

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 403 414**

51 Int. Cl.:

A01H 5/00 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.03.2005 E 05729788 (9)**

97 Fecha y número de publicación de la concesión europea: **16.01.2013 EP 1732378**

54 Título: **Plantas que sobreexpresan una proteína de tipo NAP1 con aumento del rendimiento de semillas, método para producir las mismas**

30 Prioridad:

02.04.2004 EP 04101388

20.04.2004 US 563847 P

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

17.05.2013

73 Titular/es:

CROPDESIGN N.V. (100.0%)

TECHNOLOGIEPARK 3

9052 ZWIJNAARDE, BE

72 Inventor/es:

DUDITS, DENES;

FEHER, ATTILA y

FRANKARD, VALERIE

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 403 414 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 sobreexpresan una proteína de tipo NAP1 con aumento del rendimiento de semillas, método para producir las mismas.

5 La presente invención se refiere a un método para mejorar características de crecimiento, y en particular para aumentar el rendimiento de una planta. Más específicamente, la presente invención se refiere a un método para aumentar el rendimiento modulando la expresión en una planta de un ácido nucleico que codifica para una proteína homóloga a la proteína de ensamblaje de nucleosomas 1 (proteína de tipo NAP1). La presente invención también se refiere a plantas que tienen modulada la expresión de un ácido nucleico que codifica para una proteína de tipo NAP1, plantas que tienen un aumento del rendimiento con relación a plantas de tipo natural correspondientes.

10 Dada la población mundial cada vez mayor, y la disminución del área de tierra disponible para la agricultura, sigue siendo un objetivo principal mejorar la eficacia de la agricultura y aumentar la diversidad de plantas en horticultura. Los medios convencionales para las mejoras de cultivos y en horticultura utilizan técnicas de reproducción selectiva para identificar plantas que tienen características deseables. Sin embargo, tales técnicas de reproducción selectivas tienen varios inconvenientes, concretamente que estas técnicas requieren normalmente mucha mano de obra y dan
15 como resultado plantas que contienen a menudo complementos genéticos heterogéneos que pueden no dar siempre como resultado el rasgo deseable que se transmite desde las plantas parentales. Los avances en la biología molecular han permitido a la humanidad manipular el germoplasma de animales y plantas. La ingeniería genética de plantas conlleva el aislamiento y la manipulación de material genético (normalmente en forma de ADN o ARN) y la posterior introducción de ese material genético en una planta. Tal tecnología ha conducido al desarrollo de plantas
20 que tienen diversos rasgos económicos, agronómicos u hortícolas mejorados. Rasgos de particular interés económico son características de crecimiento tales como alto rendimiento.

Las proteínas NAP forman una familia de proteínas relacionadas que se conocen en animales y se notifica que están implicadas en actividades relacionadas con la cromatina. La familia de proteínas NAP se caracteriza por la presencia de una secuencia conservada conocida como dominio NAP. El dominio NAP se describe en las bases de datos
25 Pfam (registro PF00956) e Interpro (registro IPR002164). NAP es un componente de un complejo multifactorial que media el empaquetamiento de ADN en nucleosomas (Krude, T. y Keller, C. (2001) Cell. Mol. Life Sci. 58, 665-672). Durante la fase S del ciclo de división celular en eucariotas, el ADN recién replicado se ensambla rápidamente en la cromatina. Este proceso requiere la acción coordinada de varios factores. En las etapas iniciales, CAF1 (factor 1 de ensamblaje de cromatina) se une a las proteínas histonas H3 y H4 y las dirige a la horquilla de replicación a través de la unión a PCNA. La deposición posterior de las proteínas histonas H2A y H2B está mediada por proteínas NAP1.
30 NAP1 se describió por primera vez en células HeLA (von Lindern *et al.* (1992) Mol. Cell. Biol. 12, 3346-3355) y se encontró más tarde conservada en todas las eucariotas. Además, se cree que las proteínas NAP regulan la transcripción génica y pueden influir en el desarrollo y la diferenciación celulares.

Las proteínas SET están sumamente relacionadas con las proteínas NAP y desempeñan un papel en diversos procesos celulares en seres humanos. En células humanas, se ha mostrado que las SET están asociadas con
35 diversos complejos de CDK-ciclina durante la regulación del ciclo celular, tal como la transición G2/M. SET es un potente inhibidor de la proteína fosfatasa 2A (PP2A) que está implicada en varias rutas de señalización. La actividad inhibidora de SET podría atribuirse a un dominio C-terminal ácido (Canela *et al.* (2003) J. Biol. Chem. 278, 1158-1164). Otros informes muestran la implicación de SET en la transcripción y la reparación de ADN. SET es parte de un complejo que tiene actividades de unión y curvatura del ADN mediadas por la proteína asociada a la cromatina
40 HMG2. HMG2 facilita el ensamblaje de estructuras de orden superior de nucleoproteínas mediante la curvatura y formación de bucles de ADN o mediante la estabilización de ADN subenrollado. HMG2 precipita conjuntamente con SET (Fan *et al.* (2002) Mol. Cell. Biol. 22, 2810-2820). También se notifica que SET inhibe la desmetilación activa de ADN (Cervoni *et al.* (2002) J. Biol. Chem. 277, 25026-25031). La oncoproteína Set/TAF-1, implicada en la inhibición de la acetilación de histonas, también inhibe la desmetilación de ADN metilado de manera ectópica que da como resultado el silenciamiento génico. Se sugiere que Set/TAF-1 desempeña un papel en la integración de estados epigenéticos de histonas y ADN en la regulación génica.
45

La actividad de las proteínas NAP1 está regulada en parte mediante fosforilación. Se mostró que la localización subcelular de NAP1 en *Drosophila* depende de su estado de fosforilación, que puede controlarse por caseína cinasa II (Rodríguez *et al.* (2000) J. Mol. Biol. 298, 225-238). Se notifica que los mamíferos presentan varias proteínas NAP1, mientras que en levaduras sólo existe una proteína NAP1 conocida.
50

Los ortólogos de NAP1 de plantas siguen siendo desconocidos en gran medida, aunque se notificaron proteínas NAP1 de la soja (Yoon *et al.* (1995) Mol. Gen. Genet. 249, 465-473), *Arabidopsis*, tabaco, maíz y arroz (Dong *et al.* (2003) Plant 216, 561-570). El análisis filogenético de genes de tipo NAP1 de plantas ha revelado que existen dos subgrupos, uno relacionado con NAP1 y el otro con la proteína SET (figura 1). De la manera más probable, puede haberse producido divergencia de secuencia posterior puesto que el agrupamiento de las dos secuencias de *Arabidopsis*, las dos de maíz y las dos de tabaco señalan en conjunto a un efecto de duplicación génica más reciente. El genoma de *Saccharomyces cerevisiae* contiene sólo un gen que codifica para NAP, que combina las
55

propiedades funcionales de ambos subgrupos de NAP1 y SET. De manera similar, el factor 1 de activación de molde (TAF-1), un homólogo de NAP1, combina tanto actividad de inhibición de PP2a (Saito *et al.*, Biochem. Biophys. Res. Comm. 259, 471-475, 1999) como actividad de remodelación de cromatina (Kawase *et al.*, Genes Cells 1, 1045-1056, 1996). Por tanto, es probable que las proteínas de plantas de la familia NAP/SET sean redundantes en gran medida en su función, particularmente en el grupo de proteínas SET en el que se observa un menor grado de divergencia en comparación con el grupo de NAP. Además, existen pruebas estructurales de que las proteínas NAP y SET pertenecen a la misma familia puesto que comparten el dominio NAP que va seguido por una región ácida C-terminal.

Se conoce poco sobre la función de proteínas de tipo NAP1 en plantas, aunque se ha propuesto un papel en la mitosis y la citocinesis (Dong *et al* 2003). De la manera más probable, los ortólogos de plantas de la proteína NAP1 desempeñan un papel diferente que sus homólogos de animales. Basándose en su localización nuclear y en las similitudes de secuencia con la proteína SET de mamífero, puede esperarse un papel en la remodelación de cromatina para las proteínas de plantas. Además, el grupo de proteínas NAP/SET de plantas podría estar implicado en la regulación de PP2A en plantas. PP2A es una de las principales fosfatasa en plantas, que actúa en gran medida sobre factores de transcripción y proteínas cinasas, y se ha propuesto que regula la actividad de proteínas implicadas en una variedad de procesos celulares, incluyendo el ciclo celular (Ayaydin *et al.* (2000) Plant J. 23, 85-96), acciones hormonales tales como el movimiento de estomas mediado por ABA, la germinación (Kwak *et al.* (2002) Plant Cell 14, 2849-2861), o el transporte de auxina y el desarrollo de las raíces (Garbers *et al* 1996 EMBO J. 15, 2115-2124). Se notifica además que PP2A está implicada en la fotosíntesis y la señalización luminosa (Sheen (1993) EMBO J. 12, 3497,3505) y en la asimilación de nitrógeno (Hirose y Yamaya (1999) Plant Physiology 121, 805-812).

Hasta la fecha, no se han descrito efectos sobre rasgos agronómicos con la modulación de la expresión de proteínas de tipo NAP1 en plantas. Se ha descubierto ahora sorprendentemente que la modulación de la expresión de un ácido nucleico que codifica para una proteína de tipo NAP1 en una planta proporciona que las plantas tengan un crecimiento y desarrollo mejorados, en particular el aumento del rendimiento, más particularmente el aumento del rendimiento de semillas en comparación con plantas de tipo natural correspondientes. Por tanto, según una primera realización de la presente invención se proporciona un método para mejorar el rendimiento de semillas tal como se expone en la reivindicación 1.

El término proteína de tipo NAP1, tal como se define en el presente documento, se refiere a tal como se define en la reivindicación 1 que comprende un dominio NAP y una región C-terminal ácida y que tienen actividad de inhibición de la fosfatasa PP2a. El término "dominio NAP" tal como se usa en el presente documento es tal como se define por la base de datos Pfam mediante el número de registro PF00956 (<http://www.sanger.ac.uk/Software/Pfam/>; Bateman *et al.*, Nucleic Acid Research 30(1):276-280 (2002), véase por ejemplo la tabla 1). Las secuencias de proteína de tipo NAP1 útiles en la presente invención tienen un dominio NAP que comprende una secuencia distintiva (T/S)FF(TIN/S/E/D)(W/F)(L/F) y/o la secuencia de aminoácidos conservada facilitada en SEQ ID NO:33. Más preferiblemente, el dominio NAP es tal como se representa por SEQ ID NO: 32 El término "región C-terminal ácida" o "extremo C-terminal ácido" tal como se usa en el presente documento se refiere al extremo carboxi-terminal de la proteína, extremo carboxi-terminal que tiene aproximadamente de 20 a 25 aminoácidos de largo, del que al menos 13 residuos son ácido glutámico y/o aspártico.

Tabla 1: Ejemplos de proteínas de *Arabidopsis* que comprenden un dominio NAP1

ID del gen	Perfil de Pfam	Posición	Puntuación	Valor de e	N.º de ID de SET:
at1g18800	PF00956	27-224	147,7	2e-40	20, 21
at1g74560	PF00956	31-229	135,0	1,3e-36	1, 2
at2g 19480	PF00956	52-300	457,4	1,2e-133	26, 27
at5g56950	PF00956	52-300	473,2	2,2e-138	28, 29
at4g26110	PF00956	52-301	503,4	1,7e-147	24, 25
at3g 13782	PF00956	69-311	300,7	1,7e-86	30, 31

Opcionalmente, la proteína de tipo NAP1 útil en los métodos de la presente invención tiene, además de ser un inhibidor de fosfatasa PP2a, también actividades de remodelación de cromatina. Se conocen en la técnica métodos

para medir la inhibición de fosfatasa PP2a y comprenden, por ejemplo, ensayos basados en sustratos marcados con colorante o marcados con fluorocromo disponibles comercialmente o ensayos basados en la medición de radiactividad en fracciones solubles en TCA tras tratamiento con fosfatasa PP2a de histona H1 marcada con [³²P] (Saito *et al.* (1999) *Biochem. Biophys. Res. Comm.* 259, 471-475) o de radiactividad liberada por proteína básica de mielina marcada (U *et al.*, *J. Biol. Chem.* 271, 11059-11062, 1996). Se facilita un método alternativo para medir la actividad de proteína de tipo NAP1 en el ejemplo 6, que se basa en el método descrito por Ulloa *et al.* (FEBS Letters 330, 85-89, 1993). Pueden someterse a ensayo las actividades de remodelación de cromatina de varias maneras, tales como la medición de la actividad de unión al ADN en un ensayo de retardo en gel (Fan *et al.*, 2002) o como la medición de la actividad de unión a histonas usando ELISA (Rodríguez *et al.* (1997) *Genomics* 44, 253-265). Puede determinarse la actividad de curvatura del ADN en un ensayo de circularización mediada por ligasa (Fan *et al.*, 2002) o en un ensayo de superenrollamiento (Fujii-Nakata *et al.* (1992) *J. Biol. Chem.* 267, 20980-20986; Yoon *et al.* (1995), *Mol. Gen. Gen.* 249, 465-473).

Preferiblemente, la proteína de tipo NAP1, que comprende un dominio NAP y una región C-terminal ácida tal como se describió anteriormente, es de origen vegetal. La proteína de tipo NAP1 es preferiblemente de una planta dicotiledónea, preferiblemente de la familia de *Brassicaceae*, más preferiblemente de *Arabidopsis thaliana*, lo más preferiblemente la proteína de tipo NAP1 es una proteína representada por SEQ ID NO: 2 o es un homólogo, derivado o fragmento activo de la misma tal como se define en la reivindicación 1, comprendiendo los homólogos, derivados o fragmentos activos un dominio NAP y el extremo C-terminal ácido tal como se describió anteriormente, y teniendo además los homólogos, derivados o fragmentos activos actividad de inhibición de PP2A. Preferiblemente, las proteínas de tipo NAP1 están codificadas por un ácido nucleico representado por SEQ ID NO: 1 o un ácido nucleico que puede hibridarse con el mismo en condiciones rigurosas.

El término "proteína de tipo NAP1" incluye proteínas homólogas a la proteína presentada en SEQ ID NO: 2. Los homólogos que van a usarse en los métodos de la presente invención comprenden un dominio NAP y un extremo C-terminal ácido tal como se describió anteriormente. Los homólogos tienen un dominio NAP que comprende una secuencia distintiva (T/S)FF(T/N/S/E/D)(W/F)(L/F) y/o la secuencia conservada de SEQ ID NO: 33, y un extremo C-terminal ácido de 20 a 25 residuos que comprende al menos 13 residuos de ácido aspártico y/o glutámico. Adicionalmente, los homólogos de NAP1 tienen actividad de inhibición de PP2a y presentan opcionalmente también actividades de remodelación de cromatina, que pueden medirse tal como se describió anteriormente.

Pueden encontrarse homólogos de SEQ ID NO: 2 en diversos organismos eucariotas. Los homólogos más próximos se encuentran generalmente en el reino vegetal. Los homólogos de SEQ ID NO: 2 adecuados en los métodos de la presente invención incluyen dos proteínas del tabaco (SEQ ID NO: 7 y 9), una proteína del tomate (SEQ ID NO: 23), una proteína de la alfalfa (SEQ ID NO: 11), la proteína de *Arabidopsis* representada por SEQ ID NO: 21. Otros homólogos adecuados para poner en práctica el método según la invención incluyen, por ejemplo, los homólogos de *Zea mays* nfa104 (n.º de registro AF384036, SEQ ID NO: 13) y nfa103 (n.º de registro AF384035, SEQ ID NO: 19); o los homólogos de *Oryza sativa* (SEQ ID NO: 15 y 17).

Los métodos para la búsqueda e identificación de homólogos de tipo NAP1 estarían completamente dentro del campo de los expertos en la técnica. Tales métodos comprenden la comparación de las secuencias representadas por SEQ ID NO: 1 ó 2, en un formato legible por ordenador, con secuencias que están disponibles en bases de datos públicas tales como MIPS (<http://mips.gsf.de/>), GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>) o la base de datos de secuencias de nucleótidos EMBL (<http://www.ebi.ac.uk/embl/index.html>), usando algoritmos bien conocidos en la técnica para la alineación o la comparación de secuencias, tales como GAP (Needleman y Wunsch, *J. Mol. Biol.* 48; 443-453 (1970)), BESTFIT (usando el algoritmo de homología local de Smith y Waterman (Advances in Applied Mathematics 2; 482-489 (1981))), BLAST (Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J., *J. Mol. Biol.* 215:403-410 (1990)), FASTA y TFASTA (W. R. Pearson y D. J. Lipman *Proc.Natl. Acad.Sci. USA* 85:2444- 2448 (1988)). El software para realizar análisis de BLAST está disponible para el público a través del Centro Nacional para Información Biotecnológica. Se identificaron los homólogos mencionados anteriormente usando los parámetros por defecto de Blast (matriz BLOSUM62, penalización por apertura de hueco de 11 y penalización por extensión de hueco de 1) y preferiblemente se usan las secuencias de longitud completa para el análisis. Alternativamente, puede usarse sólo el dominio NAP1 para la comparación, puesto que este dominio comprende la parte principal de la proteína.

Los "homólogos" de una proteína de tipo NAP1 engloban péptidos, oligopéptidos, polipéptidos, proteínas y enzimas que tienen sustituciones, deleciones y/o inserciones de aminoácidos con relación a la proteína sin modificar en cuestión y que tienen actividad biológica y funcional similar a la de la proteína sin modificar de la que se derivan. Para producir tales homólogos, pueden sustituirse aminoácidos de la proteína por otros aminoácidos que tienen propiedades similares (tales como similar hidrofobicidad, hidrofiliidad, antigenicidad, propensión a formar o romper estructuras de hélice α o estructuras de lámina β). Se conocen bien en la técnica tablas de sustituciones conservativas (véase, por ejemplo, Creighton (1984) *Proteins*. W.H. Freeman and Company).

Los homólogos útiles en el método según la invención tienen una identidad de secuencia de al menos el 50% con la secuencia de proteína representada en SEQ ID NO: 2 (registro GenBank NP177596), alternativamente una similitud

o identidad de secuencia de al menos el 60% o el 70% con SEQ ID NO: 2. Normalmente, los homólogos tienen una similitud o identidad de secuencia de al menos el 80%, preferiblemente una similitud o identidad de secuencia de al menos el 85%, más preferiblemente una similitud o identidad de secuencia de al menos el 90%, lo más preferiblemente una similitud o identidad de secuencia de al menos el 95%, 96%, 97%, 98% o el 99% con SEQ ID NO: 2.

El porcentaje de identidad puede calcularse partiendo de secuencias de proteína de longitud completa o de determinadas regiones (preferiblemente, conservadas) de una secuencia de este tipo, usando programas de alineación basados en el algoritmo de Needleman y Wunsch (tales como GAP), usando la matriz BLOSUM62, una penalización por apertura de hueco de 10 y una penalización por extensión de hueco de 0,5. Por ejemplo, puede usarse el dominio NAP facilitado en SEQ ID NO: 32 como consulta. La identificación de tales dominios en una secuencia de proteína estaría completamente dentro del campo del experto en la técnica e implicaría un formato legible por ordenador de los ácidos nucleicos usados en los métodos de la presente invención, el uso de programas de software de alineación y el uso de información disponible para el público sobre dominios de proteína, cajas y motivos conservados. Puede realizarse una búsqueda integrada usando la base de datos INTERPRO (Mulder *et al.*, (2003) Nucl. Acids Res. 31, 315-318, <http://www.ebi.ac.uk/interpro/scan.html>) que combina varias bases de datos sobre familias de proteínas, dominios y sitios funcionales, tales como las bases de datos PRODOM (Servant *et al.*, (2002) Briefings in Bioinformatics 3, 246-251, <http://prodes.toulouse.inra.fr/prodom/2002,1/html/home.php>), PIR (Huang *et al.* (2003) Nucl. Acids Res. 31, 390-392, <http://pir.georgetown.edu/>) o Pfam (Bateman *et al.* (2002) Nucl. Acids Res. 30, 276-280, <http://pfam.wustl.edu/>). Pueden usarse programas de análisis de secuencias diseñados para la búsqueda de motivos para la identificación de fragmentos conservados, regiones y dominios tal como se mencionó anteriormente. Los programas informáticos adecuados incluyen, por ejemplo, MEME (Bailey y Elkan (1994) Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, págs. 28-36, AAAI Press, Menlo Park, California, <http://meme.sdsc.edu/meme/website/intro.html>).

Pueden agruparse las proteínas homólogas en "familias de proteínas". Una familia de proteínas puede definirse mediante análisis funcional y de similitud de secuencia, tal como, por ejemplo, Clustal W. Puede generarse un árbol del vecino más cercano (*neighbour-joining*) de las proteínas homólogas a la de tipo NAP1 mediante el programa Clustal W y proporciona una buena visión general de su relación estructural y ancestral. En una realización particular de la presente invención, el/los homólogos de tipo NAP1 pertenece(n) a la misma familia de proteínas que la proteína correspondiente a SEQ ID NO: 2. En el genoma de *Arabidopsis*, se identificaron dos miembros de la familia de la proteína de tipo NAP1 (NM177596 (SEQ ID NO: 2), NP564063 (SEQ ID NO: 21)). También pueden identificarse miembros de la familia de la proteína de tipo NAP1 en otras plantas tales como arroz u otras plantas monocotiledóneas. Ventajosamente, estos miembros de la familia también son útiles en los métodos de la presente invención.

Dos formas especiales de homología, ortóloga y paróloga, son conceptos evolutivos usados para describir las relaciones ancestrales de genes. El término "parólogo" se refiere a genes homólogos que resultan de una o más duplicaciones génicas dentro del genoma de una especie. El término "ortólogo" se refiere a genes homólogos en diferentes organismos debido a una relación ancestral de estos genes. El término "homólogo" tal como se usa en el presente documento también engloba parálogos y ortólogos de las proteínas útiles en los métodos según la invención.

Pueden identificarse genes ortólogos realizando una consulta en una o más bases de datos de genes con un gen de interés de consulta, usando por ejemplo el programa BLAST. Los genes objeto de mayor clasificación que resultan de la búsqueda se someten de nuevo a análisis de BLAST, y sólo aquellos genes objeto que coinciden de nuevo con el gen de consulta se conservan como verdaderos genes ortólogos. Por ejemplo, para hallar un ortólogo de arroz de un gen de *Arabidopsis thaliana*, puede realizarse un análisis de BLASTN o TBLASTX en una base de datos de arroz (tal como (pero sin limitarse a) la base de datos de *Oryza sativa* Nipponbare disponible en el NCBI (<http://www.ncbi.nlm.nih.gov>) o las secuencias genómicas del arroz (cultivares índica o japónica)). En una siguiente etapa, se usan las secuencias de arroz obtenidas en un análisis de BLAST inverso usando una base de datos de *Arabidopsis*. Los resultados pueden refinarse adicionalmente cuando se analizan las secuencias resultantes con ClustalW y se visualizan en un árbol del vecino más cercano. Este método puede usarse para identificar ortólogos de muchas especies diferentes.

Los "homólogos" de una proteína de tipo NAP1 engloban proteínas que tienen sustituciones, inserciones y/o deleciones de aminoácidos con relación a la proteína sin modificar en cuestión. "Variantes de sustitución" de una proteína son aquéllas en las que se ha eliminado al menos un residuo en una secuencia de aminoácidos y se ha insertado en su lugar un residuo diferente. Las sustituciones de aminoácidos son normalmente de residuos individuales, pero pueden agruparse dependiendo de restricciones funcionales puestas sobre el polipéptido; las inserciones serán habitualmente del orden de aproximadamente 1 a 10 residuos de aminoácidos, y las deleciones oscilarán entre aproximadamente 1 y 20 residuos. Preferiblemente, las sustituciones de aminoácidos comprenden sustituciones de aminoácidos conservativas. "Las variantes de inserción" de una proteína son aquéllas en las que se introducen uno o más residuos de aminoácido en un sitio predeterminado en una proteína. Las inserciones pueden comprender fusiones amino-terminales y/o carboxi-terminales así como inserciones intra-secuencia de aminoácidos

individuales o múltiples. Generalmente, las inserciones dentro de la secuencia de aminoácidos serán menores que las fusiones amino o carboxi-terminales. Ejemplos de proteínas o péptidos de fusión amino o carboxi-terminal incluyen el dominio de unión o el dominio de activación de un activador transcripcional tal como se usa en el sistema de dos híbridos en levadura, proteínas de la cubierta de fagos, cola de 6 (histidina), cola de glutatión-S-transferasa, proteína A, proteína de unión a maltosa, dihidrofolato reductasa, epítipo Tag-100, epítipo de c-myc, epítipo FLAG[®], lacZ, CMP (péptido de unión a calmodulina), epítipo de HA, epítipo de proteína C y epítipo de VSV. Las “variantes de delección” de una proteína se caracterizan por la eliminación de uno o más aminoácidos de la proteína. Pueden prepararse fácilmente variantes de aminoácidos de una proteína usando técnicas de síntesis de péptidos bien conocidas en la técnica, tales como síntesis de péptidos en fase sólida y similares, o mediante manipulaciones ADN recombinante. Se conocen bien en la técnica métodos para la manipulación de secuencias de ADN para producir variantes de sustitución, inserción o delección de una proteína. Por ejemplo, se conocen bien técnicas para producir mutaciones de sustitución en sitios predeterminados en el ADN por los expertos en la técnica e incluyen mutagénesis con M13, mutagénesis *in vitro* del gen de T7 (USB, Cleveland, OH), mutagénesis dirigida al sitio QuickChange (Stratagene, San Diego, CA), mutagénesis dirigida al sitio mediada por PCR u otros protocolos de mutagénesis dirigida al sitio.

El término “derivados” de una proteína de tipo NAP1 se refiere a péptidos, oligopéptidos, polipéptidos, proteínas y enzimas que pueden comprender sustituciones, delecciones o adiciones de residuos de aminoácido que se producen y no se producen de manera natural en comparación con la secuencia de aminoácidos de la forma que se produce de manera natural de la proteína de tipo NAP1 (tal como la proteína presentada en SEQ ID NO: 2). Los “derivados” de una proteína engloban péptidos, oligopéptidos, polipéptidos, proteínas y enzimas que pueden comprender residuos de aminoácido que se producen de manera natural alterados, glicosilados, acilados o que no se producen de manera natural en comparación con la secuencia de aminoácidos de una forma que se produce de manera natural del polipéptido. Un derivado también puede comprender uno o más sustituyentes distintos a aminoácido en comparación con la secuencia de aminoácidos de la que se deriva, por ejemplo una molécula indicadora u otro ligando, unido de manera covalente o no covalente a la secuencia de aminoácidos tal como, por ejemplo, una molécula indicadora que se une para facilitar su detección, y residuos de aminoácido que no se producen de manera natural con relación a la secuencia de aminoácidos de una proteína que se produce de manera natural.

Los “fragmentos activos” de una proteína de tipo NAP1 engloban al menos la cantidad de residuos de aminoácido, suficiente para conservar una actividad biológica y/o funcional similar en comparación con la proteína que se produce de manera natural. Un fragmento activo preferido de una proteína de tipo NAP1 comprende al menos un dominio NAP (Pfam 00956) y un dominio C-terminal ácido enriquecido en residuos D y/o E tal como se describió anteriormente, un fragmento activo más preferido comprende además la secuencia conservada de SEQ ID NO: 33.

El término gen/ácido nucleico de tipo NAP1, tal como se define en la reivindicación 1 en el presente documento, se refiere a cualquier ácido nucleico que codifica para una proteína de tipo NAP1 tal como se define. El ácido nucleico puede derivarse (o bien directa o bien indirectamente (si se modifica posteriormente)) de cualquier fuente siempre que el ácido nucleico, cuando se expresa en una planta, conduzca a la expresión modulada de un gen/ácido nucleico de tipo NAP1 o la actividad y/o los niveles modulados de una proteína de tipo NAP1. El ácido nucleico puede aislarse de una fuente eucariota, tal como levadura u hongos, plantas (incluyendo algas) o animales (incluyendo seres humanos). Este ácido nucleico puede modificarse sustancialmente a partir de su forma nativa en composición y/o entorno genómico a través de manipulación humana deliberada. La secuencia de ácido nucleico es preferiblemente una secuencia de ácido nucleico homóloga, es decir una secuencia de ácido nucleico que codifica para una proteína relacionada estructural y/o funcionalmente con SEQ ID NO: 2, preferiblemente obtenida de una planta, ya sea de la misma especie de planta o de una diferente. Preferiblemente, el ácido nucleico es tal como se representa por SEQ ID NO: 1 o una parte de la misma o una secuencia de ácido nucleico que puede hibridarse con el mismo, codificando la secuencia de hibridación para proteínas que tienen actividad de proteína de tipo NAP1 tal como se describió anteriormente; o es un ácido nucleico que codifica para un aminoácido representado por SEQ ID NO: 2 o que codifica para un homólogo, derivado o fragmento activo del mismo. Este término también engloba variantes del ácido nucleico que codifican para una proteína de tipo NAP1 debido a la degeneración del código genético; variantes alélicas; y diferentes variantes de corte y empalme del ácido nucleico que codifican para una proteína de tipo NAP1, incluyendo variantes que están interrumpidas por una o más secuencias intermedias.

Ventajosamente, el método tal como se da a conocer en el presente documento también puede ponerse en práctica usando partes de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 definida anteriormente, tal como la proteína de tipo NAP1 codificada por SEQ ID NO: 1, o usando secuencias que se hibridan con una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 definida anteriormente, tal como SEQ ID NO: 1, preferiblemente en condiciones rigurosas, (codificando las secuencias de hibridación para proteínas que tienen actividad de tipo NAP1), o usando ácidos nucleicos que codifican para homólogos, derivados o fragmentos activos de una proteína de tipo NAP1, tal como la representada por SEQ ID NO: 2.

Partes de una secuencia de ADN se refieren a un trozo de ADN derivado o preparado a partir de una molécula de ADN original (más grande), dando lugar la parte de ADN, cuando se expresa en una planta, a plantas que tienen características de crecimiento mejoradas. Preferiblemente, las características de crecimiento mejoradas son el

aumento del rendimiento, más preferiblemente el aumento del rendimiento de semillas, lo más preferiblemente el aumento del rendimiento de semillas comprende uno o más de aumento del índice de cosecha, aumento del peso total de semillas y aumento del número de semillas llenas. La parte puede comprender muchos genes, con o sin elementos de control adicionales, o puede contener sólo secuencias espaciadoras, etc.

5 La presente invención también engloba secuencias de ácido nucleico que pueden hibridarse en condiciones rigurosas con una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1, pudiendo ser también las secuencias de ácido nucleico útiles en la puesta en práctica de los métodos según la invención. El término "hibridación" tal como se define en el presente documento es un proceso en el que sustancialmente secuencias de nucleótidos complementarias homólogas se aparean entre sí. El proceso de hibridación puede producirse
10 completamente en disolución, es decir ambos ácidos nucleicos complementarios están en disolución. Las herramientas en biología molecular que se basan en tal proceso incluyen la reacción en cadena de la polimerasa (PCR; y todos los métodos basados en la misma), hibridación sustractiva, extensión con cebadores aleatorios, mapeo con nucleasa S1, extensión con cebadores, transcripción inversa, síntesis de ADNc, presentación diferencial de ARN y determinación de la secuencia de ADN. El proceso de hibridación también puede producirse con uno de los ácidos nucleicos complementarios inmovilizados en una matriz tal como perlas magnéticas, perlas de Sepharose o cualquier otra resina. Las herramientas en biología molecular que se basan en tal proceso incluyen el aislamiento de ARNm de poli (A+). El proceso de hibridación puede producirse además con unos de los ácidos nucleicos complementarios inmovilizados en un soporte sólido tal como una membrana de nitrocelulosa o nailon o inmovilizados mediante, por ejemplo, fotolitografía en, por ejemplo, un soporte de vidrio síliceo (esto último conocido como alineamientos o microalineamientos de ácido nucleico o como chips de ácido nucleico). Las herramientas en biología molecular que se basan en tal proceso incluyen análisis de transferencia en gel de ARN y ADN, hibridación de colonias, hibridación de placas, hibridación *in situ* e hibridación de microalineamientos. Para permitir que se produzca la hibridación, las moléculas de ácido nucleico generalmente se desnaturalizan térmica o químicamente para fundir una cadena doble en dos cadenas sencillas y/o para eliminar horquillas u otras estructuras secundarias a partir de ácidos nucleicos monocatenarios. La rigurosidad de hibridación está influida por condiciones tales como temperatura, concentración salina y composición del tampón de hibridación. Para aplicaciones que requieren alta selectividad, se deseará normalmente emplear condiciones relativamente rigurosas para formar los híbridos, por ejemplo, se seleccionarán condiciones de relativamente baja concentración salina y/o alta temperatura, tales como las proporcionadas por NaCl aproximadamente de 0,02 M a aproximadamente 0,15 M a temperaturas de aproximadamente 50°C a aproximadamente 70°C. Las condiciones de alta rigurosidad para la hibridación incluyen, por tanto, alta temperatura y/o baja concentración salina (las sales incluyen NaCl y citrato de Na₃) pero también pueden estar influidas por la inclusión de formamida en el tampón de hibridación y/o la disminución de la concentración de compuestos tales como SDS (dodecilsulfato de sodio) en el tampón de hibridación y/o la exclusión de compuestos tales como sulfato de dextrano o polietilenglicol (que promueven la aglomeración molecular) del tampón de hibridación. Se prefieren particularmente condiciones de hibridación de rigurosidad suficientemente baja para el aislamiento de ácidos nucleicos homólogos a las secuencias de ADN útiles en los métodos de la invención definidas anteriormente. Los elementos que contribuyen a la homología incluyen el alelismo, la degeneración del código genético y las diferencias en la utilización de codones preferidos. La presencia de cationes monovalentes en la disolución de hibridación reduce la repulsión electrostática entre las dos cadenas de ácido nucleico promoviendo de ese modo la formación de híbridos; este efecto es visible para concentraciones de sodio de hasta 0,4 M. La formamida reduce la temperatura de fusión de dúplex de ADN-ADN y ADN-ARN con de 0,6 a 0,7°C para cada tanto por ciento de formamida, y la adición de formamida al 50% permite que se realice la hibridación a de 30 a 45°C, aunque disminuirá la tasa de hibridación. Los apareamientos erróneos de pares de bases reducen la tasa de hibridación y la estabilidad térmica de los dúplex. En promedio y para sondas grandes, la T_f (temperatura a fuerza iónica y pH definidos, a la que el 50% de la secuencia diana se hibrida con una sonda con apareamiento perfecto) disminuye aproximadamente 1°C por % de apareamiento erróneo de bases.

Las "condiciones de hibridación rigurosas" y "condiciones de lavado de hibridación rigurosas" en el contexto de experimentos de hibridación de ácidos nucleicos tales como hibridaciones de tipo Southern y Northern dependen de la secuencia y son diferentes en diferentes parámetros ambientales. Por ejemplo, secuencias más largas se hibridan específicamente a mayores temperaturas. La especificidad es normalmente la función de los lavados tras la hibridación. Los factores críticos de tales lavados incluyen la fuerza iónica y la temperatura de la disolución de lavado final.

Generalmente, se selecciona que las condiciones rigurosas sean aproximadamente 50°C menores que el punto de fusión térmico (T_f) para la secuencia específica a una fuerza iónica y un pH definidos. La T_f depende de las condiciones de disolución y la composición de bases de la sonda, y puede calcularse usando la siguiente ecuación:

$$T_f = 79,8^\circ\text{C} + (18,5 \times \log[\text{Na}^+]) + (58,4^\circ\text{C} \times \% \text{ de } [\text{G}+\text{C}]) - (820 \times (\text{n.}^\circ \text{ de pb en dúplex})^{-1}) - (0,5 \times \% \text{ de formamida})$$

Se conocen en la técnica fórmulas alternativas, dependiendo de los tipos de híbridos, por ejemplo:

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

ES 2 403 414 T3

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

- Híbridos de ADN-ARN o ARN-ARN:

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

- Híbridos de oligo-ADN u oligo-ARN^d:

5 Para < 20 nucleótidos: $T_f = 2 (I_n)$

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

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

^b sólo es preciso para el % de GC en el intervalo del 30% al 75%.

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

10 ^d Oligo, oligonucleótido; I_n , longitud eficaz del cebador = $2 \times (\text{n.}^\circ \text{ de G/C}) + (\text{n.}^\circ \text{ de A/T})$.

Nota. Para cada 1% de formamida, se reduce la T_f en aproximadamente de 0,6 a 0,7°C, mientras que la presencia de urea 6 M reduce la T_f en aproximadamente 30°C

15 Condiciones rigurosas más preferidas son cuando la temperatura es de 20°C por debajo de la T_f , y las condiciones rigurosas más preferidas son cuando la temperatura es de 10°C por debajo de la T_f . también puede controlarse la unión no específica usando una cualquiera de varias técnicas conocidas tales como, por ejemplo, el bloqueo de la membrana con disoluciones que contienen proteína, adiciones de ADN, ARN heterólogo y SDS al tampón de hibridación, y tratamiento con ARNasa.

20 Las condiciones de lavado se realizan normalmente a o por debajo de la rigurosidad de hibridación. Generalmente, condiciones rigurosas adecuadas para ensayos de hibridación de ácidos nucleicos o procedimientos de detección de amplificación génica son tal como se expusieron anteriormente. También pueden seleccionarse condiciones más o menos rigurosas.

25 Para los fines de definir el nivel de rigurosidad, puede hacerse referencia convenientemente a Sambrook *et al.* (2001) *Molecular Cloning: a laboratory manual*, 3ª edición Cold Spring Harbor Laboratory Press, CSH, Nueva York o a *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989). Un ejemplo de condiciones de baja rigurosidad es SSC 4-6x / SDS al 0,1-0,5% p/v a 37-45°C durante 2-3 horas. Dependiendo de la fuente y la concentración del ácido nucleico implicado en la hibridación, pueden emplearse condiciones de rigurosidad alternativas tales como condiciones rigurosas del medio. Ejemplos de condiciones rigurosas del medio incluyen SSC 1-4x / SDS al 0,25% p/v a $\geq 45^\circ\text{C}$ durante 2-3 horas. Un ejemplo de las condiciones de alta rigurosidad incluye SSC 0,1-1x / SDS al 0,1% p/v a 60°C durante 1-3 horas. El experto conoce diversos parámetros que pueden alterarse durante la hibridación y el lavado y que o bien mantendrán o bien cambiarán las condiciones de rigurosidad. Por ejemplo, otra condición de hibridación rigurosa es la hibridación a SSC 4x a 65°C, seguida por un lavado con SSC 0,1x a 65°C durante aproximadamente una hora. Un ejemplo alternativo de condiciones de hibridación rigurosas es en formamida al 50%, SSC 4x a 42°C. Todavía otro ejemplo de condiciones rigurosas incluyen la hibridación a 62°C en SSC 6x, BLOTTO 0,05x y lavado a SSC 2x, SDS al 0,1% p/v a 62°C.

35 Los métodos según la presente invención también pueden ponerse en práctica usando una variante alternativa de corte y empalme de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1. El término "variante alternativa de corte y empalme" tal como se usa en el presente documento engloba variantes de una secuencia de ácido nucleico en la que se han cortado, reemplazado o añadido intrones y/o exones seleccionados. Tales variantes serán aquéllas en las que la actividad biológica de la proteína permanece sin afectar, lo que puede lograrse mediante la conservación selectiva del fragmento funcional de la proteína. Tales variantes de corte y empalme pueden encontrarse en la naturaleza o pueden ser artificiales. Se conocen bien en la técnica métodos para producir tales variantes de corte y empalme. Por tanto según otro aspecto de la presente invención, se proporciona, un método para mejorar las características de crecimiento de plantas, que comprende modular la expresión en una planta de una variante alternativa de corte y empalme de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 y/o modular la actividad y/o los niveles de una proteína de tipo NAP1 codificada por la variante alternativa de corte y empalme Preferiblemente, la variante de corte y empalme es una variante de corte y empalme de la secuencia representada por SEQ ID NO: 1. Preferiblemente las características de crecimiento mejoradas son el aumento del rendimiento, más preferiblemente aumento del rendimiento de semillas, lo más preferiblemente el aumento del rendimiento de semillas comprende aumento del índice de cosecha, aumento del número de semillas llenas y/o aumento del peso total de semillas.

Ventajosamente, los métodos según la presente invención también pueden ponerse en práctica usando variantes alélicas de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1, preferiblemente una variante alélica de una secuencia representada por SEQ ID NO: 1. Las variantes alélicas existen en la naturaleza y el uso de estos alelos naturales está englobado dentro de los métodos de la presente invención. Las variantes alélicas engloban polimorfismos de un solo nucleótido (SNP), así como polimorfismos de pequeña inserción/delección (INDEL). El tamaño de los INDEL es habitualmente inferior a 100 pb. Los SNP e INDEL forman la mayor proteína de tipo NAP1 de las variantes de secuencia en cepas polimórficas que se producen de manera natural de la mayoría de organismos. Son útiles en el mapeo de genes y el descubrimiento de genes y funciones génicas. Además son útiles en la identificación de loci genéticos, por ejemplo genes de plantas, implicados en la determinación de procesos tales como la velocidad de crecimiento, el tamaño de la planta y el rendimiento de la planta, el vigor de la planta, la resistencia a enfermedades, la tolerancia al estrés, etc.

La actividad de una proteína de tipo NAP1 o un homólogo de la misma también puede modularse (aumentarse o disminuirse) introduciendo una modificación genética (preferiblemente, en el locus de un gen de tipo NAP1). El locus de un gen tal como se define en el presente documento se toma como que significa una región genómica, que incluye el gen de interés y 10 KB en el sentido de 5' o de 3' de la región codificante. El término "modificación genética" se refiere a un cambio mediante intervención humana en el contenido genético de una célula en comparación con una célula de tipo natural e incluye técnicas como ingeniería genética, reproducción o mutagénesis. El cambio en el contenido genético comprende modificaciones del genoma e incluye la adición, delección y sustitución de material genético en los cromosomas de una célula vegetal así como en episomas. El término también engloba la adición de información extracromosómica a una célula vegetal. Preferiblemente, la modificación genética da como resultado la expresión modulada de un ácido nucleico de tipo NAP1, más preferiblemente la expresión aumentada de un ácido nucleico de tipo NAP1.

La modificación genética puede introducirse, por ejemplo, mediante uno cualquiera (o más) de los siguientes métodos: activación de ADNT, TILLING, mutagénesis dirigida al sitio, recombinación homóloga o introduciendo y expresando en una célula vegetal un ácido nucleico que codifica para una proteína de tipo NAP1 o un homólogo de la misma. Tras la introducción de la modificación genética, sigue una etapa de selección de actividad modulada de una proteína de tipo NAP1. Preferiblemente, la actividad modulada de una proteína de tipo NAP1 es un aumento de actividad, proporcionando el aumento de actividad plantas que tienen características de crecimiento mejoradas tales como aumento del rendimiento de semillas. La etapa de selección puede basarse en la monitorización de la presencia o ausencia de características de crecimiento mejoradas, o en la monitorización de la presencia o ausencia de genes marcadores seleccionables o examinables unidos a un ácido nucleico de interés introducido.

El marcaje con etiqueta de la activación de ADN-T (Hayashi *et al.* Science (1992) 1350-1353) implica la inserción de ADN-T que contiene habitualmente un promotor (también puede ser un potenciador de la traducción o un intrón), en la región genómica del gen de interés o 10 KB en el sentido de 5' o de 3' de la región codificante de un gen en una configuración tal que el promotor dirige la expresión del gen dirigido. Normalmente, la regulación de la expresión del gen dirigido por su promotor natural se altera y el gen cae bajo el control del promotor recién introducido. El promotor se incluye normalmente en un ADN-T. Este ADNT se inserta aleatoriamente en el genoma de la planta, por ejemplo, a través de infección por *Agrobacterium* y conduce a la sobreexpresión de genes cerca del ADN-T insertado. Las plantas transgénicas resultantes muestran fenotipos dominantes debido a la sobreexpresión de genes cerca del promotor introducido. El promotor que va a introducirse puede ser cualquier promotor que pueda dirigir la expresión de un gen en el organismo deseado, en este caso una planta. Por ejemplo, los promotores constitutivos, preferidos según el tejido, preferidos según el tipo celular e inducibles son todos adecuados para su uso en la activación de ADN-T.

También puede introducirse una modificación genética en el locus de un gen de tipo NAP1 usando la técnica de TILLING (*Targeted Induced Local Lesions IN Genomes*, lesiones locales inducidas de manera dirigida en genomas). Esto es una tecnología de mutagénesis útil para generar y/o identificar, y aislar eventualmente variantes mutagenizadas de un ácido nucleico de tipo NAP1 que codifica para un polipéptido que puede presentar actividad de tipo NAP1. TILLING también permite la selección de plantas que portan tales variantes mutantes. Estas variantes mutantes pueden presentar incluso mayor actividad de tipo NAP1 que la presentada por el gen en su forma natural. TILLING combina mutagénesis de alta densidad con métodos de selección de alto rendimiento. Las etapas que se siguen normalmente en TILLING son: (a) mutagénesis con EMS (Redei y Koncz, 1992; Feldmann *et al.*, 1994; Lightner y Caspar, 1998); (b) preparación y combinación de ADN y de individuos; (c) amplificación por PCR de una región de interés; (d) desnaturalización y apareamiento para permitir la formación de heterodúplex; (e) DHPLC, en la que se detecta la presencia de un heterodúplex en una combinación como un pico adicional en el cromatograma; (f) identificación del individuo mutante; y (g) secuenciación del producto de PCR mutante. Se conocen bien en la técnica métodos para TILLING (McCallum Nat Biotechnol. Abril de 2000; 18(4):455-7, revisado por Stemple 2004 (TILLING-a high-throughput harvest for functional genomics. Nat Rev Genet febrero de 2004;5(2):145-50.)).

Puede usarse mutagénesis dirigida al sitio para generar variantes de ácidos nucleicos de tipo NAP1 o partes de los mismos. Están disponibles varios métodos para lograr la mutagénesis dirigida al sitio, siendo los más comunes los métodos basados en PCR (current protocols in molecular biology. Wiley Eds.

<http://www.4ulr.com/products/currentprotocols/index.html>).

La activación de ADNT, TILLING y la mutagénesis dirigida al sitio son ejemplos de tecnologías que permiten la generación de variantes de tipo NAP1 y alelos novedosos.

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 ciencias biológicas para organismos inferiores tales como levadura o el musgo *Physcomitrella*. Se han descrito métodos para realizar recombinación homóloga en plantas no sólo para plantas modelo (Offringa *et al.* Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium-mediated transformation. 1990 EMBO J. octubre de 1990; 9(10):3077-84) sino también para plantas de cultivo, por ejemplo arroz (Terada R, Urawa H, Inagaki Y, Tsugane K, Iida S. Efficient gene targeting by homologous recombination in rice. Nat Biotechnol. 2002. Iida y Terada: A tale of two Integrations, transgene and T-DNA: gene targeting by homologous recombination in rice. Curr Opin Biotechnol. abril de 2004; 15(2):132-8). El ácido nucleico que va a dirigirse (que puede ser un ácido nucleico de tipo NAP1 o variante del mismo tal como se definió anteriormente en el presente documento) no es necesario que se dirija al locus de un gen de tipo NAP1, 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 reemplazar el gen endógeno o puede introducirse además del gen endógeno.

Las proteínas de tipo NAP1 tienen una organización de dominios típica, que consiste en un dominio NAP seguido por una región C-terminal ácida. Por tanto, se prevé que la modificación mediante ingeniería genética de los dominios de la proteína de tipo NAP1 de tal manera que la actividad de la proteína de tipo NAP1 se conserve o modifique, es útil para realizar los métodos de la invención. Un tipo preferido de variantes incluye las generadas mediante la delección de dominios, apilamiento o intercambio de ADN (véanse, por ejemplo, He *et al.*, Science 288, 2360-2363, 2000; o las patentes estadounidenses 5.811.238 y 6.395.547), siempre que la proteína de tipo NAP1 resultante comprenda un dominio NAP y una región C-terminal ácida de 20 a 25 aminoácidos que comprende al menos 13 residuos de ácido glutámico y/o aspártico. También puede usarse la evolución dirigida para generar variantes de ácidos nucleicos que codifican para una proteína de tipo NAP1. Esto consiste en iteraciones de intercambio de ADN seguido por el examen y/o la selección apropiados para generar variantes de ácidos nucleicos de tipo NAP1 o partes de los mismos que codifican para polipéptidos de tipo NAP1 u homólogos o partes de los mismos que tienen una actividad biológica modificada (Castle *et al.*, (2004) Science 304(5674): 1151-4; patentes estadounidenses 5.811.238 y 6.395.547).

Por consiguiente, como otro aspecto de la invención, se proporciona un método para mejorar el rendimiento de semillas en comparación con plantas de tipo natural correspondientes, que comprende modular la expresión, preferiblemente aumentar la expresión en una planta de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1, en el que la secuencia de ácido nucleico codifica para una proteína de tipo NAP1 tal como se define en la reivindicación 1.

Preferiblemente, el aumento del rendimiento de semillas comprende uno o más de aumento del índice de cosecha, aumento del número de semillas llenas y/o aumento del peso total de semillas.

En los métodos de la presente invención, se prevé el aumento de la expresión de un ácido nucleico introduciendo una secuencia de ácido nucleico que codifica para la proteína de tipo NAP1. Modular la expresión (aumentar o disminuir la expresión) de un ácido nucleico que codifica para una proteína de tipo NAP1 o modular la actividad y/o los niveles de la propia proteína de tipo NAP1 engloba la expresión alterada de un gen y/o la actividad alterada y/o los niveles de un producto génico, concretamente un polipéptido, en células o tejidos específicos. La expresión alterada (aumentada o disminuida) de un gen y/o la actividad alterada (aumentada o disminuida) y/o los niveles de un producto génico puede efectuarse, por ejemplo con medios recombinantes. Modular la expresión de un gen y/o los niveles de un producto génico y/o modular la actividad de un producto génico puede efectuarse directamente a través de la modulación de la expresión de un gen que codifica para una proteína de tipo NAP1 y/o directamente a través de la modulación de la actividad y/o los niveles de una proteína de tipo NAP1. La expresión modulada puede resultar de los niveles de la expresión alterada de un gen endógeno de tipo NAP1 y/o puede resultar de la expresión modulada de una proteína de tipo NAP1 que codifica para ácido nucleico que se introdujo previamente en una planta. De manera similar, los niveles y/o la actividad modulados de una proteína de tipo NAP1 pueden ser el resultado de niveles de expresión alterados de un gen endógeno de tipo NAP1 y/o puede resultar de la expresión alterada de una proteína de tipo NAP1 que codifica para ácido nucleico que se introdujo previamente en una planta. Adicional o alternativamente, la modulación de la expresión tal como se mencionó anteriormente se efectúa de manera indirecta, por ejemplo puede efectuarse como resultado de una disminución o un aumento de los niveles y/o la actividad de factores que controlan la expresión de un gen de tipo NAP1 o que influye en la actividad y/o los niveles de la proteína de tipo NAP1.

Según una realización preferida de la presente invención, el aumento de la expresión de un ácido nucleico que codifica para una proteína de tipo NAP1 y/o la modulación de la actividad y/o los niveles de la propia proteína de tipo NAP1 se efectúa mediante medios recombinantes. Tales medios recombinantes pueden comprender un enfoque

directo y/o indirecto para la modulación de la expresión de un ácido nucleico y/o para la modulación de la actividad y/o los niveles de una proteína.

Un enfoque directo y preferido para modular la expresión de un gen de tipo NAP1 o modular la actividad y/o los niveles de una proteína de tipo NAP1, comprende introducir en una planta una secuencia aislada de ácido nucleico que codifica para una proteína de tipo NAP1 tal como se define en la reivindicación 1. El ácido nucleico puede introducirse en una planta mediante, por ejemplo, transformación. Por tanto, según un aspecto preferido de la presente invención, se proporciona un método para aumentar el rendimiento de semillas de una planta que comprende una modificación genética de la planta, comprendiendo la modificación genética introducir un ácido nucleico que codifica para una proteína de tipo NAP1 en una planta. El aumento del rendimiento de la planta es aumento del rendimiento de semillas, más preferiblemente comprende uno o más de aumento del índice de cosecha, aumento del número de semillas llenas o aumento del peso total de semillas.

Según un aspecto preferido de la presente invención, se prevé la potenciación o el aumento de la expresión de un ácido nucleico. Están bien documentados en la técnica métodos para obtener la potenciación o el aumento de la expresión de genes o productos génicos e incluyen, por ejemplo, sobreexpresión dirigida por un promotor adecuado (preferiblemente, fuerte), el uso de potenciadores de la transcripción o potenciadores de la traducción. 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 el sentido codificante con respecto al promotor al que está unido operativamente. Preferiblemente, el ácido nucleico que va a sobreexpresarse codifica para una proteína de tipo NAP1 definida en la reivindicación 1, además preferiblemente la secuencia de ácido nucleico que codifica para la proteína de tipo NAP1 se aísla de una planta dicotiledónea, preferiblemente de la familia *Brassicaceae*, además preferiblemente la secuencia se aísla de *Arabidopsis thaliana*, lo más preferiblemente la secuencia de ácido nucleico es tal como se representa por SEQ ID NO: 1 o una parte de la misma, o codifica para una secuencia de aminoácidos tal como se representa por SEQ ID NO: 2 o un homólogo, derivado o fragmento activo de la misma. Alternativamente, la secuencia de ácido nucleico que codifica para la proteína de tipo NAP1 es tal como se representa en SEQ ID NO: 20 (número de registro GenBank NM101738) o es una parte de la misma, o codifica para una secuencia de aminoácidos representada en SEQ ID NO: 21 (número de registro GenBank NP564063) o codifica para un homólogo, derivado o fragmento activo de la misma. Debe indicarse que la aplicabilidad de la invención no se basa en el uso del ácido nucleico representado por SEQ ID NO: 1, ni en 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 que codifican para homólogos, derivados o fragmentos activos de SEQ ID NO: 2, o partes de SEQ ID NO: 1, o secuencias que se hibridan con SEQ ID NO: 1, en los métodos de la presente invención. En particular, también son útiles homólogos de otras especies tales como tabaco (SEQ ID NO: 7 ó 9), maíz (SEQ ID NO: 13 ó 19), *Medicago sativa* (SEQ ID NO: 11), tomate (SEQ ID NO: 23) o arroz (SEQ ID NO: 15 ó 17) en los métodos de la presente invención.

La presente invención se refiere a métodos para mejorar características de crecimiento de una planta o a métodos para producir plantas con características de crecimiento mejoradas, en los que las características de crecimiento incluyen aumento del rendimiento, que comprende uno cualquiera o más de: aumento del número de vástagos, aumento del número de primeras panículas (que son la panícula más alta y todas las panículas que se solapan con la panícula más alta cuando se alinean en vertical), aumento del número de segundas panículas, aumento del número total de semillas, aumento del número de semillas llenas, aumento del peso de semilla total por planta, aumento del índice de cosecha, aumento del peso de mil granos. La presente invención también proporciona métodos para mejorar una de las características de crecimiento mencionadas anteriormente, sin provocar una penalización en una de las demás características de crecimiento, por ejemplo aumento del número de semillas llenas cuando se conserva el mismo número de espiguillas por panícula.

El término "aumento del rendimiento" engloba un aumento en la biomasa en una o más partes de una planta con relación a la(s) parte(s) correspondiente(s) de plantas de tipo natural. El término engloba un aumento en el rendimiento de semillas, incluyendo un aumento en la biomasa de la semilla (peso de semilla) y/o un aumento en el número de semillas (llenar) y/o en el tamaño de las semillas y/o un aumento en el volumen de semilla, cada uno con relación a plantas de tipo natural correspondientes. Dependiendo del cultivo, las partes de la planta en cuestión pueden ser biomasa por encima del suelo (por ejemplo, maíz, cuando se usa como ensilado, caña de azúcar), raíces (por ejemplo, remolacha azucarera), frutos (por ejemplo, tomate), fibras de algodón, o cualquier otra parte de la planta que sea de valor económico. Tomando el arroz como ejemplo, un aumento de rendimiento puede manifestarse mediante un aumento en uno o más de los siguientes: número de panículas por planta, número de espiguillas por panícula, número de flores por panícula, aumento en la tasa de llenado de semillas, aumento en el peso de mil granos, entre otros. Para el maíz, el aumento del rendimiento de semillas puede reflejarse, por ejemplo, en un aumento de filas (de semillas) por mazorca y/o un aumento del número de granos por fila. Un aumento en el tamaño y/o volumen de semilla también puede influir en la composición de semillas. Un aumento en el rendimiento de semillas podría deberse a un aumento en el número y/o el tamaño de las flores. Un aumento en el rendimiento también podría aumentar el índice de cosecha, que se expresa como una razón del rendimiento de partes cosechables, tales como semillas, con respecto a la biomasa total; o el peso de mil granos. El aumento del

rendimiento también engloba la capacidad para la plantación a mayor densidad (número de plantas por hectárea o acre).

También la división celular modificada puede contribuir al aumento de rendimiento. El término "división celular modificada" engloba un aumento o una disminución en la división celular o una citocinesis/división celular anómala, plano de división alterado, polaridad celular alterada, diferenciación celular alterada. El término también comprende fenómenos tales como endomitosis, acitocinesis, poliploidía, politenia y endorreduplicación.

Puede preverse que plantas que tienen un aumento del rendimiento también presenten una velocidad de crecimiento modificada en comparación con plantas de tipo natural correspondientes. El término "velocidad de crecimiento modificada" tal como se usa en el presente documento engloba, pero no se limita a, una velocidad de crecimiento más rápida en una o más partes de una planta (incluyendo semillas), en una o más etapas en el ciclo vital de una planta. Las plantas con crecimiento mejorado pueden mostrar una curva de crecimiento modificada y pueden tener valores modificados para su T_{mid} o T_{90} (respectivamente el tiempo necesario para alcanzar la mitad de su tamaño máximo o el 90% de su tamaño máximo, cada uno con relación a plantas de tipo natural correspondientes). El término "crecimiento mejorado" engloba vigor potenciado, floración más temprana, tiempo de ciclación modificado.

Según una característica preferida de la presente invención, la realización de los métodos según la presente invención da como resultado plantas que tienen un aumento del rendimiento, en particular plantas que tienen aumento del rendimiento de semillas. Preferiblemente, el aumento del rendimiento de semillas incluye al menos un aumento en uno cualquiera o más del número de semillas llenas, el peso total de semillas y el índice de cosecha, cada uno con relación a plantas control. Por tanto, según la presente invención, se proporciona un método para aumentar el rendimiento de plantas, comprendiendo el método modular la expresión de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 definida en la reivindicación 1 en una planta. Preferiblemente cuando la proteína de tipo NAP1 está codificada por una secuencia de ácido nucleico representada por SEQ ID NO: 1 o una parte de la misma o por secuencias que pueden hibridarse con la misma o cuando la proteína de tipo NAP1 está representada por SEQ ID NO: 2 o es un homólogo, derivado o fragmento activo de la misma. Alternativamente, la proteína de tipo NAP1 puede estar codificada por secuencias de ácido nucleico representadas en SEQ ID NO: 20 (número de registro GenBank NM101738), SEQ ID NO: 6, 8, 10, 12, 14, 16, 18 ó 22, o la proteína de tipo NAP1 puede ser tal como se representa en SEQ ID NO: 21 (número de registro GenBank NP564063), SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 23.

Los métodos de la presente invención son favorables para aplicarse a plantas de cultivo porque los métodos de la presente invención se usan para aumentar uno o más del peso total de semillas, el número de semillas llenas y el índice de cosecha de una planta. Por tanto, los métodos de la presente invención son particularmente útiles para plantas de cultivo cultivadas para sus semillas, tales como cereales, girasol, soja, guisante, lino, altramuces, canola. Por consiguiente, una realización particular de la presente invención se refiere a un método para aumentar el rendimiento de semillas (aumento del peso total de semillas, aumento del número de semillas llenas y/o para aumentar el índice de cosecha) de un cereal.

Según una realización adicional, se proporcionan constructos genéticos y vectores para facilitar la introducción y/o la expresión de las secuencias de nucleótidos útiles en los métodos. Por tanto, según una segunda realización, se proporciona un constructo génico para la expresión en una planta, que comprende:

(i) una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1;

(ii) una o más secuencias de control que pueden dirigir la expresión en una planta de la secuencia de ácido nucleico de (i); y opcionalmente

(iii) una secuencia de terminación de la transcripción.

Pueden crearse constructos útiles en los métodos según la presente invención usando tecnología de ADN recombinante bien conocida por los expertos en la técnica. Los constructos génicos pueden insertarse en vectores, que pueden estar disponibles comercialmente, ser adecuados para la transformación en plantas y adecuados para la expresión del gen de interés en las células transformadas. El constructo genético puede ser un vector de expresión en el que la secuencia de ácido nucleico está unida operativamente a una o más secuencias de control que permiten la expresión en células huésped procariontas y/o eucariotas.

Según una realización preferida de la invención, el constructo genético es un vector de expresión diseñado para sobreexpresar la secuencia de ácido nucleico. La secuencia de ácido nucleico que puede modular la expresión de un ácido nucleico que codifica para una proteína de tipo NAP1 es una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 tal como se define en la reivindicación 1. Una secuencia de ácido nucleico preferida es la secuencia representada por SEQ ID NO: 1 o una parte de la misma o secuencias que pueden hibridarse con la misma en condiciones rigurosas o una secuencia de ácido nucleico que codifica para una proteína representada por

SEQ ID NO: 2. Preferiblemente, este ácido nucleico se clona en la orientación sentido con relación a la secuencia de control a la que está unido operativamente.

Se transforman 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 de tipo NAP1), estando la secuencia unida operativamente a una o más secuencias de control (al menos un promotor). Los términos “elemento regulador”, “secuencia de control” y “promotor” se usan todos en el presente documento de manera intercambiable y han de tomarse en un contexto amplio para referirse a secuencias reguladoras de ácido nucleico que pueden efectuar la expresión de las secuencias a las que se ligan. Están englobadas por los términos mencionados anteriormente, secuencias reguladoras de la transcripción derivadas de un gen genómico eucariota clásico (incluyendo la caja TATA que se requiere para la iniciación de la transcripción exacta, con o sin una secuencia de caja CCAAT) y elementos reguladores adicionales (es decir, secuencias de activación en el sentido de 5', potenciadores y silenciadores) que alteran la expresión génica en respuesta a estímulos de desarrollo y/o externos, o de manera específica del tejido. También está incluida dentro del término una secuencia reguladora de la transcripción de un gen procarionta clásico, en cuyo caso puede incluir una secuencia de caja -35 y/o secuencias reguladoras de la transcripción de caja -10. El término “elemento regulador” también engloba una molécula de fusión sintética o un derivado que confiere, activa o potencia la expresión de una molécula de ácido nucleico en una célula, un tejido o un órgano. El término “unido operativamente” tal como se usa en el presente documento se refiere a una unión funcional entre la secuencia promotora y el gen de interés, de manera 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. Preferiblemente, la secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 está unida operativamente a un promotor constitutivo. El término “constitutivo” tal como se define en el presente documento se refiere a un promotor que se expresa predominantemente en al menos un tejido u órgano y predominantemente en cualquier etapa vital de la planta. Preferiblemente, el promotor se expresa predominantemente en la totalidad de la planta. Preferiblemente, el promotor constitutivo es el promotor GOS2 del arroz, o un promotor de fuerza similar y/o un promotor con un patrón de expresión similar. Alternativamente, pueden usarse promotores específicos de tejido. Por ejemplo, en los casos en los que se prevé el aumento del rendimiento de semillas, puede contemplarse el uso de promotores preferidos para semillas, preferidos para flores, promotores preferidos para meristemo o promotores activos en células en división. Pueden analizarse la fuerza del promotor y/o el patrón de expresión por ejemplo acoplando el promotor a un gen indicador y someter 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. Se presentan ejemplos de promotores alternativos con su patrón de expresión respectivo en la tabla 2, y estos promotores o derivados de los mismos pueden ser útiles para los métodos de la presente invención.

Tabla 2: Ejemplos de promotores para su uso en la realización de la presente invención

Fuente del gen	Patrón de expresión	Bibliografía
Actina	constitutivo	McElroy <i>et al</i> , Plant Cell, 2: 163-171, 1990
35S de VMCo	constitutivo	Odell <i>et al</i> , Nature, 313: 810-812, 1985
19S de VMCo	constitutivo	Nilsson <i>et al.</i> , Physiol. Plant. 100: 456-462, 1997
GOS2	constitutivo	de Pater <i>et al</i> , Plant J Nov;2(6):837-44, 1992
ubiquitina	constitutivo	Christensen <i>et al</i> , Plant Mol. Biol. 18: 675-689, 1992
ciclofilina del arroz	constitutivo	Buchholz <i>et al</i> . Plant Mol Biol. 25(5): 837-43, 1994

ES 2 403 414 T3

(continuación)

Fuente del gen	Patrón de expresión	Bibliografía
histona H3 del maíz	constitutivo	Lepetit <i>et al</i> , Mol. Gen. Genet 231: 276-285, 1992
actina 2	constitutivo	An <i>et al</i> , Plant J. 10(1); 107-121, 1996
genes específicos de semilla	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.
legumina	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 LMW y HMW del trigo	endospermo	Mol Gen Genet 216:81-90, 1989; NAR 17:461-2, 1989
SPA del trigo	semilla	Albani <i>et al</i> , Plant Cell, 9: 171-184, 1997
α , β , γ -gliadinas del trigo	endospermo	EMBO 3:1409-15, 1984
promotor de Itr1 de la cebada	endospermo	
hordeína B1, C, D de la cebada	endospermo	Theor Appl Gen 98:1253-62, 1999; Plant J 4:343-55,1993;MolGen Genet 250:750-60, 1996
DOF de la cebada	endospermo	Mena <i>et al</i> , The Plant Journal, 116(1): 53-62, 1998
blz2	endospermo	Documento EP99106056.7
promotor sintético	endospermo	Vicente-Carbajosa <i>et al.</i> , Plant J. 13: 629-640, 1998.
prolamina NRP33 del arroz	endospermo	Wu <i>et al</i> , Plant Cell Physiology 39(8) 885-889, 1998

ES 2 403 414 T3

(continuación)

Fuente del gen	Patrón de expresión	Bibliografía
α -globulina Glb-1 del arroz	endospermo	Wu <i>et al.</i> , Plant Cell Physiology 39(8) 885-889, 1998
OSH1 del arroz	embrión	Sato <i>et al.</i> , Proc. Natl. Acad. Sci. USA, 93: 8117-8122, 1996
α -globulina REB/OHP-1 del arroz	endospermo	Nakase <i>et al.</i> Plant Mol. Biol. 33: 513-522, 1997
ADP-glucosa PP del arroz	endospermo	Trans Res 6:157-68, 1997
familia de genes ESR del maíz	endospermo	Plant J 12:235-46, 1997
γ -kafirina del sorgo	endospermo	PMB 32:1029-35, 1996
KNOX	embrión	Postma-Haarsma <i>et al.</i> , PlantMol. Biol. 39:257-71, 1999
oleosina del arroz	embrión y aleurona	Wu et at, J. Biochem., 123:386, 1998
oleosina del girasol	semilla (embrión y semilla seca)	Cummins, <i>et al.</i> , Plant Mol. Biol. 19: 873-876,1992
AtPRP4	flores	http://salus.medium.edu/mmg/tierney/html
chalcona sintasa (chsA)	flores	Van der Meer, <i>et al.</i> , Plant Mol. Biol. 15, 95-109, 1990.
LEAFY	meristemo de brote	Weigel <i>et al.</i> , Cell 69:843-859, 1992.
knat1 de <i>Arabidopsis thaliana</i>	meristemo de brote	Número de registro AJ131822
kn1 de <i>Malus domestica</i>	meristemo de brote	Número de registro Z71981
CLAVATA1	meristemo de brote	Número de registro AF049870
Ciclina B1;1 del tabaco (<i>N. sylvestris</i>)	Células en división /tejido meristemático	Trehin <i>et al.</i> 1997 Plant Mol.Biol. 35, 667-672.
Ciclina mitótica CYS (tipo A) y CYM (tipo B) de <i>Catharanthus roseus</i>	Células en división /tejido meristemático	Ito <i>et al.</i> 1997 Plant J. 11, 983-992
cyc1At (=cyc B1;1) y cyc3aAt (tipo A) de <i>Arabidopsis</i>	Células en división /tejido meristemático	Shaul <i>et al.</i> 1996 Proc.Natl.Acad.Sci.U.S.A 93,4868-4872.
Caja del promotor tef1 de <i>Arabidopsis</i>	Células en división /tejido meristemático	Regad <i>et al.</i> 1995 Mol.Gen.Genet 248, 703-711.
cyc07 de <i>Catharanthus roseus</i>	Células en división /tejido meristemático	Ito <i>et al.</i> 1994 Plant Mol. Biol. 24, 863-878.

Opcionalmente, también pueden usarse una o más secuencias terminadoras en el constructo introducido en una

planta. El término “terminador” engloba una secuencia de control que es una secuencia de ADN en el extremo de una unidad de transcripción que señala el procesamiento en 3' y la poliadenilación de un transcrito primario y la terminación de la transcripción. Elementos reguladores adicionales pueden incluir potenciadores transcripcionales así como traducionales. Los expertos en la técnica conocerán secuencias terminadoras y potenciadoras que pueden ser adecuadas para su uso en la realización de la invención. Tales secuencias las conocería o puede obtenerlas fácilmente un experto en la técnica.

Los constructos genéticos de la invención pueden incluir además una secuencia de origen de replicación que se requiere para el mantenimiento y/o la replicación en un tipo celular específico. Un ejemplo es cuando se requiere que se mantenga un constructo genético en una célula bacteriana como un elemento genético episómico (por ejemplo, molécula de plásmido o cósmido). Los orígenes de replicación preferidos incluyen, pero no se limitan a, f1-ori y colE1.

El constructo genético puede comprender opcionalmente un gen marcador seleccionable. Tal como se usa en el presente documento, el término “gen marcador seleccionable” incluye cualquier gen que confiera un fenotipo a una célula en la que se expresa para facilitar la identificación y/o selección de células que se transfectan o transforman con un constructo de ácido nucleico de la invención. Pueden seleccionarse marcadores adecuados de marcadores que confieren resistencia a antibióticos o herbicidas, que introducen un nuevo rasgo metabólico o que permiten la selección visual. Los ejemplos de genes marcadores seleccionables incluyen genes que confieren resistencia a antibióticos (tales como *npfII* que fosforila neomicina y kanamicina, o *hpt*, que fosforila higromicina), a herbicidas (por ejemplo, *bar* que proporciona resistencia a Basta; *aroA* o *gox* que proporciona resistencia frente a glifosato), o genes que proporcionan un rasgo metabólico (tales como *manA* que permite que las plantas usen manosa como única fuente de carbono). Los genes marcadores visuales dan como resultado la formación de color (por ejemplo, β -glucuronidasa, GUS), luminiscencia (tales como luciferasa) o fluorescencia (proteína fluorescente verde, GFP, y derivados de la misma).

En una realización preferida, el constructo genético tal como se mencionó anteriormente, comprende un gen de tipo NAP1 en orientación sentido acoplado a un promotor que es preferiblemente un promotor constitutivo, tal como el promotor GOS2 del arroz. Por tanto, otro aspecto de la presente invención es un constructo de vector que porta un casete de expresión esencialmente similar a SEQ ID NO: 3, que comprende el promotor GOS2 del arroz, el gen de tipo NAP1 de *Arabidopsis* y la zeína T + secuencia de terminación de la transcripción T-rubisco deltaGA. Una secuencia esencialmente similar a SEQ ID NO: 3 engloba un ácido nucleico que codifica para una proteína homóloga a SEQ ID NO: 2 o que se hibrida con SEQ ID NO: 1, estando el ácido nucleico unido operativamente a un promotor GOS2 del arroz o un promotor con un patrón de expresión similar y/o uniéndose el ácido nucleico a una secuencia de terminación de la transcripción.

Por tanto según otro aspecto, se proporciona un constructo génico, que comprende un casete de expresión en el que se ubica una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1, elegida del grupo que comprende:

(i) una secuencia de ácido nucleico representada por SEQ ID NO: 1 o la cadena complementaria de la misma;

(ii) una secuencia de ácido nucleico que codifica para una secuencia de aminoácidos representada por SEQ ID NO: 2 u homólogos, derivados o fragmentos activos de la misma;

(iii) una secuencia de ácido nucleico que puede hibridarse (preferiblemente en condiciones rigurosas) con una secuencia de ácido nucleico de (i) o (ii) anteriormente, codificando preferiblemente la secuencia de hibridación para una proteína que tienen actividad de proteína de tipo NAP1;

(iv) una secuencia de ácido nucleico según (i) a (iii) anteriormente que es degenerada como resultado del código genético;

(v) secuencia de ácido nucleico que es una variante alélica para las secuencias de ácido nucleico según (i) a (iii);

(vi) secuencia de ácido nucleico que es una variante alternativa de corte y empalme para las secuencias de ácido nucleico según (i) a (iii);

plantas que pueden obtenerse mediante los métodos según la presente invención. Las plantas que pueden obtenerse mediante los métodos de la presente invención, teniendo las plantas un aumento del rendimiento y teniendo las plantas actividad y/o niveles de proteína de tipo NAP1 modulados y/o expresión modulada de un ácido nucleico que codifica para una proteína de tipo NAP1. Preferiblemente, las plantas son plantas transgénicas que comprenden una secuencia aislada de ácido nucleico que codifica para una proteína de tipo NAP1, caracterizadas porque la planta transgénica se ha seleccionado porque tienen un aumento del rendimiento. Además preferiblemente, la planta transgénica se ha seleccionado por la expresión modulada de un ácido nucleico que

codifica para una proteína de tipo NAP1. Además preferiblemente, la planta transgénica se ha seleccionado por la expresión modulada de un ácido nucleico que codifica para una proteína de tipo NAP1 representada por SEQ ID NO: 2

5 Según una tercera realización de la presente invención, se proporciona un método para la producción de plantas transgénicas que tienen características de crecimiento mejoradas, que comprende introducción y la expresión en una planta de un ácido nucleico que codifica para una proteína de tipo NAP1 definida en la reivindicación 1. Las características de crecimiento mejoradas comprenden el aumento del rendimiento, más preferiblemente aumento del rendimiento de semillas, lo más preferiblemente que comprende uno o más de aumento del número de semillas llenas, aumento del índice de cosecha o aumento del peso total de semillas.

10 Se da a conocer un método para la producción de plantas transgénicas que tienen un aumento del rendimiento, comprendiendo el método:

(i) introducir en una célula vegetal una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1;

(ii) regenerar y/o hacer crecer una planta a partir de una célula de planta transgénica.

15 La propia proteína de tipo NAP1 definida en la reivindicación 1 y/o el propio ácido nucleico de tipo NAP1 tal como se define en la reivindicación 1 pueden introducirse directamente en una célula vegetal o en la propia planta (incluyendo la 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 una parte de la misma o secuencias que pueden hibridarse con el mismo, o es un ácido nucleico que codifica para una secuencia de aminoácidos representada por SEQ ID NO: 2 o un homólogo, derivado o fragmento activo del mismo. Alternativamente, la secuencia de ácido nucleico es tal como se representa en SEQ ID NO: 20 (número de registro GenBank NM101738), SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 22 o una parte de las mismas o secuencias que pueden hibridarse con cualquiera de las secuencias mencionadas anteriormente. La secuencia de aminoácidos puede ser alternativamente una secuencia representada en SEQ ID NO: 21 (número de registro GenBank NP564063), SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 23 u homólogos, derivados o fragmentos activos de las mismas.

20 El término "transformación" tal como se hace referencia en el presente documento engloba la transferencia de un polinucleótido exógeno a una célula huésped, independientemente del método usado para la transferencia. Puede transformarse tejido de la planta que puede dar una propagación clonal posterior, ya sea mediante organogénesis o embriogénesis, con un constructo genético de la presente invención y regenerarse a partir del mismo una planta completa. El tejido particular elegido variará dependiendo de los sistemas de propagación clonal disponibles y más idóneos para la especie particular que esté transformándose. Las dianas tisulares a modo de ejemplo incluyen discos foliares, polen, embriones, cotiledones, hipocótilos, megagametofitos, tejido calloso, tejido meristemático existente (por ejemplo, meristemo apical, yemas axilares y meristemas de raíz), 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 no integrado, por ejemplo, como un plásmido. Alternativamente, puede integrarse en el genoma del huésped. Entonces puede usarse la célula vegetal transformada resultante para regenerar una planta transformada de manera conocida por los expertos en la técnica.

30 La transformación de una especie de planta es ahora una técnica bastante rutinaria. Ventajosamente, puede usarse cualquiera de varios métodos de transformación para introducir el gen de interés en una célula ancestral 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. Pueden seleccionarse los métodos del método con calcio/poli-etilenglicol para protoplastos (Krens, F.A. *et al.*, 1882, 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 ADN o ARN (Klein T.M. *et al.*, 1987, Nature 327, 70) infección con virus (no integradores) y similares. Se producen preferiblemente plantas de arroz transgénicas que expresan un gen de tipo NAP1 a través de una transformación mediada por *Agrobacterium* usando cualquiera de los métodos bien conocidos para la transformación de arroz, tal como los descritos en cualquiera de los siguientes: solicitud de patente europea EP 1198985 A1, 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), incorporándose las descripciones como referencia al presente documento como si se expusieran totalmente. En el caso de la transformación de maíz, el método preferido es tal como se describe o bien en Ishida *et al.* (Nat. Biotechnol. junio de 1996; 14(6): 745-50) o bien en Frame *et al.* (Plant Physiol. mayo de 2002; 129(1): 13-22), incorporándose las descripciones como referencia al presente documento como si se expusieran totalmente.

55 Generalmente tras la transformación, se seleccionan células vegetales o agrupaciones de células por la presencia de uno o más marcadores que están codificados por genes que pueden expresarse en plantas transferidos

conjuntamente con el gen de interés, tras lo cual se regenera el material transformado para dar una planta completa.

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

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

Los organismos transformados generados pueden adoptar una variedad de formas. Por ejemplo, pueden ser quimeras de células transformadas y células no transformadas; transformantes clonales (por ejemplo, para que todas las células transformadas contengan el casete de expresión); injertos de tejidos transformados y sin transformar (por ejemplo, en plantas, un portainjerto transformado injertado a una púa sin transformar).

Se da a conocer una célula vegetal o planta producida mediante cualquiera de los métodos descritos en el presente documento, y para todas las partes de la planta y propágulos de la misma. La presente invención se extiende además para englobar la progenie de una célula, tejido, órgano o planta completa primario transformado o transfectado que se ha producido mediante cualquiera de los métodos mencionados anteriormente, siendo el único requisito que la progenie presente la(s) misma(s) característica(s) genotípica(s) y/o fenotípica(s) que la(s) producida(s) en el progenitor mediante los métodos según la invención. La invención también incluye células huésped que contienen una molécula de ácido nucleico aislada que codifica para una proteína que puede modular los niveles y/o la actividad de una proteína de tipo NAP1, preferiblemente en las que la proteína es una proteína de tipo NAP1. Células huésped preferidas según la invención son células vegetales. Por tanto, la invención también engloba células huésped o plantas transgénicas que tienen características de crecimiento alteradas, caracterizadas porque la célula huésped o planta transgénica tiene la expresión modulada de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 y/o actividad y/o nivel modulados de una proteína de tipo NAP1. Preferiblemente, las características de crecimiento alteradas comprenden aumento del rendimiento, más preferiblemente aumento del rendimiento de semillas.

Se dan a conocer partes cosechables de una planta tales como, pero sin limitarse a, semillas, hojas, frutos, flores, tallos o cultivos de tallos, rizomas, raíces, tubérculos y bulbos. La invención se refiere además a productos derivados directamente de una parte cosechable de una planta de este tipo, tales como polvos o gránulos secos, aceite, grasa y ácidos grasos, almidón o proteínas.

El término "planta" tal como se usa en el presente documento engloba plantas completas, ancestros y progenie de las plantas y partes de las plantas, células, tejidos y órganos vegetales. El término "planta" también engloba, 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 microesporas. 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 forraje o leguminosas forrajeras, plantas ornamentales, cultivos de alimentos, 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., *Omithopus* 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., *Pnmus* 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., *Triticoseca/e rimpaul*, *Triticum* spp., *Vaccinium* spp., *Vicia* spp., *Vigna* spp., *Vitis* spp., *Zea mays*, *Zizania palustris*, *Ziziphus* spp., entre otros.

Según una característica preferida de la presente invención, la planta es una planta de cultivo que comprende soja, girasol, canola, alfalfa, colza o algodón. Además preferiblemente, 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, avenas, sorgo.

- 5 La presente invención también engloba el uso de ácidos nucleicos que codifican para una proteína de tipo NAP1, tal como se define en la reivindicación 1 para aumentar el rendimiento de semillas tal como se expone en las reivindicaciones 5 y 6.

10 El uso se refiere a la mejora del rendimiento de semillas. El rendimiento de semillas puede incluir uno o más de los siguientes: aumento del número de semillas, aumento del número de semillas llenas, aumento del peso total de semillas, aumento del índice de cosecha, aumento del peso de mil granos, tasa de llenado de semillas, entre otros. Preferiblemente, la proteína de tipo NAP1 o el ácido nucleico que codifica para una proteína de tipo NAP1 es de origen vegetal, más preferiblemente de una planta dicotiledónea, además preferiblemente de la familia de *Brassicaceae*, lo más preferiblemente, la proteína de tipo NAP1 está codificada por SEQ ID NO: 1 o es tal como se representa por SEQ ID NO: 2.

15 Los ácidos nucleicos de tipo NAP1 o variantes de los mismos o polipéptidos de tipo NAP1 u homólogos de los mismos pueden encontrar uso en programas de reproducción en los que se identifica un marcador de ADN que puede estar unido genéticamente a un gen de tipo NAP1 o variante del mismo. El gen de tipo NAP1 o variantes del mismo o proteína de tipo NAP1 u homólogos de la misma pueden usarse para definir un marcador molecular. Este marcador de proteína o ADN puede usarse entonces en programas de reproducción para seleccionar plantas que
20 tienen características de crecimiento mejoradas. El gen de tipo NAP1 o variante del mismo puede ser, por ejemplo, un ácido nucleico tal como se representa por SEQ ID NO: 1, o un ácido nucleico que codifica para cualquiera de los homólogos mencionados anteriormente.

25 Variantes alélicas de un gen que codifica para una proteína de tipo NAP1 también pueden encontrar uso en programas de reproducción asistidos por marcadores. Tales programas de reproducción requieren a veces la introducción de una variación alélica mediante tratamiento mutagénico de las plantas, usando por ejemplo mutagénesis con EMS; alternativamente, el programa puede iniciarse con una colección de variantes alélicas de las denominadas de origen "natural" producidas involuntariamente. Entonces tiene lugar la identificación de variantes alélicas mediante, por ejemplo, PCR. Esto va seguido por una etapa de selección para la selección de variantes alélicas superiores de la secuencia en cuestión y que dan lugar a características de crecimiento mejoradas en una
30 planta, tales como aumento del rendimiento de semillas. La selección se lleva a cabo normalmente monitorizando el rendimiento de crecimiento de plantas que contienen diferentes variantes alélicas de la secuencia en cuestión, por ejemplo, diferentes variantes alélicas de SEQ ID NO: 1, o de ácidos nucleicos que codifican para cualquiera de los homólogos de planta mencionados anteriormente. El rendimiento de crecimiento puede monitorizarse en un invernadero o en el campo. Etapas opcionales adicionales incluyen el cruzamiento de plantas, en las que se
35 identificó la variante alélica superior que daba como resultado un aumento del rendimiento de semillas, con otra planta. Esto podría usarse, por ejemplo, para preparar una combinación de características fenotípicas interesantes.

40 Un ácido nucleico de tipo NAP1 o variante del mismo también puede usarse como sondas para mapear genética y físicamente los genes de los que son parte, y como marcadores para rasgos relacionados con esos genes. Tal información puede ser útil en la reproducción de plantas con el fin de desarrollar líneas con fenotipos deseados. Tal uso de ácidos nucleicos de tipo NAP1 o variantes de los mismos sólo requiere una secuencia de ácido nucleico de al menos 10 nucleótidos de longitud. Los ácidos nucleicos de tipo NAP1 o variantes de los mismos pueden usarse como marcadores de polimorfismos de longitud de fragmento de restricción (RFLP). Pueden tratarse con sonda las transferencias de tipo Southern de ADN genómico de planta digerido por restricción con los ácidos nucleicos de tipo NAP1 o variantes de los mismos. Los patrones de bandas resultantes pueden someterse entonces a análisis
45 genéticos usando programas informáticos tales como MapMaker (Lander *et al.* (1987) *Genomics* 1, 174-181) con el fin de construir un mapa genético. Además, los ácidos nucleicos pueden usarse para tratar con sonda transferencias de tipo Southern que contienen ADN genómicos tratados con endonucleasa de restricción de un conjunto de individuos que representan los progenitores y la progenie de un cruzamiento genético definido. Se observa segregación de los polimorfismos de ADN y se usa para calcular la posición del ácido nucleico de tipo NAP1 o variante del mismo en el mapa genético obtenido previamente usando esta población (Botstein *et al.* (1980) *Am. J. Hum. Genet* 32, 314-331).
50

55 La producción y el uso de sondas derivadas de genes de plantas para su uso en el mapeo genético se describen en Bematzky y Tanksley (*Plant Mol. Biol. Reporter* 4, 37-41, 1986). Numerosas publicaciones describen mapeo genético de clones de ADNc específicos usando la metodología expuesta anteriormente o variaciones de la misma. Por ejemplo, pueden usarse para el mapeo poblaciones de intercruzamiento F2, poblaciones de retrocruzamiento, poblaciones apareadas aleatoriamente, líneas casi isogénicas y otros conjuntos de individuos. Tales metodologías las conocen bien los expertos en la técnica.

También pueden usarse sondas de ácido nucleico para mapeo físico (es decir, colocación de secuencias en mapas

físicos; véanse Hoheisel *et al.* En: Nonmammalian Genomic Analysis: A Practical Guide, Academic Press 1996, págs. 319-346, y la bibliografía citada en el mismo).

5 En otra realización, pueden usarse sondas de ácido nucleico en mapeo mediante hibridación *in situ* con fluorescencia directa (FISH) (Trask (1991) Trends Genet 7, 149-154). Aunque los métodos actuales de mapeo mediante FISH favorecen el uso de clones grandes (de varios a varios cientos de kb; véase Laan *et al.* (1995) Genome Res. 5, 13-20), las mejoras en la sensibilidad pueden permitir la realización de mapeo mediante FISH usando sondas más cortas.

10 Puede llevarse a cabo una variedad de métodos basados en la amplificación de ácidos nucleicos de mapeo genético y físico usando los ácidos nucleicos. Los ejemplos incluyen amplificación específica de alelo (Kazazian (1989) J. Lab. Clin. Med. 11, 95-96), polimorfismo de fragmentos amplificados por PCR (CAPS; Sheffield *et al.* (1993) Genomics 16, 325-332), ligamiento específico de alelo (Landegren *et al.* (1988) Science 241, 1077-1080), reacciones de extensión de nucleótidos (Sokolov (1990) Ácido nucleico Res. 18, 3671), mapeo de híbridos por radiación (Walter *et al.* (1997) Nat. Genet 7, 22-28) y mapeo HAPPY (Dear y Cook (1989) Nucleic Acid Res. 17, 6795-6807). Para estos métodos, se usa la secuencia de un ácido nucleico para diseñar y producir pares de cebadores para su uso en la reacción de amplificación o en reacciones de extensión con cebadores. El diseño de tales cebadores lo conocen bien los expertos en la técnica. En métodos que emplean mapeo genético basado en PCR, puede ser necesario identificar diferencias en la secuencia de ADN entre los progenitores del cruzamiento de mapeo en la región correspondiente a la presente secuencia de ácido nucleico. Esto, sin embargo, generalmente no es necesario para métodos de mapeo.

20 De esta manera, pueden realizarse la generación, la identificación y/o el aislamiento de plantas modificadas con la actividad de tipo NAP1 alterada que presentan características de crecimiento mejoradas.

25 Los ácidos nucleicos de tipo NAP1 o variantes de los mismos o polipéptidos de tipo NAP1 u homólogos de los mismos también pueden encontrar uso como reguladores del crecimiento. Puesto que estas moléculas han mostrado que son útiles en la mejora de las características de crecimiento de plantas, también serán útiles como reguladores del crecimiento, tales como herbicidas o estimuladores del crecimiento. La presente invención proporciona, por tanto, una composición que comprende un ácido nucleico de tipo NAP1 o variante del mismo o un polipéptido de tipo NAP1 u homólogo del mismo, junto con un portador, diluyente o excipiente adecuado, para su uso como regulador del crecimiento, preferiblemente como promotor del crecimiento, más preferiblemente para aumentar el rendimiento.

30 Los métodos según la presente invención dan como resultado plantas que tienen un rendimiento de semillas mejorados tal como se describió anteriormente en el presente documento. Estas características de crecimiento ventajosas también pueden combinarse con otros rasgos ventajosos desde el punto de vista económico, tales como rasgos de potenciación del rendimiento adicionales, tolerancia a diversos estrés, rasgos que modifican diversas características arquitectónicas y/o características bioquímicas y/o fisiológicas. Por consiguiente, los métodos de la presente invención también pueden usarse en los denominados procedimientos de "apilamiento de genes".

Descripción de las figuras

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

40 Figura 1. Árbol filogenético que representa las relaciones entre proteínas NAP y SET de levadura, hombre y plantas. Se estableció el árbol mediante el programa AlignX de VNTI serie 5.5 (Informax, <http://www.informaxinc.com/>). La matriz usada para generar la alineación múltiple es Blosum62 y los parámetros de alineación usados fueron: penalización por apertura de hueco, 10; penalización por extensión de hueco, 0,5; intervalo de penalización por separación de hueco, 8; % de identidad para el retardo de alineación, 40. Se construye el árbol filogenético usando el método del vecino más cercano de Saitou y Nei. Los números de registro GenBank y MIPS (para *Arabidopsis thaliana*) de las secuencias usadas en la alineación se indican en el árbol. At: *Arabidopsis thaliana*, Gm: *Glycine max*, Nt: *Nicotiana tabacum* (secuencias derivadas del documento WO 03/085115), Os: *Oryza sativa*, Ps: *Pisum sativum*, Zm: *Zea mays*, Hs: *Homo sapiens*, Sc: *Saccharomyces cerevisiae*.

50 Figura 2. Presentación esquemática del clon de entrada p68, que contiene CDS0406 dentro de los sitios AttL1 y AttL2 para la clonación Gateway® en la estructura principal de pDONR201. CDS0406 es el código interno para la secuencia codificante de la proteína de tipo NAP1 de *Arabidopsis*. Este vector contiene también un casete de resistencia a kanamicina bacteriano y un origen de replicación bacteriano.

Figura 3. Vector binario para la expresión en *Oryza sativa* del gen de tipo NAP1 de *Arabidopsis* (CDS0406) bajo el control del promotor GOS2 (PRO0129). Este vector contiene un ADN-T 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)). Desde el borde izquierdo al borde derecho, este ADN-T contiene: un casete para la selección de antibióticos de plantas

transformadas; un casete de marcador seleccionable para el examen visual de plantas transformadas y el casete de terminación doble PRO0129 - CDS0406 -zeína y rbcS-deltaGA para la expresión del gen de tipo NAP1 de *Arabidopsis*. 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.

5 Figura 4. La proteína de tipo NAP1 de *Medicago* se expresó en *E. coli* y se purificó a partir del extracto celular en bruto mediante cromatografía de afinidad a través de la cola de 6XHIS. Se visualiza la elución de la proteína de 34 kD a diferentes concentraciones de imidazol a partir de la resina de níquel-agarosa mediante inmunotransferencia de tipo Western usando anticuerpo anti-6xHIS (Sigma, St Louis, EE.UU.).

10 Figura 5. Alineación de la proteína de tipo NAP1 de *Arabidopsis thaliana* (de tipo AtNAP1) y la proteína de tipo NAP1 de *Medicago sativa* (de tipo MsNAP1) usando el algoritmo de Needleman y Wunsch. Se fijó la penalización por apertura de hueco en 10, la penalización por extensión de hueco era de 0,5. Con estos ajustes, la identidad de secuencia era del 71,2% y la homología de secuencia era del 84,5%.

15 Figura 6. La proteína de tipo NAP1 tiene una localización nuclear en plantas. A) La proteína de tipo NAP1 de *Medicago* ha mostrado que se localiza en el núcleo de células de alfalfa cultivadas mediante inmunofluorescencia indirecta usando un anticuerpo producido frente a la proteína purificada (fotografía a la izquierda del panel A). Para confirmar la localización nuclear, se tiñeron los núcleos en paralelo con el colorante fluorescente DAPI, (fotografía a la derecha del panel A). En la inserción, la flecha apunta a una célula en metafase. Una fluorescencia débil indica baja abundancia de la proteína de tipo NAP1 alrededor de los cromosomas en las células en metafase sin un compartimento nuclear. B) La proteína de tipo NAP1 de *Arabidopsis* expresada de manera transitoria, fusionada a GFP, se localiza en el núcleo de células de *Arabidopsis* tras una captación mediada por PEG del constructo génico en protoplastos.

20

Figura 7. La proteína de tipo NAP1 de *Medicago* purificada inhibe *in vitro* la actividad de desfosforilación de la fosfohistona H2B de PP2A (purificada a partir de músculo esquelético de conejo), pero no tiene influencia sobre la desfosforilación de la glucógeno fosforilasa por la misma enzima.

25 Figura 8. Ejemplos de secuencias útiles en la realización de los métodos según la presente invención.

Ejemplos

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

30 A menos que se establezca de otro modo, se realizan técnicas de ADN recombinante según protocolos convencionales descritos en (Sambrook (2001) Molecular Cloning: a laboratory manual, 3ª edición Cold Spring Harbor Laboratory Press, CSH, Nueva York) o en los volúmenes 1 y 2 de Ausubel *et al.* (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 Labfase (1993) de R.D.D. Croy, publicado por BIOS Scientific Publications Ltd (R.U.) y Blackwell Scientific Publications (R.U.).

35 Ejemplo 1: Clonación de genes

Se amplificó la proteína de tipo NAP1 de *Arabidopsis* (referencia interna CDS0406) por PCR usando como molde una biblioteca de ADNc de plántulas de *Arabidopsis thaliana* (Invitrogen, Paisley, R.U.). Tras transcripción inversa del ARN extraído de plántulas, se clonaron los ADNc en pCMV Sport 6.0. El tamaño de inserto promedio del banco era de 1,5 kb, y el número original de clones era de $1,59 \times 10^7$ ufc. Se determinó que el título original era de $9,6 \times 10^5$ ufc/ml, tras la primera amplificación de 6×10^{11} ufc/ml. Tras la extracción de plásmidos, se usaron 200 ng de molde en una mezcla para PCR de 50 μ l. Se usaron los cebadores prm1505 (SEQ ID NO: 4) y prm1506 (SEQ ID NO: 5), que incluyen los sitios AttB para la recombinación Gateway, para la amplificación por PCR. Se realizó la PCR usando Hifi Taq ADN polimerasa en condiciones convencionales. Se amplificó un fragmento de PCR de 771 pb y se purificó usando también métodos convencionales. Entonces se realizó la primera etapa del procedimiento Gateway, la reacción de BP, durante la cual el fragmento de PCR se recombina *in vivo* con el plásmido pDONR201 para producir, según la terminología de Gateway, un "clon de entrada", p68 (figura 2). Se adquirió el plásmido pDONR201 de Invitrogen, como parte de la tecnología Gateway®.

40

45

Ejemplo 2: Construcción de vectores

Posteriormente se usó el clon de entrada p68 en una reacción de LR con p0640, un vector de destino usado para la transformación en *Oryza sativa*. Este vector contiene como elementos funcionales dentro de los bordes del ADN-T: un marcador seleccionable de plantas; un casete de expresión de marcador visual; y un casete Gateway destinado para LR recombinación *in vivo* con la secuencia de interés ya clonada en el clon de entrada. Un promotor GOS2

50

para expresión constitutiva (PRO0129) está ubicado en el sentido de 5' de este casete Gateway. Tras la etapa de recombinación mediante LR, el vector de expresión resultante p73 (figura 3) puede transformarse en *Agrobacterium* cepa LBA4404 y posteriormente en plantas de *Oryza sativa*.

Ejemplo 3: Transformación de arroz

5 Se descascarillaron semillas secas maduras de *Oryza sativa* cultivar japónica, 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 germinaron las semillas estériles en un medio que contenía 2,4-D (medio de inducción de callo). Tras una incubación durante 4 semanas en la oscuridad, se cortaron los callos derivados del escutelo, embriogénicos y se propagaron en el mismo medio. Dos semanas después se multiplicaron o propagaron mediante subcultivo en el mismo medio durante otras dos semanas. 3 días antes del cocultivo, se subcultivaron trozos de callo embriogénico en medio nuevo para reforzar la actividad de división celular. Se usó *Agrobacterium* cepa LBA4404 que alberga el vector binario p73 para el cocultivo. Se cultivó la cepa de *Agrobacterium* durante 3 días a 28°C en medio AB con los antibióticos apropiados. Entonces se recogieron las bacterias y se suspendieron en medio de cocultivo líquido a una DO₆₀₀ de aproximadamente 1. Se transfirió la suspensión a una placa Petri y se sumergieron los callos en la suspensión durante 15 minutos. A continuación, se secaron los tejidos callosos sobre un papel de filtro, se transfirieron a medio de cocultivo solidificado y se incubaron durante 3 días en la oscuridad a 25°C. Después de eso, se hizo crecer el callo cocultivado 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 de crecimiento rápido. Tras la transferencia de este material a un medio de regeneración e incubación con luz, se liberó el potencial embriogénico y se desarrollaron brotes en las siguientes cuatro a cinco semanas. Se cortaron los brotes del callo y se incubaron durante de 2 a 3 semanas en un medio que contenía auxina desde el que se transfirieron al suelo. Se hicieron crecer los brotes endurecidos con alta humedad y pocos días en un invernadero. Finalmente, se cosecharon las semillas de tres a cinco meses tras el trasplante. El método produjo transformantes de un solo locus a una tasa de más del 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 transformantes: mediciones de crecimiento vegetativo

Se generaron aproximadamente de 15 a 20 transformantes T0 independientes. Se transfirieron los transformantes primarios desde cámaras de cultivo tisular a un invernadero para el crecimiento y el cosechado de las semillas T1. Se conservaron cinco acontecimientos a partir de los cuales la progenie T1 se segregó 3:1 para determinar la presencia/ausencia del transgén. Para cada uno de estos acontecimientos, se seleccionaron 10 plántulas T1 que contenían el transgén (hetero y homocigotas), y 10 plántulas T1 que carecían del transgén (nulicigotas), mediante examen de marcadores visuales. Se transfirieron las plantas T1 seleccionadas a un invernadero. Cada planta recibió una etiqueta de código de barras única para relacionar inequívocamente los datos de fenotipado con la planta correspondiente. Se hicieron crecer las plantas T1 seleccionadas en el suelo en macetas de 10 cm de diámetro con los siguientes ajustes ambientales: fotoperiodo = 11,5 h, intensidad de luz diurna = 30.000 lux o más, temperatura durante el día = 28°C o mayor, temperatura durante la noche = 22°C, humedad relativa = 60-70%. Se hicieron crecer plantas transgénicas y las nulicigotas correspondientes unas al lado de otras en posiciones aleatorias. Desde la etapa de siembra hasta la etapa de madurez, se hicieron pasar las plantas varias veces a través de un recinto de obtención de imágenes digitales. En cada punto de tiempo, se tomaron imágenes digitales (2048x1536 píxeles, 16 millones de colores) de cada planta desde al menos 6 ángulos diferentes. Se cosecharon las panículas primarias maduras, se embolsaron, se etiquetaron con código de barras y se secaron entonces durante tres días en el horno a 37°C. Se trillaron entonces las panículas y se recogieron todas las semillas. Se separaron las cáscaras llenas de las vacías usando un dispositivo de soplado de aire. Tras la separación, se contaron entonces ambos lotes de semillas usando una máquina contadora disponible comercialmente. Se desecharon las cáscaras vacías. Se pesaron las cáscaras llenas en una balanza analítica y se midió el área de la sección transversal de las semillas usando obtención de imágenes digitales. Este procedimiento dio como resultado el conjunto de parámetros relacionados con las semillas descrito anteriormente.

Estos parámetros se derivaron de forma automatizada a partir de las imágenes digitales usando software de análisis de imágenes y se analizaron estadísticamente. Se usó ANOVA (análisis de la varianza) de dos factores corregidos para el diseño desequilibrado como modelo estadístico para la evaluación global de las características fenotípicas de las plantas. Se llevó a cabo una prueba F con todos los parámetros medidos de todas las plantas de todos los acontecimientos transformados con ese gen. Se llevó a cabo la prueba F para comprobar un efecto del gen sobre todos los acontecimientos de transformación y para verificar un 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 está provocando el efecto. El umbral para la significación para un efecto génico global se fija en el nivel de probabilidad del 5% para la prueba F.

Para comprobar un efecto de los genes dentro de un acontecimiento, es decir, para un efecto específico de línea, se

realizó una prueba de la t dentro de cada acontecimiento usando conjuntos de datos de las plantas transgénicas y las plantas nulas correspondientes. “Plantas nulas” o “segregantes nulos” o “nulicigotas” son las plantas tratadas de la misma manera que la planta transgénica, pero de las que se ha segregado el transgén. Las plantas nulas también pueden describirse como las plantas transformadas negativas homocigotas. El umbral para la significación para la prueba de la t se fija en el nivel de probabilidad del 10%. Los resultados para algunos acontecimientos pueden estar por encima o por debajo de este umbral. Esto se basa en la hipótesis de que un gen podría tener sólo un efecto en determinadas posiciones en el genoma, y que la aparición de este efecto dependiente de la posición no es poco común. Esta clase de efecto génico también se denomina en el presente documento a “un efecto de línea del gen”. Se obtiene este valor de p comparando el valor de t de la distribución de t o alternativamente, comparando el valor de F para la distribución de F. El valor de p proporciona entonces la probabilidad de que la hipótesis nula (es decir, que no existen ningún efecto del transgén) sea correcta.

Se midieron el crecimiento vegetativo y el rendimiento de semillas según los métodos descritos anteriormente. Los inventores hallaron sorprendentemente que el peso total de semillas, el número de semillas llenas y el índice de cosecha aumentaron en las plantas de arroz transformadas con el gen de tipo NAP1 en comparación con las plantas control sin el gen de tipo NAP1.

Se confirmaron los datos obtenidos en el primer experimento en un segundo experimento con plantas T2. Se seleccionaron cuatro líneas que tenían el patrón de expresión correcto para su análisis adicional. Se examinaron lotes de semillas de las plantas positivas (tanto hetero como homocigotas) en T1, monitorizando la expresión de marcadores. Para cada acontecimiento seleccionado, se conservaron entonces los lotes de semillas heterocigotas para la evaluación de T2. Dentro de cada lote de semillas, se hizo crecer un número igual de plantas positivas y negativas en el invernadero para su evaluación.

Se evaluó un número total de 160 plantas de tipo NAP1 transformadas, en la generación T2, es decir 40 plantas por acontecimiento de las que 20 eran positivas para el transgén, y 20 negativas.

Ejemplo 5: Evaluación de transformantes: medición de parámetros relacionados con las semillas

Tras el análisis de las semillas tal como se describió anteriormente, los inventores hallaron que las plantas transformadas con el constructo génico de tipo NAP1 tenían un mayor número de semillas llenas, un mayor peso total de semillas y un aumento del índice de cosecha en comparación con plantas que carecían del transgén de tipo NAP1. Los resultados positivos obtenidos para plantas en la generación T1 se obtuvieron de nuevo en la generación T2. No sólo las líneas transgénicas individuales puntuaron significativamente mejor que las líneas de control nulicigotas correspondientes, sino que también hubo un efecto global positivo significativo cuando se evaluaron todas las plantas de todos los acontecimientos T2 sometidos a prueba, lo que indica con fuerza un efecto génico global. Se facilita en la tabla 3 una visión general de los datos.

Tabla 3

parámetro	Plantas T1		Plantas T2		Análisis combinado
	% de aumento	valor de p	% de aumento	valor de p	valor de p
Peso total de semillas	24	0,0076	28	0,0013	0,0006
Número de semillas llenas	17	0,0352	28	0,0013	0,0003
Índice de cosecha	15	0,0085	17	0,0025	0,0003
El % de aumento presenta el aumento promedio para todos los acontecimientos sometidos a prueba. El valor de p representa el valor de p derivado de la prueba F.					

Número de semillas Llenas

Se determinó el número de semillas llenas mediante recuento del número de cáscaras llenas que quedaron tras la etapa de separación. 4 de las 5 líneas sometidas a prueba mostraron un aumento en el número de semillas llenas, que ascendía al 37%. Hubo un aumento global del 17% en el número de semillas llenas producidas por las plantas transgénicas con relación los segregantes nulos correspondientes, siendo el aumento estadísticamente significativo (valor de p de 0,0352). En la generación T2, hubo un aumento para todas las líneas sometidas a prueba, que osciló

entre el 14 y el 46%. El aumento medio para las líneas T2 fue del 28%, siendo también este aumento medio estadísticamente significativo (valor de p de 0,0013). El análisis combinado de los datos de T1 y T2 también confirmó que el efecto génico global era altamente significativo (valor de p de 0,0003).

Rendimiento total de semillas

5 Se midió el rendimiento total de semillas (peso total de semillas) por planta pesando todas las cáscaras llenas cosechadas de una planta. Todas las líneas T1 transgénicas mostraron un aumento en el peso total de semillas, que varió entre el 8 y el 43%. En promedio, el aumento en el rendimiento de semillas fue del 24% y este efecto global por la presencia del transgén sobre el rendimiento de semillas fue significativo, tal como se evidencia mediante un valor de p para la prueba F de 0,0076. Estos resultados también se observaron en la generación T2. Las 4 líneas sometidas a prueba tuvieron un aumento de rendimiento de entre el 14 y el 48% con un promedio del 28%. Este aumento medio fue estadísticamente significativo (valor de p de 0,013) y también el análisis combinado de las plantas T1 y T2 mostró que hubo un efecto génico global (valor de p de 0,0006).

Índice de cosecha

15 El índice de cosecha en la presente invención se define como la razón entre el rendimiento total de semillas y el área sobre el terreno (mm^2), multiplicado por un factor de 10^6 . 4 de las 5 líneas sometidas a prueba mostraron un aumento del índice de cosecha, que osciló entre el 9 y el 48%. Hubo un efecto génico global significativo (un efecto asociado con la presencia del transgén) sobre el índice de cosecha (un aumento global del 15%), con un valor de p estadísticamente significativo para la prueba F de 0,085. Se obtuvieron resultados similares para las plantas T2. El índice de cosecha para las líneas individuales aumentó entre el 12 y el 34% con una media significativa del 17% (valor de p de 0,0025). En este caso también, el análisis combinado de los datos de T1 y T2 mostraron un efecto génico global (valor de p 0,003).

25 Los expertos en la técnica saben que la expresión de transgenes en plantas, y así también el efecto fenotípico debido a la expresión de tal transgén, puede diferir entre diferentes líneas transgénicas obtenidas independientemente y la progenie de las mismas. Los transgenes presentes en diferentes plantas transgénicas obtenidas independientemente difieren unos de otros en el locus de inserción cromosómica así como en el número de copias del transgén insertadas en ese locus y la configuración de esas copias del transgén en ese locus. Diferencias en los niveles de expresión pueden atribuirse a la influencia desde el contexto cromosómico del transgén (el denominado efecto de posición) o de mecanismos de silenciamiento desencadenados por determinadas configuraciones de transgén (por ejemplo, las inserciones en tándem orientadas hacia el interior de transgenes son propicias para el silenciamiento a nivel transcripcional o postranscripcional). A pesar de estas posibles causas de variación, los datos muestran que las plantas transgénicas que expresan el gen de tipo NAP1 de manera sistemática proporcionan un mayor número de semillas llenas, un mayor peso total de semillas, así como un aumento del índice de cosecha, cada uno con relación a plantas no transgénicas correspondientes. Los aumentos observados fueron significativos tanto en la generación T1 como en la T2, lo que es un argumento sólido para un efecto génico global tal como se evidencia mediante los valores de p del análisis combinado.

Ejemplo 6: Caracterización de una proteína de tipo NAP1 de *Medicago sativa*:

Materiales y métodos

Aislamiento del clon de ADNc de longitud completa del supuesto inhibidor de PP2A de alfalfa

40 Se ha usado un fragmento de ADNc aislado que codifica para una parte de una supuesta proteína de tipo NAP1 de alfalfa (*Medicago sativa*) para aislar el clon de longitud completa de una biblioteca de ADNc de fago λ -ZAP de nódulos de las raíces de alfalfa (Savoure *et al.* Plant Mol. Biol. 27, 1059-1070; 1995) usando procedimientos de examen convencionales tal como se describen por el fabricante (Stratagene). Se examinaron 400.000 placas de lisis, se conservaron 20 clones, de los que 18 eran positivos en el segundo examen de hibridación. Se seleccionaron 8 de estos clones para trabajo adicional y se convirtieron en fagémidos a partir de fagos individuales. Se secuenciaron cuatro clones y dos de ellos demostraron ser los clones de ADNc de longitud completa de la supuesta proteína de tipo NAP1. Se usó uno de los clones (Ms10.1) para trabajo adicional (SEQ ID NO: 10, que codifica para la proteína de SEQ ID NO: 11).

Producción y purificación de la proteína de tipo NAP1 de *Medicago*

50 Se insertó la secuencia de ADNc que codifica para la proteína de tipo NAP1 de *Medicago sativa* en el sitio *NcoI/XhoI* del vector pENTRY4 GATEWAY[®] (Invitrogen) y posteriormente se introdujo en el vector de expresión bacteriano pDEST17. El vector pDEST17 permitió la expresión de la proteína de tipo NAP1 en células de BL21 *E. coli* como una proteína con cola de 6xHIS. Se purificó la proteína de tipo NAP1 de 34 kDa mediante cromatografía de afinidad usando una resina de níquel-agarosa (Sigma) (figura 4).

Mediciones de la actividad fosfatasa

5 Se sometió a prueba *in vitro* la posible actividad de inhibición de fosfatasa de la proteína de tipo NAP1 de *Medicago sativa* sobre subunidades catalíticas de proteína fosfatasa 2A (PP2A) purificadas a partir de músculo esquelético de conejo usando glucógeno fosforilasa marcada con el isótopo ³²P y proteínas histonas H2A como sustratos según Ulloa *et al.* (1993).

Localización intracelular de las proteínas de tipo MsNAP1 y AtNAP1

10 Se produjeron anticuerpos policlonales anti-proteína de tipo MsNAP1 en conejos frente a la proteína con cola de 6xHIS purificada usando un protocolo de inmunización convencional. Se aislaron los protoplastos de células de alfalfa (*Medicago sativa*) cultivadas en suspensión y fijadas mediante formaldehído al 6%. Entonces se unieron las células a portaobjetos de vidrio recubiertos con poli-L-lisina y se expusieron al antisuero anti-proteína de tipo MsNAP1 (diluido 200x con PBS), se lavaron y se expusieron a anticuerpo secundario de cabra anti-conejo conjugado con FITC (SIGMA, dilución 100x). Se tiñeron los núcleos con DAPI (0,02 mg/ml) en paralelo y se fotografiaron con un microscopio de fluorescencia TE300 de Nikon y una cámara CCD a color SPOT II. Se insertó en el marco la región codificante del ortólogo de *Arabidopsis thaliana* de la proteína de tipo NAP1 de *Medicago sativa* (SEQ ID NO: 1), con la proteína fluorescente verde (GFP) en el vector de expresión de plantas compatible con GATEWAY[®] (pK7WGF2). Se aislaron los protoplastos y se transfectaron con el ADN de plásmido purificado usando procedimientos convencionales. Se registró la expresión transitoria uno o dos días tras la transfección mediante microscopía de fluorescencia.

Resultados

20 Las proteínas de tipo NAP1 de *Arabidopsis* y *Medicago* se localizan en el núcleo.

25 La proteína de tipo NAP1 de *Medicago sativa* era altamente homóloga a la proteína de tipo NAP1 de *Arabidopsis*: la identidad de secuencia era del 71,2% y la similitud de secuencia era del 84,5% (figura 5). Usando los anticuerpos anti-proteína de tipo MsNAP1, la inmunofluorescencia indirecta reveló que los anticuerpos reconocen una proteína que se localizaba en los núcleos de células de alfalfa cultivadas en suspensión. Se verificó esta localización mediante la tinción nuclear, DAPI. Se asoció la fluorescencia débil con los cromosomas en células en metafase (figura 6.A, inserción). La proteína de tipo NAP1 de *Arabidopsis* marcada con etiqueta de GFP también se localizaba exclusivamente en los núcleos de células de *Arabidopsis* cultivadas en suspensión (figura 6.B).

2) La proteína de tipo NAP1 de alfalfa inhibe *in vitro* la actividad fosfatasa PP2A sobre un sustrato de fosfo-histona

30 Se añadió la proteína de tipo NAP1 de alfalfa purificada a diversas concentraciones a mezclas de reacción que contenían las subunidades catalíticas de PP2A de músculo esquelético de conejo e histona H2A fosforilada, o glucógeno fosforilasa como sustrato. Se observó que la proteína de tipo NAP1 no tenía influencia sobre la desfosforilación de la glucógeno fosforilasa ni siquiera a una concentración de 500 mM, pero ya una concentración de 2,5 mM de la proteína de tipo NAP1 inhibía eficazmente la actividad PP2A sobre el sustrato de fosfo-histona H2A (disminución del 50% en la actividad) (figura 7).

35 Conclusión

Las proteínas de tipo NAP1 de *Medicago sativa* y *Arabidopsis thaliana* muestran semejanza tanto estructural como funcionalmente. Las proteínas de tipo NAP1 de plantas inhiben *in vitro* la actividad fosfatasa (PP2A) sobre sustratos de histona, indicando un posible papel *in vivo* sobre la organización de la cromatina y la transcripción génica.

LISTA DE SECUENCIAS

40 <110> CropDesign N.V.

<120> Plantas que tienen características de crecimiento mejoradas y método para producir las mismas

<130> CD-110-PCT

<150> Documento EP 04101388.9

<151> 02-04-2004

45 <150> Documento US 60/563.847

ES 2 403 414 T3

<151> 20-04-2004

<160> 33

<170> PatentIn versión 3.3

<210> 1

5 <211> 1075

<212> ADN

<213> *Arabidopsis thaliana*

<400> 1

```

gtgacggaac cacaagaaga gaagagacca aaaggaggag ccaaaatcct ctctttcttt      60
ggaattaggg tttcctcaaa ggaagtgaac tgaaaatggt cgcggaacaag agcaagaagt      120
cgaaaattga agagaaaggc gaagaagaaa acttgagca aatcgacgca gagcttgttc      180
tctcaattga gaagcttcag gagattcaag acgacctcga gaagattaac gaaaaggcca      240
gtgacgaggt cttggaagta gaggcagaaat ataacgtgat acggaaacct gtctatgaca      300
agcgcaatga agttatccaa tcgattcctg gcttttgat gactgctttt ttgagtcac      360
ctgccttagg cgacctcttg actgaagaag accaaaagat ttttaagtac ttgaactctc      420
tggaagtgga ggatgcaaaa gatgtgaaat ctggatactc tataactttt cacttcactt      480
caaaccggtt ctttgaggat gccaaagctta ccaagacatt tactttcctt gaagaaggaa      540
caacaaaaat cactgcaact cctatcaaat ggaaggaggc caagggcttg ccaaatggag      600
tgaaccatga tgataagaaa ggaataaac gtgcattgcc agaggagagt ttctttactt      660
ggtttactga tgctcaacat aaggaagatg ctggggatga gattcatgat gaggttgctg      720
atattatcaa ggaagatctc tggccaacc ctctcaccta cttcaacaat gatgctgatg      780
aagaggattt tgatggagat gatgatggtg acgaagaggg agaagaagac gatgacgatg      840
aagaggagga agatggtgag gaatgatggg agcccaaaga taacacattg ctggcttgct      900
tctataacag atgtgtaaag tttgtgttat gaggttctca attttagcaa tgatgagact      960
aagctttctc ttttgaata tttagtttat ttactatcaa tagctacatt ctgtttgtac     1020
gaaccatgtc atcatccatg tctaaatct tgccataact acatctgttt ttcgc         1075

```

10 <210> 2

<211> 256

<212> PRT

<213> *Arabidopsis thaliana*

<400> 2

```

Met Val Ala Asp Lys Ser Lys Lys Ser Lys Ile Glu Glu Lys Gly Glu
1                               10                    15
Glu Glu Asn Leu Glu Gln Ile Asp Ala Glu Leu Val Leu Ser Ile Glu
20                          25                    30
Lys Leu Gln Glu Ile Gln Asp Asp Leu Glu Lys Ile Asn Glu Lys Ala
35                          40                    45

```

15

ES 2 403 414 T3

Ser Asp Glu Val Leu Glu Val Glu Gln Lys Tyr Asn Val Ile Arg Lys
 50 . 55 60

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

Trp Met Thr Ala Phe Leu Ser His Pro Ala Leu Gly Asp Leu Leu Thr
 85 90 95

Glu Glu Asp Gln Lys Ile Phe Lys Tyr Leu Asn Ser Leu Glu Val Glu
 100 105 110

Asp Ala Lys Asp Val Lys Ser Gly Tyr Ser Ile Thr Phe His Phe Thr
 115 120 125

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

Leu Glu Glu Gly Thr Thr Lys Ile Thr Ala Thr Pro Ile Lys Trp Lys
 145 150 155 160

Glu Gly Lys Gly Leu Pro Asn Gly Val Asn His Asp Asp Lys Lys Gly
 165 . 170 175

Asn Lys Arg Ala Leu Pro Glu Glu Ser Phe Phe Thr Trp Phe Thr Asp
 180 185 190

Ala Gln His Lys Glu Asp Ala Gly Asp Glu Ile His Asp Glu Val Ala
 195 200 205

Asp Ile Ile Lys Glu Asp Leu Trp Ser Asn Pro Leu Thr Tyr Phe Asn
 210 215 220

Asn Asp Ala Asp Glu Glu Asp Phe Asp Gly Asp Asp Asp Gly Asp Glu
 225 230 235 240

Glu Gly Glu Glu Asp Asp Asp Asp Glu Glu Glu Glu Asp Gly Glu Glu
 245 250 255

<210> 3

<211> 3560

<212> ADN

5 <213> Secuencia artificial

<220>

<223> casete de expresión

<400> 3

ES 2 403 414 T3

aatccgaaaa	gtttctgcac	cgttttcacc	ccctaactaa	caatataggg	aacgtgtgct	60
aaatataaaa	tgagacctta	tatatgtagc	gctgataact	agaactatgc	aagaaaaact	120
catccaccta	ctttagtggc	aatcgggcta	aataaaaaag	agtcgctaca	ctagtttcgt	180
tttccttagt	aattaagtgg	gaaaatgaaa	tcattattgc	ttagaatata	cgttcacatc	240
tctgtcatga	agttaaatta	ttcgaggtag	ccataattgt	catcaaactc	ttcttgaata	300
aaaaaatctt	tctagctgaa	ctcaatgggt	aaagagagag	atTTTTTTta	aaaaaataga	360
atgaagatat	tctgaacgta	ttggcaaaga	tttaaacata	taattatata	atTTTTatagT	420
ttgtgcattc	gtcatatcgc	acatcattaa	ggacatgtct	tactccatcc	caatTTTTat	480
ttagtaatta	aagacaattg	acttattttt	attatttatc	ttttttcgat	tagatgcaag	540
gtacttacgc	acacactttg	tgctcatgtg	catgtgtgag	tgcacctcct	caatacacgt	600
tcaactagca	acacatctct	aatatcactc	gcctatttaa	tacatttagg	tagcaatatc	660

ES 2 403 414 T3

tgaattcaag	cactccacca	tcaccagacc	acttttaata	atatctaaaa	tacaaaaaat	720
aattttacag	aatagcatga	aaagtatgaa	acgaactatt	taggtttttc	acatacaaaa	780
aaaaaaagaa	ttttgctcgt	gcgcgagcgc	caatctccca	tattgggcac	acaggcaaca	840
acagagtggc	tgcccacaga	acaaccaca	aaaacgatg	atctaacgga	ggacagcaag	900
tccgcaacaa	ccttttaaca	gcaggctttg	cggccaggag	agaggaggag	aggcaaaaga	960
aaccaagcat	cctcctcctc	ccatctataa	attcctcccc	ccttttcccc	tctctatata	1020
ggaggcatcc	aagccaagaa	gagggagagc	accaaggaca	cgcgactagc	agaagccgag	1080
cgaccgcctt	cttcgatcca	tatcttcctg	tcgagttctt	ggtcgatctc	ttccctcctc	1140
cacctcctcc	tcacagggta	tgtgcccttc	ggttgttctt	ggatttattg	ttctaggttg	1200
tgtagtacgg	gcgttgatgt	taggaaaggg	gatctgtatc	tgtgatgatt	cctgttcttg	1260
gatttgggat	agagggggtt	ttgatgttgc	atgttatcgg	ttcggtttga	ttagtagtat	1320
ggttttcaat	cgtctggaga	gctctatgga	aatgaaatgg	tttagggtag	ggaatcttgc	1380
gattttgtga	gtaccttttg	tttgaggtaa	aatcagagca	ccggtgattt	tgcttgggtg	1440
aataaaagta	cggttgtttg	gtcctcgatt	ctggtagtga	tgcttctcga	tttgacgaag	1500
ctatcctttg	tttattccct	attgaacaaa	aataatccaa	ccttgaagac	ggtcccgttg	1560
atgagattga	atgattgatt	cttaagcctg	tccaaaatct	cgagctggc	ttgtttagat	1620
acagtagtcc	ccatcacgaa	attcatggaa	acagttataa	tcctcaggaa	caggggattc	1680
cctgttcttc	cgatttgctt	tagtcccaga	attttttttc	ccaaatatct	taaaaagtca	1740
ctttctgggt	cagttcaatg	aattgattgc	tacaaataat	gcttttatag	cgttatccta	1800
gctgtagtcc	agttaatagg	taataccctt	atagtttagt	caggagaaga	acttatccga	1860
tttctgatct	ccatttttaa	ttatatgaaa	tgaactgtag	cataagcagt	attcatttgg	1920
attatttttt	ttattagctc	tcacccttcc	attattctga	gctgaaagtc	tggcatgaac	1980
tgtcctcaat	tttgttttca	aattcacatc	gattatctat	gcattatcct	cttgtatcta	2040
cctgtagaag	tttctttttg	gttattcctt	gactgcttga	ttacagaaag	aaatttatga	2100
agctgtaatc	gggatagtta	tactgcttgt	tcttatgatt	catttccttt	gtgcagttct	2160
tgggtgtagct	tgccactttc	accagcaaa	ttcatttaaa	tcaactaggg	atatcacaag	2220
tttgtaaaaa	aaagcagget	tcacaatggt	cgcggacaag	agcaagaagt	cgaaaatgta	2280
agagaaaggc	gaagaagaaa	acttggagca	aatcgacgca	gagcttggtc	tctcaatgta	2340
gaagcttcag	gagattcaag	acgacctcga	gaagattaac	gaaaaggcca	gtgacgaggt	2400
cttggaagta	gagcagaaat	ataacgtgat	acggaaacct	gtctatgaca	agcgcaatga	2460
agttatccaa	tcgattcctg	gcttttggat	gactgctttt	ttgagtcatc	ctgccttagg	2520
cgacctcttg	actgaagaag	accaaaaagat	ttttaagtac	ttgaactctc	tggagtgga	2580
ggatgccaaa	gatgtgaaat	ctggatactc	tataactttt	cacttcactt	caaaccctgt	2640
ctttgaggat	gccaaactta	ccaagacatt	tactttcctt	gaagaaggaa	caacaaaaat	2700
cactgcaact	cctatcaaat	ggaaggaggg	caaggccttg	ccaaatggag	tgaacctatga	2760
tgataagaaa	ggaaataaac	gtgcattgcc	agaggagagt	ttctttactt	ggtttactga	2820
tgctcaacat	aaggaagatg	ctggggatga	gattcatgat	gaggttgctg	atattatcaa	2880
ggaagatctc	tggccaacc	ctctcaccta	cttcaacaat	gatgctgatg	aagaggattt	2940
tgatggagat	gatgatggtg	acgaagaggg	agaagaagac	gatgacgatg	aagaggagga	3000
agatggtgag	gaatgatggg	accagcttt	cttgtaaaaa	gtgggtgat	cacaagcccg	3060
ggcggctctc	tagggataac	agggtaat	tatccctcta	gatcacaagc	ccgggcggctc	3120
ttctacgatg	attgagtaat	aatgtgtcac	gcataccat	gggtggcagt	gtcagtgtga	3180
gcaatgacct	gaatgaacaa	ttgaaatgaa	aagaaaaaaa	gtactccatc	tgttccaaat	3240
taaaattcat	tttaaccttt	taataggttt	atacaataat	tgatatatgt	tttctgtata	3300
tgtctaattt	gttatcatcc	gggcggctct	ctagggataa	cagggttaatt	atatccctct	3360
agacaacaca	caacaaataa	gagaaaaaac	aaataatatt	aatttgagaa	tgaacaaaag	3420
gaccatatca	ttcattaact	cttctccatc	catttccatt	tcacagttcg	atagcgaaaa	3480
ccgaataaaa	aacacagtaa	attacaagca	caacaaatgg	tacaagaaa	acagttttcc	3540
caatgccata	atactcgaac					3560

<210> 4

<211> 51

<212> ADN

5 <213> Secuencia artificial

<220>

ES 2 403 414 T3

<223> cebador: prm1505

<400> 4

ggggacaagt ttgtacaaaa aagcaggctt cacaatggtc gcggaacaaga g 51

<210> 5

5 <211> 48

<212> ADN

<213> Secuencia artificial

<220>

<223> cebador: prm1506

10 <400> 5

ggggaccact ttgtacaaga aagctgggtc ccatcattcc tcaccatc 48

<210> 6

<211> 1101

<212> ADN

15 <213> *Nicotiana tabacum*

<400> 6

agcggctggt	accggtccgg	aattcccggg	ataticgtcga	cccacgcgtc	cgaaagaaga	60
gaaaaagatg	ggtgctgaca	aaggaagaa	gcaaaaagtg	gaggaagaga	acaacacccat	120
tgatggtgag	ctcgtttttt	ccattgaaaa	attgcaagaa	atacaagacg	agctcgagaa	180
gatcaatgag	gaagcaagtg	ataaagtatt	ggaagtggaa	cagaagtaca	atgagatccg	240
caagcctgtc	tatgacaaac	gaaacgacat	cattaaagct	atcccggact	tctggttgac	300
tgcttttttg	agtcatcctg	tcctaggtga	acttctaact	gaagaagacc	aaaagatcct	360
caagtttcta	agttctattg	aagttgaaga	ctctaaagat	gtgaagtcgg	gctactcgat	420
aacctttaac	ttcaatgcga	atccttattt	tgaaaataca	aagctcacia	agacctatac	480
cttccttgaa	gatggacca	caaagatttc	tgctacaaca	ataaaatgga	aagaaggcat	540
gggcattcct	aatggatttg	cacatgagaa	gaaaggaaac	aagcgatctc	atgctgagga	600
aagcttcttc	acatggttca	gtgaagtcaa	tcaaaaagat	gaggatgagg	atgaggccct	660
agagattcag	gatgaggtcg	ctgacataat	taaggatgac	ttgtggccga	accctctcac	720
ctattttaac	aacgagcctg	atgaagaaga	ttttgatggt	gacgagggaa	aggacagtga	780
aggctctgaa	gacgaagagg	aagaagaaga	ggaggatgag	gatggtgatg	aagaatgaag	840
gcagtaaact	gttcaagacc	cctatthttg	gatctcgtct	tcagcggttt	taatcatcag	900
ggtttaatgt	ctgtaaagag	gctttgaatg	ttgccaaga	acagaataac	tgtggtgact	960
ataccttttc	ttctcttgta	tggttataac	ttataagcaa	aatatctaat	tccggaggtt	1020
ccaaaatggt	ttcattaggc	tagttcgatt	aatgaagtgt	ttgtctggca	aaaactgata	1080
atgtaggtt	attgagttat	g				1101

<210> 7

<211> 256

20 <212> PRT

<213> *Nicotiana tabacum*

<400> 7

ES 2 403 414 T3

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

<210> 8

<211> 1061

5 <212> ADN

<213> *Nicotiana tabacum*

<400> 8

ES 2 403 414 T3

tttgtacaaa	aaagcaggct	ggtaccggtc	cggaattccc	gggatatcgt	cgacccacgc	60
gtccgagaaa	ttagcagtta	gagacactga	gaagcagcag	ctctcttcct	cagctgctgt	120
gtgcttaggc	aaagaataaa	atggggggcag	acaaggggaa	gaagcagaaa	gtggatgagg	180
aaaacaacaa	tgttattgat	gaaaagctca	ttttttccat	tgaaaaattg	caagagatac	240
aagacgagct	cgagaagatc	aatgaaaaag	caagcgacga	agtgttgga	gtagaacaga	300
agtacaacga	gatccgcaag	cctgtctacg	ataagcgaaa	tgatgtcatt	agctctattt	360
ctgacttctg	gttgactgct	tttttgagtc	atcctgttct	tggtaacctt	ctcactgaag	420
aggaccaaaa	gattttcaaa	tttghtaagt	ctattgaagt	ggaagactca	aaggatgtga	480
aatcgggtca	ttcaatcacg	tttaacttta	agcccaatcc	ttattttgaa	aattcaaagc	540
tctcaaagac	gtataccttc	cttgaagatg	gacctacaaa	aattacagct	acaacaataa	600
aatggaaaga	aggcatgggc	attcctaatt	gagttgctga	caagaagaaa	ggaaacaagc	660
ggtcccacgc	tgaagaaagt	ttctttacat	ggttcagtga	agtcaatcaa	aaaggtgatg	720
tggatgatga	cgaaaatgag	attctggaca	ttcaggatga	tgaggttgct	gaaataatca	780
aggatgactt	gtggcctaac	cctctcaatt	atthtgacca	tgagcctgat	gaagaagata	840
ttgagggcga	tgagggaaag	gacagcggag	gctctgaaga	ggaagaagaa	gaggaagatg	900
atgaagatga	agaagacgaa	tgaactgttg	gtagaccttg	tgthtgattt	gagttctcat	960
cagtgtttca	atcatcagag	ttgggtctctg	taaagaggtt	tcggatattg	cagaaaaatt	1020
gaatgacata	tagtggtgac	tctaattttt	agthtcagtg	a		1061

<210> 9

<211> 260

<212> PRT

5 <213> *Nicotiana tabacum*

<400> 9

ES 2 403 414 T3

Met Gly Ala Asp Lys Gly Lys Lys Gln Lys Val Asp Glu Glu Asn Asn
 1 5 10 15
 Asn Val Ile Asp Glu Lys Leu Ile Phe Ser Ile Glu Lys Leu Gln Glu
 20 25 30
 Ile Gln Asp Glu Leu Glu Lys Ile Asn Glu Lys Ala Ser Asp Glu Val
 35 40 45
 Leu Glu Val Glu Gln Lys Tyr Asn Glu Ile Arg Lys Pro Val Tyr Asp
 50 55 60
 Lys Arg Asn Asp Val Ile Ser Ser Ile Ser Asp Phe Trp Leu Thr Ala
 65 70 75 80
 Phe Leu Ser His Pro Val Leu Gly Asn Leu Leu Thr Glu Glu Asp Gln
 85 90 95
 Lys Ile Phe Lys Phe Val Ser Ser Ile Glu Val Glu Asp Ser Lys Asp
 100 105 110
 Val Lys Ser Gly His Ser Ile Thr Phe Asn Phe Lys Pro Asn Pro Tyr
 115 120 125
 Phe Glu Asn Ser Lys Leu Ser Lys Thr Tyr Thr Phe Leu Glu Asp Gly
 130 135 140
 Pro Thr Lys Ile Thr Ala Thr Thr Ile Lys Trp Lys Glu Gly Met Gly
 145 150 155 160
 Ile Pro Asn Gly Val Ala Asp Lys Lys Lys Gly Asn Lys Arg Ser His
 165 170 175
 Ala Glu Glu Ser Phe Phe Thr Trp Phe Ser Glu Val Asn Gln Lys Gly
 180 185 190
 Asp Val Asp Asp Asp Glu Asn Glu Ile Leu Asp Ile Gln Asp Asp Glu
 195 200 205
 Val Ala Glu Ile Ile Lys Asp Asp Leu Trp Pro Asn Pro Leu Asn Tyr
 210 215 220
 Phe Asp His Glu Pro Asp Glu Glu Asp Ile Glu Gly Asp Glu Gly Lys
 225 230 235 240
 Asp Ser Gly Gly Ser Glu Glu Glu Glu Glu Glu Glu Asp Asp Glu Asp
 245 250 255
 Glu Glu Asp Glu
 260

<210> 10

<211> 1210

<212> ADN

5 <213> *Medicago sativa*

<220>

ES 2 403 414 T3

<221> misc_feature

<222> (1014)..(1014)

<223> n e s a, c, g, o t

<400> 10

ggcacgagca	aaaccctaac	acttccctca	ttcacgctcg	aagaaaagaa	cacaaatctc	60
tccactgcg	tagggtttga	aacccaacac	cttctgtttc	ttcaaccatg	gtggccgaca	120
agtctaagaa	gttgaaagtt	tctgaaaagg	gtgaaaacgc	tgaagagatc	gacggagaac	180
ttgttctctc	cattgaaaag	ttgcaggaaa	ttcaagatga	gattgaaaag	atcaatgagg	240
aggctagcga	taaagttctc	gaaatagagc	agaagtacaa	tgaagtaagg	aaaccggtgt	300
atgacaagcg	caatgatgtg	atcaagtcca	ttcccgattt	ctggctaact	gcgtttttga	360
gccatcctgt	tcttggtgat	ctcttgaatg	aagaagatca	gaagatattt	aagcatttaa	420
tctctcttga	ggtggaagat	cataaagatg	ttaaatacagg	ctattcaatc	acatttaact	480
tcgactccaa	tccctttttt	gaggattcaa	aacttgttaa	gacttttacc	ttccttgaag	540
aaggaaccac	aaagcttacc	gccacaccca	taaaatggaa	agagggcaag	ggcattccca	600
atggagttat	tcatgagaag	aaagggaaaca	agcgagctgc	ttctgatatc	agtttcttta	660
cctggttttg	tgacactgag	cagaaagatg	aaatgggtga	cattcatgat	gagattgctg	720
aaatgatcaa	ggatgattta	tggccgaatc	cactcaatta	tttcaacagt	gaggaccctg	780
atgaagcaga	ggaggaggat	gatgaagctg	gtgatgcggg	aaaggatgat	gatgactctg	840
aagatgatga	tgatcaagag	gatgacgacg	atgacgagga	agaagaatag	tgtaaaatgc	900
tttaaaatag	taatacttgg	ttttaattta	tttattttaa	ggttattata	ggagtatctt	960
agtggtcttt	aggggatgat	gaaagaccaa	ggttggctat	tggttttccc	cctntgggcg	1020
taaaccttat	ttattgtgct	ttgaagggtga	tttctggttt	tatctttgtg	cgcttctttc	1080
aagataccaa	tgatacatcg	gattttatct	tagtcctata	ttgaaacat	atagtagtta	1140
aaatgtagta	tattcagtg	atagctgcgt	aatcagtatc	atthttattgc	tatcacaact	1200
5 ttacagtacc						1210

<210> 11

<211> 260

<212> PRT

<213> *Medicago sativa*

10 <400> 11

ES 2 403 414 T3

Met Val Ala Asp Lys Ser Lys Lys Leu Lys Val Ser Glu Lys Gly Glu
 1 5 10 15
 Asn Ala Glu Glu Ile Asp Gly Glu Leu Val Leu Ser Ile Glu Lys Leu
 20 25 30
 Gln Glu Ile Gln Asp Glu Ile Glu Lys Ile Asn Glu Glu Ala Ser Asp
 35 40 45
 Lys Val Leu Glu Ile Glu Gln Lys Tyr Asn Glu Val Arg Lys Pro Val
 50 55 60
 Tyr Asp Lys Arg Asn Asp Val Ile Lys Ser Ile Pro Asp Phe Trp Leu
 65 70 75 80
 Thr Ala Phe Leu Ser His Pro Val Leu Gly Asp Leu Leu Asn Glu Glu
 85 90 95
 Asp Gln Lys Ile Phe Lys His Leu Ile Ser Leu Glu Val Glu Asp His
 100 105 110
 Lys Asp Val Lys Ser Gly Tyr Ser Ile Thr Phe Asn Phe Asp Ser Asn
 115 120 125
 Pro Phe Phe Glu Asp Ser Lys Leu Val Lys Thr Phe Thr Phe Leu Glu
 130 135 140
 Glu Gly Thr Thr Lys Leu Thr Ala Thr Pro Ile Lys Trp Lys Glu Gly
 145 150 155 160
 Lys Gly Ile Pro Asn Gly Val Ile His Glu Lys Lys Gly Asn Lys Arg
 165 170 175
 Ala Ala Ser Asp Ile Ser Phe Phe Thr Trp Phe Cys Asp Thr Glu Gln
 180 185 190
 Lys Asp Glu Met Gly Asp Ile His Asp Glu Ile Ala Glu Met Ile Lys
 195 200 205
 Asp Asp Leu Trp Pro Asn Pro Leu Asn Tyr Phe Asn Ser Glu Asp Pro
 210 215 220
 Asp Glu Ala Glu Glu Glu Asp Asp Glu Ala Gly Asp Ala Gly Lys Asp
 225 230 235 240
 Asp Asp Asp Ser Glu Asp Asp Asp Asp Gln Glu Asp Asp Asp Asp
 245 250 255
 Glu Glu Glu Glu
 260

<210> 12

<211> 907

5 <212> ADN

<213> *Zea mays*

<400> 12

ES 2 403 414 T3

gttcctacct	tcttcctcc	gtctcccagc	tcgcgaggc	aggcgacaca	gcgacgctaa	60
aaaccctaga	gcgaggaggc	gtgccaggcc	agcggtttgc	gatgacggca	ccggcggaca	120
aggggaagaa	ggccaagacc	gatgcccagc	gcggcgagga	gaacgagcaa	atcgacggcg	180
cccttgctt	ctccatcgag	aagctccagg	agattcagga	cgagctcgag	aaggttaatg	240
aggaagcaag	tgacaagggt	atggagggtg	agcagaaata	cagtgagatt	cgcagacctg	300
tctatctcaa	gaggggtgac	attatcaaga	ccatcccgga	cttttggtc	acagcgtttt	360
tgagccatcc	tctactaagt	gagcttctga	ctgaagagga	tcagaagata	ttcaagtact	420
tggactccat	tgatgtcgat	gattctgatg	ttaaggcagg	atattccatt	taccttaact	480
tctctgagaa	cccgtacttt	gaagacacaa	agcttacaaa	gacctattcc	tttgttgatg	540
atggaacaac	cacaataaaa	gcttctcaaa	ttaagtggaa	ggatggaatg	ggacctgcaa	600
atggaaatgg	tattaacaag	aagggaaaca	agcggccatt	agtagtggaa	agttttttct	660
cctggtttag	tgatacagag	ctcaagagtc	ttgctgatgg	tgtgcaagat	gaggtggcgg	720
agatcatcaa	ggaagacttg	tggcctaatc	ctttgaagta	cttcaacaat	gaggttgaag	780
atgaatttga	aggagatgaa	gaagatgatg	acgacgacga	cgacgatgat	aatttggatg	840
gtgatgacaa	tgacgatgat	ggggaccagg	agaactgagc	cctcgcgttt	aggcggggaa	900
ttatgtg						907

<210> 13

<211> 258

5 <212> PRT

<213> *Zea mays*

<400> 13

ES 2 403 414 T3

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

Glu Asn

<210> 14

<211> 780

5 <212> ADN

<213> *Oryza sativa*

<400> 14

ES 2 403 414 T3

atggcggcgg	cggagcagaa	ggggaagaag	ccgaggaccg	acggcgcgga	ggccgagccc	60
gtcgcacgcc	ccctgctgca	gtccatcgag	aagctccagg	agatccagga	cgagatcgag	120
aaggttaatg	aggaagcatg	tgataaagtt	ctggagttgg	aacagaaata	caacgagggtt	180
cgcagaccag	tttatgttcg	acggaataaa	attatcaagc	aaattcctga	cttctggctg	240
acagcgtttc	ttagccatcc	tatgcttggg	gaactattaa	ctgaagatga	tcaaaagatt	300
ttcaaact	tggagtctat	cgacgtggat	gactcagaag	atatcaaatac	aggctactcc	360
attactctca	cattctcccc	caatccatat	tttgaagata	caaagcttac	aaaaacatat	420
tcctttagt	acgatgaagc	agtcaaagta	aaggctacct	ccatcagggtg	gaagaaagga	480
atggatattg	ccaatgatcg	tgcgtacacg	aagaaagggg	acaagcgaat	cttaattgat	540
gaaagtttct	ttacttgggt	caatagtga	aagaacagaa	gttttgctca	tggagctatg	600
gatgaggtgg	cagatgtcat	caaggaagat	ctgtggccta	atcctttgaa	gtacttcaac	660
aatgaatttg	aagaagaatt	agagctactg	gatgacgatg	acgaggtatc	tgatgatgac	720
gatgaggagg	aggatgatga	agaccaaggt	gaaggagagg	aggatggaga	ggagaactga	780

<210> 15

<211> 259

5 <212> PRT

<213> *Oryza sativa*

<400> 15

ES 2 403 414 T3

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

Glu Glu Asn

<210> 16

<211> 1315

5 <212> ADN

<213> *Oryza sativa*

<400> 16

ES 2 403 414 T3

```

ctccgctctc ctccagctcc gcctccgacg cgcgcacgcc tctccctccc ctcctcctcc      60
gcctcgcttc gcagtgtgga agaaaggaag gaaggctaaa accctagcga gcgcgcgagc      120
gagcgagggc tctctgcttc cttgcgatga cggcgccggc ggacaagggg aagaaggcca      180
agaccgacgc cgacggcggc gccgcccagg agaacgagca gatcgacggc gccctcgtcc      240
tctccatcga gaagctccag gagatccagg acgagctcga gaaggtcaat gaggaagcta      300
gtgacaaggt tttggaggtc gagcagaaat acagtgagat tcgcagacct gtctatctcc      360
gaaggagtga cgttatccaa acaatccccg acttctggct gacagcgttt ctgagtcatt      420
ctctacttag tgagcttttg accgaagagg atcaaaagat gttcaagtac ctggagtctg      480
tcgacgtgga tgattctaaa gatgtcaagt caggctactc cataactctt accttctccg      540
agaacccgta ctttgaagac aaagagctca cgaagacata tgccttcgct gatgacggaa      600
caaccacaat aaatgctact agcattaagt ggaaagaagg aatggaaatt gcaaatggga      660
atgccaagaa gaaagggagc aagcgaccat tggttgagga aagtttcttc acctggttta      720
ctgatacaga gcacaagagt cttgctgatg gtgtgcaaga tgaggtggct gagatcatca      780
aggaagacct gtggcccaat ccattgaagt atttcaataa tgaggctgaa gagttaggag      840
aggatgacga cgaagagggg tctgatgctg atgaggggtg agaggatgag gaggaggaga      900
actgagtcta ggatgtcaga ttgcatggtt gccgatcgtc tgcattttgt ggatgctgtc      960
actctgaagg gcgaagtgtc gtgaccctcg gttgcttctt tcttttttct ttttgatgac     1020
ttagctggaa cccttaggaa ctgtttaatg ccttatggag tccgtcgtat tttcgactca     1080
aaggagacac ctctatatca taatctgcgt ataaccatgg aagacatttt aacctgctga     1140
tgtgtggttc attgcgctgc ctctggtgct gtaggggtgt cgttcctttg tgctctctgt     1200
cttttttttt tttttttgtg tgtgtgtggt cgcgctggca ttgttgccag tctgatgggc     1260
tgttatttct ccccctagaa agagtgaaaa acctggcttg tgatcattgt ttacg           1315

```

<210> 17

<211> 252

<212> PRT

5 <213> *Oryza sativa*

<400> 17

```

Met Thr Ala Pro Ala Asp Lys Gly Lys Lys Ala Lys Thr Asp Ala Asp
1           5           10           15

Gly Gly Ala Ala Glu Glu Asn Glu Gln Ile Asp Gly Ala Leu Val Leu
          20           25           30

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

Glu Glu Ala Ser Asp Lys Val Leu Glu Val Glu Gln Lys Tyr Ser Glu
          50           55           60

Ile Arg Arg Pro Val Tyr Leu Arg Arg Ser Asp Val Ile Gln Thr Ile
65           70           75           80

```

ES 2 403 414 T3

Pro Asp Phe Trp Leu Thr Ala Phe Leu Ser His Pro Leu Leu Ser Glu
85 90 95

Leu Leu Thr Glu Glu Asp Gln Lys Met Phe Lys Tyr Leu Glu Ser Val
100 105 110

Asp Val Asp Asp Ser Lys Asp Val Lys Ser Gly Tyr Ser Ile Thr Leu
115 120 125

Thr Phe Ser Glu Asn Pro Tyr Phe Glu Asp Lys Glu Leu Thr Lys Thr
130 135 140

Tyr Ala Phe Ala Asp Asp Gly Thr Thr Thr Ile Asn Ala Thr Ser Ile
145 150 155 160

Lys Trp Lys Glu Gly Met Glu Ile Ala Asn Gly Asn Ala Lys Lys Lys
165 170 175

Gly Ser Lys Arg Pro Leu Val Glu Glu Ser Phe Phe Thr Trp Phe Thr
180 185 190

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

Glu Ile Ile Lys Glu Asp Leu Trp Pro Asn Pro Leu Lys Tyr Phe Asn
210 215 220

Asn Glu Ala Glu Glu Leu Gly Glu Asp Asp Asp Glu Glu Gly Ser Asp
225 230 235 240

Ala Asp Glu Gly Glu Glu Asp Glu Glu Glu Glu Asn
245 250

<210> 18

<211> 958

<212> ADN

5 <213> *Zea mays*

<400> 18

ccaaaagggc cacagttccg cctccttttc ctgccttctt cctcactagt cgctcccccg 60
cggctcgcgc aggcgggcca cacaacgagg ctaaaccctt atcgcgagga ggcgtgtgag 120
gccagcggct ttgcgatgac agcaccagcg gacaagggga agaaggccaa gactgatgcc 180
gacggcggcg aggagaacga acagatcgac ggcgtcctcg tcctctccat cgagaagctc 240
caggagatac aggacgagct cgagaaggta aatgaggaag caagtgacaa ggttatggag 300
gtggagcaga aatacagtga gatccgcaga cctgtctatc tcaagagggg tgacattatc 360
aagaccatcc cggacttttg gctcacagcg tttatgagcc atcctctatt aagtgagctt 420
ctgactgaag aggaccagaa gatattcaag tacttagact ccattgatgt ggatgattct 480
gatgtaagc caggatactc cattcatctt aacttctctg agaaccgta ctttgaggac 540
acaaagcttg caaagaccta tatctttgct gatgatggaa caaccacaat aaaagcttcc 600
gaaattaagt ggaaggaagg aatgggacct gcaaattgaa atggtattaa caagaagggg 660
agtaagcggc cattagtaga ggaaagtttt tttagctggt ttggtgatac agagctcaag 720
agtcttgctg atggtgtgca agatgaggtg gcgagatca taaaggaaga tttgtggcct 780
aatcctttga agtacttcaa caatgaggtt gacgatgaat ttgaaggaga tgaagatgat 840
gatgatttgg atggtgatga tgacgatgaa ggcgatgatt tggagaactg agcccttgcg 900
cttggttcag aatggtgtcc gtggatgatg tggctgggcg gaactgtgac ctttttgg 958

ES 2 403 414 T3

<210> 19

<211> 251

<212> PRT

<213> *Zea mays*

5 <400> 19

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

ES 2 403 414 T3

<210> 20

<211> 1257

<212> ADN

<213> *Arabidopsis thaliana*

5 <400> 20

```

agattcacgc atcacacaat cgagttttta gggtttttagc ggttgctctc tcggaagcca      60
gagagaagag ggaagaggaa gtctaattcc tctgcgtttt ttgcaattag ggttttctca      120
attggaatcg aaaatgggta cagacaagag caagaaggcg aaaaccgaag aagaaaacgt      180
cgagcaaatc gatgcagagc ttgtcctctc aatcgaaaag cttcaagaga tccaagacga      240
cctcgagaag ataaatgaaa aggctagtga tgaagtgttg gaagtggagc agaaatataa      300
tgtgataagg aaacctgttt atgacaagcg taacgagatc atcaaaacca tccctgattt      360
ctggttaact gctttcttga gtcaccctgc tttaggtgaa cttttgactg aagaagacca      420
aaagattttc aaatatctta gctctcttga tgttgaggat gccaaagatg tgaaatctgg      480
atactctatc actttttcct tcaatcccaa tccatttttt gaagatggaa aactgacaaa      540
gacttttacc tttctcgaag aagggacaac caaaatcaca gccacgccta tcaaatggaa      600
agagggcaaa ggcctggcga atggagtgaa tcatgagaag aatggaaaca aacgtgcact      660
acctgaagag agcttcttta cctgggttag tgatgctcaa cacaaggagg atgttgagga      720
tgagatgcaa gacgagcagg ttgcagatat catcaaggaa gatttgtggc ccaaccctct      780
cacctacttc aacaatgacg ctgatgaaga ggactttgat ggagacgatg atggagatga      840
agaggagaaa gaaggtgact ctgatgaaga tgatgacgaa gaagacgaag ttggtgagga      900
atgatggcag ggataccag aaaccacatt tgcttacatg tcttctctat aacagagtgt      960
gtaaagtttt gtgtgttgaa aggtttttta ttttaagcaa aagtggatta tgacgacaac     1020
agacaagctt ttaattttat tttaccgtaa tagttatac ttggtgtaag aaaccatttt     1080
cagccttttg ttgaaaaatc ctgcttaaat ggtttttgag tcttacataa tagcttcttc     1140
atcttttgtc ttcttaaaga gaattatatt tgtaatttca tgtctgttgt gtttctttga     1200
ctttactgaa tagagaatth gtgtgtttat ggtgaaaata tagccgatct gcttgac      1257

```

<210> 21

<211> 256

10 <212> PRT

<213> *Arabidopsis thaliana*

<400> 21

ES 2 403 414 T3

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

<210> 22

<211> 1003

5 <212> ADN

<213> *Lycopersicon esculentum*

<400> 22

ES 2 403 414 T3

```

gcgaaatcaa gaaaatcagt taagcagctc tgtaactcag gtgggaaaaa ggcaaaaaat    60
aatggtggtt gacaaagggg agaagcagaa agtgggaagag gaaagctaca ttgatgaaaa    120
gctcattttt tccattgaaa aattgcaaga aatacaagac gaccttgaca agatcaatga    180
gaaagcaagt gaggaagtgt tggaaataga acagaagtac aacaagatcc gcaagcctgt    240
ttatgataag cggaatgata tcattaactc tatttctgac ttctggttga ctgctttttt    300
gagtcacccg gttcttggtg accttctaac tgaagaggac caaaagattt tcaaattcct    360
aagttctatt gaagtgggag actcgaaaga tgtgaaatth ggttactcaa tcacgtttta    420
ctttaagccc aatcctttct ttgaaaattc aaagctctca aagacctata ccttccttga    480
agatggacct acaaaaaatca cagctacacc aataaaatgg aaagaaggca aaggcattcc    540
taatggcggt gctcaggaga agaaaggaaa caagcgatcc catgctggaag agagcttctt    600
cacctgggtc agtgaagtca ataaaaaaga tgatagcgat gatgatgaaa atgaggttct    660
ggagattcag gatgagggtg ctgaaataat caaggatgac ttgtggccaa accctttaac    720
ttattttacc aatgaacctg atgaagaaga ttttgagggt gatgaagggt gtgatgaggg    780
ggaggactct gaagatgaag gtgatgagga ggaagaggaa gacgacgaag atgaagatga    840
caaatgaact gttaatggac ctcatatttg atttgatttc tcttcttcaa tgtttcaatt    900
atcatagttg gtatctgtaa agaagcttaa tattgcagat aaaatcgaat tatatatagt    960
ggtgactgct ttttttctaa aaaaaaaaaa aaaaaaaaaa aaa                                1003

```

<210> 23

<211> 261

<212> PRT

5 <213> *Lycopersicon esculentum*

<400> 23

```

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

```

ES 2 403 414 T3

Ser His Pro Val Leu Gly Asp Leu Leu Thr Glu Glu Asp Gln Lys Ile
85 90 95

Phe Lys Phe Leu Ser Ser Ile Glu Val Glu Asp Ser Lys Asp Val Lys
100 105 110

Phe Gly Tyr Ser Ile Thr Phe Asn Phe Lys Pro Asn Pro Phe Phe Glu
115 120 125

Asn Ser Lys Leu Ser Lys Thr Tyr Thr Phe Leu Glu Asp Gly Pro Thr
130 135 140

Lys Ile Thr Ala Thr Pro Ile Lys Trp Lys Glu Gly Lys Gly Ile Pro
145 150 155 160

Asn Gly Val Ala Gln Glu Lys Lys Gly Asn Lys Arg Ser His Ala Glu
165 170 175

Glu Ser Phe Phe Thr Trp Phe Ser Glu Val Asn Lys Lys Asp Asp Ser
180 185 190

Asp Asp Asp Glu Asn Glu Val Leu Glu Ile Gln Asp Glu Val Ala Glu
195 200 205

Ile Ile Lys Asp Asp Leu Trp Pro Asn Pro Leu Thr Tyr Phe Thr Asn
210 215 220

Glu Pro Asp Glu Glu Asp Phe Glu Gly Asp Glu Gly Gly Asp Glu Gly
225 230 235 240

Glu Asp Ser Glu Asp Glu Gly Asp Glu Glu Glu Glu Asp Asp Glu
245 250 255

Asp Glu Asp Asp Lys
260

<210> 24

<211> 1557

<212> ADN

5 <213> *Arabidopsis thaliana*

<400> 24

ES 2 403 414 T3

```

gtccctagtc ttttctgctt cttcttcttc aaaatctctc ttttcaccaa atcctcagaa      60
gatgagcaac gacaaggata gtttcaacgt ctccgatctt actgctgctc ttaaggacga      120
ggatcgagct ggccttgtca atgctctaaa gaacaagctg cagaatctgg ctggtcagcg      180
ttctgatgtg ctcgagaatc tgactcccaa tgtgagaaag cgcgttgatg ccttgagggg      240
tatacagagc caacatgatg aactagagggc aaaattccgt gaggagagag ctattcttga      300
agccaagtat caaacgctgt atcagccttt gtatgtcaag cgttatgaga ttgtgaatgg      360
cactactgaa gttgaactgg ctccagagga tgataccaag gtggaccaag gagaggagaa      420
aactgcagaa gagaaaggag ttccaagttt ctggctgaca gctctgaaaa ataacgatgt      480
tatttccgag gaggtcacag agcgtgatga aggggctctc aaatatctta aagatattaa      540
gtggtgcaag attgaagagc ctaaaggatt caaacttgag tttttctttg acacgaatcc      600
gtattttaag aacactgtct tgacaaagtc ttatcatatg attgatgaag atgagccact      660
gcttgagaag gctatgggga cagaaattga ttggtatcct ggaaagtgtc taactcagaa      720
gattctcaag aagaagccta agaaaggttc aaagaatact aaaccaatca ccaaactcga      780
agattgtgaa agcttcttca acttcttttag tcctccagaa gttccggatg aagatgaaga      840
tatcgacgag gaaagagctg aggatcttca aaacctgatg gaacaagatt atgacatcgg      900
atctactatt cgggaaaaga ttattcctcg tgctgtctca tggtttactg gtgaggctat      960
ggaagcagag gattttgaaa tagatgacga tgaggaagat gacattgatg aggatgaaga     1020
tgaggaagac gaagaggatg aggaggacga tgatgatgag gatgaagaag aaagcaagac     1080
caaaaagaag ccatcaatcg gcaacaagaa gggagggaga tctcagatag ttggtgaagg     1140
taaacaagat gagaggccac ccgaatgcaa gcaacagtaa tcttttacta cgctctacca     1200
gacataaaaag gattgctgta aaatataatt caggtcattc tctgttcacg aagaatgagg     1260
attgagaaaa ggttttgagg tttttaaaag tgaaattcat cttgtaggag tttcgttcgt     1320
ttttctattg gtgtgtttat tttctctaaa gcactttaat aatatacctt ggtatttaat     1380
ttatgaatca agcatcatca tccctagtct ctgcattcac tacttcatcc cctacctaaa     1440
ctttgtcgac gaaagagatt ttaataacca tttagatagt aatgggtagt gggaatgatc     1500
attattcttt tgttcacctg cttttgattt tcaatggtaa ccattttgtt gtgtaag      1557

```

<210> 25

<211> 372

5 <212> PRT

<213> *Arabidopsis thaliana*

<400> 25

ES 2 403 414 T3

Met Ser Asn Asp Lys Asp Ser Phe Asn Val Ser Asp Leu Thr Ala Ala
1 5 10 15

Leu Lys Asp Glu Asp Arg Ala Gly Leu Val Asn Ala Leu Lys Asn Lys
20 25 30

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

Pro Asn Val Arg Lys Arg Val Asp Ala Leu Arg Asp Ile Gln Ser Gln
50 55 60

His Asp Glu Leu Glu Ala Lys Phe Arg Glu Glu Arg Ala Ile Leu Glu
65 70 75 80

Ala Lys Tyr Gln Thr Leu Tyr Gln Pro Leu Tyr Val Lys Arg Tyr Glu
85 90 95

Ile Val Asn Gly Thr Thr Glu Val Glu Leu Ala Pro Glu Asp Asp Thr
100 105 110

Lys Val Asp Gln Gly Glu Glu Lys Thr Ala Glu Glu Lys Gly Val Pro
115 120 125

Ser Phe Trp Leu Thr Ala Leu Lys Asn Asn Asp Val Ile Ser Glu Glu
130 135 140

Val Thr Glu Arg Asp Glu Gly Ala Leu Lys Tyr Leu Lys Asp Ile Lys
145 150 155 160

Trp Cys Lys Ile Glu Glu Pro Lys Gly Phe Lys Leu Glu Phe Phe Phe
165 170 175

Asp Thr Asn Pro Tyr Phe Lys Asn Thr Val Leu Thr Lys Ser Tyr His
180 185 190

Met Ile Asp Glu Asp Glu Pro Leu Leu Glu Lys Ala Met Gly Thr Glu
195 200 205

Ile Asp Trp Tyr Pro Gly Lys Cys Leu Thr Gln Lys Ile Leu Lys Lys
210 215 220

ES 2 403 414 T3

Lys Pro Lys Lys Gly Ser Lys Asn Thr Lys Pro Ile Thr Lys Leu Glu
 225 230 235 240
 Asp Cys Glu Ser Phe Phe Asn Phe Phe Ser Pro Pro Glu Val Pro Asp
 245 250 255
 Glu Asp Glu Asp Ile Asp Glu Glu Arg Ala Glu Asp Leu Gln Asn Leu
 260 265 270
 Met Glu Gln Asp Tyr Asp Ile Gly Ser Thr Ile Arg Glu Lys Ile Ile
 275 280 285
 Pro Arg Ala Val Ser Trp Phe Thr Gly Glu Ala Met Glu Ala Glu Asp
 290 295 300
 Phe Glu Ile Asp Asp Asp Glu Glu Asp Asp Ile Asp Glu Asp Glu Asp
 305 310 315 320
 Glu Glu Asp Glu Glu Asp Glu Glu Asp Asp Asp Asp Glu Asp Glu Glu
 325 330 335
 Glu Ser Lys Thr Lys Lys Lys Pro Ser Ile Gly Asn Lys Lys Gly Gly
 340 345 350
 Arg Ser Gln Ile Val Gly Glu Gly Lys Gln Asp Glu Arg Pro Pro Glu
 355 360 365
 Cys Lys Gln Gln
 370

<210> 26

<211> 1575

<212> ADN

5 <213> *Arabidopsis thaliana*

<400> 26

ES 2 403 414 T3

gtgtcttatt	tcggtctggt	cattttctca	aagccctttt	agttatztat	atatatattc	60
tctgtctcgt	atttgtcccc	aaaaatctag	ggttttaagg	tttcttatcc	ttcctcttcc	120
tccgccagat	tcttttcttg	cgaagatgag	caacgacaag	gacagcatga	acatgtccga	180
tctctccacc	gctcttaacg	aggaggatcg	tgccgggctt	gttaatgctc	ttaagaacaa	240
gttgcagaat	ttggctggac	aacactctga	tgtccttgaa	aacttgactc	caccagtcag	300
gaagcgtgtc	gagtttctaa	gagagattca	gaaccaatat	gatgagatgg	aagcaaaatt	360
ctttgaggag	agagcagctc	ttgaagctaa	gtatcaaaaag	ttatatcagc	ctttatatac	420
caagcgatat	gagattgtga	atgggtgtgg	cgaagttgaa	ggtgcagctg	aagaagtaaa	480
atccgaacaa	ggagaagata	aatcagctga	agagaaagga	gtaccagatt	tctggcttat	540
tgcattgaag	aacaatgaaa	ttactgcgga	agagataact	gagcgagatg	aaggggctct	600
caagtatctc	aaagatatca	agtggagtag	ggttgaagaa	ccaaaagggg	tcaagcttga	660
gtttttcttt	gatcagaatc	cttacttcaa	gaacactgct	ttgaccaaga	catatcacat	720
gattgatgaa	gatgagccta	tccttgagaa	ggccctcggg	acggagattg	agtggtatcc	780
tgaaaagtgt	ttgacacaga	agattctaaa	aaagaagcca	aagaaaggat	ccaaaaacac	840
aaagccgatc	actaagactg	aggactgtga	gagtttcttc	aactttttca	gtccacctca	900
agttcctgac	gatgatgagg	atcttgatga	tgacatggct	gatgaactcc	aaggacaaat	960
ggagcatgat	tatgatatcg	gttcaacaat	caaagagaaa	atcatctcgc	atgctgtgtc	1020
atggttcact	ggtgaagctg	ttgaggcaga	tgaccttgat	attgaggacg	acgatgatga	1080
gattgatgaa	gatgatgatg	aagaggacga	ggaagatgat	gaggatgacg	aggaggagga	1140
tgatgaggat	gatgacgagg	aggaagaagc	agatcaagga	aagaagagca	aaaagaagtc	1200
atcagctggg	cacaagaagg	ctggaagaag	tcaacttgcg	gaaggtcaag	caggtgagag	1260
gccaccgga	tgtaagcagc	agtgaagaag	tgaagaatct	tggcttagtt	atgatgaaga	1320
agaagagtga	agagtgtctt	tgagccgagg	ttgtgtttct	ttaatttgca	gagtcatggt	1380
ccggtttatt	atatatcagt	tttgggtgat	tggtttgcta	tttaaaaaaa	aaaaatgggt	1440
tctttggttt	ggtttgtgtc	tcttgatfff	tccttttgta	atgatcttat	gaatttgttt	1500
cgagttaatg	tcgttctctg	gtcagatfff	gaattcaatt	ctatttatcc	tcctctgtta	1560
atgagagaat	ttgtg					1575

<210> 27

<211> 379

5 <212> PRT

<213> *Arabidopsis thaliana*

<400> 27

ES 2 403 414 T3

Met Ser Asn Asp Lys Asp Ser Met Asn Met Ser Asp Leu Ser Thr Ala
1 5 10 15

Leu Asn Glu Glu Asp Arg Ala Gly Leu Val Asn Ala Leu Lys Asn Lys
20 25 30

Leu Gln Asn Leu Ala Gly Gln His Ser Asp Val Leu Glu Asn Leu Thr
35 40 45

Pro Pro Val Arg Lys Arg Val Glu Phe Leu Arg Glu Ile Gln Asn Gln
50 55 60

Tyr Asp Glu Met Glu Ala Lys Phe Phe Glu Glu Arg Ala Ala Leu Glu
65 70 75 80

Ala Lys Tyr Gln Lys Leu Tyr Gln Pro Leu Tyr Thr Lys Arg Tyr Glu
85 90 95

Ile Val Asn Gly Val Val Glu Val Glu Gly Ala Ala Glu Glu Val Lys
100 105 110

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

Phe Trp Leu Ile Ala Leu Lys Asn Asn Glu Ile Thr Ala Glu Glu Ile
130 135 140

Thr Glu Arg Asp Glu Gly Ala Leu Lys Tyr Leu Lys Asp Ile Lys Trp
145 150 155 160

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

Gln Asn Pro Tyr Phe Lys Asn Thr Val Leu Thr Lys Thr Tyr His Met
180 185 190

Ile Asp Glu Asp Glu Pro Ile Leu Glu Lys Ala Leu Gly Thr Glu Ile
195 200 205

Glu Trp Tyr Pro Gly Lys Cys Leu Thr Gln Lys Ile Leu Lys Lys Lys
210 215 220

Pro Lys Lys Gly Ser Lys Asn Thr Lys Pro Ile Thr Lys Thr Glu Asp
225 230 235 240

Cys Glu Ser Phe Phe Asn Phe Phe Ser Pro Pro Gln Val Pro Asp Asp
245 250 255

ES 2 403 414 T3

Asp Glu Asp Leu Asp Asp Asp Met Ala Asp Glu Leu Gln Gly Gln Met
260 265 270

Glu His Asp Tyr Asp Ile Gly Ser Thr Ile Lys Glu Lys Ile Ile Ser
275 280 285

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

Asp Ile Glu Asp Asp Asp Asp Glu Ile Asp Glu Asp Asp Asp Glu Glu
305 310 315 320

Asp Glu Glu Asp Asp Glu Asp Asp Glu Glu Glu Asp Asp Glu Asp Asp
325 330 335

Asp Glu Glu Glu Glu Ala Asp Gln Gly Lys Lys Ser Lys Lys Lys Ser
340 345 350

Ser Ala Gly His Lys Lys Ala Gly Arg Ser Gln Leu Ala Glu Gly Gln
355 360 365

Ala Gly Glu Arg Pro Pro Glu Cys Lys Gln Gln
370 375

<210> 28

<211> 1624

<212> ADN

5 <213> *Arabidopsis thaliana*

<400> 28

ES 2 403 414 T3

```

gccacccaga aaaaaccctc aagtcttctt cttcttctc aatctctcca cctcttttca 60
aaccttcttc acactctctc tcaatcaatc ctttttcttc tcaaactctt cagttttgat 120
ctctaaatth ccagaaaatg agcaacgata aggacagttt caatgtcagc gatctcactt 180
ctgctcttaa agatgaggat cgagctgggc ttgtcaacgc tcttaagaac aagctccaga 240
atctagctgg acaacattct gatgtgctcg agaactctgac tcctaaaatt agaaggcgtg 300
ttgaggtttt gcgggagatt cagggcaaac atgatgaaat agagacaaaa ttccgcgagg 360
agagagctgc tcttgaagcc aagtatcaaa agttatatca gcctttgtat aacaagcgtt 420
atgagattgt gaatggagct actgaagttg aaggggctcc agaggatgct aagatggacc 480
aaggagacga gaaaactgca gaagagaaag gagtccctag tttctggctg actgctctga 540
aaaataatga tgttatatct gaagagatca cagagcgtga tgaaggagcc cttatatatc 600
ttaaagatat caagtgggtc aagattgaag aaccaaaggg attcaaactt gagtttttct 660
tcgaccagaa tccttacttc aaaaacaccc tattaacaaa ggcgtatcat atgattgatg 720
aagatgagcc tctgcttgag aaggctattg ggacagagat tgattgggat cctggaaaat 780
gcttaactca gaagattctt aagaagaagc ctaagaaagg tgcaaagaat gccaagccaa 840
ttacaaaaac tgaagattgt gaaagcttct tcaacttctt caatcctccc caagttcctg 900
attatgatga agacattgac gaagaaagag ccgaggaact tcagaatctg atggaacaag 960
attatgacat tggttctaca atccgggaga agatcatacc tcatgctgtc tcatggttta 1020
ctggtgagcc tattgaggga gaggagtttg aaatagacaa tgacgatgaa gatgatctg 1080
atgaggatga agatgaggat gaagaagatg aagacgaaga tgaggaagaa gacgacgaag 1140
atgaggagga agaagtaagc aagacaaaa agaagccatc agtcttacac aagaaaggag 1200
ggagacctca ggttaccgat gatcaacaag gagagaggcc tcctgaatgc aaacaacagt 1260
aaacaaaatc gaaaagtcta aacgaaaacc agtaaaagaa aaacaaatgt tttgggtttt 1320
gagtgaagtt tcatggccta gttttttgct tccatgtaag gcaaaatgtt ttgaagactg 1380
ctcataggaa tgttgctgta ggcaaaagag tgagtttctc catgtggaga tacttgataa 1440
attatthttg gtgcatttgt tttttttttt tttaatcact aagttgaatt ttggtgtgtt 1500
cgtcaaaatt atatcttttt accacttgaa ttaagtctct tttggtttct ttaatttaaa 1560
aataaataaa tcttatcatt gttttttttg tgtggacata agtgtattat tcttattgta 1620
aacc 1624

```

<210> 29

<211> 374

<212> PRT

5 <213> *Arabidopsis thaliana*

<400> 29

ES 2 403 414 T3

Met Ser Asn Asp Lys Asp Ser Phe Asn Val Ser Asp Leu Thr Ser Ala
 1 5 10 15

Leu Lys Asp Glu Asp Arg Ala Gly Leu Val Asn Ala Leu Lys Asn Lys
 20 25 30

Leu Gln Asn Leu Ala Gly Gln His Ser Asp Val Leu Glu Asn Leu Thr
 35 40 45

Pro Lys Ile Arg Arg Arg Val Glu Val Leu Arg Glu Ile Gln Gly Lys
 50 55 60

His Asp Glu Ile Glu Thr Lys Phe Arg Glu Glu Arg Ala Ala Leu Glu
 65 70 75 80

Ala Lys Tyr Gln Lys Leu Tyr Gln Pro Leu Tyr Asn Lys Arg Tyr Glu
 85 90 95

Ile Val Asn Gly Ala Thr Glu Val Glu Gly Ala Pro Glu Asp Ala Lys
 100 105 110

Met Asp Gln Gly Asp Glu Lys Thr Ala Glu Glu Lys Gly Val Pro Ser
 115 120 125

Phe Trp Leu Thr Ala Leu Lys Asn Asn Asp Val Ile Ser Glu Glu Ile
 130 135 140

Thr Glu Arg Asp Glu Gly Ala Leu Ile Tyr Leu Lys Asp Ile Lys Trp
 145 150 155 160

Cys Lys Ile Glu Glu Pro Lys Gly Phe Lys Leu Glu Phe Phe Phe Asp
 165 170 175

Gln Asn Pro Tyr Phe Lys Asn Thr Leu Leu Thr Lys Ala Tyr His Met
 180 185 190

Ile Asp Glu Asp Glu Pro Leu Leu Glu Lys Ala Ile Gly Thr Glu Ile
 195 200 205

Asp Trp Tyr Pro Gly Lys Cys Leu Thr Gln Lys Ile Leu Lys Lys Lys
 210 215 220

Pro Lys Lys Gly Ala Lys Asn Ala Lys Pro Ile Thr Lys Thr Glu Asp
 225 230 235 240

Cys Glu Ser Phe Phe Asn Phe Phe Asn Pro Pro Gln Val Pro Asp Asp
 245 250 255

Asp Glu Asp Ile Asp Glu Glu Arg Ala Glu Glu Leu Gln Asn Leu Met
 260 265 270

Glu Gln Asp Tyr Asp Ile Gly Ser Thr Ile Arg Glu Lys Ile Ile Pro
 275 280 285

ES 2 403 414 T3

His Ala Val Ser Trp Phe Thr Gly Glu Ala Ile Glu Gly Glu Glu Phe
 290 295 300
 Glu Ile Asp Asn Asp Asp Glu Asp Asp Ile Asp Glu Asp Glu Asp Glu
 305 310 315 320
 Asp Glu Glu Asp Glu Asp Glu Asp Glu Glu Glu Asp Asp Glu Asp Glu
 325 330 335
 Glu Glu Glu Val Ser Lys Thr Lys Lys Lys Pro Ser Val Leu His Lys
 340 345 350
 Lys Gly Gly Arg Pro Gln Val Thr Asp Asp Gln Gln Gly Glu Arg Pro
 355 360 365
 Pro Glu Cys Lys Gln Gln
 370

<210> 30

<211> 990

<212> ADN

5 <213> *Arabidopsis thaliana*

<400> 30

```

atgagcaacg aagaaaacat caaatctgat aataagagcg gcgattcctc tgatctcctc      60
accattcccg ccttagatat tggggcagag gaatgtgatc ttcttcgaga gcttaaggca      120
agtcacttca aattgltgat aaaaattcac acaaacctaa ccttaaagcg accatttgat      180
gtgaaaaaac tctcacctaa agttaccaa cgtgttctgt tcctcaaaga cattcagggt      240
acacacgatg aactcgaaga gaagtttctt gctgagaaat ctgcattgga ggcaacatat      300
gataatctct acaagccgct ttttgctaag aggtatgaaa ttgtgaatgg tgtggtcgaa      360
gctgaagcag agaaagaagg agttcccaat ttctggttga ttgcaatgaa aaccaatgaa      420
atgctcgcaa atgagataac ggaaagagat gaggcagcat tgaagtatct taaggacatc      480
agatcttgca gagttgaaga cacttcaaga aatttcaagc tggagtttct ctttgattct      540
aatctttact tcaagaactc ggttctgtct aaaacttacc atgtgaacga tgaagatggt      600
cctgttcttg agaaagtgat tggaacggac atagaatggt ttccaggtaa atgtttgact      660
cataaggttg ttgtgaagaa gaaaacaaag aaagggccaa agaaggtcaa caacatcccc      720
atgaccaaaa cagaaaactg cgagagtttc ttcaatttct tcaagccacc tgagattcct      780
gagattgatg aagttgacga ttacgatgat tttgatacca ttatgacgga agaactacaa      840
aacctgatgg accaagacta tgacattgct gtgacaatcc gagataaact gatccctcat      900
gcagtttcat ggtttacggg agaggctctt gttgatgaag acgattctga tgataatgat      960
gatgatgata atgatgagaa gagtgactaa
  
```

<210> 31

<211> 329

10 <212> PRT

<213> *Arabidopsis thaliana*

<400> 31

ES 2 403 414 T3

<220>

<223> Dominio NAP

<400> 32

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

5 <210> 33

<211> 10

<212> PRT

<213> Secuencia artificial

<220>

10 <223> Secuencia distintiva

<400> 33

Ser Ile Glu Lys Leu Gln Glu Ile Gln Asp
 1 5 10

REIVINDICACIONES

- 5 1. Método para aumentar el rendimiento de semillas de una planta con relación a plantas de tipo natural correspondientes, que comprende aumentar la expresión de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 en una planta introduciendo dicha secuencia de ácido nucleico en una planta mediante transformación, en el que dicha secuencia de ácido nucleico que codifica para dicha proteína de tipo NAP1 se selecciona del grupo que consiste en:
- (i) un ácido nucleico representado por SEQ ID NO: 1;
- (ii) un ácido nucleico que codifica para una proteína de tipo NAP1 representada por SEQ ID NO: 2, y
- 10 (iii) un ácido nucleico que codifica para una proteína que es un homólogo para la proteína representada en SEQ ID NO: 2; en el que dicha proteína que es un homólogo tiene una identidad de secuencia de al menos el 50% con la proteína representada por SEQ ID NO: 2,
- (iv) una variante alternativa de corte y empalme de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 definida en (i), (ii) o (iii);
- 15 (v) una variante alélica de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 definida en (i), (ii) o (iii); y
- (vi) una secuencia de ácido nucleico que puede hibridarse con un ácido nucleico que codifica para un polipéptido mostrado en SEQ ID NO: 2 en condiciones rigurosas;
- 20 en el que la proteína de tipo NAP1 comprende un dominio NAP, comprendiendo dicho dominio NAP una secuencia distintiva (T/S) FF (T/N/S/E/D) (W/F) (L/F) y/o la secuencia de aminoácidos conservada facilitada en SEQ ID NO: 33; y
- en el que la proteína de tipo NAP1 comprende un extremo C-terminal ácido, siendo dicho extremo C-terminal ácido el extremo carboxi-terminal de dicha proteína que tiene de 20 a 25 aminoácidos de largo, del que al menos 13 residuos son ácido glutámico y/o aspártico; y
- en el que la proteína de tipo NAP1 tiene actividad de inhibición de la fosfatasa PP2a.
- 25 2. Método según la reivindicación 1, en el que dicho ácido nucleico se deriva de una planta, preferiblemente de una planta dicotiledónea, más preferiblemente de la familia *Brassicaceae*, lo más preferiblemente de *Arabidopsis thaliana*.
- 30 3. Método según las reivindicaciones 1 y 2, en el que dicho aumento del rendimiento de semillas comprende uno o más de aumento del número de semillas llenas, aumento del peso total de semillas o aumento del índice de cosecha, cada uno con relación a plantas de tipo natural correspondientes.
4. Método según cualquiera de las reivindicaciones 1 a 3, en el que dicho ácido nucleico que codifica para una proteína de tipo NAP1 está unido operativamente a un promotor constitutivo, preferiblemente un promotor GOS2.
- 35 5. Uso de una secuencia de ácido nucleico que codifica para una proteína de tipo NAP1 según la reivindicación 1, para aumentar el rendimiento de semillas de una planta, en el que se aumenta el rendimiento de semillas de la planta aumentando la expresión de dicha secuencia de ácido nucleico, en el que la expresión de dicho ácido nucleico se aumenta introduciendo dicha secuencia de ácido nucleico en la planta mediante transformación.
- 40 6. Uso de una proteína de tipo NAP1 codificada por un ácido nucleico según la reivindicación 1, para aumentar el rendimiento de semillas de una planta, en el que se aumenta el rendimiento de semillas de la planta aumentando el nivel de dicha proteína de tipo NAP1, en el que se aumenta el nivel de dicha proteína de tipo NAP1 introduciendo el ácido nucleico que codifica para dicha proteína de tipo NAP1 en la planta mediante transformación.
7. Uso según la reivindicación 5 ó 6, en el que dicho aumento del rendimiento de semillas comprende uno o más del peso total de semillas, el número de semillas llenas y el índice de cosecha.
8. Uso de cualquiera de las reivindicaciones 5 a 7, en el que la secuencia de dicho ácido nucleico que codifica para una proteína de tipo NAP1 o de la propia proteína de tipo NAP1 es tal como se representa en SEQ ID NO: 1 ó 2.

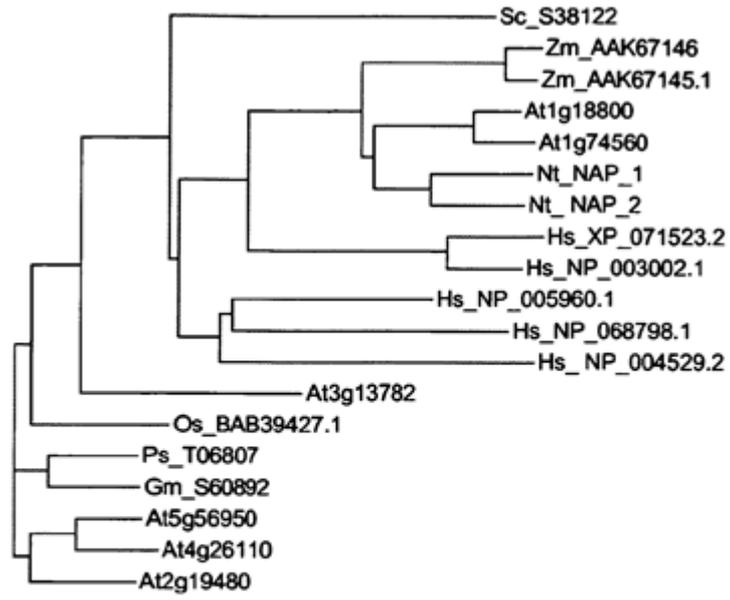


FIGURA 1

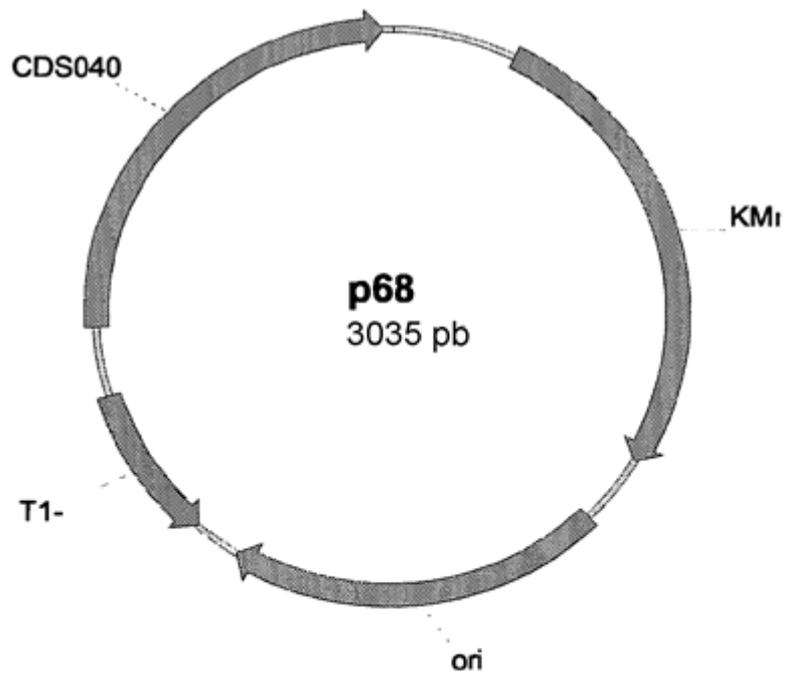


FIGURA 2

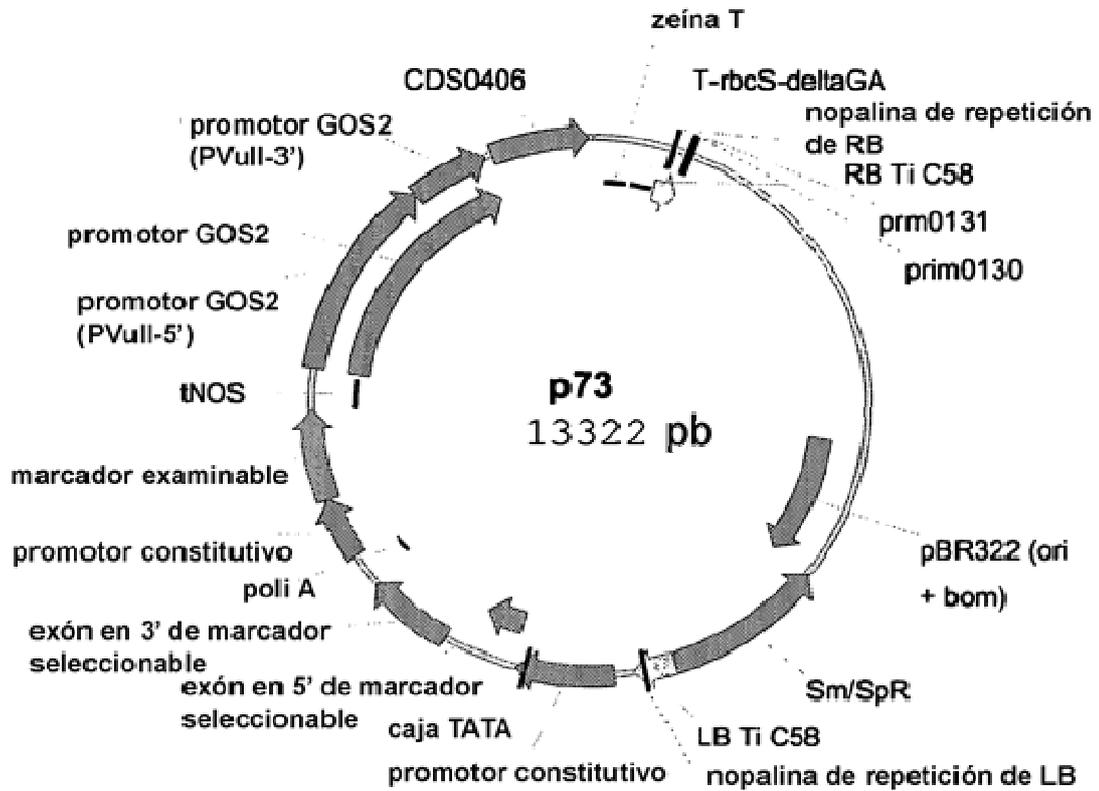


FIGURA 3

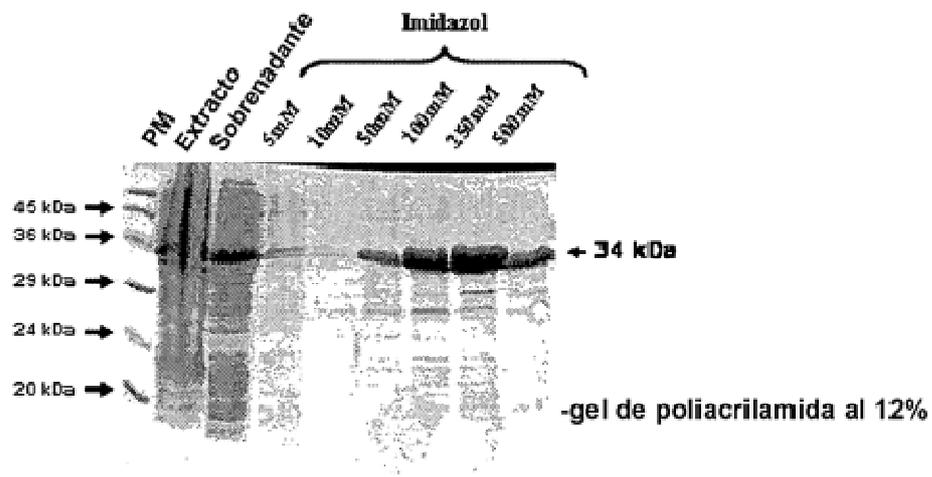
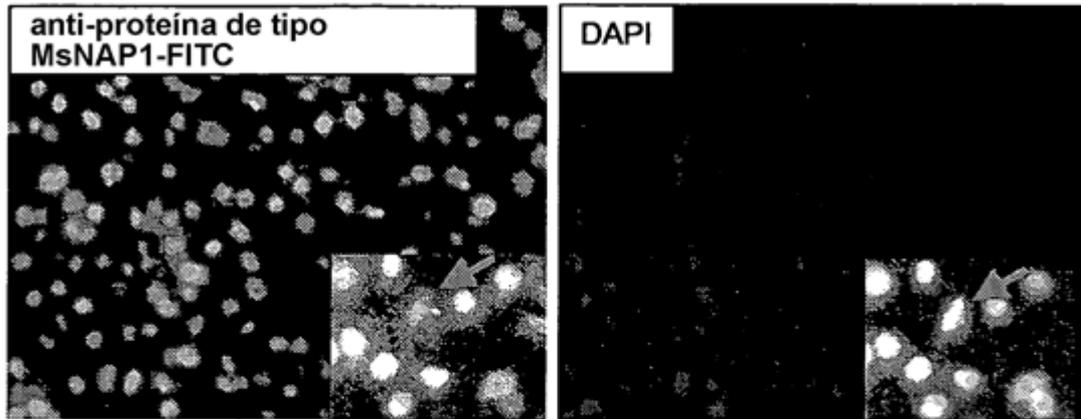


FIGURA 4

A



B

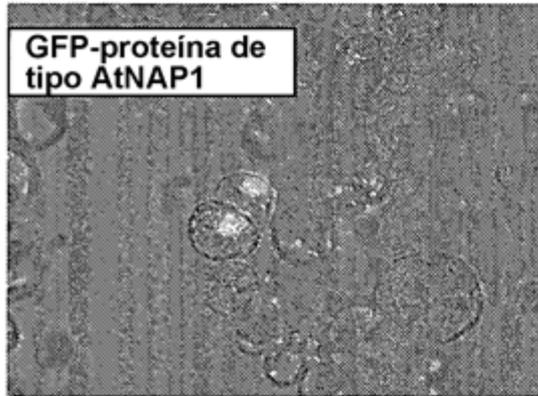


FIGURA 6

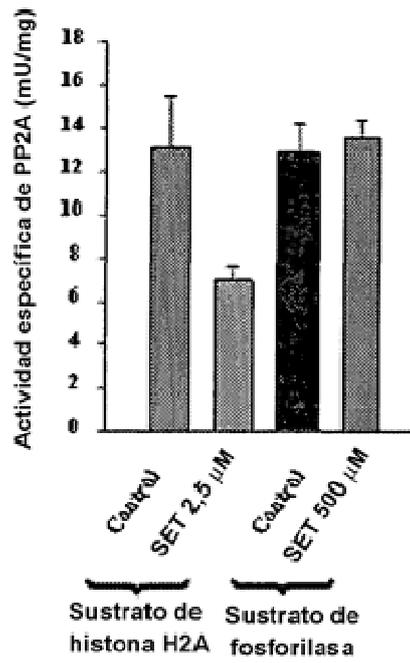


FIGURA 7

SEQ ID NO: 1, secuencia codificante para proteína de tipo NAP1 de *Arabidopsis thaliana* (AtNAP1), codón de iniciación y de terminación en negrita

gtgacggaaccacaagaagagaagagaccaaaggaggagccaaaatcctctctttcttt
ggaattagggtttctcctcaaaggaagtgaactgaaaat**tggt**tcgaggacaagagcaagaagt
cgaaaattgaagagaaaggcgaagaagaaaacttggagcaaatcgacgcagagcttggtc
tctcaattgagaagcttcaggagattcaagacgacctcgagaagattaacgaaaaggcca
gtgacgaggtccttggaaagtagagcagaaaataaacgtgatacggaaacctgtctatgaca
agcgcaatgaagttatccaatcgattcctggcttttggatgactgcttttttgagtcac
ctgccttagggcacctcttgaactgaagaagaccaaagatttttaagtaacttgaactctc
tggaaagtggaggatgccaaagatgtgaaatctggatactctataacttttcacttcactt
caaaccogtcttttgaggatgccaaagcttaccagacatttactttccttgaagaaggaa
caacaaaaatcactgcaactcctatcaaatggaaggagggaagggttgcacaaatggag
tgaacatgatgataaagaaggaaataaacgtgacattgccagaggagagtttctttactt
ggtttactgatgctcaacataaggaagatgctggggatgagattcatgatgaggttgctg
atattatcaaggaagatctctgggtccaacctctcacctacttcaacaatgatgctgatg
aagaggattttgatggagatgatgatgggtgacgaagaggagagaagaagacgatgacgatg
aagaggaggaagatgggtgaggaa**tga**tgggagcccaaagataaacacattgctggcttgc
tctataacagatgtgtaaagtttgggttatgaggttctcaatttttagcaatgatgagact
aagctttctcttttggaaatatttagtttatttactatcaatagctacattctgtttgtac
gaacatgtcatcatccatgtcctaataatcttgcataactacatctgtttttcgc

SEQ ID NO: 2, secuencia de proteína de tipo NAP1 de *Arabidopsis thaliana* (AtNAP1), secuencia de proteína deducida

MVADKSKKSKIEEKGEEENLEQIDAELVLSIEKLQEIQDDLEKINEKASDEVLEVEQKYN
VIRKPVYDKRNEVIQSI PGFWMTAFLSHPALGDLLEEDQKIFKYLNSLEVEDAKDKVKS
YSITFHFTSNPFFEDAKLTKTFTFLEEGTKITATPIKWKEGKGLPNGVNHDDKKGKRA
LPEESFFTWFTDAQHKEDAGDEIHDEVADI IKEDLWSNPLTYFNNDAADEEDFDGDDDGDE
EGEEDDDDEEEEDGE

SEQ ID NO: 3, casete de expresión (PRO0129: : CDS0406 (proteína de tipo NAP1 de *Arabidopsis thaliana* (AtNAP1a)) - zeína T - terminador T-rbsc-deltaGA; 1-2193: promotor GOS2, 2246-3016: CDS de tipo NAP1, 3126-3320: terminador de zeína T, 3364-3560: terminador T-rbs

aatccgaaaagtttctgcaccgttttcaccocctaactaacaatatagggaaacgtgtgct
aaatataaaaatgagaccttatatagtagcgctgataactagaactatgcaagaaaaact
catccacctactttagtggaatcgggctaaataaaaaagagtcgctacactagtttctg
tttcttagtaattaagtgggaaaatgaaatcattattgcttagaatatacgttcacatc
tctgtcatgaagttaaattattcgaggtagccataattgtcatcaaactcttcttgaata
aaaaaatctttctagctgaactcaatgggtaaagagagagatttttttaaaaaataga
atgaagatattctgaacgtattggcaaagatttaaacatataattatataattttatagt

FIGURA 8

ttgtgcattcgtcatatcgcacatcattaaggacatgtcttactccatcccaatTTTTat
 ttagtaattaaagacaattgacttatttttatttattatcttttttcgattagatgcaag
 gtacttacgcacacactttgtgctcatgtgcatgtgtgagtgacacctcctaatacacgt
 tcaactagcaacacatctctaatacactcgcctatttaatacatttaggtagcaatc
 tgaattcaagcactccaccatcaccagaccacttttaataatactctaaatacaaaaaat
 aattttacagaatagcatgaaaagtatgaaacgaactatttaggtttttcacatacaaaa
 aaaaaagaattttgctcgtgcgcgagcgccaatctcccatattgggcacacaggcaaca
 acagagtggctgccacagaacaaccacaacaaaaacgatgatctaacggaggacagcaag
 tccgcaacaaccttttaacagcaggctttgcgccaggagagaggaggagaggcaagaa
 aaccaagcatcctcctcctcccatctataaattcctcccccttttcccctctctatata
 ggaggcatccaagccaagaagaggagagcaccaaggacacgcgactagcagaagccgag
 cgaccgccttcttcgatccatctctccggctcgagttcttggctgatctcttcccctcctc
 cacctcctcctcacagggtatgtgcccttcgggtgttcttggatttattgttctaggttg
 tgtagtacggcggtgatgttaggaaaggggatctgtatctgtgatgattcctgttcttg
 gatttgggatagaggggttctttagtgttgcatttatcggttcgggttgattagtagtat
 ggttttcaatcgtctggagagctctatgaaatgaaatggtttagggtaggaatcttgc
 gatttgtgagtagcttttggttgaggtaaaatcagagcaccgggtgattttgcttgggtg
 aataaaagtacgggtgtttggctcctcgattctggtagtgatgcttctcgatttgacgaag
 ctatcctttggttattccctattgaacaaaaataatccaactttgaagacgggtcccggtg
 atgagattgaatgattgattcttaagcctgtccaaaatttcgcagctggcttgttagat
 acagtagtccccatcacgaaatcatggaaacagttataatcctcaggaacaggggattc
 cctgttcttccgatttgctttagtcccagaatttttttcccaaatcttataaaagtca
 ctttctgggttcagttcaatgaattgattgctacaaataatgcttttatagcgttatccta
 gctgtagttcagttaataggaataaccctatagtttagtcaggagaagaacttatccga
 tttctgatctccatttttaattatataatgaaatgaaactgtagcataagcagtagttcatttg
 attatttttttattagctctcacccttcattattctgagctgaaagtctggcatgaac
 tgtcctcaattttgttttcaaatcacatcgattatctatgcattatcctcttgtatcta
 cctgtagaagtttcttttgggtattccttgactgcttgattacagaaagaaatttatga
 agctgtaatcgggatagttatactgcttgttcttatgattcatttcttcttggcagttct
 tgggtgtagcttgccactttcaccagcaaagttcatttaaatcaactagggatatacaag
 tttgtacaaaaaagcaggctTCACAatggctcgcggacaagagcaagaagtcgaaaattga
 agagaaaggcgaagaagaaaacttggagcaaatcgacgcagagcttgttctctcaattga
 gaagcttcaggagattcaagacgacctcgagaagattaacgaaaaggccagtgacgaggt
 cttggaagttagagcagaaatataacgtgatcggaaacctgtctatgacaagcgcaatga
 agttatccaatcgattcctggcttttggatgactgcttttttgagtcactcctgccttagg
 cgacctcttgactgaagaagacccaaaagatttttaagtacttgaactctctggaagtga
 ggatgccaaagatgtgaaatctggatactctataacttttcaactcacttcaaacccggtt
 ctttgaggatgccaaagcttaccagacatttactttccttgaagaaggaacaacaaaaat
 cactgcaactcctatcaaatggaaggaggcaagggttgccaaatggagtgaacctatga
 tgataagaaaggaaataaacgtgcattgcccagaggagagtttcttacttgggttactga
 tgctcaacataaggaagatgctggggatgagattcatgatgaggttgctgatattatcaa
 ggaagatctctggtccaacctctcacctacttcaacaatgatgctgatgaagaggattt
 tgatggagatgatgatggtgacgaagaggagagaagaagacgatgacgatgaagaggagga
 agatggtgaggaatgatgggaccagcttcttgtacaaagtgggtgatatacaagcccg

FIGURA 8 (continuación)

ggcgggtcttctagggataaacagggtaattatatccctctagatcacaagcccgggcggtc
 ttctacgatgattgagtaataatgtgtcacgcatcaccaatgggtggcagtgctcagtggtga
 gcaatgacctgaatgaacaattgaaatgaaaagaaaaaaagtactccatctgttccaaat
 taaaattcattttaaccttttaatagggttatacaataattgatatagttttctgtata
 tgtctaatttgttatcattccgggcggtcttctagggataaacagggtaattatatccctct
 agacaacacacaacaaa taagagaaaaaacaataatattaatttgagaatgaacaaaag
 gaccatatcattcattaactcttctccatccatttccatttcaagttcgatagcgaaaa
 ccgaataaaaaaacacagtaattacaagcacaacaaatgggtacaagaaaaacagttttcc
 caatgccataatactcgaac

SEQ ID NO:4, cebador sentido prm1505

GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACAATGGTCGCGGACAAGAG

SEQ ID NO:5, cebador inverso prm1506

GGGGACCAC TTTGTACAAGAAAGCTGGGTCCCATCATTCTCACCATC

SEQ ID NO:6, proteína de tipo NAP1 de *Nicotiana tabacum*, codón de iniciación y terminación en negrita y subrayado (NtNAP1a)

agcggctggtaaccggtccggaattcccgggatatcgtcgacccacgcgctccgaaagaaga
 gaaaaagatgggtgctgaca aaggaagaagcaaaaagtgagggaagagaacaacaccat
 tgatgggtgagctcgtttttccattgaaaaattgcaagaaatacaagacgagctcgagaa
 gatcaatgaggaagcaagtgataaagtattggaagtggaaacagaagtacaatgagatccg
 caagcctgtctatgacaaacgaaacgacatcattaaagctatcccggactctcggttgac
 tgcTTTTtgagtcactctgtccttaggtgaaacttctaactgaagaagacccaaaagatctt
 caagtttctaagttctattgaagttgaagactctaaagatgtgaagtcgggctactcgat
 aacctttaacttcaatgCGAATCCTTATTTGAAAATACAAAGCTCACAAAGACCTATAC
 ctctcttgaagatggaccacaagatttctgctacaacaataaaaatggaaagaaggcat
 gggcattcctaattggatttgcacatgagaagaaaggaacaagcagatctcatgctgagga
 aagcttcttcacatgggttcagtgaagtcaatcaaaaagatgaggatgaggatgaggccct
 agagattcaggatgaggctcgctgacataattaaggatgacttgtggccgaaccctctcac
 ctattttaacaacgagcctgatgaagaagatttggatggatgacgagggaaaggacagtga
 aggtctgaagacgaagaggaagaagaagaggaggatgaggatgggtgatgaagaatgaag
 gcagtaaaactgttcaagaccctattttgggatctcgtcttcagcggttttaatcatcag
 ggtttaatgtctgtaaagaggctttgaaatgttgccaaagaacagaataactgtgggtgact
 ataccttttctctctgtatggttataacttataagcaaaaatctaatccgggaggtt
 ccaaaatgttttcattaggtagttcgattaatgaagtgtttgtctggcaaaaactgata
 atgttaggttattgagttatg

FIGURA 8 (continuación)

**SEQ ID NO: 7, proteína de tipo NAP1 de *Nicotiana tabacum* (NtNAP1a),
secuencia de proteína deducida**

MGADKGGKQKVEEENNTIDGELVFSIEKLQEIQDELEKINEEASDKVLEVEQKYNEIRKP
 VYDKRNDI I K A I P D F W L T A F L S H P V L G E L L T E E D Q K I F K F L S S I E V E D S K D V K S G Y S I T F
 N F N A N P Y F E N T K L T K T Y T F L E D G P T K I S A T T I K W K E G M G I P N G F A H E K K G N K R S H A E E S F
 F T W F S E V N Q K D E D E D E A L E I Q D E V A D I K D D L W P N P L T Y F N N E P D E E D F D G D E G K D S E G S
 E D E E E E E E E D E D G D E E

**SEQ ID NO: 8, proteína de tipo NAP1 de *Nicotiana tabacum*, codón de
 iniciación y terminación en negrita y subrayado (NtNAP1b)**

tttgtaaaaaaagcaggctggtaccggtccggaattcccgggatatcgctcgacccacgc
 gtccgagaaattagcagttagagacactgagaagcagcagctctcttcctcagctgctgt
 gtgcttaggcaagaataaaa**atg**ggggcagacaaaggaagaagcagaaagtggatgagg
 aaaacaacaatgttattgatgaaaagctcatttttccattgaaaaattgcaagagatac
 aagacgagctcgagaagatcaatgaaaaagcaagcgacgaagtgttggaaagtagaacaga
 agtacaacgagatccgcaagcctgtctacgataagcgaaatgatgtcattagctctattt
 ctgacttctggttgactgcttttttgagtcatcctgttcttggtaaccttctcactgaag
 aggacaaaagattttcaaatttgtaagttctattgaagtggaaactcaaaggatgtga
 aatcgggtcattcaatcacgtttaactttaagcccaatccttattttgaaaattcaaagc
 tctcaaagacgtataccttcttgaagatggacctacaaaattacagctacaacaataa
 aatggaaagaaggcatggcattcctaattggagttgctgacaagaagaaggaaacaagc
 ggtcccacgctgaagaaagtttctttacatgggttcagtgaagtcaatcaaaaaggtgatg
 tggatgatgacgaaaatgagattctggacattcaggatgatgaggttgctgaaataatca
 aggatgacttggtggcctaaccctctcaattattttgaccatgagcctgatgaagaagata
 ttgagggcgatgagggaaaggacagcggaggctctgaagaggaagaagaagaggaagatg
 atgaagatgaagaagacgaa**tga**actgttggtagacctgtgtttgatttgagttctcat
 cagtgtttcaatcatcagagttggtctctgtaaagaggttccggatattgcagaaaaatt
 gaatgacatatagtggtgactctaatttttagtttcagtga

**SEQ ID NO: 9, proteína de tipo NAP1 de *Nicotiana tabacum* (NtNAP1b),
 secuencia de proteína deducida**

MGADKGGKQKVD E E N N N V I D E K L I F S I E K L Q E I Q D E L E K I N E K A S D E V L E V E Q K Y N E I R K
 P V Y D K R N D V I S S I S D F W L T A F L S H P V L G N L L T E E D Q K I F K F V S S I E V E D S K D V K S G H S I T
 F N F K P N P Y F E N S K L S K T Y T F L E D G P T K I T A T T I K W K E G M G I P N G V A D K K K G N K R S H A E E S
 F F T W F S E V N Q K G D V D D D E N E I L D I Q D D E V A E I I K D D L W P N P L N Y F D H E P D E E D I E G D E G K
 D S G G S E E E E E E D D E D E D E E

FIGURA 8 (continuación)

SEQ ID NO:10, proteína de tipo NAP1 de *Medicago sativa* (Ms10.1), codón de iniciación y de terminación en negrita y subrayado

GGCACGAGCAAACCCCTAACACTTCCCTCATTACGCTCGAAGAAAAGAACAACAAATCTC
TCCACTGCGCTAGGGTTTCAAACCCAACACCTTCTGTTTCTTCAACC**ATG**GTGGCCGACA
AGTCTAAGAAGTTGAAAGTTTCTGAAAAGGGTGAACCGCTGAAGAGATCGACGGAGAAC
TTGTTCTCTCCATTGAAAAGTTGCAGGAAATTCAAGATGAGATTGAAAAGATCAATGAGG
AGGCTAGCGATAAAGTTCTCGAAATAGAGCAGAAGTACAATGAAGTAAGGAAACCGGTGT
ATGACAAGCGCAATGATGTGATCAAGTCCATTCCCGATTTCTGGCTAACTGCGTTTTTGA
GCCATCCTGTTCTTGGTGATCTCTTGAATGAAGAAGATCAGAAGATATTTAAGCATTAA
TCTCTCTTGAGGTGGAAGATCATAAAGATGTTAAATCAGGCTATTCAATCACATTTAACT
TCGACTCCAATCCCTTTTTTGAGGATTCAAACCTTGTTAAGACTTTTACCTTCTTGAAG
AAGGAACCACAAAGCTTACCGCCACACCCATAAAATGGAAAGAGGGCAAGGGCATTCCCA
ATGGAGTTATTCATGAGAAGAAAGGGAACAAGCGAGCTGCTTCTGATATCAGTTCTTTA
CCTGGTTTTGTGACACTGAGCAGAAAGATGAAATGGGTGACATTCATGATGAGATTGCTG
AAATGATCAAGGATGATTTATGGCCGAATCCACTCAATTATTTCAACAGTGAGGACCCTG
ATGAAGCAGAGGAGGAGGATGATGAAGCTGGTGATGCGGGAAAGGATGATGATGACTCTG
AAGATGATGATGATCAAGAGGATGACGACGATGACGAGGAAGAAGAA**TAG**TGTAATAATGC
TTTAAATAGTAATACTTGGTTTTAATTTATTTATTTAAGGTTATTATAGGAGTATCTT
AGTGGTCTTTAGGGGATGATGAAAGACCAAGGTTGGCTATTGGTTTTCCCCCTNTGGGCG
TAAACCTTATTTATTGTGCTTTGAAGGTGATTTCTGGTTTTATCTTTGTGCGCTTCTTTC
AAGATACCAAT**TGA**TACATCGGATTTTATCTTAGTCCATATTGAAACCATATAGTAGTTA
AAATGTAGTATATTCAGTGTATAGCTGCGTAATCAGTATCATTTTATTGCTATCACAAC
TTACAGTACC

SEQ ID NO:11, proteína de tipo NAP1 de *Medicago sativa* (Ms10.1), secuencia de proteína deducida

MVADKSKKLKLVSEKGENAEEIDGELVLSIEKLQEIQDEIEKINEEASDKVLEIEQKYNEV
RKPVDKRNVDIKSIPDFWLTAFLSHPVLDLLNEEDQKIFKHLISLEVEDHKDVKSGYS
ITFNFDSPFFEDSKLVKTFTFLEEGTTLTATPIKWKEGKGI PNGVIHEKKGKRAASD
ISFFTWFCDTEQKDEMGDHDEIAEMIKDDLWPNPLNYFNSEDPDEAEEDDEAGDAGKD
DDSEDDDDQEDDDDEEEE

SEQ ID NO:12, secuencia codificante de nfa104 de *Zea mays* (registro GenBank AF384036)

gttccctaccttcttccctccgtctcccagctcgcgccaggcaggcgacacagcgacgctaa
aaaccctagagcgaggaggcgtgccaggccagcggttgcgatgacggcaccggcgga
aggggaagaaggccaagaccgatgccgacggcgaggagaacgagcaaatcgacggcg
cccttgtcttctccatcgagaagctccaggagattcaggacgagctcgagaaggtaatg
aggaagcaagtgacaaggttatggagggtggagcagaaatacagtgagattcgacagacctg
tctatctcaagaggggtgacattatcaagaccatcccggacttttggtcacagcgtttt
tgagccatcctctactaagtgagcttctgactgaagaggatcagaagatattcaagtact

FIGURA 8 (continuación)

tggactccattgatgctgatgattctgatgttaaggcaggatattccatttaccttaact
tctctgagaaccgctactttgaagacacaaagcttacaagacctattcctttgttgatg
atggaacaaccacaataaaagcttctcaaattaagtggaaggatggaatgggacctgcaa
atggaaatggtattaacaagaagggaaacaagcggccatttagtagtggaagtttttct
cctggtttagtgatacagagctcaagagtcttgctgatgggtgtgcaagatgagggtggcg
agatcatcaaggaagacttgtggcctaatacctttgaagtacttcaacaatgagggtgaag
atgaatttgaaggagatgaagaagatgatgacgacgacgacgacgatgataatttggatg
gtgatgacaatgacgatgatggggaccaggagaactgagccctcgcttttaggcggggaa
ttatgtg

**SEQ ID NO:13, secuencia de proteína deducida de nfa104 de *Zea mays*
(registro GenBank AAK67146)**

MTAPADKGGKAKTDADGGEENEQIDGALVFSIEKLQEIQDELEKVNNEASDKVMEVEQKY
SEIRRPVYLKRGDI IKTI PDFWLTAFLSHPLLSSELLTEEDQKIFKYLDSIDVDDSDVKAG
YSIYLNFSENPYFEDTKLTKTYSFVDDGTTTIKASQIKWKDGMGPANGNGINKKGNKRPL
VVEFFSWFSDELKSLADGVQDEVAEI IKEDLWPNPLKYFNNEVEDEFEGDEEDDDDDDD
DDNLDGDDNDDGDQEN

**SEQ ID NO:14, proteína de tipo NAP1 de *Oryza sativa* (OsNAP1a),
secuencia codificante (registro GenBank XM_472746)**

atggcggcgcgagcagaaggggaagaagccgaggaccgacggcgcgaggccgagccc
gtcgacgcccctgctgcagtcctcgcagaagctccaggagatccaggacgagatcgag
aaggtaatgaggaagcatgtgataaagtctggagttggaacagaaatacaacgaggtt
cgcagaccagtttatgttcgacggaataaaattatcaagcaaattcctgacttctggctg
acagcgtttcttagccatcctatgcttgggtgaactattaactgaagatgatcaaaagatt
ttcaaacacttggagtcctatcgacgtggatgactcagaagatatcaaatcaggctactcc
attactctcacattctcccccaatccatattttgaagatacaaaagcttacaaaaacatat
tccttttagtgacgatgaagcagtc aaagtaaaggctacctccatcagggtggaagaaagga
atggatattgccaatgatcgtgcgtacacgaagaaaggggacaagcgaatcttaattgat
gaaagtttcttacttggttcaatagtgaaaagaacagaagttttgctcatggagctatg
gatgagggtggcagatgtcatcaaggaagatctgtggcctaatacctttgaagtacttcaac
aatgaatttgaagaagaattagagctactggatgacgatgacgaggatctgatgatgac
gatgaggaggaggatgatgaagaccaaggtgaaggagaggaggatggagaggagaactga

**SEQ ID NO:15, proteína de tipo NAP1 de *Oryza sativa* (OsNAP1a),
secuencia de proteína deducida (registro GenBank XP_472746)**

MAAAEQKGGKPRTDGAEAEPVDAALLQSIEKLQEIQDEIEKVNNEACDKVLELEQKYNEV
RRPVYVRRNKIIKQIPDFWLTAFLSHPMLGELLTEDDQKIFKHLESIDVDDSEDIKSGYS
ITLTFSPNPYFEDTKLTKTYSFSDDEAVKVKATSIRWKKGMDIANDRAYTKKGDKRILID
ESFFTWFNSEKNRSFAHGAMDEVADVIKEDLWPNPLKYFNNEFEELELLDDDDDEVSDDD
DEEEDDEDQGEEDGEEN

FIGURA 8 (continuación)

aagaccatcccggacttttggtcacagcgtttatgagccatcctctattaagtgagctt
ctgactgaagaggaccagaagatattcaagtacttagactccattgatgtggatgattct
gatgtaaggcaggatactccattcatcttaacttctctgagaaccgactttgaggac
acaagcttgcaagacctatatactttgctgatgatggaacaaccacaataaaagcttcc
gaaattaagtggaaggaaggaatgggacctgcaaatggaaatggtattaacaagaagggg
agtaagcggccattagtagaggaaagttttttagctggtttggtgatacagagctcaag
agtcttgctgatgggtgcaagatgaggtggcggagatcataaaggaagatttgtggcct
aatcctttgaagtacttcaacaatgaggttgacgatgaatttgaaggagatgaagatgat
gatgatttggatggatgatgacgatgaaggcgatgatttggagaactgagcccttgcg
cttggttcagaatgttgtccgtggatgatgtggctgggcggaactgtgacccttttgg

**SEQ ID NO:19, secuencia de proteína deducida de nfa103 de *Zea mays*
(registro GenBank AAK67145)**

MTAPADKGGKAKTDADGGEENEQIDGVLVLSIEKLQEIQDELEKVNEEASDKVMEVEQKY
SEIRRPVYLKRGDI IKTI PDFWL TAFMSHPLLSELLTEEDQKIFKYLDSIDVDDSDVKAG
YSIHLNFSENPYFEDTKLAKTYIFADDGTTTIKASEIKWKEGMGPANGNGINKKGSKRPL
VEESFFSWFGDTELKSLADGVQDEVAEIIKEDLWPNPLKYFNNEVDDEFEGDEDDDDLDG
DDDDEGDLEN

**SEQ ID NO:20, secuencia codificante para proteína de tipo NAP1 de
Arabidopsis thaliana (registro GenBank NM_101738)**

agattcacgcacacacaatcgagtttttagggtttagcggttgctctctcggaagcca
gagagaagaggggaagaggaagtctaattcctctgcgtttttgaattagggttttctca
attggaatcgaaaatgggtgacagacaagagcaagaaggcgaaaaccgaagaagaaaacgt
cgagcaaatcgatgcagagcttgtcctctcaatcgaaaagcttcaagagatccaagacga
cctcgagaagataaatgaaaaggctagtgatgaagtgttggaaagtgagcagaaaataa
tgtgataaggaacctgtttatgacaagcgtaacgagatcatcaaaacctcctgattt
ctggtaactgctttcttgagtcaccctgcttaggtgaactttgactgaagaagacca
aaagattttcaaatatcttagctctcttgatgttgaggatgccaaagatgtgaaatctgg
atactctatcacttttcttcaatcccaatccatttttgaagatggaaaactgacaaa
gacttttacctttctcgaagaagggacaaccaaatacacagccacgcctatcaaatggaa
agagggcaaaggcctggcgaatggagtgaatcatgagaagaatggaaacaaacgtgcact
acctgaagagagcttctttacctggttagtgatgctcaacacaaggaggatgttgagga
tgagatgcaagacgagcaggttgacagatatcatcaaggaagatttgtggcccaaccctct
cacctacttcaacaatgacgctgatgaagaggactttgatggagacgatgatggagatga
agaggagaaaagaaggtgactctgatgaagatgatgacgaagaagacgaagttgggtgagga
atgatggcagggataccagaaaaccacatttgcttcatgtcttctctataacagagtg
gtaaagttttgtgtgttgaaggtttttaattttaagcaaaaagtgattatgacgacaac
agacaagcttttaattttatctaccgtaatagttatatcttgttgtaagaaccatttt
cagccttttgttggaaaatcctgcttaaatggtttttgagcttacataatagcttcttc
atcttttgtcttctaaagagaattatatttgtaatttcatgtctgttgtgtttcttga
ctttactgaatagagaatttgtgtgtttatgggtgaaaatatagccgatctgcttgac

FIGURA 8 (continuación)

SEQ ID NO:21 , proteína de tipo NAP1 de *Arabidopsis thaliana*, secuencia de proteína deducida (registro GenBank NP_564063)

MVTDKSKKAKTEENVEQIDAELVLSIEKLQEIQDDLEKINEKASDEVLEVEQKYNVIRK
PVYDKRNEI IKTI PDFWLTAFLSHPALGELLTEEDQKIFKYLSSLDVEDAKDVKSGYSIT
FSFNPFPFFEDGKLTFTFLEEGTTKITATPIKWKEGKGLANGVNHEKNGNKRALPEES
FFTWFSDAQHKEDVEDEMQDEQVADI IKEDLWPNPLTYFNNDAADEEDFDGDDDGDEEKE
GDSDEDDDEEDEVGEE

SEQ ID NO:22 , secuencia codificante de proteína de tipo NAP1 de *Lysopersicon esculentum* (LeNAP1, registro GenBank BT013059)

gcgaaatcaagaaaatcagttaagcagctctgtaactcaggtgggaaaaaggcaaaaaat
aatgggtggttgacaaaagggaagaagcagaaagtggaagaggaaagctacattgatgaaaa
gctcattttttccattgaaaaattgcaagaaatacaagacgacctgacaagatcaatga
gaaagcaagtgaggaagtgttgaaatagaacagaagtacaacaagatccgcaagcctgt
ttatgataagcggaatgatatcattaactctatttctgacttctgggtgactgctttttt
gagtcacctgttcttggtgaccttctaactgaagaggacccaaaagattttcaaattctt
aagttctattgaagtggaagactcgaaagatgtgaaatttggttactcaatcacgtttaa
ctttaagcccaatcctttctttgaaaattcaaagctctcaaagacctataccttcttga
agatggacctacaaaaatcacagctacaccaataaaaatggaaagaaggcaaggcattcc
taatggcgttgctcaggagaagaaaggaaacaagcgatcccatgctgaagagagcttctt
cacctggttcagtgaaagtcaataaaaaagatgatagcgatgatgatgaaaatgaggttct
ggagattcaggatgaggttgctgaaataatcaaggatgacttggtggccaaacctttaa
ttattttaccaatgaacctgatgaagaagattttgagggtgatgaaggtggtgatgaggg
ggaggactctgaagatgaaggtgatgaggaggaagaggaagacgacgaagatgaagatga
caaatgaactgttaatggacctcataatttgatttgatttcttcttcaatgtttcaatt
atcatagttggtatctgtaaagaagcttaatatgacagataaaaatcgaattatatatagt
ggtgactgctttttttctaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

SEQ ID NO:23 , secuencia de proteína deducida de proteína tipo NAP1 de *Lysopersicon esculentum* (LeNAP1)

MVVDKGGKQKVEEESYIDEKLI FSIEKLQEIQDDLDKINEKASEEVLEIEQKYNKIRKPV
YDKRNDIINSISDFWLTAFLSHPVLGDLLEEDQKIFKFLSSIEVEDSKDVKFGYSITFN
FKPNPFFENS KLSKTYTFLEDGPTKITATPIKWKEGKGI PNGVAQEKKGNKRSHAEESFF
TWFSEVNKKDDSDDDENEVLEIQDEVAEIIKDDLWPNPLTYFTNEPDEEDFEGDEGGDEG
EDSEDEGDEEEEDDEDEDDK

SEQ ID NO:24 , secuencia codificante para proteína de tipo NAP1 de *Arabidopsis thaliana* (registro GenBank NM_118744)

gtccctagctctcttctgcttcttcttcttcaaaaatctctctcttcaccaaattcctcagaa
gatgagcaacgacaaggatagcttcaacgtctccgatcttactgctgctcttaaggacga
ggatcgagctggccttgctcaatgctctaaagaacaagctgcagaatctggctggtcagcg

FIGURA 8 (continuación)

ttctgatgtgctcgagaatctgactcccaatgtgagaaagcgcggttgatgccttgagggga
 tatacagagccaacatgatgaactagaggcaaaattccgtgaggagagagctattcttga
 agccaagtatcaaacgctgtatcagcctttgtatgtcaagcgttatgagattgtgaatgg
 cactactgaagttgaactggctccagaggatgataccaaggtggaccaaggagaggagaa
 aactgcagaagagaaaggagttccaagtttctggctgacagctctgaaaaataacgatgt
 tatttccgaggaggtcacagagcgtgatgaagggctctcaaatacttaagatattaa
 gtggtgcaagattgaagagcctaaaggattcaaacttgagtttttcttgacacgaatcc
 gtattttaagaacactgtcttgacaaagtcttatcatatgattgatgaagatgagccact
 gcttgagaaggctatggggacagaaattgattgggtatcctggaaagtgtctaactcagaa
 gattctcaagaagaagcctaagaaagggtcaaagaataactaaaccaatcaccaaaactcga
 agattgtgaaagcttcttcaacttctttagtctccagaagttccggatgaagatgaaga
 tatcgacgaggaaagagctgaggatcttcaaacctgatggaacaagattatgacatcgg
 atctactattcgggaaaagattattcctcgtgctgtctcatggtttactggtgaggctat
 ggaagcagaggattttgaatatagatgacgatgaggaagatgacattgatgaggatgaaga
 tgaggaagacgaagaggatgaggaggacgatgatgatgaggatgaagaagaaagcaagac
 caaaaagaagccatcaatcggcaacaagaaggaggagatctcagatagttggtgaagg
 taacaagatgagagggccaccgcaatgcaagcaacagtaatcttttactacgctctacca
 gacataaaaggattgctgtaaaatataattcagggtcattctctgttcatcaagaatgagg
 attgagaaaaggtttgggatttttaaaagtgaattcatctttaggagttcgttcgt
 tttctattggtgtgtttattttctctaaagcactttaataatataccttggatttaaat
 ttatgaatcaagcatcatcatccctagtctctgcattcactactcatcccctacctaaa
 ctttgcgacgaaagagattttaataaccatttagatagtaatgggtagtgggaatgatc
 attattcttttggtcaccgctctttgattttcaatggtaaccattttggtgtgtaag

SEQ ID NO:25

, proteína de tipo NAP1 de *Arabidopsis thaliana*, secuencia de proteína deducida (registro GenBank NP_194341)

msndkdsfnvsdltaalkdedraglvnalknklqnlagqrsdvlentpnvrkrvdalrd
 iqsqhdeleakfreeraileakyqtllyqplyvkryeivngtlevelapedtkvdqgeek
 taekgvpsfwltalknndviseevterdegalkylkdikwckieepkgfklefffdtnp
 yfknvtlksyhmidedepllekamgteidwypgkcltqkilkkkpkkgskntkpitkle
 dcesffnffsppevpdededideeraedlqnlmeqdydigstirekiipravswftgeam
 eaedfeidddeeddidedededeedeedeedddededeesktkkkpsignkkggrsqivgeg
 kqderppeckqq

SEQ ID NO:26, secuencia codificante para proteína de tipo NAP1 de *Arabidopsis thaliana* (registro GenBank NM_127506)

GTGTCTTATTTCCGGTCTGGTCATTTTCTCAAAGCCCTTTTAGTTATTTATATATATATTC
 TCTGTCTCGTATTTGTCCCCAAAATCTAGGGTTTTAAGGTTTCTTATCCTTCTCTTCC
 TCCGCCAGATTCTTTTCTTGCGAAGATGAGCAACGACAAGGACAGCATGAACATGTCCGA
 TCTCTCCACCGCTCTTAACGAGGAGGATCGTGCCGGGCTTGTTAATGCTCTTAAGAACAA
 GTTGCAGAAATTTGGCTGGACAACACTCTGATGTCCTTGAAAACCTTGACTCCACCAGTCAG
 GAAGCGTGTGCGAGTTTCTAAGAGAGATTCAGAACCAATATGATGAGATGGAAGCAAATT
 CTTTGAGGAGAGAGCAGCTCTTGAAGCTAAGTATCAAAGTTATATCAGCCTTTATATAC

FIGURA 8 (continuación)

CAAGCGATATGAGATTGTGAATGGTGTGGTTCGAAGTTGAAGGTGCAGCTGAAGAAGTAAA
 ATCCGAACAAGGAGAAGATAAATCAGCTGAAGAGAAAGGAGTACCAGATTTCTGGCTTAT
 TGCATTGAAGAACAATGAAATTACTGCGGAAGAGATAACTGAGCGAGATGAAGGGGCTCT
 CAAGTATCTCAAAGATATCAAGTGGAGTAGGGTTGAAGAACCAAAGGGTTCAAGCTTGA
 GTTTTTCTTTGATCAGAATCCTTACTTCAAGAACACTGTCTTGACCAAGACATATCACAT
 GATTGATGAAGATGAGCCTATCCTTGAGAAGGCCCTCGGGACGGAGATTGAGTGGTATCC
 TGGAAAGTGTGGACACAGAAGATTCTAAAAAAGAAGCCAAAGAAAGGATCCAAAAACAC
 AAAGCCGATCACTAAGACTGAGGACTGTGAGAGTTTCTTCAACTTTTTTCAGTCCACCTCA
 AGTTCCTGACGATGATGAGGATCTTGATGATGACATGGCTGATGAACTCCAAGGACAAAT
 GGAGCATGATTATGATATCGGTTCAACAATCAAAGAGAAAATCATCTCGCATGCTGTGTC
 ATGGTTCACCTGGTGAAGCTGTTGAGGCAGATGACCTTGATATTGAGGACGACGATGATGA
 GATTGATGAAGATGATGATGAAGAGGACGAGGAAGATGATGAGGATGACGAGGAGGAGGA
 TGATGAGGATGATGACGAGGAGGAAGAAGCAGATCAAGGAAAAGAAGAGCAAAAAGAAGTC
 ATCAGCTGGGCACAAGAAGGCTGGAAGAAGTCAACTTGCAGGAAAGTCAAGCAGGTGAGAG
 GCCACCGGAATGTAAGCAGCAGTGAAGAAGTGAAGAATCTTGGCTTAGTTATGATGAAGA
 AGAAGAGTGAAGAGTGTCTTTGAGCCGAGGTTGTGTTTCTTTAATTTGCAGAGTCATGGT
 CCGGTTTATTATATATCAGTTTTGGGTGATTGGTTTGCTATTTAAAAAATAATGGGT
 TCTTTGGTTGGTTTGTGTCTCTTGATTTTTCTTTTGTAAATGATCTTATGAATTTGTTT
 CGAGTTAATGTCGTTCTCTGGTCAGATTTCAATTCTATTTATCTCCCTCGTTA
 ATGAGAGAATTTGTG

**SEQ ID NO:27, proteína de tipo NAP1 de *Arabidopsis thaliana*,
 secuencia de proteína deducida (registro GenBank NP_179538)**

MSNDKDSMNMSDLSTALNEEDRAGLVNALKNKLQNLQAGQSDVLENLTPPVRKRVEFLRE
 IQNQYDEMEAKFFEERAALAEAKYQKLYQPLYTKRYEIVNGVVEVEGAAEEVKSEQGEDKS
 AEEKGVPDFWLIALKNNEITAEEITERDEGALKYLKDIKWSRVEEPKGFKLEFFDQNPY
 FKNTVLTKTYHMI DEDEPILEKALGTEIEWYPGKCLTQKILKKPKKGSKNTPITKTED
 CESFFNFFSPPQVPDDDEDLDDDMADELQGMEDHYDIGSTIKEKII SHAVSWFTGEAVE
 ADDLDIEDDDDEI DEDDDEEDEDDEEDEDDEEDEDDEEADQGKSKSKKSSAGHKKAG
 RSQLAEGQAGERPPECKQQ

**SEQ ID NO:28, secuencia codificante para proteína de tipo NAP1
 de *Arabidopsis thaliana* (registro GenBank NM_125077)**

GCCACCCAGAAAAACCTCAAGTCTTCTTCTTCTCCTCAATCTCTCCACCTCTTTTCA
 AACTTCTTCACTCTCTCAATCAATCCTTTTCTTCTCAATCTTTTCAGTTTTGAT
 CTCTAAATTTCCAGAAAATGAGCAACGATAAGGACAGTTTCAATGTCAGCGATCTCACTT
 CTGCTCTTAAAGATGAGGATCGAGCTGGTCTTGTCACGCTCTTAAGAACAAGCTCCAGA
 ATCTAGCTGGACAACATTCTGATGTGCTCGAGAATCTGACTCCTAAAATTAGAAGGCGTG
 TTGAGGTTTTGCGGGAGATTCAGGGCAAACATGATGAAATAGAGACAAAATTCGCGGAGG
 AGAGAGCTGCTCTTGAAGCCAAGTATCAAAGTTATATCAGCCTTTGTATAACAAGCGTT
 ATGAGATTGTGAATGGAGCTACTGAAGTTGAAGGGGCTCCAGAGGATGCTAAGATGGACC
 AAGGAGACGAGAAAACCTGCAGAAGAGAAAGGAGTCCCTAGTTTCTGGCTGACTGCTCTGA

FIGURA 8 (continuación)

AAAATAATGATGTTATATCTGAAGAGATCACAGAGCGTGATGAAGGAGCCCTTATATATC
 TTAAAGATATCAAGTGGTGAAGATTGAAGAACCAAAGGGATTCAAACCTTGAGTTTTTCT
 TCGACCAGAATCCTTACTTCAAAAACACCCTATTAACAAAGGCGTATCATATGATTGATG
 AAGATGAGCCTCTGCTTGAGAAGGCTATTGGGACAGAGATTGATTGGTATCCTGGAAAAT
 GCTTAACTCAGAAGATTCTTAAGAAGAAGCCTAAGAAAGGTGCAAAGAATGCCAAGCCAA
 TTACCAAAACTGAAGATTGTGAAAGCTTCTTCAACTTCTTCAATCCTCCCCAAGTTCCTG
 ATGATGATGAAGACATTGACGAAGAAAGAGCCGAGGAACCTCAGAATCTGATGGAACAAG
 ATTATGACATTGGTTCTACAATCCGGGAGAAGATCATACTCATGCTGTCTCATGGTTTA
 CTGGTGAGGCTATTGAGGGAGAGGAGTTTGAATAGACAATGACGATGAAGATGATATCG
 ATGAGGATGAAGATGAGGATGAAGAAGATGAAGACGAAGATGAGGAAGAAGACGACGAAG
 ATGAGGAGGAAGAAGTAAGCAAGACCAAAAAGAAGCCATCAGTCTTACACAAGAAAGGAG
 GGAGACCTCAGGTTACCGATGATCAACAAGGAGAGAGGCCCTCCTGAATGCAAACAACAGT
 AAACAAAATCGAAAAGTCTAAACGAAAACAGTAAAAGAAAAACAAATGTTTTGGGTTTT
 GAGTGAAGTTTCATGGCCTAGTTTTTTTGTCTCCATGTAAGGCAAAATGTTTTGAAGACTG
 CTCATAGGAATGTTGCTGTAGGCAAAAAGAGTGAGTTTCTCCATGTGGAGATACTTGATAA
 ATTATTTTTGGTGCATTTGTTTTTTTTTTTTTAATCACTAAGTTGAATTTTGGTGTGTT
 CGTCAAAATTATATCTTTTTTACCCTTGAATTAAGTCTCTTTTGGTTTCTTTAATTTAAA
 AATAAATAAATCTTATCATTGTTTTTTTTTGTGTGGACATAAGTGTATTATTCTTATTGTA
 AACC

**SEQ ID NO: 29, proteína de tipo NAP1 de *Arabidopsis thaliana*,
 secuencia de proteína deducida (registro Genbank NM_568844)**

MSNDKDSFNVSDLTSALKDEDRAGLVNALKNKLQNLQAGQSDVLENLTPKIRRRVEVLRE
 IQGKHDEIETKFEERAALEAKYQKLYQPLYNKRYEIVNGATEVEGAPEDAKMDQGDEKT
 AEEKGVPSFWLTALKNNDVISEEITERDEGALIYLKDIKWCKIEEPKGFKLEFFDQNPY
 FKNTLLTKAYHMI DEDEPLEKAIGTEIDWYPGKCLTQKILKKPKKGAKNAKPIKTED
 CESFFNFFNPQVPDDDEDI DEERAEELQNLMEQDYDIGSTIREKII PHAVSWFTGEAIE
 GEEFEIDNDDDDI DEDEDEDEDEDEDEDEDEDEDEEEVSKTKKKPSVLHKKGGRPQVT
 DDQQGERPPECKQQ

**SEQ ID NO: 30, secuencia codificante para proteína de tipo NAP1
 de *Arabidopsis thaliana* (registro Genbank NM_112230)**

ATGAGCAACGAAGAAAACATCAAATCTGATAATAAGAGCGGCGATTCCCTCTGATCTCCCT
 ACCATTCCCGCCTTAGATATTGGGGCAGAGGAATGTGATCTTCTTGCAGAGCTTAAGGCA
 AGTCACTTCAAATTGTTGATAAAAATTCACACAAACCTAACCTTAAAGCGACCATTTGAT
 GTGAAAAACTCTCACCTAAAGTTACCAAACGTGTTCTGTTCCTCAAAGACATTCAGGTT
 ACACACGATGAACTCGAAGAGAAGTTTCTTGCTGAGAAATCTGCATTGGAGGCAACATAT
 GATAATCTCTACAAGCCGCTTTTTGCTAAGAGGTATGAAATTGTGAATGGTGTGGTTCGAA
 GCTGAAGCAGAGAAAGAAGGAGTTCCCAATTTCTGGTTGATTGCAATGAAAACCAATGAA
 ATGCTCGCAAATGAGATAACGGAAAGAGATGAGGCAGCATTGAAGTATCTTAAGGACATC
 AGATCTTGCAGAGTTGAAGACACTTCAAGAAATTTCAAGCTGGAGTTTCTTTTATTCT
 AATCTTTACTTCAAGAACTCGGTTCTGTCTAAAACCTTACCATGTGAACGATGAAGATGGT

FIGURA 8 (continuación)

CCTGTTCTTGAGAAAGTGATTGGAACGGACATAGAATGGTTTCCAGGTAAATGTTTGACT
CATAAGGTTGTTGTGAAGAAGAAAACAAAGAAAGGGCCAAAGAAGGTCAACAACATCCCC
ATGACCAAAACAGAAAACCTGCGAGAGTTTCTTCAATTTCTTCAAGCCACCTGAGATTCT
GAGATTGATGAAGTTGACGATTACGATGATTTTGATACCATTATGACGGAAGAATAACAA
AACCTGATGGACCAAGACTATGACATTGCTGTGACAATCCGAGATAAACTGATCCCTCAT
GCAGTTTCATGGTTTACGGGAGAGGCTCTTGTTGATGAAGACGATTCTGATGATAATGAT
GATGATGATAATGATGAGAAGAGTGACTAA

SEQ ID NO: 31, proteína de tipo NAP1 de *Arabidopsis thaliana*,
secuencia de proteína deducida (registro Genbank NM_187993)

MSNEENIKSDNKSGDSSDLPTIPALDIGAECDLLAELKASHFKLLIKIHTNLTLRPF
VKKLSPKVTKRVLFLKDIQVTHDELEEKFLAEKSALEATYDNLYKPLFAKRYEIVNGVVE
AEAEKEGVPNFWLIAMKTNEMLANEITERDEAALKYLKDIRSCRVEDTSRNFKLEFLFDS
NLYFKNSVLSKTYHVNDEEDGPVLEKVIIGTDIEWFPGKCLTHKVVVKKKTKKGPVNNIP
MKTENCESFFNFFKPPEIPEIDEVDDYDDFDTIMTEELQNLMDQDYDIAVTIRDKLI
PHAVSWFTGEALVDEDDSDNDNDDDDNDEKSD

SEQ ID NO: 32, dominio NAP de la secuencia de proteína codificada por
at1g74560 (SEQ ID NO: 2), la secuencia distintiva está indicada en cursiva

IEKLQEIQDDLEKINEKASDEVLEVEQKYNVIRKPVYDKRNEVIQSI PGFWMTAFLSHPA
LGDLLTEEDQKIFKYLNSLEVEDAKDVKSGYSITFHFTSNPFFEDAKLTKTFTFLEEGTT
KITATPIKWKEGKGLPNGVNHDDKKGKRALPEES*FFTWFT*DAQHKEDAGDEIHDEVADI
IKEDLWSNPLTYFNDADE

SEQ ID NO: 33
SIEKLQEIQD

FIGURA 8 (continuación)