: 363 - 373); electroporación de protoplastos (Shillito R. D. et al. (1985) *Bio/Technol* 3, 1099 - 1102); microinyección en material de la planta (Crossway A et al, (1986) *Mol. Gen Genet* 202: 179 - 185); bombardeo con partículas recubiertas con ADN o ARN (Klein T. M. et al, (1987) *Nature* 327: 70) infección con virus (que no se integran) y similares. Las plantas transgénicas, que incluyen plantas de cultivo transgénicas, son producidas Preferiblemente mediante transformación mediada por *Agrobacterium*. Un método de transformación conveniente es la transformación *in planta*. Para este propósito, es posible, por ejemplo, permitir que las agrobacterias actúen en las semillas de la planta o inocular el meristema de la planta con las agrobacterias. Ha probado ser particularmente conveniente de acuerdo con la invención, permitir que una suspensión de agrobacterias transformadas actúen sobre la planta intacta o al menos sobre los primordios de la flor. Posteriormente se deja que se desarrolle la planta hasta obtener las semillas de la planta tratada (Clough y Bent, *Plant J.* (1998) 16, 735 - 743). Los métodos para transformación del arroz mediados por *Agrobacterium* incluyen métodos bien conocidos para la transformación del arroz, tales como los descritos en cualquiera de las siguientes documentos: solicitud de patente europea EP 1198985 A1, Aldemita y Hodges (*Planta* 199: 612 - 617, 1996); Chan et al. (*Plant Mol Biol* 22 (3): 491 - 506, 1993), Hiei et al. (*Plant J* 6 (2): 271 - 282, 1994). En el caso de la transformación del maíz, el método preferido es el descrito ya sea en Ishida et al. (*Nat. Biotechnol* 14(6): 745 - 50, 1996) o en Frame et al. (*Plant Physiol* 129(1): 13 - 22, 2002). Dichos métodos son descritos además a modo de ejemplo en B. Jenes et al., *Techniques for Gene Transfer*, en: *Transgenic Plants*, Vol. 1, Engineering and Utilization, eds. S. D. Kung y R. Wu, Academic Press (1993) 128 - 143 y en Potrykus *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 42 (1991) 205 - 225). Las secuencias de ácido nucleico o el constructo que va a ser expresado es clonado preferiblemente en un vector, el cual es adecuado para la transformación de *Agrobacterium tumefaciens*, por ejemplo pBin19 (Bevan et al., *Nucl. Acids Res.* 12 (1984) 8711). Las agrobacterias transformadas por dicho vector pueden ser luego utilizadas en una forma conocida para la transformación de plantas, tales como las plantas utilizadas como modelo, del tipo *Arabidopsis thaliana* no es considerada dentro del alcance de la presente invención como una planta de cultivo), o plantas de cultivo tales como, por ejemplo, plantas de tabaco, por ejemplo, mediante la inmersión de las hojas magulladas o picadas en una solución de agrobacterias y luego se cultivan en medios adecuados. La transformación de las plantas por medio del *Agrobacterium tumefaciens* es descrita, por ejemplo, en Höfgen y Willmitzer en *Nucl. Acid Res.* (1988) 16, 9877 o se conoce, entre otros, a partir de F. F. White, *Vectors for Gene Transfer in Higher Plants*; en *Transgenic Plants*, Vol. 1, Engineering and Utilization, eds. S. D. Kung y R. Wu, Academic Press, 1993, páginas 15 - 38.

Además de la transformación de células somáticas, las cuales tienen que ser luego regeneradas en plantas intactas, también es posible transformar las células de meristemas de la planta y en particular aquellas células que se desarrollan en gametos. En este caso, los gametos transformados siguen el desarrollo natural de la planta, dando lugar a plantas transgénicas. De este modo, por ejemplo, se tratan las semillas de *Arabidopsis* con agrobacterias y se obtienen las semillas de la planta en desarrollo de las cuales una cierta proporción se transforma y de este modo se vuelven transgénicas. [Feldman, K. A. y Marks M. D. (1987). *Mol Gen Genet* 208: 274 - 289; Feldmann K (1992). En: C Koncz, N-H Chua y J Shell, eds, *Methods in Arabidopsis Research*. World Scientific, Singapur, páginas 274 - 289]. Los métodos alternativos se basan en la remoción repetida de las inflorescencias y la incubación del sitio de escisión en el centro de la roseta con las agrobacterias transformadas, por lo que pueden obtenerse igualmente

semillas transformadas en un momento posterior en el tiempo (Chang (1994). Plant J. 5: 551 - 558; Katavic (1994). Mol Gen Genet, 245: 363 - 370). Sin embargo, un método especialmente efectivo es el método de infiltración al vacío con sus modificaciones tales como el método de "inmersión floral". En el caso de infiltración al vacío de *Arabidopsis*, se tratan plantas intactas bajo presión reducida con una suspensión de agrobacterias [Bechthold, N (1993). C R Acad Sci Paris Life Sci, 316: 1194 - 1199], mientras que en el caso del método de la "inmersión floral" se incubaba el tejido floral en desarrollo brevemente con una suspensión de agrobacterias tratada con un tensoactivo [Clough, S. J. y Bent A. F. (1998). The Plant J. 16, 735 - 743]. Se recolecta una cierta proporción de semillas transgénicas en ambos casos, y estas semillas pueden distinguirse de las semillas no transgénicas mediante el cultivo bajo las condiciones selectivas anteriormente descritas. Además, la transformación estable de los plástidos es ventajosa porque los plástidos se heredan del progenitor en la mayoría de los cultivos reduciendo o eliminando el riesgo de flujo transgénico a través del polen. La transformación del genoma del cloroplasto se logra generalmente mediante un proceso que ha sido presentado esquemáticamente en Klaus et al., 2004 [Nature Biotechnology 22 (2), 225 - 229]. En resumen, las secuencias que van a ser transformadas se clonan junto con un gen marcador seleccionable entre secuencias de flanqueo homólogas al genoma del cloroplasto. Estas secuencias de flanqueo homólogas dirigen la integración específica del sitio en el plastoma. La transformación plastidial ha sido descrita para muchas especies de plantas diferentes y se presenta una visión de conjunto en Bock (2001) Transgenic plastids in basic research and plant biotechnology. J Mol Biol. 2001 Sep 21; 312 (3): 425 - 38 o en Maliga, P (2003) Progress towards commercialization of plastid transformation technology. Trends Biotechnol. 21, 20 - 28. Otro avance biotecnológico ha sido informado recientemente en la forma de transformantes del plástido libres de marcadores, que pueden producirse mediante un gen marcador cointegrado en forma transitoria (Klaus et al., 2004, Nature Biotechnology 22(2), 225 - 229).

Marcación de la activación del ADN-T

La marcación de la activación del ADN-T (Hayashi et al. Science (1992) 1350 - 1353), involucra la inserción del ADN-T, que contiene usualmente un promotor (puede ser también un mejorador de la traducción o un intrón), en la región genómica del gen de interés o 10 kb secuencia arriba o secuencia abajo de la región de codificación de un gen en una configuración tal que el promotor dirige la expresión del gen objetivo. Típicamente, la regulación de la expresión del gen objetivo por su promotor natural se interrumpe y el gen cae bajo el control del promotor recientemente introducido. El promotor está típicamente embebido en un ADN-T. Este ADN-T se inserta al azar en el genoma de la planta, por ejemplo, a través de la infección con *Agrobacterium* y conduce a la expresión modificada de los genes cerca del ADN-T insertado. Las plantas transgénicas resultantes muestran fenotipos dominantes debido a la expresión modificada de los genes próximos al promotor introducido.

TILLING

El término "TILLING" (por sus siglas en inglés) es una abreviatura de "detección de lesiones locales inducidas en genomas" y se refiere a una tecnología de mutagénesis útil para generar y/o identificar secuencias de ácido nucleico que codifican proteínas con expresión y/o actividad modificada. TILLING también permite la selección de plantas que transportan tales variantes mutantes. Estas variantes mutantes pueden exhibir expresión modificada, ya sea en intensidad o en localización o en sincronización (si las mutaciones afectan al promotor por ejemplo). Estas variantes mutantes pueden exhibir mayor actividad que la exhibida por el gen en su forma natural. TILLING combina mutagénesis de alta densidad con métodos de selección de alto rendimiento. Las etapas típicamente seguidas en TILLING son: (a) mutagénesis EMS (Redei G. P. y Koncz C (1992) en Methods in Arabidopsis Research, Koncz C, Chua NH, Schell J, eds. Singapur, World Scientific Publishing Co, páginas 16 - 82; Feldmann et al., (1994) en Meyerowitz E. M., Somerville C. R., eds, Arabidopsis. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, páginas 137 - 172; Lightner J y Caspar T (1998) en J Martinez-Zapater, J Salinas, eds. Methods on Molecular Biology, Vol. 82. Humana Press, Totowa, NJ, páginas 91 - 104); (b) preparación y reunión del ADN de individuos; (c) amplificación por PCR de una región de interés; (d) desnaturalización y apareamiento para permitir la formación de heterodúplex; (e) DHPLC, donde la presencia de un heterodúplex en un consolidado se detecta como un pico extra en el cromatograma; (f) identificación del individuo mutante; y (g) secuenciación del producto de PCR mutante. Los métodos para TILLING son bien conocidos en el arte (McCallum et al., (2000) Nat Biotechnol 18: 455 - 457; revisado por Stemple (2004) Nat Rev Genet 5(2): 145 - 50).

Recombinación homóloga

La recombinación homóloga permite la introducción en un genoma de una secuencia de ácido nucleico seleccionada en una posición seleccionada definida. La recombinación homóloga es una tecnología estándar utilizada en forma rutinaria en ciencias biológicas para organismos inferiores tales como levadura o el musgo *Physcomitrella*. Los métodos para llevar a cabo la recombinación homóloga en plantas han sido descritos no sólo para las plantas modelo (Offringa et al. (1990) EMBO J 9(10): 3077 - 84) sino también para plantas de cultivo, por ejemplo arroz (Terada et al. (2002) Nat Biotech 20(10): 1030 - 4; Lida y Terada (2004) Curr Opin Biotech 15(2): 132 - 8).

Rendimiento

El término "rendimiento" en general significa una producción medible de valor económico, típicamente relacionado

con un cultivo específico, con un área, y con un periodo de tiempo. Las partes de plantas individuales contribuyen en forma directa al rendimiento con base en su cantidad, tamaño y/o peso, o el rendimiento real es el rendimiento por acre para un cultivo y el año, que se determina mediante la división de la producción total (incluye tanto producción cosechada como tasada) por los acres plantados. El término 'rendimiento' de una planta puede referirse a la biomasa vegetal, a órganos reproductivos, y/o a propágulos (tales como semillas) de esa planta.

Vigor temprano

El término "vigor temprano" se refiere a un crecimiento bien balanceado, activo y saludable, especialmente durante las etapas tempranas del desarrollo de la planta, y puede resultar de una mayor aptitud de la planta, debido, por ejemplo, a que las plantas se adaptan mejor a su ambiente (es decir, la optimización del uso de los recursos energéticos y la repartición entre brote y raíz). Las plantas que tienen vigor temprano también muestran una mayor supervivencia de las plántulas y un mejor establecimiento del cultivo, lo que a menudo resulta en campos muy uniformes (con el desarrollo del cultivo en forma uniforme, es decir, donde la mayoría de las plantas alcanzan las diferentes etapas del desarrollo sustancialmente al mismo tiempo), y a menudo con un rendimiento mejor y superior. Por lo tanto, el vigor temprano puede ser determinado midiendo diferentes factores, tales como el peso de mil granos, el porcentaje de germinación, el porcentaje de emergencia, el desarrollo de las plántulas, la altura de las plántulas, la longitud de la raíz, la biomasa de la raíz y de los brotes y muchos más.

Incremento / Mejoramiento / Intensificación

Los términos "incremento", "mejoramiento" o "intensificación" son intercambiables y se entienden en el sentido de la aplicación de al menos 5%, 6%, 7%, 8%, 9% o 10%, Preferiblemente de al menos 15% o 20%, más preferiblemente 25%, 30%, 35% o 40% de más rendimiento y/o desarrollo en comparación con las plantas de control según se define aquí.

Rendimiento de semillas

El incremento en el rendimiento de semillas puede manifestarse en sí mismo en una o más de las formas: a) un incremento en la biomasa de la semilla (peso total de la semilla) el cual puede ser con base en una semilla individual y/o por planta y/o por hectárea o acre, b) mayor número de flores por panícula y/o por planta; c) mayor número de semillas (llenas); d) mayor proporción de llenado de las semillas (la cual se expresa como la relación entre la cantidad de semilla llenas dividido por la cantidad total de semillas); e) mayor índice de cosecha, el cual se expresa como una relación del rendimiento de las partes cosechables, tal como las semillas, dividido por la biomasa total; y f) mayor número de panículas primarias, (g) mayor peso de mil granos (PMG), el cual se extrapola a partir del recuento de la cantidad de semillas llenas y su peso total. Un mayor PMG puede ser el resultado de un mayor tamaño de la semilla y/o del peso de la semilla, y también puede ser el resultado de un incremento en el tamaño del embrión y/o del endospermo.

Un incremento en el rendimiento de semillas puede manifestarse también como un incremento en el tamaño y/o el volumen de la semilla. Además, un incremento en el rendimiento puede manifestarse también como un incremento en el área de la semilla y/o de la longitud de la semilla y/o del ancho de la semilla y/o del perímetro de la semilla. Un mayor rendimiento puede dar como resultado también una arquitectura modificada, o puede ocurrir debido a una arquitectura modificada.

Índice de Verdor

El "índice de verdor" tal como se utiliza aquí descripción se calcula a partir de las imágenes digitales de las plantas. Para cada píxel perteneciente a la planta objetivo sobre la imagen, se calcula la relación del valor verde versus el valor rojo (en el modelo RGB para codificación del color). El índice de verdor se expresa como el porcentaje de píxeles para los cuales la relación verde con respecto a rojo excede un umbral dado. Bajo condiciones normales de desarrollo, bajo condiciones de desarrollo con estrés por salinidad, y bajo condiciones de desarrollo con una disponibilidad reducida de nutrientes, el índice de verdor de la planta se mide en la última imagen antes del florecimiento. Por el contrario, bajo condiciones de desarrollo con estrés por sequía, el índice de verdor de las plantas se mide en la primera imagen después de la sequía.

Planta

El término "planta" tal como se utiliza aquí, abarca plantas enteras, ancestros y progenie de las plantas y partes de la planta, que incluyen semillas, brotes, tallos, hojas, raíces (incluyendo tubérculos), flores, y tejidos y órganos, en donde cada uno de los anteriormente mencionados incluye la secuencia del gen/ácido nucleico de interés. El término "planta" abarca además células de la planta, cultivos en suspensión, tejido del callo, embriones, regiones meristemáticas, gametofitos, esporofitos, polen y microesporas, nuevamente en donde cada uno de los mencionados anteriormente incluye la secuencia del gen/ácido nucleico de interés.

Las plantas que son particularmente útiles en los métodos de la invención incluyen a todas las plantas que

pertenezcan a la superfamilia Viridiplantae, en particular plantas monocotiledóneas y dicotiledóneas incluyendo forraje y leguminosas forrajeras, plantas ornamentales, cultivos alimenticios, árboles o arbustos seleccionados de la lista que comprende *Acer* spp., *Actinidia* spp., *Abelmoschus* spp., *Agave sisalana*, *Agropyron* spp., *Agrostis stolonifera*, *Allium* spp., *Amaranthus* spp., *Ammophila arenaria*, *Ananas comosus*, *Annona* spp., *Apium graveolens*, *Arachis* spp., *Anfocarpus* spp., *Asparagus officinalis*, *Avena* spp. (por ejemplo, *Avena sativa*, *Avena fatua*, *Avena byzantina*, *Avena fatua* var. *sativa*, *Avena hybrida*), *Averrhoa carambola*, *Bambusa* sp., *Benincasa hispida*, *Bertholletia excelsa*, *Beta vulgaris*, *Brassica* spp. (por ejemplo, *Brassica napus*, y *Brassica rapa* spp. [canola, colza, nabo]), *Cadaba farinosa*, *Camellia sinensis*, *Canna indica*, *Cannabis sativa*, *Capsicum* spp., *Carex elata*, *Carica papaya*, *Carissa macrocarpa*, *Carya* spp., *Carthamus tinctorius*, *Castanea* spp., *Ceiba pentandra*, *Cichorium endivia*, *Cinnamomum* spp., *Citrullus lanatus*, *Citrus* spp., *Cocos* spp., *Coffea* spp., *Colocasia esculenta*, *Cola* spp., *Corchorus* sp., *Coriandrum sativum*, *Corylus* spp., *Crataegus* spp., *Crocus sativus*, *Cucurbita* spp., *Cucumis* spp., *Cynara* spp., *Daucus carota*, *Desmodium* spp., *Dimocarpus longan*, *Dioscorea* spp., *Diospyros* spp., *Echinochloa* spp., *Elaeis* (por ejemplo, *Elaeis guineensis*, *Elaeis oleifera*), *Eleusine coracana*, *Erianthus* sp., *Eriobotrya japonica*, *Eucalyptus* sp., *Eugenia uniflora*, *Fagopyrum* spp., *Fagus* spp., *Festuca arundinacea*, *Ficus* *Carica*, *Fortunella* spp., *Fragaria* spp., *Ginkgo biloba*, *Glycine* spp. (por ejemplo, *Glycine max*, *Soja hispida* o *Soja max*), *Gossypium hirsutum*, *Helianthus* spp. (por ejemplo, *Helianthus annuus*), *Hemerocallis fulva*, *Hibiscus* spp., *Hordeum* spp. (por ejemplo, *Hordeum vulgare*), *Ipomoea baratas*, *Juglans* spp., *Lactuca sativa*, *Lathyrus* spp., *Lens culinans*, *Linum usitatissimum*, *Litchi chinensis*, *Lotus* spp., *Luffa acutangula*, *Lupinus* spp., *Luzula sylvatica*, *Lycopersicon* spp. (por ejemplo, *Lycopersicon esculentum*, *Lycopersicon lycopersicum*, *Lycopersicon pyriforme*), *Macrotyloma* spp., *Morus* spp., *Malpighia emarginata*, *Mammea americana*, *Mangifera indica*, *Manihot* spp., *Manilkara zapota*, *Medicago sativa*, *Melilotus* spp., *Mentha* spp., *Miscanthus sinensis*, *Momordica* spp., *Morus nigra*, *Musa* spp., *Nicotiana* spp., *Olea* spp., *Opuntia* spp., *Ornithopus* spp., *Oryza* spp. (por ejemplo, *Oryza sativa*, *Oryza latifolia*), *Panicum miliaceum*, *Panicum virgatum*, *Passiflora edulis*, *Pastinaca sativa*, *Pennisetum* sp., *Persea* spp., *Petroselinum crispum*, *Phalaris arundinacea*, *Phaseolus* spp., *Phleum pratense*, *Phoenix* spp., *Phragmites australis*, *Physalis* spp., *Pinus* spp., *Pistacia vera*, *Pisum* spp., *Poa* spp., *Populus* spp., *Prosopis* spp., *Prunus* spp., *Psidium* spp., *Punica granatum*, *Pyrus communis*, *Quercus* spp., *Raphanus sativus*, *Rheum rhabarbarum*, *Ribes* spp., *Ricinus communis*, *Rubus* spp., *Saccharum* spp., *Salix* sp., *Sambucus* spp., *Secale cereale*, *Sesamum* spp., *Sinapis* sp., *Solanum* spp. (por ejemplo, *Solanum tuberosum*, *Solanum integrifolium* o *Solanum lycopersicum*), *Sorgo bicolor*, *Spinacia* spp., *Syzygium* spp., *Tagetes* spp., *Tamarindus indica*, *Theobroma cacao*, *Trifolium* spp., *Triticosecale rimpai*, *Triticum* spp. (por ejemplo, *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybemum*, *Triticum macha*, *Triticum sativum* o *Triticum vulgare*), *Tropaeolum minus*, *Tropaeolum majus*, *Vaccinium* spp., *Vicia* spp., *Vigna* spp., *Viola odorata*, *Vitis* spp., *Zea mays*, *Zizania palustris*, *Ziziphus* spp., entre otras.

Descripción detallada de la invención

Sorprendentemente, se ha encontrado ahora que al incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT como se define aquí, produce plantas que tienen más características relacionadas con el rendimiento con relación a las plantas de control. De acuerdo con una primera forma de realización, la presente invención provee un método para incrementar las características relacionadas con el rendimiento en plantas con relación a las plantas que sirven como control, que comprende incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT, siendo dichas características relacionadas con el rendimiento vigor temprano, biomasa, y/o rendimiento de semillas.

Un método preferido para incrementar la expresión de una secuencia de ácido nucleico que codifica un polipéptido AMT es mediante la introducción y expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT.

Cualquier referencia en adelante a una "proteína útil en los métodos de la invención" se entiende que significa un polipéptido AMT como se define aquí. Cualquier referencia que se haga en adelante a una "secuencia de ácido nucleico útil en los métodos de la invención" se entiende que significa una secuencia de ácido nucleico capaz de codificar dicho polipéptido AMT. La secuencia de ácido nucleico que va a ser introducida en una planta (y por lo tanto útil para la realización de los métodos de la invención) es cualquier secuencia de ácido nucleico que codifica el tipo de polipéptido, que será descrito ahora, en adelante también denominado "secuencia de ácido nucleico para el AMT" o "gen AMT".

Un "polipéptido AMT" como se define aquí se refiere a cualquier polipéptido que contiene un dominio que tiene en orden creciente de preferencia al menos 80%, 85%, 90%, 95%, 98%, 99% o más identidad de secuencia de aminoácidos con un Dominio Conservado (DC) como se representa mediante la SEQ ID NO: 33.

Un "polipéptido AMT" como se describe aquí se refiere a cualquier polipéptido que comprende: (i) un dominio transportador de amonio con un acceso InterPro IPR0001905; y (ii) al menos 10 hélices que abarcan toda la membrana.

Un "polipéptido AMT" como se describe aquí se refiere a cualquier secuencia de polipéptidos que cuando se utiliza en la construcción de un árbol filogenético AMT, tal como aquel descrito en la Figura 2, se agrupa con el clado de polipéptidos AMT que comprende la secuencia de polipéptidos como la reportada por la SEQ ID NO: 2 (encerrada en un círculo en la Figura 2) en vez de con cualquier otro clado AMT. Adicionalmente, un "polipéptido AMT" como se

define aquí se refiere a cualquier polipéptido que contiene en orden creciente de preferencia al menos 80%, 85%, 90%, 95%, 98%, 99% o más identidad de secuencia de aminoácidos con el polipéptido AMT como el representado por la SEQ ID NO: 2.

- 5 Un "polipéptido AMT" como se describe aquí es capaz de complementar una cepa de levadura MLY131 que carece de todos los tres transportadores de amonio de levadura nativa (Hildebrand (2005) J Phycol 41: 105 - 113).

Los términos "dominio" y "motivo" se definen aquí en la sección de "definiciones". Existen bases de datos especializadas para la identificación de dominios, por ejemplo, SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. 95, 5857 - 5864; Letunic et al. (2002) Nucleic Acids Res 30, 242 - 244, InterPro (Mulder et al., (2003) Nucl. Acids. Res. 31, 315 - 318), Prosite (Bucher y Bairoch (1994), A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation. (In) ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., páginas 53 - 61, AAAIPress, Menlo Park; Hulo et al., Nucl. Acids. Res. 32: D134 - D137, (2004)), o Pfam (Bateman et al., Nucleic Acids Research 30(1): 276 - 280 (2002)). Un conjunto de herramientas para el análisis *in silico* de secuencias de proteína está disponible en el servidor de proteómicos ExPASy (Instituto Suizo de Bioinformática (Gasteiger et al., ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res. 31: 3784 - 3788 (2003)). El análisis de la secuencias de polipéptidos de la SEQ ID NO: 2 se presenta más abajo aquí en el Ejemplo 4. Por ejemplo, un polipéptido AMT como se representa mediante la SEQ ID NO: 2 comprende un dominio transportador de amonio con un número de acceso InterPro, IPR0001905. Los dominios se pueden identificar también utilizando técnicas de rutina, tales como mediante la alineación de secuencias. Una alineación de los polipéptidos de la Tabla A aquí, se muestra en la Figura 3. Tales alineaciones son útiles para identificar los dominios más conservados entre los polipéptidos AMT, tales como el Dominio Conservado (DC) como el representado por medio de la SEQ ID NO: 33 (comprendida en la SEQ ID NO: 2).

Los métodos para el alineamiento de secuencias para comparación son bien conocidos en la técnica; tales métodos incluyen GAP, BESTFIT, BLAST, FASTA y TFASTA. GAP utiliza el algoritmo de Needleman & Wunsch ((1970) J Mol Biol 48: 443 - 453) para hallar el alineamiento global (es decir, expansión de las secuencias completas) de dos secuencias que maximizan la cantidad de coincidencias y minimizan el número de brechas. El algoritmo BLAST (Altschul et al. (1990) J Mol Biol 215: 403 - 10) calcula el porcentaje de identidad de secuencia y realiza un análisis estadístico de la similitud entre las dos secuencias. El software para realizar el análisis con BLAST está disponible al público a través del Centro Nacional para la Información Biotecnológica (NCBI por sus siglas en inglés). Los homólogos se pueden identificar fácilmente utilizando, por ejemplo, el algoritmo de alineación de secuencias múltiples ClustalW (versión 1.83), con los parámetros predeterminados de alineamiento por pares, y un método de asignación de puntajes en porcentaje. Los porcentajes globales de similitud y de identidad también se pueden determinar utilizando uno de los métodos disponibles en el paquete de software MatGAT (Campanella et al., BMC Bioinformatics, 10: 29. MatGAT: una aplicación que genera matrices de similitud/identidad utilizando secuencias de proteínas o de ADN). Se puede realizar una edición manual menor para optimizar el alineamiento entre los motivos conservados, como sería evidente para una persona normalmente capacitada en la técnica. Además, en lugar de utilizar secuencias de longitud completa para la identificación de homólogos, se pueden utilizar también dominios específicos. Los valores de identidad de secuencia se pueden determinar sobre la secuencia completa del ácido nucleico o la secuencia del polipéptido o sobre dominios seleccionados o el(los) motivo(s) conservado(s), utilizando los programas mencionados anteriormente que utilizan los parámetros predeterminados. El Ejemplo 3 describe aquí en la Tabla B el porcentaje de identidad entre el polipéptido AMT como el representado mediante la SEQ ID NO: 2 y los polipéptidos AMT enumerados en la Tabla A, el cual puede tener una identidad de secuencia de aminoácidos tan baja como del 44%. El porcentaje de identidad puede ser incrementado si el cálculo de identidad se realiza entre el Dominio Conservado (DC) como se representa por la SEQ ID NO: 33 (comprendida en la SEQ ID NO: 2 y en la SEQ ID NO: 4) y el Dominio Conservado de los polipéptidos AMT de la Tabla A y representado en la Figura 3. Los resultados de tales cálculos se presentan en la Tabla B1 de la presente solicitud.

La tarea de la predicción de la localización de la proteína subcelular es importante y bien estudiada. Conocer la localización de la proteína ayuda a elucidar su función. Los métodos experimentales para la localización de la proteína varían desde inmunolocalización hasta marcación de proteínas utilizando proteína fluorescente verde (GFP por sus siglas en inglés) o beta-glucuronidasa (GUS). Dichos métodos son precisos si bien son muy laboriosos en comparación con los métodos computacionales. Recientemente se ha realizado un avance importante en la predicción computacional de la localización de la proteína a partir de los datos de la secuencia. Entre los algoritmos bien conocidos por una persona normalmente capacitada en la técnica se encuentran disponibles en las herramientas ExPASy Proteomics de propiedad del Swiss Institute for Bioinformatics, por ejemplo, PSort, TargetP, ChloroP, LocTree, Predotar, LipoP, MITOPROT, PATS, PTS1, SignalP, TMHMM y otras. La predicción de la localización subcelular de un polipéptido AMT como el representado por la SEQ ID NO: 2 se describe en el Ejemplo 5 de la presente solicitud.

Además, los polipéptidos AMT útiles en los métodos de la presente invención (al menos en su forma nativa) típicamente son capaces de transportar amonio a través de las membranas. Existen muchos ensayos para medir tal actividad de captación, incluidos ensayos de complementación de una cepa de levadura con transportadores defectuosos de amonio endógeno (Ninneman et al. (1994) EMBO J 13: 3464-3471), ensayos de captación en

levadura, *Xenopus oocytes* (Ludewig et al. (2003) J Biol Chem 278: 45603-45610), células de la planta, raíces de la planta (Yuan et al. (2007) Plant Phys 143: 732-744), y plantas enteras (Hoque et al. (2006) Functional Plant Biology 33: 153-163).

La presente invención se ilustra mediante la transformación de plantas con la secuencia de ácido nucleico representada por la SEQ ID NO: 3 (comprendida en la SEQ ID NO: 1), que codifica la secuencia del polipéptido AMT de la SEQ ID NO: 4 (comprendida en la SEQ ID NO: 2). Sin embargo, la realización de la invención no se restringe a estas secuencias; los métodos de la invención se pueden realizar en forma conveniente utilizando cualquier secuencia de ácido nucleico que codifique un polipéptido AMT como se define aquí.

Ejemplos de secuencias de ácido nucleico que codifican polipéptidos AMT se exponen en la Tabla A del Ejemplo 1 aquí. Tales secuencias de ácido nucleico son útiles para llevar a cabo los métodos de la invención. Las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1 son ejemplos de secuencias de ortólogos y parálogos del polipéptido AMT representadas por la SEQ ID NO: 2, o por la SEQ ID NO: 4, siendo los términos "ortólogos" y "parálogos" como se los define aquí. Otros ortólogos y parálogos pueden ser fácilmente identificados llevando a cabo una búsqueda denominada blast recíproco. Típicamente, esto implica un primer BLAST que involucra el someter una secuencia de consulta a BLAST (por ejemplo utilizando cualquiera de las secuencias enlistadas en la Tabla A del Ejemplo 1) contra cualquier base de datos de secuencias, tal como la base de datos NCBI disponible al público. BLASTN o TBLASTX (que utilizan valores estándar predeterminados) se utilizan generalmente cuando se parte de una secuencia nucleotídica, y BLASTP o TBLASTN (que utilizan valores estándar predeterminados) cuando se parte de una secuencia de proteína. Los resultados de BLAST opcionalmente se pueden filtrar. Las secuencias de longitud completa ya sea de los resultados filtrados o de los resultados no filtrados, se someten luego nuevamente a BLAST (segundo BLAST) contra las secuencias del organismo a partir del cual se deriva la secuencia de consulta (donde la secuencia de consulta es la SEQ ID NO: 1 o la SEQ ID NO: 2, el segundo BLAST sería por lo tanto contra las secuencias de *Phaeodactylum tricomutum*). Se comparan luego los resultados del primero y el segundo BLAST. Se identifica un parólogo si una respuesta pertinente de alto rango del primer blast es de la misma especie que aquella de la cual se deriva la secuencia de consulta, un BLAST posterior resulta idealmente luego en la secuencia de consulta como una respuesta pertinente más alta; se identifica un ortólogo si una respuesta pertinente de alto rango en el primer BLAST no proviene de la misma especie que aquella de la cual se deriva la secuencia de consulta, y preferiblemente resulta del BLAST posterior estando la secuencia de consulta entre las respuestas pertinentes más altas.

Las respuestas pertinentes de alto rango son aquellas que tienen un valor E bajo. Cuanto más bajo es el valor E, más significativo es el puntaje (o en otras palabras, es más baja la posibilidad de que la respuesta pertinente sea hallada por casualidad). La computación del valor E es bien conocida en la técnica. Además de los valores E, las comparaciones también se les da puntaje por identidad porcentual. La identidad porcentual se refiere a la cantidad de nucleótidos idénticos (o aminoácidos) entre las dos secuencias de ácidos nucleicos (o polipéptidos) comparadas sobre una longitud particular. En el caso de grandes familias, se puede utilizar ClustalW, seguido de un árbol filogenético (*neighbour joining tree*), para ayudar a visualizar el agrupamiento de genes relacionados y para identificar ortólogos y parálogos.

Las variantes de ácido nucleico también pueden ser útiles para llevar a la práctica los métodos de la invención. Los ejemplos de tales variantes incluyen secuencias de ácidos nucleicos que codifican homólogos y derivados de cualquiera de las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1, siendo los términos "homólogo" y "derivado" como se definieron aquí. También son útiles en los métodos de la invención las secuencias de ácido nucleico que codifican homólogos y derivados de ortólogos o parálogos de cualquiera de las secuencias de polipéptidos que se exponen en la Tabla A del Ejemplo 1. Los homólogos y derivados útiles en los métodos de la presente invención tienen sustancialmente la misma actividad biológica y funcional que la proteína no modificada de la cual se derivan.

Otras variantes de ácido nucleico útiles en la práctica de los métodos de la invención incluyen porciones de secuencias de ácido nucleico que codifican polipéptidos AMT, secuencias de ácido nucleico que hibridan a secuencias de ácido nucleico que codifican polipéptidos AMT, variantes de empalme de secuencias de ácido nucleico que codifican polipéptidos AMT, variantes alélicas de secuencias de ácido nucleico que codifican polipéptidos AMT y variantes de secuencias de ácido nucleico que codifican polipéptidos AMT obtenidos por medio de transposición de genes. Los términos secuencias de hibridación, variante de empalme, variante alélica y transposición de genes son como se describe aquí.

Las secuencias de ácido nucleico que codifican polipéptidos AMT no necesitan ser secuencias de ácido nucleico de longitud completa, ya que el desempeño de los métodos de la invención no confía en el uso de secuencias de ácido nucleico de longitud completa. De acuerdo con la presente invención, se provee un método para aumentar las características relacionados con el rendimiento, en las plantas, que comprende la introducción y la expresión en una planta de una porción de cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o una porción de una secuencia de ácido nucleico que codifica un ortólogo, parólogo u homólogo de cualquiera de las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1, en donde dicha porción es una porción de la SEQ ID NO: 1 que tiene al menos 1380 o más nucleótidos de longitud.

Una porción de una secuencia de ácido nucleico se puede preparar, por ejemplo, haciendo una o más supresiones a la secuencia de ácido nucleico. Las porciones se pueden utilizar en forma aislada o se pueden fusionar a otras secuencias codificadoras (o no codificadoras) con el propósito de producir, por ejemplo, una proteína que combina diferentes actividades. Cuando se fusionan a otras secuencias codificadoras, el polipéptido resultante producido después de la traducción puede ser más grande que el predicho para la porción de proteína.

Las porciones útiles en los métodos de la invención, codifican un polipéptido AMT como se define aquí, y tienen sustancialmente la misma actividad biológica que las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1. Preferiblemente, la porción es una porción de cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o es una porción de una secuencia de ácido nucleico que codifica un ortólogo o un parólogo de cualquiera de las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1. Preferiblemente la porción es, en orden creciente de preferencia al menos de 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1380 o más nucleótidos consecutivos de longitud, siendo los nucleótidos consecutivos de cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o de una secuencia de ácido nucleico que codifica un ortólogo o un parólogo de cualquiera de las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1. Preferiblemente, la porción es una porción de una secuencia nucleica que codifica un polipéptido de una secuencia de polipéptidos que comprende un dominio que tiene en orden creciente de preferencia al menos una identidad de secuencia de aminoácidos del 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% o más con un Dominio Conservado (DC) como el representado por la SEQ ID NO: 33. Más preferiblemente, la porción es una porción de la secuencia de ácido nucleico de la SEQ ID NO: 1. Lo más preferible, la porción es como la representada por la SEQ ID NO: 3

Otra variante de la secuencia de ácido nucleico útil en los métodos de la invención es una secuencia de ácido nucleico capaz de hibridar, en condiciones de rigurosidad reducida, preferiblemente bajo condiciones rigurosas, con una secuencia de ácido nucleico que codifica un polipéptido AMT como se define aquí, o con una porción como se define aquí.

De acuerdo con la presente invención, se provee un método para incrementar las características relacionadas con el rendimiento en plantas, que comprende la introducción y expresión en una planta de una secuencia de ácido nucleico capaz de hibridar con cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o que comprende la introducción y expresión en una planta de una secuencia de ácido nucleico capaz de hibridar con una secuencia de ácido nucleico que codifica un ortólogo, parólogo u homólogo de cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, en donde dicha secuencia de ácido nucleico es capaz de hibridar con el ácido nucleico de la SEQ ID NO: 1.

Las secuencias de hibridación útiles en los métodos de la invención codifican un polipéptido AMT como se define aquí, y tienen sustancialmente la misma actividad biológica que las secuencias polipeptídicas presentadas en la Tabla A del Ejemplo 1. Preferiblemente, la secuencia de hibridación es capaz de hibridar con cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o a una porción de cualquiera de estas secuencias, siendo una porción como se definió anteriormente, o donde la secuencia de hibridación es capaz de hibridar con una secuencia de ácido nucleico que codifica un ortólogo o parólogo de una cualquiera de las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1. Preferiblemente, la secuencia de hibridación es capaz de hibridar con una secuencia de ácido nucleico que codifica una secuencia de polipéptidos que comprende un dominio que tiene en orden creciente de preferencia al menos una identidad de secuencia de aminoácidos del 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% o más con un Dominio Conservado (DC) como se representa mediante la SEQ ID NO: 33. Más preferiblemente, la secuencia de hibridación es capaz de hibridar con una secuencia de ácido nucleico como la representada mediante la SEQ ID NO: 1 o una porción de la misma.

Otra variante de la secuencia de ácido nucleico útil en los métodos de la invención es una variante de empalme que codifica un polipéptido AMT como se definió aquí anteriormente, siendo una variante de empalme como se definió aquí.

De acuerdo con la presente invención, se divulga un método para incrementar las características relacionadas con el rendimiento, que comprende introducir y expresar en una planta una variante de empalme de cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o una variante de empalme de una secuencia de ácido nucleico que codifica un ortólogo, parólogo u homólogo de cualquiera de las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1.

Las variantes de empalme preferidas son variantes de empalme de una secuencia de ácido nucleico representada por la SEQ ID NO: 1, o una variante de empalme de una secuencia de ácido nucleico que codifica un ortólogo o un parólogo de la SEQ ID NO: 2. Preferiblemente, la variante de empalme es una variante de empalme de una secuencia de ácido nucleico que codifica una secuencia de polipéptidos que comprende un dominio que tiene en orden creciente de preferencia de al menos una identidad de secuencia de aminoácidos del 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% o más con un Dominio Conservado (DC) como se representa mediante la SEQ ID NO: 33.

Otra variante de secuencia de ácido nucleico útil en la realización de los métodos de la invención es una variante alélica de una secuencia de ácido nucleico que codifica un polipéptido AMT como se definió aquí anteriormente, siendo una variante alélica como la definida aquí.

De acuerdo con la presente invención, se provee un método para incrementar las características relacionadas con el rendimiento, que comprende introducir y expresar en una planta una variante alélica de cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o que comprende introducir y expresar en una planta una variante alélica de una secuencia de ácido nucleico que codifica un ortólogo, parálogo u homólogo de cualquiera de las secuencias polipeptídicas presentadas en la Tabla A del Ejemplo 1.

Las variantes alélicas útiles en los métodos de la presente invención tienen sustancialmente la misma actividad biológica que el polipéptido AMT de la SEQ ID NO: 2 y cualquiera de las secuencias polipeptídicas descritas en la Tabla A del Ejemplo 1. Existen variantes alélicas en la naturaleza, y comprendido dentro de los métodos de la presente invención está el uso de estos alelos naturales. Preferiblemente, la variante alélica es una variante alélica de la SEQ ID NO: 1 o una variante alélica de una secuencia de ácido nucleico que codifica un ortólogo o un parálogo de la SEQ ID NO: 2. Preferiblemente, la variante alélica es una variante alélica de una secuencia de polipéptidos que comprende un dominio que tiene en orden creciente de preferencia al menos una identidad de secuencia de aminoácidos del 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% o más con un Domino Conservado (DC) como se representa mediante la SEQ ID NO: 33

La transposición de genes o la evolución dirigida también se puede utilizar para generar variantes de secuencias de ácido nucleico que codifican polipéptidos AMT como se definió anteriormente, siendo el término "transposición de genes" como se definió aquí.

De acuerdo con la presente invención, se divulga un método para incrementar las características relacionadas con el rendimiento, que comprende introducir y expresar en una planta una variante de cualquiera de las secuencias de ácido nucleico presentadas en la Tabla A del Ejemplo 1, o que comprende introducir y expresar en una planta una variante de una secuencia de ácido nucleico que codifica un ortólogo, parálogo u homólogo de cualquiera de las secuencias de polipéptidos presentadas en la Tabla A del Ejemplo 1, en donde la variante de la secuencia de ácido nucleico se obtiene por medio de transposición de genes.

Preferiblemente, la variante de la secuencia de ácido nucleico obtenida por transposición de genes codifica una secuencia de polipéptidos que comprende un dominio que tiene en orden creciente de preferencia al menos una identidad de secuencia de aminoácidos del 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% o más con un Domino Conservado (DC) como el representado mediante la SEQ ID NO: 33.

Además, también se pueden obtener variantes de la secuencia de ácido nucleico por mutagénesis dirigida al sitio. Varios métodos se encuentran disponibles para lograr una mutagénesis dirigida al sitio, siendo los más comunes los métodos basados en PCR (Current Protocols in Molecular Biology. Wiley Eds.).

Las secuencias de ácido nucleico que codifican polipéptidos AMT pueden derivarse de cualquier fuente natural o artificial. La secuencia de ácido nucleico puede ser modificada a partir de su forma nativa en composición y/o ambiente genómico a través de manipulación humana deliberada. La secuencia de ácido nucleico que codifica un polipéptido AMT es del dominio eucariota, preferiblemente del reino Cromalveolata, preferiblemente además de la Heterocontophyta phylum. Más preferiblemente, la secuencia de ácido nucleico que codifica un polipéptido AMT es de la clase Bacillariophyceae (diatomeas), y por ejemplo, de los siguientes ordenes: Achnanthales, Bacillariales, Centrales (tales como Thalassiosira pseudonana), Cymbellales, Eunotiales, Mastogloiales, Naviculales, Pennales (tales como Pheodactylum tricornutum), Rhopalodiales, Surirellales, o Thalassiosiphysales. Lo más preferible, la secuencia de ácido nucleico codifica un polipéptido AMT es de Pheodactylum tricornutum.

El desempeño de los métodos de la invención producen plantas que tienen un incremento de características relacionadas con el rendimiento con relación a las plantas que sirven de control. Los términos "rendimiento" y "rendimiento de semilla" se describen con más detalle en la sección de "definiciones" aquí.

Tomando el maíz como ejemplo, un incremento de rendimiento se puede manifestar como uno o más de lo siguiente: aumento en el número de plantas establecidas por hectárea o acre, un aumento en el número de mazorcas por planta, un aumento en el número de filas, en el número de granos por fila, peso del grano, peso de mil granos, longitud/diámetro de la mazorca, aumento en la tasa de llenado de la semilla (que es el número de semillas llenas dividido por el número total de semillas y multiplicado por 100), entre otros. Tomando el arroz como ejemplo un aumento en el rendimiento se puede manifestar en sí mismo como un aumento en uno o más de lo siguiente: número de plantas por hectárea o acre, número de panículas por planta, número de espiguillas por panícula, número de flores (floretes) por panícula, que se expresa como una relación del número de semillas llenas sobre el número de panículas primarias), aumento en la tasa de llenado de semilla (que es el número de semillas llenas dividido por el número total de semillas y multiplicado por 100), aumento en el peso de mil granos, entre otros.

La presente invención proporciona un método para incrementar las características relacionadas con el crecimiento

de las plantas con relación a las plantas que sirven de control, cuyo método comprende incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT como se definió aquí, en donde las características relacionadas con el rendimiento son vigor temprano, biomasa, y/o rendimiento de semillas.

- 5 Ya que las plantas transgénicas de acuerdo con la presente invención tienen un aumento de características relacionadas con el rendimiento, es probable que estas plantas exhiban una mayor tasa de crecimiento (durante al menos parte de su ciclo de vida), con relación a la tasa de crecimiento de las plantas que sirven como control en una etapa correspondiente en su ciclo de vida.
- 10 La mayor tasa de crecimiento puede ser específica para una o más partes de una planta (incluyendo las semillas), o puede ser sustancialmente a través de la planta completa. Las plantas con una mayor velocidad de crecimiento pueden tener un ciclo vital más corto. El ciclo de vida de una planta puede entenderse como el tiempo necesario para desarrollarse desde una semilla madura seca hasta la etapa donde la planta ha producido semillas maduras secas, en forma similar al material de partida. Este ciclo de vida puede estar influido por factores tales como el vigor
- 15 temprano, la velocidad de crecimiento, el índice de verdor, el tiempo de floración y la velocidad de maduración de la semilla. El incremento en la velocidad del crecimiento puede tener lugar en una o más etapas en el ciclo de vida de una planta o sustancialmente durante todo el ciclo de vida de la planta. La mayor velocidad de crecimiento durante las etapas tempranas en el ciclo de vida de una planta puede reflejar un mayor vigor (temprano). El incremento en la velocidad de crecimiento puede alterar el ciclo de cosecha de una planta permitiendo que las plantas sean sembradas más tarde y/o cosechadas más pronto de lo que sería posible de otro modo (se puede obtener un efecto similar con un tiempo de floración más temprano; la demora en la floración no es usualmente una característica deseable en los cultivos). Si la velocidad del crecimiento se incrementa lo suficiente, puede permitir el sembrado posterior de semillas de la misma especie de planta (por ejemplo el sembrado y cosecha de plantas de arroz seguido por el sembrado y cosecha de otras plantas de arroz todo dentro de un periodo de crecimiento
- 20 convencional). De manera similar, si se incrementa suficientemente la velocidad de crecimiento, permitirá el posterior sembrado de diferentes especies de planta (por ejemplo sembrado y cosecha de plantas de maíz, seguido por ejemplo, del sembrado y de cosecha opcional de soja, patata o cualquier otra planta adecuada). También se puede cosechar varias veces a partir del mismo rizoma en el caso de algunas plantas de cultivo. La alteración del ciclo de cosecha de una planta puede conducir a un incremento en la producción anual de biomasa por acre (debido a un incremento en el número de veces (digamos en un año) que cualquier planta particular puede ser cultivada y cosechada). Un incremento en la velocidad de crecimiento también puede permitir el cultivo de plantas transgénicas en un área geográfica más amplia que sus contrapartes de tipo silvestre, dado que las limitaciones territoriales para el crecimiento de un cultivo a menudo están determinadas por condiciones ambientales adversas ya sea en el momento de la plantación (a principios de la estación) o en el momento de la cosecha (finales de la estación). Estas
- 25 condiciones adversas se pueden evitar si se acorta el ciclo de la cosecha. La velocidad de crecimiento puede determinarse por medio de la derivación de diferentes parámetros a partir de las curvas de crecimiento; dichos parámetros pueden ser: T-Mid (el tiempo que les toma a las plantas alcanzar 50% de su tamaño máximo) y T-90 (el tiempo que les toma a las plantas alcanzar 90% de su tamaño máximo), entre otros.
- 30 De acuerdo con una característica preferida de la presente invención, el desempeño de los métodos de la invención produce plantas que tienen una mayor tasa de crecimiento con respecto a las plantas de control. Por lo tanto, de acuerdo con la presente invención, se provee un método para incrementar la tasa de crecimiento de las plantas, en donde dicho método comprende incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT como se definió aquí.
- 35 Un incremento de las características relacionadas con el rendimiento se presenta si la planta está bajo condiciones no estresadas o si la planta es expuesta a varios tipos de estrés en comparación con las plantas control cultivadas bajo condiciones comparables. Las plantas típicamente responden a la exposición al estrés creciendo en forma más lenta. En condiciones de estrés severo, la planta puede incluso detener el crecimiento completamente. Un estrés moderado por otro lado se define aquí como cualquier estrés al cual se expone una planta que no produce como resultado que la planta cese su crecimiento completamente sin la capacidad de reanudarlo. Un estrés moderado en el sentido de la invención conduce a una reducción en el crecimiento de las plantas estresadas de menos del 40%, 35% o 30%, preferiblemente menos del 25%, 20% o 15%, más preferiblemente menos de 14%, 13%, 12%, 11% o 10% o menos en comparación con las plantas control bajo condiciones no estresantes. Debido a los avances en las
- 40 prácticas agrícolas (irrigación, fertilización, tratamientos con pesticidas) no se encuentran a menudo estreses severos en las plantas de cultivo sembradas. En consecuencia, el crecimiento comprometido inducido por un estrés moderado a menudo es una característica indeseable para la agricultura. Los estreses moderados son los estreses bióticos y/o abióticos (ambientales) de cada temporada a los cuales está expuesta una planta. Los estreses abióticos pueden deberse a sequía o exceso de agua, estrés anaeróbico, estrés por salinidad, toxicidad química, estrés oxidativo y temperaturas calurosas, frías o congelantes. El estrés abiótico puede ser un estrés osmótico causado por un estrés por agua (particularmente debido a sequía), estrés por salinidad, estrés oxidativo o un estrés iónico. Los estreses bióticos son típicamente aquellos estreses causados por patógenos, tales como bacterias, virus, hongos, nemátodos, e insectos. El término condiciones "no estresantes" como se usa aquí son aquellas condiciones ambientales que permiten el óptimo crecimiento de las plantas. Las personas normalmente capacitadas en el arte
- 45 son conscientes de las condiciones normales de suelo y las climáticas para una ubicación dada.
- 50
- 55
- 60
- 65

El desempeño de los métodos de la invención produce plantas cultivadas bajo condiciones no estresantes o bajo condiciones de estrés moderado que tienen más características relacionadas con el rendimiento, con respecto a plantas de control cultivadas bajo condiciones comparables. Por lo tanto, de acuerdo con la presente invención, se provee un método para incrementar las características relacionadas con el rendimiento en plantas cultivadas bajo condiciones no estresantes o bajo condiciones de estrés moderado, en donde el método comprende incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT.

El desempeño de los métodos de acuerdo con la presente invención da lugar a plantas cultivadas bajo condiciones de estrés abiótico que tienen más características relacionadas con el rendimiento con respecto a plantas control cultivadas bajo condiciones de estrés comparables. De acuerdo a lo reportado en Wang et al. (Planta (2003) 218: 1 - 14), un estrés abiótico conduce a una serie de cambios morfológicos, fisiológicos, bioquímicos y moleculares que afectan adversamente el crecimiento y la productividad de la planta. Se sabe que la sequía, salinidad, temperaturas extremas y estrés oxidativo están interconectados y pueden inducir daño celular y de crecimiento a través de mecanismos similares. Rabbani et al. (Plant Physiol (2003) 133: 1755 - 1767) describe un grado particularmente elevado de "interferencia" entre el estrés por sequía y el estrés por alta salinidad. Por ejemplo, la sequía y/o la salinización se manifiestan principalmente como estrés osmótico, produciendo la interrupción de la homeostasis y la distribución de iones en la célula. El estrés oxidativo, que con frecuencia acompaña a temperaturas altas o bajas, al estrés por salinidad o por sequía, puede causar desnaturalización de proteínas funcionales y estructurales. Como consecuencia, estos diversos estreses ambientales a menudo activan rutas de señalización celular y respuestas celulares similares, tales como la producción de proteínas por estrés, favorecimiento de la expresión de antioxidantes, acumulación de solutos compatibles y detención del crecimiento. Dado que diversos estreses ambientales activan rutas similares, el ejemplo de la presente invención con estrés por sequía no debe interpretarse como una limitación al estrés por sequía, sino más bien como una selección para indicar la participación de los polipéptidos AMT como se definió anteriormente, en el incremento de las características relacionadas con el rendimiento con respecto a plantas de control cultivadas bajo condiciones de estrés comparable, en estreses abióticos en general.

El término "estrés abiótico" como se define aquí se entiende que significa uno o más de: estrés por agua (causado por sequía o exceso de agua), estrés anaeróbico, estrés por salinidad, estrés por temperatura (causada por temperaturas elevadas, frías o congelantes), estrés por toxicidad química y estrés oxidativo. De acuerdo con un aspecto de la invención, el estrés abiótico es un estrés osmótico, seleccionado de estrés por agua, estrés por salinidad, estrés oxidativo y estrés iónico. Preferiblemente, el estrés por agua es estrés por sequía. El término estrés por salinidad no se restringe a la sal común (NaCl), sino que puede ser cualquier estrés causado por una o más de: NaCl, KCl, LiCl, MgCl₂, CaCl₂, entre otras.

El desempeño de los métodos de la invención produce plantas que tienen más características relacionadas con el rendimiento, bajo condiciones de estrés abiótico con respecto a las plantas de control cultivadas bajo condiciones de estrés comparables. Por lo tanto, de acuerdo con la presente invención, se provee un método para incrementar las características relacionadas con el rendimiento, en plantas cultivadas bajo condiciones de estrés abiótico, en donde el método comprende incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT. De acuerdo con un aspecto de la invención, el estrés abiótico es un estrés osmótico, seleccionado de uno o más de los siguientes: estrés por agua, estrés por salinidad, estrés oxidativo y estrés iónico.

Otro ejemplo de estrés ambiental abiótico es la disponibilidad reducida de uno o más nutrientes que deben ser asimilados por las plantas para su crecimiento y desarrollo. Debido a la fuerte influencia de la eficiencia de la utilización de la nutrición sobre el rendimiento de la planta y la calidad del producto, se vierte una enorme cantidad de fertilizante en los campos para optimizar el crecimiento y calidad de las plantas. La productividad de las plantas en general está limitada por tres nutrientes principales, fósforo, potasio y nitrógeno, donde usualmente entre estos tres se encuentra el elemento limitante de la velocidad de crecimiento de las plantas. Por lo tanto el elemento nutricional principal requerido para el crecimiento de las plantas es el nitrógeno (N). Es un constituyente de numerosos compuestos importantes hallados en las células vivientes, incluidos aminoácidos, proteínas (enzimas), ácidos nucleicos y clorofila. 1,5% a 2% del material seco de una planta es nitrógeno y aproximadamente 16% de la proteína total de la planta. Por lo tanto, la disponibilidad de nitrógeno es un factor limitante principal para el crecimiento y producción de las plantas de cultivo (Frink et al. (1999) Proc Natl Acad Sci USA 96(4): 1175 - 1180), y tiene también un impacto principal en la acumulación de proteínas y composición de los aminoácidos. Por lo tanto, son de gran interés las plantas de cultivo con más características relacionadas con el rendimiento, cuando se cultivan bajo condiciones limitantes de nitrógeno.

El desempeño de los métodos de la invención produce plantas que crecen bajo condiciones de disponibilidad reducida de nutrientes, particularmente bajo condiciones de disponibilidad reducida de nitrógeno, que tienen más características relacionadas con el rendimiento con respecto a las plantas de control cultivadas bajo condiciones de estrés comparables. Por lo tanto, de acuerdo con la presente invención, se provee un método para incrementar las características relacionadas con el rendimiento en plantas cultivadas bajo condiciones de disponibilidad reducida de nutrientes, preferiblemente disponibilidad reducida de nitrógeno, en donde el método comprende incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT. La disponibilidad reducida de nutrientes puede resultar de una deficiencia o exceso de nutrientes tales como nitrógeno, fosfatos y

otros compuestos que contienen fósforo, potasio, calcio, cadmio, magnesio, manganeso, hierro y boro, entre otros. Preferiblemente la disponibilidad reducida de nutrientes es disponibilidad reducida de nitrógeno.

- 5 La presente invención abarca plantas o partes de las mismas (incluidas semillas) o células de las mismas que pueden ser obtenidas mediante los métodos de acuerdo con la presente invención. Las plantas o partes de las mismas o células de las mismas comprenden un transgén de ácido nucleico que codifica un polipéptido AMT, como se definió anteriormente.
- 10 La invención también divulga construcciones y vectores genéticos para facilitar la introducción y/o mayor expresión en plantas de secuencias de ácido nucleico que codifican polipéptidos AMT como se define aquí. Las construcciones génicas se pueden insertar en los vectores, que pueden estar comercialmente disponibles, adecuados para la transformación en plantas y para la expresión del gen de interés en las células transformadas. La invención también provee el uso de una construcción génica como se define aquí en los métodos de la invención.
- 15 En forma más específica, la presente invención divulga una construcción que comprende:
- (a) una secuencia de ácido nucleico que codifica un polipéptido AMT como se definió anteriormente;
 - (b) una o más secuencias control capaces de incrementar la expresión de la secuencia del ácido nucleico de (a); y
 - 20 (c) una secuencia de terminación de la transcripción.
- Preferiblemente, la secuencia de ácido nucleico que codifica un polipéptido AMT es como se definió anteriormente. El término "secuencia control" y "secuencia de terminación" son como se definen aquí.
- 25 Preferiblemente, una de las secuencias control de una construcción es un promotor constitutivo aislado de un genoma de una planta. Un ejemplo de un promotor constitutivo es un promotor GOS2, preferiblemente un promotor GOS2 de arroz, más preferiblemente un promotor GOS2 como el representado mediante la SEQ ID NO: 34.
- 30 Las plantas se transforman con un vector que comprende cualquiera de las secuencias de ácido nucleico descritas anteriormente. La persona calificada en la técnica es muy consciente de los elementos genéticos que deben estar presentes en el vector a fin de transformar, seleccionar y propagar exitosamente las células huéspedes que contienen la secuencia de interés. La secuencia de interés está operativamente enlazada a una o más secuencias de control (al menos a un promotor).
- 35 Convenientemente, se puede utilizar cualquier tipo de promotor, ya sea natural o sintético, para incrementar la expresión de la secuencia de ácido nucleico. Un promotor constitutivo es particularmente útil en los métodos, preferiblemente un promotor constitutivo aislado de un genoma de una planta. El promotor constitutivo de una planta dirige la expresión de una secuencia codificadora en un nivel que está en todos los casos por debajo de aquel obtenido bajo el control de un promotor viral 35S del CaMV.
- 40 Otros promotores específicos del órgano, por ejemplo para la expresión preferida en hojas, tallos, tubérculos, meristemas, semillas (embrión y/o endospermo), son útiles para llevar a cabo los métodos de invención. Véase la sección "Definiciones" aquí para las definiciones de los diferentes tipos de promotores.
- 45 Debe aclararse que la aplicabilidad de la presente invención no está restringida a una secuencia de ácido nucleico que codifica el polipéptido AMT, como el representado mediante la SEQ ID NO: 1 o mediante la SEQ ID NO: 3, ni la aplicabilidad de la invención está restringida a la expresión de una secuencia de ácido nucleico que codifica el polipéptido AMT cuando está dirigida por un promotor constitutivo.
- 50 Opcionalmente, se pueden utilizar una o más secuencias terminadoras en la construcción introducida en una planta. Los elementos reguladores adicionales pueden incluir mejoradores de la transcripción así como también de traducción. Las personas normalmente capacitadas en la técnica serán conscientes de las secuencias terminadoras y mejoradores que puedan ser adecuadas para uso en la realización de la invención. También se puede añadir una secuencia de intrón a la región no traducida (UTR) 5' o en la secuencia de codificación para incrementar la cantidad
- 55 del mensaje maduro que se acumula en el citosol, como se describe en la sección de las definiciones. Otras secuencias control (además de las secuencias promotoras, mejoradoras, silenciadoras, del intrón, regiones 3'UTR y/o 5'UTR) pueden ser proteínas y/o elementos estabilizantes del ARN. Tales secuencias serían conocidas o pueden ser fácilmente obtenidas por una persona normalmente capacitada en la técnica.
- 60 Las construcciones genéticas divulgadas en la invención pueden incluir además un origen de secuencia de replicación que se requiere para el mantenimiento y/o replicación en un tipo de célula específico. Un ejemplo es cuando se requiere mantener una construcción genética en una célula bacteriana como un elemento genético episomal (por ejemplo, una molécula de plásmido o de cósmido). Los orígenes de replicación preferidos incluyen, pero no se limitan al f1-ori y colE1.
- 65 Para la detección de la transferencia exitosa de las secuencias de ácido nucleico tal como se utiliza en los métodos

de la invención y/o selección de planta transgénicas que comprende estas secuencias de ácido nucleico, es conveniente utilizar genes marcadores (o genes reporteros). Por lo tanto, la construcción genética puede comprender opcionalmente un gen marcador seleccionable. Los marcadores seleccionables se describen en más detalle en la sección "definiciones" aquí.

Se sabe que tras la integración estable o transitoria de secuencias de ácido nucleico en las células de la planta, sólo una minoría de las células incorpora el ADN foráneo y, si se desea, lo integra en su genoma, dependiendo del vector de expresión utilizado y de la técnica de transfección utilizada. Para identificar y seleccionar estos integrantes, un gen que codifica para un marcador seleccionable (tal como los descritos anteriormente) es usualmente introducido en las células huésped junto con el gen de interés. Estos marcadores, por ejemplo, pueden utilizarse en mutantes en los cuales estos genes no son funcionales, por ejemplo, por medio de la supresión mediante métodos convencionales. Además, se pueden introducir moléculas de secuencias de ácido nucleico que codifican un marcador seleccionable en una célula huésped sobre el mismo vector que comprende la secuencia que codifica los polipéptidos de la invención o que se utiliza en los métodos de la invención, o bien en un vector separado. Las células que han sido transfectadas en forma estable con la secuencia introducida de ácido nucleico se pueden identificar por ejemplo mediante selección (por ejemplo, las células que han integrado el marcador seleccionable sobreviven mientras que las otras células mueren). Los genes marcadores pueden ser removidos o escindidos de la célula transgénica una vez que no se los necesita. Las técnicas para remoción del gen marcador son conocidas en el arte; las técnicas útiles fueron descritas anteriormente en la sección de definiciones.

La invención también provee un método para la producción de plantas transgénicas que tienen más características relacionadas con el rendimiento con relación a las plantas que sirven como control, que comprende la introducción y expresión en una planta de cualquier secuencia de ácido nucleico que codifique un polipéptido ATM como se definió aquí anteriormente.

Más específicamente, la presente invención provee un método para la producción de plantas transgénicas que tienen más características relacionadas con el rendimiento con respecto a las plantas que sirven como control, el cual comprende:

- (i) la introducción y expresión en una planta, parte de una planta, o una célula de una planta, de una secuencia de ácido nucleico que codifica un polipéptido AMT como se definió aquí, bajo el control de un promotor constitutivo de una planta; y
- (ii) el cultivo la célula de la planta, parte de la planta o la planta en condiciones que promuevan el crecimiento y desarrollo de la planta.

La secuencia de ácido nucleico de (i) puede ser cualquiera de las secuencias de ácido nucleico capaces de codificar un polipéptido AMT como se definió aquí.

La secuencia de ácido nucleico se puede introducir directamente en una célula de una planta o dentro de la planta misma (incluyendo la introducción en un tejido, órgano o cualquier otra parte de una planta). De acuerdo con una característica preferida de la presente invención, se introduce preferiblemente la secuencia de ácido nucleico en una planta mediante transformación. El término "transformación" se describe en forma más detallada, en la sección "definiciones" en la presente descripción.

Las células de la planta genéticamente modificadas se pueden regenerar mediante todos los métodos con los cuales está familiarizado la persona capacitada en la técnica. Los métodos adecuados pueden encontrarse en las publicaciones antes mencionadas por S. D. Kung y R. Wu, Potrykus o Höfgen y Wilmitzer.

En general después de la transformación, las células de la planta o los agrupamientos celulares se seleccionan para determinar la presencia de uno o más marcadores que son codificados por genes que pueden expresarse en una planta transferidos junto con el gen de interés, después de lo cual se regenera el material transformado en una planta entera. Para seleccionar las plantas transformadas, el material vegetal obtenido en la transformación es sometido, como regla, a condiciones selectivas de modo que las plantas transformadas puedan distinguirse de las plantas no transformadas. Por ejemplo, las semillas obtenidas en la forma antes descrita se pueden plantar y, después de un período de crecimiento inicial, ser sometidas a una selección adecuada por medio de rociado. Otra posibilidad adicional consiste en el crecimiento de las semillas, si procede después de esterilización, sobre placas de agar utilizando un agente de selección adecuado de modo que únicamente las semillas transformadas pueden convertirse en plantas. En forma alternativa, se seleccionan las plantas transformadas para determinar la presencia de un marcador seleccionable tal como los descritos anteriormente.

Después de la transferencia y regeneración del ADN, las plantas putativamente transformadas también se pueden evaluar, por ejemplo utilizando el análisis tipo Southern, para determinar la presencia del gen de interés, el número de copias y/o organización genómica. En forma alternativa o adicional, se pueden monitorear los niveles de expresión del ADN recientemente introducido utilizando análisis tipo Northern y/o tipo Western, siendo ambas técnicas bien conocidas por las personas ordinariamente capacitadas en la técnica.

Las plantas transformadas generadas se pueden propagar mediante una variedad de medios, tales como por medio de técnicas de propagación clonal o de fitomejoramiento clásico. Por ejemplo, una primera generación de la planta transformada (o T1) puede ser autofecundada y los transformantes homocigotos de segunda generación (o T2) seleccionados, y las plantas T2 pueden ser luego adicionalmente propagadas a través de las técnicas clásicas de fitomejoramiento. Los organismos transformados generados pueden adoptar una variedad de formas. Por ejemplo, pueden ser quimeras de células transformadas y de células no transformadas; transformantes clonales (por ejemplo, todas las células transformadas para contener el casete de expresión); injertos de tejidos transformados y sin transformar (por ejemplo, en plantas, un rizoma transformado injertado a un vástago sin transformar).

La presente invención claramente se extiende a cualquier célula de una planta o planta producida mediante uno cualquiera de los métodos descritos aquí, y a todas las partes de la planta y propágulos de la misma. La presente invención se extiende además para abarcar la progenie de una célula primaria transformada o transfectada, tejido, órgano o planta completa que ha sido producida mediante cualquiera de los métodos antes mencionados, siendo el único requisito que la progenie exhiba la(s) misma(s) característica(s) fenotípica(s) y/o genotípica(s) que las producidas por la progenie en los métodos de acuerdo con la invención.

La invención también divulga células huéspedes que contienen una secuencia de ácido nucleico aislada que codifica un polipéptido AMT como se definió aquí anteriormente, operativamente enlazada a un promotor constitutivo. Las células huésped preferidas de acuerdo con la invención son las células de la planta. Las plantas huésped para las secuencias de ácido nucleico o el vector utilizado en el método de acuerdo con la invención, el casete de expresión o la construcción o el vector están, en principio, en forma conveniente en todas las plantas, las cuales son capaces de sintetizar los polipéptidos utilizados en el método de la invención.

Los métodos de la invención se aplican convenientemente a cualquier planta. Las plantas que son particularmente útiles en los métodos de la invención incluyen todas las plantas que pertenecen a la superfamilia *Viridiplantae*, en particular plantas monocotiledóneas y dicotiledóneas que incluyen forraje o leguminosas forrajeras, plantas ornamentales, cultivos alimenticios, árboles o arbustos. De acuerdo con una realización preferida de la presente invención, la planta es una planta de cultivo. Los ejemplos de plantas de cultivo incluyen soja, girasol, canola, alfalfa, colza, algodón, tomate, patata y tabaco. Preferiblemente además, la planta es una planta monocotiledónea. Los ejemplos de plantas monocotiledóneas incluyen a la caña de azúcar. Más preferiblemente la planta es un cereal. Los ejemplos de cereales incluyen arroz, maíz, trigo, cebada, mijo, centeno, triticale, sorgo y avena.

La invención también se extiende a las partes cosechables de una planta que comprende una secuencia aislada de ácido nucleico que codifica un AMT (como se definió aquí anteriormente) operativamente enlazado a un promotor constitutivo de una planta, tal como, pero sin limitarse a semillas, hojas, frutos, flores, tallos, rizomas, tubérculos y bulbos. La invención se refiere además a productos derivados, preferiblemente directamente derivados de una parte cosechable de dicha planta, tal como gránulos o polvos secos, aceite, grasa y ácidos grasos, almidón o proteínas.

Los métodos para incrementar la expresión de las secuencias de ácido nucleico o los genes, o productos génicos, están bien documentados en la técnica y se proveen ejemplos en la sección de definiciones.

Como se mencionó anteriormente, un método preferido para incrementar la expresión de una secuencia de ácido nucleico que codifica un polipéptido AMT es por medio de la introducción y expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT; sin embargo, los efectos de llevar a cabo el método, es decir, de incrementar las características relacionadas con el rendimiento, también pueden lograrse utilizando otras técnicas bien conocidas, que incluyen pero no se limitan a marcación de la activación del ADN-T, TILLING, recombinación homóloga. Una descripción de estas técnicas es suministrada en la sección de definiciones.

La presente invención abarca también el uso de secuencias de ácido nucleico que codifican polipéptidos AMT como se describe aquí; y el uso de estos polipéptidos AMT para incrementar cualquiera de las características relacionadas con el rendimiento anteriormente mencionadas en plantas, bajo condiciones de crecimiento normales, bajo condiciones de crecimiento de estrés abiótico (preferiblemente condiciones de crecimiento de estrés osmótico), y bajo condiciones de crecimiento de disponibilidad reducida de nutrientes, preferiblemente bajo condiciones de disponibilidad reducida de nitrógeno.

Las secuencias de ácido nucleico que codifican polipéptidos AMT descritos aquí, o los mismos polipéptidos AMT, pueden encontrar uso en programas de fitomejoramiento en los cuales se identifica un marcador de ADN que puede estar genéticamente enlazado a un gen que codifica un polipéptido AMT. Las secuencias de genes/ ácido nucleico, o los mismos polipéptidos AMT pueden ser utilizados para definir un marcador molecular. Este ADN o marcador de proteína pueden ser luego utilizados en programas de fitomejoramiento para seleccionar plantas que tienen más características relacionadas con rendimiento, como se definió aquí anteriormente en los métodos de la invención.

Las variantes alélicas de una secuencia de un gen/ácido nucleico que codifican un polipéptidos AMT pueden encontrar uso también en programas de fitomejoramiento asistidos por marcadores. Tales programas de fitomejoramiento algunas veces requieren la introducción de una variante alélica por medio de tratamiento mutagénico de las plantas, utilizando por ejemplo mutagénesis EMS; en forma alternativa, el programa puede

iniciarse con un grupo de variantes alélicas de las así llamadas de origen “natural” originadas en forma no intencional. La identificación de variantes alélicas tiene lugar entonces, por ejemplo, por medio de PCR. Esto es seguido por una etapa para la selección de variantes alélicas superiores de la secuencia en cuestión y que producen más características relacionadas con el rendimiento. La selección se lleva a cabo típicamente monitoreando el desempeño del crecimiento de las plantas que contienen diferentes variantes alélicas de la secuencia en cuestión. El desempeño del crecimiento puede ser monitoreado en un invernadero o en el campo. Otras etapas opcionales incluyen cruzamiento de plantas en las cuales se identificó la variante alélica superior con otra planta. Esto podría ser utilizado, por ejemplo, para elaborar una combinación de características fenotípicas interesantes.

Las secuencias de ácido nucleico que codifican polipéptidos AMT pueden ser utilizadas también como sondas para mapear genética y físicamente los genes de los cuales hacen parte, y como marcadores para características enlazadas a esos genes. Tal información puede ser útil en fitomejoramiento de plantas con el propósito de desarrollar líneas con fenotipos deseados. Tal uso de las secuencias de ácido nucleico que codifican un polipéptidos AMT requiere únicamente una secuencia de ácido nucleico de al menos 15 nucleótidos de longitud. Las secuencias de ácido nucleico que codifican un polipéptido AMT se pueden utilizar como marcadores de polimorfismo con la longitud del fragmento de restricción (RFLP por sus siglas en inglés). Las transferencias tipo Southern (Sambrook J, Fritsch EF y Maniatis T (1989) Molecular Cloning, A Laboratory Manual) del ADN genómico de la planta digerido con enzimas de restricción pueden ser sondeados con las secuencias de ácido nucleico que codifican un polipéptido AMT. Los patrones de bandas resultantes pueden ser luego sometidos a análisis genético utilizando programas de computador tales como MapMaker (Lander et al. (1987) Genomics 1: 174 - 181) con el propósito de construir un mapa genético. Además, las secuencias de ácido nucleico pueden ser utilizadas para examinar las transferencias tipo Southern que contienen los ADN genómicos tratados con endonucleasa de restricción de un conjunto de individuos que representan al progenitor y a la progenie de un cruce genético definido. Se observa la segregación de los polimorfismos de ADN y se utilizan para calcular la posición de la secuencia de ácido nucleico que codifica un polipéptido AMT en el mapa genético previamente obtenido utilizando esta población (Botstein et al. (1980) Am. J. Hum. Genet. 32: 314 - 331).

La producción y el uso de las sondas derivadas de genes de la planta para uso en mapeo genético se describe en Bernatzky y Tanksley (1986) Plant Mol. Biol. Reporter 4: 37 - 41. Numerosas publicaciones describen el mapeo genético de clones específicos de ADNc utilizando la metodología expuesta anteriormente o variantes de la misma. Por ejemplo, poblaciones de entrecruzamientos F2, poblaciones de retrocruzamiento, poblaciones apareadas al azar, líneas isogénicas cercanas, y otros conjuntos de individuos pueden ser utilizados para el mapeo. Estas metodologías son bien conocidas por aquellos ordinariamente capacitados en la técnica.

Las sondas de las secuencias de ácido nucleico también se pueden utilizar para mapeo físico (es decir, colocación de secuencias sobre mapas físicos; véase Hoheisel et al. En: Non-mammalian Genomic Analysis: A Practical Guide, Academic Press 1996, páginas 319 - 346, y las referencias citadas allí).

En otra realización, las sondas de secuencias de ácido nucleico pueden ser utilizadas en el mapeo de la hibridación *in situ* de fluorescencia directa (FISH por sus siglas en inglés) (Trask (1991) Trends Genet. 7: 149 - 154). Aunque los métodos actuales de mapeo FISH favorecen el uso de clones grandes (varios kb hasta varios cientos de kb, véase Laan et al. (1995) Genome Res. 5: 13 - 20), las mejoras en sensibilidad pueden permitir la realización del mapeo FISH utilizando sondas más cortas.

Una variedad de métodos basados en la amplificación de las secuencias de ácido nucleico para mapeo genético y físico pueden ser llevados a cabo utilizando las secuencias de ácido nucleico. Los ejemplos incluyen amplificación específica de alelos (Kazazian (1989) J. Lab. Clin. Med 11: 95 - 96), polimorfismo de fragmentos amplificados por PCR (CAPS; Sheffield et al. (1993) Genomics 16: 325 - 332), ligación específica de alelos (Landegren et al. (1988) Science 241: 1077 - 1080), reacciones de extensión del nucleótido (Sokolov (1990) Nucleic Acid Sequence Res. 18: 3671), Mapeo Híbrido por Radiación (Walter et al. (1997) Nat. Genet. 7: 22 - 28) y Mapeo Happy (Dear and Cook (1989) Nucleic Acid Sequence Res. 17: 6795 - 6807). Para estos métodos, se utiliza la secuencia de una secuencia de ácido nucleico para diseñar y producir pares de cebadores para uso en la reacción de amplificación o en reacciones de extensión del cebador. El diseño de tales cebadores es bien conocido por aquellos ordinariamente bien capacitados en la técnica. En los métodos que emplean mapeo genético con base en PCR, puede ser necesario identificar las diferencias en la secuencia de ADN entre los progenitores del mapeo entrecruzado en la región correspondiente a la presente secuencia de ácido nucleico. Esto, sin embargo, generalmente, no es necesario para métodos de mapeo.

Los métodos de acuerdo con la presente invención dan como resultado plantas que tienen más características relacionadas con el rendimiento, como se describió anteriormente aquí. Estas características pueden ser combinadas también con otras características económicamente ventajosas, tales como las características que incrementan adicionalmente el rendimiento, la tolerancia al estrés abiótico y biótico, tolerancia a herbicidas, insecticidas, características que modifican diferentes patrones arquitectónicos y/o características bioquímicas y/o fisiológicas.

Descripción de las figuras

La presente invención será descrita a continuación con referencia a las siguientes figuras en las cuales:

- 5 La Figura 1 representa la salida gráfica del algoritmo TMHMM2.0 para la SEQ ID NO: 2. A partir de la predicción del algoritmo, se localiza el terminal N del polipéptido sobre el lado exterior de la membrana (extracitosólico), seguido por 11 hélices que abarcan toda la membrana, el terminal C del polipéptido del polipéptido que está siendo localizado sobre el lado interno de la membrana (citósólico).
- 10 La Figura 2 muestra un árbol filogenético de polipéptidos AMT de diferentes organismos fuente: el Grupo I representa la agrupación de secuencias de polipéptidos útiles en la realización de los métodos de la invención, como se delimita por medio del paréntesis. El círculo representa el punto de ramificación en el árbol entre los polipéptidos útiles en la realización de los métodos de la invención, y los otros polipéptidos AMT.
- 15 La Figura 3 muestra una alineación de secuencia múltiple AlignX (del Vector NTI 10.3, Invitrogen Corporation) de los polipéptidos AMT de la Tabla A. El comienzo y el final del Dominio Conservado (DC), por ejemplo como el representado por la SEQ ID NO: 33, se muestra utilizando paréntesis y marcado por medio de las X bajo la secuencia de consenso. Se presenta en una caja un residuo conservado G (Gly) involucrado en la función propia de AMT (Ludewig et al., (2003) J Biol Chem 278: 45603 - 10).
- 20 La Figura 4 muestra el vector binario para mayor expresión en *Oryza sativa* de una secuencia de ácido nucleico que codifica un polipéptido AMT bajo el control de un promotor GOS2 (pGOS2) de arroz.

La Figura 5 detalla ejemplos de secuencias útiles para realizar los métodos de acuerdo con la presente invención.

25 Ejemplos

La presente invención será descrita a continuación con referencia a los siguientes y ejemplos, que se citan sólo para efectos ilustrativos. Los siguientes ejemplos no pretenden definir completamente o limitar de algún otro modo el alcance de la invención.

- 30 Manipulación del ADN: a menos que se establezca otra cosa, las técnicas de ADN recombinante se llevan acabo de acuerdo con protocolos estándar descritos en (Sambrook (2001) Molecular Cloning: a laboratory manual, 3^o Edición Cold Spring Harbor Laboratory Press, CSH, New York) o en los Volúmenes 1 y 2 de Ausubel et al. (1994), Current Protocols in Molecular Biology, Current Protocols. Los materiales estándar y los métodos para el trabajo molecular de la planta se describen en Plant Molecular Biology Labfax (1993) de R. D. D. Croy, publicado por BIOS Scientific Publications Ltd (RU) y Blackwell Scientific Publications (RU).

Ejemplo 1: Identificación de secuencias relacionadas con la secuencia de ácido nucleico utilizada en los métodos de la invención

- 40 Las secuencias (ADNc de longitud completa, los EST o genómicas) relacionadas con la secuencia de ácido nucleico usadas en los métodos de la presente invención fueron identificadas entre aquellas mantenidas en la base de datos de nucleótidos Entrez en el National Center for Biotechnology Information (NCBI) usando herramientas de búsqueda de secuencias en la base de datos, tales como la Basic Local Alignment Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215: 403 - 410; y Altschul et al. (1997) Nucleic Acids Res. 25: 3389 - 3402). El programa se utiliza para encontrar regiones de similitud local entre secuencias mediante la comparación de secuencias de ácido nucleico o de secuencias de polipéptidos con las bases de datos de secuencias y calculando el significado estadístico de las coincidencias. Por ejemplo, el polipéptido codificado por la secuencia de ácido nucleico de la presente invención se utilizó para el algoritmo TBLASTN, con parámetros predeterminados y el filtro desactivado para ignorar secuencias de baja complejidad. El resultado del análisis se obtuvo a través de la comparación por pares, y se clasificó de acuerdo con el puntaje de la probabilidad (valor E), donde el puntaje refleja la probabilidad de que ocurra una alineación particular al azar (entre más bajo el valor E, más significativa la coincidencia). Además de los valores E, también se dio puntaje a las comparaciones mediante el porcentaje de identidad. El porcentaje de identidad se refiere a la cantidad de nucleótidos (o aminoácidos) idénticos entre las dos secuencias de ácido nucleico (o polipeptídicas) comparadas sobre una longitud particular. En algunos casos, se pueden ajustar los parámetros por defecto para modificar la rigurosidad de la búsqueda. Por ejemplo, se puede incrementar el valor E para mostrar menos coincidencias estrictas. De este modo, se pueden identificar coincidencias cortas casi exactas.

- 60 La Tabla A proporciona un listado de secuencias de ácidos nucleico relacionadas con la secuencia de ácido nucleico utilizada en los métodos de la presente invención.

Tabla A: Ejemplos de secuencias de polipéptidos AMT, y secuencias codificadoras de ácido nucleico:

Nombre	Organismo fuente	Número de acceso en la base de datos pública	Secuencia de ácido nucleico SEQ ID NO:	Secuencia del polipéptido SEQ ID NO:
Phatr_AMT1_FL	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_20754_e stExt_gwp_gw1.C_c_hr_100028	1	2
Phatr_AMT1_parcial	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_20754_e stExt_gwp_gw1.C_c_hr_100028	3	4
Phatr_AMT2	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_22927_e stExt_gwp_gw1.C_c_hr_200095	5	6
Phatr_AMT3	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_20954_e stExt_gwp_gw1.C_c_hr_110018	7	8
Phatr_AMT4	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_34491_f genes1_pg.C_chr_5000291	9	10
Phatr_AMT5	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_33710_f genes1_pg.C_chr_4000065	11	12
Phatr_AMT6	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_53344_p hadr1_ua_pm.chr_1_0000004	13	14
Phatr_AMT7	<i>Phaeodactylum tricornutum</i>	jgi_Phatr2_43458_e stExt_fgenes1_pg.C_chr_20211	15	16
Cylfu_AMT1	<i>Cylindrotheca fusiformis</i>	AF360394	17	18
Cylfu_AMT1 variante	<i>Cylindrotheca fusiformis</i>	AY651852	19	20
Cylfu_AMT2a	<i>Cylindrotheca fusiformis</i>	AY651853	21	22
Thaps_AMT1	<i>Thalassiosira pseudonana</i>	jgi_Thaps3_258067 _thaps1_ua_pm.chr_7000085	23	24
Thaps_AMT2	<i>Thalassiosira pseudonana</i>	jgi_Thaps3_261441 _thaps1_ua_kg.chr_2000265	25	26
Thaps_AMT3	<i>Thalassiosira pseudonana</i>	jgi_Thaps3_7968_fg enesh1_pg.C_chr_9_000136	27	28
Thaps_AMT4	<i>Thalassiosira pseudonana</i>	jgi_Thaps3_2305_fg enesh1_pg.C_chr_2_000260	29	30
Thaps_AMT6	<i>Thalassiosira pseudonana</i>	jgi_Thaps3_257021 _thaps1_ua_pm.chr_4000455	31	32

En algunos casos, las secuencias relacionadas han sido ensambladas tentativamente y divulgadas al público por parte de instituciones de investigación, tales como The Institute for Genomic Research (TIGR). La base de datos Eukaryotic Gene Orthologs (EGO) puede ser utilizada para identificar tales secuencias relacionadas, ya sea por medio de una búsqueda de palabras clave o por medio del uso del algoritmo BLAST con la secuencia de ácido nucleico o la secuencias de polipéptidos de interés. En otros casos, se han creado bases de datos especiales de secuencias de ácido nucleico para organismos particulares, tales como por el Joint Genome Institute, por ejemplo para *Thalassiosira pseudonana* y *Phaeodactylum tricornutum*.

Ejemplo 2: Alineación de secuencias de polipéptidos AMT

La alineación de secuencias múltiples de todas las secuencias de polipéptidos AMT en la Tabla A se llevó a cabo utilizando el algoritmo AlignX (del Vector NTI 10.3, Invitrogen Corporation). Los resultados de la alineación se muestran en la Figura 3 de la presente solicitud. El comienzo y el final del Dominio Conservado (DC), por ejemplo como el representado por la SEQ ID NO: 33, se muestra utilizando paréntesis y marcado por medio de las X bajo la secuencia de consenso. Se presenta en una caja un residuo conservado G (Gly) involucrado en la función propia de AMT (Ludewig et al., (2003) J Biol Chem 278: 45603 - 10).

Ejemplo 3: Cálculo del porcentaje global de identidad entre secuencias de polipéptidos útiles en la realización de los métodos de la invención

Se determinaron los porcentajes globales de similitud e identidad entre secuencias de polipéptidos de longitud completa útiles para llevar a cabo los métodos de la invención utilizando uno de los métodos disponibles en el arte, el software MatGAT (Matrix Global Alignment Tool) (BMC Bioinformatics. 2003 4: 29. MatGAT: una aplicación que genera matrices de similitud/identidad utilizando secuencias de proteína o de ADN. Campanella J. J., Bitincka L., Smalley J; software administrado por Ledion Bitincka). El software MatGAT genera matrices de similitud/identidad para secuencias de ADN o de proteína sin necesidad de una alineación previa para los datos. El programa realiza una serie de alineaciones de pares utilizando el algoritmo de alineación global de Myers y Miller (con una

penalización por apertura de brecha de 12, y una penalización por extensión de brecha de 2), calcula la similitud y la identidad utilizando por ejemplo Blosum 62 (para polipéptidos), y luego coloca los resultados en una matriz de distancia. La similitud de secuencia se muestra en la mitad inferior de la línea divisoria y la identidad de la secuencia se muestra en la mitad superior de la línea divisoria en diagonal.

- 5 Los parámetros utilizados en la comparación fueron:
- Matriz de puntuación: Blosum62
Primera brecha: 12
Extensión de la brecha: 2
- 10 Los resultados del análisis del software se muestran en la Tabla B para la similitud e identidad global sobre la longitud completa de las secuencias de polipéptidos (excluyendo las secuencias parciales de polipéptidos).
- 15 Se hizo el mismo análisis entre el Dominio Conservado (DC) como el representado por la SEQ ID NO: 33 (y comprendida en la SEQ ID NO: 2 y el la SEQ ID NO: 4) y el Dominio Conservado de los polipéptidos de la Tabla A (como se resalta en la Figura 3) y los resultados se muestran en la Tabla B1.

Tabla B: Resultados de MatGAT para la similitud e identidad globales sobre la longitud completa de las secuencias de polipéptidos de la Tabla A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Cylfu_AMT1		94.7	68.9	50.5	46.9	49.5	44.6	47.4	45.9	61.4	50.6	46.3	49.5	44.0
2. Cylfu_AMT2a	97.7		68.8	50.1	46.7	50.8	44.8	46.5	45.4	60.5	50.3	46.0	49.0	44.0
3. Phatr_AMT1 SEQ ID NO: 2	81.8	82.7		52.6	46.7	48.9	47.5	48.9	46.8	61.7	52.8	46.4	50.0	44.3
4. Phatr_AMT2	64.9	65.7	65.9		46.8	44.0	45.5	47.8	45.1	49.7	59.7	45.3	46.2	43.7
5. Phatr_AMT3	62.9	63.1	61.0	60.6		46.9	63.3	52.2	61.9	48.4	46.7	49.4	50.3	48.4
6. Phatr_AMT4	66.5	67.3	67.0	59.9	63.3		47.0	57.8	45.7	47.3	45.5	46.7	52.0	44.2
7. Phatr_AMT5	61.0	60.7	60.5	61.0	77.8	62.3		50.0	59.4	44.7	45.2	48.5	49.6	47.5
8. Phatr_AMT6	61.2	61.6	64.3	62.6	66.5	73.5	64.4		49.3	47.8	50.9	48.4	49.6	46.2
9. Phatr_AMT7	64.7	64.4	63.1	63.3	75.0	64.6	74.1	65.7		46.2	44.6	46.4	50.4	45.4
10. Thaps_AMT1	73.1	72.9	73.3	62.7	62.2	64.4	60.3	64.3	63.6		50.5	46.0	48.2	44.0
11. Thaps_AMT2	66.5	66.2	67.1	72.0	60.8	62.6	58.9	67.0	63.8	65.2		45.5	46.4	42.8
12. Thaps_AMT3	63.0	61.5	59.9	58.7	64.4	62.6	65.3	63.5	63.7	60.6	59.9		56.8	50.4
13. Thaps_AMT4	66.1	65.6	64.7	59.9	64.7	66.7	64.2	62.8	65.9	63.7	60.6	71.5		51.5
14. Thaps_AMT6	57.9	57.6	57.0	57.2	62.2	58.7	61.9	61.1	59.9	57.6	56.0	63.4	60.6	

Tabla B.1: Resultados de MatGAT para la similitud e identidad globales sobre el Dominio Conservado de las secuencias de polipéptidos de la Tabla A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. CD_Cylfu_AMT1		95.6	74.7	56.4	56.7	57.5	55.1	56.9	53.0	68.0	56.9	57.1	56.3	55.2
2. CD_Cylfu_AMT2a	98.3		74.7	56.0	56.2	58.7	54.9	56.2	52.5	67.2	57.1	56.6	56.3	55.2
3. CD_Phattr_AMT1SEQID NO: 2	86.2	87.2		60.3	57.5	58.4	56.6	58.7	55.1	69.0	59.3	56.7	58.1	57.4
4. CD_Phattr_AMT2	72.5	72.8	74.5		55.5	54.4	52.7	56.6	53.0	57.8	68.0	54.5	55.3	54.5
5. CD_Phattr_AMT3	74.3	73.8	73.1	72.4		59.5	73.6	61.0	74.8	58.7	55.5	61.1	62.8	59.6
6. CD_Phattr_AMT4	74.6	75.1	73.8	70.7	76.3		58.9	72.6	56.3	56.9	55.3	58.5	60.9	57.6
7. CD_Phattr_AMT5	72.9	72.2	70.5	70.5	85.9	74.3		59.1	70.0	53.6	53.3	60.2	60.7	58.8
8. CD_Phattr_AMT6	73.1	73.1	74.5	71.8	77.4	86.4	75.3		57.9	58.7	58.3	58.2	60.0	58.2
9. CD_Phattr_AMT7	73.8	73.3	72.9	69.9	87.8	74.6	84.2	75.5		53.9	50.1	57.7	60.6	58.1
10. CD_Thaps_AMT1	78.8	79.3	78.8	71.7	74.4	72.9	71.0	75.0	73.9		56.9	57.1	56.4	55.8
11. CD_Thaps_AMT2	73.3	73.8	74.3	79.2	71.4	72.4	68.3	74.8	69.4	72.4		54.3	55.6	52.5
12. CD_Thaps_AMT3	74.3	73.1	72.0	68.1	76.5	74.1	76.0	74.3	75.8	72.0	68.4		67.2	63.7
13. CD_Thaps_AMT4	72.3	72.5	72.5	68.4	76.5	74.6	76.3	73.3	76.5	70.2	69.6	82.1		64.4
14. CD_Thaps_AMT6	72.4	71.9	70.5	70.2	73.8	73.8	73.1	74.1	73.8	71.4	66.8	77.0	75.3	

El porcentaje de identidad entre las secuencias de polipéptidos de longitud completa útiles en la realización de los métodos de la invención puede ser tan baja como una de identidad de aminoácido del 44% en comparación con la SEQ ID NO: 2.

- 5 El porcentaje de identidad entre el Dominio Conservado (DC) como el representado por la SEQ ID NO: 33 (y comprendida en la SEQ ID NO: 2 y el la SEQ ID NO: 4) y el Dominio Conservado de los polipéptidos de la Tabla A (como se resalta en la Figura 3) incrementa la identidad de aminoácidos hasta un 55%, como se muestra en la Tabla B1.

- 10 **Ejemplo 4:** Identificación de dominios comprendidos en las secuencias de polipéptidos útiles en la realización de los métodos de la invención.

- 15 La base de datos de Recursos Integrados de Familias de Proteínas, Dominios y Sitios (InterPro) es una interfaz integrada para las bases de datos de firma comúnmente utilizadas para búsquedas con base en textos y secuencias. La base de datos InterPro combina estas bases de datos, que utilizan diferentes metodologías y diferentes grados de información biológica sobre proteínas bien caracterizadas para derivar firmas de proteína. Las bases de datos colaboradoras incluyen SWISS-PROT, PROSITE, TrEMBL, PRINTS, Panther, ProDom y Pfam, Smart y TIGRFAMs. InterPro se encuentra alojada en el European Bioinformatics Institute en el Reino Unido.

- 20 Los resultados del barrido con InterPro de la secuencia de polipéptidos representadas por la SEQ ID NO: 2 se presentan en la Tabla C.

Tabla C: Resultados del barrido con InterPro de la secuencia de polipéptidos como la representada por la SEQ ID NO: 2

25

Nombre y número de acceso de InterPro	Nombre de la base de datos integrada	Número de acceso de la base de datos integrada	Nombre de acceso de la base de datos integrada
Transportador de Amonio IPR0001905	Panther	PTHR11730	Transportador de Amonio
Transportador de Amonio IPR0001905	TIGR	TIGR00836	Amt: Transportador de Amonio
Transportador de Amonio IPR0001905	Prosite	PS01219	Transportador de Amonio
Transportador de Amonio/proteína tipo Rh IPR010256	PFAM	F00909	Transportador de Amonio
No IPR integrado	Panther	PTHR11730:SF8	Transportador de Amonio 1
No IPR integrado	tmhmm		Regiones transmembrana

Ejemplo 5: Predicción de la localización subcelular de las secuencias de polipéptidos útiles para llevar a cabo los métodos de la invención

- 30 Los métodos experimentales para la localización de proteína varían desde inmunolocalización hasta marcación de las proteínas utilizando proteína fluorescente verde (GFP) o beta-glucuronidasa (GUS). Por ejemplo, se localizaron transportadores AMT de *Arabidopsis thaliana* en la membrana plasmática utilizando experimentos de fusión GFP (Yuan et al. (2003) Plant Cell 19: 2636 - 2652).

- 35 Se llevó a cabo también una predicción computacional de localización de proteína a partir de los datos de la secuencia. Entre los algoritmos bien conocidos por una persona normalmente capacitada en la técnica se encuentran disponibles en las herramientas de ExPASy Proteomics alojada en el Swiss Institute for Bioinformatics, por ejemplo, PSort, TargetP, ChloroP, LocTree, Predotar, LipoP, MITOPROT, PATS, PTS1, SignalP, TMHMM y otras.

- 40 Un dominio transmembrana usualmente denota una sola hélice alfa transmembrana de una proteína transmembrana. Se llama "dominio" debido a que una hélice alfa en la membrana se puede plegar independientemente del resto de la proteína. En forma más amplia, un dominio transmembrana es cualquier estructura tridimensional de una proteína que es termodinámicamente estable en la membrana. Esta puede ser una
45 sola hélice alfa, un complejo estable de varias hélices alfa transmembrana, un barril beta transmembrana, una hélice beta de gramicidina A, o cualquier otra estructura.

- 50 Las hélices transmembrana son usualmente aproximadamente de 20 aminoácidos de longitud, aunque pueden ser mucho más largas o cortas. TMHMM2.0 es un algoritmo que puede predecir hélices que abarcan toda la membrana en proteínas. El algoritmo se aloja en el servidor de la Universidad Técnica de Dinamarca. La tabla D a continuación muestra la salida de TMHMM2.0 utilizando la información de la secuencias de polipéptidos de la SEQ ID NO: 2. A partir de la predicción, se localiza el terminal N del polipéptido sobre el lado externo de la membrana

(extracitosólico), seguido por 11 hélices que abarcan toda la membrana, el terminal C del polipéptido del polipéptido que está localizado sobre el lado interno de la membrana (citósólico). La misma configuración aplica a la SEQ ID NO: 4, excepto porque la primera porción exterior es más pequeña. La Figura 1 es una representación gráfica de la salida como en la Tabla D.

5

Tabla D. Salida TMHMM2.0 utilizando la información de la secuencia de polipéptidos de la SEQ ID NO: 2

Localización	Coordenadas del aminoácido
externa	1 - 53
hélice TM	54 - 76
interna	77 - 87
hélice TM	88 - 110
externa	111 - 131
hélice TM	132 - 154
interna	155 - 160
hélice TM	161 - 183
externa	184 - 202
hélice TM	203 - 225
interna	226 - 245
hélice TM	246 - 268
externa	269 - 282
hélice TM	283 - 305
interna	306 - 317
hélice TM	318 - 340
externa	341 - 344
hélice TM	345 - 367
interna	368 - 379
hélice TM	380 - 402
externa	403 - 429
hélice TM	430 - 452
interna	453 - 521

El compartimiento subcelular predicho del polipéptido AMT como se representa mediante la SEQ ID NO: 4 utilizando el algoritmo TMHMM2.0 es la membrana.

10

Ejemplo 6: Ensayo relacionado con las secuencias de polipéptidos útiles en la realización de los métodos de la invención.

15

Los polipéptidos AMT son capaces de transportar amonio a través de las membranas. Existen muchos ensayos para medir tal actividad de captación, incluidos ensayos de complementación de una cepa de levadura con transportadores de amonio endógeno defectuosos (Ninneman et al. (1994) EMBO J 13: 3464 - 3471), ensayos de captación en levadura, *Xenopus* oocytes (Ludewig et al. (2003) J Biol Chem 278: 45603 - 45610), células de plantas, raíces de la planta (Yuan et al. (2007) Plant Phys 143: 732 - 744), y plantas completas (Hoque et al. (2006) Functional Plant Biology 33: 153 - 163). Una persona capacitada en el arte es consciente de tales procedimientos experimentales para medir la actividad de AMT, incluida la actividad de AMT de un polipéptido AMT como el representado por la SEQ ID NO: 2.

20

25

Ejemplo 7: Clonación de una secuencia de ácido nucleico como la representada por la SEQ ID NO: 1

25

A menos que se establezca otra cosa, se llevan a cabo técnicas de ADN recombinante de acuerdo con protocolos estándar que se describen en (Sambrook (2001) Molecular Cloning: a laboratory manual, 3ª edición Cold Spring Harbor Laboratory Press, CSH, Nueva York) o en los Volúmenes 1 y 2 de Ausubel et al. (1994), Current Protocols in Molecular Biology, Current Protocols. Los materiales y métodos estándar para trabajo molecular en plantas se describen en Plant Molecular Biology Labfax (1993) por R. D. D. Croy, publicado por BIOS Scientific Publications Ltd (RU) y Blackwell Scientific Publications (RU).

30

35

El ADNc de *Arabidopsis thaliana* que codifica la secuencia de polipéptidos AMT como la representada por la SEQ ID NO: 4 se amplificó por medio de PCR utilizando como molde ADNc de sintetizado a partir del ARNm extraído de *Phaeodactylum tricornutum* en diferentes etapas de la multiplicación, y bajo diferentes condiciones de crecimiento. Los siguientes cebadores, que incluyen los sitios AttB para recombinación Gateway, fueron utilizados para la amplificación por PCR:

40

1) Prm 09458 (SEQ ID NO: 35, sentido):
 5'- ggggacaagtgtgtacaaaaagcaggcttaacaaatgatgcaggccggg- 3'
 2) Prm 09459 (SEQ ID NO: 36, inversa, complementaria):

5'-ggggaccactttgtacaagaaagctgggtacacgagcagcaattaaacc- 3'

La PCR se realizó utilizando Hifii Taq ADN polimerasa en condiciones estándar. Se amplificó y purificó también un fragmento PCR de la longitud esperada (incluyendo sitios attB), empleando métodos estándar. Luego se llevó a cabo la primera etapa del procedimiento Gateway, la reacción BP, durante la cual el fragmento de la PCR recombinado *in vivo* con el plásmido pDONR201 para producir, de acuerdo con la terminología Gateway, un "clon de entrada". Se adquirió el plásmido pDONR201 a través de Invitrogen, como parte de la tecnología Gateway®.

Ejemplo 8: Construcción del vector de expresión utilizando las secuencias de ácido nucleico como las representadas por la SEQ ID NO: 1

El clon de entrada que comprende la SEQ ID NO: 3 se utilizó posteriormente en una reacción LR con un vector de destinación usado para la transformación de *Oryza sativa*. Este vector contenía como elementos funcionales dentro de los bordes del ADN-T: un marcador seleccionable de una planta; un casete de expresión de un marcador cribable; y un casete Gateway destinado a la recombinación *in vivo* de LR con la secuencia de ácido nucleico de interés ya clonada en el clon de entrada. Se localizó un promotor GOS2 de arroz (SEQ ID NO: 34) para expresión constitutiva secuencia arriba de este casete Gateway.

Después de la etapa de recombinación de LR, se transformó el vector de expresión resultante pGOS2::AMT (Figura 4) en una cepa LBA4044 de *Agrobacterium*, de acuerdo con los métodos bien conocidos en el arte.

Ejemplo 9: Transformación de plantas

Transformación de arroz

Se utilizó *Agrobacterium* que contiene el vector de expresión para transformar plantas de *Oryza sativa*. Se descascarillaron semillas secas maduras de la variedad de cultivo Japónica de arroz Nipponbare. Se realizó la esterilización incubando durante un minuto en etanol al 70%, seguido por 30 minutos en HgCl₂ al 0,2%, seguido de 6 lavados durante 15 minutos con agua destilada estéril. Las semillas estériles fueron luego germinadas en un medio que contenía 2,4-D (medio de inducción de callos). Después de la incubación en la oscuridad durante cuatro semanas, se cortaron y propagaron los callos embriogénicos derivados de escutelo sobre el mismo medio. Al cabo de dos semanas, los callos se multiplicaron o propagaron mediante un subcultivo en el mismo medio durante otras 2 semanas, se multiplicaron o propagaron los callos por medio de subcultivo sobre el mismo medio durante otras dos semanas. Se subcultivaron piezas de callo sobre medio fresco 3 días antes del cocultivo (para reforzar la actividad de la división celular).

Se utilizó la cepa LBA4404 de *Agrobacterium* que contenía cada vector de expresión individual independientemente para el cocultivo. Se inoculó *Agrobacterium* sobre medio AB con los antibióticos apropiados y se cultivó durante 3 días a 28°C. Se recolectaron luego las bacterias y se suspendieron en un medio de cocultivo líquido hasta una densidad (OD₆₀₀) de aproximadamente 1. Se transfirió luego la suspensión a una placa de Petri y se sumergieron los callos en la suspensión durante 15 minutos. Los tejidos de los callos fueron transferidos secos sobre un papel de filtro y transferidos a un medio de cocultivo solidificado e incubados durante 3 días en la oscuridad a 25°C. Los callos cocultivados crecieron en un medio que contenía 2,4-D durante 4 semanas en la oscuridad a 28°C en presencia de un agente de selección. Durante este período, se desarrollaron rápidamente islas de callos resistentes. Luego de transferir este material a un medio de regeneración e incubarlo a la luz, se liberó el potencial embriogénico y se desarrollaron brotes en las próximas cuatro a cinco semanas. Se cortaron los brotes de los callos y se incubaron durante 2 a 3 semanas en un medio que contenía auxina, desde el cual fueron transferidos al suelo. Se cultivaron los brotes encaltecidos bajo humedad alta y días cortos en un invernadero.

Se generaron aproximadamente 35 transformantes independientes de arroz T0 para cada construcción. Los transformantes primarios se transfirieron desde una cámara de cultivo de tejido a un invernadero. Luego de un análisis por PCR cuantitativo para verificar la cantidad de copias del inserto de ADN-T, se conservaron únicamente copias simples de plantas transgénicas que exhibieron tolerancia al agente de selección para la cosecha de semillas T1. Luego se cosecharon las semillas tres a cinco meses después de ser trasplantadas. El método produjo transformantes de un solo locus en una proporción superior al 50 % (Aldemita y Hodges1996, Chan et al. 1993, Hiei et al. 1994).

Ejemplo 10: Procedimiento de evaluación fenotípica

10.1 Disposición para la evaluación

Se generaron aproximadamente 35 nuevos transformantes independientes de arroz T0. Se transfirieron los transformantes primarios desde una cámara de cultivo de tejidos hasta un invernadero para crecimiento y cosecha de semilla T1. Se retuvieron seis eventos, de los cuales la progenie T1 segregó 3:1 por la presencia /ausencia del transgén. Para cada uno de estos eventos se seleccionaron aproximadamente 10 plantas de semillero T1 que contenían el transgén (hetero y homocigotos) y aproximadamente 10 plantas de semillero T1 que carecían del

transgén (nulicigotos) y por medio de control visual de expresión del marcador. Las plantas transgénicas y los correspondiente nulicigotos se cultivaron uno al lado del otro en posiciones aleatorias. Las condiciones del invernadero fueron de días cortos (12 horas de luz), 28°C en la luz y 22°C en la oscuridad, y una humedad relativa del 70%.

Se evaluaron adicionalmente cuatro eventos T1 en la generación T2 siguiendo el mismo procedimiento de evaluación que para la generación T1 pero con más individuos por evento. Desde la etapa de cultivo hasta la etapa de maduración, se pasaron las plantas varias veces a través de una cabina de formación de imágenes digitales. En cada instante de tiempo se tomaron imágenes digitales (2048 x 1536 píxeles, 16 millones de colores) de cada planta desde al menos 6 ángulos diferentes.

10.2 Análisis estadístico: prueba F

Se utilizó un ANOVA de dos factores (análisis de variantes) como un modelo estadístico para la evaluación general de las características fenotípicas de las plantas. Se realizó una prueba F sobre todos los parámetros medidos de todas las plantas de todos los eventos transformados con el gen de la presente invención. Se realizó la prueba F para controlar el efecto del gen sobre todos los eventos de transformación y para verificar el efecto total del gen, también conocido como efecto global del gen. El umbral de significación para un efecto global verdadero del gen se fijó en un nivel de probabilidad del 5% para la prueba F. Un valor significativo de la prueba F indica un efecto del gen, lo que significa que no es sólo la mera presencia o la posición del gen lo que está causando las diferencias en el fenotipo.

10.3 Parámetros medidos

Medición de parámetros relacionados con la biomasa

Desde la etapa de cultivo hasta la etapa de maduración, se pasaron las plantas varias veces a través de una cabina de formación de imágenes digitales. En cada instante de tiempo se tomaron imágenes digitales (2048 x 1536 píxeles, 16 millones de colores) de cada planta desde al menos 6 ángulos diferentes.

Se determinó el área de la planta sobre el suelo (o biomasa foliar) contando el número total de píxeles en la imágenes digitales de partes de la planta aéreas discriminadas desde el fondo. Se promedió este valor para las fotos tomada en el mismo instante de tiempo desde los diferentes ángulos y fue convertido a un valor de superficie física expresado en mm cuadrados por calibración. Los experimentos muestran que el área de la planta sobre el suelo medida de esta manera se correlaciona con la biomasa de las partes de la planta aéreas. El área por encima del suelo es el área medida en el instante de tiempo en el que la planta había alcanzado su biomasa foliar máxima. El vigor temprano es el área por encima del suelo de la planta (planta de semillero) tres semanas después de la germinación. El aumento de la biomasa de la raíz se expresa como un aumento en la biomasa total de la raíz (medida como biomasa máxima de las raíces observada durante la vida útil de una planta); o como un aumento en el índice raíz / brote (medido como la relación entre los masa de la raíz y la masa de los brotes en el período de crecimiento activo de las raíces y los brotes).

Mediciones de parámetros relacionados con la semilla

Se cosecharon las panículas primarias maduras, se contaron, se empacaron en bolsas, se etiquetaron con un código de barras y luego se secaron durante tres días en un horno a 37° C. Se trillaron luego las panículas, y se recogieron y contaron todas las semillas. Se separaron las vainas llenas de las vacías utilizando un dispositivo de soplado de aire. Las vainas vacías se descartaron y la fracción restante se contó de nuevo. Las vainas llenas se pesaron en una balanza analítica. El número de semillas llenas se determinó contando el número de vainas llenas que quedó después de la etapa de separación. El peso total de semillas por planta se midió pesando todas las vainas llenas cosechadas de una planta. Se midió la cantidad total de semillas por planta contando el número de vainas cosechadas de una planta. Se extrapoló el Peso de Mil Granos (PMG) a partir del número de semillas llenas contadas y su peso total. El Índice de Cosecha (IC) en la presente invención se define como la relación entre el peso total de semillas por planta y el área por encima del suelo (mm²), multiplicado por un factor de 10⁵. El número total de flores por panícula tal como se define en la presente invención es la relación entre el número total de semillas y el número de panículas primarias maduras. La tasa de llenado de semilla tal como se define en la presente invención es la proporción (expresada como %) del número de semillas llenas sobre el número total de semillas (o floretes).

Ejemplo 11: Resultados de la evaluación fenotípica de las plantas de arroz transgénicas que expresan la secuencia de ácido nucleico que codifica un polipéptido AMT como el representado por la SEQ ID NO: 2

Los resultados de la evaluación de plantas de arroz transgénicas generación T2 que expresan la secuencia de ácido nucleico que codifica un polipéptido AMT como el representado por la SEQ ID NO: 2, bajo el control del promotor GOS2 para expresión constitutiva, y cultivadas bajo condiciones normales de desarrollo, se presentan a continuación.

Existió un incremento significativo en el vigor temprano, en la biomasa aérea, en la biomasa de la raíz, en el rendimiento total de semilla por planta, en la tasa de llenado de la semilla, en el número de semillas llenas, en el número de flores por panícula, y en el índice de cosecha de las plantas transgénicas comparado con los correspondientes nulicigotos (controles), como se muestra en la Tabla E.

Tabla E: Resultados de la evaluación de las plantas de arroz transgénico generación T2 que expresan la secuencia de ácido nucleico que codifica un polipéptido AMT como el representado por la SEQ ID NO: 2, bajo el control del promotor GOS2 para expresión constitutiva.

Característica	Incremento total promedio en % en 4 eventos en la generación T2
Vigor temprano	39%
Biomasa aérea	16%
Biomasa de la raíz	4%
Rendimiento total de semilla por planta	31%
Tasa de llenado de semilla	12%
Número de semillas llenas	31%
Número de flores por panícula	6%
Índice d cosecha	17%

Ejemplo 12: Ejemplos de transformación de otros cultivos

Transformación de maíz

La transformación de maíz (*Zea mays*) se realiza con una modificación del método que describió por Ishida et al. (1996) Nature Biotech 14(6): 745 - 50. La transformación depende del genotipo en el maíz y únicamente genotipos específicos son responsables por la transformación y regeneración. La línea endogámica A188 (Universidad de Minnesota) o híbridos con A188 como progenitores son buenas fuentes de material donante para transformación, pero también pueden utilizarse otros genotipos exitosamente. Se cosecharon las mazorcas de plantas de maíz aproximadamente 11 días después de la polinización (DAP por sus siglas en inglés) cuando la longitud del embrión inmaduro es de aproximadamente 1 a 1,2 mm. Se cocultivaron los embriones inmaduros con *Agrobacterium tumefaciens* que contenía el vector de expresión, y se recuperaron las plantas transgénicas a través de organogénesis. Se cultivaron los embriones cortados en un medio de inducción de callos, luego en un medio de regeneración de maíz, que contiene el agente de selección (por ejemplo imidazolinona, pero pueden utilizarse diferentes marcadores de selección). Las placas de Petri se incubaron a la luz a 25°C durante 2 - 3 semanas, o hasta que crecieron brotes. Los brotes verdes se transfieren desde cada embrión a un medio enraizador de maíz y se incuban a 25°C durante 2 - 3 semanas, hasta que brotan raíces. Los brotes con raíces se trasplantan a suelo en el invernadero. Se producen semillas T1 a partir de las plantas que exhiben tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.

Transformación de trigo

La transformación de trigo se realiza con el método descrito por Ishida et al. (1996) Nature Biotech 14(6): 745 - 50. La variedad cultivada Bobwhite (disponible a través de CIMMYT, México) es comúnmente utilizada en la transformación. Los embriones inmaduros se cocultivan con *Agrobacterium tumefaciens* que contienen el vector de expresión, y se recuperan las plantas transgénicas a través de organogénesis. Después de incubación con *Agrobacterium*, se cultivan los embriones *in vitro* en un medio de inducción de callos, luego en un medio de regeneración que contiene el agente de selección (por ejemplo imidazolinona, pero pueden utilizarse diferentes marcadores de selección). Las placas de Petri se incuban a la luz a 25°C durante 2 - 3 semanas, o hasta que se desarrollaron brotes. Los brotes verdes se transfieren de cada embrión al medio enraizador y se incuban a 25°C durante 2 - 3 semanas, hasta que brotaron raíces. Los brotes con raíces se trasplantan a suelo en el invernadero. Se producen semillas T1 a partir de las plantas que exhiben tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.

Transformación de soja

La soja se transforma de acuerdo con una modificación del método que se describe en la patente de los Estados Unidos No. 5.164.310 de Texas A&M. Diferentes variedades comerciales de soja son adecuadas para la transformación mediante este método. La variedad cultivada Jack (disponible a través de la fundación de Semillas de Illinois) es comúnmente utilizada para la transformación. Las semillas de soja se esterilizan para el sembrado *in vitro*. Se cortan el hipocotilo, la radícula y un cotiledón de plántulas jóvenes de 7 días. Se cultivan luego el epicotilo y el cotiledón restante hasta desarrollar nodos axiales. Estos nodos axiales se cortan y se incuban con *Agrobacterium tumefaciens* que contenían el vector de expresión. Después del tratamiento de cocultivo, se lavaron los explantes y se los transfirió a un medio de selección. Los brotes regenerados se cortan y se colocan en un medio para alargamiento del brote. Los brotes no mayores a 1 cm se colocan en un medio de enraizamiento hasta desarrollar raíces. Los brotes con raíces se trasplantan a suelo en el invernadero. Se producen semillas T1 a partir de las

plantas que exhiben tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.

Transformación de colza / canola

- 5 Se usaron pecíolos e hipocotilos de cotiledones de una plántula joven de 5 - 6 días como explantes para el cultivo de tejidos, y se transformaron de acuerdo con Babic et al. (1998, Plant Cell Rep 17: 183 - 188). La variedad cultivada comercial Westar (Agriculture Canada) es la variedad estándar usada para la transformación, pero también pueden usarse otras variedades. Las semillas de canola se esterilizan en la superficie para la siembra *in vitro*. Se cortan los explantes de pecíolo de cotiledón con el cotiledón unido a partir de las plántulas *in vitro*, y se inoculan con
- 10 *Agrobacterium* (que contienen el vector de expresión) sumergiendo el extremo cortado del explante del pecíolo en la suspensión bacteriana. Se cultivan luego los explantes durante 2 días sobre medio MSBAP-3 que contiene 3 mg/l de BAP, 3% de sacarosa, 0,7% de Phytagar a 23°C, con 16 h de luz. Después de dos días de cocultivo con *Agrobacterium*, los explantes de pecíolos se transfieren a un medio MSBAP-3 que contiene 3 mg/l de BAP, cefotaxima, carbenicilina o Timentina (300 mg/l) durante 7 días, y luego se cultivaron sobre medio MSBAP-3 con cefotaxima, carbenicilina o Timentina y un agente de selección hasta la regeneración de los brotes. Cuando los brotes tienen 5 - 10 mm de longitud, se cortan y se transfieren al medio para alargamiento de los brotes (MSBAP-0,5, que contiene 0,5 mg/l de BAP). Se transfieren los brotes de aproximadamente 2 cm de longitud al medio de enraizado (MS0) para la inducción de la raíz. Los brotes con raíces se trasplantan a suelo en el invernadero. Se producen semillas T1 a partir de las plantas que exhiben tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.
- 20

Transformación de alfalfa

- 25 Se transforma un clon de regeneración de alfalfa (*Medicago sativa*) utilizando el método de (McKersie et al., 1999 Plant Physiol 119: 839 - 847). La regeneración y transformación de alfalfa depende del genotipo y por lo tanto se requiere una planta regeneradora. Ya se han descrito métodos para obtener plantas que se regeneran. Por ejemplo, estas pueden seleccionarse de la variedad cultivada Rangellander (Agriculture Canada) o de cualquier otra variedad comercial de alfalfa como lo describe Brown D. C. W. y A Atanassov (1985. Plant Cell Tissue Organ Culture 4: 111 - 112). Alternativamente, se ha seleccionado la variedad RA3 (Universidad de Wisconsin) para ser usada en el cultivo de tejidos (Walker et al., 1978 Am J Bot 65: 654 - 659). Los explantes de pecíolos se cocultivan con un cultivo de C58C1 pMP90 de *Agrobacterium tumefaciens* durante la noche (McKersie et al., 1999 Plant Physiol 119: 839 - 847) o LBA4404 que contiene el vector de expresión. Los explantes se cocultivan durante 3 días en la oscuridad en un medio de inducción SH que contiene 288 mg/L de Pro, 53 mg/L de tioprolina, 4,35 g/L de K₂SO₄, y 100 µm de acetosiringinona. Los explantes se lavan en un medio de Murashige-Skoog de fuerza media (Murashige y Skoog, 1962) y se siembran en placa sobre el mismo medio de inducción SH sin acetosiringinona, pero con un agente de selección adecuado y un antibiótico adecuado para inhibir el crecimiento de *Agrobacterium*. Después de varias semanas, se transfieren los embriones somáticos a un medio de desarrollo BOi2Y que no contiene reguladores de crecimiento ni antibióticos, y con 50 g/L de sacarosa. Luego se germinan los embriones somáticos en un medio de Murashige-Skoog de fuerza media. Las plántulas enraizadas se trasplantaron a macetas y se desarrollaron en un invernadero. Se producen semillas T1 a partir de las plantas que exhiben tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.
- 40

Transformación de algodón

- 45 La transformación del algodón (*Gossypium hirsutum* L.) se lleva a cabo utilizando *Agrobacterium tumefaciens*, sobre explantes de hipocotilos. Las variedades de cultivo comerciales tales como Coker 130 o Coker 312 (SeedCo, Lubbock, TX) son variedades estándares utilizadas para la transformación, pero también se pueden utilizar otras variedades. Las semillas se esterilizan en la superficie y se germinan en la oscuridad. Los explantes de hipocotilos se cortan de las plántulas germinadas hasta longitudes de aproximadamente 1 - 1,5 centímetros. El explante de hipocotilo se sumerge en el inóculo de *Agrobacterium tumefaciens* que contiene el vector de expresión, durante 5 minutos y luego se cocultivan durante 48 horas sobre MS + 1,8 mg/l de KNO₃ + 2% de glucosa a 24° C, en la oscuridad. Los explantes se transfieren al mismo medio que contiene los marcadores seleccionables apropiados bacterianas y vegetales (renovados varias veces), hasta que se observan callos embriogénicos. Los callos se separan y se subcultivan hasta que aparecen embriones somáticos. Las plántulas derivadas de los embriones somáticos se maduran sobre el medio de enraizamiento hasta que se desarrollan las raíces. Los brotes con raíz se trasplantan a suelo en macetas en el invernadero. Se producen semillas T1 a partir de las plantas que exhiben tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.
- 55

Ejemplo 13: Ejemplos de cribas de estrés abiótico

- 60 Criba por sequía
- Se cultivan plantas de un número seleccionado de eventos en tierra para macetas bajo condiciones normales hasta que se acercan a la etapa de retiro de las mismas. A continuación, se transfieren a una sección "seca", donde se suprime el riego. Se insertan sondas de humedad en macetas elegidas al azar para controlar el contenido de agua del suelo (CAS). Cuando el CAS va por debajo de ciertos umbrales, se riega nuevamente las plantas en forma
- 65

automática y continua hasta que se vuelva a alcanzar un nivel normal. Las plantas son entonces transferidas de nuevo a condiciones normales. El resto del cultivo (maduración de la planta, cosecha de semillas) es el mismo que para las plantas no cultivadas bajo condiciones de estrés abiótico. Los parámetros de crecimiento y rendimiento se registran como se detalla para el crecimiento bajo condiciones normales.

5 Criba por estrés salino

10 Las plantas se cultivan en un sustrato elaborado a base de fibras de coco y Argex (relación 3 a 1). Se utiliza una solución nutriente normal durante las dos primeras semanas después del trasplante de las plántulas en el invernadero. Después de las dos primeras semanas, se añade sal 25 mM (NaCl) a la solución nutriente, hasta que se cosechan las plantas. Los parámetros de crecimiento y rendimiento se registran como se detalla para el crecimiento bajo condiciones normales.

15 Cribado por disponibilidad reducida en nutrientes (nitrógeno)

20 Se cultivan plantas de seis eventos (semillas T2) en tierra para macetas bajo condiciones normales excepto para la solución nutriente. Las macetas se riegan desde el trasplante hasta la maduración con una solución nutriente específica que tiene un contenido reducido de nitrógeno (N), usualmente entre 7 a 8 veces menos. El resto del cultivo (maduración de la planta, cosecha de semillas) es el mismo que para las plantas no cultivadas bajo estrés abiótico. Los parámetros de crecimiento y rendimiento se registran como se detalla para el cultivo bajo condiciones normales.

LISTADO DE SECUENCIAS

25 <110> CropDesign N.V.

<120> Plantas que tienen un aumento de características relacionadas con el rendimiento y un método para elaboración de las mismas

30 <130> PF60250

<160> 36

35 <170> PatentIn versión 3.3

<210> 1

<211> 1566

<212> ADN

40 <213> Phaeodactylum tricomutum

<400> 1

atgtctttg attggacgc attctgact ggcctgacag ccgctcgag ctctctgag 60
 caggctgtgt gcgctctgca aacgatcgtg gccggagttt ctaagactgt cggaggtatt 120
 gacgcggaag ggattactgc aggtgtgat actttctcc tcattttgc ggggtgcctg 180
 gtctcatga tgcaggccgg gttcgccatg ctttgtctg gatccgtccg tcaaaagaat 240
 gtaaagaata ttatgctcaa gaactgttg gatgcctgtg gtgggtctat tggcttctac 300
 accgttggtt tcggcttcgc ttatggcggg gacgacacca ccgacaagac cttcattggc 360
 aacagctact tcgcgctccg tgattacaca aattatgcag gtttctctt ccagtttgcg 420
 ttgtctgcca ctgccgccac gattgttgc ggtacagttg ctgagcgatg caagatgtcg 480
 gcataccttt gctactctct ctttcttacg ggttcgtct atcccgtcgt tgtacgctct 540
 gtctggagct ccaacggggt cttgtcagcc ttacgtgccg acccctcca aggagtggga 600
 accgttgact ttgccggatc aggtgtggtg cacatgactg gaggactcac cgccttgatt 660
 gctgccattg ttcttgacc gcgtaagggt cggttctacg atgaggatgg caaccctctg 720
 gagacgcccg ccagcttccc agcccactct glagccctcc agatcctcgg aactttcatc 780
 ttgtggttcg gatggtacgg attcaaccct gggtcagccc tgaagattgc taacgccgat 840
 tcggccgcaa ccgccgcttt gtgtgccgtc accaccacta tggccgccgc tgccggttgt 900

gttccgcca tgtcactga ctgatcatt gacggcatgg cgaccggtga aactacgtac 960
 gatctgacca tggccatgaa tggatgcctt gctggctctg ttgccgtcac tgctggtaca 1020
 tctgtctga ccccatgggc cgcaatcatt attggagtcg ttggagggtg ggtctacatt 1080
 ggtatgtcca agcttttgat caagctcaag attgatgacg ctgctgatgc catccctgtc 1140
 catttcgcca atggtttctg ggggtgtcta gccaccggcc ttctcgccaa cgggtggattg 1200
 atggcaaccg ctgggtacaa ctcggaacac gagggctggt tctacgaatg gggaagtggc 1260
 tccggagatg gaagtcttct catctgccag ctgcttgcc tcgctggat tattggatgg 1320
 gtcaccacca ttatgacgcc ctttttata ctttgaaca tggccggtat gttccgtgtg 1380
 gacccgcttg aggaagaagt tggcttgat attcccatc accgtggatc tgcttacgat 1440
 ctttcgggac ccagcaagga ccatgttgac gagctcatgg aaattcgtgc ctgaagcac 1500
 ggcaagggtg aggttcaaaa ggaggttgcg caggctgctg atgacgccgc cgaagagact 1560
 gcttaa 1566

<210> 2
 <211> 521
 <212> PRT
 <213> Phaeodactylum tricornutum

<400> 2

Met Ser Phe Asp Leu Asp Ala Phe Cys Thr Gly Leu Thr Ala Ala Ser
 1 5 10 15

Ser Ser Ser Glu Gln Ala Val Cys Ala Leu Gln Thr Ile Val Ala Gly
 20 25 30

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

Val Asp Thr Phe Phe Leu Ile Phe Ala Gly Ala Leu Val Phe Met Met
 50 55 60

Gln Ala Gly Phe Ala Met Leu Cys Ala Gly Ser Val Arg Gln Lys Asn
65 70 75 80

Val Lys Asn Ile Met Leu Lys Asn Leu Leu Asp Ala Cys Gly Gly Ala
85 90 95

Ile Gly Phe Tyr Thr Val Gly Phe Gly Phe Ala Tyr Gly Gly Asp Asp
100 105 110

Thr Thr Asp Lys Thr Phe Ile Gly Asn Ser Tyr Phe Ala Leu Arg Asp
115 120 125

Tyr Thr Asn Tyr Ala Gly Phe Phe Phe Gln Phe Ala Phe Ala Ala Thr
130 135 140

Ala Ala Thr Ile Val Ala Gly Thr Val Ala Glu Arg Cys Lys Met Ser
145 150 155 160

Ala Tyr Leu Cys Tyr Ser Leu Phe Leu Thr Gly Phe Val Tyr Pro Val
165 170 175

Val Val Arg Ser Val Trp Ser Ser Asn Gly Phe Leu Ser Ala Phe Ser
180 185 190

Ala Asp Pro Phe Gln Gly Val Gly Thr Val Asp Phe Ala Gly Ser Gly
195 200 205

Val Val His Met Thr Gly Gly Leu Thr Ala Leu Ile Ala Ala Ile Val
210 215 220

Leu Gly Pro Arg Lys Gly Arg Phe Tyr Asp Glu Asp Gly Asn Pro Leu
225 230 235 240

Glu Thr Pro Ala Ser Phe Pro Ala His Ser Val Ala Leu Gln Ile Leu
245 250 255

Gly Thr Phe Ile Leu Trp Phe Gly Trp Tyr Gly Phe Asn Pro Gly Ser
260 265 270

Ala Leu Lys Ile Ala Asn Ala Asp Ser Ala Ala Thr Ala Ala Leu Cys
275 280 285

Ala Val Thr Thr Thr Met Ala Ala Ala Ala Gly Cys Val Ser Ala Met
290 295 300

Phe Thr Asp Ser Ile Ile Asp Gly Met Ala Thr Gly Glu Thr Thr Tyr
305 310 315 320

Asp Leu Thr Met Ala Met Asn Gly Cys Leu Ala Gly Leu Val Ala Val
325 330 335

Thr Ala Gly Thr Ser Val Val Thr Pro Trp Ala Ala Ile Ile Ile Gly
340 345 350

Val Val Gly Gly Trp Val Tyr Ile Gly Met Ser Lys Leu Leu Ile Lys
355 360 365

Leu Lys Ile Asp Asp Ala Val Asp Ala Ile Pro Val His Phe Ala Asn
370 375 380

Gly Phe Trp Gly Val Leu Ala Thr Gly Leu Phe Ala Asn Gly Gly Leu
385 390 395 400

Met Ala Thr Ala Gly Tyr Asn Ser Glu His Glu Gly Trp Phe Tyr Glu
405 410 415

Trp Gly Ser Gly Ser Gly Asp Gly Ser Leu Leu Ile Cys Gln Leu Ala
420 425 430

Cys Leu Ala Trp Ile Ile Gly Trp Val Thr Thr Ile Met Thr Pro Phe
435 440 445

Phe Ile Leu Leu Asn Met Ala Gly Met Phe Arg Val Asp Pro Leu Glu
450 455 460

Glu Glu Val Gly Leu Asp Ile Ser His His Arg Gly Ser Ala Tyr Asp
465 470 475 480

Leu Ser Gly Pro Ser Lys Asp His Val Asp Glu Leu Met Glu Ile Arg
485 490 495

Ala Ser Lys His Gly Lys Val Glu Val Pro Lys Glu Val Ala Gln Ala
500 505 510

Ala Asp Asp Ala Ala Glu Glu Thr Ala
515 520

- 5 <210> 3
<211> 1380
<212> ADN
<213> Phaeodactylum tricornutum

<400> 3

atgatgcagg ccgggttcgc catgctttgt gctggatccg tccgtcaaaa gaatgtaaag 60
 aatattatgc tcaagaactt gtggatgcc tgtgggtgtg ctattggctt ctacaccgtt 120
 gggttcggct tcgcttatgg cggtgacgac accagcgaca agaccttcat tggcaacagc 180
 tacttcgcg cccgtgatta cacaattat gcaggtttct tcttcagtt tgcgtttgct 240
 gccactgccg ccacgattgt tgccggtaca gttgctgagc gatgcaagat gtcggcatac 300
 ctttgctact ctctctttct tacgggttgc gtctatcccg tcgtgtacg ctctgtctgg 360
 agctccaacg ggttctgtc agccttcagt gccgaccctt tccaaggagt tggaaccgtt 420
 gactttgccg gatcaggtgt ggtgcacatg actggaggac tcaccgcctt gattgctgcc 480
 attgttctg gaccgcgtaa gggtcgggtc tacgatgagg atggcaacc cctggagacg 540
 cccgccagct tccagccca ctctgtagcc ctccagatcc tcggaacttt catctgtgg 600
 ttcggatggt acggattcaa ccctgggtca gccctgaaga ttgctaacgc cgattcggcc 660
 acaaccgccc ctttgtgtgc cgtcaccacc actatggccg ccgctgccgg ttgtgtttcc 720
 gccatgttca ctgactcgat cattgacggc atggcgaccg gtgaaactac gtacgatctg 780
 accatggcca tgaatggatg ccttgctggt ctggtgccg tcactgctgg tacatctgtc 840
 gtcaccccat gggccgcaat cattattgga gtcattggag gttgggtcta cattggtatg 900
 tccaagcttt tgataaagct caagattgat gacgctgtcg atgcatccc tgtccatttc 960
 gccaacggtt tctggggtgt cctagccacc ggccttttcg ccaacggtgg attgatggca 1020
 accgctgggt acaactcgga acacgagggc tggttctacg aatggggaag tggctccgga 1080
 gatggaagtc ttctcatctg ccagcttgct tgccctgcct ggattattgg atgggtcacc 1140
 accattatga cgccctttt tatcctttg aacatggccg gtatgttccg tgtggacccg 1200
 ctgaggaag aagtgggtct tgatatttc catcaccgtg gatctgctta cgatcttcg 1260
 ggacccagca aggaccatgt tgacgagctc atggaaattc gtgcctcgaa gcacggcaag 1320
 gttgaggttc caaaggaggt tgcgcaggct gctgatgacg ccgccgaaga gactgcttaa 1380

<211> 459
 <212> PRT
 <213> Phaeodactylum tricornutum

5 <400> 4

Met Met Gln Ala Gly Phe Ala Met Leu Cys Ala Gly Ser Val Arg Gln
 1 5 10 15

Lys Asn Val Lys Asn Ile Met Leu Lys Asn Leu Leu Asp Ala Cys Gly
 20 25 30

Gly Ala Ile Gly Phe Tyr Thr Val Gly Phe Gly Phe Ala Tyr Gly Gly
 35 40 45

Asp Asp Thr Ser Asp Lys Thr Phe Ile Gly Asn Ser Tyr Phe Ala Leu
 50 55 60

Arg Asp Tyr Thr Asn Tyr Ala Gly Phe Phe Phe Gln Phe Ala Phe Ala
65 70 75 80

Ala Thr Ala Ala Thr Ile Val Ala Gly Thr Val Ala Glu Arg Cys Lys
85 90 95

Met Ser Ala Tyr Leu Cys Tyr Ser Leu Phe Leu Thr Gly Phe Val Tyr
100 105 110

Pro Val Val Val Arg Ser Val Trp Ser Ser Asn Gly Phe Leu Ser Ala
115 120 125

Phe Ser Ala Asp Pro Phe Gln Gly Val Gly Thr Val Asp Phe Ala Gly
130 135 140

Ser Gly Val Val His Met Thr Gly Gly Leu Thr Ala Leu Ile Ala Ala
145 150 155 160

Ile Val Leu Gly Pro Arg Lys Gly Arg Phe Tyr Asp Glu Asp Gly Asn
165 170 175

Pro Leu Glu Thr Pro Ala Ser Phe Pro Ala His Ser Val Ala Leu Gln
180 185 190

Ile Leu Gly Thr Phe Ile Leu Trp Phe Gly Trp Tyr Gly Phe Asn Pro
195 200 205

Gly Ser Ala Leu Lys Ile Ala Asn Ala Asp Ser Ala Thr Thr Ala Ala
210 215 220

Leu Cys Ala Val Thr Thr Thr Met Ala Ala Ala Ala Gly Cys Val Ser
225 230 235 240

Ala Met Phe Thr Asp Ser Ile Ile Asp Gly Met Ala Thr Gly Glu Thr
245 250 255

Thr Tyr Asp Leu Thr Met Ala Met Asn Gly Cys Leu Ala Gly Leu Val
260 265 270

Ala Val Thr Ala Gly Thr Ser Val Val Thr Pro Trp Ala Ala Ile Ile
275 280 285

Ile Gly Val Ile Gly Gly Trp Val Tyr Ile Gly Met Ser Lys Leu Leu
290 295 300

Ile Lys Leu Lys Ile Asp Asp Ala Val Asp Ala Ile Pro Val His Phe
305 310 315 320

Ala Asn Gly Phe Trp Gly Val Leu Ala Thr Gly Leu Phe Ala Asn Gly
325 330 335

Gly Leu Met Ala Thr Ala Gly Tyr Asn Ser Glu His Glu Gly Trp Phe
340 345 350

Tyr Glu Trp Gly Ser Gly Ser Gly Asp Gly Ser Leu Leu Ile Cys Gln
355 360 365

Leu Ala Cys Leu Ala Trp Ile Ile Gly Trp Val Thr Thr Ile Met Thr
370 375 380

Pro Phe Phe Ile Leu Leu Asn Met Ala Gly Met Phe Arg Val Asp Pro
385 390 395 400

Leu Glu Glu Glu Val Gly Leu Asp Ile Ser His His Arg Gly Ser Ala
405 410 415

Tyr Asp Leu Ser Gly Pro Ser Lys Asp His Val Asp Glu Leu Met Glu
420 425 430

Ile Arg Ala Ser Lys His Gly Lys Val Glu Val Pro Lys Glu Val Ala
435 440 445

Gln Ala Ala Asp Asp Ala Ala Glu Glu Thr Ala
450 455

<210> 5
<211> 1620
<212> ADN
<213> Phaeodactylum tricornutum

<400> 5

atgtcgagta gtgctatcta tcaatcttgc gccgggtcaat tcgatagcgg tgagcagctt 60
gatcagcttc tgcagtgcct ctctacggga catgacggcg cgctaagcga tcagaccagc 120
aacctcgag gaggtatcga tgccttttat ctatctttg ccggggccct tgtctacttc 180
atgcagacag gatttgctat gctctgtgcc ggatccattc gagcgaagaa tgtaagaat 240
gtgattctct ggaacttgc tcatcatgt ggaggcggtc tcgcttctg gagtgtaggc 300
tatgcctttg cctacggtag agacaatgct gggccaaga catttgttg taacgcaggc 360
ttttctcc agggagatga cattcgactg gaaaactggt ttccaatt tgcctttgct 420
tgtgccctt cctcaattgt cgctggaacg atcgctgagc gcactcaaat gaaggcctat 480
ttgatgtatt ccgtttttt ggctgggtt gtctatccag ttgtgccca cgcattttg 540
tcgagcaacg gattctctc caacacagcc acagatccat tgtggggatc tggcgccatt 600
gacttggcgg gctctggacc agtcacatg accggagggt ttacggctt ggccgcagct 660
ctggttttg gccctcgcat tggtcgttc tacgacaagg agggcaatcc tctgaagag 720
ccagccgaat ttctcccca ttcggttgc ttgcagttc tgggaacgt ttgtctttg 780
ttggctggt acggctttaa ccaggttcc gtctcttca ttcgagtat tgaaaacggt 840
caagttgcgg cttggttagc cgtcaatacc accctagctg cgtgtgcggg tgccgtcagc 900
gccatgtta ctcaacttt attgactac tggtaaccg gcttgcacac atacgattg 960
ggctacacta tgaacggatg ttgacggga ttggtcgca taacagcggg atgcgcgact 1020
gttgaaacgt gggctgccgt cctgattggt attggcgctg gttggttta cctgttgggt 1080
tcaaagttgc ttgtctactt ccgtattgac gatgctgtcg acgccattcc cgtgcatacg 1140

gtgggtggcg cctgggggtgt gatcgcgacg ggcctttta cgaagggaga actactattg 1200
 gccgcctttg gccaggagga acatgttga tggttctacg agtggggcag tgggagcgg 1260
 aactttaccc taatcggaat tcaattgctg tctgtgctgt tcatttcgc ctggactttt 1320
 tccgtcatgg gaatttatt ttacgccctt agcttcatgg gtggttgcg catagatccg 1380
 ttggaagagg aggttggtat ggatatttcg cgccacaagg gctcggcata cgacatgact 1440
 tcagcaaata tggagcaggt acggtcgtg atggacgacc gcagtacgag caatcgaggc 1500
 aagctgcgta agtcgtccat cattgagaag ccggccaaaa caggtccgat cgagggagga 1560
 actgcctccg gagaatcaga aattaaggag tcggacgtgg aacatgccga gaggggtctaa 1620

<210> 6

<211> 539

<212> PRT

<213> Phaeodactylum tricornutum

<400> 6

Met Ser Ser Ser Ala Ile Tyr Gln Ser Cys Ala Gly Gln Phe Asp Ser
 1 5 10 15

Gly Glu Gln Leu Asp Gln Leu Leu Gln Cys Leu Ser Thr Gly His Asp
 20 25 30

Gly Ala Leu Ser Asp Gln Thr Ser Asn Leu Ala Gly Gly Ile Asp Ala
 35 40 45

Phe Tyr Leu Ile Phe Ala Gly Ala Leu Val Tyr Phe Met Gln Thr Gly
 50 55 60

Phe Ala Met Leu Cys Ala Gly Ser Ile Arg Ala Lys Asn Val Lys Asn
 65 70 75 80

Val Ile Leu Trp Asn Leu Leu Asp Ser Cys Gly Gly Gly Leu Ala Phe
 85 90 95

Trp Ser Val Gly Tyr Ala Phe Ala Tyr Gly Gly Asp Asn Ala Gly Ser
100 105 110

Lys Thr Phe Val Gly Asn Ala Gly Phe Phe Leu Gln Gly Asp Asp Ile
115 120 125

Arg Leu Glu Asn Trp Phe Phe Gln Phe Ala Phe Ala Cys Ala Leu Ser
130 135 140

Ser Ile Val Ala Gly Thr Ile Ala Glu Arg Thr Gln Met Lys Ala Tyr
145 150 155 160

Leu Met Tyr Ser Val Phe Leu Ala Gly Phe Val Tyr Pro Val Val Ala
165 170 175

His Ala Phe Trp Ser Ser Asn Gly Phe Leu Ser Asn Thr Ala Thr Asp
180 185 190

Pro Leu Trp Gly Ser Gly Ala Ile Asp Leu Ala Gly Ser Gly Pro Val
195 200 205

His Met Thr Gly Gly Val Thr Ala Leu Ala Ala Ala Leu Val Leu Gly
210 215 220

Pro Arg Ile Gly Arg Phe Tyr Asp Lys Glu Gly Asn Pro Leu Glu Glu
225 230 235 240

Pro Ala Glu Phe Pro Pro His Ser Val Ala Leu Gln Phe Leu Gly Thr
245 250 255

Phe Cys Leu Trp Phe Gly Trp Tyr Gly Phe Asn Pro Gly Ser Val Phe
260 265 270

Phe Ile Ser Ser Ile Glu Asn Gly Gln Val Ala Ala Leu Val Ala Val
275 280 285

Asn Thr Thr Leu Ala Ala Cys Ala Gly Ala Val Ser Ala Met Phe Thr
290 295 300

Ser Thr Leu Phe Asp Tyr Trp Tyr Thr Gly Leu His Thr Tyr Asp Leu
305 310 315 320

Gly Tyr Thr Met Asn Gly Cys Leu Thr Gly Leu Val Ala Ile Thr Ala
325 330 335

Gly Cys Ala Thr Val Glu Thr Trp Ala Ala Val Leu Ile Gly Ile Gly
340 345 350

Ala Gly Trp Phe Tyr Leu Leu Gly Ser Lys Leu Leu Val Tyr Phe Arg
355 360 365

Ile Asp Asp Ala Val Asp Ala Ile Pro Val His Met Val Gly Gly Ala
370 375 380

Trp Gly Val Ile Ala Thr Gly Leu Phe Thr Lys Gly Glu Leu Leu Leu
385 390 395 400

Ala Ala Phe Gly Gln Glu Glu His Val Gly Trp Phe Tyr Glu Trp Gly
405 410 415

Ser Gly Ser Gly Asn Phe Thr Leu Ile Gly Ile Gln Leu Leu Ser Val
420 425 430

Leu Phe Ile Phe Ala Trp Thr Phe Ser Val Met Gly Ile Tyr Phe Tyr
435 440 445

Ala Leu Ser Phe Met Gly Trp Leu Arg Ile Asp Pro Leu Glu Glu Glu
450 455 460

Val Gly Met Asp Ile Ser Arg His Lys Gly Ser Ala Tyr Asp Met Thr
465 470 475 480

Ser Ala Asn Met Glu Gln Val Arg Ser Leu Met Asp Asp Arg Ser Thr
485 490 495

Ser Asn Arg Gly Lys Leu Arg Lys Ser Ser Ile Ile Glu Lys Pro Ala
500 505 510

Lys Thr Gly Pro Ile Glu Gly Gly Thr Ala Ser Gly Glu Ser Glu Ile
515 520 525

Lys Glu Ser Asp Val Glu His Ala Glu Arg Val
530 535

<210> 7

<211> 1695

<212> ADN

<213> Phaeodactylum tricornutum

<400> 7

atgaatagcc gtaccttttc ttacaacagc aacgatggga atgagttgtt ggatgcatgc 60
atggcttata tcggcgctaa cgccacgacg tacgacttgc tcggctgcgt gtctgcgcaa 120
ctcagcaacg aagtcgggag tcgcgaattc tctcgttcgg tgctgctcgt ctacgctgct 180
gcactcgtat tctttatgca agccggcttc gccatgctct gtgccggagc cgtccggaaa 240
aagaacgtgc agaacaccat gctcaagaac ttgttgagc cctgtggagc cgccgtcgcc 300
ttttcatcg tgggctacgc tatagccttt ggaggcatgg aaccggaatc gccaacaag 360
acatttctcg ggaacaccaa ttcttttta atgggagtgg acgatttggc cttttggtt 420
ttcagtagc cctttccgc tgccctgca accatcgtcg cggaactt ggccgagcgc 480
tgccaaatgg ttgcctactt gtgttactcc gtaatgttga cgggatgggt ctatccgatt 540
atcgcgacg cgatttggc acccaacggt tggctctccg ccagttccgt ggatccgcta 600
tggggtgtcg gcatggtcga ttccgccggc tcgggagtag tgcacatgac tggaggcggt 660
accgccttgt tcgccactct catactcgga cctcgccgtg gacggttca cgatgaaacc 720
ggacgccggc tcgacaaacc aaaatccttt cccggacact ccgtggcctt gcagatgctc 780

ggtaccttca ttctctggtt tgggtggtat ggattcaatt gcggttcggc gttactcatc 840
 gacaagcccg gtgctaacga tattgccgcc ttggcgggcg tcaatacgac actttctgcc 900
 ggagtcgctg ggaatagttgc actctttgtc aacctctggt acctcgagcg aacaaccgga 960
 gaacctttt ttgattgac ctacgccatg aacggatccc ttccgggict cgtggccatc 1020
 acaggaggtt gcgctgttct cgagccctgg gcggcagcgg tcaactggagt cggtgccggt 1080
 atctgtaca tggttggatc acgagggttg gtcatgttgc gattggatga cgcggttgac 1140
 gcgattcccg ttcatthgt gaacggggcc tggggctca tgcggtcgg attgttcgcg 1200
 tcaccagcac gcttgctggc cgcgtacgac aacgacgctc atcctggatg gttctattcc 1260
 ctacgaaacg gcaaatcgga cgggcgtttg gtggagttc agctggtagg cattgtattt 1320
 atcgtgggat gggtcatggt gattatgttg ccttcttca ttggttga ttggaaggga 1380
 tggttccggt cggatccttt ggaagaaatt gtgggattgg acacgtcgta ccacggtggt 1440
 ctgcctttgc tgggcgggga cgatgaagtc aacctgagt acattccgc gtacaagaag 1500
 caaaggaacg aaggtacgct gcggcgtagg cataaaggga caaccagcag tgtcaagacg 1560
 ggagaagtag aatcggacga aggcgccgaa cgggtggcgc ccgaagcgac gaagcacaaa 1620
 aagattccga tcgctacaac gacagaggtt ttcaatggag gtcaagatac ggatcaaattg 1680
 tcctaccacg cctga 1695

<210> 8
 <211> 564
 <212> PRT
 <213> Phaeodactylum tricornutum
 <400> 8

Met Asn Ser Arg Thr Phe Ser Tyr Asn Ser Asn Asp Gly Asn Glu Leu
 1 5 10 15
 Leu Asp Ala Cys Met Ala Tyr Leu Gly Ala Asn Ala Thr Thr Tyr Asp
 20 25 30

Leu Leu Gly Cys Val Ser Ala Gln Leu Ser Asn Glu Val Gly Ser Arg
35 40 45

Glu Phe Ser Arg Ser Val Leu Leu Val Tyr Ala Ala Ala Leu Val Phe
50 55 60

Phe Met Gln Ala Gly Phe Ala Met Leu Cys Ala Gly Ala Val Arg Lys
65 70 75 80

Lys Asn Val Gln Asn Thr Met Leu Lys Asn Leu Leu Asp Ala Cys Gly
85 90 95

Ala Ala Val Ala Phe Phe Ile Val Gly Tyr Ala Ile Ala Phe Gly Gly
100 105 110

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

Phe Leu Met Gly Val Asp Asp Leu Ala Phe Trp Leu Phe Gln Tyr Ala
130 135 140

Phe Ser Ala Ala Ser Ala Thr Ile Val Ala Gly Thr Leu Ala Glu Arg
145 150 155 160

Cys Gln Met Val Ala Tyr Leu Cys Tyr Ser Val Met Leu Thr Gly Trp
165 170 175

Val Tyr Pro Ile Ile Ala His Ala Ile Trp Ser Pro Asn Gly Trp Leu
180 185 190

Ser Ala Ser Ser Val Asp Pro Leu Trp Gly Val Gly Met Val Asp Phe
195 200 205

Ala Gly Ser Gly Val Val His Met Thr Gly Gly Val Thr Ala Leu Phe
210 215 220

Ala Thr Leu Ile Leu Gly Pro Arg Arg Gly Arg Phe His Asp Glu Thr
225 230 235 240

Gly Arg Arg Leu Asp Lys Pro Lys Ser Phe Pro Gly His Ser Val Ala
245 250 255

Leu Gln Met Leu Gly Thr Phe Ile Leu Trp Phe Gly Trp Tyr Gly Phe
260 265 270

Asn Cys Gly Ser Ala Leu Leu Ile Asp Lys Pro Gly Ala Asn Asp Ile
275 280 285

Ala Ala Leu Ala Gly Val Asn Thr Thr Leu Ser Ala Gly Val Ala Gly
290 295 300

Ile Val Ala Leu Phe Val Asn Leu Trp Tyr Leu Glu Arg Thr Thr Gly
305 310 315 320

Glu Pro Phe Phe Asp Leu Thr Tyr Ala Met Asn Gly Ser Leu Ser Gly
325 330 335

Leu Val Ala Ile Thr Gly Gly Cys Ala Val Leu Glu Pro Trp Ala Ala
340 345 350

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

Gly Leu Val Met Leu Arg Leu Asp Asp Ala Val Asp Ala Ile Pro Val
370 375 380

His Phe Val Asn Gly Ala Trp Gly Leu Met Ser Val Gly Leu Phe Ala
385 390 395 400

Ser Pro Ala Arg Leu Leu Ala Ala Tyr Asp Asn Asp Ala His Pro Gly
405 410 415

Trp Phe Tyr Ser Leu Arg Asn Gly Lys Ser Asp Gly Arg Leu Val Gly
420 425 430

Val Gln Leu Val Gly Ile Val Phe Ile Val Gly Trp Val Met Val Ile
435 440 445

Met Leu Pro Phe Phe Ile Trp Leu Asp Trp Lys Gly Trp Phe Arg Ser
450 455 460

Asp Pro Leu Glu Glu Ile Val Gly Leu Asp Thr Ser Tyr His Gly Gly
465 470 475 480

Leu Ala Leu Leu Gly Gly Asp Asp Glu Val Asn Pro Glu Tyr Ile Ser
485 490 495

Ala Tyr Lys Lys Gln Arg Asn Glu Gly Thr Leu Arg Arg Arg His Lys
500 505 510

Gly Thr Thr Ser Ser Val Lys Thr Gly Glu Val Glu Ser Asp Glu Gly
515 520 525

Ala Glu Arg Val Ala Pro Glu Ala Thr Lys His Lys Lys Ile Pro Ile
530 535 540

Ala Thr Thr Thr Glu Val Phe Asn Gly Gly Gln Asp Thr Asp Gln Met
545 550 555 560

Ser Tyr His Ala

<210> 9

<211> 1554

<212> ADN

<213> Phaeodactylum tricornutum

<400> 9

atggacgacg cgagctttat ccaaagcctg gtggaagggt acggtacctc gagcaaccac 60

accaccgtct atggatattg ctccaatgag gcggaaggag agaccaacct tattttgcaa 120
 tgtattacgg aggtcatgga gcaaaagcag ctggaaggag atcgtaacgt gaatcggtgg 180
 ctaatgcttt ttccgggtgg gctgatcttc tcatgcaga ctggctttgc gatgctttgc 240
 gctgggttgcg tgcgcaaaaa gaacgttcaa aatacgaatgc ttaaaaaatct ttggatgcc 300
 tgtggagctg ctctaggctt ttctacta ggctacgctt ttgcgtttgg cggacaagac 360
 gaccgggacg atgttacttt catcggaact tccaacttgc taaacaccgg aaaagtcgat 420
 atgtccttct gggttttcca gtttgcattc tcggccactg ccgtgactat cgtcgcagga 480
 actttggcag agcgttgcca aatggtagca tacctttgct attccatatt ttgacaggt 540
 ttgtgtacc cgggtgccgc ccacacaatt tggctcgcga acggcttct cagcagcaca 600
 gcagtagacc ctttcaagg tgtgggagcg attgatttcg ccggatccgg tgtggttcat 660
 gtgacgggag gaaccacggc tctcgtcgcc acatatattc ttggagctag aaaaggctgt 720
 ttctacgaca atcgcggtcg ccagctggag acgccaaagt cctttccggg gcattcagtt 780
 gctctgcagc tacttggtag attcgtattg tggttggat ggtacggctt caatcctgga 840
 tcagcattgc tgctggccca cacttcggat acgggtttcg tggcatcccg agcggctgtc 900
 aatacctcgc tttagctgc ttccggagcc gtaicggcac tgatgacaaa tatgtttatg 960
 gaagagcgtt ccaccggcga atactccttt aacattatca tggccatgaa tggcgcctta 1020
 gctggtctag tgtcaattac agcagcttgt ggtaccgtac aaaactgggc agccctttgc 1080
 actggatgta ttggaggctt tatctacttg tggggttcca aaacgctggt tcgtttgaaa 1140
 ttggacgatg ctgttgacgc gattccagtg cacatgtttg caggtggctg gggattgctt 1200
 gctgtgggtt tgtaagtga tctgacctc atgcacattg cctatgggac tggcaatcat 1260
 cctggtttat tgtattcttg ggggttaggg gagtttaacg ccatcttgct gagcaatcaa 1320
 gtgttgaac tagttttcgt ggcgggatgg gcttttgga ccatgacgcc ttctttttg 1380
 ttcatcaatc gtaggggatg gttccggtcg gacagtttg aggagttggt cggtttgat 1440
 gaggcctacc acggaggtaa acatggcggg gaagaagtgg tggaactatc cgctctagaa 1500
 ggctttatca aaaacaagat cagacaatcg caaatcggt cgccttaca ttag 1554

<211> 517
 <212> PRT
 <213> Phaeodactylum tricornutum

5 <400> 10

Met Asp Asp Ala Ser Phe Ile Gln Ser Leu Val Glu Gly Tyr Gly Thr
 1 5 10 15

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

Gly Glu Thr Asn Leu Ile Leu Gln Cys Ile Thr Glu Val Met Glu Gln
 35 40 45

Lys Gln Leu Glu Gly Asp Arg Asn Val Asn Arg Trp Leu Met Leu Phe
 50 55 60

Ser Gly Gly Leu Ile Phe Phe Met Gln Thr Gly Phe Ala Met Leu Cys
 65 70 75 80

Ala Gly Cys Val Arg Lys Lys Asn Val Gln Asn Thr Met Leu Lys Asn
 85 90 95

Leu Leu Asp Ala Cys Gly Ala Ala Leu Gly Phe Phe Leu Leu Gly Tyr
 100 105 110

Ala Phe Ala Phe Gly Gly Gln Asp Asp Arg Asp Asp Val Thr Phe Ile
 115 120 125

Gly Thr Ser Asn Phe Leu Asn Thr Gly Lys Val Asp Met Ser Phe Trp
 130 135 140

Phe Phe Gln Phe Ala Phe Ser Ala Thr Ala Val Thr Ile Val Ala Gly

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

340 345 350
 Val Gln Asn Trp Ala Ala Leu Cys Thr Gly Cys Ile Gly Gly Leu Ile
 355 360 365
 Tyr Leu Trp Gly Ser Lys Thr Leu Val Arg Leu Lys Leu Asp Asp Ala
 370 375 380
 Val Asp Ala Ile Pro Val His Met Phe Ala Gly Gly Trp Gly Leu Leu
 385 390 395 400
 Ala Val Gly Leu Leu Ser Asp Pro Asp Leu Met His Ile Ala Tyr Gly
 405 410 415
 Thr Gly Asn His Pro Gly Leu Leu Tyr Ser Trp Gly Leu Gly Glu Phe
 420 425 430
 Asn Ala Ile Leu Leu Ser Asn Gln Val Leu Glu Leu Val Phe Val Ala
 435 440 445
 Gly Trp Ala Phe Gly Thr Met Thr Pro Phe Phe Leu Phe Ile Asn Arg
 450 455 460
 Met Gly Trp Phe Arg Ser Asp Ser Leu Glu Glu Leu Val Gly Leu Asp
 465 470 475 480
 Glu Ala Tyr His Gly Gly Lys His Gly Gly Glu Glu Val Val Glu Leu
 485 490 495
 Ser Ala Leu Glu Gly Phe Ile Lys Asn Lys Ile Arg Gln Ser Gln Ile
 500 505 510
 Arg Ser Pro Tyr Asn
 515

<210> 11
 <211> 1704
 <212> ADN

<213> *Phaeodactylum tricornutum*

<400> 11

atgagtcact ctagtgtgga tgtgttcggg acctgccttg cccaagtgg cgaagacgca 60
 acgacaaagg agctactaga atgcgtttcc ttagtttgt cgcgagcagt gccagacggt 120
 ctcatgagc ctagtctaa aggcttcacg cggctctatcg tigtctgtt tgcggctgcc 180
 ttggttttt tcatgcaagc cggctttgcc atgctgtgcg ctggagccgt tagggccaag 240
 aatgttcaa acaccatgct caagaatctt ttgatgctt gtggtgccgc cattgcgttc 300
 ttactgtag gctacgcctt tgcctttggc ggtacggact tcccaccga caccgacacc 360
 gacaccggca acaaccaaac aacattcatt ggcacatcga acttttctt ggtgaatgtg 420
 gacgattatt cctttggct gttcaatac gcattttccg ccgcgtccgc aaccattgtt 480
 gctggaacac tggctgaacg gtgtcaaag gccgcctacc tgggatatc cgctttgta 540
 acgggatggg tgtaccccat tgcgctcac gctgtatgga acgtccacgg ctctctatcg 600
 gctcatgcgg tagaacctt gtgggggggt ggaatggtag atttgccgg ttctggggtc 660
 gttcacgta cgggaggtgt gacagcgctc ttgcgacaa taattctggg tccccgccgt 720
 ggacgcttc acgaccaaga tggccaaaga ttgataaggc cgcaaatatt tcctggacac 780
 tccttgctc tgcaaatgct tgaaccctt atcttatggt ttgctggta cggattcaat 840
 attggcgctg ctctgtgct ggacgtgcct agttcagaca atattgcagc tctggctgcg 900
 gtgaatacga cctgtcggg cgggacggct ggcacgttg ctctctttt caattgtgg 960
 tatctggaca aaagaactgg cgaagcctat ttgacttga aatttgccat gaatgggtgc 1020
 ctctgcggtc tctggccat cactggaggc tgtggtgtgg tgaaccctg ggctgccgtt 1080
 gtgattggct ttgtgccgg ttgtgttac aacatcgga gtcgcggact tatatattg 1140
 cgttggatg atgccgtgga tgccattccg gtacattgt gcaacggctc gtggggctc 1200
 gtagcagtag ggctgtttgc cccccctcc cggctactag ttatttacgg acatagcgat 1260
 catccaggat ggttttactc attacgagat ggcgaatccg atttcgggt actggcttc 1320

cagttggtgg gccttatttt cattgtcttc tgggtcatgt tcaacatgct gcctttttc 1380
 gtttggttaa actaccgtgg ctggttcgg tccgatccgt tggaagagtt ggtgggtctc 1440
 gacctgagct atcacggcgg ttaatgctg caccaggaag ttgatcccga atacatatcg 1500
 gcctatcgca agggccaaca cgaggctcat tctgtactc tgcgacaacg gaaacgcaca 1560
 tcccacgtgc gcttggaaac agtgagcgaa cattccgttg cgccaaatgg cggagactca 1620
 gccatacta gtaacggcga ttcgctcacg gaaacgggtca ctatccgtga accggactcc 1680
 ggttggccaa gggggaacgt ataa 1704

<210> 12
 <211> 567
 <212> PRT
 <213> Phaeodactylum tricornutum
 <400> 12

Met Ser His Ser Ser Leu Asp Val Phe Gly Thr Cys Leu Ala Gln Val
 1 5 10 15

Gly Glu Asp Ala Thr Thr Lys Glu Leu Leu Glu Cys Val Ser Phe Ser
 20 25 30

Leu Ser Arg Ala Val Pro Asp Gly Leu Asp Glu Pro Ser Ser Lys Gly
 35 40 45

Phe Thr Arg Ser Ile Val Val Val Phe Ala Ala Ala Leu Val Phe Phe
 50 55 60

Met Gln Ala Gly Phe Ala Met Leu Cys Ala Gly Ala Val Arg Ala Lys
 65 70 75 80

Asn Val Gln Asn Thr Met Leu Lys Asn Leu Leu Asp Ala Cys Gly Ala
 85 90 95

Ala Ile Ala Phe Phe Thr Val Gly Tyr Ala Phe Ala Phe Gly Gly Thr

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

290	295	300
Leu Ser Gly Gly Thr Ala Gly Ile Val Ala Leu Phe Phe Asn Leu Trp		
305	310	315 320
Tyr Leu Asp Lys Arg Thr Gly Glu Ala Tyr Phe Asp Leu Lys Phe Ala		
	325	330 335
Met Asn Gly Cys Leu Cys Gly Leu Val Ala Ile Thr Gly Gly Cys Gly		
	340	345 350
Val Val Glu Pro Trp Ala Ala Val Val Ile Gly Phe Val Ala Gly Leu		
	355	360 365
Leu Tyr Asn Ile Gly Ser Arg Gly Leu Ile Tyr Leu Arg Leu Asp Asp		
	370	375 380
Ala Val Asp Ala Ile Pro Val His Leu Cys Asn Gly Ser Trp Gly Leu		
385	390	395 400
Val Ala Val Gly Leu Phe Ala Ser Pro Ser Arg Leu Leu Val Ile Tyr		
	405	410 415
Gly His Ser Asp His Pro Gly Trp Phe Tyr Ser Leu Arg Asp Gly Glu		
	420	425 430
Ser Asp Phe Arg Leu Leu Ala Ser Gln Leu Val Gly Leu Ile Phe Ile		
	435	440 445
Val Phe Trp Val Met Phe Asn Met Leu Pro Phe Phe Val Trp Leu Asn		
	450	455 460
Tyr Arg Gly Trp Phe Arg Ser Asp Pro Leu Glu Glu Leu Val Gly Leu		
465	470	475 480
Asp Leu Ser Tyr His Gly Gly Leu Met Leu His Glu Glu Val Asp Pro		

485 490 495

Glu Tyr Ile Ser Ala Tyr Arg Lys Gly Gln His Glu Ala His Ser Arg
500 505 510

Thr Leu Arg Gln Arg Lys Arg Thr Ser His Val Arg Leu Glu Ser Val
515 520 525

Ser Glu His Ser Val Ala Pro Asn Gly Gly Asp Ser Ala His Thr Ser
530 535 540

Asn Gly Asp Ser Leu Thr Glu Thr Val Thr Ile Arg Glu Pro Asp Ser
545 550 555 560

Gly Trp Pro Arg Gly Asn Val
565

<210> 13
<211> 1665
<212> ADN
<213> Phaeodactylum tricornutum

<400> 13

atgagcacat cccatggcag tctgtttcaa cagtgttctg ccgtcgctgg ggacagtgac 60
ccttcccgga ttctgcagtg tgtgtccgat gctctcgaaa ccaacaaaaa cgatcgagct 120
gccgatctca ataactggtt tctcatcatt gctggggcac tcgtctttt catgcagtcc 180
ggcttcgcca tgctgtgtgc ggggtgtgtg cggaaaaaga acgtccaaaa caccatgctc 240
aagaatcttc tcgatgcctg cgggtccgct ctcgggtttt acgtaatcgg ttacgccttg 300
gccttgggtg gacaaaacga gcgatccgat gttaccttg tgggtaccac ggacttttc 360
aactggaaca gcgccgtccc ggtaaataca gccttttggg ttctgaatt tgccttttcg 420
gccacgtccg tgacgattgt ggccgggacg ctggcggaac ggtgtcaaat ggtcgcctac 480
ctgtgctact cggatattct aacaggcttt gtgtatcccg ttgtggccca ctccatcgg 540
agcaacaacg gctttctaag tgcatttgca gccgagccct tccagggcat tggcgtcctg 600

gactttgccg ggtcgggtgt ggtacacgtg acgggcggga ccactgcctt ggtcgccacg 660
 tataatgctag gggcccgtaa ggggcgcttc tacgacgcac gtgggagaga attggaaaag 720
 cccaaagcat ttcccgcca ttccatggcc ttgcagatga tgggcacat gattctgttg 780
 ttccgatggt acggcttcaa ccttggttcg gccttgttcg tgacggcgac gtccaatata 840
 ggtggcgtag cggctctcgc ggccgtgaac acgtccctct ctgccgcttc gggcgccgtt 900
 tccgctctct ttacctttt gtatctcgaa gaacgcaaga ctggagaata ctcttcaac 960
 atcaccatgg ctatgaacgg tgctttggct ggcttagtgg gtattacggc cggctgtggc 1020
 accgtcgaaa attgggccgc ttgtgcaca gggctcgtgt ctggatgggt ctatatattt 1080
 ggcagtgcct tcttgttcg tataaagatt gacgatgcag tcgacgcaat tccggtgcat 1140
 atgtttgtg gcgcctgggg tcttattgca accggcctat tcagctcgcc acgccatacg 1200
 ttagaagcat ttggtacgga cgcacacgta ggttggttct acagtcttg ccaagattct 1260
 ttggacgcta ttctttgat gaaccagctt ttgggtctt tattattct cgggtggagc 1320
 gccatcttga tgcaccatt ttctgggtgg ctgaactaca tgggatggtt gagggccgac 1380
 tcttggagg agttggtagg tctggatcaa gcctaccacg gtggaaggga ggctggagag 1440
 ggctttgatg atgaagtgcc ttgtccaca tccaaagaca aaccagacac ttacgtcga 1500
 cggaaagggtg gcaacagcgc ggagggcacc cgcgatgatg gaagtggac cgatttgact 1560
 gcctcggctc ctcgggaaac tgcgcctctg ccagaacttt attgtgagaa cgatgcgtct 1620
 tccaaagaat ccgacgacga tccgtccgga atgcaccacg cataa 1665

<210> 14
 <211> 554
 <212> PRT
 <213> Phaeodactylum tricornutum

<400> 14

Met Ser Thr Ser Asp Gly Ser Leu Phe Gln Gln Cys Ser Ala Val Ala

1 5 10 15

Gly Asp Ser Asp Pro Ser Arg Ile Leu Gln Cys Val Ser Asp Ala Leu
20 25 30

Glu Thr Asn Gln Asn Asp Arg Ala Ala Asp Leu Asn Asn Trp Phe Leu
35 40 45

Ile Ile Ala Gly Ala Leu Val Phe Phe Met Gln Ser Gly Phe Ala Met
50 55 60

Leu Cys Ala Gly Cys Val Arg Lys Lys Asn Val Gln Asn Thr Met Leu
65 70 75 80

Lys Asn Leu Leu Asp Ala Cys Gly Ala Ala Leu Gly Phe Tyr Val Ile
85 90 95

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

Phe Val Gly Thr Thr Asp Phe Phe Asn Trp Asn Ser Ala Val Pro Val
115 120 125

Asn Gln Ala Phe Trp Phe Phe Glu Phe Ala Phe Ser Ala Thr Ser Val
130 135 140

Thr Ile Val Ala Gly Thr Leu Ala Glu Arg Cys Gln Met Val Ala Tyr
145 150 155 160

Leu Cys Tyr Ser Val Phe Leu Thr Gly Phe Val Tyr Pro Val Val Ala
165 170 175

His Ser Ile Trp Ser Asn Asn Gly Phe Leu Ser Ala Phe Ala Ala Glu
180 185 190

Pro Phe Gln Gly Ile Gly Val Leu Asp Phe Ala Gly Ser Gly Val Val
195 200 205

His Val Thr Gly Gly Thr Thr Ala Leu Val Ala Thr Tyr Met Leu Gly
210 215 220

Ala Arg Lys Gly Arg Phe Tyr Asp Ala Arg Gly Arg Glu Leu Glu Lys
225 230 235 240

Pro Lys Ala Phe Pro Gly His Ser Met Ala Leu Gln Met Met Gly Thr
245 250 255

Met Ile Leu Trp Phe Gly Trp Tyr Gly Phe Asn Pro Gly Ser Ala Leu
260 265 270

Leu Leu Thr Ala Thr Ser Asn Thr Gly Gly Val Ala Ala Leu Ala Ala
275 280 285

Val Asn Thr Ser Leu Ser Ala Ala Ser Gly Ala Val Ser Ala Leu Phe
290 295 300

Thr Ser Leu Tyr Leu Glu Glu Arg Lys Thr Gly Glu Tyr Ser Phe Asn
305 310 315 320

Ile Thr Met Ala Met Asn Gly Ala Leu Ala Gly Leu Val Gly Ile Thr
325 330 335

Ala Gly Cys Gly Thr Val Glu Asn Trp Ala Ala Cys Cys Thr Gly Leu
340 345 350

Val Ser Gly Trp Val Tyr Ile Phe Gly Ser Ala Phe Leu Leu Arg Ile
355 360 365

Lys Ile Asp Asp Ala Val Asp Ala Ile Pro Val His Met Phe Cys Gly
370 375 380

Ala Trp Gly Leu Ile Ala Thr Gly Leu Phe Ser Ser Pro Arg His Thr
385 390 395 400

Leu Glu Ala Phe Gly Thr Asp Ala His Val Gly Trp Phe Tyr Ser Leu
405 410 415

Gly Gln Asp Ser Leu Asp Ala Ile Leu Leu Met Asn Gln Leu Leu Gly
420 425 430

Leu Leu Phe Ile Leu Gly Trp Ser Ala Ile Leu Met Ser Pro Phe Phe
435 440 445

Trp Trp Leu Asn Tyr Met Gly Trp Leu Arg Ala Asp Ser Leu Glu Glu
450 455 460

Leu Val Gly Leu Asp Gln Ala Tyr His Gly Gly Arg Glu Ala Gly Glu
465 470 475 480

Gly Phe Asp Asp Glu Val Pro Phe Ala Thr Ser Lys Asp Lys Pro Asp
485 490 495

Thr Leu Arg Arg Arg Lys Gly Gly Asn Ser Ala Glu Gly Thr Arg Asp
500 505 510

Asp Gly Ser Trp Thr Asp Leu Thr Ala Ser Ala Pro Arg Glu Thr Ala
515 520 525

Pro Leu Pro Glu Leu Tyr Cys Glu Asn Asp Ala Ser Ser Lys Glu Ser
530 535 540

Asp Asp Asp Pro Ser Gly Met His His Ala
545 550

<210> 15
<211> 1620
5 <212> ADN
<213> Phaeodactylum tricornutum

<400> 15

10 atgatctcta ctggctcttc cagctccacg aatgcctacg gcacttgctc gggttcagctg 60

ggcgaaaact ctccgcgaa agagctcttg gaatgtgtt cagattatct acagaatcag 120
 gaggcgcctt tctcgtctac tctggctact acttttgcgg gcgcgattgt ctttctaag 180
 caagccggct ttgcatggg ctgtgctggg gctgtccgca caaaaaatgt ccaaaacgcc 240
 atgctcaaga atctgttga tgcttgcggg gcgtccttgg cgttctttc cattggctac 300
 gctctggggg ttgggggtat ggagcccgaa agttccaaaa agaccttgt cggacatagt 360
 caattcttc tgatggacgt tgacgactac gcttttggg tattccaata cgcctttcc 420
 gctgcatctg ccacaattgt cgcaggaacg ctgcggaac gatgtcaaat gacggcctat 480
 ctttgctact cccttatgt aaccgggtgg gtatacccg ttattctaca ctccatatgg 540
 aatcccaacg gctggctatc tgcgtactcg gttgatccgc ttggggcag cgggctggg 600
 gattttgctg gctctggcgt cgttcatgtg accggaggaa tcaccgctct gtttgctaca 660
 atggttttg gccccgacg aggacgctt cacgatgatc ttgggcatga tctggcacgg 720
 ccacgggaat ttacggctca ttgccggct ctgcaaatgc ttggtactt tattttgtgg 780
 ttggctttt acggatttaa cattggctcc gcgctgataa gcacaaagca aggctcagac 840
 gaagcagccg ccttggccgg tgtaacaca acgctctccg ctagcgcggc aggtattgtg 900
 gcgctgttt caaatctatg gtacctggag aagacgaccg gtgagcctct cttgatcta 960
 aagtatgcta tgaatggcg catctgcggg ctgttagcta ttctggcgg ctgcgggggc 1020
 ttgagccct gggccgcggg ggtcaccggg gccgtggccg gtgtgatata tctattaggt 1080
 agtcgaggac ttgtatccat gcgactggat gacgctgtgg atgcatccc tglacacctc 1140
 tgtggtgggg cctggggcat cctggcggg ggactgttcg cagcgccga acgtctcctt 1200
 tccgtgtatg gacggaataa tcaccagga ttggttata gcattcgtga aggagatata 1260
 gatggcgttc ttctcggaat tcaactaatt ggtctcatgt tcatcatggg atgggttatg 1320
 atcatcatgc taccttctt tgatggctc aactggaaag gctgggtccg atcagacccg 1380
 ctggaggaga ttctcgggct cgatttaagc taccacgtcg gattggcgtt acataccaac 1440
 aatgttcac cggaatacgt tggcagcgaa aaagatgtgg tcgacgaaat tatttctacg 1500

cgccaacgaa aggtcaacgg gagcaccact accaaggcga cctcaggtac ggaagaattg 1560

gagtatatcc cggaagttag cgacgaagac ttgagcgaga tgaaagagga atgcttgtga 1620

<210> 16

<211> 539

<212> PRT

<213> Phaeodactylum tricornutum

<400> 16

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

Ser Val Gln Leu Gly Glu Asn Ser Ser Ala Lys Glu Leu Leu Glu Cys
20 25 30

Val Ser Asp Tyr Leu Gln Asn Gln Glu Ala Pro Phe Ser Ser Thr Leu
35 40 45

Val Leu Thr Phe Ala Gly Ala Ile Val Phe Leu Met Gln Ala Gly Phe
50 55 60

Ala Met Val Cys Ala Gly Ala Val Arg Thr Lys Asn Val Gln Asn Ala
65 70 75 80

Met Leu Lys Asn Leu Leu Asp Ala Cys Gly Ala Ser Leu Ala Phe Phe
85 90 95

Ser Ile Gly Tyr Ala Leu Gly Phe Gly Gly Met Glu Pro Glu Ser Ser
100 105 110

Lys Lys Thr Phe Val Gly His Ser Gln Phe Phe Leu Met Asp Val Asp
115 120 125

Asp Tyr Ala Phe Trp Leu Phe Gln Tyr Ala Phe Ser Ala Ala Ser Ala
130 135 140

Thr Ile Val Ala Gly Thr Leu Ala Glu Arg Cys Gln Met Thr Ala Tyr
145 150 155 160

Leu Cys Tyr Ser Leu Met Leu Thr Gly Trp Val Tyr Pro Val Ile Leu
165 170 175

His Ser Ile Trp Asn Pro Asn Gly Trp Leu Ser Ala Tyr Ser Val Asp
180 185 190

Pro Leu Trp Gly Ser Gly Leu Val Asp Phe Ala Gly Ser Gly Val Val
195 200 205

His Val Thr Gly Gly Ile Thr Ala Leu Phe Ala Thr Met Val Leu Gly
210 215 220

Pro Arg Arg Gly Arg Phe His Asp Asp Leu Gly His Asp Leu Ala Arg
225 230 235 240

Pro Arg Glu Phe Gln Ala His Ser Pro Ala Leu Gln Met Leu Gly Thr
245 250 255

Phe Ile Leu Trp Phe Gly Phe Tyr Gly Phe Asn Ile Gly Ser Ala Leu
260 265 270

Ile Ser Thr Lys Gln Gly Ser Asp Glu Ala Ala Ala Leu Ala Gly Val
275 280 285

Asn Thr Thr Leu Ser Ala Ser Ala Ala Gly Ile Val Ala Leu Phe Ser
290 295 300

Asn Leu Trp Tyr Leu Glu Lys Thr Thr Gly Glu Pro Leu Phe Asp Leu
305 310 315 320

Lys Tyr Ala Met Asn Gly Ala Ile Cys Gly Leu Val Ala Ile Ser Gly
325 330 335

Gly Cys Gly Val Phe Glu Pro Trp Ala Ala Val Val Thr Gly Ala Val
340 345 350

Ala Gly Val Ile Tyr Leu Leu Gly Ser Arg Gly Leu Val Ser Met Arg
355 360 365

Leu Asp Asp Ala Val Asp Ala Ile Pro Val His Leu Cys Gly Gly Ala
370 375 380

Trp Gly Ile Leu Ala Val Gly Leu Phe Ala Ala Pro Glu Arg Leu Leu
385 390 395 400

Ser Val Tyr Gly Arg Asn Asn His Pro Gly Leu Val Tyr Ser Ile Arg
405 410 415

Glu Gly Asp Ile Asp Gly Val Leu Phe Gly Ile Gln Leu Ile Gly Leu
420 425 430

Met Phe Ile Met Gly Trp Val Met Ile Ile Met Leu Pro Phe Phe Val
435 440 445

Trp Leu Asn Trp Lys Gly Trp Phe Arg Ser Asp Pro Leu Glu Glu Ile
450 455 460

Leu Gly Leu Asp Leu Ser Tyr His Val Gly Leu Ala Leu His Thr Asn
465 470 475 480

Asn Val His Pro Glu Tyr Val Gly Ser Glu Lys Asp Val Val Asp Glu
485 490 495

Ile Ile Ser Thr Arg Gln Arg Lys Val Asn Gly Ser Thr Thr Thr Lys
500 505 510

Ala Thr Ser Gly Thr Glu Glu Leu Glu Tyr Ile Pro Glu Val Ser Asp
515 520 525

Glu Asp Leu Ser Glu Met Lys Glu Glu Cys Leu
530 535

<210> 17
 <211> 1557
 <212> ADN
 <213> *Cylindrotheca fusiformis*

5

<400> 17

atggctgagt ttgataacac gtttatcttg gatttttgca gtggtggaaa tgaatcgctg 60
 tctgacgttc aggccttgct tcaagttgct ggtcttgcaa atggcacaag tgcctctgcc 120
 gctggacttg ttgaggggat caacaccttc ttctacttt tgcaggagc tctggtcttt 180
 ctcatgcaag ctggcttcgc catgctgtgc gctggatccg tccgtcaaaa gaacgtcaag 240
 aatatcatgt tgaagaacat gtgggacgct tgtggtggtg ctattggttt ctggactatc 300
 gggatgcgt tgcctacgc tgataactcg tgggagaca aaaccttcac tggaggcaag 360
 aacttcttg tgaaccaatt ggatgagtct ggtggcgcat ggattggttt cttttccaa 420
 ttgcctttg ctgccactgc cgccactatt gtcgccgga ccggtgctga gcgctgtaag 480
 atgagtgcac accttgcta ctcaagtctc ctcaactggt ttgtctatcc tgcgttgtt 540
 cactccatct ggagtgtga tggatgggtg actgccttcc gtgatgatcc ttggaagggt 600
 gtcggtgtca ttgattcgc cggatctggt gttgtgcaca tgaccgggtg agccactgct 660
 ctgttgctg ctattgttct tggaccccggt aaggggacgt tctatgatga ggatggaaat 720
 gctctgaaa cccctgcaag ttcccagct cacagtgtgg ccttcaagt ccttggtacc 780
 ttcttcttt ggttcggatg gtatggattc aaccaggat ctgctcttgt cattgacaat 840
 gctgcgtccg cttctacctc agcactttgt gctgtcacta ccactcttc cgctgccagt 900
 ggttgtgtct gtgcgatgtt cactgatacc atcattgaaa tgatggcaac aggagaagca 960
 tcttacgatt tgaccatggc catgaacggt gctcttgag gacttgttc catcactgct 1020
 ggatgctccg ttgtacccc ttgggcttcg atcatcattg gtatcattgc tgggtgggtc 1080
 tacattgcat ttccaaact cctagtcaaa ttgaagatcg atgatgctgt tgatgccgtt 1140

cctgttcaact tcgccaatgg tatgtggggt gtcttggtg.tgggtttctt cgccgaaccc 1200
gacgccatgg ttactgccgg gtacaacgat gtcccaggag tcttctacaa gggagatggt 1260
aaacttctcc tgtgccaatt tgcgctatc atctggattt gcgcttggat cttctcttg 1320
atgactcctt tcttcgtcgt ctgaacatc ttgggtatgt tccgtgtcga tctcttgag 1380
gaagaagttg gtcttgatat tcccacatc cgtggagctg cctacgacat gaccagtgcc 1440
aagaaggaag atgtcgagga gctcatggaa caccgttctt ctaagcatgg aaaggttgag 1500
gtccccaagg aagtcaaaa ggaggacacc gcctaaggaa ttgacatcc cttctag 1557

<210> 18
<211> 511
5 <212> PRT
<213> *Cylindrotheca fusiformis*
<400> 18

Met Ala Glu Phe Asp Asn Thr Phe Ile Leu Asp Phe Cys Ser Gly Gly
1 5 10 15

Asn Glu Ser Ser Ser Asp Val Gln Ala Leu Cys Gln Val Ala Gly Leu
20 25 30

Ala Asn Gly Thr Ser Ala Ser Ala Ala Gly Leu Val Glu Gly Ile Asn
35 40 45

Thr Phe Phe Leu Leu Phe Ala Gly Ala Leu Val Phe Leu Met Gln Ala
50 55 60

Gly Phe Ala Met Leu Cys Ala Gly Ser Val Arg Gln Lys Asn Val Lys
65 70 75 80

Asn Ile Met Leu Lys Asn Met Leu Asp Ala Cys Gly Gly Ala Ile Gly
85 90 95

Phe Trp Thr Ile Gly Tyr Ala Phe Ala Tyr Ala Asp Asn Ser Ser Gly

10

100	105	110
Asp Lys Thr Phe Ile Gly Gly Lys Asn Phe Phe Val Asn Gln Leu Asp		
115	120	125
Glu Ser Gly Gly Ala Trp Ile Gly Phe Phe Phe Gln Phe Ala Phe Ala		
130	135	140
Ala Thr Ala Ala Thr Ile Val Ala Gly Thr Val Ala Glu Arg Cys Lys		
145	150	155
Met Ser Ala Tyr Leu Cys Tyr Ser Val Phe Leu Thr Gly Phe Val Tyr		
165	170	175
Pro Val Val Val His Ser Ile Trp Ser Ala Asp Gly Trp Leu Thr Ala		
180	185	190
Phe Arg Asp Asp Pro Trp Lys Gly Val Gly Val Ile Asp Phe Ala Gly		
195	200	205
Ser Gly Val Val His Met Thr Gly Gly Ala Thr Ala Leu Val Ala Ala		
210	215	220
Ile Val Leu Gly Pro Arg Lys Gly Arg Phe Tyr Asp Glu Asp Gly Asn		
225	230	235
Ala Leu Glu Thr Pro Ala Ser Phe Pro Ala His Ser Val Ala Leu Gln		
245	250	255
Val Leu Gly Thr Phe Leu Leu Trp Phe Gly Trp Tyr Gly Phe Asn Pro		
260	265	270
Gly Ser Ala Leu Val Ile Asp Asn Ala Ala Ser Ala Ser Thr Ser Ala		
275	280	285
Leu Cys Ala Val Thr Thr Thr Leu Ala Ala Ala Ser Gly Cys Val Cys		

290	295	300	
Ala Met Phe Thr Asp Thr Ile Ile Glu Met Met Ala Thr Gly Glu Ala			
305	310	315	320
Ser Tyr Asp Leu Thr Met Ala Met Asn Gly Ala Leu Gly Gly Leu Val			
325	330	335	
Ala Ile Thr Ala Gly Cys Ser Val Val Thr Pro Trp Ala Ser Ile Ile			
340	345	350	
Ile Gly Ile Ile Ala Gly Trp Val Tyr Ile Ala Phe Ser Lys Leu Leu			
355	360	365	
Val Lys Leu Lys Ile Asp Asp Ala Val Asp Ala Val Pro Val His Phe			
370	375	380	
Ala Asn Gly Met Trp Gly Val Leu Ala Val Gly Phe Phe Ala Glu Pro			
385	390	395	400
Asp Ala Met Val Thr Ala Gly Tyr Asn Asp Val Pro Gly Val Phe Tyr			
405	410	415	
Lys Gly Asp Gly Lys Leu Leu Leu Cys Gln Phe Val Ala Ile Ile Trp			
420	425	430	
Ile Cys Ala Trp Ile Phe Phe Leu Met Thr Pro Phe Phe Val Val Leu			
435	440	445	
Asn Ile Leu Gly Met Phe Arg Val Asp Pro Leu Glu Glu Glu Val Gly			
450	455	460	
Leu Asp Ile Ser His His Arg Gly Ala Ala Tyr Asp Met Thr Ser Ala			
465	470	475	480
Lys Lys Glu Asp Val Glu Glu Leu Met Glu His Arg Ser Ser Lys His			

485

490

495

Gly Lys Val Glu Val Pro Lys Glu Val Gln Lys Glu Asp Thr Ala
500 505 510

<210> 19

<211> 1622

<212> ADN

<213> Cylindrotheca fusiformis

<400> 19

atggctgagt ttgataacac gtttatcttg gatttttgca gtggtggaaa tgaatcgctg 60
tctgacgttc aggccttggt tcaagttgct ggtgtaagtt gttgccgat gttattctga 120
ccacagtata tcattttgcc ttgacttaca ccgtcgcgtc ttgatgtgt gcacttcagc 180
ttgcaaatgg cacaagtgcc tctgccgtg gactgttga gggtatcaac accttctcc 240
tactttcgc aggagctctg gtctttctca tgcaagctgg ctgcgccatg ctgtcgcgtg 300
gatccgtccg tcaaaagaac gtcaagaata tcatgttgaa gaacatgttg gacgcttggt 360
tggtgtctat tggttctgg actatcgggt atgcgttcgc ctacgctgat aactcgtcgg 420
gagacaaaac cticattgga ggcaagaact tctttgtgaa ccaattggat gactctgggt 480
gcgcatggat tggttcttt ttccaattcg cctttgctgc cactgccgcc actattgtcg 540
ccggaaccgt tgctgagcgc tgtaagatga gtgcatacct ttgctactca gtcttctca 600
ctggtttgt ctatcctgc gtgttact ccatctggag tgctgatga tgggtgactg 660
ccttccgtga tgatccttg aagggtgtcg gtgtcattga ttgcgccga tctggtgtg 720
tgacatgac cgggtggagcc actgctcttg ttgctgctat tgttcttga ccccgtaagg 780
gacgtttcta tgatgaggat ggaaatgctc ttgaaacccc tgcaagttc ccagctcaca 840
gtgtggccct tcaagtcctt ggtaccttcc ttctttggtt cggatggat ggattcaacc 900
caggatctgc tctgtcatt gacaatgctg cgtccgctc tacctcagca cttgtgctg 960
tcactaccac tcttgccgt gccagtgggt gtgtctgtgc gatgttact gataccatca 1020
ttgaaatgat ggcaacagga gaagcatctt acgatttgac catggccatg aacgggtgctc 1080

ttggaggact tgttgccatc actgctggat gctccgtgt cacccttgg gcttcgatca 1140
 tcattggtat cattgctggg tgggtctaca ttgcatttc caaactccta gtcaaattga 1200
 agatcgatga tgctgtgat gccgttctg ttacttcgc caatgggatg tggggtgtct 1260
 tggctgtggg ttcttcgcc gaacccgacg ccatgggtac tgccgggtac aacgatgtcc 1320
 caggagtctt ctacaagga gatggtaaac ttctctgtg ccaattgtc gctatcatct 1380
 ggatttcgc ttgatctc ttctgatga ctctttctt cgtcgtctg aacatcttg 1440
 gtatgtccg tgcgatcct ctgaggaag aagtgggtct tgatattcc caccatcgtg 1500
 gagctgccta cgacatgacc agtgccaaga aggaagatgt cgaggagctc atggaacacc 1560
 gttcttctaa gcatggaaag gttgaggtcc ccaaggaagt tcaaaaggag gacaccgcct 1620

aa

1622

<210> 20
 <211> 450
 <212> PRT
 <213> *Cylindrotheca fusiformis*

<400> 20

Met Gln Ala Gly Phe Ala Met Leu Cys Ala Gly Ser Val Arg Gln Lys
 1 5 10 15

Asn Val Lys Asn Ile Met Leu Lys Asn Met Leu Asp Ala Cys Gly Gly
 20 25 30

Ala Ile Gly Phe Trp Thr Ile Gly Tyr Ala Phe Ala Tyr Ala Asp Asn
 35 40 45

Ser Ser Gly Asp Lys Thr Phe Ile Gly Gly Lys Asn Phe Phe Val Asn
 50 55 60

Gln Leu Asp Glu Ser Gly Gly Ala Trp Ile Gly Phe Phe Phe Gln Phe
 65 70 75 80

Ala Phe Ala Ala Thr Ala Ala Thr Ile Val Ala Gly Thr Val Ala Glu
85 90 95

Arg Cys Lys Met Ser Ala Tyr Leu Cys Tyr Ser Val Phe Leu Thr Gly
100 105 110

Phe Val Tyr Pro Val Val Val His Ser Ile Trp Ser Ala Asp Gly Trp
115 120 125

Leu Thr Ala Phe Arg Asp Asp Pro Trp Lys Gly Val Gly Val Ile Asp
130 135 140

Phe Ala Gly Ser Gly Val Val His Met Thr Gly Gly Ala Thr Ala Leu
145 150 155 160

Val Ala Ala Ile Val Leu Gly Pro Arg Lys Gly Arg Phe Tyr Asp Glu
165 170 175

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

Ala Leu Gln Val Leu Gly Thr Phe Leu Leu Trp Phe Gly Trp Tyr Gly
195 200 205

Phe Asn Pro Gly Ser Ala Leu Val Ile Asp Asn Ala Ala Ser Ala Ser
210 215 220

Thr Ser Ala Leu Cys Ala Val Thr Thr Thr Leu Ala Ala Ala Ser Gly
225 230 235 240

Cys Val Cys Ala Met Phe Thr Asp Thr Ile Ile Glu Met Met Ala Thr
245 250 255

Gly Glu Ala Ser Tyr Asp Leu Thr Met Ala Met Asn Gly Ala Leu Gly
260 265 270

Gly Leu Val Ala Ile Thr Ala Gly Cys Ser Val Val Thr Pro Trp Ala
275 280 285

Ser Ile Ile Ile Gly Ile Ile Ala Gly Trp Val Tyr Ile Ala Phe Ser
290 295 300

Lys Leu Leu Val Lys Leu Lys Ile Asp Asp Ala Val Asp Ala Val Pro
305 310 315 320

Val His Phe Ala Asn Gly Met Trp Gly Val Leu Ala Val Gly Phe Phe
325 330 335

Ala Glu Pro Asp Ala Met Val Thr Ala Gly Tyr Asn Asp Val Pro Gly
340 345 350

Val Phe Tyr Lys Gly Asp Gly Lys Leu Leu Leu Cys Gln Phe Val Ala
355 360 365

Ile Ile Trp Ile Cys Ala Trp Ile Phe Phe Leu Met Thr Pro Phe Phe
370 375 380

Val Val Leu Asn Ile Leu Gly Met Phe Arg Val Asp Pro Leu Glu Glu
385 390 395 400

Glu Val Gly Leu Asp Ile Ser His His Arg Gly Ala Ala Tyr Asp Met
405 410 415

Thr Ser Ala Lys Lys Glu Asp Val Glu Glu Leu Met Glu His Arg Ser
420 425 430

Ser Lys His Gly Lys Val Glu Val Pro Lys Glu Val Gln Lys Glu Asp
435 440 445

Thr Ala
450
<210> 21

<211> 1536
 <212> ADN
 <213> *Cylindrotheca fusiformis*

5 <400> 21

atggctgagt tcgataacac tttatcttg gaatttgca gtggaggaaa tgaatcgctg 60
 tctgacgttc aggtcttg ccaagttgcc agtctcgcaa acggcaccag tgcctctgcc 120
 ggtggactta ctgagggtat caataccttc tttctacttt ttgctggagc tctagtcttt 180
 atcatgcaag ctggattcgc catgctgtgt gctggatccg tccgtcaaaa gaacgtcaag 240
 aatatcatgt tgaaaaacat gttggacgct tgtgggtgtg ctattggttt ctggactatc 300
 gggatgcct tcgcctacgc tgataactcg tctggaaata aaacattcat tggaggcaag 360
 aacttctttg tgaaccaact ggatgagtct ggaggtgcat ggattggttt ctcttccaa 420
 ttgcctttg ctgccactgc cgccactatt gtcgcgggaa ccgtgccga gcgctgcaag 480
 atgagtgcgt acctatgcta ctcaatcttc ctactgggt ttgtctatcc tgcgttgtt 540
 cactccatct ggagtgtga tggatgggtg actgccttgc gtgatgatcc ttggcaggg 600
 gtcggtgtca tcgattcgc cggatctggt gttgtgcata tgtcggtgg agccactgct 660
 ctgttgctg ctattgttct tggaccccggt aagggtacgt tctatgatga ggatggaaat 720
 gctcttgaag cccctgcaag ttcccagct cacagtgtgg ccctcaagt cctcggtacc 780
 ttcttcttt ggttcggatg gtagggatc aaccaggat ctgctctgt cattgacaat 840
 gctgcgtccg ctctacctc agcactttgt gctgtcacta ccactctgc cgctgccagt 900
 ggttgtgtca ccgcgatgtt caccgatacc ctattgaaa tgatggctac cggagaagca 960
 tcttatgatt tgaccatggc catgaacggt gcacttgcgg ggcttgttc catcactgcc 1020
 ggatgctccg ttgtcactcc ttgggcttct ctcatcattg gtattattgg cgggtgggtc 1080
 taccttggac ttctaaact tctaataaaa ttgaagatcg atgatgctgt tgatgccgtc 1140
 cctgttact tcggtaatgg tatgtgggtg gtcttggctg ttggcttct tgcgaaccc 1200
 gacgccatgg ttactgccgg gtacaacgat gtcccaggag tgttctacaa aggagatggt 1260

tctctctcc tgtgccaatt cgtcgctatc gttggggtt gcgctgggt cttctcttg 1320
 atgactcctt tcttcgtcgt cttgaacatt ttgggaatgt tccgtgtcga tcctcttgag 1380
 gaagaagttg gtcttgatat ttcccaccac cgtggatctg cctacgacat gactactgcc 1440
 aagaaggaag atgtcgaaga gctcatggaa caccgttctt cgaagcacgg gaagggtgag 1500
 atccccaagg aagttcagaa ggaggacacc gcctaa 1536

<210> 22
 <211> 511
 <212> PRT
 <213> *Cylindrotheca fusiformis*
 <400> 22

Met Ala Glu Phe Asp Asn Thr Phe Ile Leu Glu Phe Cys Ser Gly Gly
 1 5 10 15

Asn Glu Ser Ser Ser Asp Val Gln Val Leu Cys Gln Val Ala Ser Leu
 20 25 30

Ala Asn Gly Thr Ser Ala Ser Ala Gly Gly Leu Thr Glu Gly Ile Asn
 35 40 45

Thr Phe Phe Leu Leu Phe Ala Gly Ala Leu Val Phe Ile Met Gln Ala
 50 55 60

Gly Phe Ala Met Leu Cys Ala Gly Ser Val Arg Gln Lys Asn Val Lys
 65 70 75 80

Asn Ile Met Leu Lys Asn Met Leu Asp Ala Cys Gly Gly Ala Ile Gly
 85 90 95

Phe Trp Thr Ile Gly Tyr Ala Phe Ala Tyr Ala Asp Asn Ser Ser Gly
 100 105 110

Asn Lys Thr Phe Ile Gly Gly Lys Asn Phe Phe Val Asn Gln Leu Asp

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

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

500 505 510

<210> 23
 <211> 1563
 <212> ADN
 <213> *Thalassiosira pseudonana*

<400> 23

atggccgaac caacaacaac cataggcgac ttcaacgtca ctgcctggtg cggcgacgac 60
 gctgtagcaa cctacgaagg acaatccgtc gagaacggta tctgtgcagc ctacgcctac 120
 accgacgaaa ccaacaccgg tcttgatgtg ttctacctcc tctcgccgc tgccatggtc 180
 ttctcatgc aggctggatt cgccatgctc tgtgctggat ctgtgaggca gaagaatgtg 240
 aagaatatca tgcttaagaa catttggat gcttgtggtg gagctcttgg atttggctct 300
 gtgggatttg cgttgccta cggaggatct ggaccggaga agaagggtt catcggtaac 360
 gagggattct tcttggga ctttacaact ggaggcgatt tgatcggtg gtcttccag 420
 ttgcctttg ctgccaccgc cgccacaatt gtggctggaa ccgtagccga gcgttgcaag 480
 ttcgaggcct acctctgcta ttctctcatg ctaccgggt tctctaccc cgtgattgtg 540
 tactccatct ggtcttctc tgggttctc accgccttca acgacgaccc tgccttggga 600
 tgtggtatgc atgatttgc cggatcggga gttgttaca tgacgggagg catcactgcc 660
 ctttgggccc ccaagattct cggacctcgt atcggacgct tctacgatgc tgatggcaat 720
 gagcttcccg agccagttag ctccctccc cactccgtgg ccttcaagt tcttggtact 780
 ttcatcctt ggttgggtg gtacggattc aaccccggtt ccactcttct catcagtaac 840
 acggccgcag ctgatgtgtc tgcccttgt gccgtcacca ccaccattgc cgccgcttcg 900
 ggctctgtct ctgccatgtt cactgatatg ttcttgagc gcagaaagac cggagagact 960
 atgtacgaca ttaccatgtg tatgaacggt gctcttctg gattggtggg catcaccgct 1020
 ggatgctcaa tctgtagcc ttgggtgcc ttgtcattg gaattgtgc cggatggact 1080
 tacatcttct ggtccagtct ccttgtgaag cttaagattg atgatgccgt cgaigccatt 1140
 cctgttcaat ttgaaacgg aatgtggggc tgcattgccg ttggactctt tgccgagcct 1200

acccgtag ccaacgccta cagtgacat ggacactatg gatggttcta ctcatggggt 1260
 gctggaaacg ccgatgccca cctttggct gctcaagtct gtggtgttct ctggatcatt 1320
 ggatggggtt ccgtcattat gatcccatac tcatcctgc tcaacgtcct gggttgttc 1380
 cgtgtggatg cctcgaaga agaagttggt ttgatattc cccaccacaa gggagctgcc 1440
 tacgatattg ccggaccttc tgaagctgcc gccgagaagt ttgagatctc taggagtcag 1500
 cgtaagcttg agatccctgt ggatgtcgcg cctgctactg ctctgccga ggatgccgcg 1560
 taa 1563

<210> 24
 <211> 520
 <212> PRT
 <213> *Thalassiosira pseudonana*
 <400> 24

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

100	105	110
Glu Lys Lys Gly Phe Ile Gly Asn Glu Gly Phe Phe Leu Gly Asp Phe		
115	120	125
Thr Thr Gly Gly Asp Leu Ile Gly Trp Phe Phe Gln Phe Ala Phe Ala		
130	135	140
Ala Thr Ala Ala Thr Ile Val Ala Gly Thr Val Ala Glu Arg Cys Lys		
145	150	155
Phe Glu Ala Tyr Leu Cys Tyr Ser Leu Met Leu Thr Gly Phe Val Tyr		
165	170	175
Pro Val Ile Val Tyr Ser Ile Trp Ser Ser Ser Gly Phe Leu Thr Ala		
180	185	190
Phe Asn Asp Asp Pro Ala Phe Gly Cys Gly Met His Asp Phe Ala Gly		
195	200	205
Ser Gly Val Val His Met Thr Gly Gly Ile Thr Ala Leu Trp Ala Ala		
210	215	220
Lys Ile Leu Gly Pro Arg Ile Gly Arg Phe Tyr Asp Ala Asp Gly Asn		
225	230	235
Glu Leu Pro Glu Pro Val Ser Phe Pro Pro His Ser Val Ala Leu Gln		
245	250	255
Val Leu Gly Thr Phe Ile Leu Trp Val Gly Trp Tyr Gly Phe Asn Pro		
260	265	270
Gly Ser Thr Leu Leu Ile Ser Asn Thr Ala Ala Ala Asp Val Ser Ala		
275	280	285
Leu Cys Ala Val Thr Thr Thr Ile Ala Ala Ala Ser Gly Ser Val Ser		

290	295	300
Ala Met Phe Thr Asp Met Phe Leu Glu Arg Arg Lys Thr Gly Glu Thr		
305	310	315 320
Met Tyr Asp Ile Thr Met Cys Met Asn Gly Ala Leu Ser Gly Leu Val		
	325	330 335
Gly Ile Thr Ala Gly Cys Ser Ile Val Glu Pro Trp Ala Ala Phe Val		
	340	345 350
Ile Gly Ile Val Ala Gly Trp Thr Tyr Ile Phe Trp Ser Ser Leu Leu		
	355	360 365
Val Lys Leu Lys Ile Asp Asp Ala Val Asp Ala Ile Pro Val His Phe		
	370	375 380
Gly Asn Gly Met Trp Gly Cys Ile Ala Val Gly Leu Phe Ala Glu Pro		
385	390	395 400
Thr Arg Val Ala Asn Ala Tyr Ser Asp His Gly His Tyr Gly Trp Phe		
	405	410 415
Tyr Ser Trp Gly Ala Gly Asn Ala Asp Ala His Leu Leu Ala Ala Gln		
	420	425 430
Val Cys Gly Val Leu Trp Ile Ile Gly Trp Val Ser Val Ile Met Ile		
	435	440 445
Pro Tyr Phe Ile Leu Leu Asn Val Leu Gly Leu Phe Arg Val Asp Ala		
	450	455 460
Leu Glu Glu Glu Val Gly Leu Asp Ile Ser His His Lys Gly Ala Ala		
465	470	475 480
Tyr Asp Met Ser Gly Pro Ser Glu Ala Ala Ala Glu Lys Phe Glu Ile		

485 490 495

Ser Arg Ser Gln Arg Lys Leu Glu Ile Pro Val Asp Val Ala Pro Ala
500 505 510

Thr Ala Pro Ala Glu Asp Ala Ala
515 520

<210> 25
<211> 1608
<212> ADN
<213> *Thalassiosira pseudonana*

<400> 25

atgtctgaag atccatcaat cttgaagtg tgtactggcc agctcggaac agatcttacc 60
gtcgagcttt tacaatgtgt ctctgatgga gctgaaagcg caaaggatga tgcattcagt 120
ggcgtcaact ccttctacct catctttgca ggagccctcg tttcttcat gcaagtcggt 180
ttcgccatgc tctgtgctgg atccatccgt gagaagaacg tcaagaatgt attgctttgg 240
aatctcctcg attctgccgg tggcgcttt ggtttctgga gtattggta tgcatttct 300
tatggtggtg atgatattac caagggaaag accttcatcg gaaacgctga ctcttcttg 360
agtggagaaa ctgatatgga gtttgggtc ttccaatacg ccttgcgtg tgctctctcc 420
tccattgttg ctggaacat tgctgagcgc accaagatga tggcctact gtgctactca 480
atcttcttt gtggattcgt ctaccacgctc tggctcatg cctttggc tcagaatgga 540
ttctctctg ccttcgctgc tgaaccttg tgggggtcgg gcgtattga cttgcagga 600
tcgggaccag ttacatgtg tggaggagt gctgctcttg tcatggctat tattcttga 660
cctcgtaggg gacgtttcta tgacgacgat ggtgtcgtat tggatgagcc aaagtccatg 720
ggacctcact ccgtcacttt gcaattcctc ggaaccttg ctcttgggt cggtatggtat 780
ggattcaacc ctggtagctc catcttgatt gcgtcagctg cctctggcga cgttgcacg 840
cttgcgtccg tcaacactac gctcggatct gctgctggag cactttctgg tatgttcacc 900
tccacgatcg tcgacgaaag aaagactgga gtgtatactt gggatactac tgctgcaatg 960

aacggatgcc tcaccgggtt ggttgctatt actgctgggt gcgctaccgt cgagccttgg 1020
 gctgccttcg tcattgggct cactgctggt tgggtgtacc ttgctgcac tgctcttatg 1080
 cttcgcttca agattgatga tgctgtcgat gccatccccg ttacatggt cggaggatca 1140
 tggggagttt ttgcactgg tctttcacc agtcctcgcc gtcttattac tgcatacgga 1200
 aatgacaata atgttggtg gtctacgaa tggggacgtg gaagtggaaa cttcactctc 1260
 ctggctgcc agctcgtctc gattctctt gtctgggat ggtctgctg catctttgct 1320
 ccattctgct tggcactcaa aacctcaac tggctccgca ttgacctct cgaggaggag 1380
 gtgggatgg atatcagtcg ccataagggg cctgcctacg agtcggaggg atctgctcat 1440
 tctgatgcta tcgagaagtt gagtgcctcc cgtcgtgata ttatgaatgc ctctggaagt 1500
 ggaaggggaa aaagtttcag caggtcgact cccacgaaag ctaacgagga gcccaaaatc 1560
 gaagcaactg aggatgcagg ggcacccgct ggagaagcaa ctgcttag 1608

<210> 26

<211> 535

<212> PRT

<213> *Thalassiosira pseudonana*

<400> 26

Met Ser Glu Asp Pro Ser Ile Phe Glu Val Cys Thr Gly Gln Leu Gly
 1 5 10 15

Thr Asp Leu Thr Val Glu Leu Leu Gln Cys Val Ser Asp Gly Ala Glu
 20 25 30

Ser Ala Lys Asp Asp Val Ile Ser Gly Val Asn Ser Phe Tyr Leu Ile
 35 40 45

Phe Ala Gly Ala Leu Val Phe Phe Met Gln Val Gly Phe Ala Met Leu
 50 55 60

Cys Ala Gly Ser Ile Arg Glu Lys Asn Val Lys Asn Val Leu Leu Trp

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

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

450

455

460

Ile Ser Arg His Lys Gly Pro Ala Tyr Glu Ser Glu Gly Ser Ala His
465 470 475 480

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

Ala Ser Gly Ser Gly Arg Gly Lys Ser Phe Ser Arg Ser Thr Pro Thr
500 505 510

Lys Ala Asn Glu Glu Pro Lys Ile Glu Ala Thr Glu Asp Ala Gly Ala
515 520 525

Pro Ala Gly Glu Ala Thr Ala
530 535

<210> 27

<211> 1632

<212> ADN

<213> Thalassiosira pseudonana

<400> 27

atgtttcagg tatcaagagc tggacacgtg tcagtgtatg aggtctgcaa atccttcgtc 60
aaccagaag atttcaagc agatcagttt gatgcaatgc tccaatgtgt tggggaatcc 120
aacggcaaga gtatagacgc ttcttcctt atatacgcat catccctcgt ctcttcattg 180
caagccggct tcgcatgct ttgtgctgga tgcgttcaac acaagaatgt tcagaacagt 240
atgtgaaga acctcctaga tgcattgcgt gcagcccttg gcttctattc cgtcgggtat 300
gcatttcctt acggtggcat ggactattca gatccaaaca agacatttat cggcacagag 360
aacttttct tgatgggagt ggacgatttc atgtttggc tattccaatt tgcattcgtc 420
gcaagtgcgt ctaccattgt ggcaggaacg ttggctgaac ggtgtcaaata gacggcatac 480
ttgtgctact cagtggcagt gactggattt gtatatccag tagttgtaca ttcggtttgg 540
tctccgcagg gtttctgtg cggacaggct gtgagcccgt tatttgaggt tgggttagta 600

gactttgcag gatcatcggg tgtgcatttg actggagggt gcattgcact cattgccacg 660
 tatattctag gccaaggcg agggagattc tatgatcaca gaggagaacc tcttgagaca 720
 ccagtcgagt ttccgggtca ttcagctgca ctcaaagtc tcgggtgcctt catattatgg 780
 ttggatggg atggcttcaa tactggatca actctttcga tcaccggccc tggccaacat 840
 caagtcgca gcctttagc tgtaaacaca accctcgcgg cggcctctgc ttgtgtggct 900
 tcccttcttg ccagttatta tglcatcgaa cgaaagactg gcgaaggtag attctcgctt 960
 tcttcagcaa tgaatgggtg cttaggagga ttagtcagta tcaactgggtg atgtgcagtg 1020
 gtggagcctt gggcagccgt cgtaatcgga tcatagcag ggttgttga tctcttcacg 1080
 tcaaagctat tgattcgtt gcgaatagac gatgcagtcg atgctattcc ggtccacttg 1140
 tctaacggaa tatggggtac ggttgcgggtg gggttgttcg catcgtcgaa tcgtttgcag 1200
 ttggcttttg gaaaagtcgc tgatactggg gtgttcattg gtggaaccgg taaactgttg 1260
 ggatgtcaaa taattgggtg ctctttgtg ctggatgga ttcttcat tatgattccg 1320
 ttctctgct tcttcaacta catgggatgg cttcgatctg agtcaattga tgaagtagaa 1380
 gggcttgatt ccaagtatca tggattgcga aacaagatg agcatagaca tgacgaagaa 1440
 gaagacaata ctccatcaca ctacggcgaa ggcaattgca ggctgaggcg tagcattttg 1500
 cgccatgagg aaagaatgag acaagaggac tcggcaatac ccgcaacttt ggtatcaagt 1560
 gatgataggt tcacgtgtga ctcaggagga aatagcactt caatgctatc tactgctaag 1620
 caatacacct ag 1632

<210> 28
 <211> 543
 5 <212> PRT
 <213> *Thalassiosira pseudonana*
 <400> 28

Met Phe Gln Val Ser Arg Ala Gly His Val Ser Val Tyr Glu Val Cys
 1 5 10 15

10

Lys Ser Phe Val Asn Pro Glu Asp Ser Gln Ala Asp Gln Phe Asp Ala
20 25 30

Met Leu Gln Cys Val Gly Glu Ser Asn Gly Lys Ser Ile Asp Ala Phe
35 40 45

Phe Leu Ile Tyr Ala Ser Ser Leu Val Phe Phe Met Gln Ala Gly Phe
50 55 60

Ala Met Leu Cys Ala Gly Cys Val Gln His Lys Asn Val Gln Asn Ser
65 70 75 80

Met Leu Lys Asn Leu Leu Asp Ala Cys Gly Ala Ala Leu Gly Phe Tyr
85 90 95

Ser Val Gly Tyr Ala Phe Ala Tyr Gly Gly Met Asp Tyr Ser Asp Pro
100 105 110

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

Asp Phe Met Phe Trp Leu Phe Gln Phe Ala Phe Ala Ala Ser Ala Ala
130 135 140

Thr Ile Val Ala Gly Thr Leu Ala Glu Arg Cys Gln Met Thr Ala Tyr
145 150 155 160

Leu Cys Tyr Ser Val Ala Val Thr Gly Phe Val Tyr Pro Val Val Val
165 170 175

His Ser Val Trp Ser Pro Gln Gly Phe Leu Cys Gly Gln Ala Val Ser
180 185 190

Pro Leu Phe Gly Val Gly Val Val Asp Phe Ala Gly Ser Ser Val Val
195 200 205

His Leu Thr Gly Gly Cys Ile Ala Leu Ile Ala Thr Tyr Ile Leu Gly
210 215 220

Pro Arg Arg Gly Arg Phe Tyr Asp His Arg Gly Glu Pro Leu Glu Thr
225 230 235 240

Pro Val Glu Phe Pro Gly His Ser Ala Ala Leu Gln Met Leu Gly Ala
245 250 255

Phe Ile Leu Trp Phe Gly Trp Tyr Gly Phe Asn Thr Gly Ser Thr Leu
260 265 270

Ser Ile Thr Gly Pro Gly Gln His Gln Val Val Ser Leu Val Ala Val
275 280 285

Asn Thr Thr Leu Ala Ala Ala Ser Ala Cys Val Ala Ser Leu Leu Ala
290 295 300

Ser Tyr Tyr Val Ile Glu Arg Lys Thr Gly Glu Gly Thr Phe Ser Leu
305 310 315 320

Ser Ser Ala Met Asn Gly Cys Leu Gly Gly Leu Val Ser Ile Thr Gly
325 330 335

Gly Cys Ala Val Val Glu Pro Trp Ala Ala Val Val Ile Gly Phe Ile
340 345 350

Ala Gly Leu Leu Tyr Leu Phe Thr Ser Lys Leu Leu Ile Arg Leu Arg
355 360 365

Ile Asp Asp Ala Val Asp Ala Ile Pro Val His Leu Ser Asn Gly Ile
370 375 380

Trp Gly Thr Val Ala Val Gly Leu Phe Ala Ser Ser Asn Arg Leu Gln
385 390 395 400

Leu Ala Phe Gly Lys Val Ala Asp Thr Gly Val Phe Met Gly Gly Thr
405 410 415

Gly Lys Leu Leu Gly Cys Gln Ile Ile Gly Val Phe Phe Val Leu Gly
420 425 430

Trp Ile Ser Phe Ile Met Ile Pro Phe Phe Cys Phe Leu His Tyr Met
435 440 445

Gly Trp Leu Arg Ser Glu Ser Ile Asp Glu Val Glu Gly Leu Asp Ser
450 455 460

Lys Tyr His Gly Leu Arg Asn Lys Asp Glu His Arg His Asp Glu Glu
465 470 475 480

Glu Asp Asn Thr Pro Ser His Tyr Gly Glu Gly Asn Cys Arg Leu Arg
485 490 495

Arg Ser Ile Leu Arg His Glu Glu Arg Met Arg Gln Glu Asp Ser Ala
500 505 510

Ile Pro Ala Thr Leu Val Ser Ser Asp Asp Arg Phe Thr Cys Asp Ser
515 520 525

Gly Gly Asn Ser Thr Ser Met Leu Ser Thr Ala Lys Gln Tyr Thr
530 535 540

<210> 29
<211> 1572
<212> ADN
<213> *Thalassiosira pseudonana*

<400> 29

atgtcctcgt cggtagcgac gagcctgtac gaggcttgca aatctactca atccaactct 60

acgttttcca actctacgct cgaatgatgca ctctctcgcc aggaacagat ctccggttgc 120

atttcggaat cgaacgcaa tagtattgat acgttctttc tgctctatgc ctcacgctg 180

gtattcttta tgcaagccgg attcgcgatg ctctgtgcag gatcagtgag aaagaagaat 240
 gtgactaata ccatgttgaa gaacctctc gatgcttgcg gtgccgcact tgggttctat 300
 tctgtgggat acgcctttgc atacggagga tcagtagacg ccgggaagaa gacgtttata 360
 ggtatgagca acttcttct acaggacgtt gacaactata tgtttggct ctttcagttt 420
 gcattcgctg caacatcagc tacaatcgta gcaggaacat tggctgaaag atgtcaaatg 480
 acagcttacc ttgctactc aattgcattg acgggatttg tctaccagt gtgcgctcat 540
 tcaatatgga gtcagcaggg gttcttatca gctactgctc aagatccatt atgggggtaca 600
 gggttcattg actttgcggg atcaacagta gtacatctga cgggtggatt tacagctttg 660
 attgcgacat atcttctagg gccacgcaga ggacggttct acgatgcgaa aggcaagcag 720
 ttggaagtgc caaatccaat gcctggccat tcagctgcac ttcatatgct tggatatctc 780
 attttatggt ttggttgga tgggttcatt gttggaicag cgataactat catcgggtccg 840
 aatcaagaca agatcatctc tacctctgca gtgaatacga cactctctgc agcgatca 900
 tgcttctctg ccttactcgt caactacgct attgtcgaga ggcaatcggg agaaggagaa 960
 ttcatcttc tcgctgcaat gaacggatgc ttgagtggat tggtagcgat aacagggtgga 1020
 tgcgcagtga tagcaccatg ggcagcaatt atcgtcggac tttcgtctgg cctctgtac 1080
 ttattacgt caaaggattt ggtacgagtt cgaatcgacg atgcagtaga agccattcct 1140
 gttcacatga ctaacgggat atggggtagt ttgcagttg gactatttgc cgctccatcg 1200
 gagctgcaat tggatatacg aaaagcaa atcacgttgac tattctattc ttggcatcaa 1260
 gggagtggag atgggacgtt actcgggtgc caatgcttgg gcattctgtt ttagtgggc 1320
 tgggtgttct gtctcatgtc tcccttctc ctgtttctga actacaaagg ctgggtcaga 1380
 gccgacgtcc tcaacgagat tgctggctta gatttgagtt atcatgatgg agtggatatg 1440
 gaattggatga ctcagatagc caatcaaagg aagaactgc acgtcaacag caggaatcgc 1500
 ttagttcga atagtgcag tctcatcat acggcaacac atgtcgatac ctcaactgat 1560
 gcatcacttt ga 1572

<210> 30
 <211> 511
 <212> PRT
 <213> Thalassiosira pseudonana

5

<400> 30

Met Ser Ser Ser Val Arg Thr Ser Leu Tyr Glu Ala Cys Lys Ser Thr
 1 5 10 15

Gln Ser Asn Ser Thr Phe Ser Asn Ser Thr Leu Asp Asp Ala Leu Ser
 20 25 30

Arg Gln Glu Gln Ile Phe Arg Cys Ile Ser Glu Ser Asn Ala Asn Ser
 35 40 45

Ile Asp Thr Phe Phe Leu Leu Tyr Ala Ser Ser Leu Val Phe Phe Met
 50 55 60

Gln Ala Gly Phe Ala Met Leu Cys Ala Gly Ser Val Arg Lys Lys Asn
 65 70 75 80

Val Thr Asn Thr Met Leu Lys Asn Leu Leu Asp Ala Cys Gly Ala Ala
 85 90 95

Leu Gly Phe Tyr Ser Val Gly Tyr Ala Phe Ala Tyr Gly Gly Ser Val
 100 105 110

Asp Ala Gly Lys Lys Thr Phe Ile Gly Met Ser Asn Phe Phe Leu Gln
 115 120 125

Asp Val Asp Asn Tyr Met Phe Trp Leu Phe Gln Phe Ala Phe Ala Ala
 130 135 140

Thr Ser Ala Thr Ile Val Ala Gly Thr Leu Ala Glu Arg Cys Gln Met
 145 150 155 160

Thr Ala Tyr Leu Cys Tyr Ser Ile Ala Leu Thr Gly Phe Val Tyr Pro
165 170 175

Val Val Ala His Ser Ile Trp Ser Gln Gln Gly Phe Leu Ser Ala Thr
180 185 190

Ala Gln Asp Pro Leu Trp Gly Thr Gly Phe Ile Asp Phe Ala Gly Ser
195 200 205

Thr Val Val His Leu Thr Gly Gly Phe Thr Ala Leu Ile Ala Thr Tyr
210 215 220

Leu Leu Gly Pro Arg Arg Gly Arg Phe Tyr Asp Ala Lys Gly Lys Gln
225 230 235 240

Leu Glu Val Pro Asn Pro Met Pro Gly His Ser Ala Ala Leu Gln Met
245 250 255

Leu Gly Ile Phe Ile Leu Trp Phe Gly Trp Tyr Gly Phe Ile Val Gly
260 265 270

Ser Ala Ile Thr Ile Ile Gly Pro Asn Gln Asp Lys Ile Ile Ser Thr
275 280 285

Ser Ala Val Asn Thr Thr Leu Ser Ala Ala Ser Ser Cys Phe Ser Ala
290 295 300

Leu Leu Val Asn Tyr Val Ile Val Glu Arg Gln Ser Gly Glu Gly Glu
305 310 315 320

Phe Ser Leu Leu Ala Ala Met Asn Gly Cys Leu Ser Gly Leu Val Ala
325 330 335

Ile Thr Gly Gly Cys Ala Val Ile Ala Pro Trp Ala Ala Ile Ile Val
340 345 350

Gly Leu Phe Ala Gly Leu Leu Tyr Leu Phe Thr Ser Lys Val Leu Val
355 360 365

Arg Val Arg Ile Asp Asp Ala Val Glu Ala Ile Pro Val His Met Thr
370 375 380

Asn Gly Ile Trp Gly Ser Phe Ala Val Gly Leu Phe Ala Ala Pro Ser
385 390 395 400

Glu Leu Gln Leu Val Tyr Gly Lys Ala Asn His Val Gly Leu Phe Tyr
405 410 415

Ser Trp His Gln Gly Ser Gly Asp Gly Thr Leu Leu Gly Val Gln Cys
420 425 430

Leu Gly Ile Leu Phe Val Val Gly Trp Val Phe Cys Leu Met Ser Pro
435 440 445

Phe Phe Leu Phe Leu Asn Tyr Lys Gly Trp Phe Arg Ala Asp Val Leu
450 455 460

Asn Glu Ile Ala Gly Leu Asp Leu Ser Tyr His Asp Gly Val Asp Met
465 470 475 480

Glu Leu Val Thr Gln Ile Arg Asn Gln Arg Lys Asn Leu His Val Asn
485 490 495

Ser Arg Asn Arg Phe Ser Ser Asn Ser Ala Cys Pro His His Thr
500 505 510

<210> 31

<211> 1791

5 <212> ADN

<213> Thalassiosira pseudonana

<400> 31

10 atggcatcat caaccaccac agacacctat caaacatgcc tcagcgatct ctggcgacg 60

tcgtctaata gacccctgcc caccaccgac gccctccttc agtgcatttc ctctccttc 120
gacgctcaaa cggcctctac acatgcttca atcaacacct ttctcctct ctatgcagcc 180
accctcgtct tctcatgca agccggcttt gccatggta gcgctggatg tgtaggacg 240
aataatgttc agaatacgt gctgaagaat ctgttgacg cctgcgggtc cgctcttga 300
ttctacactg ttggatacgc ctctgcatgg ggaggatcgt tggatacggc taccaccgag 360
aggacgttca tcggtacaca gaacttctt ttgatggatg tggatagctc tcaggattca 420
tttgggtgt tcaattggc ttctgctcg gcatctgcaa cgatttgcc tgaacgtg 480
gctgaacgtt gtcaaattgt tgcctatct gcatactcca tgacattggc aggattgtc 540
taccagtcg tcgtacacag tatctggagt ccgagtggtat tctgagtgc tactcgtgag 600
accgatctct tctggatgt ggaatgatt gacttcgag gatcaaccgt cgtgcattg 660
acgggagggg tgactgcgtt gattgctacg attgtgctgg gaccgaggac gggaagggtc 720
tatgattga gaggaaatcc gtgaaagta ccaaaggagt tgcaggaca ttattggct 780
ttgcaaattg tgggggtgt catctgttg tttggatgt atggcttcaa cgctggatcg 840
atctcaaca tcaccaacga tctaatcat acaatcgtca gtacatactgc catcaacaca 900
actcttcag ctctgctgg atccatcatg actctcttc tcagtaccgt cgtagccgaa 960
agatttacgg gggagatagt gtttagtcta tcttatgcca tgaatgggtg ttgagtga 1020
ttggtggcaa tcacagctgg ctgctcagt gtagaacact gggctgcaat cataatcggg 1080
ctttaggag gggcattgta ttggcatgc tccaagttct tggtaagaa acgtatcat 1140
gatgcagtgt atggatccc tgcacttg attaatggaa tctggggaac gttgagtgc 1200
ggtctcttcg ctgtgctga gttgttgag caagtgtatg gaaggggtga tcacgtgtg 1260
tggtttaca gttggggaca aggatcagca gacgccaagt tgtgggagc tcaagtagtt 1320
ggaatctgt ttgtcagtgg ctgggtatg atcacgatgt ttcttctt ttgttctg 1380
cattatgtg gatggcttcg tgccgactct ctgaggaag tagtcggtct cgacgtgct 1440
tactccaag gtgttctca gacgcgtgcc cgtgccccaa gtgaggaaga gaacatggag 1500

cattacatca glgaatatgt taagcagcgt gaggaaaaag cattcatcaa gaagatcaac 1560
 agcaatagta ctacaggacg tactatcctt ggtgcaagca tgcattcaat gaacatcatt 1620
 aattccagca tgcattccag aaaggacagt ctaccacgtg caattgaaag tctcaacaat 1680
 tcaagacact ctgggtcaag aggatcaaga tcaattaatg atattgcgat tgacaatttg 1740
 catgggtcaat ctgaggatgg tttgtctgca cctgatgagg gcagtgccta g 1791

<210> 32
 <211> 596
 <212> PRT
 <213> *Thalassiosira pseudonana*
 <400> 32

Met Ala Ser Ser Thr Thr Thr Asp Thr Tyr Gln Thr Cys Leu Ser Asp
 1 5 10 15

Leu Ser Ala Thr Ser Ser Asn Gly Ser Ser Pro Thr Thr Asp Ala Leu
 20 25 30

Leu Gln Cys Ile Ser Ser Ser Phe Asp Ala Gln Thr Ala Ser Thr His
 35 40 45

Ala Ser Ile Asn Thr Phe Phe Leu Leu Tyr Ala Ala Thr Leu Val Phe
 50 55 60

Phe Met Gln Ala Gly Phe Ala Met Val Ser Ala Gly Cys Val Arg Thr
 65 70 75 80

Asn Asn Val Gln Asn Thr Leu Leu Lys Asn Leu Leu Asp Ala Cys Gly
 85 90 95

Ala Ala Leu Gly Phe Tyr Thr Val Gly Tyr Ala Phe Ala Trp Gly Gly
 100 105 110

Ser Leu Asp Thr Ala Thr Thr Glu Arg Thr Phe Ile Gly Thr Gln Asn

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

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

500 505 510

Lys Ala Phe Ile Lys Lys Ile Asn Ser Asn Ser Thr His Gly Arg Thr
515 520 525

Ile Leu Gly Ala Ser Met His Ser Met Asn Ile Ile Asn Ser Ser Met
530 535 540

His Ser Arg Lys Asp Ser Leu Pro Arg Ala Ile Glu Ser Leu Asn Asn
545 550 555 560

Ser Arg His Ser Gly Ser Arg Gly Ser Arg Ser Ile Asn Asp Ile Ala
565 570 575

Ile Asp Asn Leu His Gly Gln Ser Glu Asp Gly Phe Ala Ala Pro Asp
580 585 590

Glu Gly Ser Ala
595

<210> 33
<211> 407
5 <212> PRT
<213> Secuencia artificial

<220>
10 <223> Dominio conservado comprendido dentro de la SEQ ID NO: 02 y dentro de la SEQ ID NO: 04
<400> 33

Met Gln Ala Gly Phe Ala Met Leu Cys Ala Gly Ser Val Arg Gln Lys
1 5 10 15

Asn Val Lys Asn Ile Met Leu Lys Asn Leu Leu Asp Ala Cys Gly Gly
20 25 30

Ala Ile Gly Phe Tyr Thr Val Gly Phe Gly Phe Ala Tyr Gly Gly Asp
35 40 45

Asp Thr Thr Asp Lys Thr Phe Ile Gly Asn Ser Tyr Phe Ala Leu Arg
50 55 60

Asp Tyr Thr Asn Tyr Ala Gly Phe Phe Phe Gln Phe Ala Phe Ala Ala
65 70 75 80

Thr Ala Ala Thr Ile Val Ala Gly Thr Val Ala Glu Arg Cys Lys Met
85 90 95

Ser Ala Tyr Leu Cys Tyr Ser Leu Phe Leu Thr Gly Phe Val Tyr Pro
100 105 110

Val Val Val Arg Ser Val Trp Ser Ser Asn Gly Phe Leu Ser Ala Phe
115 120 125

Ser Ala Asp Pro Phe Gln Gly Val Gly Thr Val Asp Phe Ala Gly Ser
130 135 140

Gly Val Val His Met Thr Gly Gly Leu Thr Ala Leu Ile Ala Ala Ile
145 150 155 160

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

Leu Glu Thr Pro Ala Ser Phe Pro Ala His Ser Val Ala Leu Gln Ile
180 185 190

Leu Gly Thr Phe Ile Leu Trp Phe Gly Trp Tyr Gly Phe Asn Pro Gly
195 200 205

Ser Ala Leu Lys Ile Ala Asn Ala Asp Ser Ala Ala Thr Ala Ala Leu
210 215 220

Cys Ala Val Thr Thr Thr Met Ala Ala Ala Ala Gly Cys Val Ser Ala
225 230 235 240

Met Phe Thr Asp Ser Ile Ile Asp Gly Met Ala Thr Gly Glu Thr Thr
245 250 255

Tyr Asp Leu Thr Met Ala Met Asn Gly Cys Leu Ala Gly Leu Val Ala
260 265 270

Val Thr Ala Gly Thr Ser Val Val Thr Pro Trp Ala Ala Ile Ile Ile
275 280 285

Gly Val Val Gly Gly Trp Val Tyr Ile Gly Met Ser Lys Leu Leu Ile
290 295 300

Lys Leu Lys Ile Asp Asp Ala Val Asp Ala Ile Pro Val His Phe Ala
305 310 315 320

Asn Gly Phe Trp Gly Val Leu Ala Thr Gly Leu Phe Ala Asn Gly Gly
325 330 335

Leu Met Ala Thr Ala Gly Tyr Asn Ser Glu His Glu Gly Trp Phe Tyr
340 345 350

Glu Trp Gly Ser Gly Ser Gly Asp Gly Ser Leu Leu Ile Cys Gln Leu
355 360 365

Ala Cys Leu Ala Trp Ile Ile Gly Trp Val Thr Thr Ile Met Thr Pro
370 375 380

Phe Phe Ile Leu Leu Asn Met Ala Gly Met Phe Arg Val Asp Pro Leu
385 390 395 400

Glu Glu Glu Val Gly Leu Asp
405

<210> 34
 <211> 2194
 <212> ADN
 <213> Oryza sativa

5

<400> 34

aatccgaaaa gttctgcac cgtttcacc ccctaactaa caatataggg aacgtgtgct 60
 aaalataaaa tgagacctta tataatgtagc gctgataact agaactatgc aagaaaaact 120
 catccaccta ctttagtggc aatcgggcta aataaaaaag agtcgctaca ctagtctcgt 180
 tttccttagt aattaagtgg gaaaatgaaa tcattattgc ttagaatata cgttcacatc 240
 tctgtcatga agttaaatta ttcgaggtag ccataattgt catcaaactc ttctgaata 300
 aaaaaatcct tctagctgaa ctcaatgggt aaagagagag attttttta aaaaaataga 360
 atgaagatat tctgaacgta ttggcaaaga tttaaacata taattatata attttatagt 420
 ttgtgcattc gtcatatcgc acatcaltaa ggacatgtct tactccatcc caattttat 480
 ttagtaatta aagacaattg acttattttt attatttate tttttcgat tagatgcaag 540
 gtacttacgc acacactttg tgctcatgtg catgtgtgag tgcacctcct caatacacgt 600
 tcaactagca acacatctct aatatcactc gcctatttaa tacatttagg tagcaatate 660
 tgaattcaag cactccacca tcaccagacc actttaata atatctaaaa taaaaaaat 720
 aattttacag aatagcatga aaagtatgaa acgaactatt taggttttc acatacaaaa 780
 aaaaaaagaa tttgctcgt gcgcgagcgc caatctccca tattgggcac acaggcaaca 840
 acagagtggc tgccacaga acaaccaca aaaaacgatg atctaacgga ggacagcaag 900
 tccgcaaaa cctttaaca gcaggcttg cgccaggag agaggaggag aggcaaagaa 960
 aaccaagcat cctccttct ccatctataa attcctcccc cctttcccc tctctatata 1020
 ggaggcatcc aagccaagaa gagggagagc accaaggaca cgcgactagc agaagccgag 1080
 cgaccgcctt ctgatccat atctccggt cgagttcttg gtcgatctct tccctcctcc 1140
 acctcctcct cacaggggat gtgcctcctc tcgggtgttc ttggatttat tgttctaggt 1200
 tgtgtaglac gggcgtgat gtaggaaag gggatctgta tctgtgatga ttctgttct 1260
 tggatttggg atagaggggt tcttgatgtt gcatgttate ggttcgggtt gattagtagt 1320

atggtttca atcgtctgga gagctctatg gaaatgaaat ggtttaggga tcggaatctt 1380
 gcgattttgt gaggaccttt tgttgagggt aaaatcagag caccgggtgat ttgcttggt 1440
 gtaataaagt acggttggtt ggtcctcgat tctggtagtg atgcttctcg attgacgaa 1500
 gctatccttt gtttatccc tatgaacaa aaataatcca acttgaaga cggccccgtt 1560
 gatgagattg aatgattgat tcttaagcct gtccaaaatt tcgcagctgg ctgttttaga 1620
 tacagtagtc cccatcacga aattcatgga aacagttata atcctcagga acaggggatt 1680
 cctgttctt ccgatttgct ttagtcccag aatTTTTT cccaaatata ttaaaaagtc 1740
 acttctggt tcagtcaat gaattgattg ctacaaataa tgctttata gcgttatcct 1800
 agctgtagtt cagttaatag gtaatacccc tatagttag tcaggagaag aacttatccg 1860
 attctgac tccatttta attatatgaa atgaactgta gcataagcag tattcattg 1920
 gattatttt ttattagct ctcacccctt cattattctg agctgaaagt ctggcatgaa 1980
 ctgtcccaaa ttgttttc aaattcacat cgattatcta tgcatatcc tctgtatct 2040
 acctgtagaa gttctttt ggttatcct tgactgcttg attacagaaa gaaatttatg 2100
 aagctgtaat cgggtagatt atactgctg ttcttatgat tcatttcctt tgtgcagttc 2160
 ttggtgtagc ttgccactt caccagcaaa gttc 2194

<210> 35
 <211> 50
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> cebador: prm09458

<400> 35
 ggggacaagt ttgtacaaaa aagcaggctt aaacaatgat gcaggccggg 50

<210> 36
 <211> 49
 <212> ADN
 <213> Secuencia artificial

<220>
 h<223> cebador: prm09459

<400> 36
 ggggaccact ttgtacaaga aagctgggta cagagcagc aattaaacc 49

REIVINDICACIONES

1. Un método para incrementar el vigor temprano, la biomasa, y/o el rendimiento de semilla en plantas cultivadas bajo condiciones no estresadas con relación a plantas que sirven como control, que comprende incrementar la expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido transportador de amonio (AMT), en donde el polipéptido AMT comprende un dominio que tiene al menos una identidad de secuencia de aminoácidos del 80%, 85%, 90%, 95%, 98%, 99% o más con un Dominio Conservado (DC) como el representado por la SEQ IS NO: 33, y opcionalmente seleccionar las plantas que tienen mayor vigor temprano, biomasa, y/o rendimiento de semilla,
- en donde dicho polipéptido AMT tiene al menos una identidad de secuencia de aminoácidos del 80%, 85%, 90%, 95%, 98%, 99% o más con el polipéptido AMT como el representado por la SEQ ID NO: 2.
2. Un método de acuerdo con la reivindicación 1, en donde dicha secuencia de ácido nucleico que codifica un polipéptido AMT se representa por la secuencia de ácido nucleico SEQ ID NO: 1 o una porción de la misma que es de al menos 1380 o más nucleótidos consecutivos de longitud, o una secuencia capaz de hibridar con la secuencia de ácido nucleico SEQ ID NO: 1.
3. Un método de acuerdo con cualquier reivindicación precedente, en donde dicha mayor expresión se efectúa por una cualquiera o más de: marcación de la activación de ADN-T, TILLING, o recombinación homóloga.
4. Un método de acuerdo con cualquier reivindicación precedente, en donde dicha mayor expresión se efectúa por medio de la introducción y expresión en una planta de una secuencia de ácido nucleico que codifica un polipéptido AMT.
5. Un método de acuerdo con cualquier reivindicación precedente, en donde dicha secuencia de ácido nucleico está operativamente enlazada a un promotor constitutivo.
6. Un método de acuerdo con cualquier reivindicación precedente, en donde dicha secuencia de ácido nucleico que codifica un polipéptido AMT es de Heterokontophyta phylum.
7. El uso de una construcción que comprende:
 - (a) una secuencia de ácido nucleico que codifica un polipéptido AMT como se define en la reivindicación 1 o 2 y del orden de Pennales;
 - (b) una o más secuencias control capaces de dirigir la expresión de la secuencia del ácido nucleico de (a); y opcionalmente
 - (c) una secuencia de terminación de la transcripción
 en donde dicha secuencia de control es un promotor GOS2 constitutivo de una planta, en un método para elaborar plantas que tienen mayor vigor temprano, biomasa, y/o rendimiento de semilla con relación a plantas que sirven de control.
8. Un método para la producción de plantas transgénicas que tienen mayor vigor temprano, biomasa, y/o rendimiento de semilla con relación a plantas que sirven de control, que comprende:
 - (i) la introducción y expresión en una planta, parte de una planta, o una célula de una planta, de una secuencia de ácido nucleico que codifica un polipéptido AMT como se define en cualquiera de las reivindicaciones 1 a 2, bajo el control de un promotor constitutivo de una planta; y
 - (ii) el cultivo de la célula de la planta, parte de la planta, o la planta bajo condiciones no estresadas que promuevan el crecimiento y desarrollo de la planta.
9. El uso de una secuencia de ácido nucleico que codifica un polipéptido AMT como se define en cualquiera de las reivindicaciones 1 a 2 en incrementar el vigor temprano, la biomasa, y/o el rendimiento de semilla en plantas cultivadas bajo condiciones no estresadas.

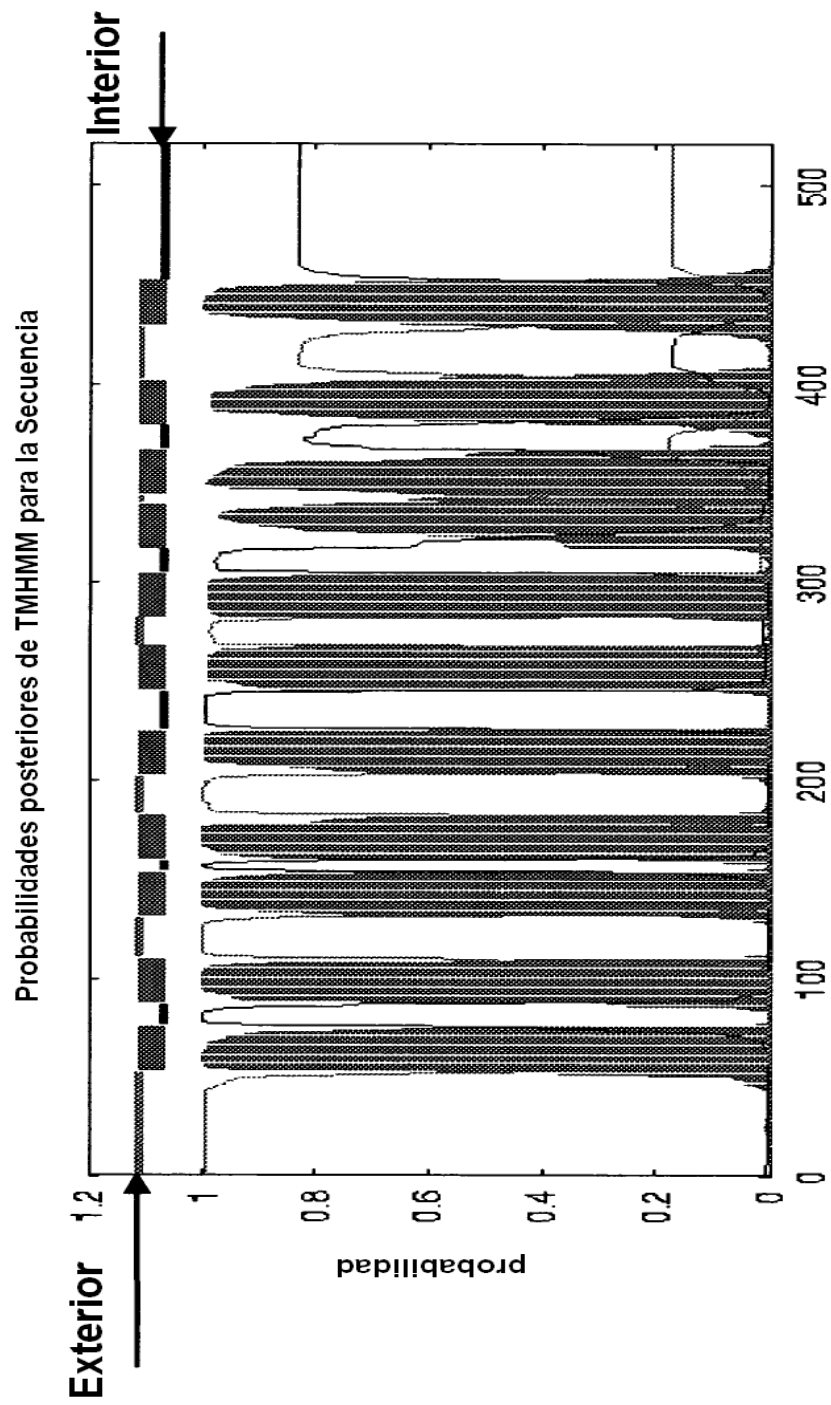


FIGURA 1

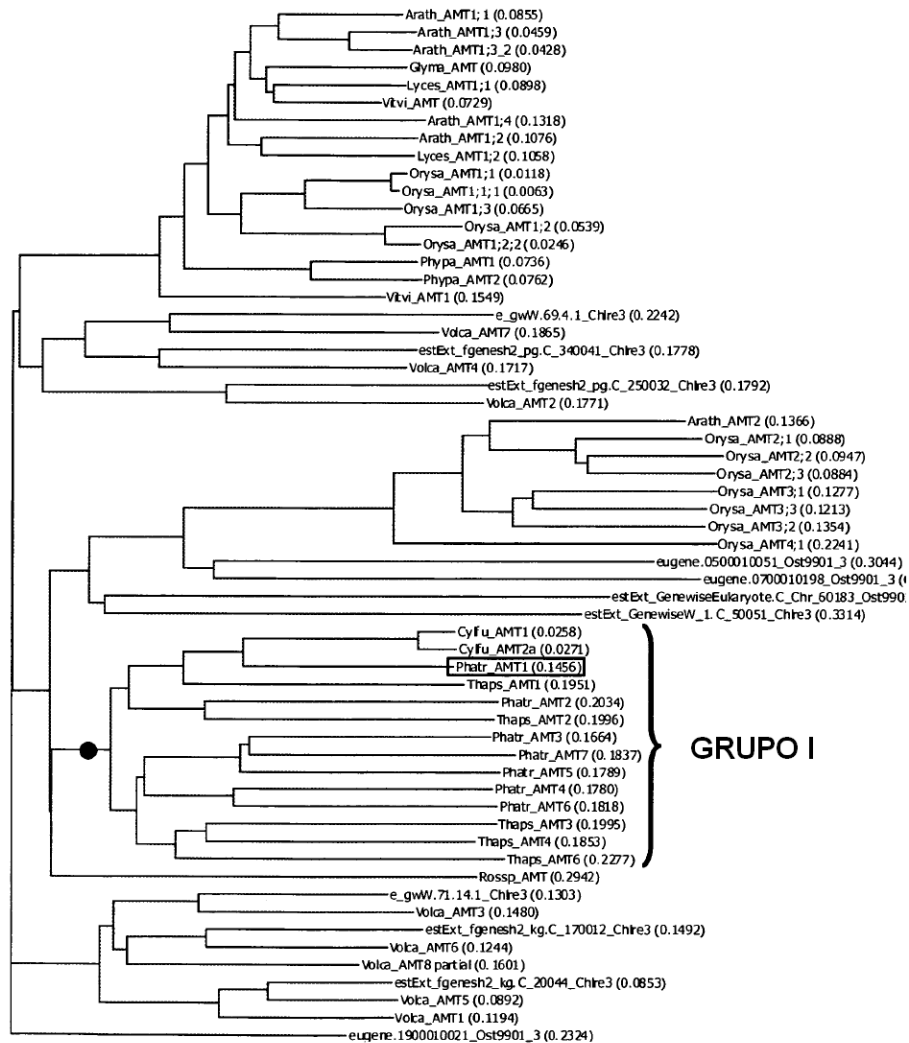


FIGURA 2

		1		50
Phatr_AMT1	(1)	-----MSFDLDAFCTGLTAA-----	SSSSEQA	
Cylfu_AMT1	(1)	-----MAEFDNTFILDFCSGGNES-----	SS-DVQA	
Cylfu_AMT2a	(1)	-----MAEFDNTFILEFCSSGGNES-----	SS-DVQV	
Thaps_AMT1	(1)	-----MAEPTTTIGDFN--VTAW-----	CGDDAVA	
Phatr_AMT2	(1)	-----MSSSAIYQSCAGQFDS-----	GEQLDQL	
Thaps_AMT2	(1)	-----MSEDPSIFEVCTGQLG-----	TDLTVEL	
Phatr_AMT4	(1)	MDDASFIQSLVEGYGTSSNHTTVYGYCSNEA-----	EGETNLI	
Phatr_AMT6	(1)	-----MSTDGSLFQCCSAVAG-----	DSDPSRI	
Phatr_AMT3	(1)	-----MNSRTFSYNSNDGNELLDACMAYLG-----	ANATTYDL	
Phatr_AMT7	(1)	-----MISTGSSTSTNAYGTCVQLG-----	ENSSAKEL	
Phatr_AMT5	(1)	-----MSHSSLDVFGTCLAQVG-----	EDATTKEI	
Thaps_AMT3	(1)	-----MFQVSRAGHVSVEVCKSFVN-----	PEDSQADQFDAM	
Thaps_AMT4	(1)	-----MSSSVRTSLYEACKSTQSNSTFSNSTLDDALSRQEQI		
Thaps_AMT6	(1)	-----MASSTTTDTYQTCLSDLSATS-----	SNGSSPTDAL	
Consenso	(1)	S T IYE C G G	S L	
		51		100
Phatr_AMT1	(23)	VCAIQTIVAGVSKTVGGIDAEGITAGVDTFFLLIFAGALVF	MQAGFAMLC	
Cylfu_AMT1	(26)	LCQVAGLANGTSASAA----GLVEGINTEFLLFAGALVF	MQAGFAMLC	
Cylfu_AMT2a	(26)	LCQVASLANGTSASAG----GLTEGINTEFLLFAGALVF	MQAGFAMLC	
Thaps_AMT1	(24)	TYEQSVENGICAAYAYT--DETNTGLDVFYLLFAAMVF	MQAGFAMLC	
Phatr_AMT2	(24)	LOCLSTGHDGALSQTSN---LAGGIDAEYLIIFAGALVF	MQAGFAMLC	
Thaps_AMT2	(24)	LQCVSDGAESAKDDVIS-----G--VNSFYLIIFAGALVF	MQAGFAMLC	
Phatr_AMT4	(39)	LOQITEVMEQKQ-----LEGDRNVNRWLMFSGGLIFF	MQAGFAMLC	
Phatr_AMT6	(25)	LQCVSDALETNQ-----NDRAADLNWFLIIAGALVF	MQAGFAMLC	
Phatr_AMT3	(34)	LGCVSAQLSNEVGS-----REFSRVLLVYAAALVF	MQAGFAMLC	
Phatr_AMT7	(30)	LECVDYLQHQE-----APFSSTLVITFAGALVF	MQAGFAMVC	
Phatr_AMT5	(26)	LECVSFSLSRAPDGLDE--PSSKGFTRSIVVFAAALVF	MQAGFAMLC	
Thaps_AMT3	(34)	LQCVGE-----SNGKSIDAEFLIYASSLVF	MQAGFAMLC	
Thaps_AMT4	(38)	FRQISE-----SNANSIDTEFLLYASSLVF	MQAGFAMLC	
Thaps_AMT6	(33)	LOQISSSFDAQTA-----STHASINTEFLLYAATLVF	MQAGFAMVS	
Consenso	(51)	L CVS I TFLLFAGALVF	MQAGFAMLC	
Comienzo del Dominio Conservado (CD)				XXXXXXXXXX
		101		150
Phatr_AMT1	(73)	AGSVRQKNVKNIMLNLLDACGGAIGFYTVGFGFAYGGDDTTD-----		
Cylfu_AMT1	(71)	AGSVRQKNVKNIMLNLLDACGGAIGFWTICAFAYADNSSGD-----		
Cylfu_AMT2a	(71)	AGSVRQKNVKNIMLNLLDACGGAIGFWTICAFAYADNSSGN-----		
Thaps_AMT1	(72)	AGSVRQKNVKNIMLNLLDACGGAIGFWSVGFAFAYGGSGPEK-----		
Phatr_AMT2	(70)	AGSIRAKNVKNVILNLLDSGGGLAFWSVGAFAYGGDNAGS-----		
Thaps_AMT2	(66)	AGSIREKNVKNVILNLLDSAGGAFGFWISICAFAYGGDDITKG-----		
Phatr_AMT4	(81)	AGCVPRKNVQNTMLNLLDACGAALGFLLCYAFAFGGQDDRD-----		
Phatr_AMT6	(67)	AGCVPRKNVQNTMLNLLDACGAALGFYVICYALAFGGQNERS-----		
Phatr_AMT3	(75)	AGAVRKNVQNTMLNLLDACGAALGFYVICYALAFGGQMEP-----E		
Phatr_AMT7	(69)	AGAVRKNVQNTMLNLLDACGAALGFYVICYALAFGGQMEP-----E		
Phatr_AMT5	(74)	AGAVRKNVQNTMLNLLDACGAALGFYVICYALAFGGQMEP-----E		
Thaps_AMT3	(69)	AGCVQHKNVQNTMLNLLDACGAALGFYVICYALAFGGQMEP-----E		
Thaps_AMT4	(73)	AGSVRKNVQNTMLNLLDACGAALGFYVICYALAFGGQMEP-----E		
Thaps_AMT6	(75)	AGCVRTNVQNTMLNLLDACGAALGFYTYCYAFWGGSLDTATT-----		
Consenso	(101)	AGSVR KNVQNTMLNLLDACGAALGFWSVGAFAYGG D		
DC (cont'd)				XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

FIGURA 3

		151		200
Phatr_AMT1	(116)	---KTFI	ENSYFA	LRDYN---YAGFF
Cylfu_AMT1	(114)	---KTFI	GKNFFVNQ	LDSEGGAWIGFF
Cylfu_AMT2a	(114)	---KTFI	GKNFFVNQ	LDSEGGAWIGFF
Thaps_AMT1	(115)	---KGF	ENEGFFLGD	FTTGG-DLIGWF
Phatr_AMT2	(113)	---KTF	VGNAGFFLQ	GDDIR---LENWF
Thaps_AMT2	(110)	---KTF	IGNADFFLSG	-ETD---MEFWF
Phatr_AMT4	(124)	--DVT	FI	GTSNFFLNTG---KV-DMSFWF
Phatr_AMT6	(110)	--DVT	FVGTDF	FNWNSAVPV-NQAFWF
Phatr_AMT3	(117)	SPNKT	FLGNTNFFLMG	-VDD---LAFWL
Phatr_AMT7	(111)	SSKKT	FVCHSQFFLMD	-VDD---YAFWL
Phatr_AMT5	(124)	NNQTT	FI	GTSNFFLVN-VDD---YSFWL
Thaps_AMT3	(113)	--NKT	FI	GTEFFLMG-VDD---FMFWL
Thaps_AMT4	(116)	--KKT	FI	MSNFFLQD-VDN---YMFWL
Thaps_AMT6	(120)	--ERT	FI	GTONFFLMD-VDSS-QDSFWL
Consenso	(151)	KTFI	GNFFL	D YAFWFFQFAAATAATIVAGTLAERCQ
DC (cont'd)		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		
		201		250
Phatr_AMT1	(159)	MSAYL	CYSLELTGF	VYPVVVR
Cylfu_AMT1	(161)	MSAYL	CYSVFLTGF	VYPVVVHSI
Cylfu_AMT2a	(161)	MSAYL	CYSIFLTGF	VYPVVVHSI
Thaps_AMT1	(161)	FEAYL	CYSLMLTGF	VYPVIVYSI
Phatr_AMT2	(157)	MKAYL	MYSVFLAGE	VYPVVAHAF
Thaps_AMT2	(153)	MMAYL	CYSIFLCGF	VYPVCAHAF
Phatr_AMT4	(168)	MVAYL	CYSIFLTGF	VYPVVAHTI
Phatr_AMT6	(157)	MVAYL	CYSVFLTGF	VYPVVAHSI
Phatr_AMT3	(163)	MVAYL	CYSVMLTGW	VYPIIAHAI
Phatr_AMT7	(157)	MTAYL	CYSLMLTGW	VYPVILHSI
Phatr_AMT5	(170)	MAAYL	CYSALLTGW	VYPIVAHAV
Thaps_AMT3	(157)	MTAYL	CYSVAVTGF	VYPVVVHSI
Thaps_AMT4	(160)	MTAYL	CYSIALTGF	VYPVVAHSI
Thaps_AMT6	(166)	MVAYL	AYSMTLAGE	VYPVVVHSI
Consenso	(201)	M AYL	CYSVFLTGF	VYPVVAHSIWS NGFLSAFA DPLWG VGVIDFAGS
DC (cont'd)		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		
PS01219				DFAGS
		251		300
Phatr_AMT1	(208)	GVVHMT	GGTALVA	AIVLGPRKGRFY
Cylfu_AMT1	(210)	GVVHMT	GGATALVA	AIVLGPRKGRFY
Cylfu_AMT2a	(210)	GVVHMT	CGGATALVA	AIVLGPRKGRFY
Thaps_AMT1	(210)	GVVHMT	GGITALLA	AKILGPRIGRFY
Phatr_AMT2	(206)	GPVHMT	GGVTALAA	ALVLGPRIGRFY
Thaps_AMT2	(202)	GPVHMT	CGGVAALV	MAIILGPRRGRFY
Phatr_AMT4	(217)	GVVHMT	GGTALVAT	YILGARKGRFY
Phatr_AMT6	(206)	GVVHMT	GGTALVAT	YMLGARKGRFY
Phatr_AMT3	(212)	GVVHMT	GGVTALFAT	LILGPRRGRFY
Phatr_AMT7	(206)	GVVHMT	GGITALFAT	MVLGPRRGRFY
Phatr_AMT5	(219)	GVVHMT	GGVTALFAT	IILGPRRGRFY
Thaps_AMT3	(206)	SVVHMT	GGTALVAT	YILGPRRGRFY
Thaps_AMT4	(209)	TVVHMT	GGTALVAT	YILGPRRGRFY
Thaps_AMT6	(216)	TVVHMT	GGTALVAT	IIVLGPRRGRFY
Consenso	(251)	GVVHMT	GGVTALVAT	IILGPRRGRFYD DGN LE PKSFPGHSVALQMLG
PS01219		GVVHMT	GGVTALVAT	IILGPR
DC (cont'd)		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		

FIGURA 3 (continuación)

		301		350
Phatr_AMT1	(258)	TFILWFEGYGFNPGSALKIANAD-SAATAALCAVTTTMAAAAGCVSAMFT		
Cylfu_AMT1	(260)	TFILWFEGYGFNPGSALVIDNAA-SASTSALCAVTTTLAAASGCVCAMFT		
Cylfu_AMT2a	(260)	TFILWFEGYGFNPGSALVIDNAA-SASTSALCAVTTTLAAASGCVTAMFT		
Thaps_AMT1	(260)	TFILWFEGYGFNPGSTLLISNTA-AADVSAALCAVTTTLAAASGSVSAMFT		
Phatr_AMT2	(256)	TFCLWFEGYGFNPGSVFTISSIE-NGQVAALVAVNTTLAACAGAVSAMFT		
Thaps_AMT2	(252)	TFALWFEGYGFNPGSSILLASAA-SGDVASLAAVNTTLGSAAGALSGMFT		
Phatr_AMT4	(267)	TFVLWFEGYGFNPGSALLAHTSDTGFAVSRRAVNTLSAASGAVSAIMT		
Phatr_AMT6	(256)	TMTLWFEGYGFNPGSALLTATSNTGGVAALAAVNTLSAASGAVSALFT		
Phatr_AMT3	(262)	TFILWFEGYGFNPGCSALLIDKPG-ANDIAALAGVNTLSAGVAGIVALEV		
Phatr_AMT7	(256)	TFILWFEGYGFNPGSALISTKQG-SDEAAALAGVNTLSASAAGIVALEFS		
Phatr_AMT5	(269)	TLILWFEGYGFNPGCAALLDVPS-SDNIAALAAVNTTLSGGTAGIVALEFF		
Thaps_AMT3	(256)	APILWFEGYGFNPGSTLSTTGPG-QHQVVSIVAVNTTLAAASACVASILA		
Thaps_AMT4	(259)	IFILWFEGYGFIVGSAITIIIGPN-QDKIIISTAVNTLSAASSCFSALLV		
Thaps_AMT6	(266)	VEILWFEGYGFNPGSILNITNDL-NHTIVSHTAINTTLAASAGSIMTLEL		
Consenso	(301)	TFILWFEGYGFNPGSALLI A SA VAALAAVNTTLAAASGAVSALFT		
DC (cont'd)		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		
		351		400
Phatr_AMT1	(307)	DSIIDGMATGETTYDLTMAMNCLAGLVAVTAGTSVVTPMAAIIIGVVG		
Cylfu_AMT1	(309)	DTIIEMMATGEASYDLTMAMNCLAGLVAITAGCSVVTPWASIIIGIAG		
Cylfu_AMT2a	(309)	DTLIEMMATGEASYDLTMAMNCLAGLVAITAGCSVVTPWASLIIGIIG		
Thaps_AMT1	(309)	DMFLERRKTGETMYDIETCMNCGALSGLVGITAGCSIVEPWAAPVIGIVAG		
Phatr_AMT2	(305)	STLFDYWTGLHTYDLGYTMNCLTGLVAITAGCATVETWAAVLIGICAG		
Thaps_AMT2	(301)	STIVDERKTGVYTWDTTAAMNCLTGLVAITAGCATVEPWAAPVIGITAG		
Phatr_AMT4	(317)	NMFMEERSTGEYSFNITAMNCGALGLVSITAAGCTVONWAAALCTCGIG		
Phatr_AMT6	(306)	SLYLEERKTGEYSFNITAMNCGALGLVGITAGCGTVENWAACTCTGVSG		
Phatr_AMT3	(311)	NLWYLERTTGEPPFDLTYAMNCSLSGLVAITGGCAVLEPWAAPVIGVAG		
Phatr_AMT7	(305)	NLWYLEKTTEGEPLFDIKYAMNCAICGLVAITSGGCGVFERWAAVTVAGVAG		
Phatr_AMT5	(318)	NLWYLDKRTGEAYFDLKFAMNCLCGLVAITGGCGVFERWAAVTVAGVAG		
Thaps_AMT3	(305)	SYVIERKTGEGETSESSAMNCLCGLVSITGGCAVVEPWAAPVIGIAG		
Thaps_AMT4	(308)	NYVIVERQSGEGETSLAAMNCLCGLVAITGGCAVIAFWAAIIVGLFAG		
Thaps_AMT6	(315)	STVVAERFTGEIVFSLSYAMNCLCGLVAITAGCSVVEPWAAPVIGLVGG		
Consenso	(351)	WIEER TGE SF'DLT AMNCGLAGLVAITAGCAVVEPWAAPVIGIVAG		
DC (cont'd)		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		
		401		450
Phatr_AMT1	(357)	WVYIGMSKLLIKEKIDDAVDATPVHFANGFWGLATGLFANGGLMATAGY		
Cylfu_AMT1	(359)	WVYIAFSKLLVKEKIDDAVDATPVHFANGFWGLAVGFFAEPDAMVTAGY		
Cylfu_AMT2a	(359)	WVYIGLSKLLIKEKIDDAVDATPVHFGNMGVLAAGFFAEPDAMVTAGY		
Thaps_AMT1	(359)	WTYIFWSSLLVKEKIDDAVDATPVHFGNMGVLAAGFFAEPTRVANAYS		
Phatr_AMT2	(355)	WVYELGSKLLVYFRIDDAVDATPVHMGCAAGVIAATGLFTKGELLAAFG		
Thaps_AMT2	(351)	WVYLAASALMLRFKIDDAVDATPVHMFSGSWGVFCTGLFTSPRRLLITAYG		
Phatr_AMT4	(367)	LIYEWGSKTLVREKIDDAVDATPVHMFAGWGLLAVGLSDPDLMHIAYG		
Phatr_AMT6	(356)	WVYIFGS AFLRLRIKIDDAVDATPVHMFCAAGLIAATGLFSSPRHTLEAFG		
Phatr_AMT3	(361)	ILYMGVSRGLVMEKIDDAVDATPVHFNCAWGLMSVGLFASPARLLAAYD		
Phatr_AMT7	(355)	VIVELGSRGLVSMRLDDAVDATPVHLCGCAAGLAVGLFAAPERILSVYG		
Phatr_AMT5	(368)	LLYNIGSRGLIYERLDDAVDATPVHLCNCSWGLVAVGLFASPSRLLIYIG		
Thaps_AMT3	(355)	LLYFTSKLLIRERIDDAVDATPVHLSNGLWGTVAAGLFASSNRLQLAFG		
Thaps_AMT4	(358)	LLYFTSKVLVVRVIDDAVEATPVHMTNGLWGSFAVGLFAAPSELQLVYG		
Thaps_AMT6	(365)	ALYLACSKFLVKKRIDDAVDGIPVHLINGLWGTLSVGLFAVPELLEQVYG		
Consenso	(401)	WLYL GSKLLVCLKIDDAVDATPVHM NG WGLVAVGLFASP LL AYG		
DC (cont'd)		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		

FIGURA 3 (continuación)

		451		500
Phatr_AMT1	(407)	NSEHEGWFYEWGSGSGDGSLEICOLACLAWTIGAVTTIMTPFEILINMAG		
Cylfu_AMT1	(409)	N-DVPGVIFYKG-----DGKILLCQFVAIIWICAWIFFLMTPEFVVLNILG		
Cylfu_AMT2a	(409)	N-DVPGVIFYKG-----DGKILLCQFVAIVVCAAVFFLMTPEFVVLNILG		
Thaps_AMT1	(409)	DHCHYGNFYSWGAGNADAHLEAAVCGVLWETICAVSVIMIPYFILLNVLG		
Phatr_AMT2	(405)	QEEHVGNFYEWGSGSGNFTLIGIOLLSVLFIFAWTFVSMGIYFYALSFMG		
Thaps_AMT2	(401)	NDNNVGNFYEWGRSGSGNFTLIGCOLVSILFVLGWSACIFAFPCIALKTLN		
Phatr_AMT4	(417)	TGNHPLGLLYSWGLGEFNAILLSNOVLELVFVAGWAFGMTPTPEFLFINRMG		
Phatr_AMT6	(406)	TDAHVGNFYSLGQDSLDAIILMNOELGLLFIILGWSAILMSPPFWLWNYMG		
Phatr_AMT3	(411)	NDAHPCWFYSLRNGKSDGRIVGVOLVGIVFIVGAVMVIMLPFFIWLWKG		
Phatr_AMT7	(405)	RNNHPLGLVYSIREGDIDGVIFGIOLIGLMFIMGVMIIMLPFFVWLWKG		
Phatr_AMT5	(418)	HSDHPCWFYSLRDGESDFRLEASOLVGLIFIVFVWFMNMLPEFVWLNYRG		
Thaps_AMT3	(405)	KVADTGVFMG-----GTGRILGCOLIGVFFVLGWSIFIMIPFCFLHYMG		
Thaps_AMT4	(408)	KANHVGLPYSWHQSGDGTILGVOCLGILFVVGWVFCILMSPPFLFLNYKG		
Thaps_AMT6	(415)	RGDHAGWFYSWGQGSADAKLEGAQVVGILFVSGWVMITMFPFCFLHYVG		
Consenso	(451)	H GWFYSWG G GDG LLG QLVGILFILGWV IM PFFVFLNYLG		
DC (cont'd)		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		
		501		550
Phatr_AMT1	(457)	MRRVDPLEEENGIDSHHRGSAYDLS-----GPSKDHVDELMEIRAS--		
Cylfu_AMT1	(453)	MRRVDPLEEENGIDSHHRGAAYDMT-----SAKKEDVEELMEHRSS--		
Cylfu_AMT2a	(453)	MRRVDPLEEENGIDSHHRGSAYDMT-----TAKKEDVEELMEHRSS--		
Thaps_AMT1	(459)	LRRVDALPEEENGIDSHHKGAAYDMS-----GPSEAAAEKFEISRS--		
Phatr_AMT2	(455)	WLRIDPLEEENGIDSRHKGSAAYDMT-----SANMEQVRSMLMDDRSTSN		
Thaps_AMT2	(451)	WLRIDPLEEENGIDSRHKGPAEYSEG-----SAHSDAIEKLSASRRDIM		
Phatr_AMT4	(467)	WFRSDSLEELAGIDAYHGGKKGHGE-----EVVELSALEGFIKNKIRQS		
Phatr_AMT6	(456)	WLRADSLPEELAGIDAYHGGREAGEG---FDDEVFPATSKDKPDTLRRRK		
Phatr_AMT3	(461)	WFRSDPLEEINGIDSYHGGLALLGG-----DDEVNPEYISAYKKQNE--		
Phatr_AMT7	(455)	WFRSDPLEEILGIDSYHVQLALHT-----NNVHPEYVGSEKDVVDEI-		
Phatr_AMT5	(468)	WFRSDPLEELAGIDSYHGGLMLHE-----EVDPEYISAYRKQHEA-		
Thaps_AMT3	(450)	WLRSESIDVEVGIDKYHGLRNKDEHRHDEEDNTPSHYEGEGNCRLLR--		
Thaps_AMT4	(458)	WFRADVLNEIAGIDSYHDCVDMELVTQIRNQKN-LHVNSRNRFFSSN--		
Thaps_AMT6	(465)	WLRADSLPEELAGIDAYSQGVLTQTRARAQSEENMEHYISEYVKQREEKA		
Consenso	(501)	WFRSDPLEEINGIDSYH G A D EYV R		
DC (final)		XXXXXXXXXXXX		
		551		600
Phatr_AMT1	(499)	-----KHGKVEVPKEVAQAADDAAEETA-----		
Cylfu_AMT1	(495)	-----KHGKVEVPKEVQKEDTA-----		
Cylfu_AMT2a	(495)	-----KHGKVEIPKEVQKEDTA-----		
Thaps_AMT1	(500)	-----QRKLEIPVDVAPATAPAEDAA-----		
Phatr_AMT2	(499)	----RGKLRKSSIIIEKPAKTGPIEGGTASGESEIKESDVEHAERV----		
Thaps_AMT2	(496)	----NASG--SGRGKSFSRSTPTKANEEPKIEATEDAGAPAGEATA----		
Phatr_AMT4	(511)	-----QIRSPYN-----		
Phatr_AMT6	(503)	-----GGNSAEGTRDDGSWTDLTASAPRETAPLPelyCENDASSKESDDD		
Phatr_AMT3	(505)	-----GTLRRRHKGTTSSVKTGEVESDEGAERVAPEATKHKKIPIAT		
Phatr_AMT7	(498)	-----ISTRQRKVNGSTTTKATSGTEELEYIPEVSDEDLSEMKEECL-		
Phatr_AMT5	(510)	-----HSRTLQRK-RTSHVRLESVSEHSVAPNGGDSAHTSNGDSLTE		
Thaps_AMT3	(498)	-----SILRHEERMROEDSAIPATLVSSDDRFTCDSGGNSTSMSTAKQY		
Thaps_AMT4	(505)	-----SACPHHT-----		
Thaps_AMT6	(515)	FIKKINSNSTHGRTILGASMHSMNIIINSSMHSRKDSLPRAIESLNNSRHS		
Consenso	(551)	R		

FIGURA 3 (continuación)

		601	632
Phatr_AMT1	(522)	-----	-----
Cylfu_AMT1	(512)	-----	-----
Cylfu_AMT2a	(512)	-----	-----
Thaps_AMT1	(521)	-----	-----
Phatr_AMT2	(540)	-----	-----
Thaps_AMT2	(536)	-----	-----
Phatr_AMT4	(518)	-----	-----
Phatr_AMT6	(548)	PSGMHHA-----	-----
Phatr_AMT3	(547)	TTEVFNGGQDTDQMSYHA-----	-----
Phatr_AMT7	(540)	-----	-----
Phatr_AMT5	(552)	TVTIREPDSGWPRGNV-----	-----
Thaps_AMT3	(543)	T-----	-----
Thaps_AMT4	(512)	-----	-----
Thaps_AMT6	(565)	GSRGSRISINDIAIDNLHGQSEDGFAAPDEGSA	-----
Consenso	(601)		

FIGURA 3 (continuación)

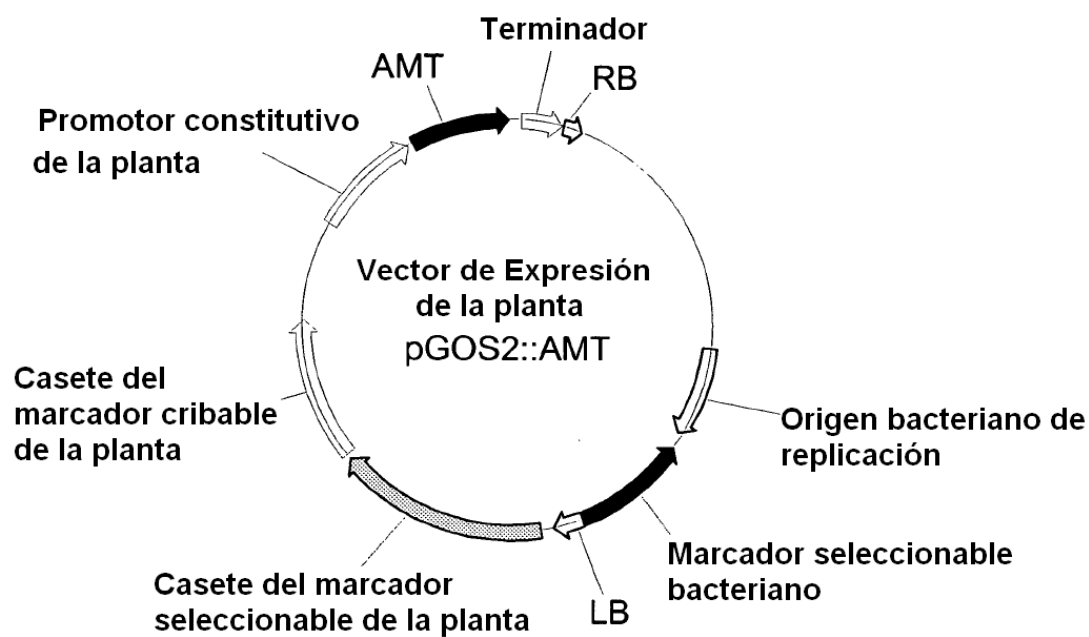


FIGURA 4

SEQ ID NO: 1 Secuencia de ácido nucleico de longitud completa Phatr_AMT1 de Phaeodactylum tricornutum

ATGTCTTTTGATTTGGACGCATTCTGCACTGGCCTGACAGCCGCGTCGAGCTCTTCTGAGCAGGCT
GTGTGCGCTCTGCAAACGATCGTGGCCGGAGTTTCTAAGACTGTCGGAGGTATTGACGCGGAAGGG
ATTACTGCAGGTGTTGATACTTTCTTCTCATTCTTTGCGGGTGCCTTGGTCTTCATGATGCAGGCC
GGGTTCCGCATGCTTTGTGCTGGATCCGTCCGTCAAAGAATGTAAAGAATATTATGCTCAAGAAC
TTGTTGGATGCCTGTGGTGGTGTCTATTGGCTTCTACACCGTTGGTTTCGGCTTCGCTTATGGCGGT
GACGACACCACCGACAAGACCTTCATTGGCAACAGCTACTTCGCGCTCCGTGATTACACAAATTAT
GCAGGTTTCTTCTTCCAGTTTGCCTTTGCTGCCACTGCCGCCACGATTGTTGCCGGTACAGTTGCT
GAGCGATGCAAGATGTCGGCATAACCTTTGCTACTCTCTCTTTCTTACGGGTTTCGTCTATCCCGTC
GTTGTACGCTCTGTCTGGAGCTCCAACGGGTTCTTGTGTCAGCCTTCAGTGCCGACCCCTTCCAAGGA
GTTGGAACCGTTGACTTTGCCGGATCAGGTGTGGTGCACATGACTGGAGGACTCACCGCCTTGATT
GCTGCCATTGTTCTTGGACCGCGTAAGGGTCGGTTCTACGATGAGGATGGCAACCCTCTGGAGACG
CCCGCCAGCTTCCCAGCCCCTCTGTAGCCCTCCAGATCCTCGGAACCTTTCATCTTGTGGTTCCGA
TGGTACGGATTCAACCCTGGTTCAGCCCTGAAGATTGCTAACGCCGATTCGGCCGCAACCGCCGCT
TTGTGTGCCGTCACCACCACTATGGCCGCCGCTGCCGTTGTGTTTCCGCCATGTTCACTGACTCG
ATCATTGACGGCATGGCGACCGGTGAACTACGTACGATCTGACCATGGCCATGAATGGATGCCTT
GCTGGTCTCGTTGCCGTCCTGCTGGTACATCTGTCTGTCACCCCATGGGCCGCAATCATTATTGGA
GTCGTTGGAGGTTGGGTCTACATTGGTATGTCCAAGCTTTTGATCAAGCTCAAGATTGATGACGCT
GTCGATGCCATCCCTGTCCATTTCCGCCAATGGTTTCTGGGGTGTCTAGCCACCGGCCTTTTCGCC
AACGGTGGATTGATGGCAACCGCTGGGTACAACCTCGGAACACGAGGGCTGGTTCTACGAATGGGGA
AGTGGCTCCGGAGATGGAAGTCTTCTCATCTGCCAGCTTGCTTGCCTCGCCTGGATTATTGGATGG
GTCACCACCATATGACGCCCTTTTTTATCCTTTTGAACATGGCCGGTATGTTCCGTGTGGACCCG
CTTGAGGAAGAAGTTGGTCTTGATATTTCCCATCACCGTGGATCTGCTTACGATCTTTCGGGACCC
AGCAAGGACCATGTTGACGAGCTCATGGAAATTCGTGCCTCGAAGCACGGCAAGGTTGAGGTTCCA
AAGGAGGTTGCGCAGGCTGCTGATGACGCCGCCGAAGAGACTGCTTAA

SEQ ID NO: 2 Secuencia del polipéptido completamente traducido Phatr_AMT1 de Phaeodactylum tricornutum

MSFDLDAFCTGLTAASSSSEQAVCALQTIIVAGVSKTVGGIDAEGITAGVDTFFLIFAGALVFMMQA
GFAMLCAGSVRQKNVKNIMLKNLLDACGGAIGFYTVGFGFAYGGDDTTDKTFIGNSYFALRDYNTY
AGFFFQFAFAATAATIVAGTVAERCKMSAYLCYSLFLTGFVYPVVVRSVWSSNGFLSAFSADPFQG
VGTVDFAAGSVVHMTGGLTALIAAIVLGRKGRFYDEDDGNPLETPASFPASHVALQILGTFILWFG
WYGFNPGSALKIANADSAATAALCAVTTTMAAAAGCVSAMFTDSIIDGMATGETTYDLTMAMNGCL
AGLVAVTAGTSVVTWAAIIIGVVGGWVYIGMSKLLIKLKIDDAVDAIPVHFANGFWGVLATGLFA
NGGLMATAGYNSEHEGWFYEWGSGDGSLLICQLACLAWIIGWVTTIMTPFFILLNMAGMFRVDP
LEEEVGLDISHHRGSAYDLSPSKDHVDELMEIRASKHGKVEVPKEVAQAADDAEETA

SEQ ID NO: 3 Secuencia parcial de ácido nucleico de longitud completa Phatr_AMT1 de Phaeodactylum tricornutum

ATGATGCAGGCCGGGTTCCGCATGCTTTGTGCTGGATCCGTCCGTCAAAGAATGTAAAGAATATT
ATGCTCAAGAACTTGTTGGATGCCTGTGGTGGTGTCTATTGGCTTCTACACCGTTGGTTTCGGCTTC
GCTTATGGCGGTGACGACACCAGCGACAAGACCTTCATTGGCAACAGCTACTTCGCGCTCCGTGAT
TACACAAATTATGCAGGTTTCTTCTTCCAGTTTGCCTTTGCTGCCACTGCCGCCACGATTGTTGCC
GGTACAGTTGCTGAGCGATGCAAGATGTCGGCATAACCTTTGCTACTCTCTCTTTCTTACGGGTTTC
GTCTATCCCGTCGTTGTACGCTCTGTCTGGAGCTCCAACGGGTTCTTGTGTCAGCCTTCAGTGCCGAC
CCCTTCCAAGGAGTTGGAACCGTTGACTTTGCCGGATCAGGTGTGGTGCACATGACTGGAGGACTC
ACCGCCTTGATTGCTGCCATTGTTCTTGGACCGCGTAAGGGTCGGTTCTACGATGAGGATGGCAAC
CCCCTGGAGACGCCGCCAGCTTCCCAGCCCCTCTGTAGCCCTCCAGATCCTCGGAACCTTTCATC
TTGTGGTTTCGGATGGTACGGATTCAACCCTGGTTCAGCCCTGAAGATTGCTAACGCCGATTCGGCC

FIGURA 5

ACAACCGCCGCTTTGTGTGCCGTACCACTATGGCCGCGCTGCCGGTTGTGTTTCCGCCATG
TTCCTGACTCGATCATTGACGGCATGGCGACCGGTGAACTACGTACGATCTGACCATGGCCATG
AATGGATGCCTTGCTGGTCTCGTTGCCGTACTGCTGGTACATCTGTCGTACCCCATGGGCCGCA
ATCATTATTGGAGTCATTGGAGGTTGGGTCTACATTGGTATGTCCAAGCTTTTGATAAAGCTCAAG
ATTGATGACGCTGTCGATGCCATCCCTGTCCATTTTCGCCAACGGTTTCTGGGGTGTCCTAGCCACC
GGCCTTTTCGCCAACGGTGGATTGATGGCAACCGCTGGGTACAACCTCGGAACACGAGGGCTGGTTC
TACGAATGGGGAAGTGGCTCCGGAGATGGAAGTCTTCTCATCTGCCAGCTTGCTTGCCCTCGCCTGG
ATTATTGGATGGGTACCACTATTATGACGCCCTTTTTTATCCTTTTGAACATGGCCGGTATGTTT
CGTGTGGACCCGCTTGAGGAAGAAGTTGGTCTTGATATTTCCCATCACCGTGGATCTGCTTACGAT
CTTTCGGGACCCAGCAAGGACCATGTTGACGAGCTCATGGAAATTCTGTGCCTCGAAGCACGGCAAG
GTTGAGGTTCCAAAGGAGGTTGCGCAGGCTGCTGATGACGCCGCCGAAGAGACTGCTTAA

SEO ID NO: 4 Secuencia parcial del polipéptido Phatr AMT1 de *Phaeodactylum tricornutum*

MMQAGFAMLCAGSVRQKNVKNIMLKNLLDACGGAIGFYTVGFGFAYGGDDTSDKTFIGNSYFALRD
YTNYAGFFFQFAFAATAATIVAGTVAERCKMSAYLCYSLFLTGFVYPVVRSVWSSNGFLSAFSAD
PFQGVGTVDFAAGSGVVHMTGGLTALIAAIVLGRPKGRFYDEEDGNPLETPASFPASVALQILGTFI
LWFGWYGFNPGSALKIANADSATTAALCAVTTTMAAAAGCVSAMFTDSIIDGMATGETTYDLTMAM
NGCLAGLVAVTAGTSVVTPWAAIIIGVIGGWVYIGMSKLLIKLKIDDAVD AIPVHFANGFWGVLAT
GLFANGGLMATAGYNSEHEGWFEWGS GSGDGSLLICQLACLAWIIGWVTTIMTPFFILLNMAGMF
RVDPLEEEVGLDISHHRGSAYDLSGSPSKDHVDELMEIRASKHKGVEVPKEVAOAADDAAEETA

SEQ ID NO: 5 Secuencia de ácido nucleico Phatr_AMT2 de Phaeodactylum tricornutum

ATGTTCGAGTAGTGCCTATCTATCAATCCTTGCGCCGGTCAATTTCGATAGCGGTGAGCAGCTTGATCAG
CTTCTGCAGTGCCTCTCTACGGGACATGACGGCGCGCTAAGCGATCAGACCAGCAACCTCGCAGGA
GGTATCGATGCCTTTTATCTTATCTTTGCCGGGGGCCCTTGTCTACTTCATGCAGACAGGATTTGCT
ATGCTCTGTGCCGGATCCATTTCGAGCGAAGAATGTCAAGAATGTGATTCTCTGGAACCTTGCTCGAT
TCATGTGGAGGCGGTCTCGCGTTCTGGAGTGTAGGCTATGCCTTTGCCCTACGGTGGAGACAATGCT
GGGTCCAAGACATTTGTTGGTAACGCAGGCTTTTTTCCCTCCAGGGAGATGACATTCGACTGGAAAAC
TGGTTTTTCCAATTTGCCTTTGCTTGTGCCCTTTCCCTCAATTGTCGCTGGAACGATCGCTGAGCGC
ACTCAAATGAAGGCCTATTTGATGTATTCCGTTTTTTTTGGCTGGGTTTGTCTATCCAGTTGTCGCC
CACGCATTTTGGTCGAGCAACGGATTCTCTCCAACACAGCCACAGATCCATTGTGGGGATCTGGC
GCCATTGACTTGGCGGGCTCTGGACCAGTCCACATGACCGGAGGTGTTACGGCTTTGGCGGCAGCT
CTGGTTTTTGGGCCCTCGCATTGGTCGCTTCTACGACAAGGAGGGCAATCCTCTCGAAGAGCCAGCC
GAATTTCCCTCCCCATTTCGGTTGCCCTTGCAGTTTTCTGGGAACGTTTTTGTCTTTGGTTTTGGCTGGTAC
GGCTTTTAACCCAGGTTCCGCTCTTCTTCATTTTCGAGTATTGAAAACGGTCAAGTTGCGGCTTTGGTA
GCCGTCAATACCACCCTAGCTGCGTGTGCGGGTGCCGTGAGCGCCATGTTTACTTCAACTTTATTT
GACTACTGGTACACCGGCTTGCACACATACGATTTGGGCTACACTATGAACGGATGTTTGACGGGA
TTGGTCGCGATAACAGCGGGATGCGCGACTGTTGAAACGTGGGCTGCCGTCTGATTGGTATTGGC
GCTGGTTGGTTTTACCTGTTGGGTTCAAAGTTGCTTGTCTACTTCCGTATTGACGATGCTGTGCGAC
GCCATTCCCGTGATATGGTGGGTGGCGCCTGGGGTGTGATCGCGACGGGCCTTTTTACGAAGGGA
GAACTACTATTGGCCGCCTTTGGCCAGGAGGAACATGTTGGATGGTTCTACGAGTGGGGCAGTGGG
AGCGGTAACCTTTACCCTAATCGGAATTCAATTGCTGTCTGTGCTGTTTCAATTTTCGCCTGGACTTTT
TCCGTCATGGGAATTTATTTTTTACGCCCTTAGCTTCATGGGTTGGTTGCGCATAGATCCGTTGGAA
GAGGAGGTTGGTATGGATATTTTCGCGCCACAAGGGCTCGGCATACGACATGACTTCAGCAAATATG
GAGCAGGTACGGTCGTTGATGGACGACCGCAGTACGAGCAATCGAGGCAAGCTGCGTAAGTCGTCC
ATCATTGAGAAGCCGGCCAAAACAGGTCCGATCGAGGGAGGAACTGCCTCCGGAGAATCAGAAATT
AAGGAGTCGGACGTGGAACATGCCGAGAGGGTCTAA

FIGURA 5 (continuación)

SEQ ID NO: 6 Secuencia traducida del polipéptido Phatr_AMT2 de *Phaeodactylum tricornutum*

MSSSAIYQSCAGQFDSGEQLDQLLQCLSTGHDGALSDQTSNLAGGIDAFYLIIFAGALVYFMQTGFA
MLCAGSIRAKNVKNVILWNLLDSCGGGLAFWSVGYAFAYGGDNAGSKTFVGNAGFFLQGDDIRLEN
WFFQAFACALSSIVAGTIAERTQMKAYLMYSVFLAGFVYPVVAHAFWSSNGFLSNTATDPLWGSG
AIDLAGSGPVHMTGGVTALAAALVLGPRIGRFYDKEGNPLEEPAEFPFPHSVALQFLGTFCFLWFGWY
GFNPGSVFFISSIENGQVAALVAVNTTLAACAGAVSAMFTSTLFDYWYTGLHTYDLGYTMNGCLTG
LVAITAGCATVETWAAVLIGIGAGWFYLLGSKLLVYFRIDDAVDAIPVHVMVGGAWGVIATGLFTKG
ELLLAAGFQEEHVGFYEWGSGSGNFTLIGIQLLSVLFIFAWTFVSVMGIYFYALSFMGWLRLIDPLE
EEVGMDISRHKGSAYDMTSANMEQVRSLMDDRSTSNRGLRKSSIIEKPAKTGPPIEGGTASGESEI
KESDVEHAERV

SEQ ID NO: 7 Secuencia de ácido nucleico Phatr_AMT3 de *Phaeodactylum tricornutum*

ATGAATAGCCGTACCTTTTCTTACAACAGCAACGATGGGAATGAGTTGTTGGATGCATGCATGGCT
TATCTCGGCGCTAACGCCACGACGTACGACTTGCTCGGCTGCGTGTCTGCGCAACTCAGCAACGAA
GTCGGGAGTCGCGAATTCTCTCGTTTCGGTGCTGCTCGTCTACGCTGCTGCACTCGTATTCTTTATG
CAAGCCGGCTTCGCCATGCTCTGTGCCGGAGCCGTCGGGAAAAGAACGTGCAGAACACCATGCTC
AAGAACTTGTTGGACGCCTGTGGAGCCGCCGTCGCCCTTTTTCATCGTGGGCTACGCTATAGCCTTT
GGAGGCATGGAACCGGAATCGCCCAACAAGACATTTCTCGGGAACACCAATTTCTTTTAAATGGGA
GTGGACGATTTGGCCTTTTGGTTGTTCCAGTACGCCCTTTCCGCTGCCCTCTGCAACCATCGTCGCG
GGAACCTTGGCCGAGCGCTGCCAAATGGTTGCCTACTTGTGTTACTCCGTAATGTTGACGGGATGG
GTCTATCCGATTATCGCGCACGCGATTTGGTCACCCAACGGTTGGCTCTCCGCCAGTTCCGTGGAT
CCGCTATGGGGTGTGCGCATGGTCGATTTCCGCCGCTCGGGAGTAGTGCACATGACTGGAGGCGTT
ACCGCCTTGTTTCGCCACTCTCATACTCGGACCTCGCCGTGGACGGTTTCACGATGAAACCGGACGC
CGGCTCGACAAACCAAATCCTTTCCCGGACACTCCGTGGCCTTGACAGATGCTCGGTACCTTCATT
CTCTGGTTTGGTTGGTATGGATTCAATTGCGGTTTCGGCGTTACTCATCGACAAGCCCGGTGCTAAC
GATATTGCCGCTTGGCGGGCGTCAATACGACACTTTCTGCCGGAGTCGCTGGGATAGTTGCACTC
TTTGTCAACCTCTGGTACCTCGAGCGAACAACCGGAGAACCTTTTTTTGATTTGACCTACGCCATG
AACGGATCCCTTTCGGGTCTCGTGGCCATCACAGGAGGTTGCGCTGTTCTCGAGCCCTGGGCGGCA
GCGGTCACTGGAGTCGGTGCCGGTATCTTGTACATGGTTGGATCACGAGGTTTGGTCATGTTGCGA
TTGGATGACGCGGTTGACGCGATTCCCCTTCATTTTGTGAACGGGGCCTGGGGTCTCATGTGCGTC
GGATTGTTTCGCGTCACCAGCACGCTTGCTGGCCGCGTACGACAACGACGCTCATCTGGATGGTTC
TATTCCCTACGAAACGGCAAATCGGACGGGCGTTTGGTTGGAGTTCAGCTGGTAGGCATTGTATTT
ATCGTGGGATGGGTCATGGTGATTATGTTGCCCTTCTTCATTTGGTTGGATTGGAAGGGATGGTTC
CGGTGCGATCCTTTGGAAGAAATTGTGGGATTGGACACGTCGTACCACGGTGGTCTCGCTTTGCTG
GGCGGGGACGATGAAGTCAACCCTGAGTACATTTCCGCGTACAAGAAGCAAAGGAACGAAGGTACG
CTGCGGCGTAGGCATAAAGGGACAACCAGCAGTGTCAAGACGGGAGAAGTAGAATCGGACGAAGGC
GCCGAACGGGTGGCGCCCGAAGCGACGAAGCACAAAAAGATTCCGATCGCTACAACGACAGAGGTT
TTCAATGGAGGTCAAGATACGGATCAAATGTCCTACCACGCCTGA

SEQ ID NO: 8 Secuencia traducida del polipéptido Phatr_AMT3 de *Phaeodactylum tricornutum*

MNSRTFSYNSNDGNELLDACMAYLGANATTYDLLGCVSAQLSNEVGSREFSRSVLLVYAAALVFFM
QAGFAMLCAGAVRKNVQNTMLKNLLDACGAAVAFFIVGYAIAFGGMEPESPNKTF LGNTNFFLMG
VDDLAFWLFQYAFSAASATIVAGTLAERCQMVAYLCYSVMLTGWVYPIIAHAIWSPNGWLSASSVD
PLWGVGMVDFAGSGVVHMTGGVTALFATLILGPRRGRFHDETGRRLDKPKSFPGHSVALQMLGTFI
LWFGWYGFNCGSALLIDKPGANDIAALAGVNTTLAGVAGIVALFVNWLWYLERTTGEFFDLTYAM
NGSLSGLVAITGGCAVLEPWAAAVTGVGAGILYMVGSRGLVMLRLDDAVDAIPVHFVNGAWGLMSV

FIGURA 5 (continuación)

GLFASPARLLAAYDNDAHPGWFYSLRNGKSDGRLVGVQLVGIVFIVGVWMVIMLPFFIWLWDKGW
RSDPLEEIVGLDTSYHGGLALLGGDDEVNPEYISAYKKQRNEGTLRRRHKGTTSSVKTGEVESDEG
AERVAPEATKHKKIPIATTTEVFNGGQDQDMSYHA

SEQ ID NO: 9 Secuencia de ácido nucleico Phatr_AMT4 de *Phaeodactylum tricornutum*

ATGGACGACGCGAGCTTTATCCAAAGCCTGGTGAAGGGTACGGTACCTCGAGCAACCACACCACC
GTCTATGGATATTGCTCCAATGAGGCGGAAGGAGAGACCAACCTTATTTTGCAATGTATTACGGAG
GTCATGGAGCAAAAGCAGCTCGAAGGAGATCGTAACGTGAATCGGTGGCTAATGCTTTTTTCCGGT
GGGCTGATCTTCTTCATGCAGACTGGCTTTGCGATGCTTTGCGCTGGTTGCGTGCGCAAAAAGAAC
GTTCAAAATACGATGCTTAAAAATCTTTTGGATGCCTGTGGAGCTGCTCTAGGCTTTTTCTACTA
GGCTACGCTTTTTCGCTTTGGCGGACAAGACGACCGGGACGATGTTACTTTTCATCGGAACCTCCAAC
TTTCTAAACACCGGAAAAGTCGATATGTCCTTCTGGTTTTTCCAGTTTGCATTCTCGGCCACTGCC
GTGACTATCGTCGCAGGAACCTTTGGCAGAGCGTTGCCAAATGGTAGCATACCTTTGCTATTCCATA
TTTTTGACAGTTTTTGTGTACCCGGTTGCCGCCACACAATTTGGTCTCGCAACGGCTTCCTCAGC
AGCACAGCAGTAGACCCCTTTCAAGGTGTGGGAGCGATTGATTTCCGCCGGATCCGGTGTGGTTTCAT
GTGACGGGAGGAACCACGGCTCTCGTCGCCACATATATTCTTGAGCTAGAAAAGGTCGTTTCTAC
GACAATCGCGGTGCGCAGCTGGAGACGCCAAAGTCCTTTCCGGGGCATTAGTTGCTCTGCAGCTA
CTTGGTACATTTCGTATTGTGGTTTTGGATGGTACGGCTTCAATCCTGGATCAGCATTGCTGCTGGCC
CACACTTCGGATACGGGTTCGTGGCATCCCGAGCGGTGTCAATACCTCGCTTTAGCTGCTTCC
GGAGCCGTATCGGCACCTGATGACAAATATGTTTATGGAAGAGCGTTCCACCGGGCAATACTCCTTT
AACATTATCATGGCCATGAATGGCGCCTTAGCTGGTCTAGTGTCAATTACAGCAGCTTGTGGTACC
GTACAAAACCTGGGCAGCCCTTTGCACTGGATGTATTGGAGGTCTTATCTACTTGTGGGGTTCCAAA
ACGCTGGTTTCGTTTGAATTTGGACGATGCTGTTGACGCGATTCCAGTGCACATGTTTGCAGGTGGC
TGGGGATTGCTTGTGTGGTTTTGTTAAGTGATCCTGACCTCATGCACATTGCCTATGGGACTGGC
AATCATCCTGGTTTTATTGTATTCTTGGGGGTAGGGGAGTTTAACGCCATCTTGCTGAGCAATCAA
GTGTTGGAAGTAGTTTTCTGTGGCGGGATGGGCTTTTGGTACCATGACGCCTTTCTTTTTTGTTCATC
AATCGTATGGGATGGTTCCGGTCGGACAGTTTGGAGGAGTTGGTTCGGTTTGGATGAGGCCTACCAC
GGAGGTAAACATGGCGGGGAAGAAGTGGTGGAACTATCCGCTCTAGAAGGCTTTATCAAAAACAAG
ATCAGACAATCGCAAATTCGGTCGCCTTACAATTAG

SEQ ID NO: 10 Secuencia traducida de polipéptido Phatr_AMT4 de *Phaeodactylum tricornutum*

MDDASFIQSLVEGYGTSSNHTTVYGYCSNEAEGETNLILQCI TEVMEQKQLEGDRNVNRWLMLFSG
GLIFFMQTGFAMLCAGCVRKKNVQNTMLKNLLDACGAALGFFLLGYAFAGGGQDDRDDVTFIGTSN
FLNTGKVDMSEFWFFQFAFSATAVTIVAGTLAERCQMVAYLCSIFLTGFVYPVAAHTIWSRNGFLS
STAVDPFQGVGAIDFAGSGVVHVTGGTTALVATYILGARKGRFYDNRGRQLETPKSFPGHSVALQL
LGTFLVLWFGWYGFNPGSALLLAHTSDTGFVASRAAVNTSLAASGAVSALMTNMFMEERSTGEYSF
NIIMAMNGALAGLVSITAACGTVQNWALCTGCIGGLIYLWGSKTLVRLKLDDAVIDAIPVHMFAGG
WGLLAVGLLSDPDLMHIAIYGTGNHPGLLYSWGLGEFNAILLSNQVLELVFVAGWAFGTMTPFLLFI
NRMGWFRSDSLEELVGLDEAYHGGKHGGEEVVELSALEGFINKKIRQSQIRSPYN

SEQ ID NO: 11 Secuencia de ácido nucleico Phatr_AMT5 de *Phaeodactylum tricornutum*

ATGAGTCACTCTAGTTTGGATGTGTTCCGGACCTGCCTTGCCCAAGTTGGCGAAGACGCAACGACA
AAGGAGCTACTAGAATGCGTTTCCTTTAGTTTGTGCGAGCAGTGCCAGACGGTCTCGATGAGCCT
AGTTCTAAAGGCTTCACGCGGTCTATCGTTGTGCTGTTTGGCGCTGCCTTGGTTTTTTTCATGCAA
GCCGGCTTTGCCATGCTGTGCGCTGGAGCCGTTAGGGCCAAGAATGTTCAAAACACCATGCTCAAG
AATCTTTTGGATGCTTGTGGTGGCGCCATTGCGTTCTTTACTGTAGGCTACGCCTTTGCCTTTGGC

FIGURA 5 (continuación)

GGTACGGACTTTCCACCGACACCGACACCGACACCGGCAACAACCAACAACATTTCATTGGCACA
 TCGAACTTTTTCTTGGTGAATGTGGACGATTATTCCTTTTGGCTGTTTCAATACGCATTTTCCGCC
 GCGTCCGCAACCATTGTTGCTGGAACACTGGCTGAACGGTGTCAAATGGCCGCCTACCTGGGATAT
 TCCGCTTTGTTAACGGGATGGGTGTACCCATTGTGCTCACGCTGTATGGAACGTCCACGGCTTC
 CTATCGGCTCATGCGGTAGAACCTTTGTGGGGGGTGGAAATGGTAGATTTTGGCGGTTCTGGGGTC
 GTTCACGTTACGGGAGGTGTGACAGCGCTCTTCGCGACAATAATTCTGGGTCCCCGCCGTGGACGC
 TTTACAGACCAAGATGGCCAAAGATTGATAAGGCCGCGAATATTTCTGGACACTCCTTTGCTCTG
 CAAATGCTTGAACCCCTTATCTTATGGTTTGGCTGGTACGGATTCAATATTGGCGCTGCTCTGCTG
 CTGGACGTGCCTAGTTTACAGACAATATTGCAGCTCTGGCTGCGGTGAATACGACCTTGTGGGGCGG
 ACGGCTGGCATCGTTGCTCTCTTTTCAATTTGTGGTATCTGGACAAAAGAACTGGCGAAGCCTAT
 TTTGACTTGAAATTTGCCATGAATGGGTGCCCTCTGCGGTCTCGTGGCCATCACTGGAGGCTGTGGT
 GTGGTCAACCCTGGGCTGCCGTTGTGATTGGCTTTGTGCGCGGTTTGTGTACAACATCGGAAGT
 CGCGGACTTATATATTTGCGTTTGGATGATGCCGTGGATGCCATTCCGGTACATTTGTGCAACGGC
 TCGTGGGGTCTCGTAGCAGTAGGGCTGTTTGCTTCCCCCTCCCGGCTACTAGTTATTTACGGACAT
 AGCGATCATCCAGGATGGTTTTACTCATTACGAGATGGCGAATCCGATTTCCGGTTACTGGCTTCC
 CAGTTGGTGGGCCTTATTTTCATTGTCTTCTGGGTGATGTTCAACATGCTGCCTTTTTTTCGTTTGG
 TTAACACTACCGTGGCTGGTTTCCGTCGATCCGTTGGAAGAGTTGGTGGGTCTCGACCTGAGCTAT
 CACGGCGGTTTAAATGCTGCACGAGGAAGTTGATCCCGAATACATATCGGCCTATCGCAAGGGCCAA
 CACGAGGCTCATTCTCGTACTCTGCGACAACGGAAACGCACATCCACGTGCGCTTGGAAATCAGTG
 AGCGAACATTCCGTTGCGCCAAATGGCGGAGACTCAGCCATACTAGTAACGGCGATTCTGCTCAG
 GAAACGGTCACTATCCGTGAACCGGACTCCGGTTGGCCAAGGGGGAACGTATAA

SEQ ID NO: 12 Secuencia traducida del polipéptido Phatr_AMT5 de *Phaeodactylum tricornutum*

MSHSSLDVFGTCLAQVGEDATTKELLECVSFSLSRVDPGLDEPSSKGFTRSIVVVFAAALVFFMQ
 AGFAMLCAGAVRAKNVQNTMLKNLLDACGAAIAFFTGVYAFAGGTDFTDTDGNNQTTFIGT
 SNFFLVNVDYSFWLFQYAFSAASATIVAGTLAERCQMAAYLGYSALLTGWVYPIVAHAVNVHGF
 LSAHAVEPLWGVGMVDFAGSGVVHVTTGGVTALFATIILGPRRGRFHDQDQRLIRPRIFPGHSFAL
 QMLGTLILWFGWYGFNIGAALLLDVPSSDNIAALAAVNTTSLGGTAGIVALFFNLWYLDKRTGEAY
 FDLKFAMNGCLCGLVAITGGCGVVEPWAAVVIGFVAGLLYNIGSRGLIYLRLLDAVDAIPVHLCNG
 SWGLVAVGLFASPSRLLVIYGHSDHPGWFYSLRDGESDFRLLASQLVGLIFIVFWMFMNLPFFVW
 LNYRGWFRSDPLEELVGLDLSYHGGLMLHEEVDPEYISAYRKGQHEAHSRTLQRKRTSHVRLESV
 SEHSVAPNGGDSAHTSNGDSLTTETVTIREPDSGWPRGNV

SEQ ID NO: 13 Secuencia de ácido nucleico Phatr_AMT6 de *Phaeodactylum tricornutum*

ATGAGCACATCCGATGGCAGTCTGTTTCAACAGTGTCTGCGCTCGCTGGGGACAGTGACCCTTCC
 CGGATTCTGCAGTGTGTGTCGATGCTCTCGAAACCAACCAAAACGATCGAGCTGCCGATCTCAAT
 AACTGGTTTCTCATCATCGCTGGGGCACTCGTCTTTTTCATGCAGTCCGGCTTCGCCATGCTGTGT
 GCGGGGTGTGTGCGGAAAAAGAACGTCCAAAACACCATGCTCAAGAATCTTCTCGATGCCTGCGGT
 GCCGCTCTCGGGTTTACGTAATCGGTTACGCCTTGGCCTTTGGTGGACAAAACGAGCGATCCGAT
 GTTACCTTTGTGGGTACCACGGACTTTTTCAACTGGAACAGCGCCGTCGCCGTAAATCAAGCCTTT
 TGGTTTTTTCGAATTTGCCTTTTCGGCCACGTCCGTGACGATTGTGGCCGGGACGCTGGCGGAACGG
 TGTCAAATGGTCGCCTACCTGTGCTACTCGGTATTCTTAACAGGCTTTGTGTATCCCGTTGTGGCC
 CACTCCATCTGGAGCAACAACGGCTTTCTAAGTGCATTTGCAGCCGAGCCCTTCCAGGGCATTTGGC
 GTCCTGGACTTTGCCGGGTGCGGTGTGGTACACGTGACGGGCGGGACCACTGCCTTGGTCGCCACG
 TATATGCTAGGGGCCCCGTAAGGGGCGCTTCTACGACGCACGTGGGAGAGAATTGAAAAGCCCCAA
 GCATTTCCCGGCCATTCCATGGCCTTGCAGATGATGGGCACCATGATTCTGTGGTTCGGATGGTAC
 GGCTTCAACCCTGGTTTCGCCCTTGTGCTGACGGCGACGTCCAATACAGGTGGCGTAGCGGCTCTC

FIGURA 5 (continuación)

GCGGCCGTGAACACGTCCCTCTCTGCCGCTTCGGGTGCCGTTTCCGCTCTCTTTACCTCTTTGTAT
CTCGAAGAACGCAAGACTGGAGAATACTCCTTCAACATCACCATGGCTATGAACGGTGCTTTGGCT
GGCTTAGTGGGTATTACGGCCGGCTGTGGCACCCTCGAAAATTGGGCCGCTTGTTCACAGGGCTC
GTGTCTGGATGGGTCTATATATTTGGCAGTGCCTTCTTGTTCGTATAAAGATTGACGATGCAGTC
GACGCAATTCCGGTGCATATGTTTTGTGGCGCCTGGGGTCTTATTGCAACCGGCCTATTGAGCTCG
CCACGCCATACGTTAGAAGCATTGTTGACGGACGCACACGTAGGTTGGTTCTACAGTCTTGCCAA
GATTCTTTGGACGCTATTCTCTTGATGAACCAGCTTTTGGGTCTTTTATTTATTCTCGGGTGGAGC
GCCATCTTGATGTCACCATTTTTCTGGTGGCTGAACTACATGGGATGGTTGAGGGCCGACTCCTTG
GAGGAGTTGGTAGGTCTGGATCAAGCCTACCACGGTGAAGGGAGGCTGGAGAGGGCTTTGATGAT
GAAGTGCCTTTTGGCACATCCAAAGACAAACCAGACACTTTACGTGACGGAAGGTGGCAACAGC
GCGGAGGGCACCCGCGATGATGGAAGTTGGACCGATTTGACTGCCTCGGCTCCTCGGGAACTGCG
CCTCTGCCAGAACTTTATTGTGAGAACGATGCGTCTTCCAAAGAATCCGACGACGATCCGTCCGA
ATGCACCACGCATAA

SEQ ID NO: 14 Secuencia traducida del polipéptido Phatr_AMT6 de *Phaeodactylum tricornutum*

MSTSDGSLFQQCSAVAGDSDPSRILQCVSDALETNQNDRADLNNWFLI IAGALVFFMQSGFAMLC
AGCVRKKNVQNTMLKNLLDACGAALGFYVIGYALAFGGQNERSDVT FVGTTDFNWN SAVPVNQAF
WFFEFAFSATSVTIVAGTLAERCQMVAYLCYSVFLTGFVYPVVAHSIWSNNGFLSAFAAEFPQIG
VLDFAGSGVVHVTGGTTALVATYMLGARKGRFYDARGRELEKPKAFPGHSMALQMMGMTMILWFGWY
GFNPGSALLLTATSNTGGVAALAAVNTSLSAASGAVSALFTSLYLEERKTGEYSFNITMAMNGALA
GLVGITAGCGTVENWAACCTGLVSGWVYIFGSAFLRLIKIDDAVDAIPVHMF CGAWGLIATGLFSS
PRHTLEAFGTDHVGWVFYSLGQDSLDAILLNQLLGLLFILGWSAILMSPFFWWLN YMGWLRADSL
EELVGLDQAYHGGREAGEGFDDEVFATSKDKPDTLRRRKGGNSAEGTRDDGSWTDLTASAPRETA
PLPELYCENDASSKESDDDP SGMHHA

SEQ ID NO: 15 Secuencia de ácido nucleico Phatr_AMT7 de *Phaeodactylum tricornutum*

ATGATCTCTACTGGCTCTTCCACGTCCACGAATGCCTACGGCACTTGCTCGGTT CAGCTGGGCGAA
AACTCTTCCGCGAAAGAGCTCTTGGAATGTGTTTCAGATTATCTACAGAATCAGGAGGCGCCTTTC
TCGTCTACTCTGGTACTCACTTTTGCGGGCGCGATTGTCTTTCTAATGCAAGCCGGCTTTGCCATG
GTCTGTGCTGGTGCTGTCCGCACAAAAAATGTCCAAAACGCCATGCTCAAGAATCTGTTGGATGCT
TGCGGTGCGTCCTTGGCGTCTTTTCCATTGGCTACGCTCTGGGGTTTGGGGGTATGGAGCCCGAA
AGTTCCAAAAGACCTTTGTGCGACATAGTCAATTCTTTCTGATGGACGTTGACGACTACGCTTTT
TGGTTATTCCAATACGCCTTTTCCGCTGCATCTGCCACAATTGTCGCAGGAACGCTCGCGGAACGA
TGTCAAATGACGGCCTATCTTTGCTACTCCCTTATGCTAACCGGGTGGGTATACCCCGTTATTCTA
CACTCCATATGGAATCCCAACGGCTGGCTATCTGCGTACTCGGTTGATCCGCTTTGGGGCAGCGGG
CTGGTGGATTTTGCTGGCTCTGGCGTCGTTTATGTGACCGGAGGAATCACCGCTCTGTTTGCTACA
ATGGTTTTTGGGCCCCCGACGAGGACGCTTTCACGATGATCTTGGGCATGATCTGGCACGGCCACGG
GAATTTT CAGGCTCATTGCGCGGCTCTGCAAATGCTTGGTACTTTTATTTTGTGGTTTGGCTTTTAC
GGATTTAACATTGGCTCCGCGCTGATAAGCACAAAGCAAGGCTCAGACGAAGCAGCCGCCTTGGCC
GGTGTCAACACAACGCTCTCCGCTAGCGCGGCAGGTATTGTGGCGCTGTTTTCAAATCTATGGTAC
CTGGAGAAGACGACCGGTGAGCCTCTCTTTGATCTAAAGTATGCTATGAATGGCGCCATCTGCGGT
CTTGTAGCTATTTCTGGCGGCTGCGGGGTCTTCGAGCCCTGGGCCGCGGTGGTCACCGGGGCGGTG
GCCGGTGTGATATATCTATTAGGTAGTCGCGGACTTGTATCCATGCGACTGGATGACGCTGTGGAT
GCCATCCCTGTACACCTCTGTGGTGGGGCCTGGGGCATCCTGGCGGTGGGACTGTTTCGACGCGCCC
GAACGTCTCCTTTCCGTGTATGGACGGAATAATCACCCAGGATTGGTTTATAGCATTCGTGAAGGA
GATATCGATGGCGTTCTTTTCGGAATTCAACTAATTGGTCTCATGTTTCATCATGGGATGGGTATG
ATCATCATGCTACCTTTCTTTGTATGGCTCAACTGGAAGGCTGGTTCGGATCAGACCCGCTGGAG

FIGURA 5 (continuación)

GAGATTCTCGGGCTCGATTTAAGCTATCACGTCGGATTGGCGTTACATACCAACAATGTTTCATCCG
GAATACGTTGGCAGCGAAAAAGATGTGGTCGACGAAATTATTTCTACGCGCCAACGAAAGGTCAAC
GGGAGCACCCTACCAAGGCGACCTCAGGTACGGAAGAATTGGAGTATATCCCGGAAGTTAGCGAC
GAAGACTTGAGCGAGATGAAAGAGGAATGCTTGTGA

SEQ ID NO: 16 Secuencia traducida del polipéptido Phatr_AMT7 de *Phaeodactylum tricornutum*

MISTGSSTSTNAYGTCVQLGENSSAKELLECVSDYLQNEAPFSSSTLVLTFAAGIIVFLMQAGFAM
VCAGAVRTKNVQNA MLKNLLDACGASLAFFSIGYALGFGGMEPESSKKTFFVGHSSQFFLMDVDDYAF
WLFQYAFSAASATIVAGTLAERCQMTAYLCYSLMLTGWVYPVILHSIWNPNGWLSAYSVDPLWGS
LVDFAGSGVVHVTGGITALFATMVLGPRRGRFHDLDLGHDLARPREFQAHSPALQMLGTFILWFGFY
GFNIGSALISTKQGSDEAAALAGVNTTLSASAAGIVALFSNLWYLEKTTGEPLFDLKYAMNGAICG
LVAISGGCGVFEPWAAVVTGAVAGVIYLLGSRGLVSMRLDDAVDAIPVHLCGGAWGILAVGLFAAP
ERLLSVYGRNNHPGLVYSIREGDI DGVLFGIQLIGLMFIMGWVMIIMLPFFVWLNWKGWFRSDPLE
EILGLDLSYHVGLALHTNNVHPEYVGSEKDVVDEIISTRQKRVNGSTTTKATSGTEELEYIPEVSD
EDLSEMKEECL

SEQ ID NO: 17 Secuencia de ácido nucleico Cylfu_AMT1 de *Cylindrotheca fusiformis*

ATGGCTGAGTTTGATAACACGTTTATCTTGGATTTTGCAGTGGTGGAAATGAATCGTCGTCTGAC
GTTTCAGGCCTTGCTCAAGTTGCTGGTCTTGCAAATGGCACAAGTGCCTCTGCCGCTGGACTTGTT
GAGGGTATCAACACCTTCTTCTACTTTTCGCAGGAGCTCTGGTCTTTCTCATGCAAGCTGGCTTC
GCCATGCTGTGCGCTGGATCCGTCCGTCAAAAGAACGTCAAGAATATCATGTTGAAGAACATGTTG
GACGCTTGTTGGTGGTGTCTATTGGTTTCTGGACTATCGGGTATGCGTTTCGCTACGCTGATAACTCG
TCGGGAGACAAAACCTTCATTGGAGGCAAGAACTTCTTTGTGAACCAATTGGATGAGTCTGGTGGC
GCATGGATTGGTTTCTTTTCCAAATTCGCCTTTGCTGCCACTGCCGCCACTATTGTGCGCCGAACC
GTTGCTGAGCGCTGTAAGATGAGTGCATACCTTTGCTACTCAGTCTTCCTCACTGGTTTTGTCTAT
CCTGTCTGTTGTTCACTCCATCTGGAGTGTGATGGATGGTTGACTGCCTTCCGTGATGATCCTTGG
AAGGGTGTCTGGTGTCTATTGATTTTCGCCGGATCTGGTGTGTGTCACATGACCGGTGGAGCCACTGCT
CTTGTTGCTGCTATTGTTCTTGGACCCCGTAAGGGACGTTTCTATGATGAGGATGGAAATGCTCTT
GAAACCCCTGCAAGTTTCCCAGCTCACAGTGTGGCCCTTCAAGTCCTTGGTACCTTCCTTCTTTGG
TTCGGATGGTATGGATTCAACCCAGGATCTGCTCTTGTCTATTGACAATGCTGCGTCCGCTTCTACC
TCAGCACTTTGTGCTGTCACTACCACTCTTGCCGCTGCCAGTGGTTGTGTCTGTGCGATGTTCACT
GATACCATCATTGAAATGATGGCAACAGGAGAAGCATCTTACGATTTGACCATGGCCATGAACGGT
GCTCTTGGAGGACTTGTTGCCATCACTGCTGGATGCTCCGTTGTCAACCCTTGGGCTTCGATCATC
ATTGGTATCATTGCTGGGTGGGTCTACATTGCATTTTCCAAACTCCTAGTCAAATTGAAGATCGAT
GATGCTGTTGATGCCGTTCTGTTCACTTCGCCAATGGTATGTGGGGTGTCTTGGCTGTGGGTTTC
TTCGCCGAACCCGACGCCATGGTTACTGCCGGGTACAACGATGTCCAGGAGTCTTCTACAAGGGA
GATGGTAAACTTCTCCTGTGCCAATTTGTGCTATCATCTGGATTTGCGCTTGATCTTCTTCTTG
ATGACTCCTTTCTTCGTCGTCTTGAACATCTTGGGTATGTTCCGTGTGATCCTCTTGAGGAAGAA
GTTGGTCTTGATATTTCCACCATCGTGGAGCTGCCTACGACATGACCAAGTCCCAAGAAGGAAGAT
GTCGAGGAGCTCATGGAACACCGTTCTTCTAAGCATGGAAAGGTTGAGGTCCCCAAGGAAGTTCAA
AAGGAGGACACCGCCTAAGGAATTTGACATCCCTTCTAG

FIGURA 5 (continuación)

SEQ ID NO: 18 Secuencia traducida del polipéptido Cylfu_AMT1 de *Cylindrotheca fusiformis*

MAEFDNTFILDFCSSGNESSSDVQALCQVAGLANGTSASAAGLVEGINTFFLLFAGALVFLMQAGF
 AMLCAGSVRQKNVKNIMLKNMLDACGGAIGFWTIGYAFAYADNSSGDKTFIGGKNFFVNQLDESGG
 AWIGFFFQFAFAATAATIVAGTVAERCKMSAYLCYSVFLTGFVYPVVVHSIWSADGWLTAFRDDPW
 KGVGVIDFAGSGVVHMTGGATALVAAIVLGPGRKGRFYDEDEGNALET PASFPAHSVALQVLGTFLW
 FGWYGFNPGSALVIDNAASASTSALCAVTTTLAAASGCVCMFTDTI IEMMATGEASYDLTMAMNG
 ALGGLVAITAGCSVVT PWASIIIGIIAGWVYIAFSKLLVKLKIDDAVDVAVPVHFANGMWGVLA VGF
 FAEPDAMVTAGYNDVPGVFYKGDGKLLLCQFVAIIWICAWIFFLMT PFFVVLNII LGMFRVDPLEEE
 VGLDISHHRGAAYDMTSAKKEDVEELMEHRSSKHGKVEVPKEVQKEDTA

SEQ ID NO: 19 Secuencia de la variante de ácido nucleico Cylfu_AMT1 de *Cylindrotheca fusiformis*

ATGGCTGAGTTTGATAACACGTTTATCTTGGATTTTGCAGTGGTGGAAATGAATCGTCGTCTGAC
 GTTCAGGCCTTGTGTCAAGTTGCTGGTGTAAAGTTGTTTGCCGATGTTATTCTGACCACAGTATATC
 ATTTTGCCTTGACTTACACCGTCGCGTCTTGATGTTGTGCACTTCAGCTTGCAAATGGCACAAGTG
 CCTCTGCCGCTGGACTTGTGAGGGTATCAACACCTTCTTCTACTTTTCGCGAGGAGCTCTGGTCT
 TTCTCATGCAAGCTGGCTTCGCCATGCTGTGCGCTGGATCCGTCGTCAAAAGAACGTCAAGAATA
 TCATGTTGAAGAACATGTTGGACGCTTGTGGTGGTGCTATTGGTTTCTGGACTATCGGGTATGCGT
 TCGCCTACGCTGATAACTCGTCGGGAGACAAAACCTTCATTGGAGGCAAGAACTTCTTTGTGAACC
 AATTGGATGAGTCTGGTGGCGCATGGATTGGTTTCTTTTTTCCAATTCGCCTTTGCTGCCACTGCCG
 CCACTATTGTCGCCGGAACCGTTGCTGAGCGCTGTAAGATGAGTGCATACCTTTGCTACTCAGTCT
 TCCTCACTGGTTTTGTCTATCCTGTCGTTGTTCACTCCATCTGGAGTGCTGATGGATGGTTGACTG
 CCTTCCGTGATGATCCTTGGAAGGGTGTGCGTGTCAATTGATTCGCCGGATCTGGTGTGTGCACA
 TGACCGGTGGAGCCACTGCTCTTGTGCTGCTATTGTTCTTGGACCCCGTAAGGGACGTTTCTATG
 ATGAGGATGGAAATGCTCTTGAAACCCCTGCAAGTTCCCAAGTCTCACAGTGTGGCCCTTCAAGTCC
 TTGGTACCTTCTTCTTTGGTTCGGATGGTATGGATTCAACCCAGGATCTGCTCTTGTCAATTGACA
 ATGCTGCGTCCGCTTCTACCTCAGCACTTTGTGCTGTCACTACCACTCTTGCCGCTGCCAGTGGTT
 GTGTCTGTGCGATGTTCACTGATACCATCATTTGAAATGATGGCAACAGGAGAAGCATCTTACGATT
 TGACCATGGCCATGAACGGTGTCTTGGAGGACTTGTGTCATCACTGCTGGATGCTCCGTTGTCA
 CCCCTTGGGCTTCGATCATCATTTGGTATCATTTGCTGGGTGGGTCTACATTGCATTTTCCAACTCC
 TAGTCAAATTGAAGATCGATGATGCTGTTGATGCCGTTCTGTTCACTTCGCCAATGGTATGTGGG
 GTGTCTTGGCTGTGGGTTTCTTCGCCGAACCCGACGCCATGGTTACTGCCGGGTACAACGATGTCC
 CAGGAGTCTTCTACAAGGGAGATGGTAACTTCTCCTGTGCCAATTTGTGCTATCATCTGGATTT
 GCGCTTGGATCTTCTTCTTGATGACTCCTTTCTTCGTCGTCTTGAACATCTTGGGTATGTTCCGTG
 TCGATCCTCTTGAGGAAGAAGTTGGTCTTGATATTTCCCAACCATCGTGGAGCTGCCTACGACATGA
 CCAGTGCCAAGAAGGAAGATGTCGAGGAGCTCATGGAACACCGTTCTTCTAAGCATGGAAAGGTTG
 AGGTCCCCAAGGAAGTTCAAAGGAGGACACCGCCTAA

SEQ ID NO: 20 Secuencia traducida de la variante del polipéptido Cylfu_AMT1 de *Cylindrotheca fusiformis*

MQAGFAMLCAGSVRQKNVKNIMLKNMLDACGGAIGFWTIGYAFAYADNSSGDKTFIGGKNFFVNQL
 DESGGAWIGFFFQFAFAATAATIVAGTVAERCKMSAYLCYSVFLTGFVYPVVVHSIWSADGWLTA
 RDDPWKGVGVIDFAGSGVVHMTGGATALVAAIVLGPGRKGRFYDEDEGNALET PASFPAHSVALQVLG
 TFLLWFGWYGFNPGSALVIDNAASASTSALCAVTTTLAAASGCVCMFTDTI IEMMATGEASYDLT
 MAMNGALGGLVAITAGCSVVT PWASIIIGIIAGWVYIAFSKLLVKLKIDDAVDVAVPVHFANGMWGV
 LAVGFFAEPDAMVTAGYNDVPGVFYKGDGKLLLCQFVAIIWICAWIFFLMT PFFVVLNII LGMFRVD
 PLEEEVGLDISHHRGAAYDMTSAKKEDVEELMEHRSSKHGKVEVPKEVQKEDTA

FIGURA 5 (continuación)

SEQ ID NO: 21 Secuencia de ácido nucleico Cylfu_AMT2a de *Cylindrotheca fusiformis*

ATGGCTGAGTTCGATAACACTTTTATCTTGGAATTTTGCAGTGGTGGAAATGAATCGTCGTCTGAC
 GTTCAGGTCTTGTGCCAAGTTGCCAGTCTCGCAAACGGCACCAGTGCCTCTGCCGGTGGACTTACT
 GAGGGTATCAATACCTTCTTTCTACTTTTTGCTGGAGCTCTAGTCTTTATCATGCAAGCTGGATTC
 GCCATGCTGTGTGCTGGATCCGTCCGTCAAAGAAGCTCAAGAATATCATGTTGAAAAACATGTTG
 GACGCTTGTGGTGGTGGTCTATTGGTTTCTGGACTATCGGGTATGCCTTCGCCTACGCTGATAACTCG
 TCTGGAAATAAAACATTTCATTGGAGGCAAGAACTTCTTTGTGAACCAACTGGATGAGTCTGGAGGT
 GCATGGATTGGTTTCTTCTTCCAATTTGCCTTTGCTGCCACTGCCGCCACTATTGTTCGCGGGAACC
 GTTGCCGAGCGCTGCAAGATGAGTGCCTACCTATGCTACTCAATCTTCCTCACTGGTTTTGTCTAT
 CCTGTCTGTTGTTCACTCCATCTGGAGTGTGATGGATGGTTGACTGCCTTTCTGTGATGATCCTTGG
 CAGGGTGTCCGGTGTTCATCGATTTCCGCCGATCTGGTGTGTGTCATATGTGCGGTGGAGCCACTGCT
 CTTGTTGCTGCTATTGTTCTTGGACCCCGTAAGGGACGTTTCTATGATGAGGATGGAAATGCTCTT
 GAAACCCCTGCAAGTTTCCCAGCTCACAGTGTGGCCCTTCAAGTCCCGGTACCTTCTTCTTTGG
 TTCGGATGGTATGGATTCAACCCAGGATCTGCTCTTGTGATTGACAATGCTGCGTCCGCTTCTACC
 TCAGCACTTTGTGCTGTCACTACCACTCTTGCCGCTGCCAGTGGTTGTGTACCGCGATGTTTACC
 GATACCCCTCATTGAAATGATGGCTACCGGAGAAGCATCTTATGATTTGACCATGGCCATGAACGGT
 GCACCTTGCAGGGGCTTGTGTCATCACTGCCGGATGCTCCGTTGTCACTCCTTGGGCTTCTCTCATC
 ATTTGGTATTATTGGCGGGTGGGTCTACCTTGGACTTTCTAAACTTCTAATCAAATTGAAGATCGAT
 GATGCTGTTGATGCCGTCCCTGTTCACTTCGGTAATGGTATGTGGGGTGTCTTGGCTGTTGGCTTC
 TTTGCCGAACCCGACGCCATGGTTACTGCCGGGTACAACGATGTCCAGGAGTGTCTACAAAGGA
 GATGGTTCTCTTCTCCTGTGCCAATTCGTCGCTATCGTTTGGGTTTGCCTTGGGTCTTCTTCTTG
 ATGACTCCTTTCTTCGTCGTCTTGAACATTTTGGGAATGTTCCGTGTGATCCTCTTGAGGAAGAA
 GTTGGTCTTGATATTTCCCACCACCGTGGATCTGCCTACGACATGACTACTGCCAAGAAGGAAGAT
 GTCGAAGAGCTCATGGAACACCGTTCTTCGAAGCACGGGAAGGTTGAGATCCCCAAGGAAGTTCAG
 AAGGAGGACACCGCCTAA

SEQ ID NO: 22 Secuencia traducida del polipéptido Cylfu_AMT2a de *Cylindrotheca fusiformis*

MAEFDNTFILEFCSGGNESSDVQVLCQVASLANGTSASAGGLTEGINTFLLFAGALVFIMQAGF
 AMLCAGSVRQKNVKNIMLKNMLDACGGAIGFWTIGYAFAYADNSSGNKTFIGGKNFFVNQLDESGG
 AWIGFFFQFAFAATAATIVAGTVAERCKMSAYLCYSIFLTGFVYPVVVHSIWSADGWLTAFRDDPW
 QGVGVIDFAGSGVVMCGGATALVAAIVLGPGRKGRFYDEEDGNALETPASFPAHSVALQVLGTFLLW
 FGWYGFNPGSALVIDNAASASTSALCAVTTTLAAASGCVTAMFTDTLIEMMATGEASYDLTMAMNG
 ALAGLVAITAGCSVVPWASLIIGIIGGWVYLGLSKLLIKLKIDDAVDVAVPVHFGNGMWGVLA VGF
 FAEPDAMVTAGYNDVPGVFYKGDGSLLLCQFVAIVWCAWVFFLMT PFFVVLNILGMFRVDPLEEE
 VGLDISHHRGSAYDMTTAKKEDVEELMEHRSSKHGKVEIPKEVQKEDTA

SEQ ID NO: 23 Secuencia de ácido nucleico Thaps_AMT1 de *Thalassiosira pseudonana*

ATGGCCGAACCAACAACAACCATAGGCGACTTCAACGTCCTGCTGGTGCAGGCGACGACGCTGTA
 GCAACCTACGAAGGACAATCCGTGCGAAGCGGTATCTGTGCAGCCTACGCCTACACCGACGAAACC
 AACACCGGTCTTGATGTGTTCTACCTCCTCTTCGCCGCTGCCATGGTCTTCTTCATGCAGGCTGGA
 TTCGCCATGCTCTGTGCTGGATCTGTGAGGCAGAAGAATGTGAAGAATATCATGCTTAAGAACATT
 TTGGATGCTTGTGGTGGAGCTCTTGGATTTTGGTCTGTGGGATTTGCGTTTGCCTACGGAGGATCT
 GGACCGGAGAAGAAGGATTCATCGGTAACGAGGGATTCTTCCTTGGTGACTTTACAACCTGGAGGC
 GATTTGATCGGATGGTTCTTCCAGTTCGCCTTTGCTGCCACCGCCGCCACAATTGTGGCTGGAACC
 GTAGCCGAGCGTTGCAAGTTCGAGGCCTACCTCTGCTATTCTCTCATGCTCACCGGGTTCGTCTAC

FIGURA 5 (continuación)

CCCGTGATTGTGTACTCCATCTGGTCTTCCTCTGGGTTCCCTCACC GCCTTCAACGACGACCCTGCC
TTTGGATGTGGTATGCATGATTTCCGCCGATCGGGAGTTGTTACATGACGGGAGGCATCACTGCC
CTTTGGGCCGCCAAGATTCTCGGACCTCGTATCGGACGCTTCTACGATGCTGATGGCAATGAGCTT
CCCGAGCCAGTGAGCTTCCCTCCCCACTCCGTGGCCCTTCAAGTTCTTGGTACTTTCATCCTTTGG
GTGGGCTGGTACGGATTCAACCCCGGTTCCACTCTTCTCATCAGTAACACGGCCGCAGCTGATGTG
TCTGCCCTTTGTGCCGTCAACCACCACCTTGCCGCCGCTTCGGGCTCTGTCTCTGCCATGTTCACT
GATATGTTCTTGGAGCGCAGAAAGACCGGAGAGACTATGTACGACATTACCATGTGTATGAACGGT
GCTCTTTCTGGATTGGTGGGCATCACCGCTGGATGCTCAATCGTTGAGCCTTGGGCTGCCTTTGTG
ATTGGAATTGTGCCGGATGGACTTACATCTTCTGGTCCAGTCTCCTTGTGAAGCTTAAGATTGAT
GATGCCGTCGATGCCATTCTGTTCACCTTTGGAAACGGAATGTGGGGCTGCATTGCCGTTGGACTC
TTTGCCGAGCCTACCCGTGTAGCCAACGCCTACAGTGACCATGGACACTATGGATGGTTCTACTCA
TGGGGTGCTGGAAACGCCGATGCCCACCTTTTGGCTGCTCAAGTCTGTGGTGTCTCTGGATCATT
GGATGGGTTTCCGTCATTATGATCCCATACTTCATCCTGCTCAACGTCCTGGGTTTGTTCGCTGTG
GATGCCCTCGAAGAAGAAGTTGGTTTGGATATCTCCCACCACAAGGGAGCTGCCTACGATATGTCC
GGACCTTCTGAAGCTGCCGCCGAGAAGTTTGAGATCTCTAGGAGTCAGCGTAAGCTTGAGATCCCT
GTGGATGTCGCGCTGCTACTGCTCCTGCCGAGGATGCCGCGTAA

SEQ ID NO: 24 Secuencia traducida del polipéptido Thaps_AMT1 de *Thalassiosira pseudonana*

MAEPTTITGDFNVTAWCGDDAVATYEGQSVENGICAAAYTDETN TGLDVFYLLFAAAMVFFMQAG
FAMLCAGSVRQKNVKNIMLKNILDACGGALGFWSVGFAYGSGPEKKGFIGNEGFFLGDFTTGG
DLIGWFFQFAFAATAATIVAGTVAERCKFEAYLCYSLMLTG FVYPVIVYSIWSSSGFLTAFNDPA
FGCGMHDFAGSGVHMTGGITALWAAKILGPRIGRFYDADGNELPEPVSFPPHSVALQVLGTFILW
VGWYGFNPGSTLLISNTAAADV SALCAVTTTIAAASGSVSAMFTDMFLERRKTGETMYDITMCMNG
ALSGLVGITAGCSIVEPWA AFVIGIVAGWTYIFWSSLLVKLKIDDAVD AIPVHFGNGMWGCIAVGL
FAEPTRVANAYS DHGHYGFYSWGAGNADAHLLAAQVCGVLWIIGWVSVIMIPYFILLNLVGLFRV
DALEEEVGLDISHHKGAAYDMSGPSEAAAEKFEISRSQRKLEIPVDVAPATAPAEDAA

SEQ ID NO: 25 Secuencia de ácido nucleico Thaps_AMT2 de *Thalassiosira pseudonana*

ATGTCTGAAGATCCATCAATCTTTGAAGTGTGTACTGGCCAGCTCGGAACAGATCTTACCGTCGAG
CTTTTACAATGTGTCTCTGATGGAGCTGAAAGCGCAAAGGATGATGTCATCAGTGGCGTCAACTCC
TTCTACCTCATCTTTGCAGGAGCCCTCGTTTTCTTCATGCAAGTCGGTTTCGCCATGCTCTGTGCT
GGATCCATCCGTGAGAAGAACGTCAAGAATGTATTGCTTTGGAATCTCCTCGATTCTGCCGGTGGT
GCCTTTGGTTTCTGGAGTATTGGTTATGCATTTGCTTATGGTGGTGTATGATATTACCAAGGGAAAG
ACCTTCATCGAAACGCTGACTTCTTCTTGAGTGGAGAACTGATATGGAGTTTGGTTCTTCCAA
TACGCCTTTGCGTGTGCTCTCTCCTCCATTGTTGCTGGAACCATTGCTGAGCGCACCAAGATGATG
GCCTACTTGTGCTACTCAATCTTCCTTTGTGGATTTCGTCTACCCAGTCTGTGCTCATGCCTTTTGG
TCTCAGAATGGATTCTCTCTGCCTTCGCTGCTGAACCTTTGTGGGGTTCGGGCGTTATTGACTTT
GCAGGATCGGGACCAGTTCACATGTGTGGAGGAGTTGCTGCTCTTGTGATGGCTATTATTCTTGA
CCTCGTAGGGGACGTTTCTATGACGACGATGGTGTGCTATTGGATGAGCCAAAGTCCATGGGACCT
CACTCCGTCACCTTTGCAATTCCTCGGAACCTTTGCTCTTTGGTTCGGATGGTATGGATTCAACCCT
GGTAGCTCCATCTTGATTGCGTCAGCTGCCTCTGGCGACGTTGCATCGCTTGCTGCCGTCAACACT
ACGCTCGGATCTGCTGCTGGAGCACTTTCTGGTATGTTACCTCCACGATCGTCGACGAAAGAAAG
ACTGGAGTGTATACTTGGGATACTACTGCTGCAATGAACGGATGCCTCACCGGTTTGGTTGCTATT
ACTGCTGGTTGCGCTACCGTCGAGCCTTGGGCTGCCTTCGTCATTGGGCTCACTGCTGGTTGGGTG
TACCTTGCTGCATCTGCTCTTATGCTTCGCTTCAAGATTGATGATGCTGTGATGCCATCCCCGTT
CACATGTTCCGAGGATCATGGGGAGTTTTTTGCACTGGTCTTTTACCAGTCTCGCCGCTTATT

FIGURA 5 (continuación)

ACTGCATACGGAAATGACAATAATGTTGGTTGGTTCTACGAATGGGGACGTGGAAGTGGAACCTTC
 ACTCTCCTTGGCTGCCAGCTCGTCTCGATTCTCTTTGTCTTGGGATGGTCTGCTTGCATCTTTGCT
 CCATTCTGCTTGGCACTCAAAACCCTCAACTGGCTCCGCATTGACCCTCTCGAGGAGGAGGTTGGT
 ATGGATATCAGTCGCCATAAGGGACCTGCCTACGAGTCGGAGGGATCTGCTCATTCTGATGCTATC
 GAGAAGTTGAGTGCCTCCCGTCGTGATATTATGAATGCCTCTGGAAGTGGAAGGGGAAAAAGTTTC
 AGCAGGTCGACTCCCACGAAAGCTAACGAGGAGCCCCAAAATCGAAGCAACTGAGGATGCAGGGGCA
 CCCGCTGGAGAAGCAACTGCTTAG

SEQ ID NO: 26 Secuencia traducida del polipéptido Thaps_AMT2 de *Thalassiosira pseudonana*

MSEDPSIFEVCTGQLGTDLTVELLQCVSDGAESAKDDVISGVNSFYLI FAGALVFFMQVGFAMLC
 GSIREKNVKNVLLWNLLDSAGGAFGFWISIGYAFAYGGDDITKGKTFIGNADFFLSGETDMEFWFFQ
 YAFACALSSIVAGTIAERTKMMAYLCYSIFLCGFVYPVCAHAFWSQNGFLSAFAAEPLWGSVIDF
 AGSGPVHMC GGVAALVMAIILGPRRGRFYDDDGVLDEPKSMGPHSVTLQFLGTFALWFGWYGFNP
 GSSILIASAASGDVASLAAVNNTLGSAAALSGMFTSTIVDERKTGVYTWDTTAAAMNGCLTGLVAI
 TAGCATVEPWAAFVIGLTAGWVYLAASALMLRFKIDDAVD AIPVHMF GGSWGVFCTGLFTSPRRLI
 TAYGNDNNVGFYEWGRGSGNFTLLGCQLVSI L FVLGWSACIFAPFCLALKTLNLRLIDPLEEEVG
 MDISRHKGPAYESEGSAHSDAIEKLSASRRDIMNASGSGRGKSFSRSTPTKANEKPKIEATEDAGA
 PAGEATA

SEQ ID NO: 27 Secuencia de ácido nucleico Thaps_AMT3 de *Thalassiosira pseudonana*

ATGTTTTCAGGTATCAAGAGCTGGACACGTGTCAGTGTATGAGGTCTGCAAATCCTTCGTCAACCCA
 GAAGATTCTCAAGCAGATCAGTTTGATGCAATGCTCCAATGTGTTGGGGAATCCAACGGCAAGAGT
 ATAGACGCTTTCTTCCTTATATACGCATCATCCCTCGTCTTCTTCATGCAAGCCGGCTTCGCCATG
 CTTTGTGCTGGATGCGTTCAACACAAGAATGTT CAGAACAGTATGTTGAAGAACCTCCTAGATGCA
 TGCGGTGCAGCCCTTGGCTTCTATTCCGTGCGGTATGCATTTGCTTACGGTGGCATGGACTATTCA
 GATCCAAACAAGACATTTATCGGCACAGAGAACTTTTCTTGATGGGAGTGGACGATTTTCATGTTT
 TGGCTATTCCAATTTGCATTCGCTGCAAGTGCTGCTACCATTGTGGCAGGAACGTTGGCTGAACGG
 TGTCAAATGACGGCATACTTGTGCTACTCAGTGGCAGTGAAGTGGATTGTATATCCAGTAGTTGTA
 CATTCGGTTTGGTCTCCGCAGGGTTTCTTGTGCGGACAGGCTGTGAGCCCGTTATTTGGAGTTGGT
 GTAGTAGACTTTGCAGGATCATCGGTTGTGCATTTGACTGGAGGGTGCATTGCACTCATTGCCACG
 TATATTCTAGGCCCAAGGCGAGGGAGATTCTATGATCACAGAGGAGAACCTCTTGAGACACCAGTC
 GAGTTTCCGGGTCATT CAGCTGCACTTCAAATGCTCGGTGCCTTCATATTATGGTTTGGATGGTAT
 GGCTTCAATACTGGATCAACTCTTTCGATCACCGGCCCTGGCCAACATCAAGTCGTCAGCCTTGTA
 GCTGTAAACACAACCCTCGCGGCGGCCTCTGCTTGTGTGGCTTCCCTTCTTGCCAGTTATTATGTC
 ATCGAACGAAAGACTGGCGAAGGTACATTCTCGCTTTCTTCAGCAATGAATGGGTGCTTAGGAGGA
 TTAGTCAGTATCACTGGTGGATGTGCAGTGGTGGAGCCTTGGGCAGCCGTCGTAATCGGATTCATA
 GCAGGGTTGTTGTATCTCTTCACGTCAAAGCTATTGATTTCGTTTGCGAATAGACGATGCAGTCGAT
 GCTATTCCGGTCCACTTGTCTAACGGAATATGGGGTACGGTTGCGGTGGGGTTGTTTCGCATCGTCG
 AATCGTTTGCAGTTGGCTTTTGAAAAGTCGCTGATACTGGTGTGTTTCATGGGTGGAACCGGTAAA
 CTGTTGGGATGTCAAATAATTGGTGTCTTCTTTGTGCTTGGATGGATTTCTTCATTATGATTCCG
 TTCTTCTGCTTCTTCACTACATGGGATGGCTTCGATCTGAGTCAATTGATGAAGTAGAAGGGCTT
 GATTCCAAGTATCATGGATTGCGAAACAAAGATGAGCATAGACATGACGAAGAAGAAGACAATACT
 CCATCACACTACGGCGAAGGCAATTGCAGGCTGAGGCGTAGCATTTTGCGCCATGAGGAAAGAATG
 AGACAAGAGGACTCGGCAATACCCGCAACTTTGGTATCAAGTGATGATAGGTTACGTGTGACTCA
 GGAGGAAATAGCACTTCAATGCTATCTACTGCTAAGCAATACACCTAG

FIGURA 5 (continuación)

SEQ ID NO: 28 Secuencia traducida del polipéptido Thaps_AMT3 de *Thalassiosira pseudonana*

MFQVSRAGHVSVEVCKSFVNPEDSQADQFDAMLQCVGESNGKSIDAFFLIYASSLVFFMQAGFAM
LCAGCVQHKNVQNSMLKNLLDACGAALGFYSVGYAFAYGGMDYSDPNKTFIGTENFFLMGVDDFMF
WLFQFAFAASAATIVAGTLAERCQMTAYLCYSVAVTGFVYPVVVHVSWSVWSPQGFLCGQAVSPLFGVG
VVDFAAGSSVHLTGGCIALIATYILGPRRGRFYDHRGEPLETPVEFPGHSAALQMLGAFILWFGWY
GFNTGSTLSITGPGQHQVVSLLVAVNTTLAAASACVASLLASYVIERKTGEGTFSLSAMNGCLGG
LVSITGGCAVVEPWAAVVIGFIAGLLYLFTSKLLIRLRIDDAVDAPVHLSNGIWTGTVAVGLFASS
NRLQLAFGKVADTGVFMGGTGKLLGCQIIGVFFVLGWISFIMIPFFCFLHYMGWLRSESIDEVEGL
DSKYHGLRNKDEHRHDEEEDNTPSHYGEENCRLLRSILRHEERMQRQEDSAIPATLVSSDDRFTCD
SGNSTSMLSTAKQYT

SEQ ID NO: 29 Secuencia de ácido nucleico Thaps_AMT4 de *Thalassiosira pseudonana*

ATGTCCTCGTCGGTACGACGAGCCTGTACGAGGCTTGCAAATCTACTCAATCCAACCTCTACGTTT
TCCAACCTCTACGCTCGATGATGCACTCTCTCGCCAGGAACAGATCTTCCGTTGCATTTTCGGAATCG
AACGCCAATAGTATTGATACGTTCTTTCTGCTCTATGCCTCATCGCTGGTATTCTTTATGCAAGCC
GGATTCGCGATGCTCTGTGCAGGATCAGTGAGAAAGAAGAATGTGACTAATACCATGTTGAAGAAC
CTTCTCGATGCTTGCAGGTGCCGCACTTGGGTTCTATTCTGTGGGATACGCCTTTGCATACGGAGGA
TCAGTAGACGCCGGGAAGAAGACGTTTATAGGTATGAGCAACTTCTTTCTACAGGACGTTGACAAC
TATATGTTTTGGCTCTTTCAGTTTGCATTCTGCTGCAACATCAGCTACAATCGTAGCAGGAACATTG
GCTGAAAGATGTCAAATGACAGCTTACCTTTGCTACTCAATTGCATTGACGGGATTTGTCTACCCA
GTTGTCTGCTCATTCAATATGGAGTCAGCAGGGGTTCTTATCAGCTACTGCTCAAGATCCATTATGG
GGTACAGGTTTTCATTGACTTTGCGGGATCAACAGTAGTACATCTGACGGGTGGATTTACAGCTTTG
ATTGCGACATATCTTCTAGGGCCACGCAGAGGACGGTTCTACGATGCGAAAGGCAAGCAGTTGGAA
GTGCCAAATCCAATGCCTGGCCATTGAGCTGCACTTCAGATGCTTGGTATCTTCATTTTATGGTTT
GGTTGGTATGGTTTTCATTGTTGGATCAGCGATAACTATCATCGGTCCGAATCAAGACAAGATCATC
TCTACCTCTGCAGTGAATACGACACTCTCTGCAGCGTCATCATGCTTCTCTGCCTTACTCGTCAAC
TACGTCATTGTGCGAGAGGCAATCGGGAGAAGGAGAATTGAGTCTTCTCGCTGCAATGAACGGATGC
TTGAGTGGATTGGTAGCGATAACAGGTGGATGCGCAGTGATAGCACCATGGGCAGCAATTATCGTC
GGACTTTTTCGCTGGCCTCTTGTACTTATTTACGTCAAAGGTATTGGTACGAGTTCGAATCGACGAT
GCAGTAGAAGCCATTCTTGTTCACATGACTAACGGGATATGGGGTAGTTTCGCAGTTGGACTATTT
GCCGCTCCATCGGAGCTGCAATTGGTATACGGAAAAGCAAATCACGTTGGACTATTCTATTCTTGG
CATCAAGGGAGTGGAGATGGGACGTTACTCGGTGTCCAATGCTTGGGCATTCTGTTTGTAGTGGGC
TGGGTGTTCTGTCTCATGTCTCCCTTCTTCTGTTTCTGAACTACAAAGGCTGGTTTCAGAGCCGAC
GTCCTCAACGAGATTGCTGGCTTAGATTTGAGTTATCATGATGGAGTGGATATGGAATTGGTGACT
CAGATACGCAATCAAAGGAAGAACTTGCACGTCAACAGCAGGAATCGCTTTAGTTTGAATAGTGCA
TGTCTCATCATACGGCAACACATGTCGATACCTCAACTGATGCATCACTTTGA

SEQ ID NO: 30 Secuencia traducida del polipéptido Thaps_AMT4 de *Thalassiosira pseudonana*

MSSSVRTSLYEACKSTQSNSTFSNSTLDDALSRQEQIFRCISESNANSIDTFFLLYASSLVFFMQA
GFAMLCAGSVRKKNVTNTMLKNLLDACGAALGFYSVGYAFAYGGSVDAGKKTFIGMSNFFLQDVDN
YMFWLFQFAFAATSATIVAGTLAERCQMTAYLCYSIALTGTVYPVVAHSIWSQQGFLSATAQDPLW
GTGFIDFAGSTVVHLTGGFTALIALYLLGPRRGRFYDAKGKQLEVPNPMPGHSAALQMLGIFILWF
GWYGFIVGSAITIIIPNQDKIISTSAVNTTLAASSCFSAALLVNYVIVERQSGEGEFSLLAAMNGC
LSGLVAITGGCAVIAPWAAIIVGLFAGLLYLFTSKVLVRVRIDDAVEAIPVHMTNGIWSFAVGLF
AAPSELQLVYGKANHVGFLFYSWHQSGDGTLLGVQCLGILFVVGWVFCFLMSPFFLFLNYKGWFRAD
VLNEIAGLDLSYHDGVDMELVTOIRNQRKNLHVNSRNRFSNSACPHHT

FIGURA 5 (continuación)

SEQ ID NO: 31 Secuencia de ácido nucleico Thaps_AMT6 de Thalassiosira pseudonana

ATGGCATCATCAACCACCACAGACACCTATCAAACATGCCTCAGCGATCTCTCGGCGACGTCGTCT
AATGGATCCTCGCCCACCACCGACGCCCTCCTTCAGTGCATCTCCTCCTCCTTCGACGCTCAAACG
GCCTCTACACATGCTTCAATCAACACCTTTTTCTCCTCTATGCAGCCACCCTCGTCTTCTTCATG
CAAGCCGGCTTTGCCATGGTTAGCGCTGGATGTGTTAGGACGAATAATGTTTCAAGATACGCTGCTG
AAGAATCTGTTGGACGCCTGCGGTGCCGCTCTTGGATTCTACACTGTTGGATACGCCTTCGCATGG
GGAGGATCGTTGGATACGGCTACCACCGAGAGGACGTTTCATCGGTACACAGAACTTCTTTTTGATG
GATGTGGATAGCTCTCAGGATTCATTTTGGTTGTTTCAATTGGCTTTCTGCTCGGCATCTGCAACG
ATTGTGGCTGGAACGTTGGCTGAACGTTGTCAAATGGTTGCCTATCTTGCATACTCCATGACATTG
GCAGGATTTGTCTACCCAGTCGTCGTACACAGTATCTGGAGTCCGAGTGGATTCTTGAGTGTCTACT
CGTGAGACCGATCTCTTCTTGGATGTGGGAATGATTGACTTCGCAGGATCAACCGTCGTGCATTTG
ACGGGAGGGATGACTGCGTTGATTGCTACGATTGTGCTGGGACCGAGGACGGGAAGGTTCTATGAT
TTGAGAGGAAATCCGTTGAAAGTACCAAAGGAGTTTGCAGGACATTCATTGGCTTTGCAAATGTTG
GGGGTGTTCATCTTGTGGTTTGGATGGTATGGCTTCAACGCTGGATCGATCCTCAACATCACCAAC
GATCTCAATCATAAATCGTCAGTCATACTGCCATCAACACAACCTCTTGCAGCTTCTGCTGGATCC
ATCATGACTCTCTCCTCAGTACCGTCGTAGCCGAAAGATTTACGGGGGAGATAGTGTTTAGTCTA
TCTTATGCCATGAATGGGTGTTTGGAGTGGATTGGTGGCAATCACAGCTGGCTGCTCAGTGGTAGAA
CACTGGGCTGCAATCATAATCGGGCTTGTAGGAGGGGCATTGTATTTGGCATGCTCCAAGTTCTTG
GTGAAGAAACGTATCGATGATGCAGTTGATGGTATCCCTGTCCACTTGATTAATGGAATCTGGGGA
ACGTTGAGTGTGCGTCTCTTCGCTGTGCCTGAGTTGTTGGAGCAAGTGATGGAAGGGGTGATCAC
GCTGGTTGGTTTTTACAGTTGGGGACAAGGATCAGCAGACGCCAAGTTGTTGGGAGCTCAAGTAGTT
GGAATCTTGTGTTGTCAGTGGCTGGGTATGATCACGATGTTTCCTTTCTTTTGTCTTCTGCATTAT
GTTGGATGGCTTCGTGCCGACTCTCTCGAGGAAGTAGTCGGTCTCGACGCTGCTTACTCCCAAGGT
GTTCTTCAGACGCGTGCCCGTGCCCAAGTGAGGAAGAGAACATGGAGCATTACATCAGTGAATAT
GTTAAGCAGCGTGAGGAAAAAGCATTCATCAAGAAGATCAACAGCAATAGTACTCACGGACGTACT
ATCCTTGGTGCAAGCATGCATTCAATGAACATCATTAATTCCAGCATGCATTCCAGAAAGGACAGT
CTACCACGTGCAATTGAAAGTCTCAACAATTCAAGACACTCTGGTTCAAGAGGATCAAGATCAATT
AATGATATTGCGATTGACAATTTGCATGGTCAATCTGAGGATGGTTTTGCTGCACCTGATGAGGGC
AGTGCTTAG

SEQ ID NO: 32 Secuencia traducida del polipéptido Thaps_AMT6 de Thalassiosira pseudonana

MASSTTTDTYQTCLSDLSATSSNGSSPTDALLQCISSSFDAQTASTHASINTFFLLYAATLVFFM
QAGFAMVSAGCVRTNNVQNTLLKNLLDACGAALGFYTVGYAFWGGSLDTATTERTFIGTQNFFLM
DVDSSQDSFWLFLQAFCSASATIVAGTLAERCQMVAYLAYSMTLAGFVYPVVVHSIWSPSGFLSAT
RETDLFLDVGMIDFAGSTVVHLTGGMTALIAITIVLGPRTGRFYDLRGNPLKVPKEFAGHSLALQML
GVFILWFGWYGFNAGSILNITNDLNHTIVSHTAINTTLAASAGSIMTLFLSTVVAERFTGEIVFSL
SYAMNGCLSLVAITAGCSVVEHWAIIIGLVGGALYLACSKFLVKKRIDDAVDGIPVHLINGIWG
TLVGLFAVPELLEQVYGRGDHAGWFYSWGQGSADAKLLGAQVVGILFVSGWVMITMFPFFCFLHY
VGWLRADSLEEVVGLDAAYSQGVLTQTRARAQSEENMEHYISEYVKQREEKAFIKKINSNSTHGRT
ILGASMHSMNIIINSSMHSRKDSLPRAIESLNNSRHSGSRGSRINDIAIDNLHGQSEDGFAAPDEG
SA

SEQ ID NO: 33 Dominio Conservado comprendido dentro de la SEQ ID NO: 2 y dentro de la SEQ ID NO: 4

MQAGFAMLCAGSVRQKNVKNIMLKNLLDACGGAIGFYTVGFGFAYGGDDTTDKTFIGNSYFALRDY
TNYAGFFFQFAFAATAATIVAGTVAERCKMSAYLCYSLFLTGFVYPVVVRSVWSSNGFLSAFSDP
FQGVGTVDFAAGSVVHMTGGLTALIAAIVLGP RKGRFYDEDGNPLETPASFPAHSVALQILGTFIL

FIGURA 5 (continuación)

WFGWYGFNPGSALKIANADSAATAALCAVTTTMAAAAGCVSAMFTDSIIDGMATGETTYDLTMAMN
GCLAGLVAVTAGTSVVTPWAAIIIGVVGGWVYIGMSKLLIKLKIDDAVDAPVHFANGFWGVLATG
LFANGGLMATAGYNSEHEGWFEWGS GSGDGSLLICQLACLAWIIGWVTTIMTPFFILLNMAGMFR
VDPLEEEVGLD

SEQ ID NO: 34 Promotor GOS2 de *Oryza sativa*

AATCCGAAAAGTTTCTGCACCGTTTTACCCCCTAACTAACAATATAGGGAACGTGTGCTAAATAT
AAAATGAGACCTTATATATGTAGCGCTGATAACTAGAACTATGCAAGAAAACTCATCCACCTACT
TTAGTGGCAATCGGGCTAAATAAAAAAGAGTCGCTACACTAGTTTCGTTTTCTTAGTAATTAAGT
GGGAAAATGAAATCATTATTGCTTAGAATATACGTTACATCTCTGTCATGAAGTTAAATTATTCTG
AGGTAGCCATAAATTGTCATCAAACCTCTTCTTGAATAAAAAAATCTTTCTAGCTGAACCTCAATGGGT
AAAGAGAGAGATTTTTTTTTAAAAAATAGAATGAAGATATTCTGAACGTATTGGCAAAGATTTAAA
CATATAATTATATAATTTTATAGTTTGTGCATTTCGTCATATCGCACATCATTAAAGGACATGTCTTA
CTCCATCCCAATTTTTATTTAGTAATTAAAGACAATTGACTTATTTTTATTATTTATCTTTTTTTCG
ATTAGATGCAAGGTACTTACGCACACACTTTGTGCTCATGTGCATGTGTGAGTGCACCTCCTCAAT
ACACGTTCAACTAGCAACACATCTCTAATATCACTCGCCTATTTAATACATTTAGGTAGCAATATC
TGAATTCAAGCACTCCACCATCACCAGACCACTTTTAATAATATCTAAAATACAAAAAATAATTTT
ACAGAATAGCATGAAAAGTATGAAACGAACCTATTTAGGTTTTTCACATACAAAAAAGAAATT
TTGCTCGTGC GCGAGCGCCAATCTCCCATATTGGGCACACAGGCAACAACAGAGTGGCTGCCACACA
GAACAACCCACAAAAACGATGATCTAACGGAGGACAGCAAGTCCGCAACAACCTTTTAACAGCAG
GCTTTGCGGCCAGGAGAGAGGAGGAGGCAAGAAAACCAAGCATCCTCCTTCTCCCATCTATAA
ATTCCTCCCCCTTTTCCCCTCTCTATATAGGAGGCATCCAAGCCAAGAAGAGGGAGAGCACCAAG
GACACGCGACTAGCAGAAGCCGAGCGACCGCCTTCTCGATCCATATCTTCCGGTCGAGTTCTTGGT
CGATCTCTTCCCTCCTCCACCTCCTCCTCACAGGGTATGTGCCCTCCCTTCGGTTGTTCTTGGATTT
ATTGTTCTAGGTTGTGTAGTACGGGCGTTGATGTTAGGAAAGGGGATCTGTATCTGTGATGATTCC
TGTTCTTGGATTTGGGATAGAGGGGTTCTTGATGTTGCATGTTATCGGTTTCGGTTTGATTAGTAGT
ATGGTTTTCAATCGTCTGGAGAGCTCTATGGAAATGAAATGGTTTAGGGATCGGAATCTTGCGATT
TTGTGAGTACCTTTTGTGTGAGGTAAATCAGAGCACCGGTGATTTTGCTTGGTGTAATAAAGTAC
GGTTGTTGGTCCCTCGATTCTGGTAGTGATGCTTCTCGATTTGACGAAGCTATCCTTTGTTTATTC
CCTATTGAACAAAAATAATCCAACCTTTGAAGACGGTCCCGTTGATGAGATTGAATGATTGATTCTT
AAGCCTGTCCAAAATTTTCGCAGCTGGCTTGTTTAGATACAGTAGTCCCCATCACGAAATTCATGGA
AACAGTTATAATCCTCAGGAACAGGGGATTCCCTGTTCTTCCGATTTGCTTTAGTCCCAGAATTTT
TTTTCCCAAATATCTTAAAAAGTCACTTTCTGGTTTCAGTTCAATGAATTGATTGCTACAAATAATG
CTTTTATAGCGTTATCCTAGCTGTAGTTTCAGTTAATAGGTAATACCCCTATAGTTTAGTCAGGAGA
AGAACTTATCCGATTTCTGATCTCCATTTTAAATTATATGAAATGAACTGTAGCATAAGCAGTATT
CATTTGGATTATTTTTTTTATTAGCTCTCACCCCTTCATTATTCTGAGCTGAAAGTCTGGCATGAA
CTGTCCTCAATTTTGTGTTTCAAATTCACATCGATTATCTATGCATTATCCTCTTGATCTACCTGT
AGAAGTTTCTTTTGGTTATTCCTTGACTGCTTGATTACAGAAAGAAATTTATGAAGCTGTAATCG
GGATAGTTATACTGCTTGTCTTATGATTCATTTCTTTGTGTCAGTTCTTGGTGTAGCTTGCCACT
TTCACCAGCAAAGTTC

SEQ ID NO: 35 Prm09458

GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAACAATGATGCAGGCCGGG

SEQ ID NO: 36 Prm09459

GGGGACCACTTTGTACAAGAAAGCTGGGTACACGAGCAGCAATTAAACC

FIGURA 5 (continuación)