

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 372 938**

51 Int. Cl.:
C12N 15/82 (2006.01)
A01H 5/00 (2006.01)
C07K 14/415 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **04816347 .1**
96 Fecha de presentación: **17.12.2004**
97 Número de publicación de la solicitud: **1694849**
97 Fecha de publicación de la solicitud: **30.08.2006**

54 Título: **PLANTAS QUE TIENEN CARACTERÍSTICAS DE CRECIMIENTO MODIFICADAS Y MÉTODO PARA PREPARAR LAS MISMAS.**

30 Prioridad:
17.12.2003 EP 03104764
22.12.2003 US 531866 P

45 Fecha de publicación de la mención BOPI:
27.01.2012

45 Fecha de la publicación del folleto de la patente:
27.01.2012

73 Titular/es:
CROPDESIGN N.V.
TECHNOLOGIEPARK 3
9052 ZWIJNAARDE, BE

72 Inventor/es:
FRANKARD, Valerie;
MIRONOV, Vladimir y
SANZ MOLINERO, Ana Isabel

74 Agente: **Carvajal y Urquijo, Isabel**

ES 2 372 938 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 características de crecimiento modificadas y método para preparar las mismas

Campo de la invención

5 La presente invención se refiere a un método para mejorar las características de crecimiento de plantas. Más específicamente, la presente invención se refiere a un método para mejorar la producción de semillas introduciendo y sobreexpresando un ácido nucleico que codifica para una GRUBX en una planta o célula vegetal y cultivando la planta o célula vegetal en condiciones que promueven el crecimiento de la planta.

10 Dado el crecimiento constante de la población mundial, y la disminución del área de tierra disponible para la agricultura, sigue siendo un objetivo importante de la investigación agrícola mejorar la eficacia de la agricultura y aumentar la diversidad de plantas en la horticultura. Los medios convencionales para mejoras hortícolas y en cosechas utilizan técnicas de cría selectiva para identificar plantas que tienen características deseadas. Sin embargo, tales técnicas de cría selectiva tienen varios inconvenientes, concretamente que estas técnicas normalmente son laboriosas y dan como resultado plantas que a menudo contienen componentes genéticos heterogéneos que no siempre dan como resultado que se transmita el rasgo deseado de las plantas originales.

15 Además, las especies donadoras adecuadas para proporcionar un rasgo deseado pueden ser escasas. Los avances en biología molecular han permitido a la humanidad manipular el germoplasma de animales y plantas. La modificación por ingeniería genética de plantas conlleva el aislamiento y la manipulación del material genético (normalmente en forma de ADN o ARN) y la introducción posterior del material genético en una planta. Tal tecnología ha conducido al desarrollo de plantas que tienen diversos rasgos económicos, agronómicos u hortícolas mejorados. Rasgos de interés económico particular son características de crecimiento tales como alta producción. Producción se define normalmente como el producto medible de valor económico de un cultivo. Esto puede definirse en términos de cantidad y/o calidad. La producción de un cultivo se ve influida de manera adversa por el estrés típico al que se ven sometidos las plantas o los cultivos. Tales estreses incluyen estreses abióticos, tales como estreses de temperatura provocados por temperaturas altas o bajas atípicas; estreses provocados por deficiencia de nutrientes; estreses provocados por una falta de o exceso de agua (sequía, inundación), estreses provocados por productos químicos tales como fertilizantes o insecticidas. Los estreses típicos también incluyen estreses bióticos, que pueden imponerse a las plantas por otras plantas (malas hierbas, o los efectos de plantación de alta densidad), por plagas de animales (incluyendo estreses provocados por el pastoreo), y por patógenos. La producción de un cultivo no sólo puede aumentarse combatiendo uno o más de los estreses a los que se ve sometido el cultivo o la planta, sino que también puede aumentarse modificando los mecanismos de crecimiento inherentes de una planta.

20 Los mecanismos de crecimiento inherentes de una planta se controlan a varios niveles y mediante diversos procesos metabólicos. Un proceso de este tipo es el control de los niveles de proteína en una célula mediante la degradación de proteínas mediada por ubiquitina.

35 Ubiquitinación se refiere a una modificación de proteínas mediante conjugación a moléculas de ubiquitina. El término ubiquitinación a menudo se extiende a procesos que median la unión de proteínas de ubiquitina o de proteínas que imitan la función de la ubiquitina. La ubiquitinación es una herramienta versátil para que las células eucariotas controlen la estabilidad, función y la localización subcelular de las proteínas. Este mecanismo desempeña un papel central en degradación de proteínas, control del ciclo celular, respuestas al estrés, reparación del ADN, transducción de señales, regulación de la transcripción y tráfico vesicular. Puesto que la degradación de proteínas mediada por ubiquitina está en las bases de muchos procesos celulares, está altamente regulada y requiere alta especificidad de sustrato y amplia diversidad en efectores posteriores. Se conocen varias proteínas de unión a ubiquitina. Estas proteínas tienen a menudo una arquitectura de dominios modular. Por ejemplo, las proteínas de unión a ubiquitina combinan normalmente un dominio de unión a ubiquitina con un dominio de efector variable. Luego existen otras que no contienen un dominio de unión a ubiquitina, pero tiene una estructura terciaria similar a ubiquitina y por tanto pueden imitar ciertos aspectos de la ubiquitinación (dominios similares a ubiquitina).

40 El número de dominios y motivos relacionados con ubiquitina presentes en ubiquitina y proteínas similares a ubiquitina está creciendo a medida que está disponible más información sobre secuencias del genoma. Algunos prototipos de esos dominios son por ejemplo UBA, UBD, UIM y UBX (véase por ejemplo la base de datos Pfam; Bateman *et al.*, *Nucleic Acids Research* 30(1):276-280 (2002)). El dominio UBX es una secuencia de aproximadamente 80 residuos de aminoácido de longitud, es de función desconocida y está presente en proteínas de diversos organismos. La mayoría de estas proteínas pertenecen a una de cinco familias conservadas en la evolución mostradas a modo de ejemplo por las proteínas humanas FAF1, p47, Y33K, REP8 y UBXD1 (Buchberger *et al.* (2001) *J. Mol. Biol.* 307, 17-24; Carim-Todd *et al.* (2001) *Biochim. Biophys. Acta* 1517, 298-301). Normalmente, el dominio UBX se sitúa en el extremo C-terminal de una proteína.

55 Las pruebas estructurales sugieren una función del dominio UBX en procesos relacionados con ubiquitina; en particular el dominio UBX puede estar implicado en interacciones proteína-proteína. Se pronostica habitualmente que proteínas que comprenden dominios UBX están presentes principalmente en el citoplasma, aunque también se han notificado otras localizaciones subcelulares. Por ejemplo, la fosforilación, que es una modificación de proteínas

específica usada para regular la actividad de muchas proteínas, se ha mostrado que también influye en el transporte al núcleo de FAF-1 (Olsen *et al.* (2003) FEBS Lett. 546, 218-222.). En resumen, se ha propuesto que proteínas que contienen UBX animales podrían estar implicadas en el aumento de la expresión de genes relacionados con apoptosis, ciclo celular o direccionamiento de proteínas para degradación.

5 En *Arabidopsis*, planta cuyo genoma se ha secuenciado completamente, existen al menos 15 proteínas que contienen UBX. Pueden clasificarse según la similitud de secuencia en los grupos FAF1, p47, Y33K y UBXD1, sólo el grupo correspondiente a REP8 parece no estar presente en plantas (véase la figura 1). Como en el reino animal, los dominios UBX en proteínas vegetales están presentes en combinación con otros dominios, como por ejemplo SEP, G6PD, PUG o dedos de zinc. Se han descrito proteínas que contienen UBX y la estructura de dominios de estas proteínas (véase Buchberger (2002) Trends Cell Biol. 12, 216-221) y pueden identificarse mediante búsqueda usando bases de datos especializadas tales como SMART (Schultz *et al.* (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864; Letunic *et al.* (2002) Nucleic Acid Res 30, 242-244).

15 Los dominios PUG (en Peptide:N-Glicanasas and other supuesta nuclear UBX-domain-containing proteins; Doerks *et al.* (2002) Genome Research 12, 47-56) se producen conjuntamente en proteínas con dominios que son centrales para proteólisis mediada por ubiquitina, incluyendo dominios UBX (en mamíferos y plantas), UBA (en plantas) y UBC (en *Plasmodium*). Se cree que las proteínas que contienen PUG tales como PNGasas desempeñan un papel en la respuesta de proteínas no plegadas, un sistema de vigilancia de control de calidad del retículo endoplasmático (RE) que distingue proteínas aberrantes de proteínas correctamente plegadas. En algunos casos, se ha mostrado que estas proteínas mal plegadas y/o no plegadas se degradan mediante un denominado mecanismo de degradación asociado al RE, que implica el sistema ubiquitina-proteasoma (Suzuki *et al.* (2000) J. Cell Biol. 149, 1039-1052). También están presentes formas divergentes de dominios PUG en cinasas del tipo IRE1p que se sabe que funcionan en los estadios iniciales de la respuesta de proteínas no plegadas (Shamu y Walter (1996) EMBO J. 15, 3028-3039).

25 Una proteína de *Arabidopsis* recientemente caracterizada que comprende un dominio UBX es PUX1 (Rancour *et al.* (2004) J. Biol. Chem., publicación en línea 10.1074/jbc.M405498200). PUX1 es un gen único en *Arabidopsis* y probablemente se expresa de manera ubicua en plantas. Se mostró que la proteína era un inhibidor no competitivo de la ATPasa CDC48 de tipo AAA. PUX1 se asocia a través de su dominio UBX con la forma no hexamérica de CDC48, pero no con la CDC48 hexamérica. Se postula que PUX1 facilita el desensamblaje de CDC48 hexamérica activa y que se requiere el dominio N-terminal de la proteína para este proceso. Las plantas deficientes en *pux1* mostraron un desarrollo más rápido hasta la madurez pero no tenían anomalías morfológicas macroscópicas. Además de PUX1, se mostro que otras dos proteínas que comprenden dominios UBX, PUX2 y PUX3, interaccionan con CDC48 (Rancour *et al.*, 2004). PUX2 (At2g01650) se dio a conocer anteriormente en el documento WO 03/085115 (secuencia de gen y proteína descritas como SEQ ID NO: 1 y SEQ ID NO: 2 respectivamente).

35 Se ha encontrado ahora que introducir y sobreexpresar un ácido nucleico que codifica para una proteína GRUBX (proteína que comprende el dominio UBX relacionado con el crecimiento), y en particular un ácido nucleico que codifica para la proteína GRUBX mostrada a modo de ejemplo por SEQ ID NO: 2, en una planta proporciona plantas que tienen un índice de cosecha mejorado con respecto a plantas de tipo natural correspondientes. Por tanto, según una primera realización de la presente invención se proporciona un método para mejorar la producción de semillas de una planta, que comprende introducir y sobreexpresar en una planta una secuencia de ácido nucleico que codifica para una proteína GRUBX, seleccionándose dicha secuencia de ácido nucleico del grupo:

- una secuencia de ácido nucleico tal como se representa por una cualquiera de SEQ ID NO:1, 3, 6;
- un ácido nucleico que codifica para un polipéptido tal como se representa por una cualquiera de SEQ ID NO: 2, 4, 7;
- un ácido nucleico que codifica para un polipéptido que tiene al menos el 80% de identidad de secuencia con respecto a una cualquiera de SEQ ID NO: 2, 4, 7, teniendo dicho polipéptido acción reguladora de ubiquitina dentro de la ruta de ubiquitinación y cultivar la planta o célula vegetal en condiciones que promueven el crecimiento de la planta. Opcionalmente, pueden seleccionarse plantas que tienen una producción de semillas mejorada.

50 La recombinación homóloga permite la introducción en un genoma de un ácido nucleico seleccionado en una posición seleccionada definida. La recombinación homóloga es una tecnología convencional usada de manera rutinaria en las ciencias biológicas para organismos inferiores tales como levadura y el musgo *Physcomitrella*. Se han descrito métodos para realizar recombinación homóloga en plantas no sólo para plantas modelo (Offringa *et al.* (1990) EMBO J. 9, 3077-3084) sino también para plantas de cultivo, por ejemplo arroz (Terada *et al.*, (2002) Nature Biotechnol. 20, 1030-1034; o lida y Terada (2004) Curr. Opin. Biotechnol. 15, 132-138). El ácido nucleico que va a dirigirse (que puede ser una molécula de ácido nucleico GRUBX o variante de la misma tal como se definió anteriormente en el presente documento) no necesita dirigirse al locus de un gen GRUBX, sino que puede introducirse en, por ejemplo, regiones de alta expresión. El ácido nucleico que va a dirigirse puede ser un alelo

mejorado usado para sustituir el gen endógeno o puede introducirse en adición al gen endógeno.

El ácido nucleico puede introducirse en una planta mediante, por ejemplo, transformación.

El término proteína GRUBX, tal como se define en el presente documento, se refiere a un ácido nucleico que codifica para un polipéptido que tiene al menos el 80% de identidad de secuencia con respecto a una cualquiera de SEQ ID NO: 2, 4, 7, teniendo dicho polipéptido acción reguladora de ubiquitina dentro de la ruta de ubiquitinación, y opcionalmente también un dominio de dedos de zinc. Preferiblemente, la proteína GRUBX está relacionada estructuralmente con la proteína UBXD1 humana (SPTREMBL AAH07414). Preferiblemente, la proteína GRUBX es de una planta. Preferiblemente de manera adicional, la proteína GRUBX es de la familia de Solanaceae, más preferiblemente la GRUBX es una proteína de *Nicotiana tabacum*, lo más preferiblemente la GRUBX es una proteína tal como se representa por SEQ ID NO: 2 o un homólogo, derivado o fragmento activo de la misma, homólogos, derivados o fragmentos activos que tienen actividad biológica similar a la de SEQ ID NO: 2. Sin embargo, debería entenderse que podría usarse proteínas GRUBX de las plantas monocotiledóneas igualmente bien en los métodos de la presente invención, incluyendo proteínas GRUBX de *Zea mays*, *Saccharum officinarum* (SEQ ID NO 4), *Oryza sativa* (SEQ ID NO 7), *Triticum sp.*, *Hordeum sp.*, y *Sorghum sp.*, puesto que estas secuencias están relacionadas con SEQ ID NO 2 (véase la figura 1b).

Una de las actividades de una proteína GRUBX es aumentar la producción de semillas, en particular aumentar el índice de cosecha, cuando un ácido nucleico que codifica para tal proteína GRUBX se expresa en arroz bajo el control de un promotor de prolamina tal como se usa en la presente invención. Ventajosamente, una proteína GRUBX puede interactuar con proteínas CDC48 vegetales en condiciones descritas en Rancour *et al.* (2004).

Se analizaron las proteínas GRUBX de *Nicotiana tabacum* con la herramienta SMART y se usaron para examinar la base de datos Pfam (versión 11.0, noviembre de 2003; Bateman *et al.* (2002) Nucl. Acids Res. 30, 276-280) e InterPro (versión 7.0, 22 de julio de 2003; Mulder *et al.* (2003) Nucl. Acids. Res. 31, 315-318). Las proteínas GRUBX comprenden un dominio UBX (PF00789, SM00166, IPR001012) y un dominio PUG (SM00580, IPR006567). El dominio UBX, tal como se define en InterPro, se encuentra en proteínas reguladoras de ubiquitina, que son miembros de la ruta de ubiquitinación, así como varias otras proteínas incluyendo FAF-1 (factor 1 asociado a FAS), la proteína de reproducción Rep-8 humana y varias proteína hipotéticas de levadura. En *Arabidopsis*, existen aproximadamente veinte proteínas que se pronostica que comprenden este dominio. El dominio PUG se encuentra en proteínas cinasas, N-glicanasas y otras proteínas nucleares en eucariotas y se postula que está implicado en interacciones proteína-proteína (para una revisión véase Suzuki & Lennarz (2003) Biochem Biophys Res Commun. 302,1-5 y Biochem Biophys Res Commun. 303, 732) y en unión a ARN (Doerks *et al.*, 2002). A menudo se encuentran dominios PUG junto con dominios UBA o UBX en proteínas de *Arabidopsis* (Doerks *et al.*, 2002). Se proporciona una secuencia consenso para los dominios UBX y PUG, tal como se define en la base de datos SMART (software versión 4.0, actualización de la base de datos de secuencias del 15 de septiembre de 2003) en la figura 2a; la figura 2b muestra los dominios UBX y PUG de respectivamente SEQ ID NO 2 y SPTREMBL Q9ZU93; la figura 2c muestra una alineación de BLAST de estas 2 proteínas; y las figuras 2d y 2e presentan una alineación entre SEQ ID NO 2 y SEQ ID NO 4, y SEQ ID NO 4 y SEQ ID NO 7, respectivamente. Se indican los dominios PUG y UBX.

Opcionalmente, puede estar presente un dominio de dedos de zinc en la proteína GRUBX. Los dominios de dedos de zinc, tal como se define en InterPro, son estructuras proteicas de unión a ácido nucleico que se identificaron por primera vez en el factor de transcripción de *Xenopus laevis* TFIIIA. Estos dominios se han encontrado desde entonces en numerosas proteínas de unión a ácido nucleico. Un dominio de dedo de zinc está compuesto por de 25 a 30 residuos de aminoácido incluyendo 2 residuos Cys conservados y 2 His conservados en un motivo de tipo C-2-C-12-H-3-H. Los 12 residuos que separan la segunda Cys y la primera His son principalmente polares y básicos, lo que indica que esta región está implicada en la unión a ácido nucleico. El motivo de dedo de zinc es un dominio de autoplegamiento, inusualmente pequeño, en el que el Zn es un componente crucial de su estructura terciaria. Todos los dominios de dedos de zinc unen un átomo de Zn en una red tetrahédrica dando como resultado la formación de una proyección similar a un dedo que puede interactuar con los nucleótidos en el surco mayor del ácido nucleico. El Zn se une a los residuos de Cys y His conservados. Se ha encontrado que los dedos se unen a aproximadamente 5 pares de bases de ácido nucleico que contiene tramos cortos de residuos de guanina, y tienen la capacidad de unirse tanto a ARN como a ADN. El dedo de zinc puede representar por tanto la proteína de unión a ácido nucleico original. También se ha sugerido que un dominio con Zn centrado podría usarse en una interacción de proteínas, por ejemplo en la proteína cinasa C. Muchas clases de dedos de zinc se caracterizan según el número y las posiciones de los residuos de histidina y cisteína implicados en el posicionamiento espacial del átomo de zinc. En la primera clase que va a caracterizarse, denominada C2H2 (IPR007087), el primer par de residuos de coordinación de zinc consiste en cisteínas, mientras que el segundo par son histidinas. Otro dominio de dedos de zinc (IPR006642) puede ser del tipo encontrado en la proteína de *Saccharomyces cerevisiae* Rad18. En este caso también, el dominio de dedos de zinc es una supuesta secuencia de unión a ácido nucleico. El dominio de dedo de zinc opcional en la proteína GRUBX tal como se define en el presente documento no se limita sin embargo al tipo C2H2 o Rad18, sino que puede ser cualquier tipo de dominio de dedos de zinc.

La expresión gen/ácido nucleico *GRUBX*, tal como se define en el presente documento, se refiere a cualquier ácido

nucleico que codifica para una proteína GRUBX tal como se definió anteriormente, o el complemento de la misma. El ácido nucleico puede derivarse (o bien directa o bien indirectamente (si se modifica posteriormente)) de cualquier fuente natural o artificial siempre que el ácido nucleico, cuando se expresa en una planta, conduzca a la expresión de un gen/ácido nucleico GRUBX aislado. El ácido nucleico puede aislarse de una fuente microbiana, tal como bacterias, levaduras u hongos, o de una fuente vegetal, de algas o animal (incluyendo humana). Preferiblemente el ácido nucleico se deriva de un organismo eucariota. Preferiblemente el ácido nucleico GRUBX es de origen vegetal, preferiblemente de manera adicional de origen de plantas monocotiledóneas o dicotiledóneas, más preferiblemente el ácido nucleico GRUBX codifica para una proteína GRUBX de la familia de Solanaceae, además preferiblemente el ácido nucleico GRUBX es una secuencia de ácido nucleico de *Nicotiana tabacum*, lo más preferiblemente el ácido nucleico GRUBX es una secuencia de ácido nucleico tal como se representa por SEQ ID NO: 1, y también abarca ácidos nucleicos que codifican para una secuencia de aminoácidos representada por SEQ ID NO: 2. Alternativamente, el ácido nucleico que codifica para una proteína GRUBX puede derivarse de la familia de la Poaceae, preferiblemente de *Oryza sativa*. Este ácido nucleico puede modificarse sustancialmente de su forma nativa en composición y/o entorno genómico a través de manipulación humana deliberada. La secuencia de ácido nucleico preferiblemente es una secuencia de ácido nucleico homóloga, es decir una secuencia de ácido nucleico estructural y/o funcionalmente relacionada, obtenida preferiblemente de una planta, ya sea de la misma especie vegetal o diferente.

La secuencia representada por SEQ ID NO: 6 era hasta la fecha desconocida.

La secuencia representada por SEQ ID NO: 4 se ensambló a partir de 4 secuencias EST (CA154270, CA144028, BQ535511 & CA184742) y era hasta la fecha desconocida. Los métodos según la presente invención también pueden ponerse en práctica introduciendo en una planta al menos una parte de un cromosoma (natural o artificial) (tal como un cromosoma artificial bacteriano (BAC)), cromosoma que contiene al menos una secuencia de ácido nucleico/gen que codifica para una proteína GRUBX tal como se definió anteriormente (tal como SEQ ID NO: 1 o SEQ ID NO 3), preferiblemente junto con uno o más miembros de la familia génica relacionada y/o secuencia(s) de ácido nucleico que codifica(n) para proteínas reguladoras para la expresión y/o actividad de GRUBX. Por tanto, según un aspecto adicional de la presente invención, se proporciona un método para mejorar la producción de semillas de plantas introduciendo en una planta al menos una parte de un cromosoma que comprende al menos un gen/ácido nucleico que codifica para una proteína GRUBX.

Según otro aspecto de la presente invención, puede aprovecharse el ácido nucleico que codifica para una proteína GRUBX en programas de cría. La secuencia de ácido nucleico puede estar en un cromosoma, o una parte del mismo, que comprende al menos la secuencia de ácido nucleico que codifica para la proteína GRUBX y preferiblemente también uno o más miembros de la familia relacionados. En un ejemplo de un programa de cría de este tipo, se identifica un marcador de ADN que puede estar genéticamente ligado a un gen que puede modular la expresión de un ácido nucleico que codifica para una proteína GRUBX en una planta, gen que puede ser un gen que codifica para la propia proteína GRUBX o cualquier otro gen que puede influir directa o indirectamente en la expresión del gen que codifica para una proteína GRUBX y/o actividad de la propia proteína GRUBX. Este marcador de ADN puede usarse entonces en programas de cría para seleccionar plantas que tienen una producción de semillas mejorada.

La presente invención por tanto se extiende al uso de una secuencia de ácido nucleico que codifica para una proteína GRUBX en programas de cría.

Según la presente invención, se prevé el aumento o la potenciación de la expresión de un ácido nucleico. Están bien documentados en la técnica métodos para obtener el aumento o la potenciación de la expresión de genes o productos genéticos e incluyen, por ejemplo, sobreexpresión dirigida por un promotor (fuerte), el uso de potenciadores de la transcripción o potenciadores de la traducción. Pueden introducirse ácidos nucleicos aislados que sirven como elementos promotores o potenciadores en una posición apropiada (normalmente en el sentido de 5') de una forma no hereditaria de un polinucleótido de modo que se regule por incremento la expresión de un ácido nucleico GRUBX tal como se definió anteriormente. Por ejemplo, pueden alterarse promotores endógenos *in vivo* mediante mutación, delección, y/o sustitución (véase Kmiec, patente estadounidense n.º 5.565.350; Zarlign *et al.*, PCT/US93/03868), o pueden introducirse promotores aislados en una célula vegetal en la orientación y distancia apropiadas de un gen de la presente invención de modo que se controle la expresión del gen. Preferiblemente, los ácidos nucleicos útiles en la presente invención se sobreexpresan en una planta o célula vegetal. El término sobreexpresión tal como se usa en el presente documento significa cualquier forma de expresión que es adicional al nivel de expresión de tipo natural original. Preferiblemente, el ácido nucleico que va a introducirse en la planta y/o el ácido nucleico que va a sobreexpresarse en las plantas está en una dirección sentido con respecto al promotor al que está operativamente unido. Preferiblemente, el ácido nucleico que va a sobreexpresarse codifica para una proteína GRUBX, preferiblemente de manera adicional la secuencia de ácido nucleico que codifica para la proteína GRUBX se aísla de una planta dicotiledónea, preferiblemente de la familia Solanaceae, preferiblemente de manera adicional en la que la secuencia se aísla de *Nicotiana tabacum*, lo más preferiblemente la secuencia de ácido nucleico es tal como se representa por SEQ ID NO: 1, o codifica para una secuencia de aminoácidos tal como se representa por SEQ ID NO: 2. Alternativamente, la secuencia de ácido nucleico que codifica para la proteína

GRUBX es tal como se representa por MIPS No. At2g01650, SEQ ID NO: 3 ó 6, o codifica para una secuencia de aminoácidos tal como se representa por Q9ZU93, SEQ ID NO: 4 ó 7. Debe observarse que la aplicabilidad de la invención no descansa en el uso del ácido nucleico representado por SEQ ID NO: 1, ni sobre la secuencia de ácido nucleico que codifica para la secuencia de aminoácidos de SEQ ID NO: 2, sino que pueden usarse otras secuencias de ácido nucleico en los métodos de la presente invención. En particular, los ácidos nucleicos útiles en los métodos de la presente invención codifican para proteínas que tienen al menos el 80% de identidad de secuencia con respecto a una cualquiera de SEQ ID NO: 2, 4, 7, teniendo dicho polipéptido acción reguladora de ubiquitina dentro de la ruta de ubiquitinación, y Opcionalmente también un dominio de dedos de zinc.

Se transforman las plantas con un vector que comprende la secuencia de interés (es decir, la secuencia de ácido nucleico que puede modular la expresión de ácido nucleico que codifica para una proteína GRUBX), secuencia que está operativamente unida a una o más secuencias de control (al menos un promotor). Las expresiones "elemento regulador", "secuencia de control" y "promotor" se usan todas en el presente documento de manera intercambiable y han de tomarse en un contexto amplio para referirse a secuencias de ácido nucleico reguladoras que pueden efectuar la expresión de las secuencias a las que están ligadas. Se abarcan por las expresiones mencionadas anteriormente secuencias reguladoras de la transcripción derivadas de un gen genómico eucariota clásico (incluyendo la caja de TATA que se requiere para la iniciación precisa de la transcripción, con o sin una secuencia de caja CCAAT) y elementos reguladores adicionales (es decir, silenciadores, potenciadores y secuencias de activación en el sentido de 5') que alteran la expresión génica en respuesta a estímulos del desarrollo y/o externos, o de una manera específica de tejido. También se incluye dentro de la expresión 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. La expresión "elemento regulador" también abarca una molécula de fusión sintética o derivado que confiere, activa o potencia la expresión de una molécula de ácido nucleico en una célula, tejido u órgano. La expresión "operativamente unido" tal como se usa en el presente documento se refiere a un ligamiento funcional entre la secuencia promotora y el gen de interés, de modo que la secuencia promotora puede iniciar la transcripción del gen de interés. Ventajosamente, puede usarse cualquier tipo de promotor para dirigir la expresión de la secuencia de ácido nucleico dependiendo del resultado deseado. Los promotores adecuados incluyen promotores que son activos en plantas monocotiledóneas tales como arroz o maíz.

La secuencia de ácido nucleico que codifica para una proteína GRUBX está operativamente unida a un promotor preferido en semillas. La expresión "preferido en semillas" tal como se define en el presente documento se refiere a un promotor que se expresa predominantemente en tejido de semillas, pero no necesariamente de manera exclusiva en este tejido. La expresión "preferido en semillas" abarca todos los promotores que son activos en semillas. Tejido de semillas abarca cualquier parte de la semilla incluyendo el endospermo, aleurona o embrión. Preferiblemente, el promotor preferido en semillas es un promotor de prolamina, o un promotor de potencia similar y/o un promotor con un patrón de expresión similar. Lo más preferiblemente, el promotor de prolamina es tal como se representa por los nucleótidos 1-654 en la casete de expresión de SEQ ID NO: 5. Pueden analizarse la potencia del promotor y/o el patrón de expresión por ejemplo acoplando el promotor a un gen indicador y sometiendo a ensayo la expresión del gen indicador en diversos tejidos de la planta. Un gen indicador adecuado bien conocido por un experto en la técnica es *beta-glucuronidasa* bacteriana. Se presentan ejemplos de otros promotores preferidos de semillas en la tabla 1, y estos promotores son útiles para los métodos de la presente invención.

TABLA 1: Ejemplos de promotores preferidos de semillas para su uso en la realización de la presente invención:

FUENTE DEL GEN	PATRÓN DE EXPRESIÓN	REFERENCIA
genes específicos de semillas	semilla	Simon, <i>et al.</i> , Plant Mol. Biol. 5: 191, 1985; Scofield, <i>et al.</i> , J. Biol. Chem. 262: 12202, 1987.; Baszczynski, <i>et al.</i> , Plant Mol. Biol. 14: 633, 1990.
albúmina de nuez de Brasil	semilla	Pearson, <i>et al.</i> , Plant Mol. Biol. 18: 235-245, 1992.
legúmina	semilla	Ellis, <i>et al.</i> , Plant Mol. Biol. 10: 203-214, 1988.
glutelina (arroz)	semilla	Takaiwa, <i>et al.</i> , Mol. Gen. Genet. 208: 15-22, 1986; Takaiwa, <i>et al.</i> , FEBS Letts. 221: 43-47, 1987.
zeína	semilla	Matzke <i>et al</i> Plant Mol Biol, 14(3):323-32 1990
napA	semilla	Stalberg, <i>et al</i> , Planta 199: 515-519, 1996.
glutenina-1 de APM y BPM de trigo	endospermo	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
SPA de trigo	semilla	Albani <i>et al</i> , Plant Cell, 9: 171-184, 1997

(continuación)

FUENTE DEL GEN	PATRÓN DE EXPRESIÓN	REFERENCIA
α, β, γ -gliadinas de trigo	endospermo	EMBO J. 3:1409-15, 1984
promotor de <i>ltr1</i> de cebada	endospermo	
hordeína B1, C, D, de cebada	endospermo	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55, 1993; Mol Gen Genet 250: 750-60, 1996
DOF de cebada	endospermo	Mena <i>et al</i> , The Plant Journal, 116(1): 53-62, 1998
<i>blz2</i>	endospermo	Documento EP99106056.7
promotor sintético	endospermo	Vicente-Carbajosa <i>et al.</i> , Plant J. 13: 629-640, 1998.
prolamina de arroz NRP33	endospermo	Wu <i>et al</i> , Plant Cell Physiology 39(8) 885-889, 1998
α -globulina de arroz Glb-1	endospermo	Wu <i>et al</i> , Plant Cell Physiology 39(8) 885-889, 1998
OSH1 de arroz	embrión	Sato <i>et al</i> , Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
α -globulina de arroz REB/OHP-1	endospermo	Nakase <i>et al</i> . Plant Mol. Biol. 33: 513-522, 1997
ADP-glucosa de arroz PP	endospermo	Trans Res 6:157-68, 1997
familia génica ESR de maíz	endospermo	Plant J 12:235-46, 1997
γ -kafirina de sorgo	endospermo	PMB 32:1029-35, 1996
KNOX	embrión	Postma-Haarsma <i>et al</i> , Plant Mol. Biol. 39:257-71, 1999
oleosina de arroz	embrión y aleurona	Wu <i>et al</i> , J. Biochem., 123:386, 1998
oleosina de girasol	semilla (embrión y semilla seca)	Cummins, <i>et al.</i> , Plant Mol. Biol. 19: 873-876, 1992
PRO0117, supuesta proteína ribosómica 40S de arroz	débil en endospermo	Documento WO2004/070039
PRO0135, alfa-globulina de arroz	fuerte en endospermo	
PRO013, alanina aminotransferasa de arroz	débil en endospermo	
PRO0147, inhibidor de tripsina ITR1 (cebada)	débil en endospermo	
PRO0151, WSI18 de arroz	embrión + estrés	Documento WO2004/0700
PRO017, RAB21 de arroz	embrión + estrés	Documento WO2004/070039
PRO0218, oleosina de arroz 18kd	aleurona + embrión	

5 También puede añadirse una secuencia de intrón a la región no traducida en 5' o la secuencia codificante de la secuencia codificante parcial para aumentar la cantidad del mensajero maduro que se acumula en el citosol. La inclusión de un intrón que puede cortarse y empalmarse en la unidad de transcripción en constructos de expresión tanto vegetales como animales ha mostrado que aumenta la expresión génica tanto a los niveles de ARNm como de proteínas hasta 1000 veces (Buchman y Berg, Mol. Cell Biol. 8, 4395-4405 (1988); Callis *et al.*, Genes Dev. 1, 1183-1200 (1987)). Tal potenciación por intrón de la expresión génica normalmente es máxima cuando se coloca cerca del extremo 5' de la unidad de transcripción. Se conoce en la técnica el uso de los intrones de maíz intrón Adh1-S 1, 2, y 6, el intrón Bronze-1. Véase generalmente, The Maize Handbook, capítulo 116, Freeling y Walbot, Eds., Springer, N.Y. (1994).

15 La propia proteína GRUBX y/o el propio ácido nucleico GRUBX puede introducirse directamente en una célula vegetal o en la propia planta (incluyendo introducción en un tejido, órgano o cualquier otra parte de la planta). Según una característica preferida de la presente invención, el ácido nucleico se introduce preferiblemente en una planta

mediante transformación. El ácido nucleico es preferiblemente tal como se representa por SEQ ID NO: 1, o es un ácido nucleico que codifica para una secuencia de aminoácidos representada por SEQ ID NO: 2. Alternativamente, la secuencia de ácido nucleico es tal como se representa por cualquiera de MIPS No. At2g01650, SEQ ID NO: 3, SEQ ID NO 6. La secuencia de aminoácidos puede ser alternativamente una secuencia tal como se representa por cualquiera de SPTREMBL Q9ZU93, n.º de registro de GenBank AAR01744, SEQ ID NO: 4, SEQ ID NO 7.

El término “transformación” tal como se le hace referencia en el presente documento abarca la transferencia de un polinucleótido exógeno a una célula huésped, independientemente del método usado para la transferencia. El tejido vegetal que puede someterse a la propagación clonal posterior, ya sea por organogénesis o por embriogénesis, puede transformarse con un constructo genético de la presente invención y regenerarse una planta completa a partir del mismo. El tejido particular elegido variará dependiendo de los sistemas de propagación clonal disponibles para, y que mejor se adapten a, la especie particular que está transformándose. Dianas tisulares a modo de ejemplo incluyen discos de hojas, polen, embriones, cotiledones, hipocotilos, megagametofitos, tejido calloso, tejido meristemáticamente existente (por ejemplo, meristemo apical, yemas axilares y meristemos radiculares), y tejido de meristemo inducido (por ejemplo, meristemo de cotiledón y meristemo de hipocótilo). El polinucleótido puede introducirse de manera transitoria o estable en una célula huésped y puede mantenerse sin integrarse, por ejemplo, como un plásmido. Alternativamente, puede integrarse en el genoma del huésped. La célula vegetal transformada resultante puede usarse entonces para regenerar una planta transformada de una manera conocida por los expertos en la técnica.

La transformación de una especie vegetal es ahora una técnica bastante rutinaria. Ventajosamente, puede usarse cualquiera de varios métodos de transformación para introducir el gen de interés en una célula antecesora adecuada. Los métodos de transformación incluyen el uso de liposomas, electroporación, productos químicos que aumentan la captación de ADN libre, inyección del ADN directamente en la planta, bombardeo con pistola de partículas, transformación usando virus o polen y microproyección. Los métodos pueden seleccionarse del método de calcio/poli(etilenglicol) para protoplastos (Krens, F.A. *et al.*, 1982, *Nature* 296, 72-74; Negrutiu I. *et al.*, junio de 1987, *Plant Mol. Biol.* 8, 363-373); electroporación de protoplastos (Shillito R.D. *et al.*, 1985 *Bio/Technol* 3, 1099-1102); microinyección en material vegetal (Crossway A. *et al.*, 1986, *Mol. Gen Genet* 202, 179-185); bombardeo con partículas recubiertas con ARN o ADN (Klein T.M. *et al.*, 1987, *Nature* 327, 70) infección con virus (no integrativos) y similares. Se producen preferiblemente plantas de arroz transgénicas que sobreexpresan un gen *GRUBX* por medio transformación mediada por *Agrobacterium* usando cualquiera de los métodos bien conocidos para la transformación de arroz, tal como se describe en cualquiera de los siguientes: solicitud de patente europea publicada 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 transformación de maíz, el método preferido es tal como se describe en o bien Ishida *et al.* (*Nat. Biotechnol.* 1996 Jun; 14(6): 745-50) o bien Frame *et al.* (*Plant Physiol.* Mayode 2002; 129(1): 13-22).

Generalmente tras la transformación, se seleccionan células vegetales o agrupaciones celulares para detectar la presencia de uno o más marcadores que se codifican por genes expresables en plantas transferidos conjuntamente con el gen de interés, tras lo que se regenera el material transformado dando una planta completa.

Tras la transferencia y regeneración del ADN, pueden evaluarse las plantas supuestamente transformadas, por ejemplo usando análisis de tipo Southern, para detectar la presencia del gen de interés, el número de copias y/o la organización genómica. Alternativa o adicionalmente, los niveles de expresión del ADN recién introducido pueden monitorizarse usando análisis de tipo Northern y/o Western, siendo ambas técnicas bien conocidas por los expertos habituales en la técnica.

Las plantas transformadas generadas pueden propagarse mediante una variedad de medios, tales como mediante propagación clonal o técnicas de cría clásicas. Por ejemplo, una planta transformada de primera generación (o T1) pueden autofecundarse para dar transformantes de segunda generación (o T2) homocigotos, y las plantas T2 propagarse adicionalmente a través de técnicas de cría clásicas.

Los organismos transformados generados pueden tomar una variedad de formas. Por ejemplo, pueden ser quimeras de células transformadas y células no transformadas; transformantes clonales (por ejemplo, todas las células transformadas para que contengan el casete de expresión); injertos de tejidos transformados y no transformados (por ejemplo, en plantas, un rizoma transformado injertado en un esqueje no transformado). La presente invención se extiende claramente a cualquier célula vegetal o planta producida por cualquiera de los métodos descritos en el presente documento, y a todas las partes de la planta, propágulos y progenie de la misma. El término “planta” tal como se usa en el presente documento abarca plantas completas, antecesores y progenie de las plantas, partes de la planta, células vegetales, tejidos y órganos. El término “planta” también abarca por tanto cultivos en suspensión, embriones, regiones meristemáticas, tejido calloso, hojas, flores, frutos, semillas, raíces (incluyendo rizomas y tubérculos), brotes, bulbos, tallos, gametofitos, esporofitos, polen y microsporas. Las plantas que son particularmente útiles en los métodos de la invención incluyen algas, helechos y todas las plantas que pertenecen a la superfamilia Viridiplantae, en particular plantas monocotiledóneas y dicotiledóneas, incluyendo pasto o leguminosas forrajeras, plantas ornamentales, cultivos alimenticios, árboles o arbustos seleccionados de la lista que

comprende *Abelmoschus spp.*, *Acer spp.*, *Actinidia spp.*, *Agropyron spp.*, *Allium spp.*, *Amaranthus spp.*, *Ananas comosus*, *Annona spp.*, *Apium graveolens*, *Arabidopsis thaliana*, *Arachis spp.*, *Artocarpus spp.*, *Asparagus officinalis*, *Avena sativa*, *Averrhoa carambola*, *Benincasa hispida*, *Bertholletia excelsea*, *Beta vulgaris*, *Brassica spp.*, *Cadaba farinosa*, *Camellia sinensis*, *Canna indica*, *Capsicum spp.*, *Carica papaya*, *Carissa macrocarpa*, *Carthamus tinctorius*, *Carya spp.*, *Castanea spp.*, *Cichorium endivia*, *Cinnamomum spp.*, *Citrullus lanatus*, *Citrus spp.*, *Cocos spp.*, *Coffea spp.*, *Cola spp.*, *Colocasia esculenta*, *Corylus spp.*, *Crataegus spp.*, *Cucumis spp.*, *Cucurbita spp.*, *Cynara spp.*, *Daucus carota*, *Desmodium spp.*, *Dimocarpus longan*, *Dioscorea spp.*, *Diospyros spp.*, *Echinochloa spp.*, *Eleusine coracana*, *Eriobotrya japonica*, *Eugenia uniflora*, *Fagopyrum spp.*, *Fagus spp.*, *Ficus carica*, *Fortunella spp.*, *Fragaria spp.*, *Ginkgo biloba*, *Glycine spp.*, *Gossypium hirsutum*, *Helianthus spp.*, *Hibiscus spp.*, *Hordeum spp.*, *Ipomoea batatas*, *Juglans spp.*, *Lactuca sativa*, *Lathyrus spp.*, *Lemna spp.*, *Lens culinaris*, *Linum usitatissimum*, *Litchi chinensis*, *Lotus spp.*, *Luffa acutangula*, *Lupinus spp.*, *Macrotyloma spp.*, *Malpighia emarginata*, *Malus spp.*, *Mammea americana*, *Mangifera indica*, *Manihot spp.*, *Manilkara zapota*, *Medicago sativa*, *Melilotus spp.*, *Mentha spp.*, *Momordica spp.*, *Morus nigra*, *Musa spp.*, *Nicotiana spp.*, *Olea spp.*, *Opuntia spp.*, *Ornithopus spp.*, *Oryza spp.*, *Panicum miliaceum*, *Passiflora edulis*, *Pastinaca sativa*, *Persea spp.*, *Petroselinum crispum*, *Phaseolus spp.*, *Phoenix spp.*, *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.*, *Rubus spp.*, *Saccharum spp.*, *Sambucus spp.*, *Secale cereale*, *Sesamum spp.*, *Solanum spp.*, *Sorghum bicolor*, *Spinacia spp.*, *Syzygium spp.*, *Tamarindus indica*, *Theobroma cacao*, *Trifolium spp.*, *Triticosecale rimpaii*, *Triticum spp.*, *Vaccinium spp.*, *Vicia spp.*, *Vigna spp.*, *Vitis spp.*, *Zea mays*, *Zizania palustris*, *Ziziphus spp.*, entre otras.

Según una característica preferida de la presente invención, la planta es una planta de cultivo que comprende soja, girasol, canola, alfalfa, colza o algodón. Preferiblemente de manera adicional, la planta según la presente invención es una planta monocotiledónea tal como caña de azúcar, lo más preferiblemente un cereal, tal como arroz, maíz, trigo, mijo, cebada, centeno, sorgo o avena. Sin embargo, se prevé que los métodos de la presente invención pueden aplicarse a una amplia variedad de plantas, puesto que la conservación de dominios entre los homólogos de GRUBX eucariota conocidos sugiere una función igualmente conservada en el metabolismo celular.

Ventajosamente, la realización de los métodos según la presente invención da como resultado plantas que tienen una producción de semillas mejoradas con respecto a plantas de tipo natural correspondientes.

La presente invención se refiere a métodos para mejorar el índice de cosecha.

El índice de cosecha mejorado se expresa como una razón de la biomasa total con respecto a la producción de partes cosechables, tales como semillas; o peso de mil granos. Según la presente invención, la realización de los métodos según la presente invención da como resultado plantas que tienen un aumento en la producción de semillas, con respecto a plantas de control. Por tanto, según la presente invención, se proporciona un método para aumentar la producción de semillas, método que comprende introducir y sobreexpresar una secuencia de ácido nucleico que codifica para una proteína GRUBX tal como se definió anteriormente, preferiblemente en el que la proteína GRUBX se codifica por una secuencia de ácido nucleico representada por SEQ ID NO: 1 o en el que la proteína GRUBX se representa por SEQ ID NO: 2. Alternativamente, la GRUBX puede codificarse por una secuencia de ácido nucleico representada por cualquiera de MIPS No. At2g01650, SEQ ID NO: 3, o en el que la GRUBX se representa por cualquiera de SPTreMBL Q9ZU93, SEQ ID NO: 4.

Los métodos de la presente invención son favorables para aplicarse a plantas de cultivo debido a que los métodos de la presente invención se usan para aumentar la producción de semillas de una planta. Por tanto, los métodos de la presente invención son particularmente útiles para plantas de cultivo cultivadas por sus semillas, tales como cereales. Por consiguiente, una realización particular de la presente invención se refiere a un método para aumentar la producción de semillas de un cereal.

Se produce un aumento en la producción y/o el crecimiento ya esté la planta en condiciones no estresantes o se exponga la planta a diversos estreses en comparación con plantas de control. Las plantas normalmente responden a la exposición al estrés creciendo más lentamente. En condiciones de estrés agudo, la planta puede incluso detener su crecimiento completamente. El estrés leve se define por otro lado en el presente documento como cualquier estrés al que se expone una planta que no da como resultado el cese del crecimiento de la planta completamente sin la capacidad de reanudar el crecimiento. Debido a los avances en las prácticas agrícolas (irrigación, fertilización, tratamientos con pesticidas), no se encuentra a menudo estrés agudo en plantas de cultivo cultivadas. Como consecuencia, el crecimiento comprometido inducido por el estrés leve es a menudo una característica indeseable para la agricultura. El estrés leve son los estreses típicos a los que puede estar expuesta una planta. Este estrés puede ser el estrés biótico y/o abiótico diario (medioambiental) al que está expuesta una planta. Los estreses abióticos o medioambientales típicos incluyen estreses por temperatura provocados por temperaturas frías/de congelación o cálidas atípicas, estrés salino, estrés por agua (sequía o exceso de agua). Los estreses abióticos también pueden provocarse por productos químicos. Estreses bióticos son normalmente los estreses provocados por patógenos, tales como bacterias, virus, hongos o insectos.

La presente invención también se refiere al uso de un ácido nucleico que codifica para una proteína GRUBX en la mejora de la producción de semillas de plantas. La presente invención también se refiere al uso de una proteína GRUBX en la mejora de la producción de semillas de plantas. La secuencia de ácido nucleico es preferiblemente tal como se representa por SEQ ID NO: 1, 3, ó 6, o codifica para una secuencia de aminoácidos representada por SEQ ID NO: 2, 4, 7.

La presente invención se describirá ahora con referencia a las siguientes figuras en las que:

Figura 1a. Árbol filogenético que representa proteínas de *Arabidopsis thaliana* y proteínas de referencia animales que comprenden un dominio UBX, tal como se reconoce mediante la herramienta SMART. Las proteínas humanas se representan por sus números de registro de GenBank NP_079517 (dominio UBX de *Homo sapiens* que contiene 1 (UBXD1)), AAP97263 (ARNm del factor proteico asociado a Fas FAF1 de *Homo sapiens*), NP_005662 (reproducción 8 de *Homo sapiens* (D8S2298E), REP8) y una proteína de rata mediante NP_114187 (proteína p47 de *Rattus norvegicus*). Los otros identificadores (excepto para SEQ ID NO 2, SEQ ID NO 4 y SEQ ID NO 7) tienen números de registro de GenBank o SPTreMBL para proteínas de *Arabidopsis thaliana*.

Figura 1b. Árbol filogenético que representa proteínas vegetales que comprenden un dominio PUG, tal como se reconoce mediante la herramienta SMART. Se comparan SEQ ID NO 2 y SEQ ID NO 4 con proteínas de *Arabidopsis thaliana* (números de registro de SPTreMBL Q9ZU93 (proteína expresada), Q9FKI1 (similitud con la metaloproteinasas de zinc), Q9MAT3 (proteína F13M7.16), Q9FKC7 (ADN genómico, cromosoma 5, clon TAC: K24G6), Q9SF12 (proteína hipotética), Q9C5S2 (endorribonucleasa/proteína cinasa IRE1), Q8RX75 (AT5g24360/K16H17_7), Q94IG5 (homologo-1 de Ire1)), y con la proteína de arroz SPTreMBL Q7XIT1 (OsIre1p).

Figura 2a. Definición de los dominios UBX1 y PUG por sus secuencias consenso (base de datos SMART). CONSENSO/50%, respectivamente /65% y /80% son las secuencias de consenso para el 50, 65 y 80% superiores de las secuencias de referencia que comprenden el dominio UBX1 o PUG. Las letras mayúsculas son los códigos de la IUPAC de una única letra convencionales para los diversos aminoácidos, las otras letras simbolizan la naturaleza de los aminoácidos tal como se explica resumidamente a continuación:

Clase	Clave	Residuos
Alcohol	o	S,T
Alifático	l	I,L,V
Cualquier	.	A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
Aromático	a	F,H,W,Y
Cargado	c	D,E,H,K,R
Hidrófobo	h	A,C,F,G,H,I,K,L,M,R,T,V,W,Y
Negativo	-	D,E
Polar	p	C,D,E,H,K,N,Q,R,S,T
Positivo	+	H,K,R
Pequeño	s	A,C,D,G,N,P,S,T,V
Diminuto	u	A,G,S
Giro	t	A,C,D,E,G,H,K,N,Q,R,S,T

Figura 2b. Secuencias de los dominios UBX y PUG presentes en SEQ ID NO 2 y en Q9ZU93.

Figura 2c. Alineación de Q9ZU93 y SEQ ID NO 2, dominios PUG subrayados, dominios UBX en negrita.

Figura 2d. Alineación de SEQ ID NO 2 y SEQ ID NO 4, dominios PUG subrayados, dominios UBX en negrita.

Figura 2e. Alineación de SEQ ID NO 4 y SEQ ID NO 7, dominios PUG subrayados, dominios UBX en negrita.

Figura 3. Presentación esquemática del clon p77 de entrada, que contiene CDS0669 dentro de los sitios AttL1 y AttL2 para clonación Gateway® en la estructura principal de pDONR201. CDS0669 es el código interno de la secuencia codificante de GRUBX de tabaco. Este vector contiene también un casete de resistencia a kanamicina bacteriano y un origen bacteriano de replicación.

Figura 4. Vector binario para la expresión en *Oryza sativa* del gen GRUBX de tabaco (CDS0669) bajo el control del promotor de prolamina (PRO0090) Este vector contiene un T-ADN derivado del plásmido Ti, limitado por un borde izquierdo (repetición de LB, LB Ti C58) y un borde derecho (repetición de RB, RB Ti C58). Del borde izquierdo al

borde derecho, este T-ADN contiene: un casete para la selección por antibióticos de plantas transformadas; un casete para la selección visual de plantas transformadas; el casete terminador doble PRO0090 CDS0669-zeína y rbcS-deltaGA para la expresión del gen *GRUBX* de tabaco. Este vector también contiene un origen de replicación de pBR322 para la replicación bacteriana y un marcador seleccionable (Spe/SmeR) para la selección bacteriana con espectinomomicina y estreptomomicina.

Figura 5. Ejemplos de secuencias útiles en la presente invención. SEQ ID NO: 1 y SEQ ID NO: 2 son las secuencias del ácido nucleico *GRUBX* y la proteína GRUBX respectivamente que se usaron en los ejemplos. SEQ ID NO: 3 y SEQ ID NO: 4 representan la secuencia codificante y la secuencia de proteína del ortólogo de GRUBX de la caña de azúcar, SEQ ID NO: 5 es la secuencia del casete de expresión que se usó en las plantas de arroz transformadas, SEQ ID NO: 6 y SEQ ID NO: 7 representan la secuencia codificante respectivamente y la secuencia de proteína del ortólogo de GRUBX de arroz

Ejemplos

La presente invención se describirá ahora con referencia a los siguientes ejemplos, que son a modo de ilustración solo.

Manipulación del ADN: a menos que se establezca lo contrario, se realizan técnicas de ADN recombinante según protocolos convencionales descritos en (Sambrook (2001) Molecular Clonación: a laboratory manual, 3ª edición Cold Spring Harbor Laboratory Press, CSH, Nueva York) o en los volúmenes 1 y 2 de Ausubel *et al.* (Current Protocols in Molecular Biology. Nueva York: John Wiley y Sons, 1998). Se describen materiales y métodos convencionales para el trabajo molecular con plantas en Plant Molecular Biology Labfax (1993) por R.D.D. Croy, publicado por BIOS Scientific Publications Ltd (RU) y Blackwell Scientific Publications (RU).

Ejemplo 1: Clonación de la secuencia CDS0669

Clonación del fragmento del gen *GRUBX* a partir de tabaco

Se realizó un experimento de ADNc-AFLP en un cultivo celular BY2 de tabaco sincronizado (*Nicotiana tabacum* L. cv. Bright Yellow-2), y se eligieron etiquetas de secuencia expresada de BY2 que estaban moduladas por el ciclo celular para la clonación adicional. Se usaron las etiquetas de secuencia expresada para seleccionar una biblioteca de ADNc de tabaco y para aislar el ADNc de longitud completa de interés, concretamente uno que codifica para el gen *GRUBX* (CDS0669).

Sincronización de células BY2.

Se sincronizó una suspensión celular cultivada de BY2 de tabaco (*Nicotiana tabacum* L. cv. Bright Yellow-2) mediante células que se bloqueaban en la fase S temprana con afidicolina tal como sigue. Se mantuvo la suspensión celular de *Nicotiana tabacum* L. cv. Bright Yellow 2 tal como se describió (Nagata *et al.* Int. Rev. Cytol. 132, 1-30, 1992). Para la sincronización, se diluyó un cultivo estacionario de 7 días de edad 10 veces en medio fresco complementado con afidicolina (Sigma-Aldrich, St. Louis, MO; 5 mg/l), un fármaco inhibidor de ADN-polimerasa α . Tras 24 h, se liberaron las células del bloqueo mediante varios lavados con medio fresco tras lo cual se reanudó su progresión en el ciclo celular.

Extracción de ARN y síntesis de ADNc.

Se preparó ARN total usando precipitación con LiCl (Sambrook *et al.*, 2001) y se extrajo ARN poli(A+) a partir de 500 μ g de ARN total usando columnas Oligotex (Qiagen, Hilden, Alemania) según las instrucciones del fabricante. Partiendo de 1 μ g de ARN poli(A+), se sintetizó ADNc de primera cadena mediante transcripción inversa con un cebador de oligo-dT₂₅ biotinilado (Genset, Paris, Francia) y Superscript II (Life Technologies, Gaithersburg, MD). Se realizó la síntesis de la segunda de cadena mediante desplazamiento de cadena con ligasa de *Escherichia coli* (Life Technologies), ADN-polimerasa I*(USB, Cleveland, OH) y ARNasa-H (USB).

Análisis de ADNc-AFLP.

Se usó quinientos ng de ADNc bicatenario para análisis de AFLP tal como se describió (Vos *et al.*, Nucleic Acids Res. 23 (21) 4407-4414, 1995; Bachem *et al.*, Plant J. 9 (5) 745-53, 1996) con modificaciones. Las enzimas de restricción usadas fueron BstYI y MseI (Biolabs) y se realizó la digestión en dos etapas separadas. Tras la primera digestión de restricción con una de las enzimas, se atraparon los fragmentos de extremo 3' en perlas Dyna (Dyna, Oslo, Noruega) por medio de su cola biotinilada, mientras que los otros fragmentos se eliminaron con el lavado. Tras la digestión con la segunda enzima, se recogieron los fragmentos de restricción liberados y se usaron como moldes en las etapas de AFLP posteriores. Para las amplificaciones previas, se combinó un cebador de *Msil* sin nucleótidos selectivos con un cebador de BstYI que contenía o bien una T o bien una C como nucleótido mayoritario en 3'. Las

condiciones de PCR fueron tal como se describe (Vos *et al.*, 1995). Se diluyeron las mezclas de amplificación obtenidas 600 veces y se usaron 5 µl para amplificaciones selectivas usando un cebador de *Bst*YI marcado con P³³ y la polimerasa Amplitaq-Gold (Roche Diagnostics, Bruselas, Bélgica). Se separaron los productos de amplificación en geles de poliacrilamida al 5% usando el sistema Sequigel (Biorad). Se expusieron los geles secos a películas Kodak Biomax así como se escanearon en un instrumento PhosphorImager (Amersham Pharmacia Biotech, Little Chalfont, RU).

Caracterización de los fragmentos de AFLP.

Se aislaron del gel bandas que correspondían a transcritos expresados de manera diferente, entre los que estaba el transcrito (parcial) que correspondía a SEQ ID NO 1 (o CDS0669), y volvió a amplificarse el ADN eluido en las mismas condiciones que para la amplificación selectiva. Se obtuvo la información de secuencia o bien mediante secuenciación directa del producto de reacción en cadena de la polimerasa reamplificado con el cebador de *Bst*YI selectivo o bien tras la clonación de los fragmentos en pGEM-T easy (Promega, Madison, WI) y secuenciación de clones individuales. Se compararon las secuencias obtenidas frente a secuencias de nucleótidos y proteínas presentes en la base de datos públicamente disponible mediante alineaciones de secuencias BLAST (Altschul *et al.*, Nucleic Acids Res. 25 (17) 3389-3402 1997). Cuando estaban disponibles, se reemplazaron las secuencias de etiqueta por EST más largos o secuencias de ADNc aisladas para aumentar la posibilidad de encontrar una homología significativa. El clon de ADNc físico que correspondía a SEQ ID NO 1 (CDS0669) se amplificó posteriormente de una biblioteca de ADNc de tabaco comercial tal como sigue:

Clonación del gen *GRUBX* (CDS0669)

Se preparó una biblioteca de ADNc con un tamaño promedio de insertos de 1.400 pb de ARN poli(A+) aislado de células de tabaco BY2 no sincronizadas, en división activa. Se clonaron estos insertos de biblioteca en el vector pCMVSPORT6.0, que comprende un casete attB Gateway (Life Technologies). De esta biblioteca, se seleccionaron 46.000 clones, se dispusieron en placas de microtitulación de 384 pocillos y se colocaron posteriormente por duplicado en filtros de nailon. Se seleccionaron los clones dispuestos usando conjuntos de varios cientos de etiquetas marcadas radiactivamente como sondas (incluyendo la etiqueta BY2 que corresponde a la secuencia CDS0669, SEQ ID NO 1). Se aislaron clones positivos (entre los que estaba el clon que correspondía a CDS0669, SEQ ID NO 1), se secuenciaron y se alinearon con la secuencia de etiqueta. Cuando la hibridización con la etiqueta fallaba, se seleccionó el ADNc de longitud completa que correspondía a la etiqueta mediante amplificación por PCR: se diseñaron cebadores específicos de etiqueta usando el programa primer3 (http://www-genoma.wi.mit.edu/genoma_software/other/primer3.html) y se usaron en combinación con un cebador de vector común para amplificar insertos de ADNc parciales. Se usaron conjuntos de ADN de 50.000, 100.000, 150.000 y 300.000 clones de ADNc como moldes en las amplificaciones por PCR. Se aislaron entonces los productos de amplificación de geles de agarosa, se clonaron, se secuenciaron y se alineó su secuencia con las de las etiquetas. A continuación, se clonó el ADNc de longitud completa que correspondía a la secuencia de nucleótidos de la SEQ ID NO 1 del vector de biblioteca pCMVSPORT6.0 en pDONR201, un vector donador Gateway® (Invitrogen, Paisley, RU) por medio de una reacción LR, dando como resultado el clon de entrada p77 (figura 3).

Ejemplo 2: Construcción del vector

Se usó posteriormente el clon de entrada p77 en una reacción LR con p0830, se usó un vector de destinación para la transformación de *Oryza sativa*. Este vector contenía como elementos funcionales dentro de los bordes de T-ADN: un marcador seleccionable vegetal; un casete de expresión de marcador visual; y un casete Gateway destinado a recombinación LR *in vivo* con la secuencia de interés ya clonada en el clon de entrada. Un promotor de prolamina para la expresión preferida en semillas (PRO0090) estaba en el sentido de 5' de este casete Gateway. Tras la etapa de recombinación LR, se transformó el vector de expresión resultante p72 (figura 4) en la cepa de *Agrobacterium* LBA4404 y posteriormente en plantas de *Oryza sativa*.

Ejemplo 3: Transformación de arroz con el constructo PRO0090-CDS0669

Se descascarillaron semillas secas maduras de *Oryza sativa* japónica cultivar Nipponbare. Se realizó la esterilización incubando las semillas durante un minuto en etanol al 70%, seguido por 30 minutos en HgCl₂ al 0,2% y mediante 6 lavados de 15 minutos con agua destilada estéril. Entonces se hicieron germinar las semillas estériles en un medio que contenía 2,4-D (medio de inducción de callos). Tras una incubación de 4 semanas en la oscuridad, se escindieron callos derivados del escutelo, embriogénicos y se propagaron en el mismo medio. Dos semanas después, se multiplicaron los callos o se propagaron mediante subcultivo en el mismo medio durante otras 2 semanas. 3 días antes del cocultivo, se subcultivaron trozos de callos embriogénicos en medio fresco para estimular la actividad de división celular. Se usó la cepa de *Agrobacterium* LBA4414 que albergaba el vector binario p72 para el cocultivo. Se cultivó la cepa de *Agrobacterium* durante 3 días a 28°C en medio AB con los antibióticos apropiados. Se recogieron entonces las bacterias y se suspendieron en medio de cocultivo líquido a una DO₆₀₀ de aproximadamente 1. Se transfirió la suspensión a una placa de Petri y se sumergieron los callos en la suspensión durante 15 minutos. A continuación, se secaron los tejidos callosos en un papel de filtro, se transfirieron a un medio

de cocultivo solidificado y se incubaron durante 3 días en la oscuridad a 25°C. Después de eso, se hicieron crecer callos cocultivados en medio que contenía 2,4-D durante 4 semanas en la oscuridad a 28°C en presencia de un agente selectivo a una concentración adecuada. Durante este periodo, se desarrollaron islas callosas resistentes en crecimiento rápido. Tras la transferencia de este material a un medio de regeneración y la incubación en la luz, se liberó el potencial embriogénico y se desarrollaron brotes en las siguientes cuatro a cinco semanas. Se escindieron los brotes de los callos y se incubaron durante de 2 a 3 semanas en un medio que contenía auxina del que se transfirieron al suelo. Se hicieron crecer los brotes endurecidos bajo alta humedad y días cortos en un invernadero. Finalmente, se cosecharon las semillas de tres a cinco meses tras el trasplante. El método produjo transformantes de locus único a una tasa superior al 50% (Aldemita y Hodges, Plant 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).

Ejemplo 4: Evaluación de arroz transgénico transformado con el constructo PRO0090-CDS0669

Se generaron de aproximadamente 15 a 20 transformantes de arroz T0 independientes. Se transfirieron los transformantes primarios de las cámaras de cultivo de tejido a un invernadero para el crecimiento y la cosecha de semilla T1. Se conservaron 6 acontecimientos, de los que la progenie T1 segregó 3:1 para presencia/ausencia del transgén. Para cada uno de estos acontecimientos, se seleccionaron aproximadamente 10 plántulas T1 que contenían el transgén (hetero y homocigotas), y aproximadamente 10 plántulas T1 que carecían del transgén (nulicigotas), monitorizando la expresión del marcador visual. Se evaluaron varios parámetros relacionados con el crecimiento vegetativo y la producción de semillas y se analizaron estadísticamente todos los datos tal como se explica a continuación:

20 Análisis estadístico: prueba de la t y prueba F:

Se usó un ANOVA de dos factores (análisis de variantes) como modelo estadístico para la evaluación global de las características fenotípicas de la planta. Se llevó a cabo una prueba F sobre todos los parámetros medidos de todas las plantas de todos los acontecimientos transformados con el gen de la presente invención. Se lleva a cabo la prueba F para comprobar el efecto del gen sobre todos los acontecimientos de transformación y para verificar el efecto global del gen, también denominado en el presente documento "efecto génico global". Si el valor de la prueba F muestra que los datos son significativos, entonces se concluye que existe un efecto "génico", lo que significa que no sólo la presencia o la posición del gen provoca las diferencias en el fenotipo. El umbral para la significación para un efecto génico verdadero global se fija al nivel de probabilidad del 5% para la prueba F.

4.1 Mediciones del crecimiento vegetativo:

30 Se transfirieron las plantas T1 seleccionadas (aproximadamente 10 con el transgén y aproximadamente 10 sin el transgén) a un invernadero. Cada planta recibió una etiqueta de código de barras única para vincular inequívocamente los datos de fenotipado con la planta correspondiente. Se hicieron crecer en el suelo plantas T1 seleccionadas en macetas de 10 cm de diámetro bajo los siguientes parámetros medioambientales: fotoperiodo = 11,5 h, intensidad de luz diurna = 30.000 lux o más, temperatura diurna = 28°C o superior, temperatura nocturna = 22°C, humedad relativa = 60-70%. Se hicieron crecer las plantas transgénicas y los nulicigotas correspondientes lado a lado en posiciones aleatorias. Desde la fase de siembra hasta la fase de madurez cada planta se pasó varias veces a través de una cabina de obtención de imágenes digitales y se obtuvieron imágenes. En cada punto de tiempo, se tomaron imágenes digitales (2048x1536 píxeles, 16 millones de colores) de cada planta desde al menos 6 ángulos diferentes. Pueden deducirse varios parámetros de una forma automatizada a partir de todas las imágenes digitales de todas las plantas, usando software de análisis de imagen.

4.2 Mediciones de parámetros relacionados con las semillas:

Se cosecharon las panículas primarias maduras, se guardaron en bolsas, se marcaron con códigos de barras y entonces se secaron durante tres días en el horno a 37°C. Entonces se trillaron las panículas y se recogieron todas las semillas y se contaron. Se separaron las cáscaras llenas de las vacías usando un dispositivo de soplado de aire. Se desecharon las cáscaras vacías y se contó de nuevo la fracción restante. Se pesaron las cáscaras llenas en una balanza analítica. Este procedimiento permite deducir un conjunto de parámetros relacionados con las semillas.

Índice de cosecha de plantas

El índice de cosecha en la presente invención se define como la razón entre la producción de semillas total y el área por encima del suelo (mm^2), multiplicado por un factor 10^6 . Se midió la producción de semillas total por planta pesando todas las cáscaras llenas cosechadas de una planta tal como se describió anteriormente. Se determinó el área por encima del suelo de la planta contando el número total de píxeles de las imágenes digitales de partes de la planta por encima del suelo diferenciadas del fondo. Se promedió este valor para las fotografías tomadas en el mismo punto de tiempo desde los diferentes ángulos y se convirtió en un valor de superficie expresado en mm^2 mediante calibración. Los experimentos mostraron que el área de la planta por encima del suelo medida

de esta manera se correlaciona con la biomasa de las partes de la planta por encima del suelo.

5 Se confirmaron los datos obtenidos en el primer experimento en un segundo experimento con plantas T2. Se seleccionaron tres líneas que tenían el patrón de expresión correcto para análisis adicionales. Se seleccionaron lotes de semillas de las plantas positivas (tanto hetero como homocigotas) en T1, monitorizando la expresión del marcador. Para cada acontecimiento escogido, se conservaron entonces los lotes de semillas heterocigotas para la evaluación T2. En cada lote de semillas se hizo crecer un número igual de plantas positivas y negativas en el invernadero para la evaluación.

Se evaluó un número total de 120 plantas transformadas con *GRUBX* en la generación T2, es decir, 40 plantas por acontecimiento de las cuales 20 eran positivas para el transgén y 20 negativas.

10 Debido a que se han llevado acabo dos experimentos con acontecimientos solapantes, se realizó un análisis combinado. Esto es útil para comprobar la consistencia de los efectos a lo largo de los dos experimentos, y si este es el caso, para acumular pruebas a partir de ambos experimentos con el fin de aumentar la confianza de la conclusión. El método usado era una aproximación de modelo mixto que tiene en cuenta la estructura multinivel de los datos (es decir experimento – acontecimiento - segregantes). Se obtuvieron valores de P comparando pruebas de razón de probabilidad con distribuciones de chi cuadrado.

15

En un primer experimento, se evaluaron seis líneas en la generación T1. Hubo un aumento promedio del índice de cosecha y dos líneas tuvieron un aumento significativo del 50% o más en comparación con las líneas nulicigotas (Tabla 2).

Tabla 2: Evaluación de los dos acontecimientos T1 con mejor rendimiento

Índice de cosecha :					
Línea	TR	nul.	dif.	% de dif.	Valor de p
10	74,9	49,9	24,97	50	0,039
4	35	21,7	13,28	61	0,0656

20 Se facilitan valores absolutos medios de las mediciones del índice de cosecha para las líneas transgénicas (TR) y plantas de control (nul.) en la generación T1 en las columnas 2 y 3, la diferencia absoluta en la columna 4 y la diferencia en % en la columna 5, la significación, expresada como valor de p obtenido en una prueba de la t, se facilita en la columna 6.

25 Se confirmaron los resultados obtenidos para la generación T1 en la generación T2; el aumento promedio para el índice de cosecha era del 13% y una prueba F mostró que este aumento era significativo (valor de p de 0,0447). Además, se volvieron a evaluar estos datos de T2 en un análisis combinado con los resultados para la generación T1, y el valor de p obtenido a partir de una prueba F mostró de nuevo que los efectos observados eran significativos (valor de p 0,0181).

30

LISTA DE SECUENCIAS

<110> CropDesign N.V.

<120> Plantas que tienen características de crecimiento modificadas y método para preparar las mismas

<130> CD-107-PCT

5 <150> documento EP 03104764.0

<151> 17-12-2003

<150> documento US 60/531.866

<151> 22-12-2003

<160> 7

10 <170> PatentIn versión 3.3

<210> 1

<211> 1380

<212> ADN

<213> *Nicotiana tabacum*

15 <400> 1

```

atgggtgaca tgaaggataa agtcaaaggg ttcattgaaa aagtcacatc ttctttcttca      60
ggtaagttta aaggccaagg taggggttttg ggtgggttcat cttcttcagg accctcaaat      120
catgtcaata atttttcatc acatccccta aatacaaggc aagatcaaca accttcatat      180
acaaaaactt cgcctcaaaa accaagtaat tctgatcaaa gaattgagaa tataatgtgaa      240
attcagttca acaaaaagtga atcaaaggat ggttttgatc catttggtga attagtcact      300
tctgggaaga gaaacccaaa agggatttca cttactaatg tgtttgaatg ccctgtctgt      360
ggtagtgggt ttgtttctga agaagaggtg tcaactcata ttgatagctg tttaaagttct      420
gaagtgtcct ctaatttggg agttgaaagt aaagttgaag ttaaaagtga attggaaca      480
tgtgttagtg catatgtttc agggaagccc tcagaagggt cagttgaagt ggtcattaag      540
ttgttaaaga atattgtgaa ggaaccagag aatgccaagt ttaggaaaat aaggatgggg      600
aatccaaaaa taaaaggtgc tataggtgat gttgtaggag gagtggagct attggaattt      660
gttggtttg agttgaaaga agaaggtggg gaaatttggg ctgtgatgga tgttccttct      720
gaagaacaac ttgttatgct taagaatgta gtttactctc tggaaaccgaa gaaggttgaa      780
gagttggcgt ccttatccca agttaaggcg agtgaaccag ttgagccgaa gaagattgat      840
agacagattc gagtgttctt ttctgttccc gagagcgtag cagcaaaaat tgagctacct      900
gattccttct ttaacctctc acgtgaggaa ttgagaagag aagcagagat gaggaagaag      960
aaattagaag attccaaatt attgattcct aaatcttatc gggaaaagca ggcaaaagct     1020
gcaagaaaga agtacacaaa atccattatc cgtgtacagt ttccagatgg agcattgctt     1080
caaggtgtct ttctaccttc ggagccaact agtgcctttt atgagtttgt gagcgcagcg     1140
ttaaaggaac caagcttaga gttcgaattg ttacatccgg tgcttgtaa aaagcgggtg     1200
attccccatt ttccagctgc tggggagagg gctgtaacag ttgaagagga ggatttggtt     1260
cctgcagctc tactcaaatt taaacctatc gaaacagatt ctggtgtttt tactggctct     1320
tgtaatgagc ttcttgaaat tagcgagccc ctcgagaccg gatcagttgc ttctctgtaa     1380

```

<210> 2

<211> 459

<212> PRT

20 <213> *Nicotiana tabacum*

<400> 2

```

Met Gly Asp Met Lys Asp Lys Val Lys Gly Phe Met Lys Lys Val Thr
1           5           10           15
Ser Ser Ser Ser Gly Lys Phe Lys Gly Gln Gly Arg Val Leu Gly Gly

```


ES 2 372 938 T3

Gln Ala Lys Ala Ala Arg Lys Lys Tyr Thr Lys Ser Ile Ile Arg Val
 340 345 350

Gln Phe Pro Asp Gly Ala Leu Leu Gln Gly Val Phe Leu Pro Ser Glu
 355 360 365

Pro Thr Ser Ala Leu Tyr Glu Phe Val Ser Ala Ala Leu Lys Glu Pro
 370 375 380

Ser Leu Glu Phe Glu Leu Leu His Pro Val Leu Val Lys Lys Arg Val
 385 390 395 400

Ile Pro His Phe Pro Ala Ala Gly Glu Arg Ala Val Thr Val Glu Glu
 405 410 415

Glu Asp Leu Val Pro Ala Ala Leu Leu Lys Phe Lys Pro Ile Glu Thr
 420 425 430

Asp Ser Val Val Phe Thr Gly Leu Cys Asn Glu Leu Leu Glu Ile Ser
 435 440 445

Glu Pro Leu Glu Thr Gly Ser Val Ala Ser Ser
 450 455

<210> 3

<211> 1311

<212> ADN

5 <213> *Saccharum officinarum*

<220>

<221> misc_feature

<222> (277)..(279)

<223> n puede ser cualquier nucleótido

10 <400> 3

ES 2 372 938 T3

```

atgatgaag acaagatgaa ggagttcatg aagaaggcca cctcctccgg gtccgggacc      60
ccctcctcct tcaagggcac ctcccacgtc ctcggtccg gccctcccc ctctcctcc      120
caccgccgtg cccgctcctc aaaccctagc ccaaacctca ggcccgtcc taagcggacc      180
tcgccacctc ccccgcccac ttaaccacc gatttgacct ccttcacgcc cctcgtctgc      240
tactcctccc gccgcccga cgcgaacggc accgcgnnng ccgtcgccac cgtcgcgtgc      300
cccagctgcg gagacgcgtt tccgtccgag ctgcgcgtct ccgagcatct cgacggctgc      360
ctcgcgtcgg cggggggcgc ccgcgcgcgc gccgcgcggt acctcgcgcg cgacccgcct      420
ccgcccgcgg cctccgtaga ggtggtcaaa cgctcgtcgg gcaacctgct ccgggagccc      480
ggcaacgata agttcaggcg ggtgagattg ggtaaccgcg ggatcaagga ggccctggca      540
gacagggatg gcggggtgga gtcctcggag gccgtcggct tcacagttgg ggatgagggc      600
ggggagccct tcgccgtgat ggacgaagtg cctagcgcacc ctaggctcaa cgggatcagg      660
agggccgtcc tctgctcga gggggcacac ccctctgcgc ctccagttaa ggccggaggct      720
gaggccaagg agagctgcag caatgtgtct gacgtgcagg aggggtgtaa gactattgat      780
cggcagattc gggatattgt ctctgttctt gggagttcta tggcacaata tgatgtacca      840
gattcttttt acaagcttag tggtagaggc ataaggaatg aagcaaagat gaggagggaa      900
aggttagaac aatctcgatt gctgatacca aagtcttaca aggagaaaca ggcattggct      960
gctcgacaga agtataaaca agcagtcatt cgagttcagt ttccagatag aatgattctt     1020
cagggcatat tcctaccagg agaggccact agttcactgt atgagttcgt cacatctgct     1080
ctgaagcaat caggtttgga attcgaactt atctctccag ccatacctaa gccacgtgtg     1140
gtgccccatt ttccaaaccg gggagagcgg gcacgcacct tgcaagagga ggagctggtc     1200
ccatctgcgc tcctcaagtt cattcccaag gagactgatt ccatggtttt caccggtttg     1260
cttgatgagc ttctcatggc cagtgagccg cttcctcgtg catcacaatg a             1311

```

<210> 4

<211> 436

<212> PRT

5 <213> *Saccharum officinarum*

<220>

<221> MISC_FEATURE

<222> (93)..(93)

<223> Xaa puede ser cualquier aminoácido

10 <400> 4

ES 2 372 938 T3

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

ES 2 372 938 T3

cttctacatc	ggcttaggtg	tagcaacacg	actttattat	tattattatt	attattatta	60
ttattttaca	aaaatataaa	atagatcagt	ccctcaccac	aagtagagca	agttgggtgag	120
ttattgtaaa	gttctacaaa	gctaatttaa	aagttattgc	attaacttat	ttcatattac	180
aaacaagagt	gtcaatggaa	caatgaaaac	catatgacat	actataatTT	tgTTTTtatt	240
attgaaatta	tataattcaa	agagaataaa	tccacatagc	cgtaaagtTC	tacatgtggt	300
gcattaccaa	aatatatata	gcttacaaaa	catgacaagc	ttagtttgaa	aaattgcaat	360
ccttatcaca	ttgacacata	aagtgagtga	tgagtcataa	tattattttc	tttgctaccc	420
atcatgtata	tatgatagcc	acaaagtTtC	tttgatgatg	atatcaaaga	acatttttag	480
gtgcacctaa	cagaatatcc	aaataatatg	actcacttag	atcataatag	agcatcaagt	540
aaaactaaca	ctctaaagca	accgatggga	aagcatctat	aaatagacaa	gcacaatgaa	600
aatcctcatc	atccttcacc	acaattcaaa	tattatagtt	gaagcatagt	agtaatttaa	660
atcaactagg	gatatcacia	gtttgtacaa	aaaagcaggc	tggtaccggT	ccggaattcc	720
cgggatatcg	tcgacccacg	cgTccgccaa	tatcagattt	ctttcatgaa	ctccacttcc	780

aatttctcat	tgcttcttct	tcccatttcc	acctccaaag	ccatccttcc	agaaaacctt	840
gttccttaca	tttcttagcc	ccaaaaaaga	ttcccatctc	aattccacaa	aaaaacacaa	900
ggagatctaa	ggaaattccc	cgctctata	tatagagagg	tggaattggt	cctgaatttg	960
gtttgaattg	attgattgac	agattttggT	gagaggggtg	tattgaaaaa	atgggtgaca	1020
tgaaggataa	agtcaaaggg	ttcatgaaaa	aagtcacatc	ttcttcttca	ggtaagtTta	1080
aaggccaagg	tagggTTTTg	ggTggTtcat	cttcttcagg	accctcaaat	catgtcaata	1140
atTTTTcatc	acatccccta	aatacaaggc	aagatcaaca	accttcatat	acaaaaactt	1200
cgctcaaaa	accaagtaat	tctgatcaaa	gaattgagaa	tatatgtgaa	attcagTtca	1260
acaaaagtga	atcaaaggat	ggTTTTgatc	catttggtga	attagTcact	tctgggaaga	1320
gaaacccaaa	agggtattca	cttactaatg	tgTttgaaTg	ccctgtctgt	ggtagTggtt	1380
ttgTttctga	agaagaggTg	tcaactcata	ttgatagctg	tttaagTtct	gaagTgtctt	1440
ctaattTggg	agTtgaaagt	aaagTtgaag	ttaaaagTga	attggaaaca	tgtgttagTg	1500
catatgtTtc	agggaagccc	tcagaagggt	cagTtgaagt	ggTcattaag	ttgtTaaaga	1560
atattgtgaa	ggaaccagag	aatgccaagt	ttaggaaaat	aaggatgggg	aatccaaaaa	1620
taaaaggTgc	tataggtgat	gtTgtaggag	gagTggagct	attggaattt	gtTggattTg	1680
agTtgaaaga	agaaggTggg	gaaattTggg	ctgtgatgga	Tgttccttct	gaagaacaac	1740
ttgttatgct	taagaatgta	gtTtTactct	Tggaaccgaa	gaaggTtgaa	gagTtggcgt	1800
ccttatccca	agTtaaggcg	agtgaaccag	ttgagccgaa	gaagattgat	agacagattc	1860
gagTgttctt	ttctgttccc	gagagcgtag	cagcaaaaat	Tgagctacct	gattccttct	1920
Ttaacctctc	acgtgaggaa	ttgagaagag	aagcagagat	gaggaagaag	aaattagaag	1980
attccaaatt	attgattcct	aaatcttatac	gggaaaagca	ggcaaaaagct	gcaagaaaga	2040
agtacaaaa	atccattatc	cgtgtacagt	ttccagatgg	agcattgctt	caaggTgtct	2100
ttctaccttc	ggagccaact	agtgtctttt	atgagTttgt	gagcgcagcg	Ttaaagggaac	2160
caagcttaga	gttcgaattg	ttacatccgg	TgcttgTtaa	aaagcgggtg	attccccatt	2220
ttccagctgc	Tggggagagg	gctgtaacag	ttgaagagga	ggattTgggt	cctgcagctc	2280
tactcaaat	Taaacctatc	gaaacagatt	ctgtTgtTtt	tactgtTctt	Tgtaatgagc	2340
ttctTgaaat	tagcgagccc	ctcgagaccg	gatcagTtgc	ttcctcgtaa	gctctaaatt	2400
acatcagact	ttgaattctt	ctgagTgtTg	gaaaccttat	aaaactctct	gcgcccgggaa	2460
Tgctgcggcc	gctctagagt	atccctcgag	gggcccagc	ttacgcgtac	ccagctTtct	2520
TgtacaaaagT	ggTgatatca	caagcccggg	cggTcttcta	gggataacag	ggtaattata	2580
Tccctctaga	Tcacaagccc	gggcggtctt	ctacgatgat	Tgagtaataa	Tgtgtcacgc	2640
atcaccatgg	gtggcagTgt	cagTgtgagc	aatgacctga	atgaacaatt	gaaatgaaaa	2700
gaaaaaaagT	actccatctg	Ttccaaatta	aaattcattt	TaacctTtTa	ataggtTtTa	2760
acaataattg	atataTgtTt	TctgtatagT	TctaattTgt	Tatcatccgg	gcggTcttct	2820
agggataaca	gggtaattat	atccctctag	acaacacaca	acaaataaga	gaaaaaacia	2880
ataatattaa	tttgagaatg	aacaaaagga	ccatatcatt	cattaactct	Tctccatcca	2940
tttccatttc	acagTtctgat	agcgaaaacc	gaataaaaaa	cacagTaaat	Tacaagcaca	3000
acaaatggta	caagaaaaac	agTtttccca	atgccataat	actcgaac		3048

<210> 6

<211> 1302

5 <212> ADN

<213> *Oryza sativa*

ES 2 372 938 T3

<400> 6

atgatgaagg	aaaagatgaa	ggatctcatg	aggaaggtca	cctcctcctc	ctcctcctcc	60
tcgtcgtcct	cctccttcaa	gggcaccgcc	cacgtcctcg	gctccggccc	cgaccctcc	120
tcccgccct	ccaaccctac	ccctagtcgc	cccgtgccc	cccggcgaga	ggccgcccgc	180
tccgcgaggc	cgccctcctc	cggcttcgcc	ccctactccc	cgctcatctc	cacctcctcc	240
cgccgcaccg	accaccccg	gggggcgggg	gcgggggagg	acgacgcggt	cgcggtcccc	300
agctgcgccc	agccgttccc	ctccgagctg	gcggtgtcgg	accacctcga	cggtgcctc	360
gcggcggcgg	ggggagcccc	cccccgcgcg	gccgcctacc	tggccggcga	cccccccgcg	420
tccgccgtgg	aggtggtgaa	gaggctgctc	gggaacctgc	tctccgacce	ccggaacgac	480
aagtacagga	aggtcaggct	cgggaacctg	aggatcaagg	aggccctggc	ggacagggag	540
ggcgggggtg	atctcctcga	ggccgtgggg	ttcagggtcg	ccgacgaggg	cggggagctc	600
ttcgccctca	tggacgaggt	gcccggggag	gcgaggtcgc	gcggcatcag	gcagggcgtg	660
ctcctgctcg	agagggcccc	gccatcgacg	ccgccgcaga	cacaggcaga	tgccaaagag	720
acttgcccga	atggagttag	cgaagagcag	gggattaaga	agccggttga	tcgtcagatt	780
cgggtgttct	tctctgttgc	tgcaagttct	gttgacgaaa	atgatctacc	agattctttc	840

tatagcctta	gtaatgagga	gatcaggaat	gaggcaaaga	tgaggagggg	gaggctagaa	900
caatctcggg	tggttgattcc	aaagtcatat	aaggagaagc	aggcactggc	tgcccgacag	960
aagtataaac	aagctctgat	tcgaattcag	tttccggatg	gagtaattct	gcaggggtgtg	1020
ttccttcccg	cggagcccat	tagttcacta	tatgagtttg	tcgcatcttc	tttgaagcag	1080
ccaagtttgg	aattcgatct	tatctgtcca	gcggttccta	ggacacgcgt	catacctcct	1140
tttccaaaac	caggggaaca	ggcacgcaca	ctgcccggatg	aagacctagt	cccttctgct	1200
cggctcacct	tcaagcccaa	agagactgat	tcggttgtgt	tcaactggtct	gttggacgag	1260
cttcttgaga	ccagtgagcc	gttcacatct	gcacctcat	ga		1302

<210> 7

5 <211> 433

<212> PRT

<213> *Oryza sativa*

<400> 7

ES 2 372 938 T3

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

ES 2 372 938 T3

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

REIVINDICACIONES

1. Método para mejorar la producción de semillas de una planta con respecto a plantas de tipo natural correspondientes, comprendiendo dicho método:

5 - introducir y sobreexpresar en una planta o célula vegetal una secuencia de ácido nucleico aislada que codifica para una proteína GRUBX, seleccionándose dicha secuencia de ácido nucleico del grupo:

- una secuencia de ácido nucleico tal como se representa por una cualquiera de SEQ ID NO:1, 3, 6;

- un ácido nucleico que codifica para un polipéptido tal como se representa por una cualquiera de SEQ ID NO: 2, 4, 7;

10 - un ácido nucleico que codifica para un polipéptido que tiene al menos el 80% de identidad de secuencia con una cualquiera de SEQ ID NO: 2, 4, 7, teniendo dicho polipéptido acción reguladora de ubiquitina dentro de la ruta de ubiquitinación

- cultivar la planta o célula vegetal en condiciones que promueven el crecimiento de la planta.

15 2. Método según la reivindicación 1, en el que dicho ácido nucleico se deriva de un organismo eucariota, preferiblemente de una planta.

3. Método según la reivindicación 2, en el que dicho ácido nucleico se deriva de una planta dicotiledónea, preferiblemente de la familia Solanaceae, más preferiblemente de *Nicotiana tabacum*.

4. Método según la reivindicación 2, en el que dicho ácido nucleico se deriva de una planta monocotiledónea, preferiblemente de la familia Poaceae, más preferiblemente de *Oryza sativa*.

20 5. Método según cualquiera de las reivindicaciones 1 a 4, en el que la expresión de dicho ácido nucleico que codifica para una proteína GRUBX esta dirigida por un promotor preferido en semillas, preferiblemente un promotor de prolamina.

25 6. Método según cualquiera de las reivindicaciones 1 a 5, en el que dicho aumento de la producción de semillas comprende uno o más de (i) aumento de la biomasa de semillas, (ii) aumento del número total de semillas, (iii) aumento del número de semillas llenas, (iv) aumento del tamaño de semilla, (v) aumento del volumen de semilla, (vi) aumento del índice de cosecha, y (vii) aumento del peso de mil granos, todo con respecto a plantas de tipo natural correspondientes.

7. Uso de una secuencia de ácido nucleico aislada que codifica para una proteína GRUBX, seleccionándose dicha secuencia de ácido nucleico del grupo:

30 - una secuencia de ácido nucleico tal como se representa por una cualquiera de SEQ ID NO:1, 3, 6;

- un ácido nucleico que codifica para un polipéptido tal como se representa por una cualquiera de SEQ ID NO: 2, 4, 7;

35 - un ácido nucleico que codifica para un polipéptido que tiene al menos el 80% de identidad de secuencia con respecto a una cualquiera de SEQ ID NO: 2, 4, 7, teniendo dicho polipéptido acción reguladora de ubiquitina dentro de la ruta de ubiquitinación, en el aumento de la producción de semillas de una planta con respecto a plantas de tipo natural correspondientes.

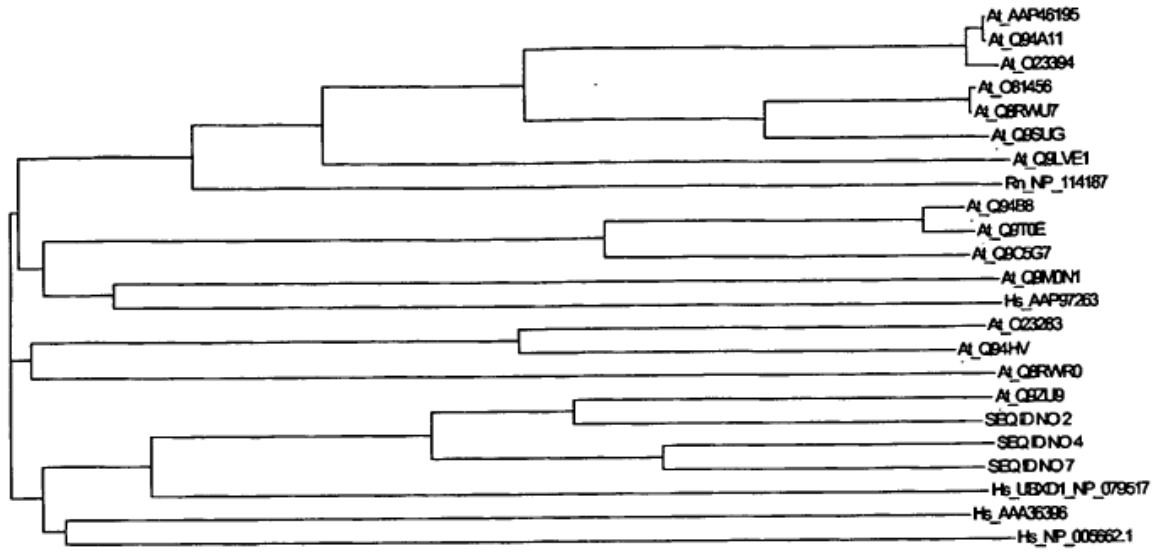


FIGURA 1 A

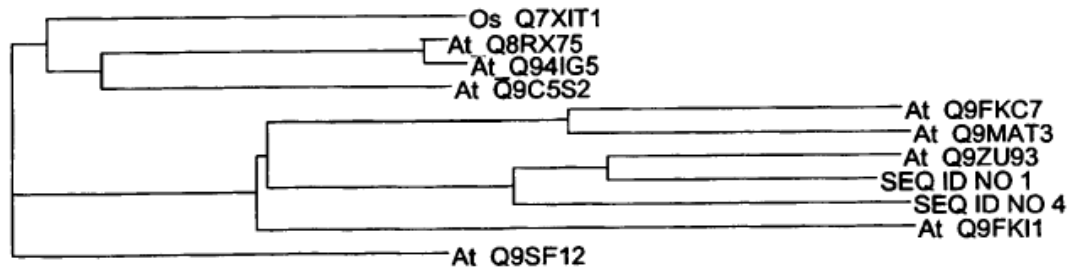


FIGURA 1 B

Términos de consenso para el dominio UBX

```

YJE8_YEAST      KGFLAQ-NYCTLQLKLPNGYTISNTFP-PQTKLHKVRMWL----DYNCYDDGTP-----
  consenso /80%  s.t..t...splplRhssGpph.ppF..spp.lptlhthl.....s.sss.....
  consenso /65%  stpptt.stsplQlRhssGpphsppFs.uppslpplhphl....t...s.ssss.....
  consenso /50%  Pppps.sps+LQlRhPDGpplscCFs.us-sLcplapaV....sptpssusss.....

YJE8_YEAST      -----YLFHRNIPRVTLTRNDELKSLQEL-DLLPRSTLILEP
  consenso /80%  .....apl.pshPch.hspt-.hph.p.thhspsslll..
  consenso /65%  .....FpL.pshP+c.hsps-.ppl.ch.tlhPsuslllt.
  consenso /50%  .....FpLpssaPR+slocsD..KoLp-l.sLhPsuoLLlct

AAF74723        EASKLLTYADNILRNPSDEKYRSIRIGNTAFSTRLLPVRGAVECLFEMGFEEG-----
  consenso /80%  tslt.LLphhcNhhpp.p-.....s..hphlh.h.tsh..h...tF.ph.....
  consenso /65%  psllpsLL+hl+Nhhcc.pE.....stthpctltpl.sshphhhtFscI.h....
  consenso /50%  solccLLRhLRnlh+HhcEt.....tsppl+EplGslP-uhcpaFsssFPcLhh....

AAF74723        --ETHLIFPK
  consenso /80%  ...phh.h.p
  consenso /65%  ...phathhp
  consenso /50%  ...csYplss
  
```

FIGURA 2 A

Dominio UBX presente en SEQ ID N0 2

339 KAARKKYTKSIIRVQFPDQALLQGVFLPSEPTSALYEFVSAALKEPSLEFELLHPVLVKKRVI PHF
 PAAGERAVTVEEEDLVPAALLKFK 428

Dominio PUG presente en SEQ ID N0 2 :

173 GSVEVVIKLLKNIVKEPENAKFRKIRMGNPKIKAIGDVVGGVELLEFVGFELKEEGGEIWA VMDV
 PS 240

Dominio UBX presente en Q9ZU93

345 KAARKRYKRSMIRVQFPDGVVLQGVFAPWEPTFALYEFVSSALKEPSLQFELLDPVLVKKRVI PHF
 PAPGQKPITLEDDELVPSALIRFRP 435

Dominio PUG presente en Q9ZU93

181 SSIDVLLRLFKNIVKEPENAKFRKVRMSNAKIKAIGDVAGGVELLELVGFELKEENDEIWA VMDV
 PS 248

FIGURA 2 B

ES 2 372 938 T3

Q9ZU93	1	MDDVKDKLKGFMKKVNLSSSSGKFKGQGRVLGSSSSSSAPPVNPIONRFN	50
SEQ_ID_NO_2	1	MGMKDKVKGFMKKVT-SSSSGKFKGQGRVLGSSSSSG--PSNHV-NNFS	46
Q9ZU93	51	S-----SQAPNPTPRPKPNPNPLPEKPLSSSDQKISGSTR---NPDHD	90
SEQ_ID_NO_2	47	SHPLNTRQDQQPSYT-----KTSPOKQK-SENSDQRIENICEIQFNKSES	88
Q9ZU93	91	PVRAPQDGFDPYGAFITSSNRSONGYSL-SMFECPICKNPFTEEEVSVH	139
SEQ_ID_NO_2	89	-----KDGFDPFGLVTSGRNPKGYSLTNVFECPCVCGSGFVSEEEVSTH	133
Q9ZU93	140	VESCLG---DTN-GDESSFEDKNDNDKSEMEKLVVVYLSGKPSSESSIDV	185
SEQ_ID_NO_2	134	IDSCLSSSEVSSNLGVESKVEV-----KSELETCVSAYVSGKPSSEGSVEV	177
Q9ZU93	186	<u>LLRLFKNIVKEPENAKFRKVRMSNAKIKEAIGDVAGGVELLELVGFELKE</u>	235
SEQ_ID_NO_2	178	<u>VIKLLKNIVKEPENAKFRKIRMGNPKIKGAIGDVVGGVELLEFVGFELKE</u>	227
Q9ZU93	236	<u>ENDEIWAVMDVPSEEQSILINKVVGYLEKRKTESGSSAQV--MEPVAPK</u>	283
SEQ_ID_NO_2	228	<u>EGGEIWAVMDVPSEEQLVMLKNVVSLLLEPKKVEELASLSQVKASEPVEPK</u>	277
Q9ZU93	284	KIDREIRVFFSVSENVASRIEVPDSFYLSLDAEIKREADLRRKIAESQL	333
SEQ_ID_NO_2	278	KIDRQIRVFFSVPEVSAAKIELPDSFFNLSREELRREAEMRKKKLEDSKL	327
Q9ZU93	334	<u>LIPRSYKEKQAKAARKRYKRSMIRVQFPDGVVLQGVFAPWEPTFALYEFV</u>	383
SEQ_ID_NO_2	328	<u>LIPKSYREKQAKAARKRYTKSIIRVQFPDGALLQGVFLPSEPTFALYEFV</u>	377
Q9ZU93	384	<u>SSALKEPSLQFELLDPVLVKRRVI PHTPAPGQKPTLEDEELVPSALIRF</u>	433
SEQ_ID_NO_2	378	<u>SAALKEPSLEFELLHPVLVKKRVI PHFPAAGERAVTVEEEDLVPAALLKF</u>	427
Q9ZU93	434	RPIETDSLVTGLRNELLEISEPLS	458
SEQ_ID_NO_2	428	KPIETDSVVFTGLCNELLEISEPLETGSVASS	459

FIGURA 2 C

SEQ_ID_NO2	1	MGDMKDKVKGFMMKVTSSSSG---KFKGQGRVLGGSSSSGSPSNHVNPFSS	47
		. .	
SEQ_ID_NO4_C	1	mmkdkmkefmkkvtsssgsgtpssfkgtshvlg----sgpsps----ss	40
SEQ_ID_NO2	48	HPL----NTRQDQQPSYTKTSPQKPSNSDQRIENICEIQFNKSESKDGF	93
	: : :: ...: ...: ...: ...: ...: ...:	
SEQ_ID_NO4_C	41	hpaarssnpsnlrpapkrtspptltdlts-----	74
SEQ_ID_NO2	94	PFGELVTSGKRNPKGYS---LTNVFECPCVCGSGFVSEEEVSTHIDSC	140
		
SEQ_ID_NO4_C	75	-ftplvcyssrrpdangtaCavatvacpscgdafpselavsehldgclas	123
SEQ_ID_NO2	141	EVSSNLGVESKVEVKSELETCVSAYVSG--KPSEGSVEVVIKLLKNIVKE	188
	: ...: ...: ...: ...: ...: ...: ...: ...: ...:	
SEQ_ID_NO4_C	124	ag----garar-----aaaylaadppppaasvevvkrlilgnllre	159
SEQ_ID_NO2	189	<u>PENAKFRKIRMGNPKIKGAIGDVVGGVELLEFVGFELKEEGGEIWA</u>	237
		<u>MD- . . . : </u>	
SEQ_ID_NO4_C	160	<u>pgndkfrvrlnprikealadrddgvelleavgftvgdeggepfavmde</u>	209
SEQ_ID_NO2	238	VPSEEQLVMLKNVVSLE-----PKKVEELASLSQVKASEPVE-PKKI	279
		: :	
SEQ_ID_NO4_C	210	vpsdprlngirravlllegahpsappvkaeaeakescsnsvdvgqegakti	259
SEQ_ID_NO2	280	DRQIRVFFSVSPESVAAKIELPDSFFNLSREELRREAEMRKKKLEDSKLLI	329
		: :	
SEQ_ID_NO4_C	260	drqirvfvsvpgssmaqndvpdsfyklsgeeirneakmrrerleqsrlli	309
SEQ_ID_NO2	330	PKSYREKQAKAARKKYTKSIIRVQFPDGALLQGVFLPSEPTSALYEFVSA	379
		: :	
SEQ_ID_NO4_C	310	pksykekqalaarqkyqavirvqfdrmilqgflpgeatsslyefvts	359
SEQ_ID_NO2	380	ALKEPSLEFELLHFVLVKKRVIPIPFPAAGERAVTVEEEDLVPAALLKFKP	429
		: :	
SEQ_ID_NO4_C	360	alkqsglefelispaipkprvvphfpnngerartlqeeelvpsallkfp	409
SEQ_ID_NO2	430	IETDSVVFTGLCNELLEISEPLETGSVASS 459	
		. .	
SEQ_ID_NO4_C	410	ketdsmvftgllldellmaseplpaasq 436	

FIGURA 2 D

ES 2 372 938 T3

SEQ_ID_NO_4	1 mmkdkmkefmkkvt--ssgsqtpssfkgtshvlsgspssshpaarssn	48
SEQ_ID_NO_7	1 MMKEKMKDLMRKVTSSSSSSSSSSSFKGTAHVLSGSPDPSS-----RPSN	45
SEQ_ID_NO_4	49 pspnlrpapkrtsptptltdltsftplvcyssrr--pdangtaCava	96
SEQ_ID_NO_7	46 PTPSRPAAPRREAAAASARPPSSGFAPYSPLISTSSRRTDPPAGAGAGEDD	95
SEQ_ID_NO_4	97 tvacpscgdafpselavsehldgclasaggararaaaylaadppppaasv	146
SEQ_ID_NO_7	96 AVACPSCAEPPFSELAVSDHLDGCLAAAGGARPRAAAYLAGD--PPASAV	143
SEQ_ID_NO_4	147 <u>evvkrlngnllrepngndkfrvrngnprikealadrdggvelleavgftv</u>	196
SEQ_ID_NO_7	144 <u>EVVKRLLGNLLSDPRNDKYRKVRLGNPRIKEALADREGGVDLLEAVGERV</u>	193
SEQ_ID_NO_4	197 <u>gdeggepfavmdevpsdprlngirravlllegahpsappvkaeaeakesc</u>	246
SEQ_ID_NO_7	194 <u>ADEGGELFALMDEVPGDARLGGIRQAVLLLERARPSTPP-QTQADAKETC</u>	242
SEQ_ID_NO_4	247 sn-vsdvqegaktidrqrirvfvsvpgssmagndvpdsfyklsgeeirnea	295
SEQ_ID_NO_7	243 PNGVSEEQGIKKPVDRQIRVFFSVAASSVAENDLPDSFYLSNEEIRNEA	292
SEQ_ID_NO_4	296 kmrrerleqsrllipksykekqalaarqkyqavirvqfpdrmilqgifl	345
SEQ_ID_NO_7	293 KMRRERLEQSRLLIPKSYKEKQALAAARQKYQALIRIQFPDGVILQGVFL	342
SEQ_ID_NO_4	346 <u>pgeatslyefvtsalkqsglefelispaipkprvvphfpnpperartlq</u>	395
SEQ_ID_NO_7	343 <u>PAEPISSLYEFVASSLKQPSLEFDLICPAGPRTRVIPFPKPGEQARTLR</u>	392
SEQ_ID_NO_4	396 <u>eeelvpsallkfipketdsmvftgllldellmaseplpaasq</u>	436
SEQ_ID_NO_7	393 <u>DEDLVPSARLTFKPKETDSVVFTGLLDELLETSEPFTSASS</u>	433

FIGURA 2 E

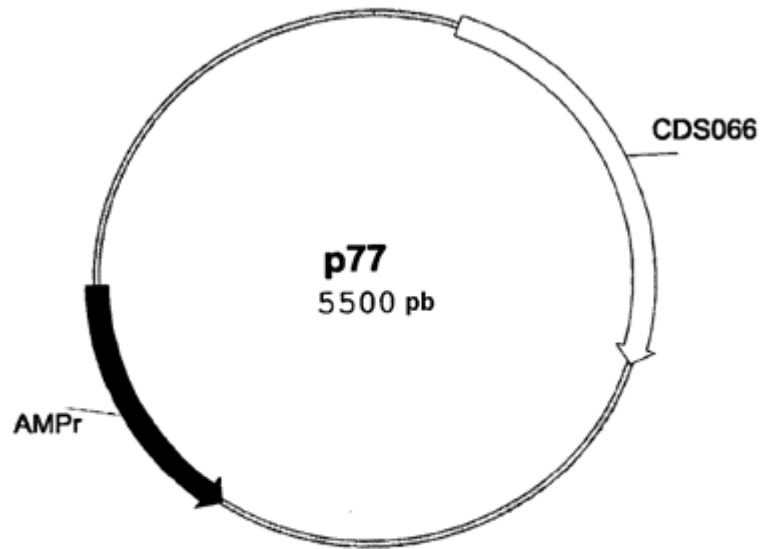


FIGURA 3

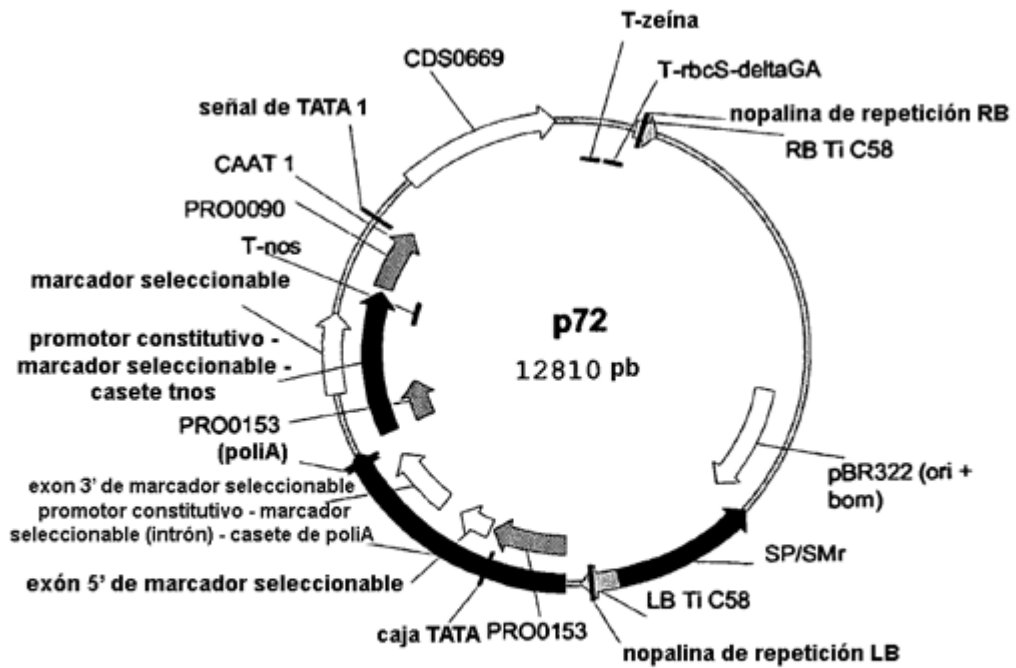


FIGURA 4

SEQ ID NO 1: Secuencia que codifica para GRUBX

atgggtgacatgaaggataaagtcaaaggggttcatgaaaaagtcacatcttcttctcaggttaagttta
aaggccaaggtaggggttttgggtgggttcatcttctcaggaccctcaaatacatgtcaataatcttctc
acatcccctaaatacaaggcaagatcaacaacctcatatacaaaaacttcgcctcaaaaaccaagtaat
tctgatcaaagaattgagaatatatgtgaaattcagttcaacaaaagtgaatcaaaggatgggtttgatc
catttgggtgaattagtcacttctgggaagagaaacccaaaaggggtattcacttactaatgtgtttgaatg
ccctgtctgtgtagtgggtttgtttctgaagaagaggtgtcaactcatattgatagctgtttaagttct
gaagtgtcttctaatttgggagttgaaagtaaagttgaagttaaaagtgaattggaaacatgtgttagtg
catatgttccagggaagccctcagaagggtcagttgaagtggtcattaagttgttaaagaatattgtgaa
ggaaccagagaatgccaagtttaggaaaataaggatggggaatccaaaaataaaaggtgctataggtgat
gtttagaggaggagtgagctattggaatttgttggatttgagttgaaagaagaaggtgggaaatttggg
ctgtgatggatgttcttctgaagaacaacttgttatgcttaagaatgtagtttctccttggaaaccgaa
gaaggttgaagagttggcgctcttatcccaagtttaaggcgagtgaaaccagttgagccgaagaagattgat
agacagattcgagtggttcttctgttcccgagagcgtagcagcaaaaattgagctacctgattccttct
ttaacctctcacgtgaggaattgagaagagaagcagagatgaggaagaagaattagaagattccaaatt
attgattcctaaatcttatcgggaaaagcaggcaaaaagctgcaagaaagaagtacacaaaatccattatc
cgtgtacagtttccagatggagcattgcttcaaggtgtcttctaccttcggagccaactagtgctctt
atgagtttgtgagcgcagcgttaaaggaaccaagcttagagttcgaattgttacatccgggtgcttgttaa
aaagcgggtgattccccatttccagctgctgggagagggctgtaacagttgaagaggagatttgggt
cctgcagctctactcaaatttaacctatcgaaacagattctgttgttttactggctcttgtaatgagc
tcttgaatagcagaccctcgagaccggatcagttgcttctcgttaa

SEQ ID NO 2: GRUBX, secuencia de proteína deducida

MGDMKDKVKGFMKKVTSSSSGKFKGQGRVLGSSSSSGPSNHVNNFSSHPLNTRQDQPPSYTKTSPQKPSN
SDQRIENICEIQFNKSESKDGFDPFGEVLTSGKRNPKGYSLTNVFECVPCGSGFVSEEVSTHIDSLSS
EVSSNLGVEKVEVKSELETCVSAYVSGKPSEGSVEVVIKLLKNIKKEPENAKFRKIRMGPNPKIKGAIGD
VVGVELLEFVGFELKEEGGEIWAVMDVPSEQLVMLKNVVSLLPEPKVEELASLSQVKASEPVEPKKID
RQIRVFFSVPEVAAKIELPDSFFNLSREELRREAEMRKKKLEDSKLLIPKSYREKQAKAARKKYTKSI
RVQFPDQALLQGVFLPSEPTSAFYEFVSAALKEPSLEFELLHPVLVKKRVI PHFPAAGERAVTVEEDLV
PAALLKFKPIETDSVFTGLCNELLEISEPLETGSVASS

SEQ ID NO 3: Ortólogo de GRUBX de caña de azúcar

atgatgaaggacaagatgaaggagttcatgaagaaggtcacctcctccgggtccgggacccccctcctcct
tcaagggcacctcccacgtcctcggctccggccccctccccctcctcctcccaccccctgcccgtcctc
aaacctagcccaaacctcaggcccgtcctaagcggacctcgccacctaccccggcccactttaaccacc
gatttgacctccttcacgccccctcgtctgctactcctcccgcgccccgacgcgaacggcacccgcnng
cgtcgcaccctcgcgtgccccagctgaggagacgcttccgtccgagctcgcctcctcggagcatct
cgacggctgcctcgcgtcggcggggggcgcccgcgcgcgccgctcctcgcgcccgaaccgctcct
ccgcccggggcctccgtagaggtggtcaaacgcctgctgggcaacctgctccgggagcccggcaacgata
agttcaggcgggtgagattgggtaaccgcggatcaaggaggccctggcagacagggatggcgggggtgga
gctcctggaggccgtcggctcacagttggggatgagggcggggagcccttcgcccgtgatggacgaagt
cctagcagaccctagggtcaacgggatcaggagggccgtcctcctgctcgagggggcacaccccctcgcgc
ctccagtgaaaggcggaggtgagccaaggagagctgcagcaatgtgtctgacgtgcaggaggggtgtaa
gactattgatcggcagattcgggtattgtctctgctcctgggagttctatggcacaataatgatgtacca
gattctttttacaagcttagtggtgaggagataaggaatgaagcaaatgaggagggaaaggttagaac
aatctcgattgctgataccaaagctttacaaggagaaacaggcattggctgctcgacagaagtataaaca
agcagtcattcagagttcagtttccagatagaatgattcttcagggcataattcctaccaggagaggccact
agttcactgtatgagttcgtcacatctgctctgaagcaatcaggtttggaaatcgaacttatctctccag
ccatacctaagccacgtgtggtgccccatttccaaaccccgggagagcgggacgcaccttgaagagga
ggagctggtcccctcgtcgtcctcaagttcattcccaggagactgattccatggttttcaccgggttg
cttgatgagcttctcatggccagtgagccgttctcgtgctcatcacaatga

FIGURA 5

SEQ ID N0 4: ortólogo de GRUBX de caña de azúcar, secuencia de proteína deducida
 MMKDKMKEFMKKVTS SSGSGT PSSF KGTSHVLGSGSPSSSH PAARSSNPSPNLRPAPKRTSPPTPPTLTT
 DLTSFTPLVCYSSRRPDANGTAXAVATVACPSCGDAFPSELAVSEHLDGCLASAGGARARAAA YLAADFP
 PPAASVEVVKRL LGNLLREPGNDKFRVRVRLGNPRIKEALADRDGGVELLEAVGFTVGDEGEPPFAVMDEV
 PSDPRLNGIRRAVLLLEGAHPSAPPVKAEEAEKESCSNVSDVQEGAKTIDRQIRV FVSVPGSSMAQNDVP
 DSFYKLSGEEIRNEAKMRRERLEQSRL LI PKSYKEKQALAA RQKYKQAVIRVQFPDRMILQGI FLPG EAT
 SSLYEFVTSALKQSGLEFELI SPAI PKPRVVP HFPN PGERARTLQEEELVPSALLKFI PKETDSMVF TGL
 LDELLMASEPLPAASQ

SEQ ID N0 5: casete de expresión que comprende GRUBX (1011-2390) operativamente unido al promotor de prolamina (1-654) y el terminador T-zeína + Rubisco deltaG (2615-2808 y 2852-3048):

cttctacatcggtcttaggtgtgcaaacacgactttattattattattattattattattattattattattattattttaca
 aaaatataaaatagatcagtcacctcaccacaagtagagcaagttgggtgagttattgtaaagttctacaaa
 gctaatttaaaagttattgcatctaacttatttcatattacaaacaagagtgatcaatggaacaatgaaaac
 catatgacatactataattttgtttttattattgaaattatataattcaaagagaataaatccacatagc
 cgtaaagttctacatgtgggtgacattacaaaatatatagcttacaaaacatgacaagcttagtttgaa
 aaattgcaatccttatcacattgacacataaaagtgagtgatgagtcataatatttttctttgctaccc
 atcatgtatataatgatagccacaaaagttactttgatgatgatatacaagaacatttttaggtgacctaa
 cagaatatccaaaataatgactcacttagatcataatagagcatcaagtaaaaactaacactctaaagca
 accgatgggaaagcatctataaatagacaagcacaatgaaaatcctcatcctccaccacaattcaaa
 tattatagttgaagcatagtagtaatttaaatcaactagggatatacacaagttgtacaaaaaagcaggc
 tggtagcgggtccggaattcccgggatatacgtcgacccacgcgtccgccaatatacagatttcttcatgaa
 ctccacttccaatttctcattgcttcttcttccatttccacctccaaagccatccttccagaaaacctt
 gttccttacatttcttagcccaaaaaagattcccatctcaattccacaaaaaacacaaggagatctaa
 ggaaattccccgcctctatatatagagaggtggaattggtcctgaatttggtttgaattgattgattgac
 agattttgggtgagaggggtgttattgaaaaaatggggtgacatgaaggataaaagtc aaagggttcatgaaaa
 aagtcacatcttcttctcaggttaagtttaaggccaaggtagggtttgggtgggtcatcttctcagg
 acctcaaatctgtcaataatttttcatcacatcccctaaatacaaggcaagatcaacaaccttcatat
 acaaaaaacttcgctcaaaaaaccaagtaattctgatcaagaattgagaataatgtgaaattcagttca
 acaaaagtgaatcaaaaggatgggtttgatccatttggtgaattagtcaactctgggaagagaaacccaaa
 agggattcacttactaatgtgtttgaatgccctgtctgtggtagtggttttgttctgaagaagagggtg
 tcaactcatattgatagctgtttaagttctgaagtgcttctaatttgggagttgaaagtaagttgaag
 ttaaaagtgaattggaacatgtgttagtgcatatgtttcagggaaagccctcagaagggtcagttgaagt
 ggtcattaagttgtaaaagaatattgtgaaggaaccagagaatgccaaagtttaggaaaaaaggatgggg
 aatccaaaaataaaagggtgctataggtgatgtttagtaggaggagtgagctattggaattgttggatttg
 agttgaaagaagaaggtggggaaatttgggctgtgatggatgttcttctgaagaacaacttgttatgct
 taagaatgtagtttctacttggaaaccgaagaaggttgaagagttggcgtccttatccaagttaaggcg
 agtgaaccagttgagccgaagaagattgatagacagattcagagttcttcttctgttcccgagagcgtag
 cagcaaaaaattgagctacctgatccttctttaaacctctcacgtgaggaattgagaagagaagcagagat
 gaggaagaagaattagaagattccaaattattgattcctaaatcttatcgggaaaagcaggcaaaagct
 gcaagaagaagtagacacaaaatccattatccgtgtacagtttccagatggagcattgcttcaaggtgctt
 ttctaccttccggagccaactagtgctctttatgagtttgtgagcgcagcgttaaaggaaccaagcttaga
 gttcgaattgttacatccggtgcttgttaaaaagcgggtgatccccattttccagctgctggggagagg
 gctgtaacagttgaagaggaggatttgggtcctgacgctctactcaaatttaaacctatcgaaacagatt
 ctggtgttttactgggtctttgtaatgagcttcttgaatttagcgagcccctcgagaccggatcagttgc
 ttctcgttaagctctaaattacatcagactttgaattcttctctgagttgtgaaaccttataaaactctct
 gcgcccggaaatgctcggccgctctagagtatccctcgaggggccaagcttacgcgtacccagcttctt
 tgtacaaagttggtgatatacaagcccggcggtcttctagggataaacagggtaatatatccctctaga
 tcacaagcccggggcgttcttctacgatgattgagtaataatgtgtcacgcatcccatgggtggcagtg

FIGURA 5 (continuación)

