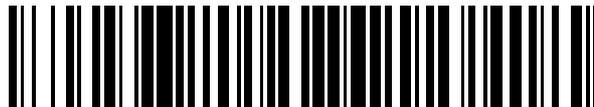


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 542 536**

51 Int. Cl.:

C07K 14/415 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.08.2010 E 10740676 (1)**

97 Fecha y número de publicación de la concesión europea: **29.04.2015 EP 2467394**

54 Título: **Plantas que tienen rasgos potenciados relacionados con el rendimiento y un procedimiento de producción de las mismas**

30 Prioridad:

19.08.2009 EP 09168166 19.08.2009 EP 09168159
19.08.2009 EP 09168213 21.08.2009 EP 09168392
21.08.2009 EP 09168433 21.08.2009 EP 09168402
14.10.2009 US 251309 14.10.2009 US 251307
14.10.2009 US 251308 14.10.2009 US 251312
14.10.2009 US 251306 14.10.2009 US 251311

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
06.08.2015

73 Titular/es:

BASF PLANT SCIENCE COMPANY GMBH
(100.0%)
67056 Ludwigshafen, DE

72 Inventor/es:

SANZ MOLINERO, ANA ISABEL;
HATZFELD, YVES;
FRANKARD, VALERIE y
REUZEAU, CHRISTOPHE

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 542 536 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Plantas que tienen rasgos potenciados relacionados con el rendimiento y un procedimiento de producción de las mismas

5 La presente divulgación se refiere, en general, al campo de la biología molecular, y se refiere a un procedimiento de potenciación de los rasgos relacionados con el rendimiento en plantas mediante la modulación de la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 (Fruto Sant/Myb). La presente divulgación también se refiere a plantas que tienen la expresión modulada de un ácido nucleico que codifica un polipéptido de tipo FSM1, plantas que tienen rasgos potenciados relacionados con el rendimiento en comparación con las plantas de tipo natural correspondientes u otras plantas de control. La invención también proporciona construcciones útiles en los procedimientos de la invención.

10 La población mundial en continuo crecimiento y el suministro cada vez menor de tierra cultivable disponible para la agricultura potencian la investigación sobre el aumento de la eficacia de la agricultura. Los medios convencionales para obtener mejoras en los cultivos y la horticultura utilizan técnicas de reproducción selectivas para identificar plantas que tengan características deseables. Sin embargo, dichas técnicas de reproducción selectivas tienen varias desventajas; en concreto, que estas técnicas normalmente emplean mucha mano de obra y producen plantas que a menudo contienen componentes genéticos heterogéneos que no siempre pueden hacer que el rasgo deseable se haga pasar desde las plantas progenitoras. Los avances en la biología molecular han permitido modificar el germoplasma de animales y plantas. La ingeniería genética de plantas supone 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. Dicha tecnología tiene la capacidad de aportar cosechas o plantas que tienen mejores rasgos económicos, agronómicos u hortícolas.

15 Un rasgo de particular interés económico es el aumento del rendimiento. El rendimiento se define normalmente como la producción medible de valor económico de una cosecha. Esto se puede definir en términos de cantidad y/o de calidad. El rendimiento depende directamente de varios factores, por ejemplo, del número y del tamaño de los órganos, de la arquitectura de la planta (por ejemplo, del número de ramas), de la producción de semillas, de la senescencia de las hojas y más. El desarrollo de las raíces, la absorción de nutrientes, la tolerancia al estrés y el vigor temprano también pueden ser factores importantes para determinar el rendimiento. Por lo tanto, la optimización de los factores anteriormente mencionados puede contribuir a aumentar el rendimiento de una cosecha.

20 El rendimiento de las semillas es un rasgo particularmente importante, ya que las semillas de muchas plantas son importantes para la nutrición de los seres humanos y de los animales. Cosechas tales como maíz, arroz, trigo, colza y soja representan más de la mitad del aporte calórico total de los seres humanos, ya sea a través del consumo directo de las propias semillas o a través del consumo de productos cárnicos generados con semillas procesadas. También son una fuente de azúcares, aceites y muchos tipos de metabolitos usados en procedimientos industriales. Las semillas contienen un embrión (la fuente de nuevos vástagos y raíces) y un endospermo (la fuente de nutrientes para el crecimiento del embrión durante la germinación y durante el crecimiento temprano de las plántulas). El desarrollo de una nueva semilla implica muchos genes, y requiere la transferencia de metabolitos desde las raíces, las hojas y los tallos hasta la semilla en crecimiento. El endospermo, en particular, asimila los precursores metabólicos de los hidratos de carbono, los aceites y las proteínas, y los sintetiza como macromoléculas de almacenamiento para engordar el grano.

30 Otro rasgo importante para muchos cultivos es el vigor temprano. La mejora del vigor temprano es un importante objetivo de los programas modernos de reproducción del arroz en variedades de cultivo de arroz tanto templados como tropicales. Las raíces largas son importantes para el anclaje apropiado al suelo del arroz sembrado en agua. Cuando el arroz se siembra directamente en campos anegados, y cuando las plantas deben emerger rápidamente a través del agua, los vástagos más largos se asocian con el vigor. Cuando se practica la siembra en línea, los mesocotilos y los coleóptilos más largos son importantes para una buena emergencia de las plántulas. La capacidad para diseñar por ingeniería genética el vigor temprano en plantas sería de gran importancia en la agricultura. Por ejemplo, un bajo vigor temprano ha sido una limitación para la introducción de híbridos de maíz (*Zea mays L.*) basados en el germoplasma del cinturón del maíz atlántico europeo.

35 Un rasgo importante adicional es el de la tolerancia mejorada al estrés abiótico. El estrés abiótico es una causa principal de pérdida de cosechas en todo el mundo, reduciendo los rendimientos medios para la mayoría de las principales plantas de cultivo en más de 50 % (Wang *et al.*, "Planta" (2003) 218: 1-14). Los estreses abióticos pueden estar provocados por la sequía, la salinidad, las temperaturas extremas, la toxicidad química y el estrés oxidativo. La capacidad para mejorar la tolerancia de las plantas al estrés abiótico sería de un gran beneficio económico para los agricultores de todo el mundo y permitiría el cultivo de cosechas en condiciones adversas y en territorios en los que el cultivo de las cosechas no se puede realizar de otro modo.

40 Por lo tanto, el rendimiento de las cosechas se puede aumentar optimizando uno de los factores anteriormente mencionados.

45 Dependiendo del uso final, se puede favorecer la modificación de ciertos rasgos de rendimiento frente a otros. Por

ejemplo, para aplicaciones tales como la producción de forraje o madera, o los recursos biocombustibles, se puede desear un aumento de las partes vegetativas de una planta, y para aplicaciones tales como la producción de harina, almidón o aceite, se puede desear particularmente el aumento de los parámetros de las semillas. Incluso entre los parámetros de las semillas, se pueden favorecer algunos frente a otros, dependiendo de la aplicación. Diversos mecanismos pueden contribuir a aumentar el rendimiento de las semillas, ya sea en forma del aumento de tamaño de las semillas o del aumento del número de semillas.

Una metodología para aumentar el rendimiento (rendimiento de semillas y/o biomasa) en plantas puede ser a través de la modificación de los mecanismos de crecimiento inherentes de una planta, tales como el ciclo celular o diversas vías de señalización implicadas en el crecimiento de las plantas o en mecanismos de defensa.

Se ha encontrado ahora que es posible mejorar diversos rasgos relacionados con el rendimiento en plantas mediante la modulación de la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 (Fruto Sant/Myb).

Antecedentes

1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

Las proteínas de tipo FSM1 son factores de transcripción que se caracterizan por la presencia de un dominio SANT/MYB. Fueron descritos por primera vez en el tomate (Barg *et al.*, "Planta" 221, 197-211, 2005); y se ha informado de que participan en la determinación de la simetría floral (Corley *et al.*, *Proc. Natl. Acad. Sci.* 102, 5068-5073, 2005). La proteína RADIALIS, por ejemplo, una proteína de tipo FSM1 en *Antirrhinum*, se expresa en la región dorsal del meristemo floral, donde interactúa con CYC y DICH para controlar la asimetría de la flor (Baxter *et al. Plant J.* 52, 105-113, 2007). FSM1 del tomate se expresa en el fruto durante las etapas de desarrollo más tempranas. La expresión ectópica de fsm1 del tomate en *Arabidopsis* produjo graves alteraciones del desarrollo que se manifestaron en un crecimiento retardado, y una dominancia apical reducida durante el desarrollo de plántulas de tomate y *Arabidopsis* (Barg *et al.*, 2005).

Sumario

1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

Sorprendentemente, ahora se ha encontrado que la modulación de la expresión de un ácido nucleico que codifica un polipéptido de tipo FSM1 proporciona plantas que tienen mejores rasgos relacionados con el rendimiento, en particular, un mayor rendimiento con respecto a plantas de control.

De acuerdo con una realización, se proporciona un procedimiento de mejora de rasgos relacionados con el rendimiento en plantas con respecto a plantas de control, que comprende modular la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.

La presente invención proporciona la materia objeto expuesta en todos y cada uno de los artículos 1 a 12 que se presentan a continuación.

1. Un procedimiento de aumento del rendimiento de la biomasa y/o de las semillas en plantas en comparación con plantas de control, que comprende la modulación de la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1, en el que dicho polipéptido de tipo FSM1 está representado por SEC ID N° 2 o uno de sus homólogos con al menos un 50 % de identidad de secuencia con SEC ID N° 2, y que comprende además un dominio SANT (acceso SMART SM0017) y uno o más de los siguientes motivos:

(i) Motivo 1:

W[TS][PA]K[QE]NK[LA]FE[RK]ALAVYD[KR][DE]TPDRW[HSQ]N[VI]A[RK]A (SEC ID N°: 283),

(ii) Motivo 2: GGK[ST][AV][ED]EV[KR]RHYE[IL]L (SEC ID N°: 284),

(iii) Motivo 3: D[VL][KF][HF][ED][SN]G[RM]VPFP[NK]Y (SEC ID N°: 285),

en el que dicho aumento del rendimiento de la biomasa y/o de las semillas se obtiene en condiciones sin estrés, en el que dicho ácido nucleico está unido operativamente al promotor GOS2 del arroz y en el que dicha expresión modulada se efectúa mediante la introducción y la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.

2. Procedimiento de acuerdo con el artículo 1, en el que dicho ácido nucleico que codifica un polipéptido de tipo FSM1 codifica una cualquiera de las proteínas enumeradas en la Tabla A1 o es una parte de dicho ácido nucleico, o un ácido nucleico capaz de hibridarse con dicho ácido nucleico.

3. Procedimiento de acuerdo con el artículo 1 o 2, en el que dicha secuencia de ácido nucleico codifica un ortólogo o parólogo de cualquiera de las proteínas dadas en la Tabla A1.

4. Procedimiento de acuerdo con uno cualquiera de los artículos 1 a 3, en el que dicho ácido nucleico que codifica un polipéptido de tipo FSM1 es de origen vegetal, preferentemente, de una planta dicotiledónea, más preferentemente, de la familia *Solanaceae*, más preferentemente del género *Lycopersicon*, y lo más preferentemente de *Solanum lycopersicum*.
5. Construcción que comprende:
- (i) ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en el artículo 1;
 - (ii) una o más secuencias de control capaces de dirigir la expresión de la secuencia de ácido nucleico de (a), siendo una de dichas secuencias de control el promotor GOS2 del arroz y, opcionalmente,
 - (iii) una secuencia de terminación de la transcripción.
6. Uso de una construcción de acuerdo con el artículo 5 en un procedimiento de producción de plantas que tienen un mayor rendimiento, particularmente, un mayor rendimiento de la biomasa y/o un mayor rendimiento de las semillas en comparación con plantas de control.
7. Planta, parte de planta o célula vegetal transformada con una construcción de acuerdo con el artículo 5.
8. Procedimiento de producción de una planta transgénica que tiene un mayor rendimiento, particularmente, un mayor rendimiento de la biomasa y/o un mayor rendimiento de las semillas en comparación con plantas de control, que comprende
- (i) introducción y expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en el artículo 1 unido operativamente al promotor GOS2 del arroz; y
 - (ii) cultivo de la célula vegetal en condiciones que potencian el crecimiento y el desarrollo vegetal.
9. Planta transgénica que tiene un mayor rendimiento, particularmente, un mayor rendimiento de la biomasa y/o de las semillas en comparación con plantas de control, que procede de la expresión modulada de un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en el artículo 1 unido operativamente al promotor GOS2 del arroz, o una célula de planta transgénica derivada de dicha planta transgénica, en la que dicha expresión modulada se realiza mediante la introducción y la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.
10. Planta transgénica de acuerdo con el artículo 7 o 9, o una célula de planta transgénica de la misma, siendo dicha planta una planta de cultivo tal como remolacha azucarera, o una planta monocotiledónea tal como caña de azúcar, o un cereal tal como arroz, maíz, trigo, cebada, mijo, centeno, triticale, sorgo, escanda, espelta, secale, escanda menor, teff, milo y avena.
11. Partes cosechables de una planta de acuerdo con el artículo 10, siendo dichas partes cosechables preferentemente biomasa de los brotes y/o de las semillas, comprendiendo dichas partes cosechables un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en el artículo 1 unido operativamente con el promotor GOS2 del arroz.
12. Uso de un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en el artículo 1 unido operativamente con el promotor GOS2 del arroz para aumentar el rendimiento de las semillas y/o de la biomasa en plantas cultivadas en condiciones sin estrés en comparación con plantas de control.

Definiciones

A lo largo de la presente memoria descriptiva, se usaran las siguientes definiciones.

Polipéptido/s /Proteína/s

- 40 Los términos "polipéptido" y "proteína" se usan indistintamente en el presente documento, y se refieren a aminoácidos en forma polimérica de cualquier longitud, ligados entre sí por enlaces peptídicos

Polinucleótido/s /ácido/s nucleico/s /secuencia/s de ácido nucleico/secuencia/s de nucleótidos

- 45 El término "polinucleótido/s", y las expresiones "secuencia/s de ácido nucleico", "secuencia/s de nucleótidos", "ácido/s nucleico/s" y "molécula de ácido nucleico" se usan indistintamente en el presente documento, y se refieren a nucleótidos, bien ribonucleótidos o desoxirribonucleótidos, o a una combinación de ambos, en una forma polimérica no ramificada de cualquier longitud.

Homólogo/s

- 50 Los "homólogos" de una proteína engloban péptidos, oligopéptidos, polipéptidos, proteínas y enzimas que tienen sustituciones, deleciones y/o inserciones de aminoácidos con relación a la proteína no modificada en cuestión y que tienen una actividad biológica y funcional similar a la proteína no modificada de la que se obtienen.

Una deleción se refiere a la eliminación de uno o más aminoácidos de una proteína.

Una inserción se refiere a la introducción de uno o más restos de aminoácido en un sitio predeterminado de una proteína. Las inserciones pueden comprender fusiones N-terminales y/o C-terminales, así como inserciones intrasecuenciales de uno solo o múltiples aminoácidos. En general, las inserciones dentro de la secuencia de aminoácidos serán más cortas que las fusiones N- o C-terminales, del orden de aproximadamente 1 a 10 restos. Los ejemplos de proteínas o péptidos de fusión N- o C-terminales incluyen el dominio de unión o el dominio de activación de un activador de la transcripción como el que se usa en el sistema de dos híbridos de levadura, proteínas de recubrimiento de fagos, marcador de (histidina)-6, marcador de glutatona S-transferasa, proteína A, proteína de unión a la maltosa, dihidrofolato reductasa, epítipo Tag-100, epítipo de c-myc, epítipo FLAG®, lacZ, CMP (péptido de unión a la calmodulina), epítipo de HA, epítipo de proteína C y epítipo de VSV.

Una sustitución se refiere al reemplazo de aminoácidos de la proteína por otros aminoácidos que tienen propiedades similares (tales como hidrofobia, hidrofilia, antigenicidad, propensión a formar o romper estructuras α -helicoidales o estructuras de lámina β similares. Por lo general, las sustituciones de aminoácidos son de restos individuales, pero pueden agruparse dependiendo de las restricciones funcionales impuestas al polipéptido, pudiendo variar de 1 a 10 aminoácidos. Habitualmente, las inserciones serán del orden de aproximadamente 1 a 10 restos de aminoácido. Preferentemente, las sustituciones de aminoácidos son sustituciones de aminoácidos conservadoras. Las tablas de sustituciones conservadoras son muy conocidas en la técnica (véanse, por ejemplo, Creighton (1984) "Proteins". W. H. Freeman and Company y la siguiente Tabla 1).

Tabla 1: Ejemplos de sustituciones de aminoácidos conservadoras

Resto	Sustitución conservadora	Resto	Sustitución conservadora
Ala	Ser	Leu	Ile; Val
Arg	Lys	Lys	Arg; Gln
Asn	Gln; His	Met	Leu; Ile
Asp	Glu	Phe	Met; Leu; Tyr
Gln	Asn	Ser	Thr; Gly
Cys	Ser	Thr	Ser; Val
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp; Phe
His	Asn; Gln	Val	Ile; Leu
Ile	Leu, Val		

Las sustituciones, deleciones y/o inserciones de aminoácidos se pueden realizar fácilmente usando técnicas de síntesis de péptidos muy conocidas en la materia, tales como la síntesis de péptidos en fase sólida y similares, o mediante manipulación de ADN recombinante. Los procedimientos de manipulación de secuencias de ADN para producir variantes de sustitución, inserción o deleción de una proteína son muy conocidos en la materia. Por ejemplo, las técnicas para realizar mutaciones de sustitución en sitios predeterminados del ADN son muy conocidas por los expertos en la materia, e incluyen mutagénesis M13, mutagénesis *in vitro* del gen T7 (USB, Cleveland, OH), mutagénesis dirigida QuickChange (Stratagene, San Diego, CA), mutagénesis dirigida mediada por PCR u otros protocolos de mutagénesis dirigida.

Derivados

Los "derivados" incluyen péptidos, oligopéptidos, polipéptidos que pueden, en comparación con la secuencia de aminoácidos de la forma natural de la proteína, tal como la proteína de interés, comprender sustituciones de aminoácidos con restos de aminoácido no naturales, o adiciones de restos de aminoácidos no naturales. Los "derivados" de una proteína también engloban péptidos, oligopéptidos, polipéptidos que comprenden restos de aminoácidos alterados naturales (glucosilados, acilados, prenilados, fosforilados, miristoilados, sulfatados etc.) o alterados no naturales en comparación con la secuencia de aminoácidos de una forma natural del polipéptido. Un derivado también puede comprender uno o más sustituyentes o adiciones que no son de aminoácido en comparación con la secuencia de aminoácidos de la que deriva, por ejemplo, una molécula indicadora u otro ligando, unido covalentemente o no covalentemente a la secuencia de aminoácidos, tal como una molécula indicadora que se une para facilitar su detección, y restos de aminoácido no naturales con respecto a la secuencia de aminoácidos de una proteína natural. Además, los "derivados" también incluyen fusiones de la forma natural de la proteína con péptidos de marcaje tales como FLAG, HIS6 o tiorredoxina (para consultar información sobre los péptidos de marcaje, véase Terpe, *Appl. Microbiol. Biotechnol.* 60, 523-533, 2003).

Ortólogo/s /parólogo/s

Los ortólogos y los parálogos engloban conceptos evolutivos usados para describir las relaciones ancestrales de los genes. Los parálogos son genes dentro de la misma especie que se han originado mediante la duplicación de un gen ancestral; y los ortólogos son genes procedentes de organismos diferentes que se han originado por especiación, y que también derivan de un gen ancestral común.

Dominio, motivo/secuencia de consenso/distintivo

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

El término "motivo" o "secuencia de consenso" o "distintivo" se refiere a una región conservada corta en la secuencia de proteínas relacionadas evolutivamente. Con frecuencia, los motivos son partes muy conservadas de dominios, pero también pueden incluir solo parte del dominio, o estar situados fuera del dominio conservado (si todos los aminoácidos del motivo están fuera de un dominio definido).

Existen bases de datos especializadas para la identificación de dominios, por ejemplo, SMART (Schultz *et al.* (1998) *Proc. Natl. Acad. Sci. EE.UU.* 95, 5857-5864; Letunic *et al.* (2002) *Nucleic Acids Res* 30, 242-244), InterPro (Mulder *et al.*, (2003) *Nucl. Acids. Res.* 31, 315-318), Prosite (Bucher y Bairoch (1994), "A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation". (In) ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. Altman R., Brutlag D., Karp P., Lathrop R., Searls D., Eds., pág. 53-61, AAAI Press, Menlo Park; Hulo *et al.*, *Nucl. Acids. Res.* 32:D134-D137, (2004)), o Pfam (Bateman *et al.*, *Nucleic Acids Research* 30(1): 276-280 (2002)). Hay un conjunto de herramientas para el análisis *in silico* de secuencias de proteínas disponible en el servidor de proteómica ExPASy (Instituto suizo de Bioinformática (Gasteiger *et al.*, ExPASy: the proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Res.* 31:3784-3788(2003)). Los dominios o los motivos también se pueden identificar usando técnicas habituales tales como mediante la alineación de secuencias.

son bien conocidos en la técnica, incluyendo dichos procedimientos GAP, BESTFIT, BLAST, FASTA y TFASTA. GAP usa el algoritmo de Needleman y Wunsch ((1970) *J. Mol Biol* 48: 443-453) para encontrar la alineación global (es decir, que engloba las secuencias completas) de dos secuencias que aumenta al máximo el número de coincidencias y reduce al mínimo el número de huecos. El algoritmo BLAST (Altschul *et al.* (1990) *J Mol Biol* 215: 403-10) calcula el porcentaje de identidad de secuencia y realiza un análisis estadístico de la similitud entre las dos secuencias. El programa informático para realizar el análisis BLAST está disponible para el público en general a través del Centro Nacional de Información Biotecnológica (NCBI). Los homólogos se pueden identificar fácilmente usando, por ejemplo, el algoritmo de alineación de secuencias múltiples ClustalW (versión 1.83), con los parámetros de alineación por pares por defecto, y un procedimiento de puntuación en porcentaje. Los porcentajes globales de similitud e identidad también se pueden determinar mediante uno de los procedimientos disponibles en el paquete informático MatGAT (Campanella *et al.*, BMC Bioinformatics. 10 de julio de 2003; 4:29. MatGAT: una aplicación que genera matrices de similitud/identidad usando proteínas o secuencias de ADN). Se puede realizar la edición manual secundaria para optimizar la alineación entre motivos conservados, como sería evidente para un experto en la materia. Además, en lugar de usar secuencias de longitud completa para la identificación de homólogos, también se pueden usar dominios específicos. Los valores de identidad de secuencia se pueden determinar en toda la secuencia de aminoácidos o ácido nucleico, o en los dominios seleccionados o motivo/s conservado/s, usando los programas mencionados anteriormente con los parámetros por defecto. Para las alineaciones locales, el algoritmo de Smith-Waterman es particularmente útil (Smith T. F., Waterman M. S. (1981) *J. Mol Biol* 147 (1); 195-7).

BLAST recíproco

Por lo general, esto implica un primer BLAST que implica procesar con BLAST una secuencia de consulta (por ejemplo, usando cualquiera de las secuencias enumeradas en la Tabla A del apartado de ejemplos) frente a cualquier base de datos de secuencias, tal como la base de datos del NCBI disponible para el público en general. BLASTN o TBLASTX (usando los valores por defecto convencionales) se usan generalmente cuando se parte de una secuencia de nucleótidos, y BLASTP o TBLASTN (usando los valores por defecto convencionales), cuando se parte de una secuencia de proteína. Opcionalmente, los resultados de BLAST se pueden filtrar. Las secuencias de longitud completa de los resultados filtrados o de los resultados no filtrados se vuelven a procesar con BLAST (segundo BLAST) frente a las secuencias del organismo del que se ha obtenido la secuencia de consulta. A continuación, se comparan los resultados del primer y segundo BLAST. El parólogo se identifica si un acierto de alto rango del primer BLAST es de la misma especie de la que deriva la secuencia de consulta, entonces lo ideal es que la repetición del BLAST genere la secuencia de consulta entre los mayores aciertos; el ortólogo se identifica si un

acierto de alto rango del primer BLAST no es de la misma especie de la que deriva la secuencia de consulta, estando preferentemente los resultados de la repetición del BLAST en la secuencia de consulta entre los mayores aciertos.

5 Los aciertos de alto rango son aquellos que tienen un valor de E bajo. Cuanto menor es el valor de E, más significativo es el marcador (o en otras palabras, menor será la probabilidad de encontrar el acierto por casualidad). El cálculo del valor de E es muy conocido en la técnica. Además de los valores de E, las comparaciones también se califican en porcentaje de identidad. El porcentaje de identidad se refiere al número de nucleótidos (o aminoácidos) idénticos entre las dos secuencias de ácido nucleico (o polipéptido) comparadas en una determinada longitud. En el caso de las familias de gran tamaño, se puede usar ClustalW, seguido de un árbol de unión de vecinos, para ayudar a visualizar el agrupamiento de los genes relacionados e identificar los ortólogos y los parálogos.

Hibridación

15 El término "hibridación", como se define en el presente documento, es un procedimiento en el que secuencias de nucleótidos complementarias sustancialmente homólogas se hibridan entre sí. El procedimiento de hibridación se puede producir totalmente en solución, es decir, ambos ácidos nucleicos complementarios están en solución. El procedimiento de hibridación también se puede producir con uno de los ácidos nucleicos complementarios inmovilizado a una matriz tal como perlas magnéticas, perlas de sefarosa y cualquier otra resina. El procedimiento de hibridación se puede producir además con uno de los ácidos nucleicos complementarios inmovilizado a un soporte sólido tal como una membrana de nitrocelulosa o de nailon, o inmovilizado, por ejemplo, mediante fotolitografía a, por ejemplo, un soporte vítreo silíceo (conocido esto último como matrices o micromatrices de ácidos nucleicos, o como microplacas de ácidos nucleicos). En general, para permitir que se produzca la hibridación, las moléculas de ácido nucleico se desnaturalizan térmica o químicamente para fundir una doble hebra en dos hebras simples y/o para retirar las horquillas u otras estructuras secundarias de ácidos nucleicos monocatenarios.

20 El término "rigurosidad" se refiere a las condiciones en las que tiene lugar una hibridación. La rigurosidad de la hibridación está influida por condiciones tales como la temperatura, la concentración de sal, la fuerza iónica y la composición del tampón de hibridación. En general, las condiciones de baja rigurosidad se seleccionan para que sean temperatura aproximadamente 30 °C inferior al punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y un pH definidos. Las condiciones de rigurosidad media son cuando la temperatura es 20 °C inferior a la T_m , y las condiciones de alta rigurosidad son cuando la temperatura es 10 °C inferior a la T_m . Normalmente, las condiciones de hibridación de alta rigurosidad se usan para aislar secuencias que se hibridan que tienen una gran similitud de secuencia con la secuencia de ácido nucleico diana. Sin embargo, los ácidos nucleicos pueden desviarse en la secuencia y aún así codificar un polipéptido sustancialmente idéntico, debido a la degeneración del código genético. Por lo tanto, a veces, se pueden requerir condiciones de hibridación de rigurosidad media para identificar dichas moléculas de ácido nucleico.

35 La T_m es la temperatura bajo una fuerza iónica y un pH definidos a la que el 50 % de la secuencia diana se hibrida a una sonda perfectamente coincidente. La T_m depende de las condiciones de la solución y la composición de las bases y la longitud de la sonda. Por ejemplo, las secuencias más largas se hibridan específicamente a temperaturas superiores. El grado máximo de hibridación se obtiene a una temperatura de aproximadamente 16 °C a 32 °C por debajo de la T_m . La presencia de cationes monovalentes en la solución de hibridación reduce la repulsión electrostática entre las dos hebras del ácido nucleico, potenciando así la formación del híbrido; este efecto es visible para concentraciones de sodio de hasta 0,4 M (para concentraciones superiores, este efecto se puede ignorar). La formamida reduce la temperatura de fusión de los dúplex de ADN-ADN y ADN-ARN con de 0,6 a 0,7 °C por cada porcentaje de formamida, y la adición de formamida al 50 % permite que la hibridación se realice a una temperatura de 30 a 45 °C, aunque el grado de hibridación se reducirá. Las faltas de coincidencia de los pares de bases reducen el grado de hibridación y la estabilidad térmica de los dúplex. De media y para las sondas de gran tamaño, la T_m disminuye aproximadamente 1 °C por % de falta de coincidencia de las bases. La T_m se puede calcular usando las siguientes ecuaciones, dependiendo de los tipos de híbridos:

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

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

2) Híbridos de ADN-ARN o ARN-ARN:

50
$$T_m = 79,8 + 18,5 (\log_{10}[\text{Na}^+]^a) + 0,58 (\% \text{G/C}^b) + 11,8 (\% \text{G/C}^b)^2 - 820/\text{L}^c$$

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

Para < 20 nucleótidos: $T_m = 2 (I_n)$

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

^a o para otro catión monovalente, pero solo exacto en el intervalo 0,01-0,4 M.

55 ^b solo exacto para % GC en el intervalo del 30 % al 75 %.

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

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

- La unión no específica se puede controlar usando una cualquiera de una serie de técnicas conocidas tales como, por ejemplo, el bloqueo de la membrana con soluciones que contienen proteínas, adiciones de ARN, ADN heterólogos y SDS al tampón de hibridación, y tratamiento con Rnasa. Para las sondas no homólogas, se puede realizar una serie de hibridaciones variando uno de (i) reducir progresivamente la temperatura de hibridación (por ejemplo de 68 °C a 42 °C) o (ii) reducir progresivamente la concentración de formamida (por ejemplo, del 50 % a 10 %). El experto en la materia conoce diversos parámetros que se pueden alterar durante la hibridación y que bien mantendrán o cambiarán las condiciones de rigurosidad.
- Además de las condiciones de hibridación, la especificidad de la hibridación normalmente también depende de la función de los lavados posteriores a la hibridación. Para eliminar el fondo resultante de la hibridación no específica, las muestras se lavan con soluciones salinas diluidas. Los factores críticos de dichos lavados incluyen la fuerza iónica y la temperatura de la solución de lavado final: cuanto menor es la concentración de sal y la temperatura de lavado, mayor será la rigurosidad del lavado. Las condiciones de lavado se realizan normalmente en o por debajo de la rigurosidad de hibridación. Una hibridación positiva da una señal que es al menos dos veces la del fondo. En general, condiciones de rigurosidad adecuadas para los ensayos de hibridación de ácidos nucleicos o procedimientos de detección de amplificación génica son como se han indicado anteriormente. También se pueden seleccionar condiciones más o menos rigurosas. El experto en la materia conoce diversos parámetros que se pueden modificar durante el lavado y que bien mantendrán o cambiarán las condiciones de rigurosidad.
- Por ejemplo, las condiciones de hibridación de alta rigurosidad típicas para los híbridos de ADN más largos de 50 nucleótidos comprenden la hibridación a 65 °C en 1 x SSC o a 42 °C en 1 x SSC y formamida al 50 %, seguido del lavado a 65 °C en 0,3 x SSC. Los ejemplos de condiciones de hibridación de rigurosidad media para los híbridos de ADN más largos de 50 nucleótidos comprenden la hibridación a 50 °C en 4 x SSC o a 40 °C en 6 x SSC y formamida al 50 %, seguido del lavado a 50 °C en 2 x SSC. La longitud del híbrido es la longitud anticipada para el ácido nucleico que se hibrida. Cuando se hibridan ácidos nucleicos de secuencia conocida, es posible determinar la longitud del híbrido alineando las secuencias e identificando las regiones conservadas descritas en el presente documento. 1 x SSC es NaCl 0,15 M y citrato sódico 15 mM; la solución de hibridación y las soluciones de lavado pueden incluir además 5 x reactivo de Denhardt, SDS al 0,5-1,0 %, 100 µg/ml de ADN de esperma de salmón desnaturalizado fragmentado, pirofosfato sódico al 0,5 %.
- Para definir el nivel de rigurosidad, cabe hacer referencia a Sambrook *et al.* (2001) "Molecular Cloning: a laboratory manual", III Edición, Cold Spring Harbor Laboratory Press, CSH, Nueva York o a "Current Protocols In Molecular Biology", John Wiley & Sons, N. Y. (1989 y actualizaciones anuales).

Variante de corte y empalme

- La expresión "variante de corte y empalme", como se usa en el presente documento, engloba variantes de una secuencia de ácido nucleico en la que intrones y/o exones seleccionados se han cortado, reemplazado, desplazado o añadido, o en la que los intrones se han acortado o alargado. Dichas variantes serán aquellas en las que la actividad biológica de la proteína se conserva sustancialmente; esto se puede realizar conservando selectivamente segmentos funcionales de la proteína. Dichas variantes de corte y empalme se pueden encontrar en la naturaleza o pueden ser artificiales. Los procedimientos para predecir y aislar dichas variantes de corte y empalme son muy conocidos en la técnica (véase, por ejemplo, Foissac y Schiex (2005), BMC Bioinformatics; 6: 25).

Variante alélica

- Los alelos o las variantes alélicas son formas alternativas de un gen dado, situadas en la misma posición cromosómica. Las variantes alélicas engloban polimorfismos de un solo nucleótido (SNP), así como polimorfismos de inserción/delección pequeña (INDEL). El tamaño de los INDEL es habitualmente inferior a 100 pb. Los SNP y los INDEL forman el mayor conjunto de variantes de secuencia en las cepas polimórficas naturales de la mayoría de los organismos.

Gen endógeno

- La referencia en el presente documento a un gen "endógeno" no solo se refiere al gen en cuestión como se encuentra en una planta en su forma natural (es decir, sin que haya intervención humana), sino que también se refiere al mismo gen (o un ácido nucleico/gen sustancialmente homólogo) en una forma aislada posteriormente (re)introducido en una planta (un transgén). Por ejemplo, una planta transgénica que contiene dicho transgén puede encontrar una reducción sustancial de la expresión del transgén y/o una reducción sustancial de la expresión del gen endógeno. El gen aislado se puede aislar de un organismo o puede ser artificial, por ejemplo, mediante síntesis química.

55

Combinación de genes/evolución dirigida

La combinación de genes o evolución dirigida consiste en iteraciones de combinación de ADN seguidas del rastreo y/o selección apropiados para generar variantes de ácidos nucleicos o partes de los mismos que codifican proteínas que tienen una actividad biológica modificada (Castle *et al.*, (2004) *Science* 304 (5674): 1151-4; patentes de EE. UU. N° 5.811.238 y 6.395.547).

Construcción

Los elementos reguladores adicionales pueden incluir potenciadores de la transcripción así como de la traducción. Los expertos en la materia conocerán las secuencias terminadoras y potenciadoras que pueden ser adecuadas para su uso en la realización de la invención. También se puede añadir una secuencia de intrón a la región no traducida 5' (UTR) o en la secuencia de codificación para aumentar la cantidad del mensaje de madurez que se acumula en el citosol, como se describe en el apartado de definiciones. Otras secuencias de control (además de promotor, potenciador, silenciador, secuencias de intrones, regiones 3'UTR y/o 5'UTR) pueden ser elementos de estabilización de proteínas y/o ARN. Dichas secuencias serían conocidas o se podrían obtener fácilmente por un experto en la materia.

Las construcciones genéticas de la invención pueden incluir además un origen de secuencia de replicación que se requiere para el mantenimiento y/o la replicación en un tipo de célula específico. Un ejemplo es cuando se requiere una construcción genética para mantenerla en una célula bacteriana como un elemento genético episomal (por ejemplo, molécula de plásmido o cósmido). Los orígenes preferidos de replicación incluyen, pero sin limitación, el f1-ori y colE1.

Para la detección de la transferencia correcta de las secuencias de ácidos nucleicos como se usa en los procedimientos de la invención y/o selección de plantas transgénicas que comprenden estos ácidos nucleicos, es ventajoso usar genes marcadores (o genes indicadores). Por lo tanto, la construcción genética puede comprender opcionalmente un gen marcador seleccionable. Los marcadores seleccionables se describen en más detalle en el apartado de "Definiciones" del presente documento. Los genes marcadores se pueden suprimir o retirar de la célula transgénica, una vez que ya no son necesarios. Las técnicas para la eliminación de marcadores son conocidas en la materia, describiéndose las técnicas útiles anteriormente en el apartado de definiciones.

Elemento regulador/secuencia de control/promotor

Las expresiones "elemento regulador" y "secuencia de control", y el término "promotor" se usan todos indistintamente en el presente documento y se han de tomar en un contexto amplio para referirse a secuencias de ácido nucleico reguladoras capaces de afectar a la expresión de las secuencias a las que están ligadas. El término "promotor" normalmente se refiere a una secuencia de control de ácido nucleico situada cadena arriba del inicio de la transcripción de un gen, y que está implicada en el reconocimiento y la unión de ARN polimerasa y otras proteínas, dirigiendo de ese modo la transcripción de un ácido nucleico conectado operativamente. Están englobadas por las expresiones y el término mencionados anteriormente las secuencias reguladoras de la transcripción derivadas de un gen genómico eucariota clásico (incluyendo la "caja" TATA, que se requiere para la iniciación exacta de la transcripción, con o sin una secuencia de "caja" CCAAT) y elementos reguladores adicionales (es decir, secuencias activadoras, estimuladores y silenciadores cadena arriba) que alteran la expresión génica en respuesta a estímulos de desarrollo y/o externos, o de una manera específica de un tejido. También se incluye dentro de la expresión una secuencia reguladora de la transcripción de un gen procarionta clásico, en cuyo caso puede incluir una secuencia de la caja -35 y/o secuencias reguladoras de la transcripción de la caja -10. La expresión "elemento regulador" también engloba una molécula de fusión sintética o derivado que confiere, activa o estimula la expresión de una molécula de ácido nucleico en una célula, un tejido o un órgano.

Un "promotor vegetal" comprende elementos reguladores, que median en la expresión de un segmento de secuencia codificante en células vegetales. Por consiguiente, un promotor vegetal no necesita ser de origen vegetal, sino que puede proceder de virus o microorganismos, por ejemplo, de virus que ataquen células vegetales. El "promotor vegetal" también puede proceder de una célula vegetal, por ejemplo, de la planta que se transforma con la secuencia de ácido nucleico que se va a expresar en el procedimiento de la invención y se describe en el presente documento. Esto también se aplica a otras señales reguladoras "vegetales" tales como terminadores "vegetales". Los promotores secuencia arriba de las secuencias de nucleótidos útiles en los procedimientos de la presente invención se pueden modificar mediante una o más sustituciones, inserciones y/o deleciones de nucleótidos sin interferir con la funcionalidad ni la actividad de ninguno de los promotores, el marco de lectura abierto (ORF) o la región reguladora 3' tales como terminadores u otras regiones reguladoras 3' que están situados más allá del ORF. Por otra parte, es posible que la actividad de los promotores se aumente mediante la modificación de su secuencia, o que se reemplacen completamente por promotores más activos, incluso por promotores procedentes de organismos heterólogos. Para la expresión en plantas, la molécula de ácido nucleico, como se ha descrito anteriormente, debe estar conectada operativamente a o comprender un promotor adecuado que exprese el gen en el punto temporal correcto y con el patrón de expresión espacial requerido.

Para la identificación de promotores funcionalmente equivalentes, se pueden analizar la fuerza del promotor y/o el patrón de expresión de un promotor candidato, por ejemplo, conectando operativamente el promotor a un gen indicador, y ensayando el nivel de expresión y el patrón del gen indicador en diversos tejidos de la planta. Los genes indicadores muy conocidos adecuados incluyen, por ejemplo, β -glucuronidasa o β -galactosidasa. La actividad del promotor se ensaya midiendo la actividad enzimática de la β -glucuronidasa o β -galactosidasa. A continuación, se pueden comparar la fuerza y/o el patrón de expresión del promotor con los de un promotor de referencia (tal como el usado en los procedimientos de la presente invención). Como alternativa, se puede ensayar la fuerza del promotor cuantificando los niveles de ARNm o comparando los niveles de ARNm del ácido nucleico usado en los procedimientos de la presente invención con los niveles de ARNm de genes constitutivos tales como ARNr de 18S, usando procedimientos conocidos en la técnica, tales como transferencia Northern con análisis densitométricos de autorradiogramas, PCR o RT-PCR cuantitativa en tiempo real (Held *et al.*, 1996 "Genome Methods" 6: 986-994). En general, la expresión "promotor débil" pretende significar un promotor que dirige la expresión de una secuencia codificante a bajo nivel. La expresión "bajo nivel" pretende significar niveles de aproximadamente 1/10.000 transcripciones a aproximadamente 1/100.000 transcripciones, hasta aproximadamente 1/500.000 transcripciones por célula. Por el contrario, un "promotor potente" dirige la expresión de una secuencia codificante a alto nivel, o de aproximadamente 1/10 transcripciones a aproximadamente 1/100 transcripciones a aproximadamente 1/1.000 transcripciones por célula. En general, la expresión "promotor medio" pretende significar un promotor que dirige la expresión de una secuencia codificante a un nivel inferior al del promotor potente, en particular, a un nivel que es, en todos los casos, inferior al obtenido cuando está bajo el control de un promotor CaMV de 35S.

20 Unido operativamente

La expresión "unido operativamente", como se usa en el presente documento, se refiere a un enlace funcional entre la secuencia promotora y el gen de interés, de modo que la secuencia promotora sea capaz de iniciar la transcripción del gen de interés.

Promotor constitutivo

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

Tabla 2a: Ejemplos de promotores constitutivos

Origen del gen	Referencia
Actina	McElroy <i>et al.</i> , <i>Plant Cell</i> , 2: 163-171, 1990
HMGP	WO 2004/070039
CAMV 35S	Odell <i>et al.</i> , <i>Nature</i> , 313: 810-812, 1985
CaMV 19S	Nilsson <i>et al.</i> , <i>Physiol. Plant.</i> 100:456-462, 1997
GOS2	de Pater <i>et al.</i> , <i>Plant J Nov</i> , 2(6):837-44, 1992, WO 2004/065596
Ubiquitina	Christensen <i>et al.</i> , <i>Plant Mol. Biol.</i> 18: 675-689, 1992
Ciclofilina del arroz	Buchholz <i>et al.</i> , <i>Plant Mol Biol.</i> 25(5): 837-43, 1994
Histona H3 del maíz	Lepetit <i>et al.</i> , <i>Mol. Gen. Genet.</i> 231:276-285, 1992
Histona H3 de la alfalfa	Wu <i>et al.</i> , <i>Plant Mol. Biol.</i> 11:641-649, 1988
Actina 2	An <i>et al.</i> , <i>Plant J.</i> 10(1); 107-121, 1996
34S FMV	Sanger <i>et al.</i> , <i>Plant. Mol. Biol.</i> , 14, 1990: 433-443
Subunidad pequeña de rubisco	US 4.962.028
OCS	Leisner (1988) <i>Proc Natl Acad Sci</i> , EE.UU. 85(5): 2553
SAD1	Jain <i>et al.</i> , <i>Crop Science</i> , 39 (6), 1999: 1696
SAD2	Jain <i>et al.</i> , <i>Crop Science</i> , 39 (6), 1999: 1696
nos	Shaw <i>et al.</i> (1984) <i>Nucleic Acids Res.</i> 12(20):7831-7846
V-ATPasa	WO 01/14572
Promotor súper	WO 95/14098
Proteínas de la caja G	WO 94/12015

Promotor ubicuo

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

Promotor regulado con respecto al desarrollo

5 Un promotor regulado con respecto al desarrollo es activo durante ciertas fases del desarrollo o en partes de la planta que sufren cambios del desarrollo.

Promotor inducible

10 Un promotor inducible tiene una iniciación de la transcripción inducida o aumentada como respuesta a un producto químico (para más información, véase Gatz 1997, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 48:89-108), un estímulo ambiental o físico, o puede ser "inducible por estrés", es decir activarse cuando una planta se expone a diversas condiciones de estrés, o "inducible por patógenos", es decir activarse cuando una planta se expone a diversos patógenos.

Promotor específico de un órgano/específico de un tejido

15 Un promotor específico de un órgano o específico de un tejido es aquel que es capaz de iniciar preferentemente la transcripción en ciertos órganos o tejidos, tales como las hojas, las raíces, el tejido seminal, etc. Por ejemplo, un "promotor específico de las raíces" es un promotor que es activo para la transcripción predominantemente en las raíces de las plantas, sustancialmente hasta la exclusión de cualquier otra parte de una planta, aunque sin embargo permite cualquier expresión filtrante en estas otras partes de la planta. Los promotores capaces de iniciar la transcripción en ciertas células se denominan en el presente documento "específicos de una célula".

Los ejemplos de promotores específicos de raíz se listan en la tabla 2b a continuación:

20

Tabla 2b: Ejemplos de promotores específicos de la raíz

Origen del gen	Referencia
RCc3	<i>Plant Mol Biol.</i> Enero de 1995;27(2):237-48
PHT1 de <i>Arabidopsis</i>	Koyama <i>et al.</i> <i>J Biosci Bioeng.</i> Enero de 2005; 99(1):38-42.; Mudge <i>et al.</i> (2002, <i>Plant J.</i> 31:341)
Transportador de fostato de <i>Medicago</i>	Xiao <i>et al.</i> , 2006, <i>Plant Biol</i> (Stuttg). Julio de 2006; 8(4):439-49
Pyk10 de <i>Arabidopsis</i>	Nitz <i>et al.</i> (2001) <i>Plant Sci</i> 161(2): 337-346
Genes expresables en la raíz	Tingey <i>et al.</i> , <i>EMBO J.</i> 6: 1, 1987.
Gen inducible por la auxina del tabaco	Van der Zaal <i>et al.</i> , <i>Plant Mol. Biol.</i> 16, 983, 1991.
β -tubulina	Oppenheimer, <i>et al.</i> , <i>Gene</i> 63: 87, 1988.
Genes específicos de la raíz del tabaco	Conkling, <i>et al.</i> , <i>Plant Physiol.</i> 93: 1203, 1990.
Gen G1-3b de <i>B. napus</i>	Patente de EE.UU. N° 5. 401.836
SbPRP1	Suzuki <i>et al.</i> , <i>Plant Mol. Biol.</i> 21: 109-119, 1993.
LRX1	Baumberger <i>et al.</i> 2001, <i>Genes & Dev.</i> 15:1128
BTG-26 de <i>Brassica napus</i>	US 20050044585
LeAMT1 (tomate)	Lauter <i>et al.</i> (1996, <i>PNAS</i> 3:8139)
El LeNRT1-1 (tomate)	Lauter <i>et al.</i> (1996, <i>PNAS</i> 3:8139)
Gen de la patatina de clase I (patata)	Liu <i>et al.</i> , <i>Plant Mol. Biol.</i> 17 (6): 1139-1154
KDC1 (<i>Daucus carota</i>)	Downey <i>et al.</i> (2000, <i>J. Biol. Chem.</i> 275:39420)
Gen TobRB7	W. Song (1997) Tesis doctoral, Universidad estatal de Carolina del Norte, Raleigh, NC EE.UU.
OsRAB5a (arroz)	Wang <i>et al.</i> 2002, <i>Plant Sci.</i> 163:273
ALF5 (<i>Arabidopsis</i>)	Diener <i>et al.</i> (2001, <i>Plant Cell</i> 13:1625)
NRT2;1Np (<i>N. plumbaginifolia</i>)	Quesada <i>et al.</i> (1997, <i>Plant Mol. Biol.</i> 34:265)

5 Un promotor específico de la semilla es activo para la transcripción predominantemente en el tejido seminal, pero el promotor puede estar activo durante el desarrollo y/o durante la germinación de las semillas. El promotor específico de la semilla puede ser específico del endospermo/aleurona/embrión. Los ejemplos de promotores específicos de la semilla (específicos del endospermo/aleurona/embrión) se muestran de la Tabla 2c a la Tabla 2f que se presentan a continuación. Otros ejemplos de promotores específicos de la semilla se dan en Qing Qu y Takaiwa (*Plant Biotechnology*. J. 2, 113-125, 2004).

Tabla 2c: Ejemplos de promotores específicos de la semilla

Origen del gen	Referencia
Genes específicos de la semilla	Simon <i>et al.</i> , <i>Plant Mol. Biol.</i> 5: 191, 1985
	Scofield <i>et al.</i> , <i>J. Biol. Chem.</i> 262: 12202, 1987
	Baszczynski <i>et al.</i> , <i>Plant Mol. Biol.</i> 14: 633, 1990
Albúmina de la nuez de Brasil	Pearson <i>et al.</i> , <i>Plant Mol. Biol.</i> 18: 235-245, 1992
legumina	Ellis <i>et al.</i> , <i>Plant Mol. Biol.</i> 10: 203-214, 1988
glutelina (arroz)	Takaiwa <i>et al.</i> , <i>Mol. Gen. Genet.</i> 208: 15-22, 1986
	Takaiwa <i>et al.</i> , <i>FEBS Letts.</i> 221: 43-47, 1987
zeína	Matzke <i>et al.</i> , <i>Plant Mol Biol</i> , 14(3):323-32 1990
napA	Stalberg <i>et al.</i> , <i>Planta</i> 199: 515-519, 1996
Glutenina-1 LMW y HMW del trigo	<i>Mol Gen Genet</i> 216:81-90, 1989; NAR 17:461-2, 1989
SPA del trigo	Albani <i>et al</i> , <i>Plant Cell</i> , 9: 171-184, 1997
α , β , γ -gliadinas del trigo	<i>EMBO J.</i> 3:1409-15, 1984
promotor ltr1 de la cebada	Diaz <i>et al.</i> (1995) <i>Mol Gen Genet</i> 248(5):592-8
B1, C, D, hordeína de la cebada	<i>Theor Appl Gen</i> 98:1253-62, 1999; <i>Plant J</i> 4:343-55, 1993; <i>Mol Gen Genet</i> 250:750-60, 1996
DOF de la cebada	Mena <i>et al.</i> , <i>The Plant Journal</i> , 116(1): 53-62, 1998
blz2	EP99106056.7
Promotor sintético	Vicente-Carbajosa <i>et al.</i> , <i>Plant J.</i> 13: 629-640, 1998.
Prolamina NRP33 del arroz	Wu <i>et al.</i> , <i>Plant Cell Physiology</i> 39(8) 885-889, 1998
α -globulina Glb-1 del arroz	Wu <i>et al.</i> , <i>Plant Cell Physiology</i> 39(8) 885-889, 1998
OSH1 del arroz	Sato <i>et al.</i> , <i>Proc. Natl. Acad. Sci. EE.UU.</i> , 93: 8117-8122, 1996
α -globulina REB/OHP-1 del arroz	Nakase <i>et al.</i> <i>Plant Mol. Biol.</i> 33: 513-522, 1997
ADP-glucosa pirofosforilasa del arroz	<i>Trans Res</i> 6:157-68, 1997
Familia de genes ESR del maíz	<i>Plant J</i> 12:235-46, 1997
α -kafirina del sorgo	DeRose <i>et al.</i> , <i>Plant Mol. Biol</i> 32:1029-35, 1996
KNOX	Postma-Haarsma <i>et al.</i> , <i>Plant Mol. Biol.</i> 39:257-71, 1999
Oleosina del arroz	Wu <i>et al.</i> , <i>J. Biochem.</i> 123:386, 1998
Oleosina del girasol	Cummins <i>et al.</i> , <i>Plant Mol. Biol.</i> 19: 873-876, 1992
PRO0117, proteína ribosomal 40S putativa del arroz	WO 2004/070039
PRO0136, alanina aminotransferasa del arroz	Sin publicar
PRO0147, inhibidor de la tripsina ITR1 (cebada)	Sin publicar

(Continuación)

Origen del gen	Referencia
PRO0151, WSI18 del arroz	WO 2004/070039
PRO0175, RAB21 del arroz	WO 2004/070039
PRO005	WO 2004/070039
PRO0095	WO 2004/070039
α -amilasa (Amy32b)	Lanahan <i>et al.</i> , <i>Plant Cell</i> 4:203-211, 1992; Skriver <i>et al.</i> , <i>Proc Natl Acad Sci EE.UU.</i> 88:7266-7270, 1991
Gen de tipo β de la catepsina	Cejudo <i>et al.</i> , <i>Plant Mol Biol</i> 20:849-856, 1992
Ltp2 de la cebada	Kalla <i>et al.</i> , <i>Plant J.</i> 6:849-60, 1994
Chi26	Leah <i>et al.</i> , <i>Plant J.</i> 4:579-89, 1994
Maize B-Peru	Selinger <i>et al.</i> , <i>Genetics</i> 149:1125-38, 1998

Tabla 2d: Ejemplos de promotores específicos del endospermo

Origen del gen	Referencia
glutelina (arroz)	Takaiwa <i>et al.</i> (1986) <i>Mol Gen Genet</i> 208:15-22; Takaiwa <i>et al.</i> (1987) <i>FEBS Letts.</i> 221:43-47
zeína	Matzke <i>et al.</i> , (1990) <i>Plant Mol Biol</i> 14(3): 323-32
glutenina-1 LMW y HMW del trigo	Colot <i>et al.</i> (1989) <i>Mol Gen Genet</i> 216:81-90, Anderson <i>et al.</i> (1989) <i>NAR</i> 17:461-2
SPA del trigo	Albani <i>et al.</i> (1997) <i>Plant Cell</i> 9:171-184
Gliadinas del trigo	Rafalski <i>et al.</i> (1984) <i>EMBO</i> 3:1409-15
Promotor ltr1 de la cebada	Diaz <i>et al.</i> (1995) <i>Mol Gen Genet</i> 248(5):592-8
B1, C, D, hordeína de la cebada	Cho <i>et al.</i> (1999) <i>Theor Appl Genet</i> 98:1253-62; Muller <i>et al.</i> (1993) <i>Plant J</i> 4:343-55; Sorenson <i>et al.</i> (1996) <i>Mol Gen Genet</i> 250:750-60
DOF de la cebada	Mena <i>et al.</i> , (1998) <i>Plant J</i> 116(1): 53-62
blz2	Onate <i>et al.</i> (1999) <i>J Biol Chem</i> 274(14):9175-82
Promotor sintético	Vicente-Carbajosa <i>et al.</i> (1998) <i>Plant J</i> 13:629-640
prolamina NRP33 del arroz	Wu <i>et al.</i> , (1998) <i>Plant Cell Physiol</i> 39(8) 885-889
globulina Glb-1 del arroz	Wu <i>et al.</i> (1998) <i>Plant Cell Physiol</i> 39(8) 885-889
globulina REB/OHP-1 del arroz	Nakase <i>et al.</i> (1997) <i>Plant Molec Biol</i> 33: 513-522
ADP-glucosa pirofosforilasa del arroz	Russell <i>et al.</i> (1997) <i>Trans Res</i> 6:157-68
Familia de genes ESR del maíz	Opsahl-Ferstad <i>et al.</i> (1997) <i>Plant J</i> 12:235-46
Kafirina del sorgo	DeRose <i>et al.</i> (1996) <i>Plant Mol Biol</i> 32:1029-35

Tabla 2e: Ejemplos de protómeros específicos del embrión:

Origen del gen	Referencia
OSH1 del arroz	Sato <i>et al.</i> , <i>Proc. Natl. Acad. Sci. EE.UU.</i> , 93: 8117-8122, 1996
KNOX	Postma-Haarsma <i>et al.</i> , <i>Plant Mol. Biol.</i> 39:257-71, 1999
PRO0151	WO 2004/070039
PRO0175	WO 2004/070039

(Continuación)

Origen del gen	Referencia
PRO005	WO 2004/070039
PRO0095	WO 2004/070039

Tabla 2f: Ejemplos de promotores específicos del aleurona:

Origen del gen	Referencia
α -amilasa (Amy32b)	Lanahan <i>et al.</i> , <i>Plant Cell</i> 4:203-211, 1992; Skriver <i>et al.</i> , <i>Proc Natl Acad Sci EE.UU.</i> 88:7266-7270, 1991
Gen de tipo β de la cathepsina	Cejudo <i>et al.</i> , <i>Plant Mol Biol</i> 20:849-856, 1992
Ltp2 de la cebada	Kalla <i>et al.</i> , <i>Plant J.</i> 6:849-60, 1994
Chi26	Leah <i>et al.</i> , <i>Plant J.</i> 4:579-89, 1994
B-Peru del maíz	Selinger <i>et al.</i> , <i>Genetics</i> 149:1125-38, 1998

5 Un promotor específico del tejido verde, como se define en el presente documento, es un promotor que es activo para la transcripción predominantemente en el tejido verde, excluyendo sustancialmente cualquier otra parte de la planta, aun permitiendo cualquier expresión filtrante en estas otras partes de la planta.

Los ejemplos de promotores específicos del tejido verde que se pueden usar para realizar los procedimientos descritos en el presente documento se muestran en la siguiente Tabla 2g.

Tabla 2g: Ejemplos de promotores específicos del tejido verde

Gen	Expresión	Referencia
Ortofosfato diquinasa del maíz	Específica de la hoja	Fukavama <i>et al.</i> , <i>Plant Physiol.</i> Noviembre de 2001;127(3):1136-46
Fosfoenolpiruvato carboxilasa del maíz	Específica de la hoja	Kausch <i>et al.</i> , <i>Plant Mol Biol.</i> Enero de 2001;45(1):1-15
Fosfoenolpiruvato carboxilasa del arroz	Específica de la hoja	Lin <i>et al.</i> , 2004 <i>DNA Seq.</i> Agosto de 2004;15(4):269-76
Subunidad pequeña de la Rubisco del arroz	Específica de la hoja	Nomura <i>et al.</i> , <i>Plant Mol Biol.</i> Septiembre de 2000;44(1):99-106
β expansina EXBP9 del arroz	Específica de los brotes	WO 2004/070039
Subunidad pequeña de la Rubisco del guandú	Específica de la hoja	Panguluri <i>et al.</i> , <i>Indian J Exp Biol.</i> Abril de 2005; 43(4):369-72
RBCS3A del guisante	Específica de la hoja	

10 Otro ejemplo de un promotor específico del tejido es un promotor específico del meristemo, que es activo para la transcripción predominantemente en el tejido meristemático, excluyendo sustancialmente cualquier otra parte de la planta, aun permitiendo cualquier expresión filtrante en estas otras partes de la planta. Los ejemplos de promotores específicos del meristemo verde que se pueden usar para realizar los procedimientos descritos en el presente documento se muestran en la siguiente tabla 2h.

Tabla 2h: Ejemplos de promotores específicos del meristemo

Origen del gen	Patrón de expresión	Referencia
OSH1 del arroz	Meristemo apical de la raíz, de la fase globular del embrión a la fase de plántula	Sato <i>et al.</i> (1996) <i>Proc. Natl. Acad. Sci. EE.UU.</i> , 93: 8117-8122
Metalotioneína del arroz	Específica del meristemo	BAD87835.1
WAK1 y WAK 2	Meristemos apicales de los brotes y de la raíz, y en hojas y sépalos en expansión	Wagner y Kohorn (2001) <i>Plant Cell</i> 13(2): 303-318

15

Terminador

El término "terminador" engloba una secuencia de control que es una secuencia de ADN al final de una unidad de transcripción que señala el procesamiento y la poliadenilación 3' de una transcripción primaria y la terminación de la transcripción. El terminador puede proceder del gen natural, de una variedad de otros genes vegetales o de ADN-T. El terminador que se va a añadir puede proceder, por ejemplo, de los genes de nopalina sintasa u octopina sintasa, o como alternativa de otro gen vegetal, o menos preferentemente de cualquier otro gen eucariota.

(Gen) marcador seleccionable/gen indicador

"Marcador seleccionable", "gen marcador seleccionable" o "gen indicador" incluye cualquier gen que confiere un fenotipo sobre una célula en la que se expresa para facilitar la identificación y/o la selección de células que se transfecan o transforman con una construcción de ácido nucleico de la invención. Estos genes marcadores permiten la identificación de una transferencia correcta de las moléculas de ácido nucleico a través de una serie de diferentes principios. Los marcadores adecuados se pueden seleccionar entre marcadores que confieren resistencia antibiótica o herbicida, que introducen un nuevo rasgo metabólico o que permiten la selección visual. Los ejemplos de genes marcadores seleccionables incluyen genes que confieren resistencia a los antibióticos (tales como nptII que fosforila la neomicina y la kanamicina, o hpt, que fosforila la higromicina, o genes que confieren resistencia a, por ejemplo, la bleomicina, estreptomycin, tetraciclina, cloranfenicol, ampicilina, gentamicina, geneticina (G418), espectinomycin o blasticidina), a herbicidas (por ejemplo, bar, que proporciona resistencia a Basta[®]; aroA o gox, que proporcionan resistencia contra el glifosato, o los genes que confieren resistencia a, por ejemplo, la imidazolinona, la fosfinotricina o la sulfonilurea), o genes que proporcionan un rasgo metabólico (tal como manA, que permite a las plantas usar manosa como única fuente de carbono o la xilosa isomerasa para la utilización de xilosa, o marcadores antinutritivos tales como la resistencia a la 2-desoxiglucosa). La expresión de genes marcadores visuales genera la formación de color (por ejemplo, β -glucuronidasa, GUS o β -galactosidasa con sus sustratos de color, por ejemplo, X-Gal), luminiscencia (tal como el sistema luciferina/luciferasa) o fluorescencia (proteína verde fluorescente, GFP, y sus derivados). Esta lista solo representa un pequeño número de posibles marcadores. El experto en la materia está familiarizado con dichos marcadores. Se prefieren marcadores diferentes, dependiendo del organismo y del procedimiento de selección.

Se sabe que, tras la integración estable o transitoria de los ácidos nucleicos en células vegetales, solo una minoría de las células absorbe el ADN foráneo y, si se desea, lo integra en su genoma, dependiendo del vector de expresión usado y de la técnica de transfección usada. Para identificar y seleccionar estos integrantes, normalmente se introduce un gen que codifica un marcador seleccionable (tal como los descritos anteriormente) en las células huésped junto con el gen de interés. Estos marcadores se pueden usar, por ejemplo, en mutantes, en los que estos genes no son funcionales, por ejemplo, mediante la eliminación por procedimientos convencionales. Por otro lado, las moléculas de ácido nucleico que codifican un marcador seleccionable se pueden introducir en una célula huésped en el mismo vector que comprende la secuencia que codifica los polipéptidos descritos en el presente documento o usados en los procedimientos de la invención, o bien en un vector separado. Las células que se han transfecado de forma estable con el ácido nucleico introducido se pueden identificar, por ejemplo, mediante la selección (por ejemplo, las células que llevan integrado el marcador seleccionable sobreviven mientras que las otras células mueren).

Dado que los genes marcadores, en particular, los genes de resistencia a los antibióticos y herbicidas, ya no son necesarios o no se desean en la célula huésped transgénica una vez que los ácidos nucleicos se han introducido correctamente, el procedimiento de acuerdo con la invención para introducir los ácidos nucleicos emplea ventajosamente técnicas que permiten la eliminación o la escisión de estos genes marcadores. Uno de dichos procedimientos es el que se conoce como cotransformación. El procedimiento de cotransformación emplea dos vectores simultáneamente para la transformación, un vector que porta el ácido nucleico según lo descrito en el presente documento y un segundo que porta el/los gen/es marcador/es. Una gran proporción de las transformaciones recibe o, en el caso de las plantas, comprende (hasta el 40 % o más de las transformaciones), ambos vectores. En el caso de la transformación con *Agrobacteria*, las transformaciones normalmente solo reciben una parte del vector, es decir, la secuencia flanqueada por el ADN-T, que generalmente representa el casete de expresión. Los genes marcadores se pueden retirar posteriormente de la planta transformada mediante la realización de cruces. En otro procedimiento, los genes marcadores integrados en un transposón se usan para la transformación junto con el ácido nucleico deseado (conocido como la tecnología Ac/Ds). Las transformaciones se pueden cruzar con una fuente de transposasa o las transformaciones se transforman con una construcción de ácido nucleico que confiere la expresión de una transposasa, de manera transitoria o estable. En algunos casos (aproximadamente el 10 %), el transposón salta del genoma de la célula huésped una vez que la transformación se ha producido correctamente, y se pierde. En una serie adicional de casos, el transposón salta a una ubicación diferente. En estos casos el gen marcador se debe eliminar mediante la realización de cruces. En microbiología, se desarrollaron técnicas que hacen posible o facilitan la detección de dichos hechos. Otro procedimiento ventajoso se basa en lo que se conoce como sistemas de recombinación; cuya ventaja es que se puede prescindir de la eliminación por cruzamiento. El sistema más conocido de este tipo es lo que se conoce como el sistema Cre/lox. Cre1 es una recombinasa que elimina las secuencias situadas entre las secuencias loxP. Si el gen marcador está integrado entre las secuencias loxP, se elimina una vez que la transformación ha tenido lugar correctamente, mediante la expresión de la recombinasa. Otros sistemas de recombinación adicionales son el sistema HIN/HIX,

FLP/FRT y REP/STB (Tribble *et al.*, *J. Biol. Chem.*, 275, 2000: 22255-22267; Velmurugan *et al.*, *J. Cell Biol.*, 149, 2000: 553-566). Es posible una integración específica del sitio en el genoma de las secuencias de ácidos nucleicos de la planta según lo descrito en el presente documento. Naturalmente, estos procedimientos también se pueden aplicar a los microorganismos tales como levaduras, hongos o bacterias.

5 Transgénico/transgén/recombinante

A los efectos de la invención, "transgénico", "transgén" o "recombinante" significa con respecto a, por ejemplo, una secuencia de ácido nucleico, un casete de expresión, una construcción génica o un vector que comprende la secuencia de ácido nucleico o un organismo transformado con las secuencias de ácido nucleico, casetes de expresión o vectores de acuerdo con la invención, todas aquellas construcciones generadas mediante procedimientos recombinantes en las que bien

- (a) las secuencias de ácido nucleico que codifican proteínas útiles en los procedimientos de la invención, o
- (b) las secuencia/s de control genético que está/n unida/s operativamente con la secuencia de ácido nucleico de acuerdo con lo descrito en el presente documento, por ejemplo, un promotor, o
- (c) a) y b)

15 no se encuentran en su entorno genético natural o se han modificado mediante procedimientos recombinantes, siendo posible que la modificación adopte la forma de, por ejemplo, una sustitución, adición, delección, inversión o inserción de uno o más restos de nucleótidos. La expresión "entorno genético natural" pretende significar el locus genómico o cromosómico natural de la planta original o la presencia en una biblioteca genómica. En el caso de una biblioteca genómica, el entorno genético natural de la secuencia de ácido nucleico preferentemente se conserva, al menos parcialmente. El entorno flanquea la secuencia de ácido nucleico al menos en un lado y tiene una longitud de secuencia de al menos 50 pb, preferentemente de al menos 500 pb, especialmente preferentemente de al menos 1.000 pb y lo más preferentemente de al menos 5.000 pb. Un casete de expresión de origen natural, por ejemplo, la combinación natural del promotor natural de las secuencias de ácido nucleico con la secuencia de ácido nucleico correspondiente que codifica un polipéptido útil en los procedimientos de la presente invención, como se ha definido anteriormente, se convierte en un casete de expresión transgénico cuando este casete de expresión se modifica mediante procedimientos sintéticos no naturales ("artificiales") tales como, por ejemplo, el tratamiento mutagénico. Los procedimientos adecuados se describen, por ejemplo, en el documento US 5.565.350 o WO 00/15815.

Por lo tanto, a los efectos de la invención, la expresión "planta transgénica" pretende significar, como se ha explicado anteriormente, que los ácidos nucleicos usados en el procedimiento de la invención no están presentes en ni proceden del genoma de dicha planta, o están presentes en el genoma de dicha planta, pero no en su locus natural en el genoma de dicha planta, siendo posible que los ácidos nucleicos se expresen de manera homóloga o heteróloga. Sin embargo, como se ha mencionado, transgénico también significa que, mientras que los ácidos nucleicos de acuerdo con la invención o usados en el procedimiento de la invención están en su posición natural en el genoma de una planta, la secuencia se ha modificado con respecto a la secuencia natural y/o que las secuencias reguladoras de las secuencias naturales se han modificado. Preferentemente, el término "transgénico" pretende significar la expresión de los ácidos nucleicos de acuerdo con la invención en un locus no natural en el genoma, es decir, que tiene lugar la expresión homóloga o preferentemente heteróloga de los ácidos nucleicos. En el presente documento, se mencionan las plantas transgénicas preferidas.

Además, cabe señalar que, en el contexto de la presente invención, la expresión "ácido nucleico aislado" o "polipéptido aislado" se pueden considerar en algunos casos sinónimos de un "ácido nucleico recombinante" o un "polipéptido recombinante", respectivamente, y se refieren a un ácido nucleico o polipéptido que no se encuentra en su entorno genético natural y/o que se ha modificado mediante procedimientos recombinantes.

Modulación

El término "modulación" significa, en relación con la expresión o la expresión génica, un procedimiento en el que el nivel de expresión es cambiado por dicha expresión génica en comparación con la planta de control, preferentemente el nivel de expresión se aumenta o se reduce. La expresión no modulada original puede ser de cualquier tipo de expresión de un ARN estructural (ARNr, ARNt) o ARNm con la posterior traducción. La expresión "modular la actividad" significará cualquier cambio de la expresión de las secuencias de ácido nucleico o las proteínas codificadas de la invención que conduce al aumento del rendimiento y/o al aumento del crecimiento de las plantas. La expresión puede aumentar de cero (ausencia de expresión o expresión no medible) a una cierta cantidad, o puede reducir de una cierta cantidad a cantidades pequeñas no medibles o cero.

Expresión

El término "expresión" o la expresión "expresión génica", como se usan en el presente documento, significan la transcripción de un gen específico o genes específicos o construcción genética específica. El término "expresión" o la expresión "expresión génica", en particular, significan la transcripción de un gen o genes o construcción genética en ARN estructural (ARNr, ARNt) o ARNm con o sin la posterior traducción de este último en una proteína. El procedimiento incluye la transcripción del ADN y el procesamiento del producto de ARNm resultante.

Expresión aumentada/sobreexpresión

La expresión "expresión aumentada" o "sobreexpresión", como se usa en el presente documento, significa cualquier forma de expresión que sea adicional al nivel de expresión del tipo natural original. A los efectos de la presente invención, el nivel de expresión de tipo natural original también podría ser cero (ausencia de expresión o expresión no medible).

Los procedimientos para aumentar la expresión de genes o productos génicos están bien documentados en la técnica e incluyen, por ejemplo, la sobreexpresión dirigida por promotores apropiados, el uso de promotores de la transcripción o promotores de la traducción. Los ácidos nucleicos aislados que sirven como elementos promotores o potenciadores se pueden introducir en una posición apropiada (normalmente, cadena arriba) de una forma no heteróloga de un polinucleótido con el fin de regular positivamente la expresión de un ácido nucleico que codifica el polipéptido de interés. Por ejemplo, los promotores endógenos se pueden alterar *in vivo* mediante mutación, delección y/o sustitución (véanse Kmiec, US 5.565.350; Zarling *et al.*, WO9322443), o se pueden introducir promotores aislados en una célula vegetal en la orientación y la distancia adecuadas desde un gen como se describe en el presente documento con el fin de controlar la expresión del gen.

Si se desea la expresión de un polipéptido, en general, es deseable incluir una región de poliadenilación en el extremo 3' de una región codificante de un polinucleótido. La región de poliadenilación se puede derivar del gen natural, de una variedad de otros genes vegetales o de ADN-T. La secuencia del extremo 3' que se va a añadir se puede derivar, por ejemplo, de los genes de nopalina sintasa u octopina sintasa, o como alternativa, de otro gen vegetal, o menos preferentemente de cualquier otro gen eucariota.

También se puede añadir una secuencia intrónica a la región no traducida 5' (UTR) o la secuencia codificante de la secuencia codificante parcial para aumentar la cantidad del mensaje maduro que se acumula en el citosol. Se ha demostrado que la inclusión de un intrón empalmable en la unidad de transcripción en construcciones de expresión tanto vegetales como animales aumenta la expresión génica a niveles tanto de ARNm como de proteína hasta 1.000 veces (Buchman y Berg (1988) *Mol. Cell Biol.* 8: 4395-4405; Callis *et al.* (1987) *Genes Dev* 1:1183-1200). Dicha potenciación intrónica de la expresión génica normalmente es mayor cuando se coloca cerca del extremo 5' de la unidad de transcripción. El uso de los intrones de maíz intrón Adh1-S 1, 2 y 6, el intrón Bronze-1 se conoce en la técnica. Para una información general, véase: "The Maize Handbook", Capítulo 116, Freeling y Walbot, Eds., Springer, N. Y. (1994).

Expresión reducida

Se considera que la referencia en el presente documento a "expresión reducida" o "reducción o eliminación sustancial" de la expresión significa una reducción de la expresión de un gen endógeno y/o los niveles de polipéptido y/o la actividad de un polipéptido con respecto a plantas de control. La reducción o eliminación sustancial se produce en orden creciente de preferencia al menos 10 %, 20 %, 30 %, 40 % o 50 %, 60 %, 70 %, 80 %, 85 %, 90 % o 95 %, 96 %, 97 %, 98 %, 99 % o más reducida en comparación con la de plantas de control.

Para la reducción o eliminación sustancial de la expresión de un gen endógeno en una planta, se requiere una longitud suficiente de nucleótidos sustancialmente contiguos de una secuencia de ácido nucleico. Para realizar un silenciamiento génico, esto puede ser tan poco como 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10 o menos nucleótidos, como alternativa, esto puede ser tanto como todo el gen (incluyendo la UTR 5' y/o 3', bien en parte o en su totalidad). El tramo de nucleótidos sustancialmente contiguos se puede derivar del ácido nucleico que codifica la proteína de interés (gen diana) o de cualquier ácido nucleico capaz de codificar un ortólogo, parálogo u homólogo de la proteína de interés. Preferentemente, el tramo de nucleótidos sustancialmente contiguos es capaz de formar enlaces de hidrógeno con el gen diana (hebra bien con sentido o antisentido), más preferentemente, el tramo de nucleótidos sustancialmente contiguos tiene, en orden de preferencia creciente, 50 %, 60 %, 70 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 100 % de identidad de secuencia con el gen diana (hebra bien con sentido o antisentido). No se requiere una secuencia de ácido nucleico que codifique un polipéptido (funcional) para los diversos procedimientos analizados en el presente documento para la reducción o la eliminación sustancial de la expresión de un gen endógeno.

Esta reducción o eliminación sustancial de la expresión se puede lograr usando herramientas y técnicas habituales. Un procedimiento preferido para la reducción o eliminación sustancial de la expresión del gen endógeno es mediante la introducción y expresión en una planta de una construcción genética en la que el ácido nucleico (en este caso, un tramo de nucleótidos sustancialmente contiguos derivados del gen interés o de cualquier ácido nucleico capaz de codificar un ortólogo, parálogo u homólogo de una cualquiera de las proteínas de interés) se clona como una repetición invertida (en parte o completamente), separada por un espaciador (ADN no codificante).

En dicho procedimiento preferido, la expresión del gen endógeno se reduce o se elimina sustancialmente a través del silenciamiento mediado por ARN usando una repetición invertida de un ácido nucleico o una parte del mismo (en este caso, un tramo de nucleótidos sustancialmente contiguos derivados del gen de interés o de cualquier ácido nucleico capaz de codificar un ortólogo, parálogo u homólogo de la proteína de interés), preferentemente capaz de formar una estructura de horquilla. La repetición invertida se clona en un vector de expresión que comprende

secuencias de control. Una secuencia de ácido nucleico de ADN no codificante (un espaciador, por ejemplo, un fragmento de la región de enlace a la matriz (MAR), un intrón, un policonector, etc.) se sitúa entre los dos ácidos nucleicos invertidos formando la repetición invertida. Tras la transcripción de la repetición invertida, se forma un ARN quimérico con una estructura autocomplementaria (parcial o completa). Esta estructura de ARN bicatenario se denomina ARN de horquilla. El ARN de horquilla es procesado por la planta en ARNip que se incorporan a un complejo silenciador inducido por ARN (RISC). El RISC escinde además las transcripciones de ARNm, reduciendo sustancialmente de ese modo el número de transcripciones de ARNm que se van a traducir en polipéptidos. Para detalles adicionales generales, véanse, por ejemplo, Grierson *et al.* (1998) WO 98/53083; Waterhouse *et al.* (1999) WO 99/53050).

El comportamiento de los procedimientos descritos en el presente documento no se basa en la introducción y la expresión en una planta de una construcción genética en la que el ácido nucleico se clona como una repetición invertida, sino que, para alcanzar los mismos efectos, se pueden usar uno cualquiera o más procedimientos de "silenciamiento génico" bien conocidos.

Uno de dichos procedimientos para la reducción de la expresión de un gen endógeno es el silenciamiento mediado por ARN de la expresión del gen (regulación negativa). El silenciamiento en este caso es desencadenado en una planta por una secuencia de ARN bicatenario (ARNbc) que es sustancialmente similar al gen endógeno diana. Este ARNbc es procesado adicionalmente por la planta en de aproximadamente 20 a aproximadamente 26 nucleótidos denominados ARN interferentes pequeños (ARNip). Los ARNip se incorporan en un complejo silenciador inducido por ARN (RISC) que escinde la transcripción de ARNm del gen diana endógeno, reduciendo sustancialmente de ese modo el número de transcripciones de ARNm que se van a traducir en un polipéptido. Preferentemente, la secuencia de ARN bicatenario corresponde a un gen diana.

Otro ejemplo de procedimiento de silenciamiento por ARN implica la introducción de secuencias de ácido nucleico o partes de las mismas (en este caso, un tramo de nucleótidos sustancialmente contiguos derivados del gen de interés o de cualquier ácido nucleico capaz de codificar un ortólogo, parálogo u homólogo de la proteína de interés) en una orientación con sentido en una planta. "Orientación con sentido" se refiere a una secuencia de ADN que es homóloga a una transcripción de ARNm de la misma. Introducida en una planta sería, por tanto, al menos una copia de la secuencia de ácido nucleico. La secuencia de ácido nucleico adicional reducirá la expresión del gen endógeno, dando lugar a un fenómeno conocido como cosupresión. La reducción de la expresión génica será más pronunciada si se introducen en la planta varias copias adicionales de una secuencia de ácido nucleico, ya que hay una correlación positiva entre los altos niveles de transcripciones y el desencadenamiento de la cosupresión.

Otro ejemplo de un procedimiento de silenciamiento por ARN implica el uso de secuencias de ácido nucleico antisentido. Una secuencia de ácido nucleico "antisentido" comprende una secuencia de nucleótidos que es complementaria a una secuencia de ácido nucleico "con sentido" que codifica una proteína, es decir, complementaria a la hebra codificante de una molécula de ADN bicatenario o complementaria a una secuencia de transcripción de ARNm. Preferentemente, la secuencia de ácido nucleico antisentido es complementaria al gen endógeno que se va a silenciar. La complementariedad puede estar situada en la "región codificante" y/o en la "región no codificante" de un gen. La expresión "región codificante" se refiere a una región de la secuencia de nucleótidos que comprende codones que se traducen en restos de aminoácido. La expresión "región no codificante" se refiere a secuencias 5' y 3' que franquean la región codificante que se transcriben, pero no se traducen en aminoácidos (también denominadas regiones no traducidas 5' y 3').

Las secuencias de ácido nucleico antisentido se pueden diseñar de acuerdo con las reglas de apareamiento de bases de Watson y Crick. La secuencia de ácido nucleico antisentido puede ser complementaria a toda la secuencia de ácido nucleico (en este caso, un tramo de nucleótidos sustancialmente contiguos derivados del gen de interés o de cualquier ácido nucleico capaz de codificar un ortólogo, parálogo u homólogo de la proteína de interés), pero también puede ser un oligonucleótido que sea antisentido a solo una parte de la secuencia de ácido nucleico (incluyendo la UTR 5' y 3' de ARNm). Por ejemplo, la secuencia oligonucleotídica antisentido puede ser complementaria a la región que rodea el sitio de inicio de la traducción de una transcripción de ARNm que codifica un polipéptido. La longitud de una secuencia oligonucleotídica antisentido adecuada se conoce en la técnica y puede comenzar desde aproximadamente 50, 45, 40, 35, 30, 25, 20, 15 o 10 nucleótidos de longitud o menos. Se puede construir una secuencia de ácido nucleico antisentido usando síntesis química y reacciones de ligación enzimáticas usando procedimientos conocidos en la técnica. Por ejemplo, una secuencia de ácido nucleico antisentido (por ejemplo, una secuencia oligonucleotídica antisentido) se puede sintetizar químicamente usando nucleótidos naturales o nucleótidos diversamente modificados diseñados para aumentar la estabilidad biológica de las moléculas o para aumentar la estabilidad física del dúplex formado entre las secuencias de ácido nucleico antisentido y con sentido, pudiéndose usar, por ejemplo, derivados de fosforotioato y nucleótidos sustituidos con acridina. Los ejemplos de nucleótidos modificados que se pueden usar para generar las secuencias de ácido nucleico antisentido son muy conocidos en la técnica. Las modificaciones de nucleótidos conocidas incluyen metilación, ciclación y "protecciones", y la sustitución de uno o más de los nucleótidos naturales con un análogo tal como inosina. Otras modificaciones de nucleótidos son muy conocidas en la técnica.

La secuencia de ácido nucleico antisentido se puede producir biológicamente usando un vector de expresión en el que una secuencia de ácido nucleico se ha subclonado en una orientación antisentido (es decir, el ARN transcrito a

partir del ácido nucleico insertado será de una orientación antisentido con respecto a un ácido nucleico diana de interés). Preferentemente, la producción de secuencias de ácido nucleico antisentido en plantas se produce por medio de una construcción de ácido nucleico integrada de manera estable que comprende un promotor, un oligonucleótido antisentido conectado operativamente y un terminador.

- 5 Las moléculas de ácido nucleico usadas para el silenciamiento en los procedimientos descritos en el presente documento (bien introducidas en una planta o generadas *in situ*) se hibridan con o se unen a transcripciones de ARNm y/o ADN genómico que codifica un polipéptido para inhibir de ese modo la expresión de la proteína, por ejemplo, inhibiendo la transcripción y/o la traducción. La hibridación puede ser mediante complementariedad de nucleótidos convencional para formar un dúplex estable o, por ejemplo, en el caso de una secuencia de ácido nucleico antisentido que se une a dúplex de ADN, a través de interacciones específicas en la hendidura principal de la doble hélice. Las secuencias de ácido nucleico antisentido se pueden introducir en una planta mediante transformación o inyección directa en una zona de tejido específica. Como alternativa, las secuencias de ácido nucleico antisentido se pueden modificar con respecto a células diana seleccionadas y, a continuación, administrarse sistémicamente. Por ejemplo, para la administración sistémica, las secuencias de ácido nucleico antisentido se pueden modificar de modo que se unan específicamente a receptores o antígenos expresados sobre una superficie celular seleccionada, por ejemplo, mediante el enlace de la secuencia de ácido nucleico antisentido a péptidos o anticuerpos que se unen a receptores o antígenos de la superficie celular. Las secuencias de ácido nucleico antisentido también se pueden administrar a las células usando los vectores descritos en el presente documento.
- 20 Como se describe en el presente documento, la secuencia de ácido nucleico antisentido es una secuencia de ácido nucleico α -anomérica. Una secuencia de ácido nucleico α -anomérica forma híbridos bicatenarios específicos con ARN complementario en el que, al contrario de las unidades b habituales, las hebras van paralelas entre sí (Gaultier *et al.* (1987) *Nucl Ac Res* 15: 6625-6641). La secuencia de ácido nucleico antisentido también puede comprender un 2'-o-metilribonucleótido (Inoue *et al.* (1987) *Nucl Ac Res* 15, 6131-6148) o un análogo de ARN-ADN quimérico (Inoue *et al.* (1987) *FEBS Lett.* 215, 327-330).

- La reducción o la eliminación sustancial de la expresión de un gen endógeno también se puede realizar usando ribozimas. Las ribozimas son moléculas de ARN catalíticas con actividad de ribonucleasa que son capaces de escindir una secuencia de ácido nucleico monocatenaria, tal como un ARNm, con la que tienen una región complementaria. Así pues, las ribozimas (por ejemplo, las ribozimas de cabeza de martillo (descritas en Haselhoff y Gerlach (1988) *Nature* 334, 585-591)) se pueden usar para escindir catalíticamente transcripciones de ARNm que codifican un polipéptido, reduciendo de ese modo sustancialmente el número de transcripciones de ARNm que se ha de traducir en un polipéptido. Se puede diseñar una ribozima que tenga especificidad hacia una secuencia de ácido nucleico (véanse, por ejemplo: la patente de EE.UU. N° 4.987.071 de Cech *et al.* y la patente de EE.UU. N° 5.116.742 de Cech *et al.*). Como alternativa, se pueden usar transcripciones de ARNm correspondientes a una secuencia de ácido nucleico para seleccionar un ARN catalítico que tenga una actividad de ribonucleasa específica de un conjunto de moléculas de ARN (Bartel y Szostak (1993) *Science* 261, 1411-1418). El uso de ribozimas para el silenciamiento génico en plantas se conoce en la técnica (por ejemplo, WO 94/00012 de Atkins *et al.* (1994); WO 95/03404 de Lenne *et al.* (1995); WO 00/00619 de Lutziger *et al.* (2000); WO 97/13865 de Prinsen *et al.* (1997) y WO 97/38116 de Scott *et al.* (1997)).

- 40 El silenciamiento génico también se puede realizar mediante mutagénesis de inserción (por ejemplo, inserción de ADN-T o inserción de un transposón) o mediante estrategias como las descritas por, entre otros, Angell y Baulcombe ((1999) *Plant J* 20 (3): 357-62), (WO 98/36083 de Amplicon VIGS) o Baulcombe (WO 99/15682).

- El silenciamiento génico también se puede producir si hay una mutación en un gen endógeno y/o una mutación en un gen aislado/ácido nucleico introducido posteriormente en una planta. La reducción o la eliminación sustancial pueden estar provocadas por un polipéptido no funcional. Por ejemplo, un polipéptido se puede unir a diversas proteínas interactivas; una o más mutaciones y/o truncamientos pueden proporcionar, por tanto, un polipéptido que siga siendo capaz de unirse a proteínas interactivas (tales como proteínas receptoras), pero que no pueda presentar su función normal (tal como ligando de señalización).

- 50 Una metodología adicional para el silenciamiento génico es la dirección de secuencias de ácido nucleico complementarias a la región reguladora del gen (por ejemplo, el promotor y/o los potenciadores) para formar estructuras de triple hélice que eviten la transcripción del gen en células diana. Véanse Helene, C., *Anticancer Drug Res.* 6, 569-84, 1991; Helene *et al.*, *Ann. N. Y. Acad. Sci.* 660, 27-36 1992; y Maher, L. J. *Bioassays* 14, 807-15, 1992.

- 55 Otros procedimientos, tales como el uso de anticuerpos dirigidos a un polipéptido endógeno para inhibir su función en la planta, o la interferencia en la vía de señalización en la que participa un polipéptido, serán muy conocidos para el experto. En particular, se puede prever la utilidad de moléculas artificiales para inhibir la función biológica de un polipéptido diana o para interferir con la vía de señalización en la que participa el polipéptido diana.

Como alternativa, se puede establecer un programa de rastreo para identificar, en una población de plantas, variantes naturales de un gen, variantes que codifican polipéptidos con actividad reducida. Dichas variantes

naturales también se pueden usar, por ejemplo, para realizar recombinación homóloga.

Se pueden usar microARN (miARN) artificiales y/o naturales para desactivar la expresión génica y/o la traducción del ARNm. Los miARN endógenos son ARN pequeños monocatenarios normalmente de 19-24 nucleótidos de longitud. Funcionan principalmente para regular la expresión génica y/o la traducción del ARNm. La mayoría de los microARN (miARN) de planta tienen una complementariedad perfecta o casi perfecta con sus secuencias diana. Sin embargo, hay dianas naturales con hasta cinco faltas de coincidencia. Se procesan a partir de ARN no codificantes más largos con estructuras replegadas características mediante RNasas bicatenarias específicas de la familia Dicer. Tras el procesamiento, se incorporan en el complejo silenciador inducido por ARN (RISC) uniéndose a su componente principal, una proteína Argonauta. Los miARN sirven como los componentes de especificidad de RISC, ya que tienen apareamiento de bases con ácidos nucleicos diana, principalmente ARNm, en el citoplasma. Los hechos reguladores posteriores incluyen la escisión y destrucción y/o la inhibición de la traducción del ARNm diana. Por lo tanto, los efectos de la sobreexpresión del miARN se reflejan a menudo en niveles de ARNm reducidos de los genes diana.

Los microARN artificiales (miARNA), que tienen normalmente 21 nucleótidos de longitud, se pueden diseñar mediante ingeniería genética de forma específica para regular negativamente la expresión génica de genes de interés individuales o múltiples. Los determinantes de la selección como diana de microARN vegetal son muy conocidos en la técnica. Se han definido parámetros empíricos para el reconocimiento de dianas, y se pueden usar para ayudar en el diseño de miARNA específicos, (Schwab et al., *Dev. Cell* 8, 517-527, 2005). También están disponibles para el público herramientas convenientes para el diseño y la generación de miARNA y sus precursores, (Schwab et al., *Plant Cell* 18, 1121-1133, 2006).

Para una realización óptima, las técnicas de silenciamiento génico usadas para reducir la expresión en una planta de un gen endógeno requieren el uso de secuencias de ácido nucleico procedentes de plantas monocotiledóneas para la transformación de plantas monocotiledóneas, y procedentes de plantas dicotiledóneas para la transformación de plantas dicotiledóneas. Preferentemente, una secuencia de ácido nucleico procedente de cualquier especie de planta dada se introduce en la misma especie. Por ejemplo, una secuencia de ácido nucleico procedente del arroz se transforma en una planta de arroz. Sin embargo, no es un requisito indispensable que la secuencia de ácido nucleico que se vaya a introducir proceda de la misma especie vegetal que la planta en la que se introduce. Basta con que haya una homología sustancial entre el gen diana endógeno y el ácido nucleico que se vaya a introducir.

Se describen anteriormente ejemplos de diversos procedimientos para la reducción o la eliminación sustancial de la expresión en una planta de un gen endógeno. Un experto en la materia sería capaz de adaptar fácilmente los procedimientos mencionados anteriormente para el silenciamiento con el fin de realizar la reducción de la expresión de un gen endógeno en una planta entera o en partes de la misma mediante el uso de un promotor apropiado, por ejemplo.

El término "introducción" o "transformación", al que se hace referencia en el presente documento, engloba la transferencia de un polinucleótido exógeno en una célula huésped, independientemente del procedimiento usado para la transferencia. El tejido vegetal capaz de una propagación clonal posterior, ya sea mediante organogénesis o embriogénesis, se puede transformar con una construcción genética de la presente invención, regenerándose la planta entera a partir del mismo. El tejido elegido en particular variará dependiendo de los sistemas de propagación clonal disponibles y más adecuados para la especie que se esté transformando en concreto. Los ejemplos de dianas tisulares incluyen discos foliares, polen, embriones, cotiledones, hipocótilos, megagametofitos, tejido calloso, tejido meristemático existente (por ejemplo, meristemo apical, yemas axilares y meristemas radiculares), y tejido meristemático inducido (por ejemplo, meristemo de cotiledones y meristemo de hipocótilos). El polinucleótido se puede introducir transitoria o establemente en una célula huésped y puede mantenerse no integrado, por ejemplo, como un plásmido. Como alternativa, puede integrarse en el genoma huésped. La célula vegetal transformada resultante se puede usar a continuación para regenerar una planta transformada de una manera conocida por los expertos en la materia.

La transferencia de genes foráneos al genoma de una planta se denomina transformación. La transformación de una especie vegetal actualmente es una técnica bastante habitual. La ventaja es que se puede usar cualquiera de varios procedimientos de transformación para introducir el gen de interés en una célula antecesora adecuada. Los procedimientos descritos para la transformación y regeneración de plantas a partir de tejidos vegetales o células vegetales se pueden usar para la transformación transitoria o estable. Los procedimientos de transformación incluyen el uso de liposomas, electroporación, productos químicos que aumentan la captación de ADN libre, inyección del ADN directamente en la planta, bombardeo con pistola de partículas, transformación usando virus o polen y microproyección. Los procedimientos se pueden seleccionar entre los procedimientos de calcio/poli-etilenglicol para protoplastos (Krens, F. A. et al., (1982) *Nature* 296, 72-74; Negrutiu I et al. (1987) *Plant Mol Biol* 8: 363-373); electroporación de protoplastos (Shillito R. D. et al. (1985) *Bio/Technol* 3, 1099-1102; microinyección en material vegetal (Crossway A et al., (1986) *Mol. Gen Genet* 202: 179-185); bombardeo de partículas recubiertas con ADN o ARN (Klein T. M. et al., (1987) *Nature* 327: 70), infección con virus (no integradores) y similares. Las plantas transgénicas, incluyendo plantas de cultivo transgénicas, se producen preferentemente a través de LA transformación mediada por *Agrobacterium*. Un procedimiento de transformación ventajoso es la transformación en planta. Con este fin, es posible, por ejemplo, permitir que las agrobacterias actúen

sobre las semillas de la planta o inocular en el meristemo de la planta agrobacterias. Ha demostrado ser particularmente conveniente de acuerdo con la invención permitir que una suspensión de agrobacterias transformadas actúe sobre la planta intacta o al menos sobre los primordios de las flores. Posteriormente, se cultiva la planta hasta que se obtienen las semillas de la planta tratada (Clough y Bent, *Plant J.* (1998) 16, 735-743). Los procedimientos para la transformación mediada por *Agrobacterium* del arroz incluyen procedimientos muy conocidos para la transformación del arroz tales como los descritos en cualquiera de los siguientes: solicitud de patente europea EP 1198985 A1, Aldemita y Hodges ("Planta" 199: 612-617, 1996); Chan *et al.* (*Plant Mol Biol* 22 (3): 491-506, 1993), Hiel *et al.* (*Plant J* 6 (2): 271-282, 1994). En el caso de la transformación del maíz, 14 (6): 745-50, 1996) o Frame *et al.* (*Plant Physiol* 129 (1): 13-22, 2002). Dichos procedimientos se describen además a modo de ejemplo en B. Jenes *et al.*, "Techniques for Gene Transfer", en: *Transgenic Plants*, Vol. 1, Engineering and Utilization, eds. S. D. Kung y R. Wu, Academic Press (1993) 128-143 y en Potrykus *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 42 (1991) 205-225). Los ácidos nucleicos o la construcción que se van a expresar se clonan preferentemente en un vector, que es adecuado para transformar *Agrobacterium tumefaciens*, por ejemplo, pBin19 (Bevan *et al.*, *Nucl. Acids Res.* 12 (1984) 8711). A continuación, se pueden usar las agrobacterias transformadas por dicho vector de manera conocida para la transformación de plantas, tales como plantas usadas como un modelo, como *Arabidopsis* (*Arabidopsis thaliana*, en el alcance de la presente invención, no se considera una planta de cultivo), o plantas de cultivo tales como, a modo de ejemplo, plantas de tabaco, por ejemplo, sumergiendo hojas machacadas u hojas cortadas en una solución de agrobacterias y, a continuación, cultivándolas en un medio adecuado. La transformación de plantas por medio de *Agrobacterium tumefaciens* se describe, por ejemplo, por Höfgen y Willmitzer en *Nucl. Acid Res.* (1988) 16, 9877 o se conoce, entre otros, a partir de F. F. White, "Vectors for Gene Transfer in Higher Plants"; en *Transgenic Plants*, Vol. 1, Engineering and Utilization, eds. S. D. Kung y R. Wu, Academic Press, 1993, pág. 15-38.

Además de la transformación de células somáticas, que luego se han de regenerar en plantas intactas, también es posible transformar las células de meristemos vegetales y, en particular, aquellas células que se desarrollan en gametos. En este caso, los gametos transformados siguen el desarrollo natural de la planta, dando lugar a plantas transgénicas. Así pues, por ejemplo, se tratan semillas de *Arabidopsis* con agrobacterias y se obtienen semillas de plantas en desarrollo de las que una cierta parte se transforma, siendo así transgénicas [Feldman, K. A. y Marks M. D. (1987). *Mol Gen Genet* 208:1-9; Feldmann K. (1992). En: C Koncz, N-H Chua y J Shell, eds, "Methods in Arabidopsis Research". Word Scientific, Singapur, pág. 274-289]. Los procedimientos alternativos se basan en la retirada repetida de las inflorescencias y la incubación de la zona de escisión en el centro de la roseta con agrobacterias transformadas, con lo que asimismo se pueden obtener semillas transformadas en un punto temporal posterior (Chang (1994). *Plant J.* 5: 551-558; Katavic (1994). *Mol Gen Genet*, 245: 363-370). Sin embargo, un procedimiento especialmente eficaz es el procedimiento de infiltración al vacío con sus modificaciones, tal como el procedimiento del "baño floral". En el caso de la infiltración al vacío de *Arabidopsis*, se tratan plantas intactas bajo presión reducida con una suspensión agrobacteriana [Bechthold, N. (1993). *C. R. Acad Sci Paris Life Sci*, 316: 1194-1199], mientras que en el caso del procedimiento del "baño floral", el tejido floral en desarrollo se incuba brevemente con una suspensión agrobacteriana tratada con tensioactivo [Clough, S. J. y Bent, A. F. (1998). *The Plant J.* 16, 735-743]. En ambos casos, se recoge una cierta proporción de semillas transgénicas, y estas semillas se pueden distinguir de las semillas no transgénicas al crecer en las condiciones selectivas descritas anteriormente. Además, la transformación estable de plastidios es ventajosa debido a que los plastidios son hereditarios maternalmente en la mayoría de las cosechas reduciendo o eliminando el riesgo de flujo de transgenes a través del polen. Por lo general, la transformación del genoma de los cloroplastos se alcanza mediante un procedimiento que se ha presentado esquemáticamente en Klaus *et al.*, 2004 [*Nature Biotechnology* 22 (2), 225-229]. En resumen, las secuencias que se van a transformar se clonan junto con un gen marcador seleccionable entre secuencias de flaqueo homólogas al genoma del cloroplasto. Estas secuencias de flaqueo homólogas dirigen la integración específica para el sitio en el plastoma. La transformación plastídica se ha descrito para muchas especies de plantas diferentes y en Bock (2001) "Transgenic plastids in basic research and plant biotechnology". *J Mol Biol.* 21 de septiembre de 2001; 312(3):425-38 o Malign, P. (2003) "Progress towards commercialization of plastid transformation technology". *Trends Biotechnol.* 21, 20-28, se ofrece una visión general. Se ha presentado recientemente un avance biotecnológico adicional en forma de transformantes plastídicos exentos de marcador, que se pueden producir mediante un gen marcador cointegrado transitorio (Klaus *et al.*, 2004, *Nature Biotechnology* 22 (2), 225-229). Las células de las plantas modificadas genéticamente se pueden regenerar mediante todos los procedimientos con los que el experto está familiarizado. Los procedimientos adecuados se pueden encontrar en las publicaciones anteriormente mencionadas por S. D. Kung y R. Wu, Potrykus o Höfgen y Willmitzer.

En general, tras la transformación, se seleccionan células vegetales o agrupaciones de células vegetales en busca de la presencia de uno o más marcadores que son codificados por genes expresables en plantas cotransferidos con el gen de interés, tras lo que el material transformado se regenera en una planta entera. Para seleccionar plantas transformadas, el material vegetal obtenido en la transformación se somete, por regla general, a condiciones selectivas de modo que las plantas transformadas se puedan distinguir de las plantas no transformadas. Por ejemplo, las semillas obtenidas de la manera descrita anteriormente se pueden plantar y, tras un período de crecimiento inicial, someterse a una selección adecuada por pulverización. Una posibilidad adicional consiste en el crecimiento de las semillas, en su caso después de la esterilización, en placas de agar usando un agente de selección adecuado para que solo las semillas transformadas puedan crecer en las plantas. Como alternativa, las plantas transformadas se exploran para detectar la presencia de un marcador seleccionable tal como los descritos

anteriormente.

Tras la transferencia de ADN y la regeneración, las plantas transformadas putativamente también se pueden evaluar, por ejemplo, usando análisis Southern, en busca de la presencia del gen de interés, el número de copias y/o la organización genómica. Como alternativa o además, se pueden monitorizar los niveles de expresión del ADN recién introducido usando análisis Northern y/o Western, siendo ambas técnicas bien conocidas por los expertos habituales en la materia.

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

15 Marcaje de activación con ADN-T

El marcaje de activación con 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 secuencia arriba o abajo de la región codificante de un gen en una configuración tal que el promotor dirija la expresión del gen diana. Por lo general, la regulación de la expresión del gen diana mediante su promotor natural se interrumpe, y el gen queda bajo el control del promotor recién introducido. El promotor está incluido normalmente en un ADN-T. Este ADN-T está insertado aleatoriamente en el genoma de la planta, por ejemplo, a través de infección por *Agrobacterium*, y conduce a una expresión modificada de los genes próximos al ADN-T insertado. Las plantas transgénicas resultantes muestran fenotipos dominantes debido a la expresión modificada de los genes cercanos al promotor introducido.

25 TILLING

El término "TILLING" es la abreviatura de "Targeted Induced Local Lesions In Genomes" (Inducción Dirigida de Lesiones Locales en el Genoma) y se refiere a una tecnología de mutagénesis útil para generar y/o identificar ácidos nucleicos que codifican proteínas con expresión y/o actividad modificada. El TILLING también permite la selección de plantas que llevan dichas variantes mutantes. Estas variantes mutantes pueden presentar expresión modificada, en fuerza o en localización o en tiempo (si las mutaciones afectan, por ejemplo, al promotor). Estas variantes mutantes pueden presentar mayor actividad que la mostrada por el gen en su forma natural. El TILLING combina mutagénesis de alta densidad con procedimientos de detección de alto rendimiento. Las etapas seguidas típicamente en el TILLING son: (a) mutagénesis EMS (Redei GP y Koncz C (1992) In *Methods in Arabidopsis Research*, Koncz C, Chua NH, Schell J, eds. Singapore, World Scientific Publishing Co, páginas 16-82; Feldmann y col., (1994) In Meyerowitz EM, Somerville CR, eds, *Arabidopsis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, páginas 137-172; Lightner J y Caspar T (1998) In J Martinez-Zapater, J Salinas, eds, *Methods on Molecular Biology*, Vol. 82. Humana Press, Totowa, NJ, páginas 91-104); (b) preparación y agrupamiento de ADN en individuos; (c) amplificación por PCR de una región de interés; (d) desnaturalización e hibridación para permitir la formación de heteroduplex; (e) DHPLC, donde la presencia de un heteroduplex en un grupo se detecta como un pico extra en el cromatograma; (f) identificación del mutante individual; y (g) secuenciación del producto PCR mutante. Los procedimientos de TILLING son bien conocidos en la técnica (McCallum y col., (2000) *Nat Biotechnol* 18: 455-457; revisado por Stemple (2004) *Nat Rev Genet* 5(2): 145-50).

Recombinación homóloga

La recombinación homóloga permite la introducción en un genoma de un ácido nucleico seleccionado en una posición definida seleccionada. La recombinación homóloga es una tecnología convencional utilizada rutinariamente en ciencias biológicas para organismos inferiores tales como levaduras o el musgo *Physcomitrella*. Se han descrito procedimientos para realizar recombinación homóloga en plantas no solo para el modelo de plantas (*Offringa* y col. (1990) *EMBO J* 9(10): 3077-84) sino también para plantas de cultivo, por ejemplo, arroz (Terada y col. (2002) *Nat Biotech* 20(10): 1030-4; Iida y Terada (2004) *Curr Opin Biotech* 15(2): 132-8), y existen estrategias que son generalmente aplicables independientemente del organismo diana (Miller y col, *Nature Biotechnol.* 25, 778-785, 2007).

Rasgos relacionados con el rendimiento

Los rasgos relacionados con el rendimiento comprenden uno o más entre momento de floración temprano, rendimiento, biomasa, rendimiento seminal, vigor temprano, índice de verdor, aumento de la tasa de crecimiento, mejores rasgos agronómicos (tales como mejor eficacia del uso del agua, eficacia del uso del nitrógeno, etc.).

Rendimiento

El término "rendimiento", en general, significa una producción medible de valor económico, por lo general, en relación con un determinado cultivo, con una zona y con un período de tiempo. Cada parte de la planta contribuye directamente al rendimiento en función de su número, tamaño y/o peso, o el rendimiento real es la producción por metro cuadrado para un cultivo y un año, que se determina dividiendo la producción total (incluyendo tanto la producción cosechada como la producción valorada) entre los metros cuadrados plantados. Las expresiones "rendimiento de una planta" y "rendimiento vegetal" se usan indistintamente en el presente documento y se entiende que se refieren a la biomasa vegetal tal como la biomasa radicular y/o de los brotes, a los órganos reproductores y/o a propágulos tales como semillas de esa planta.

Tomando como ejemplo el maíz, un aumento del rendimiento se puede manifestar como uno o más de los siguientes: aumento en el número de plantas establecidas por metro cuadrado, un aumento en el número de mazorcas por planta, un aumento en el número de filas, número de granos por hilera, peso del grano, peso de mil granos, longitud/diámetro de la mazorca, aumento en la tasa de llenado de las semillas (que es el número de semillas llenas dividido entre el número total de semillas y multiplicado por 100), entre otros. Tomando como ejemplo el arroz, un aumento del rendimiento se puede manifestar como un aumento de uno o más de los siguientes: número de plantas por metro cuadrado, número de panículas por planta, longitud de las panículas, número de espiguillas por panícula, número de flores (floretes) por panícula, aumento en la tasa de llenado de las semillas (que es el número de semillas llenas dividido entre el número total de semillas y multiplicado por 100), aumento del peso de mil granos, entre otros. En el arroz, la tolerancia a la sumersión también puede dar lugar a un mayor rendimiento.

Momento de floración temprano

Las plantas que tienen un "momento de floración temprano" como se usa en el presente documento son plantas que empiezan a florecer antes que las plantas de control. Por consiguiente, esta expresión se refiere a las plantas que muestran un inicio más precoz de la floración. El momento de floración de las plantas se puede evaluar contando el número de días ("tiempo hasta la floración") entre la siembra y la emergencia de una primera inflorescencia. El "momento de floración" de una planta se puede determinar, por ejemplo, usando el procedimiento como se describe en el documento WO 2007/093444.

Vigor temprano

El vigor temprano se refiere al crecimiento bien equilibrado sano activo, especialmente durante las fases tempranas del crecimiento de la planta, y se puede deber a la mejor adecuación de la planta debido, por ejemplo, a que las plantas se adaptan mejor a su ambiente (es decir, que usan de manera óptima los recursos energéticos, y los reparten entre los brotes y las raíces). Las plantas que tienen vigor temprano también muestran una mayor supervivencia de las plántulas y un mejor establecimiento de la cosecha, lo que a menudo se traduce en campos muy uniformes (con la cosecha creciendo de un modo uniforme, es decir, con la mayoría de las plantas alcanzando las diversas fases de desarrollo sustancialmente a la vez), y a menudo un rendimiento mejor y superior. Por lo tanto, el vigor temprano se puede determinar midiendo diversos factores, tales como el peso de mil granos, el porcentaje de germinación, el porcentaje de emergencia, el crecimiento de las plántulas, la altura de las plántulas, la longitud de las raíces, la biomasa de las raíces y los brotes, y muchos más.

Aumento de la tasa de crecimiento

El aumento de la tasa de crecimiento puede ser específico de una o más partes de una planta (incluidas las semillas), o puede ser sustancialmente en toda la planta. Las plantas que tienen una mayor tasa de crecimiento pueden tener un ciclo de vida más corto. El ciclo de vida de una planta puede significar el tiempo necesario para crecer a partir de una semilla madura seca hasta la etapa en la que la planta ha producido semillas maduras secas, similar al material de partida. Este ciclo de vida puede estar influido por factores tales como la velocidad de germinación, el vigor temprano, la tasa de crecimiento, el índice de verdor, el momento de floración y la velocidad de maduración de las semillas. El aumento de la tasa de crecimiento puede tener lugar en una o más etapas del ciclo de vida de una planta o durante sustancialmente todo el ciclo de vida de la planta. El aumento de la tasa de crecimiento durante las primeras etapas del ciclo de vida de una planta puede reflejar un mayor vigor. El aumento de la tasa de crecimiento puede alterar el ciclo de cosecha de una planta, permitiendo que las plantas sean sembradas más tarde y/o cosechadas antes de lo que sería posible (se puede obtener un efecto similar con un momento de floración más temprano). Si la tasa de crecimiento se aumenta lo suficiente, puede permitir la siembra de más semillas de la misma especie vegetal (por ejemplo, la siembra y la cosecha de plantas de arroz seguidas de la siembra y la cosecha de más plantas de arroz, todo en un período de crecimiento convencional). Del mismo modo, si la tasa de crecimiento se aumenta lo suficiente, puede permitir la siembra de más semillas de diferentes especies de plantas (por ejemplo, la siembra y la cosecha de plantas de maíz de, por ejemplo, la siembra y la cosecha opcional de soja, patata o cualquier otra planta adecuada). También pueden ser posibles las recolecciones en otros momentos del mismo rizoma en el caso de algunas plantas de cultivo. La alteración del ciclo de cosecha de una planta puede conducir a un aumento de la producción anual de biomasa por metro cuadrado (debido a un aumento del número de veces (digamos en un año) que cualquier planta en particular se puede cultivar y cosechar). Un aumento en la tasa de crecimiento también puede permitir el cultivo de plantas transgénicas en una zona geográfica

más amplia que sus homólogos de tipo silvestre, pues las limitaciones territoriales para el crecimiento de un cultivo suelen estar determinadas por las condiciones ambientales adversas, bien en el momento de la siembra (inicio de la temporada) o en el momento de la cosecha (a finales de la temporada). Dichas condiciones adversas se pueden evitar acortando el ciclo de cosecha. La tasa de crecimiento se puede determinar derivando diferentes parámetros de las curvas de crecimiento, pudiendo ser dichos parámetros: T-Mid (el tiempo tomado por las plantas para alcanzar el 50 % de su tamaño máximo) y T-90 (el tiempo tomado por las plantas para alcanzar el 90 % de su tamaño máximo), entre otros.

Resistencia al estrés

Si la planta está en condiciones sin estrés o si la planta está expuesta a diferentes tipos de estrés en comparación con las plantas de control, se produce un aumento en el rendimiento y/o la tasa de crecimiento. Por lo general, las plantas responden a la exposición al estrés mediante un crecimiento más lento. En condiciones de estrés severo, la planta puede incluso dejar de crecer por completo. Por otro lado, el estrés leve se define en el presente documento como cualquier estrés al que se exponga una planta que no haga que la planta deje de crecer por completo sin la capacidad para reanudar el crecimiento. En el sentido de la invención, el estrés leve conduce a una reducción del crecimiento de las plantas estresadas inferior al 40 %, 35 %, 30 % o 25 %, más preferentemente inferior al 20 % o 15 % en comparación con la planta de control en condiciones sin estrés. Debido a los avances en las prácticas agrícolas (riego, fertilización, tratamientos con plaguicidas), normalmente, no se encuentran estreses severos en las plantas cultivadas. Por consiguiente, el crecimiento comprometido inducido por el estrés leve es a menudo una característica no deseada en la agricultura. Los "estreses leves" son los estreses bióticos y/o abióticos (ambientales) cotidianos a los que está expuesta una planta. Los estreses abióticos se pueden deber a la sequía o al exceso de agua, al estrés anaeróbico, al estrés salino, a la toxicidad química, al estrés oxidativo y al calor, al frío o a las temperaturas de congelación.

Por lo general, los "estreses bióticos" son aquellos estreses causados por patógenos, tales como bacterias, virus, hongos, nematodos e insectos.

El "estrés abiótico" puede ser un estrés osmótico causado por un estrés hídrico, por ejemplo, debido a la sequía, estrés salino o estrés por congelación. El estrés abiótico también puede ser un estrés oxidativo o estrés por frío. La expresión "estrés por congelación" pretende significar el estrés debido a las temperaturas de congelación, es decir, temperaturas a las que las moléculas de agua disponibles se congelan y se convierten en hielo. La expresión "estrés por frío", también denominado "estrés escalofriante", pretende significar las bajas temperaturas, por ejemplo, temperaturas por debajo de 10 grados, o preferentemente por debajo de 5 °C, pero a las que las moléculas de agua no se congelan. Como se ha informado en Wang *et al.* (Planta (2003) 218: 1-14), el estrés abiótico conduce a una serie de cambios morfológicos, fisiológicos, bioquímicos y moleculares que afectan negativamente al crecimiento y a la productividad de la planta. Se sabe que la sequía, la salinidad, las temperaturas extremas y el estrés oxidativo están interconectados y pueden provocar el crecimiento y el daño celular a través de mecanismos similares. Rabbani *et al.* (*Plant Physiol* (2003) 133: 1755-1767) describen un grado de "comunicación" particularmente alto entre el estrés por sequía y el estrés de alta salinidad. Por ejemplo, la sequía y/o la salinización se manifiestan principalmente como estrés osmótico, lo que produce la interrupción de la homeostasis y la distribución de iones en la célula. El estrés oxidativo, que con frecuencia acompaña al estrés por alta o baja temperatura, por salinidad o por sequía, puede provocar la desnaturalización de las proteínas funcionales y estructurales. Por consiguiente, estos diversos estreses ambientales a menudo activan vías de señalización celulares y respuestas celulares similares, tales como la producción de proteínas de estrés, la regulación positiva de los antioxidantes, la acumulación de solutos compatibles y detención del crecimiento. La expresión "condiciones sin estrés", como se usa en el presente documento, son las condiciones ambientales que permitan un crecimiento óptimo de las plantas. Los expertos en la materia conocen las condiciones del suelo y las condiciones climáticas normales para un determinado lugar. Las plantas con condiciones óptimas de crecimiento, (cultivadas en condiciones sin estrés) normalmente producen por orden creciente de preferencia al menos el 97 %, 95 %, 92 %, 90 %, 87 %, 85 %, 83 %, 80 %, 77 % o 75 % de la producción media de dicha planta en un entorno dado. La producción media se puede calcular en base a la cosecha y/o a la estación. Los expertos en la materia conocen las producciones medias de un cultivo.

En particular, los procedimientos de la presente invención se pueden realizar en condiciones sin estrés. En un ejemplo, los procedimientos de la presente invención se pueden realizar en condiciones sin estrés tales como la sequía leve, dando plantas que tienen un mayor rendimiento con respecto a las plantas de control.

En otra realización, los procedimientos según lo descrito en el presente documento se pueden realizar en condiciones de estrés.

En un ejemplo, los procedimientos descritos en el presente documento se pueden realizar en condiciones de estrés tales como sequía, dando plantas que tienen un mayor rendimiento con relación a las plantas de control. En otro ejemplo, los procedimientos descritos en el presente documento se pueden realizar en condiciones de estrés tales como déficit de nutrientes para dar plantas que tienen mayor rendimiento con respecto a las plantas de control.

El déficit de nutrientes puede provocar una falta de nutrientes tales como nitrógeno, fosfatos y otros compuestos que contienen fósforo, potasio, calcio, magnesio, manganeso, hierro y boro, entre otros.

En otro ejemplo más, los procedimientos descritos en el presente documento se pueden realizar en condiciones de estrés tales como el estrés salino para dar plantas que tienen mayor rendimiento con respecto a las plantas de control. La expresión "estrés salino" no se limita a la sal común (NaCl), sino que puede ser una cualquiera o más de entre: NaCl, KCl, LiCl, MgCl₂, CaCl₂, entre otros.

- 5 En otro ejemplo más, los procedimientos descritos en el presente documento se pueden realizar en condiciones de estrés tales como el estrés por frío o el estrés por congelación para dar plantas que tienen mayor rendimiento con respecto a las plantas de control.

Aumento/mejora/potenciación

- 10 Los términos "aumento", "mejora" o "potenciación" son intercambiables y significarán en el sentido de la solicitud al menos un 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 % o 10 %, preferentemente al menos un 15 % o 20 %, más preferentemente 25 %, 30 %, 35 % o 40 % más de rendimiento y/o crecimiento en comparación con las plantas de control según lo definido en el presente documento.

Rendimiento seminal

El aumento del rendimiento de las semillas se puede manifestar como uno o más de los siguientes:

- 15 (a) un aumento en la biomasa seminal (peso total de las semillas) que puede ser a nivel de cada semilla y/o por planta y/o por metro cuadrado;
 (b) aumento del número de flores por planta;
 (c) aumento del número de semillas y/o aumento del número de semillas llenas;
 20 (d) aumento de la tasa de llenado de las semillas (que se expresa como la proporción entre el número de semillas llenas dividido entre el número total de semillas);
 (e) aumento del índice de recolección, que se expresa como la proporción del rendimiento de las partes recolectables, tales como las semillas, dividido entre la biomasa total; y
 (f) aumento del peso de mil granos (TKW), que se extrapola del número de semillas llenas contadas y su peso total. Un aumento del TKW se puede deber a un aumento del tamaño de las semillas y/o del peso de las
 25 semillas, y también se puede deber al aumento del tamaño de los embriones y/o los endospermos.

Un aumento del rendimiento seminal también se puede manifestar como un aumento del tamaño de las semillas y/o del volumen de las semillas. Por otra parte, un aumento del rendimiento de las semillas también se puede manifestar como un aumento de la superficie de las semillas y/o de la longitud de las semillas y/o de la anchura de las semillas y/o del perímetro de las semillas.

30 Índice de verdor

- El "índice de verdor" como se usa en el presente documento se calcula a partir de imágenes digitales de plantas. Para cada píxel que pertenece al objeto de la planta sobre la imagen, se calcula la proporción del valor verde frente al valor rojo (en el modelo RGB por el que se codifica el color). El índice de verdor se expresa como el porcentaje de píxeles por el que la proporción con verde con respecto a rojo supera un umbral determinado. En condiciones de
 35 crecimiento normales, en condiciones de crecimiento de estrés salino y en condiciones de crecimiento de disponibilidad de nutrientes reducido, el índice de verdor de las plantas se mide en la última imagen antes de la floración. Por otro lado, en condiciones de crecimiento de estrés por sequía, el índice de verdor de las plantas se mide en la primera imagen después de la sequía.

Biomasa

- 40 El término "biomasa", como se usa en el presente documento, pretende referirse al peso total de una planta. Dentro de la definición de biomasa, se puede hacer una distinción entre la biomasa de una o más partes de una planta, que puede incluir:
- partes (cosechables) aéreas tales como, pero sin limitación, biomasa de los brotes, biomasa seminal, biomasa folicular, etc., y/o
 - 45 - partes (cosechables) por debajo del suelo tales como, pero sin limitación, la biomasa radicular, etc., y/o
 - biomasa vegetativa tal como la biomasa radicular, la biomasa de los brotes, etc., y/o
 - órganos reproductores y/o
 - propágulos tales como las semillas.

Reproducción asistida por marcadores

- 50 Dichos programas de reproducción algunas veces requieren la introducción de variación alélica mediante tratamiento mutagénico de las plantas, usando por ejemplo mutagénesis EMS; como alternativa, el programa puede iniciar con una colección de variantes alélicas de lo que se denomina origen "natural" causado involuntariamente. La identificación de variantes alélicas tiene lugar entonces, por ejemplo, mediante PCR. A esto, le sigue una etapa de selección de variantes alélicas superiores de la secuencia en cuestión y que dan un mayor rendimiento. Por lo

general, la selección se lleva a cabo mediante el control del rendimiento del crecimiento de las plantas que contienen diferentes variantes alélicas de la secuencia en cuestión. El rendimiento de crecimiento se puede monitorizar en un invernadero o en el campo. Otras etapas opcionales incluyen el cruce de las plantas en las que la variante alélica superior se identificó con otra planta. Esto se podría usar, por ejemplo, para realizar una combinación de características fenotípicas interesantes.

Uso como sondas en (cartografía de genes)

El uso de ácidos nucleicos que codifican la proteína de interés para la cartografía genética y física de los genes solo requiere una secuencia de ácido nucleico de al menos 15 nucleótidos de longitud. Estos ácidos nucleicos se pueden usar como marcadores del polimorfismo de longitud de fragmentos de restricción (RFLP). Las transferencias Southern (Sambrook J., Fritsch E. F. y Maniatis T. (1989) "Molecular Cloning, A Laboratory Manual") de ADN genómico de la planta digerido por restricción se pueden sondear con los ácidos nucleicos que codifican la proteína de interés. A continuación, se pueden someter los patrones de bandas resultantes a análisis 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 se pueden usar para sondear transferencias Southern que contienen ADN genómicos tratados con endonucleasas de restricción de un conjunto de individuos que representan a los padres y la progenie de un cruce genético definido. Se observa la segregación de los polimorfismos de ADN y se usa para calcular la posición del ácido nucleico que codifica la proteína de interés en el mapa genético previamente obtenido usando esta población (Botstein *et al.* (1980) *Am. J. Hum. Genet.* 32: 314-331).

La producción y el uso de sondas derivadas de genes vegetales para su uso en la cartografía genética se describe en Bernatzky y Tanksley (1986) *Plant Mol. Biol. Reporter* 4: 37-41. Numerosas publicaciones describen la cartografía genética de clones de ADNc específicos usando la metodología descrita anteriormente o variaciones de la misma. Para la cartografía, se pueden usar, por ejemplo, poblaciones de entrecruzamiento F2, poblaciones de retrocruzamiento, poblaciones apareadas al azar, líneas isogénicas cercanas y otros grupos de individuos. Dichas metodologías son bien conocidas por los expertos en la materia.

Para la cartografía física, también se pueden usar las sondas de ácido nucleico (es decir, la colocación de secuencias en mapas físicos; véase Hoheisel *et al* en: "Non-mammalian Genomic Analysis: A Practical Guide", Academic press 1996, págs. 319-346, y las referencias citadas en la misma).

En otra realización, las sondas de ácido nucleico se pueden usar en la cartografía por fluorescencia directa de hibridación *in situ* (FISH) (Trask (1991) *Trends Genet.* 7: 149-154). Aunque los procedimientos actuales de cartografía FISH favorecen el uso de grandes clones (varios kb 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 la cartografía FISH usando sondas más cortas.

Se puede llevar a cabo una variedad de procedimientos basados en la amplificación de ácido nucleico para la cartografía genética y física usando los ácidos nucleicos. Los ejemplos incluyen la amplificación específica del alelo (Kazazian (1989) *J. Lab. Clin. Med* 11:95-96), el polimorfismo de fragmentos amplificados por PCR (CAPS; Sheffield *et al* (1993) *Genomics* 16: 325-332), ligación específica del alelo (Landegren *et al* (1988) *Science* 241:1077-1080), reacciones de extensión de nucleótidos (Sokolov (1990) *Nucleic Acid Res.* 18:3671), cartografía híbrida por radiación (Walter *et al* (1997) *Nat. Genet.* 7:22-28) y la cartografía feliz (Dear y Cook (1989) *Nucleic Acid Res.* 17: 6795-6807). Para estos procedimientos, 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 del cebador. El diseño de dichos cebadores es bien conocido para los expertos en la materia. En los procedimientos que emplean la cartografía genética basada en PCR, puede ser necesario identificar diferencias en la secuencia de ADN entre los padres del cruce cartográfico en la región correspondiente a la secuencia de ácido nucleico instantánea. Esto, sin embargo, generalmente no es necesario para los procedimientos de cartografía.

Planta

El término "planta", como se usa en el presente documento, engloba plantas enteras, antecesores y progenie de las plantas y partes de plantas, incluyendo semillas, brotes, tallos, hojas, raíces (incluyendo tubérculos), flores y tejidos y órganos, comprendiendo cada uno de los mencionados anteriormente el gen/ácido nucleico de interés. El término "planta" también engloba células vegetales, cultivos en suspensión, tejido calloso, embriones, regiones meristemáticas, gametofitos, esporofitos, polen y microesporas, comprendiendo, de nuevo, cada uno de los mencionados anteriormente el gen/ácido nucleico de interés.

Las plantas que son particularmente útiles en los procedimientos de la invención incluyen todas las plantas que pertenecen a la superfamilia *Viridiplantae*, en particular, plantas monocotiledóneas y dicotiledóneas incluyendo legumbres para pasto o forraje, plantas ornamentales, cosechas alimenticias, árboles o arbustos seleccionados de la lista que comprende *Acer* sp., *Actinidia* sp., *Abelmoschus* sp., *Agave sisalana*, *Agropyron* sp., *Agrostis stolonifera*, *Allium* sp., *Amaranthus* sp., *Ammophila arenaria*, *Ananas comosus*, *Annona* sp., *Apium graveolens*, *Arachis* sp., *Artocarpus* sp., *Asparagus officinalis*, *Avena* sp. (por ejemplo, *Avena sativa*, *Avena fatua*, *Avena byzantina*, *Avena fatua* var. *sativa*, *Avena híbrida*), *Averrhoa carambola*, *Bambusa* sp., *Benincasa hispida*, *Bertholletia excelsea*, *Beta*

vulgaris, *Brassica* sp. (por ejemplo, *Brassica napus*, *Brassica rapa* ssp. [colza, colza de semilla oleaginosa, nabina]), *Cadaba farinosa*, *Camellia sinensis*, *Canna indica*, *Cannabis sativa*, *Capsicum* sp., *Carex elata*, *Carica papaya*, *Carissa macrocarpa*, *Carya* sp., *Carthamus tinctorius*, *Castanea* sp., *Ceiba pentandra*, *Cichorium endivia*, *Cinnamomum* sp., *Citrullus lanatus*, *Citrus* sp., *Cocos* sp., *Coffea* sp., *Colocasia esculenta*, *Cola* sp., *Corchorus* sp., *Coriandrum sativum*, *Corylus* sp., *Crataegus* sp., *Crocus sativus*, *Cucurbitas* sp., *Cucumis* sp., *Cynara* sp., *Daucus carota*, *Desmodium* sp., *Dimocarpus longan*, *Dioscorea* sp., *Diospyros* sp., *Echinochloa* sp., *Elaeis* (por ejemplo, *Elaeis guineensis*, *Elaeis oleifera*), *Eleusine coracana*, *Eragrostis tef*, *Erianthus* sp., *Eriobotrya japonica*, *Eucalyptus* sp., *Eugenia uniflora*, *Fagopyrum* sp., *Fagus* sp., *Festuca arundinacea*, *Ficus carica*, *Fortunella* sp., *Fragaria* sp., *Ginkgo biloba*, *Glycine* sp. (por ejemplo, *Glycine max*, *Soja hispida* or *Soja max*), *Gossypium hirsutum*, *Helianthus* sp. (por ejemplo, *Helianthus annuus*), *Hemerocallis fulva*, *Hibiscus* sp., *Hordeum* sp. (por ejemplo, *Hordeum vulgare*), *Ipomoea batatas*, *Juglans* sp., *Lactuca sativa*, *Lathyrus* sp., *Lens culinaris*, *Linum usitatissimum*, *Litchi chinensis*, *Lotus* sp., *Luffa acutangula*, *Lupinus* sp., *Luzula sylvatica*, *Lycopersicon* sp. (por ejemplo, *Lycopersicon esculentum*, *Lycopersicon lycopersicum*, *Lycopersicon pyriforme*), *Macrotyloma* sp., *Malus* sp., *Malpighia emarginata*, *Mammea americana*, *Mangifera indica*, *Manihot* sp., *Manilkara zapota*, *Medicago sativa*, *Melilotus Nicotiana* sp., *Olea* sp., *Opuntia* sp., *Ornithopus* sp., *Oryza* sp. (por ejemplo, *Oryza sativa*, *Oryza latifolia*), *Panicum miliaceum*, *Panicum virgatum*, *Passiflora edulis*, *Pastinaca sativa*, *Pennisetum* sp., *Persea* sp., *Petroselinum crispum*, *Phalaris arundinacea*, *Phaseolus* sp., *Phleum pratense*, *Phoenix* sp., *Phragmites australis*, *Physalis* sp., *Pinus* sp., *Pistacia vera*, *Pisum* sp., *Poa* sp., *Populus* sp., *Prosopis* sp., *Prunus* sp., *Psidium* sp., *Punica granatum*, *Pyrus communis*, *Quercus* sp., *Raphanus sativus*, *Rheum rhabarbarum*, *Ribes* sp., *Ricinus communis*, *Rubus* sp., *Saccharum* sp., *Salix* sp., *Sambucus* sp., *Secale cereale*, *Sesamum* sp., *Sinapis* sp., *Solanum* sp. (por ejemplo, *Solanum tuberosum*, *Solanum integrifolium* o *Solanum lycopersicum*), *Sorghum bicolor*, *Spinacia* sp., *Syzygium* sp., *Tagetes* sp., *Tamarindus indica*, *Theobroma cacao*, *Trifolium* sp., *Tripsacum dactyloides*, *Triticosecalerimpai*, *Triticum* sp. (por ejemplo, *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum*, *Triticum monococcum* or *Triticum vulgare*), *Tropaeolum minus*, *Tropaeolum majus*, *Vaccinium* sp., *Vicia* sp., *Vigna* sp., *Viola odorata*, *Vitis* sp., *Zea mays*, *Zizania palustris*, *Ziziphus* sp., entre otras.

Planta/s de control

La elección de las plantas de control adecuadas es una parte rutinaria de un entorno experimental, y puede incluir las correspondientes plantas de tipo silvestre o las correspondientes plantas sin el gen de interés. Por lo general, la planta de control es de la misma especie vegetal o incluso de la misma variedad que la planta que se va a evaluar. La planta de control también puede ser un nulicigoto de la planta que se va a evaluar. Los nulicigotos son individuos que pierden el transgén por segregación. Una "planta de control", como se usa en el presente documento, no solo se refiere a las plantas enteras, sino también a partes de la planta, incluyendo semillas y partes de semillas.

Descripción detallada de la invención

Sorprendentemente, ahora se ha encontrado que la modulación de la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 da plantas que tienen mejores rasgos relacionados con el rendimiento en comparación con las plantas de control. La presente divulgación proporciona un procedimiento de mejora de los rasgos relacionados con el rendimiento en plantas en comparación con plantas de control, que comprende modular la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 y, opcionalmente, seleccionar las plantas que tengan mejores rasgos relacionados con el rendimiento.

En el presente documento, también se describen los hasta ahora desconocidos ácidos nucleicos codificantes de tipo FSM1 y polipéptidos de tipo FSM1 útiles para conferir mejores rasgos relacionados con el rendimiento en plantas en comparación con las plantas de control.

Por lo tanto, la descripción proporciona una molécula de ácido nucleico aislada seleccionada entre:

- (i) un ácido nucleico representado por cualquiera de SEC ID N°: 17, SEC ID N°: 239, SEC ID N°: 259, SEC ID N°: 261, SEC ID N°: 265 o SEC ID N°: 277;
- (ii) el complemento de un ácido nucleico representado por cualquiera de SEC ID N°: 17, SEC ID N°: 277;
- (iii) un ácido nucleico que codifica un polipéptido de tipo FSM1, que comprende un dominio MYB/SANT y que tiene, por orden creciente de preferencia, una identidad de secuencia de al menos 50 %, 51 %, 52 %, 53 %, 54 %, 55 %, 56 %, 57 %, 58 %, 59 %, 60 %, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % con la secuencia de aminoácidos representada por una cualquiera de SEC ID N°: 18, SEC ID N°: 240, SEC ID N°: 260, SEC ID N°: 262, SEC ID N°: 266 o SEC ID N°: 278, y además o como alternativa, que comprende uno o más motivos que tienen, por orden creciente de preferencia, 4 o menos de 4, menos de 3, menos de 2, o ninguna sustitución (faltas de coincidencia de secuencia) a uno cualquiera o más de los motivos dados en SEC ID N°: 283 a SEC ID NO: 288, y todavía más preferentemente que confieren mejores rasgos relacionados con el rendimiento en comparación con las plantas de control.
- (iv) una molécula de ácido nucleico que se hibrida con una molécula de ácido nucleico de (i) a (iii) en condiciones de hibridación de alta rigurosidad y que preferentemente confiere mejores rasgos relacionados con el rendimiento en comparación con las plantas de control.

De acuerdo con una realización adicional, la descripción también proporciona un polipéptido aislado seleccionado entre:

- (i) una secuencia de aminoácidos representada por SEC ID N°: 18, SEC ID N°: 240, SEC ID N°: 260, SEC ID N°: 262, SEC ID N°: 266 o SEC ID N°: 278;
- 5 (ii) una secuencia de aminoácidos, que comprende un dominio MYB/SANT y que tiene, por orden creciente de preferencia, una identidad de secuencia de al menos 50 %, 51 %, 52 %, 53 %, 54 %, 55 %, 56 %, 57 %, 58 %, 59 %, 60 %, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % con la secuencia de aminoácidos representada por SEC ID N°: 18, SEC ID N°: 240, SEC ID N°: 260, SEC ID N°: 262, SEC ID N°: 266 o SEC ID N°: 278, y además o como alternativa, que comprende uno o más motivos que tienen, por orden creciente de preferencia, 4 o menos de 4, menos de 3, menos de 2, o ninguna sustitución (faltas de coincidencia de secuencia) a cualquiera de los motivos dados en SEC ID N°: 283 a SEC ID NO: 288, y todavía más preferentemente que confieren mejores rasgos relacionados con el rendimiento en comparación con las plantas de control;
- 10 (iii) derivados de cualquiera de las secuencias de aminoácidos dadas en (i) o (ii) anterior.
- 15

Un procedimiento preferido de modulación (preferentemente, aumento) de la expresión de un ácido nucleico que codifica un polipéptido de tipo FSM1 (Fruto Sant/Myb) es mediante la introducción y la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 (Fruto Sant/Myb).

20 En una realización, la expresión "proteína útil en los procedimientos de la invención" pretende significar un polipéptido de tipo FSM1 como se define en el presente documento. En otra realización, la expresión "ácido nucleico útil en los procedimientos de la invención" pretende significar un ácido nucleico capaz de codificar dicho polipéptido de tipo FSM1. El ácido nucleico que se va a introducir en una planta (y, por lo tanto, que es útil en la realización de los procedimientos de la invención) es cualquier ácido nucleico que codifique el tipo de proteína que se describirá ahora, también denominado de aquí en adelante "ácido nucleico de tipo FSM1" o "gen de tipo FSM1".

25 Un "polipéptido de tipo FSM1", como se define en el presente documento, se refiere a cualquier polipéptido que comprende un dominio MYB/SANT (Acceso de SMART SM00717, acceso Pfam PF00249).

Además o como alternativa, la proteína de tipo FSM1 comprende uno o más de los siguientes motivos:

- Motivo 1 (SEC ID N° 283): W[TS][PA]K[QE]NK[LA]FE[RK]ALAVYD[KR][DE]TPDR W[HSQ]N[VI]A[RK]A
- 30 Motivo 2 (SEC ID N° 284): GGK[ST][AV][ED]EV[KR]RHYE[IL]L
- Motivo 3 (SEC ID N°: 285): D[VL][KF][HF][ED][SN]G[RM]VPFP[NK]Y.

Estos motivos se obtuvieron usando el algoritmo MEME (Bailey y Elkan, Actas de la Segunda Conferencia Internacional sobre Sistemas Inteligentes de Biología Molecular, pág. 28-36, AAAI Press, Menlo Park, California, 1994.). Los motivos MEME representan los restos de aminoácidos que están presentes en al menos el 80 % de los las proteínas de tipo FSM1 dadas en el listado de secuencias. La ubicación de los motivos se muestra en la Fig. 1 (que representa la secuencia de la proteína de SEC ID N°: 2).

35

Como alternativa o además, el polipéptido de tipo FSM1 también comprende uno o más de los siguientes motivos:

- Motivo 4 (SEC ID N° 286): W(T/S)(A/V/T)(K/Q)(E/Q/D)(N/S)K(D/A)FEX(A/V)LA(E/V/T)(F/Y)D(K/R/Q)(D/E)(T/S)(P/A/R/C)(E/D/N)RWX(N/D)VA(R/K/Q/H)(A/V)(V/I)(G/E/A), en el que X de la posición 11 puede ser cualquier aminoácido, preferentemente uno entre E, R, K, S, Q, y en el que X de la posición 25 puede ser cualquier aminoácido, preferentemente, uno de S, A, Y, H, Q, K;
- 40 Motivo 5 (SEC ID N° 287): T(V/P/A/T)(E/D)(E/D)(V/A)KX(H/Q)Y(E/D)(V/I/L/H)L(L/V) en el que X de la posición 7 puede ser cualquier aminoácido, preferentemente, uno de K, R, Q, H, S;
- Motivo 6 (SEC ID N°: 288): I(E/D)(S/N)(G/D)XV(P/A)(L/F/Y)P,

45 en el que X de la posición 5 puede ser cualquier aminoácido, preferentemente uno de K, Q, R, H, M. Más preferentemente, el polipéptido de tipo FSM1 comprende, por orden creciente de preferencia, al menos 2, al menos 3, al menos 4, al menos 5 o los 6 motivos anteriores.

Además o como alternativa, el homólogo de una proteína de tipo FSM1 tiene, por orden creciente de preferencia, al menos 25 %, 26 %, 27 %, 28 %, 29 %, 30 %, 31 %, 32 %, 33 %, 34 %, 35 %, 36 %, 37 %, 38 %, 39 %, 40 %, 41 %, 42 %, 43 %, 44 %, 45 %, 46 %, 47 %, 48 %, 49 %, 50 %, 51 %, 52 %, 53 %, 54 %, 55 %, 56 %, 57 %, 58 %, 59 %, 60 %, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia general con la secuencia de aminoácidos representada por SEC ID N°: 2, siempre que la proteína homóloga comprenda un dominio MYB/SANT y preferentemente también uno cualquiera o más de los motivos conservados señalados anteriormente. La identidad de secuencia global se determina usando un algoritmo de alineación global, tal como el algoritmo de Needleman Wunsch del programa GAP (GCG Wisconsin Package, Accelrys), preferentemente con los parámetros por defecto y preferentemente con secuencias de proteínas maduras (es decir, sin tener en cuenta las señales de secreción ni los péptidos de tránsito).

50

55

- En comparación con la identidad de secuencia global, la identidad de secuencia generalmente será mayor cuando solo se consideran los dominios o motivos conservados. Preferentemente, los motivos de un polipéptido de tipo FSM1 tienen, por orden creciente de preferencia, al menos el 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad de secuencia con uno cualquiera o más de los motivos representados por la SEC ID N°: 283 a SEC ID N° 288 (motivos 1 a 6).
- Preferentemente, la secuencia de polipéptido, cuando se usa en la construcción de un árbol filogenético, tal como el representado en la Figura 3, se agrupa con el grupo de polipéptidos de tipo FSM1 que comprenden la secuencia de aminoácidos representada por SEC ID N°: 2 en lugar de con otras proteínas MYB.
- Los términos "dominio", "firma" y "motivo" se definen en el apartado de "definiciones" del presente documento.
- Además, los polipéptidos de tipo FSM1 (al menos en su forma nativa) normalmente tienen actividad de unión al ADN. Las herramientas y técnicas para medir la actividad de unión al ADN (por ejemplo, los ensayos de retardo en gel) son bien conocidas en la materia. En el Ejemplo 6, se proporcionan más detalles.
- Además, los polipéptidos de tipo FSM1, cuando se expresan en el arroz de acuerdo con los procedimientos de la presente invención como se describe en los Ejemplos 7 y 8, dan plantas que tienen mejores rasgos relacionados con el rendimiento, en particular, un mayor rendimiento seminal (Ejemplo 11).
- En cuanto a los polipéptidos de tipo FSM1, la presente invención se ilustra mediante la transformación de plantas con la secuencia de ácido nucleico representada por SEC ID N°: 1, que codifica la secuencia polipeptídica de SEC ID N°: 2. Sin embargo, la realización de la invención no se limita a estas secuencias; los procedimientos de la invención se pueden realizar ventajosamente usando cualquier ácido nucleico codificante de tipo FSM1 o polipéptido de tipo FSM1 como se define en el presente documento.
- Los ejemplos de ácidos nucleicos que codifican polipéptidos de tipo FSM1 se dan en la Tabla A1 del apartado de Ejemplos del presente documento. Dichos ácidos nucleicos son útiles en la realización de los procedimientos descritos en el presente documento. Las secuencias de aminoácidos dadas en la Tabla A1 del apartado de ejemplos son ejemplos de secuencias de ortólogos y parálogos del polipéptido de tipo FSM1 representado por SEC ID N°: 2, siendo los términos "ortólogos" y "parálogos" como se definen en el presente documento. Otros ortólogos y parálogos se pueden identificar fácilmente mediante la realización de la denominada búsqueda BLAST recíproca como se describe en el apartado de definiciones; donde la secuencia de consulta es SEC ID N°: 1 o SEC ID N°: 2, el segundo BLAST (retro-BLAST) sería contra las secuencias de *Solanum lycopersicum*.
- Las variantes de ácidos nucleicos también pueden ser útiles en la práctica de los procedimientos descritos en el presente documento. Los ejemplos de dichas variantes incluyen ácidos nucleicos que codifican homólogos y derivados de una cualquiera de las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos, siendo los términos "homólogo" y "derivado" como se definen en el presente documento. También son útiles en los procedimientos de la invención los ácidos nucleicos que codifican homólogos y derivados de ortólogos o parálogos de una cualquiera de las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos. Los homólogos y derivados útiles en los procedimientos descritos en el presente documento tienen sustancialmente la misma actividad biológica y funcional que la proteína no modificada de la que se derivan. Otras variantes útiles en la práctica de los procedimientos descritos en el presente documento son las variantes en las que el uso de codones se optimiza o en las que se eliminan los sitios diana de miARN.
- Otras variantes de ácido nucleico adicionales útiles en la práctica de los procedimientos descritos en el presente documento incluyen partes de ácidos nucleicos que codifican polipéptidos de tipo FSM1, variantes de corte y empalme de ácidos nucleicos que codifican polipéptidos de tipo FSM1, variantes alélicas de ácidos nucleicos que codifican polipéptidos de tipo FSM1 y variantes de ácidos nucleicos que codifican polipéptidos de tipo FSM1, obtenidas por combinación de genes. Las expresiones secuencia de hibridación, variante de corte y empalme, variante alélica y combinación de genes son como se describen en el presente documento.
- Los ácidos nucleicos que codifican polipéptidos de tipo FSM1 no necesitan ser ácidos nucleicos de longitud completa, ya que el rendimiento de los procedimientos descritos en el presente documento no se basa en el uso de secuencias de ácido nucleico de longitud completa. De acuerdo con la presente descripción, se proporciona un procedimiento para mejorar los rasgos relacionados con el rendimiento en plantas, que comprende introducir y expresar en una planta una parte de una cualquiera de las secuencias de ácidos nucleicos dadas en la Tabla A1 del apartado de Ejemplos, o una parte de un ácido nucleico que codifica un ortólogo, parálogo u homólogo de cualquiera de las secuencias de aminoácidos dadas en la Tabla A del apartado de Ejemplos.
- Una parte de un ácido nucleico se puede preparar, por ejemplo, haciendo una o más deleciones en el ácido nucleico. Las partes se pueden usar en forma aislada o pueden estar fusionadas a otras secuencias codificantes (o no codificante) con el fin de, por ejemplo, producir una proteína que combine varias actividades. Cuando se fusiona a otras secuencias codificantes, el polipéptido resultante producido tras la traducción puede ser más grande que el predicho para la parte de proteína.

En cuanto los polipéptidos de tipo FSM1, las partes útiles en los procedimientos descritos en el presente documento, codifican un polipéptido de tipo FSM1 como se define en el presente documento, y tienen sustancialmente la misma actividad biológica que las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos. Preferentemente, la parte es una parte de uno cualquiera de los ácidos nucleicos dados en la Tabla A1 del apartado de Ejemplos, o es una parte de un ácido nucleico que codifica un ortólogo o un parálogo de una cualquiera de las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos. Preferentemente, la parte es de al menos 150, 200, 250, 300, 350, 400 nucleótidos consecutivos de longitud, siendo los nucleótidos consecutivos de una cualquiera de las secuencias de ácidos nucleicos dadas en la Tabla A1 del apartado de Ejemplos, o de un ácido nucleico que codifica un ortólogo o un parálogo de una cualquiera de las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos. Lo más preferentemente, la parte es una parte del ácido nucleico de SEC ID N°: 1. Preferentemente, la parte codifica un fragmento de una secuencia de aminoácidos que comprende un dominio MYB/SANT, que cuando se usa en la construcción de un árbol filogenético, tal como el representado en la Figura 3, se agrupa con el grupo de polipéptidos de tipo FSM1 que comprenden la secuencia de aminoácidos representada por la SEC ID N°: 2 en lugar de con otras proteínas MYB y que comprende uno o más de los motivos 1 a 6 indicados anteriormente.

Otra variante de ácido nucleico útil en los procedimientos descritos en el presente documento es un ácido nucleico capaz de hibridarse, en condiciones de rigurosidad reducida, preferentemente en condiciones rigurosas, con un ácido nucleico que codifica un polipéptido de tipo FSM1 como se define en el presente documento, o con una parte como se define en el presente documento.

La presente descripción proporciona un procedimiento de mejora de los rasgos relacionados con el rendimiento en plantas, que comprende introducir y expresar en una planta un ácido nucleico capaz de hibridarse con uno cualquiera de los ácidos nucleicos dados en la Tabla A1 del apartado de Ejemplos, o que comprende introducir y expresar en una planta un ácido nucleico capaz de hibridarse con un ácido nucleico que codifique un ortólogo, parálogo u homólogo de cualquiera de las secuencias de ácidos nucleicos dadas en la Tabla A1 del apartado de Ejemplos.

En cuanto a los polipéptidos de tipo FSM1, las secuencias de hibridación útiles en los procedimientos de la invención codifican un polipéptido de tipo FSM1 como se define en el presente documento, que tiene sustancialmente la misma actividad biológica que las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos. Preferentemente, la secuencia de hibridación es capaz de hibridarse con el complemento de uno cualquiera de los ácidos nucleicos dados en la Tabla A1 del apartado de Ejemplos, o con una parte de cualquiera de estas secuencias, siendo una parte como se define anteriormente, o la secuencia de hibridación es capaz de hibridarse con el complemento de un ácido nucleico que codifica un ortólogo o un parálogo de una cualquiera de las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos. Lo más preferentemente, la secuencia de hibridación es capaz de hibridarse con el complemento de un ácido nucleico como el representado por SEC ID N°: 1 o una parte del mismo.

Preferentemente, la secuencia de hibridación codifica un polipéptido que comprende un dominio MYB/SANT y con una secuencia de aminoácidos que, cuando se usa en la construcción de un árbol filogenético, tal como el representado en la Figura 3, se agrupa con el grupo de polipéptidos de tipo FSM1 que comprenden la secuencia de aminoácidos representada por SEC ID N°: 2, en lugar de con otras proteínas MYB y que comprende uno o más de los motivos 1 a 6 descritos anteriormente.

Otra variante de ácido nucleico útil en los procedimientos descritos en el presente documento es una variante de corte y empalme que codifica un polipéptido de tipo FSM1 como se define anteriormente en el presente documento, estando la variante de corte y empalme definida en el presente documento.

La presente descripción proporciona un procedimiento de mejora de los rasgos relacionados con el rendimiento en plantas, que comprende introducir y expresar en una planta una variante de corte y empalme de una cualquiera de las secuencias de ácidos nucleicos dadas en la Tabla A1 del apartado de Ejemplos, o una variante de corte y empalme de un ácido nucleico que codifique un ortólogo, parálogo u homólogo de cualquiera de las secuencias de ácidos nucleicos dadas en la Tabla A1 del apartado de Ejemplos.

En cuanto a los polipéptidos de tipo FSM1, las variantes de corte y empalme preferidas son variantes de corte y empalme de un ácido nucleico representado por SEC ID N°: 1, o una variante de corte y empalme de un ácido nucleico que codifica un ortólogo o un parálogo de SEC ID N°: 2. Preferentemente, la secuencia de aminoácidos codificada por la variante de corte y empalme comprende un dominio MYB/SANT y tiene una secuencia que, cuando se usa en la construcción de un árbol filogenético, tal como el representado en la Figura 3, se agrupa con el grupo de polipéptidos de tipo FSM1 que comprenden la secuencia de aminoácidos representada por SEC ID N° 2, en lugar de con otras proteínas MYB, y que comprende uno o más de los motivos 1 a 6 señalados anteriormente.

Otra variante de ácido nucleico útil en la realización de los procedimientos descritos en el presente documento es una variante alélica de un ácido nucleico que codifica un polipéptido de tipo FSM1 como se define anteriormente en el presente documento, estando variante alélica definida en el presente documento.

La presente descripción proporciona un procedimiento de mejora de los rasgos relacionados con el rendimiento en plantas, que comprende introducir y expresar en una planta una variante alélica de uno cualquiera de los ácidos nucleicos dados en la Tabla A1 del apartado de Ejemplos, o que comprende introducir y expresar en una planta una variante alélica de un ácido nucleico que codifica un ortólogo, un parólogo o un homólogo de cualquiera de las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos.

En cuanto a los polipéptidos de tipo FSM1, los polipéptidos codificados por variantes alélicas útiles en los procedimientos de la presente invención tienen sustancialmente la misma actividad biológica que el polipéptido de tipo FSM1 de SEC ID N°: 2 y cualquiera de los aminoácidos representados en la Tabla A1 del apartado de Ejemplos. Las variantes alélicas existen en la naturaleza, y dentro de los procedimientos de la presente invención se engloba el uso de estos alelos naturales. Preferentemente, la variante alélica es una variante alélica de SEC ID N°: 1 o una variante alélica de un ácido nucleico que codifica un ortólogo o un parólogo de SEC ID N°: 2. Preferentemente, la secuencia de aminoácidos codificada por la variante alélica comprende un dominio MYB/SANT y tiene una secuencia que, cuando se usa en la construcción de un árbol filogenético, tal como el representado en la Figura 3, se agrupa con el grupo de polipéptidos de tipo FSM1 que comprenden la secuencia de aminoácidos representada por SEC ID N°: 2, en lugar de con otras proteínas MYB, y que comprende uno o más de los motivos 1 a 6 indicados anteriormente.

La combinación de genes o la evolución dirigida también se pueden usar para generar variantes de ácidos nucleicos que codifican los polipéptidos de tipo FSM1 según lo definido anteriormente, estando la expresión "combinación de genes" definida en el presente documento.

La presente descripción proporciona un procedimiento de mejora de los rasgos relacionados con el rendimiento en plantas, que comprende introducir y expresar en una planta una variante de una cualquiera de las secuencias de ácidos nucleicos dadas en la Tabla A1 del apartado de Ejemplos, o que comprende introducir y expresar en una planta una variante de un ácido nucleico que codifica un ortólogo, un parólogo o un homólogo de cualquiera de las secuencias de aminoácidos dadas en la Tabla A1 del apartado de Ejemplos, variante de ácido nucleico que se obtiene mediante la combinación de genes.

En cuanto a los polipéptidos de tipo FSM1, preferentemente, la secuencia de aminoácidos codificada por la variante de ácido nucleico obtenida por combinación de genes comprende un dominio MYB/SANT y tiene una secuencia que, cuando se usa en la construcción de un árbol filogenético, tal como el representado en la Figura 3, se agrupa con el grupo de polipéptidos de tipo FSM1 que comprenden la secuencia de aminoácidos representada por SEC ID N°: 2 en lugar de con otras proteínas MYB, y que comprende uno o más de los motivos 1 a 6 indicados anteriormente.

Además, las variantes de ácido nucleico también se pueden obtener por mutagénesis dirigida. Hay varios procedimientos disponibles para realizar la mutagénesis dirigida, siendo los más comunes los procedimientos basados en PCR ("Current Protocols in Molecular Biology". Wiley Eds.).

Los ácidos nucleicos que codifican polipéptidos de tipo FSM1 se pueden obtener de cualquier fuente natural o artificial. El ácido nucleico se puede modificar con respecto de su forma nativa en composición y/o ambiente genómico a través de la manipulación humana deliberada. Preferentemente, el ácido nucleico que codifica el polipéptido de tipo FSM1 procede de una planta, más preferentemente de una planta dicotiledónea, más preferentemente de la familia *Solanaceae*, y lo más preferentemente el ácido nucleico es de *Solanum lycopersicum*.

La realización de los procedimientos descritos en el presente documento da plantas que tienen mejores rasgos relacionados con el rendimiento. En particular, la realización de los procedimientos de la invención da plantas que tienen mayor vigor temprano y/o mayor rendimiento, especialmente mayor rendimiento seminal con respecto a las plantas de control. Los términos "rendimiento" y "rendimiento seminal" se describen más detalladamente en el apartado de "Definiciones" del presente documento.

La referencia en el presente documento a mejores rasgos relacionados con el rendimiento pretende significar un aumento del vigor temprano y/o en la biomasa (peso) de una o más partes de una planta, que pueden incluir partes (cosechables) aéreas y/o partes (cosechables) por debajo del suelo. En particular, dichas partes cosechables son partes vegetativas de la planta, pero las semillas también se incluyen en dichas partes cosechables. La realización de los procedimientos de la invención genera plantas que tienen mayor biomasa y rendimiento seminal con respecto a la biomasa y al rendimiento seminal de las plantas de control.

La presente descripción proporciona un procedimiento de mejora de los rasgos relacionados con el rendimiento, especialmente del rendimiento seminal de las plantas y/o de la biomasa (partes cosechables por debajo del suelo), con respecto a las plantas de control, procedimiento que comprende modular la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 como se define en el presente documento.

Dado que las plantas transgénicas de acuerdo con la presente invención tienen mayores rasgos relacionados con el rendimiento, es probable que estas plantas presenten una mayor tasa de crecimiento (durante al menos parte de su ciclo de vida), con respecto a la tasa de crecimiento de las plantas de control en una fase correspondiente de su ciclo de vida.

De acuerdo con una característica preferida de la presente descripción, la realización de los procedimientos de la invención da plantas que tienen una mayor tasa de crecimiento en comparación con las plantas de control. Por lo tanto, de acuerdo con la presente descripción, se proporciona un procedimiento de aumento de la tasa de crecimiento de las plantas, procedimiento que comprende modular la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en el presente documento.

La realización de los procedimientos descritos en el presente documento que da plantas cultivadas en condiciones sin estrés o en condiciones de sequía moderada aumentó del rendimiento en comparación con las plantas de control cultivadas en condiciones comparables. Por lo tanto, en el presente documento, se desvela un procedimiento de aumento del rendimiento en plantas cultivadas en condiciones sin estrés o en condiciones de sequía moderada, procedimiento que comprende modular la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.

La realización de los procedimientos descritos en el presente documento que da plantas cultivadas en condiciones de déficit de nutrientes, particularmente en condiciones de déficit de nitrógeno, aumentó el rendimiento en comparación con las plantas de control cultivadas en condiciones comparables. Por lo tanto, en el presente documento, se desvela un procedimiento de aumento del rendimiento en plantas cultivadas en condiciones de déficit de nutrientes, procedimiento que comprende modular la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.

La realización de los procedimientos descritos en el presente documento que da plantas cultivadas en condiciones de estrés salino aumentó el rendimiento en comparación con las plantas de control cultivadas en condiciones comparables. Por lo tanto, en el presente documento, se desvela un procedimiento de aumento del rendimiento en plantas cultivadas en condiciones de estrés salino, procedimiento que comprende modular la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.

La invención también proporciona construcciones genéticas y vectores para facilitar la introducción y/o la expresión en plantas de ácidos nucleicos que codifican polipéptidos de tipo FSM1. Las construcciones génicas se pueden insertar en vectores, que pueden estar disponibles en el mercado, adecuados para transformar en plantas y adecuados para la expresión del gen de interés en las células transformadas. La invención también proporciona el uso de una construcción génica según lo definido en el presente documento en los procedimientos de la invención.

Más concretamente, la presente descripción desvela una construcción que comprende:

- (a) un ácido nucleico que codifica un polipéptido de tipo FSM1 como se ha definido anteriormente;
- (b) una o más secuencias de control capaces de dirigir la expresión de la secuencia de ácido nucleico de (a); y opcionalmente
- (c) una secuencia de terminación de la transcripción.

Preferentemente, el ácido nucleico que codifica un polipéptido de tipo FSM1 es como se ha definido anteriormente. Las expresiones "secuencia de control" y "secuencia de terminación" son como se han definido en el presente documento.

Las plantas se transforman con un vector que comprende cualquiera de los ácidos nucleicos descritos anteriormente. El experto en la materia conoce bien los elementos genéticos que deben estar presentes en el vector para transformar, seleccionar y propagar correctamente células huésped que contienen la secuencia de interés. La secuencia de interés está ligada operativamente a una o más secuencias de control (al menos a un promotor).

Ventajosamente, se puede usar cualquier tipo de promotor, bien natural o sintético, para dirigir la expresión de la secuencia de ácido nucleico, pero preferentemente el promotor es de origen vegetal. Un promotor constitutivo es particularmente útil en los procedimientos. Preferentemente, el promotor constitutivo es un promotor constitutivo ubicuo de resistencia media. Véase el apartado de "Definiciones" del presente documento para consultar las definiciones de los diversos tipos de promotores.

En cuanto a los polipéptidos de tipo FSM1, debería quedar claro que la aplicabilidad de la presente invención no se limita al ácido nucleico que codifica el polipéptido de tipo FSM1 representado por SEC ID N°: 1, ni la aplicabilidad de la invención se restringe a la expresión de un ácido nucleico que codifica el polipéptido de tipo FSM1 cuando es dirigida por un promotor constitutivo.

Preferentemente, el promotor constitutivo de fuerza media se obtiene de una planta, tal como un promotor GOS2, más preferentemente es el promotor GOS2 del arroz. Por otro lado, preferentemente, el promotor constitutivo está representado por una secuencia de ácido nucleico sustancialmente similar a SEC ID N°: 289, más preferentemente el promotor constitutivo es como el representado por la SEC ID N°: 289. Véase el apartado de "Definiciones" del presente documento para obtener más ejemplos de promotores constitutivos.

Opcionalmente, se pueden usar una o más secuencias terminadoras en la construcción introducida en una planta. Preferentemente, la construcción comprende un casete de expresión que comprende un promotor GOS2, sustancialmente similar a SEC ID N°: 289, y el ácido nucleico que codifica el polipéptido de tipo FSM1.

De acuerdo con una característica preferida de la invención, la expresión modulada es el aumento de la expresión. Los procedimientos para aumentar la expresión de ácidos nucleicos o genes, o productos génicos, están bien documentados en la técnica y, en el apartado de definiciones, se proporcionan ejemplos.

5 que codifica un polipéptido de tipo FSM1 es mediante la introducción y expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1; sin embargo los efectos de realizar el procedimiento, es decir, de mejorar rasgos relacionados con el rendimiento, también se pueden obtener usando otras técnicas bien conocidas, incluyendo, pero sin limitación, el marcaje de activación de ADN-T, TILLING, recombinación homóloga. En el apartado de definiciones, se proporciona una descripción de estas técnicas.

10 La descripción también proporciona un procedimiento de producción de plantas transgénicas que tienen mejores rasgos relacionados con el rendimiento en comparación con las plantas de control, que comprende la introducción y expresión en una planta de cualquier ácido nucleico que codifica un polipéptido de tipo FSM1, como se define anteriormente en el presente documento.

Más concretamente, la presente invención proporciona un procedimiento de producción de plantas transgénicas que tiene mejor biomasa y/o rendimiento seminal, procedimiento que comprende:

- 15 (i) introducir y expresar, en una planta o célula vegetal, un ácido nucleico que codifica un polipéptido de tipo FSM1; y
(ii) cultivar la célula vegetal en condiciones que potencien el crecimiento y desarrollo de la planta.

El ácido nucleico de (i) puede ser cualquiera de los ácidos nucleicos capaces de codificar un polipéptido de tipo FSM1 según lo definido en el presente documento.

20 El ácido nucleico se puede introducir directamente en una célula vegetal o en la propia planta (incluyendo la introducción en un tejido, órgano o cualquier otra parte de una planta). De acuerdo con una característica preferida de la presente invención, el ácido nucleico se introduce preferentemente en una planta mediante transformación. El término "transformación" se describe con más detalle en el apartado de "Definiciones" del presente documento.

25 La presente invención se extiende claramente a cualquier célula vegetal o planta producida por cualquiera de los procedimientos descritos en el presente documento, y a todas las partes de la planta y propágulos de las mismas. La presente invención abarca plantas o partes de las mismas (incluyendo semillas) que se pueden obtener mediante los procedimientos de acuerdo con la presente invención. Las plantas o partes de las mismas comprenden un transgén de ácido nucleico que codifica un polipéptido de tipo FSM1 como se ha definido anteriormente. La presente invención se extiende además para abarcar la progenie de una célula transformada o transfectada primaria, tejido,
30 órgano o planta entera que se haya producido mediante cualquiera de los procedimientos anteriormente mencionados, siendo el único requisito que la progenie presente la/s misma/s característica/s genotípica/s y/o fenotípica/s que aquellas producidas por el padre en los procedimientos de acuerdo con la invención.

35 La descripción también incluye células huésped que contienen un ácido nucleico aislado que codifica un polipéptido de tipo FSM1. Las células huésped preferidas de acuerdo con la invención son células vegetales. Las plantas huésped para los ácidos nucleicos o el vector usados en el procedimiento descrito en el presente documento, [...] son capaces de sintetizar los polipéptidos usados en el procedimiento de la invención.

40 Los procedimientos de la invención son ventajosamente aplicables a cualquier planta. Las plantas que son particularmente útiles en los procedimientos de la invención incluyen todas las plantas que pertenecen a la superfamilia Viridiplantae, en particular, plantas monocotiledóneas y dicotiledóneas, incluyendo forraje o leguminosas forrajeras, plantas ornamentales, cultivos alimentarios, árboles o arbustos. De acuerdo con una realización preferida de la presente invención, la planta es una planta de cultivo. Los ejemplos de plantas de cultivo incluyen soja, girasol, alfalfa, colza, remolacha azucarera, semilla de lino, algodón, tomate, patata y tabaco. Más preferentemente, la planta es una planta monocotiledónea. Los ejemplos de plantas monocotiledóneas incluyen la caña de azúcar. Más preferentemente, la planta es un cereal. Los ejemplos de cereales son arroz, maíz, trigo,
45 cebada, mijo, centeno, triticale, sorgo, escanda, espelta, secale, escanda menor, teff, milo y avena.

50 La invención también se extiende a partes cosechables de una planta tales como, pero sin limitación, semillas, hojas, frutos, flores, tallos, raíces, rizomas, tubérculos y bulbos, partes cosechables que comprenden un ácido nucleico recombinante que codifica un polipéptido de tipo FSM1. La divulgación se refiere además a productos derivados, preferentemente derivados directamente, de una parte cosechable de dicha planta, tales como gránulos o polvos secos, aceite, grasa y ácidos grasos, almidón o proteínas. La presente invención también abarca el uso de ácidos nucleicos que codifican polipéptidos de tipo FSM1 como se describe en el presente documento y el uso de estos polipéptidos de tipo FSM1 en la mejora de cualquiera de los rasgos relacionados con el rendimiento de las plantas anteriormente mencionados. Por ejemplo, los ácidos nucleicos que codifican un polipéptido de tipo FSM1 descrito en el presente documento, o los propios polipéptidos de tipo FSM1, pueden encontrar uso en programas de
55 reproducción en los que se identifica un marcador de ADN que puede estar genéticamente ligado a un gen que codifica un polipéptido de tipo FSM1. Los ácidos nucleicos/genes, o los propios polipéptidos de tipo FSM1, se pueden usar para definir un marcador molecular. Este ADN o marcador proteico se puede usar entonces en programas de reproducción para seleccionar plantas que tengan mejores rasgos relacionados con el rendimiento

según lo definido anteriormente en el presente documento en los procedimientos de la invención. Además, las variantes alélicas de un ácido nucleico/gen que codifica un polipéptido de tipo FSM1 pueden encontrar uso en los programas de reproducción asistida por marcadores. Los ácidos nucleicos que codifican un polipéptido de tipo FSM1, también se pueden usar como sondas para cartografiar genética y físicamente los genes de los que forman parte, y como marcadores para los rasgos ligados a esos genes. Dicha información puede ser útil en la reproducción de la planta para desarrollar líneas con fenotipos deseados.

Descripción de las figuras

Ahora, se describirá la presente invención con referencia a las siguientes figuras, en las que:

- La **Figura 1** representa la estructura de dominio de la SEC ID N°: 2 con los motivos conservados subrayados y numerados. El dominio MYB/SANT (PF00249) se indica en cursiva y se extiende de Ser11 a Leu60.
- La **Figura 2** representa una alineación múltiple de varios polipéptidos de tipo FSM1. Los asteriscos indican aminoácidos idénticos entre las diversas secuencias de proteínas; los dos puntos representan sustituciones de aminoácidos altamente conservados; y los puntos representan la sustitución de aminoácidos menos conservados; en otras posiciones, no hay conservación de secuencia. Estas alineaciones se pueden usar para definir más motivos, cuando se usan aminoácidos conservados.
- La **Figura 3** muestra el árbol filogenético de los polipéptidos de tipo FSM1, construido como se describe en Ejemplo 2. El subtipo SANT/MYB está en azul y comprende todas las proteínas de tipo FSM1 dadas en la lista de secuencias, los otros subtipos representan otras proteínas MYB.
- La **Figura 4** representa el vector binario usado para aumentar la expresión en *Oryza sativa* de un ácido nucleico que codifica el tipo FSM1 bajo el control de un promotor GOS2 de L. arroz (pGOS2).
- La **Figura 5** representa la secuencia de consenso del dominio SANT como se presenta en la base de datos SMART. Los diferentes símbolos de los motivos se aclaran en la tabla, donde para cada clase de aminoácido se dan los respectivos restos con el código de una letra.

Ejemplos

La presente invención se describirá ahora con referencia a los siguientes ejemplos, que solo figuran a modo ilustrativo. Los siguientes ejemplos no pretenden definir completamente ni limitar de otro modo el alcance de la invención.

Manipulación de DNA: a menos que se indique lo contrario, las técnicas de DNA recombinante se realizan de acuerdo con protocolos convencionales descritos en (Sambrook (2001) "Molecular Cloning: a laboratory manual", III Edición, Cold Spring Harbor Laboratory Press, CSH, Nueva York) o en los Volúmenes 1 y 2 de Ausubel *et al.* (1994), "Current Protocols In Molecular Biology, Current Protocols". Los materiales y procedimientos convencionales para el trabajo molecular de las plantas se describen en "Plant Molecular Biology" Labfax (1993) de R. D. D. Croy, publicado por BIOS Scientific Publications Ltd (RU) y Blackwell Scientific Publications (RU).

Ejemplo 1: Identificación de secuencias relacionadas con la secuencia de ácido nucleico usada en los procedimientos de la invención

1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

Las secuencias (ADNc de longitud completa, EST o genómicas) relacionadas con SEC ID N° 1 y SEC ID N° 2 se identificaron entre aquellas mantenidas en la base de datos de nucleótidos Entrez del Centro Nacional para la Información Biotecnológica (NCBI) usando las herramientas de búsqueda de secuencias de base de datos, tales como la herramienta de alineación local básica (BLAST) (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410; y Altschul *et al.* (1997) *Nucleic acid Res.* 25:3389-3402). El programa se usa para encontrar regiones de similitud local entre las secuencias al comparar las secuencias de ácido nucleico o de polipéptido con las bases de datos de secuencias y al calcular la significancia estadística de las coincidencias. Por ejemplo, el polipéptido codificado por el ácido nucleico de SEC ID N° 1 se usó para el algoritmo TBLASTN, con las configuraciones por defecto y el filtro para ignorar las secuencias de baja complejidad desencadenadas. El rendimiento del análisis se ve mediante la comparación por pares, y se clasifica de acuerdo con el orden de probabilidad (valor E), donde la puntuación refleja la probabilidad de que ocurra una determinada alineación por casualidad (cuanto menor es el valor E, más significativo es el acierto). Además de los valores E, las comparaciones también se puntuaron por el porcentaje de identidad. El porcentaje de identidad se refiere al número de nucleótidos (o aminoácidos) idénticos entre las dos secuencias de ácido nucleico (o polipéptido) comparadas en una longitud particular. En algunos casos, se pueden ajustar los parámetros por defecto para modificar la rigurosidad de la búsqueda. Por ejemplo, se puede aumentar el valor E para mostrar coincidencias menos rigurosas. De este modo, se pueden identificar coincidencias cortas casi exactas.

La Tabla A1 proporciona una lista de secuencias de ácidos nucleicos relacionadas con SEC ID N° 1 y SEC ID N° 2.

Tabla A1: Ejemplos de ácidos nucleicos y polipéptidos de tipo FSM1:

Nombre	SEC ID N° de ácido nucleico	SEC ID N° de polipéptido
S.lycopersicum_TA46686_4081#1	1	2
A.majus_AJ793240#1	3	4
A.majus_TA4522_4151#1	5	6
A.thaliana_AT1G19510.1#1	7	8
A.thaliana_AT1G75250.2#1	9	10
A.thaliana_AT2G21650.1#1	11	12
At4g36570_l-box-binding-like_NA	13	14
A.thaliana_AT4G39250.1#1	15	16
B.napus_BN06MC20309_46582753@20242#1	17	18
B.napus_EE472702#1	19	20
B.napus_EL589700#1	21	22
B.rapa_L38243#1	23	24
C.annuum_BM060888#1	25	26
C.clementina_CX295458#1	27	28
C.intybus_EH706852#1	29	30
C.maculosa_EH726540#1	31	32
C.sinensis_TA19011_2711#1	33	34
E.grandis_CD669972#1	35	36
E.gunnii_CT980385#1	37	38
E.gunnii_CT985658#1	39	40
E.tereticornis_CD668687#1	41	42
F.arundinacea_DT700402#1	43	44
F.vesca_EX665737#1	45	46
G.hirsutum_DR461014#1	47	48
G.hirsutum_DW476156#1	49	50
G.hirsutum_DW497247#1	51	52
G.hirsutum_DW503354#1	53	54
G.hirsutum_TA26706_3635#1	55	56
G.hirsutum_TA31277_3635#1	57	58
G.hirsutum_TA36120_3635#1	59	60
G.hirsutum_TA37973_3635#1	61	62
G.hirsutum_TA43581_3635#1	63	64
G.max_CA852793#1	65	66
G.max_Glyma01g38250.1#1	67	68
G.max_Glyma02g06230.1#1	69	70
G.max_Glyma02g06240.1#1	71	72

ES 2 542 536 T3

(Continuación)

Nombre	SEC ID N° de ácido nucleico	SEC ID N° de polipéptido
G.max_Glyma02g18210.1#1	73	74
G.max_Glyma02g41670.1#1	75	76
G.max_Glyma03g28050.1#1	77	78
G.max_Glyma04g16390.1#1	79	80
G.max_Glyma06g03490.1#1	81	82
G.max_Glyma06g46590.1#1	83	84
G.max_Glyma12g04680.1#1	85	86
G.max_Glyma14g08090.3#1	87	88
G.max_Glyma16g25250.1#1	89	90
G.max_Glyma18g04120.1#1	91	92
H.argophyllus_EE611072#1	93	94
H.exilis_EE652824#1	95	96
H.vulgare_BQ762901#1	97	98
H.vulgare_CK125210#1	99	100
I.batatas_EE882249#1	101	102
I.nil_CJ754125#1	103	104
L.perennis_DW074785#1	105	106
L.sativa_TA7946_4236#1	107	108
L.serriola_DW122081#1	109	110
L.virosa_TA3253_75947#1	111	112
M.acuminata_ES435134#1	113	114
M.domestica_DR994824#1	115	116
M.domestica_TA45680_3750#1	117	118
M.esculenta_CK645030#1	119	120
M.guttatus_DV207881#1	121	122
M.truncatula_AC135566_10.5#1	123	124
M.truncatula_AC157502_5.5#1	125	126
M.truncatula_AC162787_11.4#1	127	128
M.truncatula_AC162787_14.4#1	129	130
M.truncatula_CT033768_12.4#1	131	132
M.truncatula_CX518743#1	133	134
N.benthamiana_EH366122#1	135	136
O.minuta_TA1082_63629#1	137	138
O.sativa_LOC_Os01g44370.1#1	139	140
O.sativa_LOC_Os01g44390.1#1	141	142
O.sativa_LOC_Os01g47370.1#1	143	144
O.sativa_LOC_Os02g47744.1#1	145	146

ES 2 542 536 T3

(Continuación)

Nombre	SEC ID N° de ácido nucleico	SEC ID N° de polipéptido
O.sativa_LOC_Os05g49240.1#1	147	148
O.sativa_LOC_Os05g50350.1#1	149	150
O.sativa_LOC_Os06g28630.1#1	151	152
O.sativa_LOC_Os07g26150.1#1	153	154
O.sativa_LOC_Os12g33950.1#1	155	156
O.sativa_Os05g0579600#1	157	158
P.americana_C0997831#1	159	160
P.americana_DT578382#1	161	162
P.coccineus_CA897021#1	163	164
P.euphratica_AJ768009#1	165	166
P.euphratica_AJ768184#1	167	168
P.euphratica_AJ768574#1	169	170
P.glauca_DR559298#1	171	172
P.pinaster_BX682614#1	173	174
P.taeda_TA26731_3352#1	175	176
P.tremula_BU823356#1	177	178
P.tremula_TA8294_113636#1	179	180
P.trichocarpa_553339#1	181	182
P.trichocarpa_557252#1	183	184
P.trichocarpa_scaff_57.171#1	185	186
P.trichocarpa_scaff_IV.1184#1	187	188
P.trichocarpa_scaff_IX.506#1	189	190
P.trichocarpa_scaff_V.1348#1	191	192
P.trichocarpa_TA23239_3694#1	193	194
S.bicolor_Sb01g040910.1#1	195	196
S.bicolor_Sb03g028950.1#1	197	198
S.bicolor_Sb03g028960.1#1	199	200
S.bicolor_Sb03g030330.1#1	201	202
S.bicolor_Sb04g030510.1#1	203	204
S.bicolor_Sb09g028790.1#1	205	206
S.bicolor_Sb09g029560.1#1	207	208
S.henryi_TA107_13258#1	209	210
S.indicum_BU668323#1	211	212
S.officinatum_CA111191#1	213	214
S.officinatum_CA283940#1	215	216
S.officinatum_TA43900_4547#1	217	218
S.officinatum_TA47846_4547#1	219	220

(Continuación)

Nombre	SEC ID N° de ácido nucleico	SEC ID N° de polipéptido
S.tuberosum_BI433702#1	221	222
S.tuberosum_CV500497#1	223	224
S.tuberosum_CV504810#1	225	226
T.aestivum_BE427243#1	227	228
T.aestivum_BM259034#1	229	230
T.aestivum_CA623533#1	231	232
T.aestivum_CV775992#1	233	234
T.aestivum_TA110455_4565#1	235	236
T.aestivum_TA95415_4565#1	237	238
T.erecta_SIN_31b-CS_SCR20-B14.b1-----@5804#1	239	240
T.pratense_BB920325#1	241	242
T.salsuginea_DN773374#1	243	244
V.vinifera_CA812415#1	245	246
V.vinifera_CB345061#1	247	248
V.vinifera_CB972578#1	249	250
V.vinifera_EE072107#1	251	252
V.vinifera_GSVIVT00014602001#1	253	254
V.vinifera_GSVIVT00028247001#1	255	256
V.vinifera_GSVIVT00036374001#1	257	258
Z.mays_c58977309gm030403@3901#1	259	260
Z.mays_c68506706gm030403@11442#1	261	262
Z.mays_DR787399#1	263	264
Z.mays_s58275634gm030403@34722#1	265	266
Z.mays_TA203683_4577#1	267	268
Z.mays_TA206461_4577#1	269	270
Z.mays_TA208256_4577#1	271	272
Z.mays_TA210755_4577#1	273	274
Z.mays_TA219702_4577#1	275	276
Z.mays_ZM07MSbpsHQ_57388861.r01@38768#1	277	278
Z.officinale_DY348802#1	279	280
Z.officinale_DY369792#1	281	282

Las secuencias han sido ensambladas provisionalmente y desveladas al público por instituciones de investigación tales como The Institute for Genomic Research (TIGR; partiendo de TA). La base de datos de ortólogos de genes eucariotas (EGO) se pueden usar para identificar dichas secuencias relacionadas, bien mediante búsqueda por palabras clave o usando el algoritmo BLAST con la secuencia de ácido nucleico o la secuencia de polipéptido de interés. Se han creado bases de datos de secuencias de ácido nucleico especiales para determinados organismos, tales como la creada por Joint Genome Institute. Por otro lado, el acceso a bases de datos patentadas ha permitido la identificación de secuencias de ácido nucleico y de polipéptido.

Ejemplo 2: Alineación de secuencias relacionadas con las secuencias polipeptídicas usadas en los procedimientos de la invención1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

5 Se realizó la alineación de una serie de secuencias de polipéptidos de tipo FSM1 usando el algoritmo ClustalW 2.0 de alineación progresiva (Thompson *et al.* (1997) *Nucleic acids Res* 25:4876-4882; Chenna *et al.* (2003). *Nucleic acids Res* 31:3497-3500) con la configuración convencional (alineación lenta, matriz de similitud: Gonnet, penalización por apertura de hueco de 10, penalización por extensión de hueco de 0,2). Se realizó la edición manual menor para optimizar más la alineación. Los polipéptidos de tipo FSM1 se alinean en la Figura 2.

10 Se construyó un árbol filogenético de los polipéptidos de tipo FSM1 (Figura 3) de la siguiente manera: se generó la alineación usando MAFFT (Kato y Toh (2008) *Briefings in Bioinformatics* 9: 286-298). Se calculó un árbol de unión a vecinos usando QuickTree (Howe *et al.* (2002), *Bioinformatics* 18(11):1546-7), 100 repeticiones de arranque. El cladograma circular se dibujó usando Dendroscope (Huson *et al.* (2007), *BMC Bioinformatics* 8(1): 460). Se indica la confianza para 100 repeticiones de arranque para la ramificación principal.

Ejemplo 3: Cálculo del porcentaje de identidad global entre las secuencias de polipéptido útiles en la realización de los procedimientos de la invención

15 Se determinan los porcentajes globales de similitud e identidad entre las secuencias polipeptídicas de longitud completa útiles en la realización de los procedimientos de la invención usando uno de los procedimientos disponibles en la técnica, el programa informático MatGAT (herramienta de alineación global de matriz) (BMC Bioinformatics. 2003 4:29. "MatGAT: an application that generates similarity/identity matrices using protein o DNA sequences".
 20 Campanella J. J., Bitincka L., Smalley J; programa propiedad de Ledion Bitincka). El programa MatGAT genera matrices de similitud/identidad para las secuencias de ADN o de proteína sin la necesidad de realizar una alineación previa de los datos. El programa realiza una serie de alineaciones por pares usando al algoritmo de alineación global de Myers y Miller (con una penalización por hueco abierto de 12 y una penalización de extensión de hueco de 2), calcula la similitud e identidad usando, por ejemplo, Blosum 62 (para los polipéptidos), y luego coloca los resultados
 25 en una matriz de distancia. Se muestra la similitud de secuencias en la mitad inferior de la línea divisoria, y se muestra la identidad de secuencias en la mitad superior de la línea divisoria diagonal.

Los parámetros usados en la comparación fueron: Matriz de clasificación: Blosum62; Primer hueco: 12; Extensión de hueco: 2.

1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

30 En la Tabla B1, se muestran los resultados del análisis para la similitud e identidad globales en la longitud completa de una serie de secuencias de polipéptidos de tipo FSM1. La identidad de secuencia (en %) entre las secuencias de los polipéptidos de tipo FSM1 útiles en la realización de los procedimientos de la invención generalmente es del 50 % o superior en comparación con SEC ID N°: 2, pero puede ser tan baja como del 19 % en el grupo de polipéptidos de tipo FSM1 dado en la lista de secuencias.

35

ES 2 542 536 T3

Tabla B1: Resultados de MatGAT para la similitud e identidad globales en la longitud completa de las secuencias polipeptídicas

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. <i>M.guttatus</i> _DV207881		59,4	55,7	59,0	61,2	57,4	52,9	57,8	60,5	63,9	62,8	62,7	63,9
2. <i>A.majus</i> _TA4522_4151	71,0		54,2	58,3	58,1	69,3	68,6	59,8	58,3	61,3	63,9	61,2	61,9
3. <i>S.lycopersicum</i> _TA46686	70,5	73,1		79,5	63,6	62,4	60,8	63,6	67,0	64,8	68,2	67,4	68,2
4. <i>N.benthamiana</i> _EH366122	74,7	73,1	86,4		69,0	57,4	59,8	73,2	69,8	74,1	72,9	78,3	78,0
5. <i>G.hirsutum</i> _DR461014	77,1	72,0	72,7	80,2		56,4	61,8	72,9	72,1	72,8	75,3	70,6	71,5
6. <i>V.vinifera</i> _EE072107	67,3	77,2	69,3	65,3	71,3		77,5	63,4	65,3	59,4	68,3	63,4	64,4
7. <i>M.domestica</i> _DR994824	66,7	82,4	71,6	66,7	70,6	88,2		68,6	69,9	61,8	72,8	69,6	71,6
8. <i>G.hirsutum</i> _TA31277_3635	72,3	73,1	76,1	84,0	87,7	70,3	73,5		80,2	79,3	80,2	80,7	81,7
9. <i>G.hirsutum</i> _DW503354	76,7	74,2	78,4	79,1	84,9	74,3	77,5	86,0		73,3	83,7	77,9	79,1
10. <i>H.exilis</i> _EE652824	77,1	71,0	73,9	80,2	80,2	66,3	66,7	86,3	81,4		76,5	79,5	81,7
11. <i>E.gunnii</i> _CT985658	77,6	78,5	79,5	81,2	84,7	77,2	79,4	85,9	90,7	80,0		84,9	84,9
12. <i>L.sativa</i> _TA7946_4236	75,9	78,5	77,3	84,3	86,7	74,3	75,5	91,6	88,4	84,3	91,8		96,4
13. <i>C.intybus</i> _EH706852	75,9	77,9	77,3	84,1	87,8	75,2	76,5	91,5	89,5	85,4	90,6	96,4	
14. <i>P.trichocarpa_sc_IV.1184</i>	64,3	80,6	73,5	73,5	74,5	82,2	84,3	77,6	80,6	71,4	79,6	80,6	80,6
15. <i>P.trichocarpa_557252</i>	70,1	76,3	78,4	79,3	82,8	76,2	76,5	85,1	89,7	82,5	89,7	90,8	88,5
16. <i>P.euphratica</i> _AJ768574	71,3	77,4	79,5	80,5	83,9	77,2	77,5	86,2	90,8	79,3	90,8	92,0	89,7
17. <i>M.truncatula</i> _AC135566	65,9	82,8	73,6	72,5	71,4	72,3	75,5	71,4	72,5	67,0	71,4	72,5	72,5
18. <i>G.max_Glyma06g46590.1</i>	65,2	83,9	72,8	70,7	76,1	76,2	80,4	77,2	77,2	67,4	73,9	78,3	78,3
19. <i>G.max_Glyma04g16390.1</i>	66,3	83,9	75,0	70,7	75,0	78,2	82,4	77,2	78,3	69,6	75,0	79,3	79,3
20. <i>G.max_Glyma02g18210.1</i>	65,6	81,7	69,9	69,9	71,0	76,2	78,4	72,0	75,3	65,6	72,0	75,3	74,2
21. <i>M.truncatula</i> _AC157502	64,1	80,6	71,7	70,7	78,3	80,2	81,4	76,1	77,2	70,7	76,1	79,3	78,3
22. <i>G.max_Glyma03g28050.1</i>	64,9	80,4	69,1	68,0	70,1	81,2	80,4	71,1	74,2	70,1	76,3	77,3	76,3
23. <i>A.thaliana</i> _At4g39250	63,0	77,0	68,0	65,0	67,0	84,2	85,3	69,0	72,0	65,0	736,0	73,0	72,0
24. <i>T.salsuginea</i> _DN773374	58,0	76,0	65,0	67,0	66,0	81,2	84,3	69,0	73,0	64,0	70,0	69,0	71,0
25. <i>B.napus</i> _BN06MC20309	56,0	73,0	62,0	64,0	63,0	79,2	81,4	66,0	70,0	62,0	68,0	68,0	68,0
26. <i>A.thaliana</i> _At2g21650	57,4	78,2	65,3	63,3	66,3	82,2	84,3	68,3	72,3	64,4	69,3	70,3	69,3

ES 2 542 536 T3

	14	15	16	17	18	19	20	21	22	23	24	25	26
1. M.guttatus_DV207881	52,0	56,3	57,5	54,7	52,6	53,7	51,6	55,3	51,5	49,0	43,0	43,0	44,6
2. A.majus_TA4522_4151	66,7	62,9	63,9	64,9	64,2	65,3	66,0	69,1	64,9	57,0	56,4	55,4	58,8
3. S.lycopersicum_TA46686	60,2	62,5	63,6	60,0	62,1	62,1	52,6	56,4	58,6	56,0	52,0	50,0	54,5
4.N.benthamiana_EH366122	64,3	67,8	69,0	56,4	55,9	57,0	53,2	58,1	60,2	55,0	53,0	51,0	56,4
5. G.hirsutum_DR461014	60,2	69,0	70,1	60,4	65,6	64,5	58,1	64,1	59,8	55,0	54,0	50,0	55,4
6. V.vinifera_EE072107	71,3	68,3	69,3	59,8	63,7	64,7	62,4	69,3	66,3	68,3	64,4	63,4	68,6
7. M.domestica_DR994824	74,5	70,6	71,6	65,0	71,8	69,9	61,8	69,6	69,2	70,6	66,7	63,7	70,6
8.G.hirsutum_TA31277_363568,4	79,3	80,5	58,9	69,1	68,1	57,9	64,9	62,6	61,0	60,0	54,0	59,4	
9. G.hirsutum_DW503354	68,4	78,2	79,3	55,8	65,3	66,3	57,9	63,8	65,7	63,4	57,4	55,4	57,4
10. H.exilis_EE652824	67,3	74,7	73,6	58,2	55,9	57,0	51,6	58,7	61,9	55,0	52,0	49,0	54,5
11. E.gunnii_CT985658	73,5	83,9	85,1	60,6	65,6	64,5	58,5	67,7	68,4	65,3	56,4	55,4	60,4
12. L.sativa_TA7946_4236	71,4	79,5	80,7	58,9	62,1	61,1	55,8	64,9	64,6	62,4	55,0	53,5	59,4
13. C.intybus_EH706852	72,4	79,3	80,5	60,0	63,2	62,1	55,8	63,8	64,6	63,0	56,0	54,0	59,4
14. P.trichocarpa_sc_IV.1184		80,6	81,6	58,6	59,6	58,6	57,1	65,3	63,6	64,0	59,0	56,0	63,4
15. P.trichocarpa_557252	85,7		98,9	57,9	64,9	64,9	57,9	68,1	65,7	63,0	62,0	59,0	65,3
16. P.euphratica_AJ768574	86,7	98,9		58,9	66,0	64,9	58,9	67,0	64,6	64,0	61,0	58,0	64,4
17. M.truncatula_AC135566	74,5	71,4	72,5		80,6	79,6	72,3	72,0	73,5	58,4	60,4	60,4	59,8
18.G.max_Glyma06g46590.175,5	75,5	76,1	77,2	88,0		92,4	74,5	76,3	76,5	67,3	66,3	62,4	64,7
19.G.max_Glyma04g16390.176,5	76,5	78,3	78,3	90,2	96,7		77,7	77,4	75,5	67,3	66,3	62,4	65,7
20.G.max_Glyma02g18210.175,5	75,5	73,1	74,2	86,0	89,2	89,2		76,6	72,2	61,0	59,0	60,0	61,4
21. M.truncatula_AC157502	80,6	80,4	79,3	83,7	88,0	89,1	86,0		79,4	66,0	66,0	64,0	65,3
22.G.max_Glyma03g28050.179,6	179,6	77,3	76,3	82,5	87,6	89,7	85,6	90,7		65,7	62,7	61,8	64,1
23. A.thaliana_At4g39250	78,0	74,0	75,0	75,0	81,0	81,0	81,0	77,0	80,0		75,0	72,0	73,3
24. T.salsuginea_DN773374	77,0	74,0	73,0	77,0	80,0	82,0	78,0	79,0	81,0	87,0		88,0	89,1
25. B.napus_BN06MC20309	74,0	71,0	70,0	75,0	77,0	79,0	75,0	76,0	78,0	84,0	97,0		83,2
26. A.thaliana_At2g21650	76,2	74,3	73,3	73,3	77,2	80,2	78,2	78,2	80,2	84,2	94,1	91,1	

Ejemplo 4: Identificación de los dominios comprendidos en las secuencias de polipéptidos útiles en la realización de los procedimientos de la invención

5 La base de datos de fuente integrada de familias, dominios y sitios de proteínas (InterPro) es una interfaz integrada para las bases de datos características comúnmente usadas para las búsquedas basadas en secuencias y en texto. La base de datos InterPro combina estas bases de datos, que usan diferentes metodologías y diversos grados de información biológica a cerca de proteínas bien caracterizadas para obtener proteínas características. Las bases de datos que colaboran incluyen SWISS-PROT, PROSITE, TrEMBL, PRINTS, ProDom y Pfam, Smart y TIGRFAM. 10 Pfam es una gran colección de múltiples alineaciones de secuencia y modelos ocultos de Markov que cubren muchos dominios y familias de proteínas comunes. Pfam se aloja en el servidor del Instituto Sanger, en Reino Unido. Interpro se aloja en el Instituto de Bioinformática europeo, en Reino Unido.

1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

Los resultados de la exploración con InterPro de la secuencia polipeptídica representada por SEC ID N°: 2 se presentan en la Tabla C1.

15 **Tabla C1:** Resultados de la exploración con InterPro (números de acceso principales) de la secuencia polipeptídica representada por SEC ID N°: 2.

Procedimiento	Número de acceso	Nombre corto	Puntuación	Ubicación
INTERPRO	IPR001005	SANT, de unión al ADN		
SMART	SM00717	SANT	7,5E-9	[10-62]T
PROFILE	PS50090	MYB_3	0,0	[6-60]T

(Continuación)

Procedimiento	Número de acceso	Nombre corto	Puntuación	Ubicación
INTERPRO	IPR009057	de tipo homeodominio		
SUPERFAMILY	SSF46689	de tipo_homeodominio	7,0E-13	[13-67]T
INTERPRO	IPR014778	Myb, de unión al ADN		
PFAM	PF00249	Myb_de unión al ADN	4,8E-4	[11-60]T
INTERPRO	IPR015609	Chaperona molecular, proteína de choque térmico, Hsp40, DnaJ		
PANTHER	PTHR11821	Hsp40/DnaJ_Rel	9,0E-8	[13-59]T
INTERPRO	noIPR	Sin integrar		
PANTHER	PTHR11821:SF29	PTHR11821:SF29	9,0E-8	[13-59]T

Ejemplo 5: Predicción de la topología de las secuencias polipeptídicas útiles en la realización de los procedimientos de la invención

5 TargetP 1.1 predice la ubicación subcelular de las proteínas eucariotas. La asignación de ubicación se basa en la presencia pedicha de cualquiera de las presecuencias del extremo N: péptido transitorio de cloroplasto (cTP), péptido de dirección mitocondrial (mTP) o péptido señal de vía secretoria (SP). Las puntuaciones en las que se basa la predicción final no son realmente probabilidades, y no se agregan necesariamente a una. Sin embargo, la ubicación con la puntuación más alta es más probable de acuerdo con TargetP, y la relación entre las puntuaciones (la clase de confiabilidad) puede indicar cómo de cierta es la predicción. La clase de confiabilidad (RC) varía de 1 a 10 5, donde 1 indica la predicción más fiable. TargetP se mantiene en el servidor de la Universidad Técnica de Dinamarca.

Para las secuencias predichas que contienen una presecuencia N-terminal, también se puede predecir un posible sitio de escisión.

15 Se seleccionaron una serie de parámetros, tales como grupo de organismo (no vegetal o vegetal), grupos de corte (ninguno, grupo predefinido de cortes o grupo de cortes especificado por el usuario) y el cálculo de la predicción de los sitios de escisión (si o no).

1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

20 Los resultados del análisis TargetP1.1 de la secuencia polipeptídica representada por SEC ID N° 2 se presentan en la Tabla D1. Se ha seleccionado el grupo de organismo "vegetal", no se han definido los cortes y se ha solicitado la longitud predicha del péptido transitorio. La ubicación subcelular de la secuencia polipeptídica representada por SEC ID N°: 2 puede ser el citoplasma o el núcleo, y no se predice el péptido transitorio.

25 **Tabla D1:** Análisis TargetP 1.1 D de la secuencia polipeptídica representada por SEC ID N° 2. Abreviaturas: Len, longitud, cTP, péptido transitorio de cloroplasto; mTP, péptido de tránsito mitocondrial; SP, péptido señal de vía secretora; otro, otra dirección subcelular; Loc, ubicación predicha; RC, clase de confiabilidad; Tplen, péptido de tránsito predicho.

Nombre	Len	cTP	mTP	SP	otro	Loc	RC	Tplen
SEC ID N° 2	88	0,505	0,093	0,089	0,548	-	5	-
Corte		0,000	0,000	0,000	0,000			

Cuando se usa el algoritmo de predicción de la ubicación de proteínas (PLOC, Park y Kanehisa, *Bioinformatics*, 19, 1656-1663, 2003), se predice que SEC ID N° 2 tiene una ubicación nuclear, que coincide con su función postulada como factor de transcripción.

Se pueden usar muchos otros algoritmos para realizar dichos análisis, que incluyen:

- 30
- ChloroP 1.1 alojado en el servidor de la Universidad Técnica de Dinamarca;
 - Predictor de la ubicación subcelular de proteína Prowler, versión 1.2, alojado en el servidor del Instituto de Biociencia Molecular, Universidad de Queensland, Brisbane, Australia;
 - Analista de Proteomas PENCE PA-GOSUB 2.5, alojado en el servidor de la Universidad de Alberta, Edmonton, Alberta, Canadá;
- 35
- TMHMM, alojado en el servidor de la Universidad Técnica de Dinamarca;
 - PSORT (URL: psort.org).

Ejemplo 6: Ensayo relacionado con las secuencias polipeptídicas útiles en la realización de los procedimientos de la invención**1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)**

5 Los ensayos de unión de ADN son muy conocidos en la técnica, incluyendo la selección de sitios de unión a ADN asistida por PCR y un ensayo de cambio en gel de unión a DNA. Para una referencia general, véase "Current Protocols in Molecular Biology", Volúmenes 1 y 2, Ausubel *et al.* (1994), Current Protocols.

10 En Li y Parish (*Plant J.* 8, 963-972, 1995), se describe un ensayo de unión a ADN para MYB7, como ejemplo para proteínas MYB en general. En resumen, se clona la secuencia codificante de MYB7 en marco con la secuencia líder 10 del gen de T7 y se expresa en *E. coli*. Las proteínas se purifican y se analizan en un ensayo de retardo de la movilidad, usando el sitio de unión a c-myb marcado con 32P (MBS) y el sitio de unión del producto génico P del maíz (PBS). De esta manera, se demostró que MYB7 de *Arabidopsis thaliana* se unió al sitio MBS, aunque no con alta afinidad, y tuvo una preferencia de unión por el sitio de PBS.

Ejemplo 7: Clonación de la secuencia de ácido nucleico usada en los procedimientos de la invención**1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)**

15 Se amplificó la secuencia de ácido nucleico por PCR usando como plantilla una biblioteca de ADNc de plántulas de *Solanum lycopersicum* diseñada a medida (en pCMV Sport 6.0; Invitrogen, Paisley, RU). Se realizó la PCR usando Taq ADN polimerasa Hifi en condiciones convencionales, usando 200 ng de plantilla en una mezcla de PCR de 50 µl. Los cebadores usados fueron prm10562 (SEC ID N°: 290; sentido, codón de inicio en negrita): 5'-ggggacaagttgtacaataaaagcaggcttaacaatgagctcgatgctgctca-3' y prm10563 (SEC ID N°: 291; inverso, complementario): 5'-ggggaccacttgtacaagaaagctggg taagtgttaggtttgctcctt-3', que incluyen los sitios AttB para la recombinación Gateway. El fragmento de PCR amplificado se purificó usando también procedimientos convencionales. A continuación, se realizó la primera etapa del procedimiento Gateway, la reacción BP, durante el cual el fragmento de PCR se recombinó *in vivo* con el plásmido pDONR201 para producir, de acuerdo con la terminología Gateway, un "clon de entrada", de tipo FSM1. El plásmido pDONR201 se adquirió en Invitrogen, como parte de tecnología Gateway®.

25 Entonces, se usó el clon de entrada que comprende la SEC ID N°: 1 en una reacción LR con un vector de destino usado para la transformación de *Oryza sativa*. Este vector contenía como elementos funcionales dentro de los límites del ADN-T: un marcador seleccionable vegetal; un casete de expresión de marcador detectable; y un casete Gateway destinado a la recombinación LR *in vivo* con la secuencia de ácido nucleico de interés ya clonada en el clon de entrada. Cadena arriba de este casete Gateway, se encuentra un promotor GOS2 del arroz (SEC ID N°: 289) para la expresión específica constitutiva.

30 Tras la etapa de recombinación LR, el vector de expresión resultante pGOS2::de tipo FSM1 (Figura 4) se transformó en la cepa LBA4044 de *Agrobacterium* de acuerdo con procedimientos muy conocidos en la técnica.

Ejemplo 8: Transformación de plantas**35 Transformación del arroz**

La agrobacteria que contenía el vector de expresión se usó para transformar plantas de *Oryza sativa*. Se descascarillaron semillas secas maduras de la variedad cultivada Nipponbare de arroz *iaponica*. Se llevó a cabo la esterilización mediante la incubación durante un minuto en etanol al 70 %, seguido por 30 minutos en HgCl₂ al 0,2 %, seguido por un lavado de 6 veces durante 15 minutos con agua destilada estéril. Luego se germinaron las semillas estériles en un medio que contenía 2,4-D (medio de inducción de callos). Tras la incubación en la oscuridad durante cuatro semanas, se cortaron los callos derivados del escutelo, embriogénicos, y se propagaron en el mismo medio. Tras dos semanas, se multiplicaron o propagaron los callos mediante subcultivo en el mismo medio durante otras 2 semanas. Se subcultivaron pedazos de callo embriogénico en medio fresco 3 días antes del cocultivo (para reforzar la actividad de división celular).

45 Para el cocultivo, se usó la cepa LBA4404 de *Agrobacterium* que contenía el vector de expresión. Se inoculó *Agrobacterium* en medio AB con los antibióticos apropiados y se cultivó durante 3 días a 28 °C. Luego se recogieron las bacterias y se suspenden en medio de cocultivo líquido a una densidad (DO₆₀₀) de aproximadamente 1. A continuación, se transfirió la suspensión a una placa Petri y se sumergieron los callos en la suspensión durante 15 minutos. Entonces, se transfirieron los tejidos de callo secos sobre un papel de filtro y se transfirieron a medio de cocultivo solidificado, y se incubaron durante 3 días a oscuras a 25 °C. Se sembraron los callos cocultivados en medio que contenía 2,4-D durante 4 semanas a oscuras a 28 °C en presencia de un agente de selección. Durante este período, se desarrollaron rápidamente islotes de callos resistentes al crecimiento. Después de la transferencia de este material a un medio de regeneración y de la incubación a la luz, se liberó el potencial embriogénico y se desarrollaron brotes en las cuatro a cinco semanas siguientes. Se extirparon los brotes de los callos y se incubaron durante 2 a 3 semanas en un medio que contenía auxina del que se transfirieron al suelo. Los brotes endurecidos se cultivan bajo alta humedad y días cortos en un invernadero.

Se generaron aproximadamente 35 transformantes de arroz T0 independientes para una construcción. Los transformantes primarios se transfirieron desde una cámara de cultivo de tejido a un invernadero. Después de un análisis PCR cuantitativo para verificar el número de copias del inserto de ADN-T, solo se mantuvieron plantas transgénicas de una sola copia que presentaban tolerancia al agente de selección para la cosecha de la semilla T1. Luego se cosecharon las semillas de tres a cinco meses después del trasplante. El procedimiento produjo transformantes de un solo locus a una tasa superior al 50 % (Aldemita y Hodges 1999, Chan *et al.* 1993, Hiei *et al.* 1994).

Ejemplo 9: Transformación de otros cultivos

Transformación del maíz

La transformación del maíz (*Zea mays*) se realiza con una modificación del procedimiento descrito por Ishida *et al.* (1996) *Nature Biotech* 14(6): 745-50. La transformación es dependiente del genotipo en el maíz y solo genotipos específicos son susceptibles a la transformación y la regeneración. La línea endogámica A188 (Universidad de Minnesota) o los híbridos con A188 como progenitor son buenas fuentes de material donante para la transformación, pero también se pueden usar otros genotipos satisfactoriamente. Se cosechan espigas de plantas de maíz aproximadamente 11 días después de la polinización (DAP), cuando la longitud del embrión inmaduro es de aproximadamente 1 a 1,2 mm. Se cocultivan embriones inmaduros con *Agrobacterium tumefaciens* que contiene el vector de expresión, y las plantas transgénicas se recuperan a través de organogenia. Se cultivan los embriones cortados en medio de inducción de callos y luego el medio de regeneración del maíz, que contiene el agente de selección (por ejemplo, imidazolinona, pero se pueden usar diversos marcadores de selección). Se incuban las placas Petri a la luz a 25 °C durante 2-3 semanas, o hasta que se desarrollan brotes. Los brotes verdes se transfieren desde cada embrión al medio de enraizamiento del maíz y se incuban a 25 °C durante 2-3 semanas, hasta que se desarrollan raíces. Se trasplantan los brotes enraizados al suelo en el invernadero. Se producen semillas T1 de plantas que presentan tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.

Transformación del trigo

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

Transformación de la soja

Se transforma soja de acuerdo con una modificación del procedimiento descrito en la patente de EE.UU. N° 5.164.310 DE Texas A&M. Diversas variedades comerciales de soja son susceptibles a la transformación mediante este procedimiento. La variedad cultivada Jack (disponible de la fundación Seed de Illinois) se usa comúnmente para la transformación. Se esterilizan las semillas de soja para la siembra *in vitro*. Se extirpan el hipocotilo, el radículo y un cotiledón de plántulas jóvenes de siete días de vida. El epicotilo y el cotiledón restante se cultivan además para desarrollar nodos axilares. Estos nodos axilares se extirpan y se incuban con *Acrobacterium tumefaciens* que contiene el vector de expresión. Después de tratamiento del cocultivo, se lavan los explantes y se transfieren al medio de selección. Se extirpan los brotes regenerados y se ponen en un medio de elongación de brotes. Se colocan los brotes no superiores a 1 cm en medio de enraizamiento hasta que se desarrollan raíces. Las raíces enraizadas se trasplantan al suelo en el invernadero. Se producen semillas T1 de plantas que muestran tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.

Transformación de la colza

Se usan los pecíolos cotiledonarios y los hipocotilos de plántulas jóvenes de 5-6 días de vida como explantes para el cultivo del tejido y se transforman de acuerdo con Babic *et al.* (1998, *Plant cell Rep* 17: 183-188). La variedad cultivada comercial Westar (Agriculture Canada) es la variedad convencional usada para la transformación, pero también se pueden usar otras variedades. Se esterilizan superficialmente las semillas de colza para la siembra *in vitro*. Se extirpan los explantes de pecíolos cotiledones con el cotiledón unido de las plántulas *in vitro*, y se inoculan con *Agrobacterium* (que contiene el vector de expresión) sumergiendo el extremo cortado del explante de pecíolo en la suspensión bacteriana. A continuación, se cultivan los explantes durante 2 días en medio MSBAP-3 que contiene 3 mg/l de BAP, sacarosa al 3 %, Phytagar al 0,7 % a 23 °C, 16 h de luz. Después de dos días de cocultivo con *Agrobacterium*, se transfieren los explantes de pecíolo a medio MSBAP-3 que contiene 3 mg/l de BAP, cefotaxima, carbenicilina o timentina (300 mg/l) durante 7 días, y luego se cultivan en medio MSBAP-3 con cefotaxima,

carbenicilina o timentina y agente selección hasta la regeneración de los brotes. Cuando los brotes tienen una longitud de 5-10 mm, se cortan y se transfieren al medio de elongación de brotes (MSBAP-0.5, que contiene 0,5 mo/l de BAP). Los brotes de aproximadamente 2 cm de longitud se transfieren al medio de enraizamiento (MS0) para la inducción de las raíces. Los brotes enraizados se trasplantan al suelo en el invernadero. Se producen semillas T1 de plantas que muestran tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.

Transformación de la alfalfa

Se transforma un clon de regeneración de alfalfa (*Medicago sativa*) usando el procedimiento de (McKersie *et al.*, 1999 *Plant Physiol* 119: 839-847). La regeneración y transformación de la alfalfa es dependiente del genotipo y, por lo tanto, se requiere una planta regenerada. Se han descrito procedimientos para obtener plantas regenerantes. Por ejemplo, estos se pueden seleccionar del cultivo Ranglander (Agriculture Canada) o cualquier otra variedad comercial de alfalfa como según lo descrito por Brown DCW y A Atanassov (1985. *Plant Cell Tissue Organ Culture* 4: 111-112). Como alternativa, la variedad RA3 (Universidad de Wisconsin) se ha seleccionado para su uso en el cultivo de tejido (Walker *et al.*, 1978 *Am J Bot* 65:654-659). Se cocultivan durante una noche los explantes de pecíolo con un cultivo de *Agrobacterium tumefaciens* C58C1 pMP90 (McKersie *et al.*, 1999 *Plant Physiol* 119: 839-847) o LBA4404 que contiene el vector de expresión. Se cocultivan los explantes durante 3 d a oscuras en medio de inducción SH que 288 mg/l de Pro, 53 mg/l de tioprolina, 4,35 g/l de K₂SO₄, y 100 µm de acetosiringinona. Se lavan los explantes en medio Murashige-Skoog de resistencia media (Murashige y Skoog, 1962) y se ponen siembran en el mismo medio de inducción SH sin acetosiringinona, pero con un agente de selección adecuado y antibiótico adecuado para inhibir el crecimiento de *Agrobacterium*. Tras varias semanas, se transfieren embriones somáticos al medio de desarrollo BOi2Y que no contiene reguladores del crecimiento, sin antibióticos, y 50 g/l de sacarosa. Posteriormente, se germinan los embriones somáticos en medio Murashige-Skoog de resistencia media. Se trasplantan plántulas enraizadas en macetas y se cultivan en un invernadero. Se producen semillas T1 de plantas que presentan tolerancia al agente de selección y que contienen una sola copia del inserto de ADN-T.

Transformación del algodón

El algodón se transforma usando *Agrobacterium tumefaciens* de acuerdo con el método descrito en el documento US 5.159.135. Se esterilizan las semillas de algodón superficialmente en solución de hipoclorito sódico al 3 % durante 20 minutos y se lavan en agua destilada con 500 µg/ml de cefotaxima. A continuación, se transfieren las semillas a medio SH con 50 µg/ml de benomil para la germinación. Se retiran los hipocotilos de las plántulas de 4 a 6 días de vida, se cortan en trozos de 0,5 cm y se colocan en agar al 0,8 %. Se usa una suspensión de *Agrobacterium* (aprox. 10⁸ células por ml, diluidos a partir de un cultivo de una noche transformado con el gen de interés y marcadores de selección adecuados) para la inoculación de los explantes de hipocótilo. Después de 3 días a temperatura e iluminación ambiente, se transfieren los tejidos a un medio sólido (1,6 g/l de Gelrite) con sales de Murashige y Skoog con vitaminas B5 (Gamborg *et al.*, *Exp. Cell Res.* 50: 151-158 (1968)), 0,1 mg/l de 2,4-D, 0,1 mg/l de 6-furfurilaminopurina y 750 µg/ml de MgCl₂, y con 50 a 100 µg/ml de cefotaxima y 400-500 µg/ml de carbenicilina para matar las bacterias residuales. Se aíslan líneas celulares individuales después de dos a tres meses (con subcultivos cada cuatro a seis semanas) y se cultivan adicionalmente en un medio selectivo para la amplificación de tejido (30 °C, fotoperíodo de 16 h). Después, se cultivan los tejidos transformados más en medio no selectivo durante 2 a 3 meses, dando lugar a embriones somáticos. Se transfieren los embriones de aspecto saludable de al menos 4 mm de longitud a tubos con medio SH en vermiculita fina, suplementado con 0,1 mg/l de ácido indolacético, 6 furfurilaminopurina y ácido giberélico. Se cultivan los embriones a 30 °C con un fotoperíodo de 16 horas, y se transfieren las plántulas en la fase folicular 2 a 3 a macetas con vermiculita y nutrientes. Se endurecen las plantas y, posteriormente, se trasladan al invernadero para su posterior cultivo.

Ejemplo 10: Procedimiento de evaluación fenotípica

10.1 Configuración de la evaluación

Se generan aproximadamente 35 transformantes de arroz T0 independientes. Los transformantes primarios se transfieren desde una cámara de cultivo de tejido a un invernadero para la siembra y cosecha de la semilla T1. Se retienen seis sucesos, de los que la progenie T1 segregó 3:1 para la presencia/ausencia del transgén. Para cada uno de estos sucesos, se seleccionan aproximadamente 10 plántulas T1 que contienen el transgén (heterocigotos y homocigotos) y aproximadamente 10 plántulas T1 que carecen del transgén (nulicigotos) mediante la monitorización de la expresión visual del marcador. Se cultivan las plantas transgénicas y los nulicigotos correspondientes uno al lado de los otros en posiciones aleatorias. Las condiciones del invernadero son días cortos (12 horas de luz), 28 °C a la luz y 22 °C en la oscuridad, y una humedad relativa del 70 %. Las plantas que se cultivan en condiciones sin estrés se hidratan en intervalos reguladores para asegurar que el agua y los nutrientes no se reduzcan y satisfacer las necesidades de la planta para completar el crecimiento y el desarrollo.

Se evaluaron además cuatro sucesos T1 en la generación T2 siguiendo el mismo procedimiento de evaluación que para la generación T1, pero con más individuos por suceso. Desde la etapa de siembra hasta la edad de madurez, las plantas se pasan varias veces a través de una cabina de formación de imágenes digital. En cada punto temporal se toman imágenes digitales (2.048 x 1.536 píxeles, 16 millones de colores) de cada planta desde al menos 6 ángulos diferentes.

Detección de la sequía

Se cultivaron plantas a partir de semillas T2 en suelo de maceta en condiciones normales hasta que alcanzaron la etapa de partida. Luego se transfirieron a una sección "seca", donde se retuvo la irrigación. Se insertaron sondas de humedad en macetas seleccionadas aleatoriamente para monitorizar el contenido de agua del suelo (SWC). Cuando el SWC se redujo por debajo de ciertos umbrales, las plantas se rehidrataron automáticamente de manera continua hasta que se alcanzó de nuevo un nivel normal. Entonces, se transfirieron las plantas de nuevo a condiciones normales. El resto del cultivo (maduración de la planta, cosecha de las semillas) es igual que para las plantas que no se cultivan en condiciones de estrés abiótico. Los parámetros de cultivo y producción se registran como se detalla para el cultivo en condiciones normales.

10 *Detección de la eficacia de uso de nitrógeno*

Se cultivan plantas de arroz a partir de semillas T2 en suelo de maceta en condiciones normales, excepto para la solución de nutrientes. Estas macetas se hidratan desde el trasplante hasta la maduración con una solución de nutrientes específica que tiene un contenido reducido de nitrógeno N (N), normalmente entre 7 a 8 veces inferior. El resto del cultivo (maduración de la planta, cosecha de las semillas) es igual que para las plantas que no se cultivan en condiciones de estrés abiótico. Los parámetros de cultivo y producción se registran como se detalla para el cultivo en condiciones normales.

Detección de estrés salino

Se cultivan plantas en un sustrato de fibras de coco y argex (proporción 3 a 1). Se usa una solución de nutrientes normal durante las dos primeras semanas después del trasplante de las plántulas en el invernadero. Después de las dos primeras semanas, se añaden 25 mM de sal (NaCl) a la solución de nutrientes, hasta que las plantas se cosechan. Luego se miden los parámetros relacionados con las semillas.

10.2 Análisis estadístico: prueba F

Se usó ANOVA de dos factores (análisis de variantes) como un modelo estadístico para la evaluación general de las características fenotípicas de la planta. Se llevó a cabo una prueba F en todos los parámetros medidos de todas las plantas de todos los sucesos transformados con el gen de la presente invención. La prueba F se llevó a cabo para comprobar un efecto del gen sobre todos los sucesos de transformación y para verificar un efecto general del gen, también conocido como efecto global del gen. El umbral de significancia para un efecto global del gen verdadero se establece a un nivel de probabilidad del 5 % para la prueba F. Los puntos de valor de la prueba F significativos para un efecto de gen, significa que solo la mera presencia o posición del gen es lo que provoca las diferencias en el fenotipo.

Cuando se llevaron a cabo dos experimentos con sucesos superpuestos, se realizó un análisis combinado. Esto es útil para comprobar la consistencia de los efectos sobre los dos experimentos, y si se da el caso, para acumular pruebas de ambos experimentos con el fin de aumentar la confianza en la conclusión. El método usado fue una metodología de modelo mixto que tiene en cuenta la estructura multinivel de los datos (es decir, el experimento - suceso - segregantes). Los valores de p se obtuvieron mediante la comparación de la prueba de razón de verosimilitud con las distribuciones chi cuadrado

10.3 Parámetros medidos

Medición de los parámetros relacionados con biomasa

A partir de la etapa de siembra hasta la etapa de madurez, se pasaron las plantas varias veces a través de una cabina de formación de imágenes digital. Se tomaron imágenes digitales en cada punto temporal (2.048 x 1.536 píxeles, 16 millones de colores) de cada planta desde al menos 6 ángulos diferentes.

Se determinó la superficie sobre el suelo de la planta (o biomasa frondosa) contando el número total de píxeles de las imágenes digitales de las partes de planta situadas sobre el suelo discriminadas del fondo. Se calculó la media de dicho valor para las fotografías tomadas en el mismo punto temporal desde los diferentes ángulos y se convirtió en un valor de superficie física expresado en mm cuadrados por calibración. Los experimentos muestran que la superficie de la planta sobre el suelo medida de esta forma se correlaciona con la biomasa de partes de planta por encima del suelo. La superficie por encima del suelo es la superficie medida en el punto temporal en el que la planta alcanza su biomasa frondosa máxima. El vigor temprano es la superficie encima del suelo de la planta (plántula) tres semanas después de la germinación. El aumento en el biomasa radicular se expresa como un aumento de la biomasa radicular total (medido como la biomasa radicular máxima observada durante la vida útil de una planta (RootMax) o como la biomasa radicular máxima con un espesor por encima de cierto umbral observado durante la vida útil de una planta (RootThickMax); o como un aumento en el índice de raíz/brote (medido como la proporción entre la masa radicular y la masa de brotes en el período de crecimiento activo de raíz y brote).

Se determinó el vigor temprano contando el número total de píxeles de las partes de la planta situadas por encima del suelo discriminadas del fondo. Se calculó la media de dicho valor para las fotografías tomadas en el mismo punto

temporal desde los diferentes ángulos y se convirtió en un valor de superficie física expresado en mm cuadrados por calibración. Los resultados descritos más adelante son para las plantas tres semanas después de la germinación.

Mediciones de los parámetros relacionados con las semillas

5 Se cosecharon las panículas primarias maduras, se contaron, se embolsaron, se marcaron con código de barras y luego se secaron durante tres días en un horno a 37 °C. Luego se desgranaron las panículas, y se recogieron y se contaron todas las semillas. Se separaron las cáscaras llenas de las vacías usando un dispositivo de soplado de aire. Se desecharon las cáscaras vacías y se volvió a contar la fracción restante. Se pesaron las cáscaras llenas en una balanza de análisis. Se determinó el número de semillas llenas contando el número de cáscaras llenas que quedaba después de la etapa de separación. Se midió la producción total de semillas pesando todas las cáscaras llenas cosechadas de una planta. Se midió el número total de semillas por planta contando el número de cáscaras cosechadas por planta. Se extrapolo el peso de mil granos (TKW) del número de semillas llenas contadas y su peso total. En la presente invención, el índice de cosecha (HI) se define como la proporción entre la producción total de semillas y la superficie por encima del suelo (mm²), multiplicada por un factor 10⁶. El número total de flores por panícula, como se define en la presente invención, es la proporción entre el número total de semillas y el número de panículas primarias maduras. El índice de llenado de las semillas, como se define en la presente invención, es la proporción (expresada como un %) del número de semillas llenas con respecto al número total de semillas (o floretes).

Ejemplo 11: Resultados de la evaluación fenotípica de las plantas transgénicas

1. Polipéptidos de tipo FSM1 (Fruto Sant/Myb)

20 A continuación, se presentan los resultados de la evaluación de plantas de arroz transgénicas que expresan SEC ID N°: 1 en condiciones sin estrés. Las plantas T1 presentaron un aumento de la producción de semillas y un aumento de la biomasa radicular, y estas observaciones se confirmaron en una segunda evaluación con plantas T2. Las Tablas E1 y E2 enumeran los parámetros para los que el aumento global fue de al menos el 5 %.

25 **Tabla E1:** Resumen de datos para plantas de arroz transgénicas de la generación T1; para cada parámetro, se muestra el porcentaje de aumento global, siendo para cada parámetro el valor $p < 0,05$.

Parámetro	Global
Filtrado	19,9
RootThickMax	6,5

Tabla E2: Resumen de datos para plantas de arroz transgénicas de la generación T2; para cada parámetro, se muestra el porcentaje de aumento global, siendo para cada parámetro el valor $p < 0,05$.

Parámetro	Global
RootMax	7,5
Filtrado	12,5

30 Por otra parte, las plantas T1 tuvieron un aumento del peso total de las semillas (aumento global +17,1 %, $p < 0,05$), un aumento del índice de cosecha (aumento global +16,5 %, $p < 0,05$), aumento del número de semillas llenas (aumento global +17,9 %, $p < 0,05$) y un aumento del número de flores por panícula (aumento global +10,6 %, $p < 0,05$). Además, una de las líneas ensayadas en ambas generaciones T1 y T2 mostró un aumento del vigor temprano.

REIVINDICACIONES

1. Un procedimiento de aumento del rendimiento de la biomasa y/o de las semillas en plantas en comparación con plantas de control, que comprende la modulación de la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1, en el que dicho polipéptido de tipo FSM1 está representado por SEC ID N° 2 o uno de sus homólogos que tiene al menos un 50 % de identidad de secuencia con SEC ID N° 2, y que comprende además un dominio SANT (acceso SMART SM0017) y uno o más de los siguientes motivos:
- 5 (i) Motivo 1:
W[TS][PA]K[QE]NK[LA]FE[RK]ALAVYD[KR][DE]TPDRW[HSQ]N[VI]A[RK]A (SEC ID N°: 283),
- 10 (ii) Motivo 2: GGK[ST][AV][ED]EV[KR]RHYE[IL]L (SEC ID N°: 284),
(iii) Motivo 3: D[VL][KF][HF][ED][SN]G[RM]VPFP[NK]Y (SEC ID N°: 285),
- en el que dicho aumento del rendimiento de la biomasa y/o de las semillas se obtiene en condiciones sin estrés, en el que dicho ácido nucleico está unido operativamente al promotor GOS2 del arroz y en el que dicha expresión modulada se efectúa mediante la introducción y la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.
- 15 2. Procedimiento de acuerdo con la reivindicación 1, en el que dicho ácido nucleico que codifica un polipéptido de tipo FSM1 codifica una cualquiera de las proteínas enumeradas en la Tabla A1 o es una parte de dicho ácido nucleico, o un ácido nucleico capaz de hibridarse con dicho ácido nucleico.
3. Procedimiento de acuerdo con la reivindicación 1 o 2, en el que dicha secuencia de ácido nucleico codifica un ortólogo o parálogo de cualquiera de las proteínas dadas en la Tabla A1.
- 20 4. Procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que dicho ácido nucleico que codifica un polipéptido de tipo FSM1 es de origen vegetal, preferentemente, de una planta dicotiledónea, más preferentemente, de la familia *Solanaceae*, más preferentemente del género *Lycopersicon*, y lo más preferentemente de *Solanum lycopersicum*.
5. Construcción que comprende:
- 25 (i) ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en la reivindicación 1;
(ii) una o más secuencias de control capaces de dirigir la expresión de la secuencia de ácido nucleico de (a), siendo una de dichas secuencias de control el promotor GOS2 del arroz y, opcionalmente,
(iii) una secuencia de terminación de la transcripción.
- 30 6. Uso de una construcción de acuerdo con la reivindicación 5 en un procedimiento de producción de plantas que tienen un mayor rendimiento, particularmente, un mayor rendimiento de la biomasa y/o un mayor rendimiento de las semillas en comparación con plantas de control.
7. Planta, parte de planta o célula vegetal transformada con una construcción de acuerdo con la reivindicación 5.
8. Procedimiento de producción de una planta transgénica que tiene un mayor rendimiento, particularmente, un mayor rendimiento de la biomasa y/o un mayor rendimiento de las semillas en comparación con plantas de control, que comprende:
- 35 (i) introducción y expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en la reivindicación 1 unido operativamente al promotor GOS2 del arroz; y
(ii) cultivo de la célula vegetal en condiciones que potencian el crecimiento y el desarrollo vegetal.
- 40 9. Planta transgénica que tiene un mayor rendimiento, particularmente, un mayor rendimiento de la biomasa y/o de las semillas, cuando se cultiva en condiciones sin estrés y en comparación con plantas de control, que procede de la expresión modulada de un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en la reivindicación 1 unido operativamente al promotor GOS2 del arroz, o una célula de planta transgénica derivada de dicha planta transgénica, en la que dicha expresión modulada se realiza mediante la introducción y la expresión en una planta de un ácido nucleico que codifica un polipéptido de tipo FSM1.
- 45 10. Planta transgénica de acuerdo con la reivindicación 7 o 9, o una célula de planta transgénica derivada de la misma, siendo dicha planta una planta de cultivo tal como remolacha azucarera, o una planta monocotiledónea tal como caña de azúcar, o un cereal tal como arroz, maíz, trigo, cebada, mijo, centeno, triticale, sorgo, escanda, espelta, secale, escanda menor, teff, milo y avena.
- 50 11. Partes cosechables de una planta de acuerdo con la reivindicación 10, siendo dichas partes cosechables preferentemente biomasa de los brotes y/o de las semillas, comprendiendo dichas partes cosechables un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en la reivindicación 1 unido operativamente con el promotor GOS2 del arroz.

12. Uso de un ácido nucleico que codifica un polipéptido de tipo FSM1 según lo definido en la reivindicación 1 unido operativamente con el promotor GOS2 del arroz para aumentar el rendimiento de las semillas y/o de la biomasa en plantas cultivadas en condiciones sin estrés en comparación con plantas de control.


```

M.guttatus_DV207881      VEEVKKHYEVLLEDFNKIESGKVPLPNYKTTGCG----SF-----
A.majus_TA4522_4151      PEEVKKHYEILVEDIKYIESGKVPPFPNYRTTGG----NMKTDEKRFRNLK
V.vinifera_EE072107      VEEVKKRHYEILVEDIKSIDSDKVPPFPNYKTTGASGRSNMSDQEKRMNLK
G.hirsutum_TA31277_3635  AEEVKKRHYELLVADVKEYIESGQVPPFP--YRSNGN-----
G.hirsutum_DW503354      AEEVKKRHYELLVKDVKQIESGKVPPFPNYRTTGGSNTOG-----
H.exilis_EE652824        AEEVKKRHYEILVGDVKLIESGRVPPFPNYRTTN--GA-----
E.gunnii_CT985658        AEEVKKRHYELLVEDVKHIESGRVPPFPNYRTTGR-GNVNG-----
P.trichocarpa_557252     AEEVKKRHYELLVEDVKHIESGHVPPFPNYRTTGANGHARG-----
P.euphratica_AJ768574    AEEVKKRHYELLVEDVKHIESGHVPPFPNYRTTGANGHARG-----
P.trichocarpa_sc_IV.1184 AEEVKKRHYEILVEDVKHIESGRVPPFPNYRTTGANGHSKTGEEKMYIYSSL
L.serriola_DW122081      AEEVKQHYEVLVEDVKHIENGRVPPFPNYRRTTGGG-----
L.virosa_TA3253_75947    AEEVKQHYEVLVEDVKHIENGRVPPFPNYRRTTGGG-----
L.sativa_TA7946_4236     AEEVKQHYEVLVEDVKHIENGRVPPFPNYRRTTGGG-----
C.intybus_EH706852       AEEVKKHYEVLVEDVKHIENGRVPPFPNYR-TTGA-----
M.domestica_DR994824     PEEVKKRHYEVLVEDVKHIESGQVPPFDYR-TTGNSHGHMSDEEKMRMN
S.lycopersicum_TA46686   AEEVKKRHYEILLRDVFFIDNGMVPPFPKYKTTGGSHNSTSD-----
N.benthamiana_EH366122  AEEVKQHYEILVHDVLYIENGRVPPFPKYKTT-----TRE-----
T.salsuginea_DN773374   PEEAKRQYDLLVRDIESIENGHVPPFDYKSN-GGSTKGRLRDEEKMRMSM
B.napus_BN06MC20309     PDEAKRQYELLVRDIESIENGHVAFPNYKTN-GGSTKGRLRDEEKMRMSM
A.thaliana_At2g21650     PEEAKRQYDLLVRDIESIENGHVPPFDYKTTTGNNSNRGLRDEEKMRMSM
A.thaliana_At4g39250     TEEVKKRHYELLVQDINSIENGHVPPFPNYRTS-GGCTNGRLSQEEKMRNM
G.max_Glyma06g46590.1   PEEVKKRHYELLRDVRIIESGQVPPFP-YKQNG-----GS-QEEKRLRNM
G.max_Glyma04g16390.1   PEEVKKRHYELLRDVRYIESGKVPPFP-YKQSG-----GS-EEEKRLRNM
M.truncatula_AC135566   PEDVKKRHYELLRDVRIIESGQVAFPNYKNIG-----GY-DEEKRLRNL
G.max_Glyma02g18210.1   PDEVKSHYELLRDISQIESGKVPPFPNYKKS-----EHDEEEKMRNL
M.truncatula_AC157502   PEEVKKHYEVLVEDIKHIESGKVPPFPNYKKS-----VSHEEKMRNM
G.max_Glyma03g28050.1   PEEVKKRHYELLVQDVKHIESGRVPPFPNYKTT-----GSTDQEEKRLRNL
G.hirsutum_DR461014     VEEVKKHYEVLLEDFNRHIESGRVPPFPDYWTVT-----GNRQA--
::.* :*: *: *. *:. * . * *

```

FIGURA 2 (Continuación)

M.guttatus_DV207881	-----
A.majus_TA4522_4151	IR---
V.vinifera_EE072107	LR---
G.hirsutum_TA31277_3635	-----
G.hirsutum_DW503354	-----
H.exilis_EE652824	-----
E.gunnii_CT985658	-----
P.trichocarpa_557252	-----
P.euphratica_AJ768574	-----
P.trichocarpa_sc_IV.1184	-----
L.serriola_DW122081	-----
L.virosa_TA3253_75947	-----
L.sativa_TA7946_4236	-----
C.intybus_EH706852	-----
M.domestica_DR994824	LKLH-
S.lycopersicum_TA46686	-----
N.benthamiana_EH366122	-----
T.salsuginea_DN773374	RLQ--
B.napus_BN06MC20309	KLQ--
A.thaliana_At2g21650	KLQ--
A.thaliana_At4g39250	RLQ--
G.max_Glyma06g46590.1	KLQ--
G.max_Glyma04g16390.1	KLQ--
M.truncatula_AC135566	KLQ--
G.max_Glyma02g18210.1	KIQ--
M.truncatula_AC157502	SLH--
G.max_Glyma03g28050.1	NLNLQ
G.hirsutum_DR461014	-----

FIGURA 2 (Continuación)

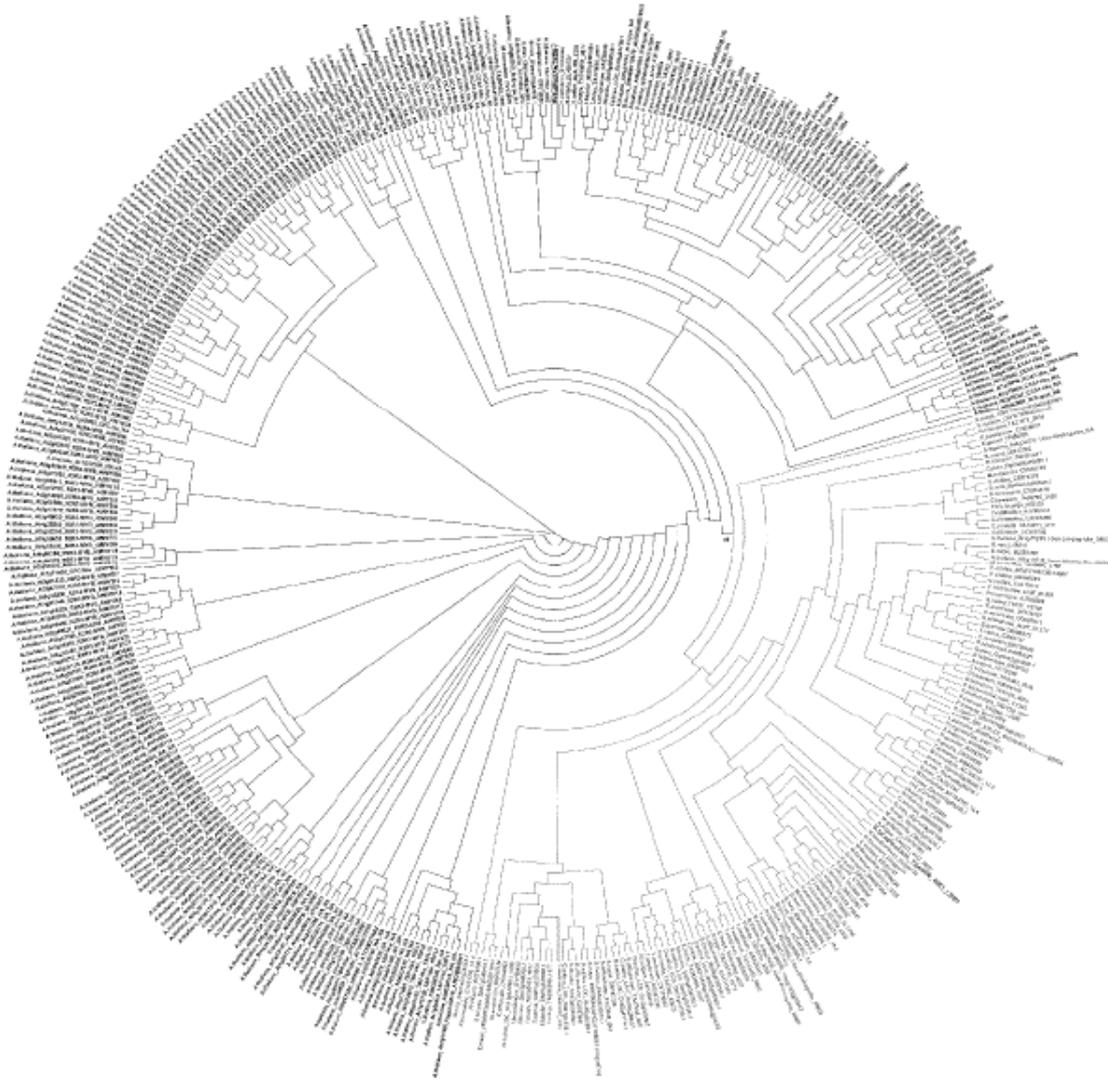


FIGURA 3

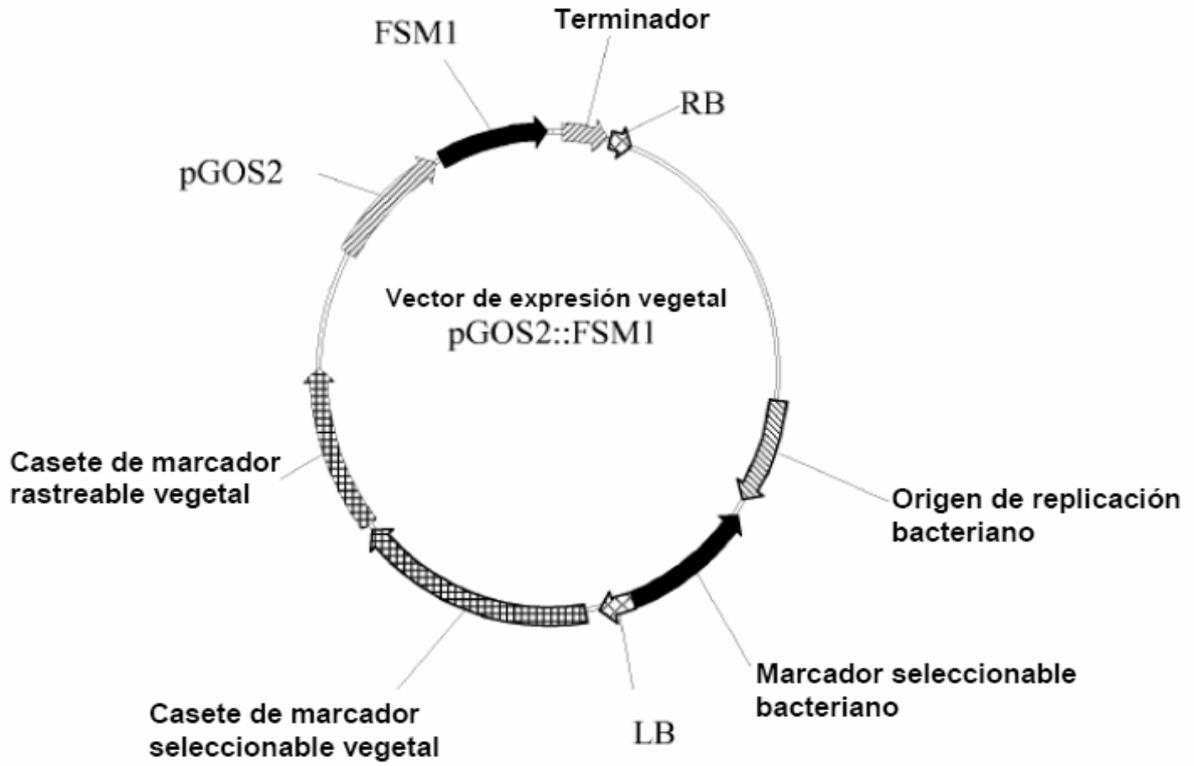


FIGURA 4

```

624028/1-50      IKGPWTKKEEDQRVIELVQKYGP-----
CONSENSO / 80 %  ttt.aopEpthhhphht.hs.....
CONSENSO / 65 %  ppstWotcEcpHlhphhphG.....
CONSENSO / 50 %  p+ssWtpEE-ctLlctlcpaG.....

624028/1-50      -----
CONSENSO / 80 %  .....
CONSENSO / 65 %  .....
CONSENSO / 50 %  .....

624028/1-50      -----KRWSLIAKHLK----
CONSENSO / 80 %  .....tpat.lst.h.....
CONSENSO / 65 %  .....tpWptIupths.....
CONSENSO / 50 %  .....scWspIacpls.....

624028/1-50      -----GRIGKQCRERWHNHLNP
CONSENSO / 80 %  .....t+s..phh.ha.phht.
CONSENSO / 65 %  .....sRosppspp+atphhpt
CONSENSO / 50 %  .....sRosppl+p+apshhpc
    
```

Clase	Clave	Restos
alcohol	o	S, T
alifático	l	I, L, V
cualquiera	.	A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y
aromático	a	F, H, W, Y
cargado	c	D, E, H, K, R
hidrófobo	h	A, C, F, G, H, I, K, L, M, R, T, V, W, Y
negativo	-	D, E
polar	p	C, D, E, H, K, N, Q, R, S, T
positivo	+	H, K, R
pequeño	s	A, C, D, G, N, P, S, T, V
fino	u	A, G, S
de tipo giro	t	A, C, D, E, G, H, K, N, Q, R, S, T

FIGURA 5