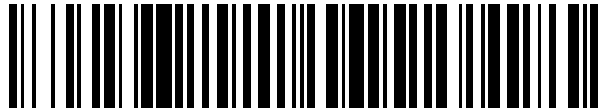


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 429 430**

51 Int. Cl.:

C07K 14/415 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.01.2006 E 06705073 (2)**

97 Fecha y número de publicación de la concesión europea: **04.09.2013 EP 1841870**

54 Título: **Gen y proteína para la detección de azúcares regulados por nitrógeno y modulación de los mismos**

30 Prioridad:

14.01.2005 US 643575 P

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

14.11.2013

73 Titular/es:

**UNIVERSITY OF GUELPH (100.0%)
BUSINESS DEVELOPMENT OFFICE UNIT 4, 130
RESEARCH LANE
GUELPH, ONTARIO N1G 5G3, CA**

72 Inventor/es:

**ROTHSTEIN, STEVEN y
BI, YONG-MEI**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 429 430 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

Gen y proteína para la detección de azúcares regulados por nitrógeno y modulación de los mismos

5 CAMPO DE LA INVENCION

[0001] La presente invención se refiere a procedimiento de modulación de los rasgos característicos agronómicos en plantas modulando la expresión del factor de transcripción GATA en las células vegetales. En particular, la presente invención se refiere a procedimientos para mejorar la utilización del nitrógeno en las plantas.

10 La presente invención también se refiere a moléculas de ácido nucleico aislados de *Arabidopsis thaliana* que comprenden secuencias de nucleótidos que codifican proteínas que detectan el azúcar y, finalmente, pueden modular la captación de nitrógeno y el metabolismo global del carbono.

ANTECEDENTES DE LA INVENCION

15

[0002] La mejora de las características agronómicas de los cultivos vegetales se ha venido desarrollando desde el inicio de la agricultura. En la actualidad se está utilizando la mayor parte de la tierra adecuada para la producción de cosechas. Puesto que la población humana sigue creciendo, será necesario mejorar las variedades de los cultivos para conseguir alimento y comida de forma adecuada (Trewavas (2001) Plant Physiol. 125: 174-179).

20 Para evitar hambrunas catastróficas y malnutrición será necesario mejorar los rendimientos de las variedades de cosechas futuras con ingresos agrícolas equivalentes. Será necesario que estas variedades cultivadas resistan condiciones adversas, como sequía, salinidad del suelo o enfermedades, de forma más eficaz, lo que será especialmente importante para que tierras marginales pueden ponerse en cultivo. Finalmente, serán necesarias variedades cultivadas con alteración de la composición de nutrientes para potenciar la nutrición humana y animal, y

25 permitir un procesamiento de alimentos y piensos más eficaz. Para todos estos rasgos característicos, la identificación de los genes que controlan la expresión fenotípica de dichas características de interés será crucial para acelerar el desarrollo del germoplasma para cultivos superiores mediante métodos convencionales o transgénicos.

30 **[0003]** Diversas estrategias altamente eficaces están disponibles para ayudar en la identificación de genes que tienen una función importante en la expresión de rasgos característicos agronómicamente importantes. Entre estas se incluyen la genética, genómica, bioinformática y genómica funcional. La genética es el estudio científico de los mecanismos de la herencia. Identificando mutaciones que alteran la ruta o respuesta de interés, la genética clásica (o avanzada) puede ayudar a identificar los genes implicados en estas rutas o respuestas. Por ejemplo, un mutante

35 con aumento de la susceptibilidad a la enfermedad puede permitir identificar un componente importante de la ruta de transducción de señales de la planta que lleve del reconocimiento de un patógeno a la resistencia a la enfermedad. La genética también es el componente central para la mejora del germoplasma mediante cultivo. Mediante análisis molecular y fenotípico de cruces genéticos, pueden mapearse rasgos característicos de interés controlados por loci y seguirse en las generaciones posteriores. El conocimiento de los genes subyacentes a una variación fenotípica entre

40 los muestreos de cosechas puede permitir el desarrollo de marcadores que mejoren en gran medida la eficacia del proceso de mejora del germoplasma, así como vías abiertas para el descubrimiento de alelos superiores adicionales.

[0004] La genómica es el estudio a nivel de sistema del genoma de un organismo, incluyendo los genes y los productos génicos correspondiente (ARN y proteínas). En un primer nivel, las técnicas genómicas han proporcionar

45 grandes grupos de datos de información de secuencia a partir de diversas especies vegetales, como secuencias de ADNc completas y parciales y la secuencia genómica completa de una especie vegetal modelo, *Arabidopsis thaliana*. Recientemente, también ha surgido la primera secuencia preliminar del genoma de un cultivo vegetal, la del arroz (*Oryza sativa*). La disponibilidad de una secuencia de genoma completa hace posible el desarrollo de herramientas para el estudio a nivel de sistema de otros complementos moleculares, como matrices y chips para su

50 uso en la determinación de las secuencias complementas de los genes expresados en un organismo en condiciones específicas. Estos datos pueden usarse como primera indicación del potencial de determinados genes para tener funciones importantes en la expresión de fenotipos vegetales diferentes.

[0005] Las técnicas de bioinformática conectan directamente con los grupos de datos genómicos de primer

55 nivel permitiendo el procesamiento de las secuencias descubiertas de interés mediante glosado u otras medidas. Usando, por ejemplo, búsqueda de similitudes, alineamientos y análisis filogenéticos, la bioinformática puede a menudo identificar homólogos de un producto génico de interés. Homólogos muy similar (p. ej., >~90% de identidad de aminoácidos a lo largo de la longitud total de la proteína) son probablemente ortólogos, es decir, comparten la misma función en diferentes organismos.

60

[0006] La genómica funcional puede definirse como la asignación de funciones a los genes y a sus productos. La genómica funcional surge de la genética, genómica y bioinformática para derivar en una vía hacia la identificación de genes importantes en una ruta en particular o respuesta de interés. El análisis de expresión, por ejemplo, utiliza micromatrices de ADN de alta densidad (a menudo derivadas de secuenciación de organismos a escala genómica)

65 para controlar la expresión del ARNm de miles de genes en un único experimento. Entre los tratamientos experimentales pueden incluirse aquellos que permiten obtener una respuesta de interés, como la respuesta de

resistencia a la enfermedad en plantas infectadas con un patógeno. Para dar ejemplos adicionales del uso de micromatrices, pueden seguirse los niveles de expresión de ARNm en tejidos diferentes durante el transcurso del desarrollo o en mutantes afectados en una respuesta de interés. La proteómica también puede ayudar a asignar una función, ensayando la expresión y las modificaciones postraduccionales de cientos de proteínas en un único experimento.

[0007] Las técnicas de proteómica son en muchos casos análogas a las técnicas utilizadas para controlar la expresión del ARNm en experimentos de micromatrices. Las interacciones proteína-proteína también pueden ayudar a asignar proteínas a una ruta o respuesta determinada, identificando proteínas que interactúan con componentes conocidos de la ruta o respuesta. En la genómica funcional se estudian a menudo las interacciones proteína-proteína usando ensayos de doble híbrido de levadura a gran escala. Otra técnica para asignar la función génica es expresar la proteína correspondiente en un huésped heterólogo, por ejemplo, la bacteria *Escherichia coli*, seguido de su purificación y de ensayos enzimáticos.

[0008] La demostración de la capacidad de un gen de interés para controlar un rasgo característico determinado puede derivar, por ejemplo, de una prueba experimental en especies vegetales de interés. La generación y análisis de plantas transgénicas para un gen de interés pueden usarse para la genómica funcional vegetal, con varias ventajas. El gen puede además estar tanto sobreexpresado como subexpresado («silenciado»), aumentando de este modo las posibilidades de observar un fenotipo que liga al gen a una ruta o respuesta de interés. Dos aspectos de la genómica funcional transgénica ayudan a dar un alto nivel de confianza a la asignación funcional mediante esta técnica. En primer lugar, las observaciones fenotípicas se llevan a cabo en el contexto de la planta viva. En segundo lugar, pueden comprobarse la gama de fenotipos observados y correlacionarse con niveles de expresión observados de los transgenes introducidos. La genómica funcional transgénica es especialmente válida para mejorar el desarrollo de la variedad cultivada. Solo genes que funcionan en una ruta o respuesta de interés y que, además, son capaces de conferir un fenotipo deseado en base a un rasgo característico, se promocionan como genes candidatos en los esfuerzos por mejorar una cosecha. En algunos casos, las líneas transgénicas desarrolladas para estudios genómicos funcionales pueden utilizarse directamente en estadios iniciales de desarrollo del producto.

[0009] Otra técnica dirigida hacia la genómica funcional vegetal supone identificar en primer lugar las líneas vegetales con mutaciones en genes de interés específicos, seguido de la evaluación fenotípica de las consecuencias de dichos silenciamiento deL gen sobre el rasgo característico estudiado. Esta técnica muestra genes esenciales para la expresión de rasgos característicos específicos.

[0010] Los genes identificados mediante la genómica funcional pueden emplearse directamente en los esfuerzos dirigidos a la mejora del germoplasma mediante técnicas transgénicas, como se describe anteriormente o utilizado para desarrollar marcadores para la identificación del seguimiento de alelos de interés en el mapeo y las poblaciones reproductoras. El conocimiento de dichos genes también pueden permitir la construcción de alelos superiores no existentes en la naturaleza, mediante cualquier de los diversos procedimientos moleculares.

[0011] El rápido aumento en el rendimiento durante los últimos 80 años de las cosechas de surcos ha sido debido aproximadamente por igual medida a la mejora genética y a mejoras en las prácticas agronómicas. En particular, en una cosecha, como el maíz, la combinación de híbridos de alto rendimiento y el uso de cantidades grandes de fertilizante nitrogenado se hace en condiciones ideales que permiten rendimientos mayores de 27600 kg/ha. Sin embargo, el uso de cantidades grandes de fertilizante nitrogenado tiene efectos secundarios negativos principalmente alrededor del aumento del coste de esta entrada para el agricultor y el coste para el medio ambiente ya que la contaminación por nitratos es un problema importante en muchas zonas agrícolas que contribuyen significativamente a la degradación tanto del agua dulce como del medio ambiente marino. El desarrollo de la genética de cosechas que usan nitrógeno de forma más eficaz a través de un entendimiento del papel del genotipo sobre el uso del nitrógeno sería altamente ventajosa reduciendo los costes de ingresos del productor así como la carga medioambiental. Esto es especialmente importante para una cosecha como el maíz que crece usando un alto nivel de fertilizante nitrogenado.

[0012] La eficacia del uso del nitrógeno puede definirse de varias formas, aunque la más simple es el rendimiento/N suministrado. Hay dos estadios en este proceso: en primer lugar, la cantidad de nitrógeno disponible que se capta, almacena y asimila como aminoácidos y otros compuestos nitrogenados importantes; en segundo lugar, la proporción de nitrógeno que se reparte a la semilla, que da lugar al rendimiento final. Se han realizado múltiples estudios de campo en diversos cultivos de importancia agrícola para estudiar este problema (Lawlor DW y col., 2001 en Lea PJ., Morot-Gaudry JF, eds. Plant Nitrogen. Berlín: Springer-Verlag 343-367; Lafitte HR y Edmeades GO 1994 Field Crops Res 39, 15-25; Lawlor DW 2002 J Exp Bot. 53, 773-87; Moll RH y col 1982 Agron J 74, 562-564). Estos experimentos han demostrado que existe un componente genético para el uso eficaz del nitrógeno, aunque no se ha comprobado que son satisfactorios para determinar qué genes son importantes para este proceso. Además, los agricultores del maíz generalmente no han tenido como objetivo el mantenimiento del rendimiento bajo condiciones de fertilizante nitrogenado limitantes. Estos tipos de experimentos de campo sobre el uso de nitrógeno son difíciles por diversas razones como la falta de uniformidad de nitrógeno accesible en un campo de prueba o entre las zonas del campo bajo cualquier régimen de tratamiento y la interacción de otros factores

ambientales que pueden dificultar la interpretación de los experimentos.

[0013] Por tanto, aunque existe evidencia experimental de la variación genética para este rasgo característico, es difícil obtener conclusiones a partir de estos experimentos sobre que produce esta variación. Sería posible y ciertamente importante desarrollar procedimientos para estudiar este rasgo característico en condiciones de campo en los cultivos. Sin embargo, pueden hacerse progresos significativos hacia la identificación, comprensión y manipulación de rasgos característicos importantes mediante el uso de un sistema modelo como el de *Arabidopsis*. En el último de los casos, estos experimentos darán pistas importantes sobre los posibles genes diana para evaluar su importancia para las cosechas. Además, también existen considerables recursos genéticos y genómicos disponibles para estudiar el arroz y sus especies también se usarán para algunos de los experimentos propuestos como una especie más similar al maíz que *Arabidopsis*.

[0014] El nitrato es la principal forma de nitrógeno disponible en el campo y existe gran cantidad de bibliografía sobre genes implicados en la captación y reducción del nitrato (Forde B G 2000 *Biochimica et Biophysica Acta* 1465, 219-235; Howitt SM y Udvardi MK 2000 *Biochimica et Biophysica Acta* 1465, 152-170; Stitt M y col 2002 *J Exp Bot.* 53, 959-70) así como sobre los genes implicados en otros aspectos del metabolismo del nitrógeno (Lea P.J., Morot-Gaudry JF, eds. 2001 *Plant Nitrogen*. Berlín: Springer-Verlag; Morot-Gaudry JF 2001 *Nitrogen assimilation by plants* Science Publishers Inc. NH, EE. UU.). También, está claro que la disponibilidad de metabolitos de carbono es crucial para el uso eficaz del nitrato del suero y existen buenas evidencias experimentales sobre la relación entre el metabolismo del carbono y del nitrógeno (Coruzzi GM y Zhou L 2001 *Curr Opin Plant Biol.* 4, 247-53). Además, algunos experimentos sugieren que GS y GOGAT están implicados en la reactivación del N desde órganos senescentes al órgano sumidero (Brouquisse R y col., 2001 en Lea P.J, Morot-Gaudry JF, eds. *Plant Nitrogen*. Berlín: Springer-Verlag 275-293; Yamaya T y col. 2002 *J Exp Bot.* 53, 917-925). Sin embargo, muchos aspectos de la regulación de estos genes siguen sin estar claros y no se tienen nociones de como esta regulación afecta al uso eficaz del nitrógeno.

[0015] Las plantas pueden detectar los niveles de metabolitos de carbono y nitrógeno y ajustar su crecimiento y desarrollo en consecuencia. Los mecanismos de percepción son complejas redes reguladoras que controlan la expresión génica para acomodar los cambios constantes en las actividades celulares dependientes de nutrientes. La posesión de un mecanismo de detección de azúcar permite a las plantas desactivar la fotosíntesis cuando los esqueletos carbonados son abundantes. Los mecanismos de detección de N permiten a las plantas detener la captación y reducción del nitrato cuando los niveles de N reducido u orgánico son altos (Coruzzi, G.M. y Zhou, L. (2001) *Curr Opin Plant Biol.* 4, 247-53).

[0016] Existen múltiples vías de transducción de señales de azúcares en plantas. La glucosa ha surgido como un regulador clave en muchos procesos vitales en plantas fotosintéticas como en la fotosíntesis y en el metabolismo del carbono y del nitrógeno (Rolland, F., Moore, B. y Sheen, J. (2002) *Plant Cell* S185-S205). Las hexoquinasas (HKK) son un punto de control importante para el metabolismo de la glucosa. No solo catalizan la fosforilación de la glucosa sino que también funcionan como detectores de glucosa para interrelacionar nutriente, luz y redes de señalización de hormonas para controlar el crecimiento y desarrollo en respuesta al entorno cambiante (Jang, J., Leon, P, Zhou, L. y Sheen, J. (1997) *Plant Cell* 9, 5-19; Dai, N., Schaffer, A., Petreikov, M., Shahak, Y., Giller, Y., Ratner, K., Levine, A. y Granot, D. (1999) *Plant Cell* 11, 1253-1266; Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W., Liu, Y., Hwang, I., Jones, T. y Sheen, J. (2003) *Science* 300, 332-336). En otros organismos se ha demostrado que las moléculas que transportan hexosa también sirven como detectores de azúcar.

[0017] También existen en las plantas múltiples señales N y rutas de detección. Las plantas tienen mecanismos para detectar el nitrato, la forma principal de fertilizante nitrogenado, como señal para el estado del N inorgánico así como para detectar metabolitos del nitrato como señales para los estados del N reducido u orgánico. La nitrato reductasa (NR) y la nitrito reductasa (NiR) son las dos primeras enzimas en el proceso de reducción del nitrato y su expresión puede estimularse por la presencia de nitrato y modularse mediante otros factores fisiológicos como algunos compuestos nitrogenados, sacarosa, luz y hormonas (Forde, B.G. (2000) *Biochimica et Biophysica Acta* 1465, 219-235; Howitt, S.M. y Udvardi, M.K. (2000) *Biochimica et Biophysica Acta* 1465, 152-170; Stitt, M., Müller, M., Matt, M., Gibon, Y., Carillo, P., Morcuende, R., Scheible, W. y Krapp, A. (2002) *J Exp Bot.* 53, 959-970; Lea, P.J. y Morot-Gaudry, J.F. eds. 2001 *Plant Nitrogen*. Berlín: Springer-Verlag; Morot-Gaudry J F 2001 *Nitrogen assimilation by plants* Science Publishers Inc. NH, EE. UU.).

[0018] Es claro que el metabolismo del carbono y del nitrógeno están relacionados de cerca y estrechamente ligados (Coruzzi, G. y Bush, D.R. (2001) *Plant Physiol* 125, 61-64). La disponibilidad de metabolitos de carbono es crucial para la utilización eficaz de los nitratos y el estado del nitrógeno es muy sensible a la fotosíntesis. A pesar del aumento del conocimiento sobre los genes estructurales implicados en el metabolismo del carbono y del nitrógeno, los factores de transactivación implicados en la regulación transcripcional de la expresión de genes C/N no se han caracterizado.

[0019] Los factores de transcripción GATA son un grupo de reguladores transcripcionales ampliamente distribuidos en eucariotas. El dominio de unión al ADN de GATA normalmente reconoce la secuencia consenso WGATAR (W=T o A, R=G o A) (Lowry, J. y Atchley, W. (2000) *J Mol Evol* 50, 103-115). Se han identificado motivos

GATA en las regiones reguladoras de muchos genes fotosensibles (que responden a la luz) (Arguello-Astorga, G. y Herrera-Estrella, L. (1998) *Annu Rev Plant Physiol Plant Mol Biol* 49, 525-555), incluyendo muchos genes implicados o relacionados con la fotosíntesis como el RBCS, CAB (proteína que une clorofila A/B) y GAP (gliceraldehído-3-fosfato deshidrogenasa) (Terzaghi, W.B. y Cashmore, A.R. (1995) *Annu Rev Plant Physiol Plant Mol Biol* 46, 445-474; Koch, K.E. (1996) *Carbohydrate-modulated gene expression in plants. Annu Rev Plant Physiol Plant Mol Biol* 47, 509-540; Jeong, M.J. y Shih, M.C. (2003) *Biochem Biophys Res Commun* 300, 555-562) así como genes implicados en la asimilación del nitrato como la nitrato reductasa, nitrito reductasa y Gln sintetasa (Jarai, G., Truong, H., Daniel-Vedele, F. y Marzluf, G. (1992) *Curr Genet* 21, 37-41; Rastogi, R., Bate, N., Sivasankar, S y Rothstein, S. (1997) *Plant Mol. Biol.* 34, 465-76; Oliveira, I.C. y Coruzzi, G.M. (1999) *Plant Physiol* 121, 301-309). Algunas proteínas reguladoras que actúan en trans que globalmente regulan genes del metabolismo del N son genes del factor de transcripción GATA. En levaduras, cuatro factores reguladores generales del nitrógeno GLN3, NIL1, NIL2 y DAL80 son proteínas que se unen a ADN que contiene un motivo único de dedo de cinc GATA, que reconocen el motivo consenso GATA (Hofman-Bang, J. (1999) *Mol Biotech* 12, 35-73). En hongos, NIT2 de *Neurospora crassa* (Tao Y y Marzluf GA 1999 *Curr Genet* 36, 153-158) y AREA de *Aspergillus nidulans* (Caddick MX, Arst HN, Jr Taylor LH, Johnson RI, Brownlee AG 1986 *Cloning of the regulatory gene areA mediating nitrogen metabolite repression in Aspergillus nidulans EMBO J* 5, 1087-1090) son los genes del factor de transcripción GATA.

[0020] En las plantas, la función *in vivo* de los factores GATA sigue estando muy mal definida, teniendo el genoma de *Arabidopsis* 30 miembros GATA (Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K. y Yu, G. (2000) *Science* 290, 2105-2110; Reyes, J.C., Muro-Pastor, M.I. y Florencio, F.J. (2004) *Plant Physiol.* 134, 1718-1732).

RESUMEN DE LA INVENCION

[0021] La invención es como se describe en las reivindicaciones adjuntas. En el intento por entender la función biológica de los miembros de la familia de genes del factor de transcripción GATA, se ha estudiado un gen del factor de transcripción GATA para entender su papel en la regulación del metabolismo del carbono y del nitrógeno. Los inventores han determinado que la expresión del gen At5g56860 está influenciada por el estado del nitrógeno y su expresión regula la expresión de genes que controlan el metabolismo del carbono así como genes implicados en otros procesos biológicos. Las plantas mutantes con pérdida de función en el gen At5g56860 daban lugar a la reducción en el nivel de clorofila y estas plantas eran hipersensibles a la glucosa exógena. Por el contrario, las plantas transgénicas con ganancia de función eran menos sensibles a la glucosa exógena.

[0022] Los azúcares son reguladores centrales de muchos procesos vitales en plantas fotosintéticas, como fotosíntesis y metabolismo del carbono y el nitrógeno. Esta regulación se alcanza mediante la regulación de la expresión génica para activar o reprimir los genes implicados. Los mecanismos por los que los azúcares controlan la expresión génica no se conocen bien. El factor de transcripción GATA descrito en este documento está implicado en la regulación de la detección de azúcares y la expresión del factor en sí está influenciada por el cambio del estado del N. El aumento de la expresión de este gen puede producir plantas con aumento del rendimiento, especialmente porque la manipulación de las rutas de señalización de azúcares puede llevar a un aumento de la fotosíntesis, así como de la asimilación del nitrógeno, y a alterar las relaciones fuente-sumidero en semillas, vasos conductores, raíces y otros órganos de almacenamiento.

[0023] Por consiguiente, la presente invención se refiere a un procedimiento de modulación de una característica en una planta o célula vegetal que comprende modular la expresión de un gen del factor de transcripción GATA en la planta o célula vegetal. En un ejemplo de la descripción, la expresión del gen del factor de transcripción GATA se modula mediante la administración a la célula de una cantidad eficaz de un agente que puede modular los niveles de expresión de un gen del factor de transcripción GATA en la célula vegetal. En otro ejemplo de la descripción, el agente potencia los niveles de expresión de un gen del factor de transcripción GATA en la célula vegetal.

[0024] La característica que se va a modular en la planta puede ser cualquier rasgo característico agronómico de interés. En un ejemplo de la descripción, la característica es cualquiera que esté afectada por el metabolismo del nitrógeno, carbono y/o azufre, biosíntesis de lípidos, percepción de nutrientes, adaptación nutricional, transporte de electrones y/o conservación de energía asociada a membranas. En un ejemplo adicional de la descripción, la característica se selecciona entre uno o más de entre utilización de nitrógeno, rendimiento, crecimiento celular, reproducción, fotosíntesis, asimilación de nitrógeno, resistencia a enfermedades, diferenciación, transducción de señales, regulación génica, tolerancia al estrés abiótico y composición nutricional. Aún en un ejemplo adicional de la descripción la característica modulada es un aumento o mejora de uno o más de entre utilización de nitrógeno, rendimiento, crecimiento celular, reproducción, fotosíntesis, asimilación de nitrógeno, resistencia a enfermedades, diferenciación, transducción de señales, regulación génica, tolerancia al estrés abiótico y composición nutricional.

[0025] La presente invención se refiere a un procedimiento de mejora de la utilización del nitrógeno en una planta o célula vegetal que comprende potenciar la expresión de un gen del factor de transcripción GATA en la planta o célula vegetal. La mejora de la utilización del nitrógeno en una planta permitirá la reducción de las

cantidades de fertilizante nitrogenado aplicado a la planta con una reducción concomitante de los costes para el agricultor y para el medio ambiente ya que la contaminación por nitratos es un problema principal en muchas zonas agrícolas que contribuye significativamente a la degradación tanto del agua dulce como del medioambiente marino.

5 **[0026]** La planta o célula vegetal puede proceder de cualquier planta donde se desee modular una característica. En una realización de la invención, la célula vegetal es una dicotiledónea, una gimnosperma o una monocotiledónea. En una realización, la dicotiledónea se selecciona a partir del grupo compuesto por soja, tabaco o algodón. En una realización adicional de la invención, la monocotiledónea se selecciona a partir de maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* sp. y teosinte.

10 **[0027]** En un ejemplo de la descripción, el agente que potencia los niveles de expresión de un gen del factor de transcripción GATA en la célula vegetal comprende una molécula de ácido nucleico que codifica un factor de transcripción GATA.

15 **[0028]** En un ejemplo de la descripción, el agente que potencia los niveles de expresión de un gen del factor de transcripción GATA en una célula vegetal comprende:

- 20 a) una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma;
- b) una secuencia de nucleótidos que codifica un polipéptido SEC ID N.º 2 o un fragmento o dominio de la misma;
- c) una secuencia de nucleótidos que tiene similitud considerable con a) o b);
- 25 d) una secuencia de nucleótidos capaz de hibridar con a), b) o c);
- e) una secuencia de nucleótidos que es complementaria a a), b), c) o d); o
- 30 f) una secuencia de nucleótidos que es la complementaria inversa de a), b), c) o d).

[0029] En un ejemplo adicional de la descripción, la molécula de ácido nucleico comprende la secuencia del gen AT5g56860 SEC ID N.º 1 o un fragmento funcional de la misma. Aún en una realización adicional de la invención, la molécula de ácido nucleico comprende la secuencia que hibrida en condiciones de rigurosidad media con el gen AT5g56860 SEC ID N.º 1 o un fragmento funcional de la misma. En otro ejemplo de la presente

35 invención, la molécula de ácido nucleico procede de la secuencia de nucleótidos del gen AT5g56860 SEC ID N.º 1 y tiene una secuencia de nucleótidos que comprende los codones específicos para su expresión en plantas. Aún en otra realización de la invención, la molécula de ácido nucleico es el ortólogo del arroz del gen AT5g56860 que comprende la secuencia SEC ID N.º 3.

40 **[0030]** En un ejemplo adicional de la invención, el agente que puede modular los niveles de expresión de un gen del factor de transcripción GATA en una célula vegetal comprende:

- 45 a) una secuencia polipeptídica listada en SEC ID N.º 2 o un fragmento, dominio, repetición o quimera funcional de la misma;
- b) una secuencia polipeptídica que tiene similitud considerable con a);
- c) una secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o que tiene similitud considerable con la secuencia de nucleótidos listada en SEC ID N.º 1, o un fragmento o dominio funcional de la
- 50 misma o una secuencia complementaria de la misma; o
- d) una secuencia polipeptídica codificada por una secuencia de nucleótidos capaz de hibridar en condiciones de rigurosidad media con una secuencia de nucleótidos listada en SEC ID N.º 1 o a una secuencia complementaria de la misma.

55 **[0031]** En una realización de la presente invención, la secuencia de ácido nucleico se expresa en una ubicación o tejido específico de la planta. La ubicación o tejido es, por ejemplo, pero sin limitaciones, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y/o flor. En una realización alternativa, la ubicación o el tejido es una semilla.

60 **[0032]** Los ejemplos de la presente descripción también se refieren al uso de una molécula de ácido nucleico combinada aleatoriamente para modular una característica de una célula vegetal, conteniendo dicha molécula de ácido nucleico combinada aleatoriamente diversos fragmentos de secuencias de nucleótidos, en la que al menos uno de los fragmentos que codifican un factor de transcripción GATA y en la que al menos dos de los diversos

65 fragmentos de secuencia están en un orden, de 5' a 3' que no es el orden en el que los diversos fragmentos se encuentran en la naturaleza en un ácido nucleico. En un ejemplo específico, todos los fragmentos de una molécula

de ácido nucleico combinada aleatoriamente que contiene diversos fragmentos de secuencias de nucleótidos proceden de un único gen. En un ejemplo más específico, los diversos fragmentos proceden de al menos dos genes diferentes. En un ejemplo más específico, el ácido nucleico combinado aleatoriamente está unido de forma operativa a una secuencia promotora. Otro ejemplo más específico es un uso de un poli nucleótido quimérico para modular una característica de una célula vegetal, incluyendo dicho polinucleótido quimérico una secuencia promotora unida de forma operativa al ácido nucleico combinado aleatoriamente. En un ejemplo más específico, el ácido nucleico combinado aleatoriamente está contenido dentro de una célula huésped. En un ejemplo específico adicional de la descripción, el fragmento que codifica un factor de transcripción consta o comprende:

- 5 a) una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma;
- b) una secuencia de nucleótidos que codifica un polipéptido SEC ID N.º 2 o un fragmento o dominio de la misma;
- c) una secuencia de nucleótidos que tiene similitud considerable con a) o b);
- 15 d) una secuencia de nucleótidos capaz de hibridar con a), b) o c);
- e) una secuencia de nucleótidos que es complementaria a a), b), c) o d); o
- 20 f) una secuencia de nucleótidos que es la complementaria inversa de a), b), c) o d).

[0033] Las realizaciones de la presente invención también contemplan un uso de un casete de expresión para modular una característica de una célula vegetal que incluye una secuencia promotora unida de forma operativa a un ácido nucleico aislado que codifica un factor de transcripción GATA. En ejemplos de la descripción, el ácido nucleico aislado que codifica un factor de transcripción GATA consta o comprende:

- a) una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma;
- b) una secuencia de nucleótidos que codifica un polipéptido SEC ID N.º 2 o un fragmento o dominio de la misma;
- 30 c) una secuencia de nucleótidos que tiene similitud considerable con a) o b);
- d) una secuencia de nucleótidos capaz de hibridar con a), b) o c);
- 35 e) una secuencia de nucleótidos que es complementaria a a), b), c) o d); o
- f) una secuencia de nucleótidos que es la complementaria inversa de a), b), c) o d).

[0034] La invención también contemplan el uso de un vector recombinante para modular una característica de una célula vegetal que comprende un casete de expresión que incluye una secuencia promotora unida de forma operativa a un ácido nucleico aislado que codifica un factor de transcripción GATA. En ejemplos de la descripción, el ácido nucleico aislado que codifica un factor de transcripción GATA consta o comprende:

- a) una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma;
- 45 b) una secuencia de nucleótidos que codifica un polipéptido SEC ID N.º 2 o un fragmento o dominio de la misma;
- c) una secuencia de nucleótidos que tiene similitud considerable con a) o b);
- 50 d) una secuencia de nucleótidos capaz de hibridar con a), b) o c);
- e) una secuencia de nucleótidos que es complementaria a a), b), c) o d); o
- f) una secuencia de nucleótidos que es la complementaria inversa de a), b), c) o d).

[0035] También abarca usos de células vegetales, que contienen casetes de expresión, según la presente descripción, y usos de las plantas que contienen estas células vegetales.

[0036] En una realización, el casete de expresión se expresa en toda la planta. En otra realización, el casete de expresión se expresa en una ubicación o tejido específico de una planta. En una realización específica, la ubicación o tejido puede ser, por ejemplo, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En una realización específica alternativa, la ubicación o el tejido es una semilla.

[0037] Los ejemplos de la presente descripción también proporcionan el uso de la semilla y el producto aislado de las plantas para modular una característica en una célula vegetal que contiene un casete de expresión que incluye una secuencia promotora unida de forma operativa a un ácido nucleico aislado que codifica un gen del factor

de transcripción GATA según la presente invención.

[0038] En una realización específica, el vector de expresión incluye uno o más elementos, como por ejemplo, pero sin limitaciones, una secuencia promotora-potenciadora, una secuencia marcadora de selección, un origen de replicación, una secuencia codificadora de epítipo etiquetado o una secuencia codificadora de purificación por afinidad etiquetada. En una realización más específica, la secuencia promotora-potenciadora puede ser, por ejemplo, el promotor 35S de CaMV, el promotor 19S de CaMV, el promotor PR-1a del tabaco, el promotor de la ubiquitina y la faseolina. En otra realización, el promotor puede funcionar en plantas y, más específicamente, es un promotor constitutivo o inducible. En otra realización específica, la secuencia marcadora de selección codifica un gen de resistencia a antibiótico. En otra realización específica, la secuencia de epítipo etiquetado codifica V5, el péptido Phe-His-His-Thr-Thr, hemaglutinina o glutatión-S-transferasa. En otra realización específica, la secuencia de purificación por afinidad etiquetada codifica una secuencia de ácido poliamino o un polipéptido. En una realización más específica, la secuencia de ácido poliamino es polihistidina. En una realización más específica, el polipéptido es el dominio de unión a quitina o glutatión-S-transferasa. En una realización más específica, la secuencia de purificación por afinidad etiquetada comprende una secuencia codificadora de inteína.

[0039] En una realización específica, el vector de expresión es un vector de expresión eucariota o un vector de expresión procariota. En una realización más específica, el vector de expresión eucariota incluye un promotor específico de tejido. Más específicamente, el vector de expresión puede funcionar en plantas.

[0040] Las realizaciones de la presente invención también se refieren a una planta modificada por un procedimiento que incluye la introducción en una planta de un ácido nucleico en el que el ácido nucleico se puede expresar en la planta en una cantidad efectiva para realizar la modificación. La modificación puede ser un aumento o disminución de una o más de los rasgos característicos de interés. La modificación puede incluir sobreexpresión, subexpresión, modulación de la cadena complementaria, supresión de la cadena sentido, expresión inducible, represión inducible o modulación inducible de un gen. En una realización de la invención, la modificación implica un aumento o mejora del rasgo característico de interés, por ejemplo, la utilización de nitrógeno.

[0041] Los ejemplos de la presente descripción proporcionan secuencias de nucleótidos y de aminoácidos aislados de *Arabidopsis thaliana*. En especial, la presente invención se refiere a un gen del factor de transcripción GATA regulado por nitrógeno necesario para la detección de azúcares.

[0042] Un ejemplo de la presente invención se refiere a un ácido nucleico aislado que comprende o consta de una secuencia de nucleótidos que comprende:

- a) una secuencia de nucleótidos listada en SEC ID N.º 1 o un fragmento o dominio de la misma;
- b) una secuencia de nucleótidos que tiene similitud considerable con a);
- c) una secuencia de nucleótidos capaz de hibridar con a);
- d) una secuencia de nucleótidos que es complementaria a a), b) o c); o
- e) una secuencia de nucleótidos que es la complementaria inversa de a), b) o c).

[0043] En un ejemplo específico, la similitud considerable es de al menos aproximadamente el 65% de identidad, específicamente aproximadamente el 80% de identidad, específicamente el 90% y más específicamente al menos aproximadamente el 95% de identidad de secuencia con la secuencia de nucleótidos listada como SEC ID N.º 1, o un fragmento o dominio de la misma.

[0044] En un ejemplo, la secuencia que tiene similitud considerable con la secuencia de nucleótidos SEC ID N.º 1, un fragmento o dominio de la misma, procede de una planta. En un ejemplo específico, la planta es una dicotiledónea. En un ejemplo más específico, la dicotiledónea se selecciona a partir del grupo compuesto por soja, tabaco o algodón. En otro ejemplo específico, la planta es una gimnosperma. En otro ejemplo específico, la planta es una monocotiledónea. En un ejemplo más específico, la monocotiledónea es un cereal. En un ejemplo más específico, el cereal puede ser, por ejemplo, maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* sp. o teosinte.

[0045] En una realización, el ácido nucleico se expresa en una ubicación o tejido específico de una planta. La ubicación o tejido es, por ejemplo, pero sin limitaciones, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En una realización alternativa, la ubicación o el tejido es una semilla. En otro ejemplo, el ácido nucleico codifica un polipéptido implicado en una función, como por ejemplo, pero sin limitaciones, el metabolismo del carbono, nitrógeno y/o azufre, utilización del nitrógeno, asimilación del nitrógeno, fotosíntesis, transducción de señales, crecimiento celular, reproducción, resistencia a enfermedades, tolerancia al estrés abiótico, composición nutricional, regulación génica y/o diferenciación.

- [0046]** En un ejemplo específico, el ácido nucleico aislado comprende o consta de una secuencia de nucleótidos capaz de hibridar con una secuencia de nucleótidos listada en SEC ID N.º 1 o un fragmento o dominio de la misma. En un ejemplo específico, la hibridación permite que la secuencia forme un dúplex en condiciones de rigurosidad media o alta. Los ejemplos de la presente descripción también abarcan una secuencia de nucleótidos complementaria a la secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma. Ejemplos de la presente invención además abarcan una secuencia de nucleótidos complementaria a una secuencia de nucleótidos que tiene similitud considerable es capaz de hibridar con una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma.
- 10 **[0047]** En un ejemplo específico, la secuencia de nucleótidos que tiene similitud considerable es una variante alélica de la secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma. En un ejemplo alternativo, la secuencia que tiene similitud considerable es una variante natural. En otro ejemplo alternativo, la secuencia que tiene similitud considerable es una variante polimórfica de la secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma.
- 15 **[0048]** En un ejemplo específico, el ácido nucleico aislado contiene diversas regiones que tienen la secuencia de nucleótidos SEC ID N.º 1 o un exón o dominio de la misma.
- [0049]** En un ejemplo específico, el ácido nucleico aislado contiene una secuencia que codifica un polipéptido. En un ejemplo más específico, la secuencia que codifica un polipéptido contiene una porción de nucleótidos de 20 pares de bases idéntica en secuencia a una porción de nucleótidos de 20 pares de bases de una secuencia de ácidos nucleicos de SEC ID N.º 1. En un ejemplo más específico, el polipéptido contiene una secuencia polipeptídica SEC ID N.º 2 o un fragmento de la misma. En un ejemplo más específico, el polipéptido es un polipéptido de una planta. En un ejemplo más específico, la planta es una dicotiledónea. En un ejemplo más específico, la planta es una gimnosperma. En un ejemplo más específico, la planta es una monocotiledónea. En un ejemplo más específico, la monocotiledónea es un cereal. En un ejemplo más específico, el cereal puede ser, por ejemplo, maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* y teosinte.
- 20 **[0050]** En un ejemplo, el polipéptido se expresa en toda la planta. En un ejemplo más específico, el polipéptido se expresa en una ubicación o tejido específico de una planta. En un ejemplo más específico, la ubicación o tejido puede ser, por ejemplo, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En un ejemplo más específico, la ubicación o el tejido es una semilla.
- 25 **[0051]** En un ejemplo específico, la secuencia del ácido nucleico aislado codifica un polipéptido útil para la generación de un anticuerpo que tiene inmunorreactividad frente a un polipéptido codificado por una secuencia de nucleótidos SEC ID N.º 2 o un fragmento o dominio de la misma.
- 30 **[0052]** En un ejemplo específico, la secuencia que tiene similitud considerable contiene una delección o inserción de al menos un nucleótido. En un ejemplo más específico, la delección o inserción es de menos de aproximadamente treinta nucleótidos. En el ejemplo más específico, la delección o inserción es de menos de aproximadamente cinco nucleótidos.
- 35 **[0053]** En un ejemplo específico, la secuencia del ácido nucleico aislado que tiene similitud considerable comprende o consta de una sustitución en al menos un codón. En un ejemplo específico, la sustitución es conservadora.
- 40 **[0054]** Los ejemplos de la presente invención también se refiere a una molécula aislada de ácido nucleico que comprende o consta de una secuencia de nucleótidos, su complementaria o su complementaria inversa, que codifica un polipéptido que incluye:
- 45 a) una secuencia polipeptídica SEC ID N.º 2 o un fragmento, dominio, repetición o quimera de la misma;
- b) una secuencia polipeptídica que tiene similitud considerable con a);
- 50 c) una secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o que tiene una similitud considerable con la secuencia de nucleótidos SEC ID N.º 1, o un fragmento o dominio de la misma o una secuencia complementaria a la misma;
- d) una secuencia polipeptídica codificada por una secuencia de nucleótidos capaz de hibridar en condiciones de rigurosidad media con una secuencia de nucleótidos SEC ID N.º 1 o a una secuencia complementaria a la misma; o
- 55 e) un fragmento funcional de a), b), c) o d).
- 60 **[0055]** En otro ejemplo específico, el polipéptido que tiene similitud considerable es una variante alélica de una secuencia polipeptídica SEC ID N.º 2 o un fragmento, dominio, repetición o quimera de la misma. En otro ejemplo específico, el ácido nucleico aislado incluye diversas regiones de la secuencia polipeptídica codificada por una

secuencia de nucleótidos idéntica o con similitud considerable con una secuencia SEC ID N.º 1 o fragmento o dominio de la misma, o una secuencia complementaria a la misma.

[0056] En otro ejemplo específico, el polipéptido es una secuencia polipeptídica SEC ID N.º 2. En otro ejemplo específico, el polipéptido es un fragmento o dominio funcional. Aún en otro ejemplo específico, el polipéptido es una quimera, en el que la quimera puede incluir dominios proteicos funcionales, como dominios, repeticiones, sitios de modificación postraduccional u otras características. En un ejemplo más específico, el polipéptido es un polipéptido de una planta. En un ejemplo más específico, la planta es una dicotiledónea. En un ejemplo más específico, la planta es una gimnosperma. En un ejemplo más específico, la planta es una monocotiledónea. En un ejemplo más específico, la monocotiledónea es un cereal. En un ejemplo más específico, el cereal puede ser, por ejemplo, maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* y teosinte.

[0057] En un ejemplo específico el polipéptido se expresa en una ubicación o tejido específico de una planta. En un ejemplo más específico, la ubicación o tejido puede ser, por ejemplo, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En otro ejemplo específico, la ubicación o el tejido es una semilla.

[0058] En un ejemplo específico, la secuencia polipeptídica codificada por una secuencia de nucleótidos que tiene similitud considerable con una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma o una secuencia complementaria a la misma, incluye una delección o inserción de al menos un nucleótido. En un ejemplo más específico, la delección o inserción es de menos de aproximadamente treinta nucleótidos. En un ejemplo más específico, la delección o inserción es de menos de aproximadamente cinco nucleótidos.

[0059] En un ejemplo específico, la secuencia polipeptídica codificada por una secuencia de nucleótidos que tiene similitud considerable con una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma o una secuencia complementaria a la misma, incluye una sustitución de al menos un codón. En un ejemplo más específico, la sustitución es conservadora.

[0060] En un ejemplo específico, las secuencias polipeptídicas que tienen similitud considerable con la secuencia polipeptídica SEC ID N.º 2 o un fragmento, dominio, repetición o quimeras de la misma incluye una delección o inserción de al menos un aminoácido.

[0061] En un ejemplo específico, las secuencias polipeptídicas que tienen similitud considerable con la secuencia polipeptídica SEC ID N.º 2 o un fragmento, dominio, repetición o quimeras de la misma incluye una sustitución de al menos un aminoácido.

[0062] Los ejemplos de la presente invención también se refieren a un ácido nucleico combinado aleatoriamente que contiene varios fragmentos de secuencias de nucleótidos, en el que al menos uno de los fragmentos corresponde a una región de una secuencia de nucleótidos SEC ID N.º 1 y en el que al menos dos de los diversos fragmentos de secuencia están en un orden, de 5' a 3' que no es el orden en el que los diversos fragmentos se encuentran en la naturaleza en un ácido nucleico. En un ejemplo más específico, todos los fragmentos de ácido nucleico combinado aleatoriamente que contiene diversos fragmentos de secuencias de nucleótidos proceden de un único gen. En un ejemplo más

[0063] específico, los diversos fragmentos proceden de al menos dos genes diferentes. En un ejemplo más específico, el ácido nucleico combinado aleatoriamente está unido de forma operativa a una secuencia promotora. Otro ejemplo más específico es un polinucleótido quimérico que incluye una secuencia promotora unida de forma operativo al ácido nucleico combinado aleatoriamente. En un ejemplo más específico, el ácido nucleico combinado aleatoriamente está contenido dentro de una célula huésped.

[0064] Los ejemplos de la presente descripción también contemplan un casete de expresión que incluye una secuencia promotora unida de forma operativa a un ácido nucleico aislado que contiene una secuencia de nucleótidos que incluye:

a) una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma;

b) una secuencia de nucleótidos que codifica un polipéptido SEC ID N.º 2 o un fragmento o dominio de la misma;

c) una secuencia de nucleótidos que tiene similitud considerable con a) o b);

60

d) una secuencia de nucleótidos capaz de hibridar con a), b) o c);

e) una secuencia de nucleótidos que es complementaria a a), b), c) o d); o

65 f) una secuencia de nucleótidos que es la complementaria inversa de a), b), c) o d).

- [0065]** Además la invención abarca un vector recombinante que comprende un casete de expresión según las realizaciones de la presente invención. También abarca células vegetales, que contienen casetes de expresión, según la presente descripción, y plantas que contienen estas células vegetales. En una realización específica, la planta es una dicotiledónea. En una realización más específica, la dicotiledónea se selecciona a partir del grupo compuesto por soja, tabaco o algodón. En otra realización específica, la planta es una gimnosperma. En otra realización específica, la planta es una monocotiledónea. En una realización más específica, la monocotiledónea es un cereal. En una realización más específica, el cereal puede ser, por ejemplo, maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* y teosinte.
- 10 **[0066]** En una realización, el casete de expresión se expresa en toda la planta. En otra realización, el casete de expresión se expresa en una ubicación o tejido específico de una planta. En una realización específica, la ubicación o tejido puede ser, por ejemplo, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En una realización específica alternativa, la ubicación o el tejido es una semilla.
- 15 **[0067]** En una realización, el casete de expresión está implicado en una función, como por ejemplo, pero sin limitaciones, el metabolismo del carbono, nitrógeno y/o azufre, utilización del nitrógeno, asimilación del nitrógeno, fotosíntesis, transducción de señales, crecimiento celular, reproducción, resistencia a enfermedades, tolerancia al estrés abiótico, composición nutricional, regulación génica y/o diferenciación. En una realización más específica, el polipéptido quimérico está implicado en una función, como utilización del nitrógeno, tolerancia al estrés abiótico, rendimiento mejorado, resistencia a enfermedades y/o composición nutricional.
- 20 **[0068]** En una realización, la planta contiene una modificación de un fenotipo o característica mensurable de la planta, siendo la modificación atribuible a la expresión de al menos un gen contenido en el casete de expresión. En una realización específica, la modificación puede ser, por ejemplo, el metabolismo del carbono, nitrógeno y/o azufre, utilización del nitrógeno, asimilación del nitrógeno, fotosíntesis, transducción de señales, crecimiento celular, reproducción, resistencia a enfermedades, tolerancia al estrés abiótico, composición nutricional, regulación génica y/o diferenciación.
- 25 **[0069]** Los ejemplos de la presente invención también proporcionan semillas y productos aislados de plantas que contienen un casete de expresión que incluye una secuencia promotora unida de forma operativa a un ácido nucleico aislado que contiene una secuencia de nucleótidos que incluye:
- 30 a) una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma;
- 35 b) una secuencia de nucleótidos que codifica un polipéptido SEC ID N.º 2 o un fragmento o dominio de la misma;
- c) una secuencia de nucleótidos que tiene similitud considerable con a) o b);
- d) una secuencia de nucleótidos capaz de hibridar con a), b) o c);
- 40 e) una secuencia de nucleótidos que es complementaria a a), b), c) o d); o
- f) una secuencia de nucleótidos que es la complementaria inversa de a), b), c) o d) según la presente descripción.
- 45 **[0070]** En un ejemplo específico, el producto aislado incluye una enzima, una proteína nutricional, una proteína estructural, un aminoácido, un lípido, un ácido graso, un polisacárido, un azúcar, un alcohol, un alcaloide, un carotenoide, un propanoide, un esteroide, un pigmento, una vitamina y una hormona vegetal.
- 50 **[0071]** Los ejemplos de la presente invención también se refieren a productos aislados producidos mediante la expresión de un ácido nucleico aislado que contiene una secuencia de nucleótidos que incluye:
- a) una secuencia de nucleótidos SEC ID N.º 1 o un fragmento o dominio de la misma;
- b) una secuencia de nucleótidos que codifica un polipéptido SEC ID N.º 2 o un fragmento o dominio de la misma;
- 55 c) una secuencia de nucleótidos que tiene similitud considerable con a) o b);
- d) una secuencia de nucleótidos capaz de hibridar con a) o b);
- 60 e) una secuencia de nucleótidos que es complementaria a a), b), c) o d); o
- f) una secuencia de nucleótidos que es la complementaria inversa de a), b), c) o d) según la presente descripción.
- [0072]** En un ejemplo específico, el producto se produce en una planta. En otro ejemplo específico, el producto se produce en un cultivo celular. En otro ejemplo específico, el producto se produce en un sistema acelular. En otro ejemplo específico, el producto incluye una enzima, una proteína nutricional, una proteína estructural, un

aminoácido, un lípido, un ácido graso, un polisacárido, un azúcar, un alcohol, un alcaloide, un carotenoide, un propanoide, un esteroide, un pigmento, una vitamina y una hormona vegetal.

[0073] En un ejemplo específico, el producto es un polipéptido que contiene una secuencia de aminoácidos
5 SEC ID N.º 2. En una realización más específica, la proteína es un factor de transcripción.

[0074] Los ejemplos de la presente invención además se refieren a un polinucleótido aislado que incluye una
10 secuencia de nucleótidos de al menos 10 bases, cuya secuencia es idéntica, complementaria o considerablemente similar a una región de cualquier secuencia SEC ID N.º 1, y en la que el polinucleótido está adaptado para cualquier

[0075] En un ejemplo específico, el polinucleótido se usa como marcador cromosómico. En otro ejemplo
específico, el polinucleótido se usa como marcador para el análisis por RFLP. En otro ejemplo específico, el
15 polinucleótido se usa como marcador para la reproducción cuantitativa asociada al rasgo característico. En otro ejemplo específico, el polinucleótido se usa como marcador para la reproducción asistida por marcador. En otro ejemplo específico, el polinucleótido se usa como secuencia señuelo en un sistema de doble híbrido para identificar secuencias que codifican polipéptidos que interactúan con el polipéptido codificado por la secuencia señuelo. En otro ejemplo específico, el polinucleótido se usa como indicador diagnóstico para el genotipado o identificación de un individuo o población de individuos. En otro ejemplo específico, el polinucleótido se usa para el análisis genético
20 para identificar los límites de genes o exones.

[0076] Los ejemplos de la presente invención también se refieren a un vector de expresión que comprende o
consta de una molécula de ácido nucleico que incluye:

25 a) un ácido nucleico que codifica un polipéptido como se lista en SEC ID N.º 2;

b) un fragmento, uno o más dominios, o regiones de SEC ID N.º 1 con la característica especial; o

30 c) una secuencia de ácido nucleico completa listada en SEC ID N. 1 o un fragmento de la misma, en combinación con una secuencia heteróloga.

[0077] En una realización específica, el vector de expresión incluye uno o más elementos, como por ejemplo,
pero sin limitaciones, una secuencia promotora-potenciadora, una secuencia marcadora de selección, un origen de
replicación, una secuencia codificadora de epítipo etiquetado o una secuencia codificadora de purificación por
35 afinidad etiquetada. En una realización más específica, la secuencia promotora-potenciadora puede ser, por ejemplo, el promotor 35S de CaMV, el promotor 19S de CaMV, el promotor PR-1a del tabaco, el promotor de la ubiquitina y la faseolina. En otra realización, el promotor puede funcionar en plantas y, más específicamente, es un promotor constitutivo o inducible. En otra realización específica, la secuencia marcadora de selección codifica un gen de resistencia a antibiótico. En otra realización específica, la secuencia de epítipo etiquetado codifica V5, el
40 péptido Phe-His-His-Thr-Thr, hemaglutinina o glutatión-S-transferasa. En otra realización específica, la secuencia de purificación por afinidad etiquetada codifica una secuencia de ácido poliamino o un polipéptido. En una realización más específica, la secuencia de ácido poliamino es polihistidina. En una realización más específica, el polipéptido es el dominio de unión a quitina o glutatión-S-transferasa. En una realización más específica, la secuencia de purificación por afinidad etiquetada comprende una secuencia codificadora de inteína.

45 **[0078]** En una realización específica, el vector de expresión es un vector de expresión eucariota o un vector de expresión procariota. En una realización más específica, el vector de expresión eucariota incluye un promotor específico de tejido. Más específicamente, el vector de expresión puede funcionar en plantas.

50 **[0079]** Los ejemplos de la presente invención también se refieren a una célula que comprende o constan de una construcción de ácido nucleico que comprende un vector de expresión y un ácido nucleico que incluye un ácido nucleico que codifica un polipéptido como se lista en SEC ID N.º 2, o una secuencia de ácido nucleico listada en SEC ID N.º 1, o un segmento de la misma, en combinación con una secuencia heteróloga.

55 **[0080]** En una realización específica, la célula es una célula bacteriana, una célula fúngica, una célula vegetal o una célula animal. En una realización específica, la célula es una célula vegetal. En una realización más específica, el polipéptido se expresa en una ubicación o tejido específico de una planta. En una realización más específica, la ubicación o tejido puede ser, por ejemplo, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En una realización más específica alternativa, la ubicación o el tejido es una semilla. En
60 un ejemplo específico, el polipéptido está implicado en una función como, por ejemplo, el metabolismo del carbono, nitrógeno y/o azufre, utilización del nitrógeno, asimilación del nitrógeno, fotosíntesis, transducción de señales, crecimiento celular, reproducción, resistencia a enfermedades, tolerancia al estrés abiótico, composición nutricional, regulación génica y/o diferenciación.

65 **[0081]** Los ejemplos de la presente invención también se refieren a polipéptidos codificados por las moléculas aisladas de ácido nucleico de la presente descripción en las que se incluye un polipéptido que contiene una

secuencia polipeptídica codificada por un ácido nucleico aislado que contenga una secuencia de nucleótidos que incluya:

- 5 a) una secuencia de nucleótidos listada en SEC ID N.º 1 o un exón o dominio de la misma;
- b) una secuencia de nucleótidos que tiene similitud considerable con a);
- c) una secuencia de nucleótidos capaz de hibridar con a);
- 10 d) una secuencia de nucleótidos que es complementaria a a), b) o c); o
- e) una secuencia de nucleótidos que es la complementaria inversa de a), b) o c);
- f) o un fragmento funcional de la misma.

15 **[0082]** Un polipéptido que contiene una secuencia polipeptídica codificada por un ácido nucleico aislado que contiene una secuencia de nucleótidos, su complementaria o su complementaria inversa, que codifica un polipéptido que incluye una secuencia polipeptídica que incluye:

- 20 a) una secuencia polipeptídica listada en SEC ID N.º 2 o un dominio, repetición o quimeras de la misma;
- b) una secuencia polipeptídica que tiene similitud considerable con a);
- c) una secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o que tiene una similitud
- 25 considerable con una secuencia de nucleótidos listada en SEC ID N.º 1, o un exón o dominio de la misma, o una secuencia complementaria a la misma;
- d) una secuencia polipeptídica codificada por una secuencia de nucleótidos capaz de hibridar en condiciones de rigurosidad media con una secuencia de nucleótidos listada en SEC ID N.º 1 o con una secuencia complementaria
- 30 de la misma; o
- e) un fragmento funcional de a), b), c) o d);
- f) o un fragmento funcional de la misma.

35 **[0083]** Los ejemplos de la presente descripción contemplan un polipéptido que contiene una secuencia polipeptídica codificada por un ácido nucleico aislado que incluye un ácido nucleico combinado aleatoriamente que contiene varios fragmentos de secuencias de nucleótidos, en el que al menos uno de los fragmentos corresponde a una región de una secuencia de nucleótidos listada en SEC ID N.º 1 y en el que al menos dos de los diversos

40 fragmentos de secuencia están en un orden, de 5' a 3' que no es el orden en el que los diversos fragmentos se dan en la naturaleza en un ácido nucleico, o fragmento funcional de la misma.

[0084] Los ejemplos de la presente descripción contemplan un polipéptido que contiene una secuencia polipeptídica codificada por un polinucleótido aislado que contiene una secuencia de nucleótidos de al menos 10

45 bases, cuya secuencia es idéntica, complementaria o considerablemente similar a una región de cualquier secuencia SEC ID N.º 1, o un fragmento funcional de la misma y en la que el polinucleótido está adaptado para un uso que incluye:

- 50 a) uso de un marcador cromosómico para identificar la ubicación del polinucleótido correspondiente o complementario en un cromosoma nativo o artificial;
- b) uso como marcador para el análisis por RFLP;
- c) uso como marcador para la reproducción cuantitativa asociada al rasgo característico;
- 55 d) uso como marcador para la reproducción asistida por marcador;
- e) uso como secuencia señuelo en un sistema de doble híbrido para identificar secuencias que codifican polipéptidos que interaccionan con el polipéptido codificado por la secuencia señuelo;
- 60 f) uso como indicador diagnóstico para el genotipado o identificación de un individuo o población de individuos; o
- g) uso para el análisis genético para identificar límites de genes o exones.

65 **[0085]** Los ejemplos de la presente descripción también contemplan un polipéptido aislado que contiene una secuencia polipeptídica que incluye:

a) una secuencia polipeptídica listada en SEC ID N.º 2 o un exón o dominio de la misma;

b) una secuencia polipeptídica que tiene similitud considerable con a);

5

c) una secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o que tiene una similitud considerable con la secuencia de nucleótidos SEC ID N.º 1, o un exón o dominio de la misma, o una secuencia complementaria a la misma;

10 d) una secuencia polipeptídica codificada por una secuencia de nucleótidos capaz de hibridar en condiciones de rigurosidad media con una secuencia de nucleótidos listada en SEC ID N.º 1 o con una secuencia complementaria de la misma; o

e) un fragmento funcional de a), b), c) o d).

15

[0086] En un ejemplo específico, la similitud considerable es de al menos aproximadamente el 65% de identidad. En un ejemplo más específico, la similitud considerable es de al menos aproximadamente el 80% de identidad. En un ejemplo más específico, la similitud considerable es de al menos aproximadamente el 95% de identidad. En un ejemplo específico, la similitud considerable es al menos un tres por ciento mayor que el porcentaje de identidad con la secuencia homóloga más próxima listada en cualquiera de los listados de secuencias.

20

[0087] En un ejemplo específico, la secuencia que tiene similitud considerable es de una planta. En un ejemplo más específico, la planta es una dicotiledónea. En un ejemplo más específico, la planta es una gimnosperma. En un ejemplo más específico, la planta es una monocotiledónea. En un ejemplo más específico, la monocotiledónea es un cereal. En un ejemplo más

25

[0088] específico, el cereal puede ser, por ejemplo, maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* y teosinte.

30 **[0089]** En un ejemplo específico el polipéptido se expresa en una ubicación o tejido específico de una planta. En un ejemplo más específico, la ubicación o tejido puede ser, por ejemplo, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En otro ejemplo específico, la ubicación o el tejido es una semilla. En un ejemplo específico, el polipéptido está implicado en una función como, por ejemplo, el metabolismo del carbono, nitrógeno y/o azufre, utilización del nitrógeno, asimilación del nitrógeno, fotosíntesis, transducción de señales, crecimiento celular, reproducción, resistencia a enfermedades, tolerancia al estrés abiótico, composición

35

[0090] En un ejemplo específico, la hibridación de una secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o con similitud considerable con una secuencia de nucleótidos listada en SEC ID N.º 1, o un exón o dominio de la misma, o una secuencia complementaria de la misma, o una secuencia polipeptídica codificada por una secuencia de nucleótidos capaz de hibridar en condiciones de rigurosidad media con una secuencia de nucleótidos listada en SEC ID N.º 1, o con una secuencia complementaria a la misma, permite que la secuencia forme un dúplex en condiciones de rigurosidad media o alta.

40

45 **[0091]** En un ejemplo específico, un polipéptido que tiene similitud considerable con una secuencia polipeptídica listada en SEC ID N.º 2, o un exón o dominio de la misma, es una variante alélica de la secuencia polipeptídica listada en SEC ID N.º 2. En otro ejemplo específico, un polipéptido que tiene similitud considerable con una secuencia polipeptídica listada en SEC ID N.º 2, o un exón o dominio de la misma, es una variante natural de la secuencia polipeptídica listada en SEC ID N.º 2. En otro ejemplo específico, un polipéptido que tiene similitud considerable con una secuencia polipeptídica listada en SEC ID N.º 2, o un exón o dominio de la misma, es una variante polimórfica de la secuencia polipeptídica listada en SEC ID N.º 2.

50

[0092] En un ejemplo específico alternativo, la secuencia que tiene similitud considerable contiene una deleción o inserción de al menos un aminoácido. En un ejemplo más específico, la deleción o inserción es de menos de aproximadamente diez aminoácidos. En un ejemplo más específico, la deleción o inserción es de menos de aproximadamente tres aminoácidos.

55

[0093] En un ejemplo específico, la secuencia que tiene similitud considerable codifica una sustitución en al menos un aminoácido.

60

[0094] También se contempla un procedimiento de producción de una planta que comprende una modificación del mismo, que incluye las etapas de: 1) proporcionar un ácido nucleico que es un ácido nucleico aislado que contiene una secuencia de nucleótidos que incluye:

65 a) una secuencia de nucleótidos listada en SEC ID N.º 1 o un exón o dominio de la misma;

b) una secuencia de nucleótidos que tiene similitud considerable con a);

c) una secuencia de nucleótidos capaz de hibridar con a);

5 d) una secuencia de nucleótidos que es complementaria a a), b) o c); o

e) una secuencia de nucleótidos que es la complementaria inversa de a), b) o c);

10 y 2) introducir el ácido nucleico en la planta, en el que el ácido nucleico se puede expresar en la planta en una cantidad eficaz para afectar a la modificación. En un ejemplo, la modificación comprende una característica alterada en la planta, en la que la característica corresponde con el ácido nucleico introducido en la planta. En otros ejemplos específicos, la característica corresponde al metabolismo del carbono, nitrógeno y/o azufre, utilización del nitrógeno, asimilación del nitrógeno, fotosíntesis, transducción de señales, crecimiento celular, reproducción, resistencia a enfermedades, tolerancia al estrés abiótico, composición nutricional, regulación génica y/o diferenciación.

15 **[0095]** En otro ejemplo, la modificación incluye un aumento o disminución de la expresión o la acumulación de un producto de la planta. Específicamente, el producto es un producto natural de la planta. De igual modo específicamente, el producto es un producto nuevo o alterado de la planta. Específicamente, el producto comprende un factor de transcripción GATA.

20 **[0096]** La descripción también abarca un procedimiento de producción de una proteína recombinante, que comprende las etapas de:

25 a) crecer células recombinantes que comprenden una construcción de ácido nucleico en condiciones de crecimiento idóneas, comprendiendo la construcción un vector de expresión y un ácido nucleico que incluye: un ácido nucleico que codifica una proteína como se lista en SEC ID N. 2, o una secuencia de ácido nucleico listada en SEC ID N.º 1 o segmentos de la misma; y

30 b) aislar a partir de las células recombinantes la proteína recombinante expresada por las mismas.

35 **[0097]** Los ejemplos de la presente descripción proporcionan un procedimiento de producción de una proteína recombinante en el que el vector de expresión incluye uno o más elementos que incluyen una secuencia promotora-potenciadora, una secuencia marcadora de selección, un origen de replicación, una secuencia codificadora de epítipo etiquetado y una secuencia codificadora de purificación por afinidad etiquetada. En un ejemplo específico, la construcción de ácido nucleico incluye una secuencia codificadora de epítipo etiquetado y la etapa de aislamiento incluye el uso de un anticuerpo específico para el epítipo etiquetado. En otro ejemplo específico, la construcción de ácido nucleico contiene una secuencia codificadora de ácido poliamino y la etapa de aislamiento incluye el uso de una resina que comprende una sustancia de unión a ácido poliamino, específicamente cuando el ácido poliamino es polihistidina y la resina de unión a poliamino es una resina de agarosa cargada con níquel. Aún en otro ejemplo específico, la construcción de ácido nucleico contiene una secuencia codificadora de un polipéptido y la etapa de aislamiento incluye el uso de una resina que contiene una sustancia de unión a polipéptido, específicamente cuando el polipéptido es un dominio de unión a quitina y la resina contiene quitina-seferosa.

45 **[0098]** Las realizaciones de la presente invención también se refieren a una planta modificada por un procedimiento que incluye la introducción en una planta de un ácido nucleico en el que el ácido nucleico se puede expresar en la planta en una cantidad eficaz para realizar la modificación. La modificación puede ser, por ejemplo, el metabolismo del carbono, nitrógeno y/o azufre, utilización del nitrógeno, asimilación del nitrógeno, fotosíntesis, transducción de señales, crecimiento celular, reproducción, resistencia a enfermedades, tolerancia al estrés abiótico, composición nutricional, regulación génica y/o diferenciación. En un ejemplo, la planta modificada tiene resistencia 50 aumentada o disminuida a un herbicida, un estrés o un patógeno. En otro

[0099] ejemplo, la planta modificada tiene mayores o menores requisitos por la luz, agua, nitrógeno u oligoelementos. Aún en otra realización, la planta modificada está enriquecida con un aminoácido esencial como una proporción de una fracción de proteína de la planta. La fracción de proteína puede ser, por ejemplo, proteína total de 55 semilla, proteína soluble, proteína insoluble, proteína extraíble en agua y proteína asociada a lípidos. La modificación puede incluir sobreexpresión, subexpresión, modulación de la cadena complementaria, supresión de la cadena sentido, expresión inducible, represión inducible o modulación inducible de un gen.

60 **[0100]** La descripción además se refiere a una semilla de una planta modificada o un producto aislado de una planta modificada, donde el producto puede ser una enzima, una proteína nutricional, una proteína estructural, un aminoácido, un lípido, un ácido graso, un polisacárido, un azúcar, un alcohol, un alcaloide, un carotenoide, un propanoide, un esteroide, un pigmento, una vitamina y una hormona vegetal.

[0101] En el resumen de la invención anterior se recogen varias realizaciones de la invención y, en muchos 65 casos, se recogen variaciones y permutaciones de estas realizaciones. El resumen constituye meramente ejemplos de los numerosos y distintos ejemplos. La mención de una o más características específicas de un ejemplo dado es

de algún modo ejemplificador. Dicho ejemplo puede existir normalmente con o sin la característica o características mencionadas; del mismo modo, esas características se pueden aplicar a otros ejemplos de la descripción, se mencionen o no este resumen. Para evitar la repetición excesiva, este resumen no menciona ni sugiere todas las posibles combinaciones de dichas características.

5

[0102] Con el fin de resumir la invención y las ventajas conseguidas sobre la técnica anterior, se han descrito a continuación ciertos objetos o ventajas de la invención. Por supuesto, se debe entender que todos estos objetos y ventajas no necesariamente pueden conseguirse de acuerdo con cualquiera de los ejemplos particulares de la invención. De este modo, por ejemplo, los expertos en la materia reconocerán que la invención puede ser realizada o llevada a cabo de forma que se consiga u optimice una ventaja o grupo de ventajas como se ha enseñado en este documento sin conseguir necesariamente otros objetos o ventajas como los que se pueden enseñar o sugerir en este documento.

10

[0103] Aspectos, características y ventajas adicionales de esta invención se harán aparentes a partir de la descripción detallada de las realizaciones específicas que siguen.

15

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0104]

20

En la figura 1 y en SEC ID N.º 1 se muestra la secuencia CDS de At5g56860 de longitud completa.

En la figura 2 y en SEC ID N.º 2 se muestra la secuencia de aminoácidos de At5g5680.

25 En la figura 3 y en SEC ID N.º 3 se muestra la secuencia del ácido nucleico del ortólogo del arroz de At5g56860.

La figura 4 es una construcción del plásmido que contiene el promotor 35S de CaMV, en ADNc de GNC y terminador de la nopalina sintetasa (nos) entre los límites derecho (BD) e izquierdo (BI) del T-ADN.

30 DEFINICIONES

[0105] Por claridad, determinados términos utilizados en la descripción se definen y presentan como sigue: «asociado con/unido de forma operativa» se refiere a dos secuencias de ácido nucleico que están física o funcionalmente relacionadas. Por ejemplo, una secuencia de ADN promotora o reguladora se dice que está «asociada con» una secuencia de ADN que codifica un ARN o una proteína si las dos secuencias están unidas de forma operativa, o están situadas de forma que la secuencia del ADN reguladora afectará al nivel de expresión de la secuencia de ADN codificadora o estructural.

35

[0106] Una «construcción quimérica» es una secuencia de ácido nucleico recombinante en la que una secuencia de ácido nucleico promotora o reguladora está unida de forma operativa, o asociada con, una secuencia de ácido nucleico que codifica un ARNm o que se expresa como proteína, de modo que la secuencia de ácido nucleico reguladora es capaz de regular la transcripción o expresión de la secuencia de ácido nucleico asociada. La secuencia de ácido nucleico reguladora de la construcción quimérica normalmente no está unida de forma operativa a la secuencia de ácido nucleico asociada cuando se encuentra en la naturaleza.

45

[0107] Un «cofactor» es un reactivo natural, tal como una molécula orgánica o un ión metálico, necesario en una reacción catalizada realizada por una enzima. Un cofactor es, por ejemplo, NAD(P), riboflavina (incluyendo FAD y FMN), folato, molibdopterina, tiamina, biotina, ácido lipoico, ácido pantoténico y coenzima A, S-adenosilmetionina, fosfato de piridoxal, ubiquinona o menaquinona. Opcionalmente, un cofactor puede regenerarse y reutilizarse.

50

[0108] Una «secuencia codificadora» es una secuencia de ácido nucleico que se transcribe en ARN como ARNm, ARNr, ARNt, ARNnp, ARN sentido o ARN complementario. Específicamente, el ARN se traduce a continuación en un organismo para producir una proteína.

55 **[0109]** Complementaria: «complementaria» se refiere a dos secuencias de nucleótidos que comprenden secuencias de nucleótidos antiparalelas capaces de aparearse entre sí tras la formación de puentes de hidrógeno entre los restos base complementarios en las secuencias de nucleótidos antiparalelas.

[0110] Actividad enzimática: en este documento significa la capacidad de una enzima para catalizar la conversión de un sustrato en un producto. El sustrato para la enzima comprende no solo el sustrato natural de dicha enzima sino que también comprende análogos del sustrato natural, que también pueden ser convertidos por la enzima en un producto o un análogo del producto. La actividad de la enzima se mide, por ejemplo, determinando la cantidad de producto en la reacción tras un determinado período de tiempo, o determinando la cantidad de sustrato remanente en la mezcla de reacción tras un determinado período de tiempo. La actividad de la enzima también puede medirse determinando la cantidad de un cofactor no utilizado de la reacción que permanece en la mezcla de reacción tras un determinado período de tiempo o determinando la cantidad de cofactor utilizado en la mezcla de

60

65

reacción tras un determinado período de tiempo. La actividad de la enzima también se mide determinando la cantidad de un donante de energía libre o molécula rica en energía (p. ej., ATP, fosfoenolpiruvato, fosfato de acetilo o fosfocreatina) que permanece en la mezcla de reacción tras un determinado período de tiempo o determinando la cantidad de un donante de energía libre usado o molécula rica en energía (p. ej., ADP, piruvato, acetato o creatina) en la mezcla de reacción tras un determinado período de tiempo.

[0111] Casete de expresión: el «casete de expresión» como se utiliza en este documento significa una molécula de ácido nucleico capaz de dirigir la expresión de una secuencia de nucleótidos en particular en una célula huésped adecuada que comprende un promotor unido de forma operativa a la secuencia de nucleótidos de interés, la cual está unida de forma operativa a señales de terminación. Típicamente también comprende secuencias requeridas para la traducción adecuada de la secuencia de nucleótidos. La región codificadora normalmente codifica una proteína de interés, aunque también puede codificar un ARN funcional de interés, por ejemplo, un ARN complementario o un ARN no traducido, en la dirección sentido o en la complementaria. El casete de expresión que comprende la secuencia de nucleótidos de interés puede ser quimérico, lo que significa que al menos uno de sus componentes es heterólogo con respecto a al menos uno de los demás componentes. El casete de expresión también puede ser aquel que aparece en la naturaleza pero se ha obtenido en una forma recombinante útil para la expresión heteróloga. Sin embargo, típicamente el casete de expresión es heterólogo con respecto al huésped, es decir, la secuencia de ADN del casete de expresión en particular no se encuentra de forma natural en la célula huésped y debe haber sido introducido en la célula huésped o en un predecesor de la célula huésped mediante un acontecimiento de transformación. La expresión de la secuencia de nucleótidos en el casete de expresión puede estar bajo el control de un promotor constitutivo o de un promotor inducible que inicia la transcripción solo cuando la célula huésped se expone a algunos estímulos externos en particular. En el caso de un organismo multicelular, como una planta, el promotor también puede ser específico de un tejido, órgano o etapa de desarrollo en particular.

[0112] El término «fragmento funcional» según se usa en este documento en relación con una secuencia de ácido nucleico o de proteína significa un fragmento o porción de la secuencia que retiene la función de la secuencia de longitud completa.

[0113] Gen: el término «gen» se usa ampliamente para referirse a cualquier segmento de ADN asociado con una función biológica. De este modo, los genes incluyen las secuencias codificadoras y/o las secuencias reguladoras requeridas para su expresión. Los genes también incluyen segmentos de ADN no expresados que, por ejemplo, forman secuencias de reconocimiento para otras proteínas. Los genes pueden obtenerse a partir de diversas fuentes, incluyendo la clonación de una fuente de interés, o sintetizarse a partir de la información de la secuencia conocida o predicha, y puede incluir secuencias diseñadas para tener los parámetros deseados.

[0114] Heterólogo/exógeno: los términos «heterólogo» y «exógeno» cuando se usan en este documento para referirse a una secuencia de ácido nucleico (p. ej., una secuencia de ADN) o un gen, se refieren a una secuencia que se origina a partir de una fuente ajena a la célula huésped en particular o, si es de la misma fuente, se modifica a partir de su forma original. Por tanto, un gen heterólogo en una célula huésped incluye un gen que es endógeno a la célula huésped en particular pero se ha modificado mediante, por ejemplo, el uso de combinación aleatoria de ADN. Los términos también incluyen múltiples copias no naturales de una secuencia de ADN natural. Por tanto, el término se refiere a un segmento de ADN que es extraño o heterólogo para la célula u homólogo para la célula pero en una posición dentro del ácido nucleico de la célula huésped en la que no se encuentra normalmente. Los segmentos de ADN exógeno se expresan hasta producir polipéptidos exógenos.

[0115] Una secuencia de ácido nucleico (p. ej., ADN) «homóloga» es una secuencia de ácido nucleico (p. ej., ADN) asociada de forma natural con una célula huésped en la que se va a introducir.

[0116] Hibridación: la frase «que hibrida específicamente con» se refiere a la unión, formación de un dúplex o hibridación de una molécula solo con un ácido nucleico en particular en condiciones rigurosas cuando esta secuencia está presente en una mezcla compleja (por ejemplo, celular total) ADN o ARN. La expresión «se une sustancialmente» se refiere a hibridación complementaria entre una sonda de ácido nucleico y un ácido nucleico diana y abarca errores menores que pueden acomodarse reduciendo la rigurosidad de los medios de hibridación para conseguir la detección deseada de la secuencia de ácido nucleico diana.

[0117] Inhibidor: sustancia química que inactiva la actividad enzimática de una proteína como una enzima biosintética, receptor, proteína de transducción de señales, producto génico estructural o proteína de transporte. El término «herbicida» (o «compuesto herbicida») se usa en este documento para definir un inhibidor aplicado a una planta en cualquier etapa del desarrollo, en el que el herbicida inhibe el crecimiento de la planta o la mata.

[0118] Interacción: calidad o estado de acción mutua de manera que la eficacia o toxicidad de una proteína o compuesto sobre otra proteína es inhibitoria (antagonistas) o potenciadora (agonistas).

[0119] Una secuencia de ácido nucleico es «isocodificadora con» una secuencia de ácido nucleico de referencia cuando la secuencia de ácido nucleico codifica un polipéptido que tiene la misma secuencia de aminoácidos que el polipéptido codificado por la secuencia del ácido nucleico de referencia.

[0120] Isogénicas: plantas que son genéticamente idénticas, excepto porque pueden diferir en la presencia o ausencia de una secuencia de ADN heteróloga.

5 **[0121]** Aislada: en el contexto de la presente invención, una molécula de ADN aislada o una enzima aislada es una molécula de ADN o enzima que, por intervención humana, puede encontrarse fuera de su entorno nativo y, por tanto, no es un producto natural. Una molécula de ADN o enzima aislada puede estar en forma purificada o puede estar en un entorno no nativo como, por ejemplo, una célula huésped transgénica.

10 **[0122]** Proteína madura: proteína de la cual se han eliminado el péptido de tránsito, el péptido señal y/o porciones propeptídicas.

[0123] Promotor mínimo: fragmento más pequeño de un promotor, como un elemento TATA, que puede ser transcrito. Un promotor mínimo típicamente tiene una actividad promotora muy reducida en ausencia de activación
15 antes del extremo 5'. En presencia de un factor de transcripción adecuado, el promotor mínimo funciona permitiendo la transcripción.

[0124] Actividad enzimática modificada: actividad enzimática diferente de la que aparece de forma natural en una planta (es decir, la actividad enzimática que aparece de forma natural en ausencia de manipulación directa o
20 indirecta de dicha actividad por el hombre), que tolera los inhibidores que inhiben la actividad enzimática natural.

[0125] Nativo: se refiere a un gen presente en el genoma de una célula vegetal no transformada.

[0126] De origen natural: el término «de origen natural» se usa para describir un objeto que puede
25 encontrarse en la naturaleza para distinguirlo del producido artificialmente por el hombre. Por ejemplo, una proteína o secuencia de nucleótidos presente en un organismo (incluyendo virus) que puede aislarse a partir de una fuente natural y que el hombre no ha modificado de manera intencionada en el laboratorio es de origen natural.

[0127] Ácido nucleico: el término «ácido nucleico» se refiere a desoxirribonucleótidos o ribonucleótidos y a
30 polímeros de los mismos en forma de cadena sencilla o doble. Siempre que no esté específicamente limitado, el término abarca ácidos nucleicos que contienen análogos de nucleótidos naturales con propiedades de unión similares a las del ácido nucleico de referencia y que se metabolizan de forma similar a los nucleótidos de origen natural. Siempre que no se indique otra cosa, una secuencia de ácido nucleico en particular también abarca implícitamente variantes modificadas de forma conservadora de la misma (p. ej., sustituciones de codon
35 degenerado) y secuencias complementarias así como la secuencia explícitamente indicada. Específicamente, las sustituciones de codon degenerado pueden lograrse generando secuencias en las que la tercera posición de uno o más (o todos) codones seleccionados se sustituye por restos con mezclas de bases y desoxiinosina (Batzner y col., Nucleic Acid Res. 19: 5081 (1991); Ohtsuka y col., J. Biol. Chem. 260: 2605-2608 (1985); Rossolini y col., Mol. Cell. Probes 8: 91-98 (1994)). Los términos «ácido nucleico» o «secuencia de ácido nucleico» también pueden usarse de
40 forma indistinta con gen, ADNc y ARNm codificado por un gen.

[0128] «ORF» significa marco de lectura abierta.

[0129] Porcentaje de identidad: las expresiones «porcentaje de identidad» o «porcentaje idéntico» en el
45 contexto de dos secuencias de ácido nucleico o proteína, se refieren a dos o más secuencias o subsecuencias que tienen, por ejemplo, el 60%, específicamente el 70%, más específicamente el 80%, aún más específicamente el 90%, incluso más específicamente el 95% y más específicamente al menos el 99% de identidad de nucleótidos o restos de aminoácidos cuando se comparan y alinean para una correspondencia máxima, según se determina usando uno de los algoritmos de comparación o mediante inspección visual. Específicamente, se observa a lo largo
50 de una región de las secuencias que el porcentaje de identidad es al menos de aproximadamente 50 restos de longitud, más específicamente a lo largo de una región de al menos aproximadamente 100 restos o más específicamente, el porcentaje de identidad se observa en al menos aproximadamente 150 restos. En una realización especialmente específica, el porcentaje de identidad existe a lo largo de la longitud completa de las regiones codificadoras.

55 **[0130]** Para la comparación de secuencia, típicamente una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencia, se introducen en un ordenador las secuencias de ensayo y de referencia, se diseñan coordenadas en consecuencia y, si es necesario, se diseñan parámetros para el programa de algoritmo de secuencia. A continuación, el algoritmo de
60 comparación de secuencias calcula el porcentaje de identidad de secuencia para la secuencia o secuencias de ensayo con respecto a la secuencia de referencia, en función de los parámetros especificados del programa.

[0131] El alineamiento óptimo de secuencias para comparación puede llevarse a cabo, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, Adv. Appl. Math, 2: 482 (1981), mediante el algoritmo de
65 alineamiento de homología de Needleman y Wunsch, J. Mol. Biol. 48: 443 (1970), mediante la búsqueda para el procedimiento de similitud de Pearson y Lipman. Proc. Nat'l. Acad. Sci. USA 85: 2444 (1988), mediante

implementaciones por ordenador de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el paquete de software para genética de Wisconsin del Genetics Computer Group, 575 Science Dr., Madison, WI) o mediante inspección visual (véase en general, Ausubel y col., a continuación).

5 **[0132]** Un ejemplo de algoritmo adecuado para determinar el porcentaje de identidad de secuencia y similitud de secuencia es el algoritmo BLAST, que se describe en Altschul y col., J. Mol. Biol. 215: 403-410 (1990). El software para realizar análisis BLAST está a disposición de público a través del Centro Nacional de Información Biotecnológica (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica en primer lugar identificar los pares de secuencia de mayor puntuación (HSP, por sus siglas en inglés) identificando las palabras cortas de longitud W en la
10 secuencia problema, que coinciden o satisfacen alguna puntuación T de umbral con valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. T se refiere como el valor umbral de la palabra vecina (Altschul y col., 1990). Estas coincidencias de palabras vecinas iniciales actúan como semillas para iniciar búsquedas para encontrar HSP las largas que las contengan. A continuación, las palabras coincidentes se extienden en ambas direcciones a lo largo de cada secuencia tanto como pueda incrementarse el valor del
15 alineamiento acumulado. Los valores acumulados se calculan usando, para la secuencias de nucleótidos, los parámetros M (valores de recompensas para un par de restos coincidentes; siempre > 0) y N (valor de penalización por restos no coincidentes, siempre < 0). Para las secuencias de aminoácidos, se usa una matriz de valoración para calcular el valor acumulativo. La extensión de las palabras coincidentes en cada dirección se detiene cuando el valor de alineamiento acumulativo desciende hasta la cantidad X desde su valor máximo alcanzado; el valor acumulativo
20 desciende hasta cero o por debajo debido a la acumulación de uno o más alineamientos de restos negativos valorados o se alcanza el final de otra secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) usa por defecto una longitud de palabra (W) de 11, un valor esperado (E) de 10 y un punto de corte de 100, M=5, N=-4 y una comparación de ambas cadenas. Para las secuencias de aminoácidos, el programa BLASTP usa por defecto una
25 longitud de palabra (W) de 3, un valor esperado (E) de 10 y la matriz de valoración BLOSUM62 (véase Henikoff y Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989)).

[0133] Además del porcentaje de identidad de secuencia calculado, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul., Proc. Nat'l. Acad. Sci.
30 USA 90: 5873-5787 (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la suma de probabilidad menor (P(N)), que proporciona una indicación de la probabilidad por la cual podría darse por casualidad una coincidencia entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, se considera que una secuencia de ácido nucleico de muestra es similar a una secuencia de referencia si la suma de la probabilidad menor en una comparación de la secuencia de ácido nucleico de ensayo con la secuencia de ácido nucleico de referencia es
35 menor de aproximadamente 0,1, más específicamente menor de aproximadamente 0,01 y aún más específicamente, menos de aproximadamente 0,001.

[0134] Preproteína: proteína que normalmente se dirige a un orgánulo celular, como un cloroplasto, y sigue comprendiendo su péptido de tránsito nativo.
40

[0135] Purificado: el término «purificado» cuando se aplica a un ácido nucleico o proteína, indica que el ácido nucleico o la proteína está esencialmente libre de otros componentes celulares con los que se asocia en estado natural. Dicho ácido nucleico o proteína está específicamente en un estado homogéneo aunque puede estar seco o en solución acuosa. La pureza y homogeneidad se determinan típicamente usando técnicas químicas analíticas
45 como electroforesis en geles de poliacrilamida o cromatografía líquida de alta resolución. Una proteína que es la especie predominante presente en una preparación está sustancialmente purificada. El término «purificado» indica que un ácido nucleico o proteína da lugar esencialmente a una banda en un gel de electroforesis. Especialmente, esto significa que el ácido nucleico o proteína está al menos aproximadamente puro al 50%, más específicamente, al menos aproximadamente puro al 85% y, más específicamente, al menos aproximadamente puro al 99%.
50

[0136] Dos ácidos nucleicos se han «recombinado» cuando las secuencias de cada uno de los dos ácidos nucleicos se combinan en un ácido nucleico progenie. Dos secuencias están «directamente» recombinadas cuando ambos ácidos nucleicos son sustratos para la recombinación. Dos secuencias están «recombinadas indirectamente» cuando las secuencias se recombinan usando un intermediario como un oligonucleótido entrecruzado. Para la
55 recombinación indirecta, no más de una de las secuencias es un sustrato real para la recombinación y, en algunos casos, ninguna secuencia es sustrato para la recombinación.

[0137] «Elementos reguladores» se refieren a secuencias implicadas en el control de la expresión de una secuencia de nucleótidos. Los elementos reguladores comprenden un promotor unido de forma operativa a la
60 secuencia de nucleótidos de interés y la señales de terminación. Típicamente, también abarcan secuencias requeridas para una traducción adecuada de la secuencia de nucleótidos.

[0138] Aumento significativo: aumento en la actividad enzimática que es mayor que el margen de error inherente a la técnica de medición, específicamente un aumento de aproximadamente 2 veces o mayor de la
65 actividad de la enzima nativa en presencia del inhibidor, más específicamente un aumento de aproximadamente 5 veces o mayor y, más específicamente, un aumento de aproximadamente 10 veces o más.

[0139] Significativamente menor: significa que la cantidad de un producto de una reacción enzimática se reduce más que el margen de error inherente a la técnica de medición, específicamente una disminución de aproximadamente 2 veces o mayor de la actividad de la enzima nativa en ausencia del inhibidor, más específicamente una disminución de aproximadamente 5 veces o mayor y, más específicamente, una disminución de aproximadamente 10 veces o mayor.

[0140] Unión específica/reactividad inmunológica cruzada: indicación de que dos secuencias de ácido nucleico o proteínas son sustancialmente idénticas si la proteína codificada por el primer ácido nucleico presenta reacción inmunológica cruzada con la proteína codificada (o se une específicamente a ella) por el segundo ácido nucleico. De este modo, típicamente una proteína es sustancialmente idéntica a una segunda proteína, por ejemplo, cuando las dos proteínas difieren sólo en las sustituciones conservadoras. La frase «se une específicamente (o selectivamente) a un anticuerpo» o «es inmunorreactivo específicamente (o selectivamente) con» cuando se refiere a una proteína o péptido, se refiere a una reacción de unión que es determinativa de la presencia de la proteína en presencia de una población heteróloga de proteínas y otros compuestos biológicos. Por tanto, en las condiciones de inmunoensayo indicadas, los anticuerpos específicos se unen a una proteína en particular y no se unen en una cantidad significativa a otras proteínas presentes en la muestra. La unión específica a un anticuerpo en estas condiciones puede requerir un anticuerpo que se selecciona por su especificidad por una proteína en particular. Por ejemplo, pueden seleccionarse anticuerpos formados frente a la proteína con la secuencia de aminoácidos codificada por cualquiera de las secuencias de ácido nucleico de la invención para obtener anticuerpos específicamente inmunorreactivos con esta proteína y no con otras proteínas, excepto en caso de variantes polimórficas. Pueden usarse diversos formatos de inmunoensayos para seleccionar anticuerpos específicamente inmunorreactivos con una proteína en especial. Por ejemplo, los inmunoensayos de tipo ELISA en fase sólida, las inmunotransferencias o las técnicas de inmunohistoquímica se usan de forma rutinaria para seleccionar anticuerpos monoclonales específicamente inmunorreactivos con una proteína. Véase Harlow y Lane (1988) *Antibodies, A Laboratory Manual*, Publicaciones Cold Spring Harbor, Nueva York («Harlow y Lane»), para una descripción de los formatos de inmunoensayos y las condiciones que pueden usarse para determinar la inmunorreactividad específica. Típicamente, una reacción específica o selectiva será al menos el doble de la señal de fondo o ruido y, más típicamente, más de 10 a 100 veces el fondo.

[0141] «Condiciones de hibridación rigurosas» y «condiciones de lavado de hibridación rigurosas» en el contexto de los experimentos de hibridación de ácidos nucleicos, tales como hibridaciones de ADN y ARN, son secuencias dependientes y son diferentes bajo parámetros ambientales diferentes. Las secuencias más largas hibridan específicamente a temperaturas más altas. En Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization with Nucleic Acid Probes* parte I, capítulo 2, «Overview of principles of hybridization and the strategy of nucleic acid probe assays», Elsevier, Nueva York, se encuentra una amplia guía para la hibridación de ácidos nucleicos. Generalmente, la hibridación y las condiciones de lavado altamente rigurosas se seleccionan para que estén a aproximadamente 5° C por debajo del punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y pH definidos. Típicamente, en «condiciones rigurosas» una sonda hibridará con su subsecuencia diana, pero no otras secuencias.

[0142] La T_m es la temperatura (en condiciones de fuerza iónica y pH definidas) a la cual el 50% de la secuencia diana hibrida con una sonda con la que coincide a la perfección. Las condiciones muy rigurosas se seleccionan para que sean iguales a la T_m para una sonda en particular. Un ejemplo de condiciones rigurosas de hibridación para la hibridación de ácidos nucleicos complementarios que tienen más de 100 restos complementarios sobre un filtro en transferencia de ARN o ADN es formamida al 50% con 1 mg de heparina a 42°C, con la hibridación realizada durante una noche. Un ejemplo de condiciones de lavado altamente rigurosas es NaCl 0.15 M a 72° C durante aproximadamente 15 minutos. Un ejemplo de condiciones de lavado rigurosas es el lavado en SSC0.2x a 65° C durante 15 minutos (véase Sambrook, a continuación, para una descripción del tampón SSC). A menudo, un lavado altamente riguroso va precedido por un lavado poco riguroso para eliminar la señal de fondo de la sonda. Un ejemplo de lavado en condiciones de rigurosidad media para la formación de un dúplex, por ejemplo, de más de 100 nucleótidos, es SSC 1x a 45°C durante 15 minutos. Un ejemplo de lavado en condiciones de rigurosidad media para la formación de un dúplex, por ejemplo, de más de 100 nucleótidos, es SSC 4-6x a 40°C durante 15 minutos. Para sondas cortas (p. ej., aproximadamente de 10 a 50 nucleótidos) las condiciones rigurosas típicamente suponen concentraciones de sales de menos de aproximadamente 1,0 M de ión Na, típicamente una concentración de ión Na de aproximadamente 0,01 a 1,0 M (u otras sales) a pH de 7,0 a 8,3, y la temperatura típicamente es de al menos aproximadamente 30° C. Las condiciones rigurosas también pueden lograrse con la adición de agentes desestabilizantes como formamida. En general, una relación entre la señal y el ruido de 2 veces (o mayor) que la observada para una sonda no relacionada en el ensayo de hibridación en particular indica la detección de una hibridación específica. Los ácidos nucleicos que no hibridan entre sí en condiciones rigurosas siguen siendo sustancialmente idénticos si las proteínas que codifican son sustancialmente idénticas. Esto ocurre, por ejemplo, cuando se genera una copia de un ácido nucleico usando la degeneración de codon máxima permitida por el código genético.

[0143] A continuación aparecen ejemplos de conjuntos de condiciones de hibridación o lavado que pueden usarse para clonar secuencias de nucleótidos que son homólogas a las secuencias de nucleótidos de referencias de

- la presente invención: una secuencia de nucleótidos de referencia hibrida específicamente con la secuencia de nucleótidos de referencia en dodecil sulfato sódico (SDS) al 7%, NaPO₄ 0,5 M, EDTA 1 mM a 50° C con lavado en SSC 2x, SDS al 0,1% a 50° C; más deseablemente en dodecil sulfato sódico (SDS) al 7%, NaPO₄ 0,5 M, EDTA 1 mM a 50°C con lavado en SSC 1x, SDS al 0,1% a 50° C, aún más deseablemente en dodecil sulfato sódico (SDS) al 7%, NaPO₄ 0,5 M, EDTA 1 mM a 50°C con lavado en SSC 0,5x, SDS al 0,1% a 50° C, específicamente en dodecil sulfato sódico (SDS) al 7%, NaPO₄ 0,5 M, EDTA 1 mM a 50°C con lavado en SSC 0,1x, SDS 0,1% a 50°C, más específicamente en dodecil sulfato sódico (SDS) al 7%, NaPO₄ 0,5 M, EDTA 1 mM a 50° C con lavado en SSC 0,1x, SDS al 0,1% a 65° C.
- 10 **[0144]** Una «subsecuencia» hace referencia a una secuencia de ácido nucleico o aminoácidos que comprende una parte de una secuencia de ácido nucleico o aminoácidos (p. ej., proteína) más larga, respectivamente.
- [0145]** Similitud considerable: El término «similitud considerable» en el contexto de dos secuencias de ácido nucleico o proteínas, se refiere a dos o más secuencias o subsecuencias que son considerablemente similares, por ejemplo, que tiene el 50%, específicamente el 60%, más específicamente el 70%, incluso más específicamente el 80%, todavía más específicamente el 90%, aún más específicamente el 95% y lo más específico el 99% de identidad de secuencia.
- 20 **[0146]** Sustrato: un sustrato es la molécula que una enzima reconoce de forma natural y se convierte en un producto en la ruta bioquímica en la que la enzima lleva a cabo su función de forma natural, o es una versión modificada de la molécula, que también es reconocido por la enzima y es convertido por la enzima en un producto en una reacción enzimática similar a la reacción natural.
- 25 **[0147]** Transformación: proceso de introducción de un ADN heterólogo en una célula vegetal, tejido vegetal o planta. Se entiende que células vegetales, tejidos vegetales o plantas transformados abarca no solo el producto final del proceso de transformación sino también a su progenie transgénica.
- [0148]** «Transformado», «transgénico» y «recombinante» se refiere a un organismo huésped como una bacteria o una planta dentro de la que se ha introducido una molécula de ácido nucleico heteróloga. La molécula de ácido nucleico puede estar integrada de forma estable en el genoma del huésped o la molécula de ácido nucleico también puede presentarse como una molécula extracromosómica. Esta molécula extracromosómica puede autoreplicarse. Se entiende que células, tejidos o plantas transformados abarca no solo el producto final del proceso de transformación sino también a su progenie transgénica. Un huésped «no transformado», «no transgénico» o «no recombinante» se refiere a un organismo nativo, por ejemplo, una bacteria o planta que no contiene la molécula de ácido nucleico heterólogo.
- [0149]** Viabilidad: «viabilidad» según se usa en este documento, se refiere a un parámetro de aptitud de una planta. En las plantas se estudia su rendimiento homocigoto de desarrollo de la planta, lo que indica que proteínas son esenciales para el crecimiento de la planta.

DESCRIPCIÓN DETALLADA DE LA INVENCION

I. Descripción general de la genómica funcional de los rasgos característicos

- 45 **[0150]** El objetivo de la genómica funcional es identificar los genes que controlan la expresión de los fenotipos de los organismos y emplear diversas metodologías, incluyendo entre otras la bioinformática, estudios de expresión génica, interacciones entre productos génicos, genética, bioquímica y genética molecular. Por ejemplo, la bioinformática puede asignar una función a un determinado gen identificando genes en organismos heterólogos con un alto grado de similitud (homología) a nivel de aminoácidos o nucleótidos. La expresión de un gen a los niveles de ARNm o proteína puede asignar una función relacionado la expresión de dicho gen con una respuesta ambiental, un proceso de desarrollo o una perturbación genética (mutacional) o de genética molecular (sobreexpresión o subexpresión génica). La expresión de un gen a nivel de ARNm puede determinarse sola (análisis de transferencia de ARN) o conjuntamente con otros genes (análisis de micromatrices), mientras que la expresión de un gen a nivel de proteína puede determinarse sola (gel de proteínas nativas o desnaturalizados o análisis por inmunotransferencia) o conjuntamente con otros genes (análisis proteómico). El conocimiento de las interacciones proteína/proteína y proteína/ADN puede permitir asignar función identificando secuencias de proteínas y ácidos nucleicos que actúan conjuntamente en el mismo proceso biológico. La genética permite asignar función a un gen demostrando que las lesiones del ADN (mutaciones) en el gen tienen un efecto cuantificable sobre el organismo, incluyendo pero sin limitaciones: su desarrollo, biosíntesis y respuesta a hormonas, crecimiento y porte (arquitectura de la planta), perfiles de expresión de ARNm, perfiles de expresión de proteínas, capacidad para resistir a las enfermedades, tolerancia es estreses abióticos, capacidad para obtener nutrientes, eficacia fotosintética, alteración del metabolismo principal y secundario y composición de los diversos órganos de la planta. La bioquímica puede asignar función demostrando que la proteína codificada por el gen, típicamente cuando se expresa en un organismo heterólogo, posee una determinada actividad enzimática, sola o en combinación con otras proteínas. La genética molecular puede asignar función mediante sobreexpresión o subexpresión del gen en la planta nativa o en

organismos heterólogos y observando efectos cuantificados como se describe en la asignación funcional mediante la genética anterior. En la genómica funcional, se utilizan cualquiera o todas estas técnicas, a menudo conjuntamente, para asignar genes a funciones a través de cualquiera de los diversos fenotipos de organismos.

5 **[0151]** Los expertos en la materia reconocen que estas diferentes metodologías pueden proporcionar cada una datos que evidencien la función de un gen en particular y que esta evidencia es más fuerte con el aumento en la cantidad de datos utilizados para la asignación funcional: específicamente a partir de una única metodología, más específicamente a partir de dos metodologías e, incluso más específicamente, a partir de más de dos metodologías. Además, los expertos en la materia son conscientes de que diferentes metodologías pueden diferir en la fuerza de la evidencia para la asignación de la función génica. Típicamente, aunque no siempre, un dato de evidencia bioquímica, genética y genética molecular se considera más potente que un dato de evidencia bioinformática o expresión génica. Finalmente, los expertos en la materia reconocen que, para genes diferentes, un único dato de una única metodología puede diferir en términos de la fuerza de la evidencia proporcionada por cada dato diferente para la asignación de la función de estos genes diferentes.

15 **[0152]** El objetivo de la genómica funcional de características de cosechas es identificar los genes de los rasgos característicos de las cosechas, es decir, genes capaces de conferir características agronómicas útiles a los cultivos vegetales. Entre estos rasgos característicos agronómicos se incluyen, pero sin limitaciones: potenciar el rendimiento, en cantidad o en calidad; potenciar la adquisición de nutrientes y potenciar la eficacia metabólica, potenciar o alterar la composición de nutrientes de los tejidos vegetales utilizados para alimentos, piensos, fibra o procesamiento, potenciar la utilidad del procesamiento agrícola o industrial; potenciar la resistencia a enfermedades vegetales; potenciar la tolerancia a condiciones ambientales adversas (estrés abióticos) incluyendo, pero sin limitaciones, sequía, frío excesivo, calor excesivo o salinidad excesiva del suelo o acidez o alcalinidad extremas; y alteraciones en la arquitectura o desarrollo de la planta incluyendo cambios en el ritmo del desarrollo. La utilización de estos genes de rasgos característicos identificados por medios transgénicos o no transgénicos podría mejorar materialmente los cultivos vegetales para beneficio de la agricultura.

[0153] Los cereales son los cultivos vegetales más importantes en el planeta, en términos tanto de consumo humano como animal. La sintenia genómica (conservación del orden de los genes dentro de segmentos cromosómicos grandes) se observa en arroz, maíz, trigo, cebada, centeno, avena y otras monocotiledóneas de importancia agrícola, lo que facilita el mapeo y aislamiento de genes ortólogos para diversas especies de cereales en función de la secuencia de un único gen del cereal. El arroz tiene el genoma más pequeño (~420 Mb) entre los cereales y, recientemente, ha tenido una atención principal de la genómica pública y privada y se han realizado esfuerzos de secuenciación por marcador de secuencia expresada (EST, por sus siglas en inglés).

35 **[0154]** Para identificar los genes del rasgo característico de cosechas en el control [rasgo característico] del genoma del arroz [trigo], se priorizaron los genes del borrador de la secuencia genética del arroz [base de datos EST del trigo] en función de una o más metodologías de genómica funcional. Por ejemplo, los estudios de expresión de todo el genoma de las plantas de arroz infectadas con el hongo del añublo del arroz (*Magnaporthe oryzae*) se usaron para priorizar los genes candidatos que controlan la resistencia a la enfermedad. Los ADNc de longitud completa y parcial de los genes del rasgo característico del arroz candidatos podrían predecirse a continuación en base al análisis de la secuencia del genoma completo del arroz, y aislarse mediante el diseño y utilización de cebadores para la amplificación mediante PCR usando un programa de recolección de cebadores de PCR disponible. Los cebadores se usaron para la amplificación por PCR de los ADNc de longitud completa o parcial a partir de 45 bibliotecas de ADNc de arroz o de la primera cadena del ADNc. Los clones de ADNc resultantes de cada técnica se usaron para la construcción de vectores diseñados para alterar la expresión de estos genes en plantas transgénicas usando metodologías de genética molecular de plantas, que se describen en detalle a continuación. La alteración del fenotipo de la planta mediante sobreexpresión o subexpresión de genes de rasgos característicos clave en plantas transgénicas es un procedimiento sólido y establecido para la asignación de funciones a los genes de plantas. Los 50 ensayos para identificar plantas transgénicas con alteraciones en los rasgos característicos de interés se usarán para asignar claramente la utilidad de estos genes para la mejora del arroz, y por extensión, de otros cereales, mediante procedimientos de cultivo transgénicos o clásicos.

II. Identificación, clonación y secuenciación de ADNc

55 **[0155]** La clonación y secuenciación de los ADNc de la presente invención se describen en el ejemplo 1.

[0156] Los ácidos nucleicos y proteínas aislados de la presente invención son útiles para una gama de plantas, monocotiledóneas y dicotiledóneas, especial monocotiledóneas, como el arroz, el trigo, la cebada y el maíz. En una realización más específica, la monocotiledónea es un cereal. En una realización más específica, el cereal puede ser, por ejemplo, maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* sp. o teosinte. En una realización más específica, el cereal es arroz. Entre otros géneros vegetales se incluyen, pero sin limitaciones *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Gragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesis*, *Pelargonium*, *Paniseum*, *Pennisetum*,

Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Pisum, Phaseolus, Lolium, Oryza, Avena, Hordeum, Secale, Allium y Triticum.

[0157] La presente invención también proporciona un procedimiento de genotipado de una planta o parte de una planta que comprende una molécula de ácido nucleico de la presente invención. Opcionalmente, la planta es una monocotiledónea como, pero sin limitaciones, arroz o trigo. El genotipado proporciona un medio para distinguir homólogos de un par de cromosomas y puede usarse para diferencias segregantes en una población vegetal. Los procedimientos de marcador molecular pueden usarse en estudios filogenéticos, en la caracterización de relaciones genéticas entre las variedades de cosechas, en la identificación de cruces o híbridos somáticos, en la localización de segmentos cromosómicos que afectan a los rasgos característicos monogénicos, mapeo en base a la clonación y en el estudio de la herencia cuantitativa (véase *Plant Molecular Biology: A Laboratory Manual*, capítulo 7, Clark ed., Springer-Verlag, Berlín 1997; Paterson, A.H., "The DNA Revolution", capítulo 2 en *Genome Mapping in Plants*, Paterson, A.H. ed., Academic Press/R. G. Lands Co., Austin, Tejas, 1996).

[0158] El procedimiento de genotipado puede emplear cualquier número de técnicas analíticas de marcadores moleculares como, pero sin limitaciones, polimorfismos de longitud de fragmentos de restricción (RFLP). Como es conocido en la materia, los RFLP se producen por diferencias en la longitud de fragmentos de restricción de ADN que resulta de las diferencias nucleotídicas entre alelos del mismo gen. Por tanto, la presente descripción proporciona un procedimiento de seguimiento de la segregación de un gen o ácido nucleico de la presente descripción o de secuencias cromosómicas genéticamente ligadas usando análisis de RFLP. Las secuencias cromosómicas ligadas están dentro de los 50 centiMorgan (50 cM), dentro de los 40 o 30 cM, específicamente dentro de 20 o 10 cM, más específicamente dentro de 5, 3, 2 o 1 CM del ácido nucleico de la descripción.

III. Rasgos característicos de interés

[0159] La presente descripción abarca la identificación y aislamiento de polinucleótidos que codifican proteínas implicadas en la detección de azúcar y, en última instancia, en la captación del nitrógeno y el metabolismo del carbono. Puede usarse la alteración de la expresión de los genes relacionados con estos rasgos característicos para mejorar o modificar las plantas y/o el grano, si se desea. En los ejemplos se describen los genes de interés aislados y los procedimientos de análisis de la alteración de expresión y sus efectos sobre las características de la planta.

[0160] Un aspecto de la presente invención proporciona composiciones y procedimientos para alterar (es decir, aumento) el nivel de moléculas de ácido nucleico y polipéptido de la presente invención en plantas. En particular, las moléculas de ácido nucleico y polipéptidos de la invención se expresan constitutiva, temporal o espacialmente, por ejemplo, en las etapas de desarrollo, en determinados tejidos y/o cantidades, que no son característicos de plantas no manipuladas genéticamente mediante recombinación. Por tanto, la presente descripción proporciona utilidad en estos ejemplos de aplicaciones alterando las características específicas identificadas anteriormente.

VI. Control de la expresión génica en plantas transgénicas

[0161] La invención además se refiere a células transformadas que comprenden las moléculas de ácido nucleico, plantas transformadas, semillas y partes de plantas, y procedimientos para modificar los rasgos característicos fenotípicos alterando la expresión de los genes de la invención.

A. Modificación de las secuencias codificadoras y secuencias adyacentes

[0162] La expresión transgénica en plantas de genes derivados de fuentes heterólogas puede implicar la modificación de esos genes para conseguir y optimizar su expresión en plantas. En particular, los ORF bacterianos que codifican enzimas independientes pero que son codificadas por la misma transcripción en el microorganismo nativo se expresan mejor en las plantas en transcripciones independientes. Para conseguir esto, cada ORF microbiano se aísla individualmente y se clona dentro de un casete que proporciona una secuencia promotora de plantas en el extremo 5' del ORF y un terminador de la transcripción de plantas en el extremo 3' del ORF. La secuencia de ORF aislado incluye específicamente el codón de inicio ATG y el codón PARADA de terminación aunque puede incluir una secuencia adicional más allá del ATG de inicio y el codón PARADA. Además, el ORF puede estar truncado, pero seguir manteniendo la actividad requerida; las versiones de ORF truncadas especialmente largas que retienen la actividad pueden ser preferibles para la expresión en organismos transgénicos. Por «promotor de plantas» y «terminador de transcripción de plantas» se pretende que signifique promotores y terminadores de transcripción que funcionan dentro de las células vegetales. Esto incluye promotores y terminadores de transcripción que pueden derivar de fuentes no vegetales como virus (un ejemplo es el virus del mosaico de la coliflor).

[0163] En algunos casos, no es necesaria la modificación de las secuencias que codifican el ORF y de la secuencia adyacente. Basta con aislar un fragmento que contenga el ORF de interés e insertarlo antes del extremo 5' de un promotor de plantas. Por ejemplo, Gaffney y col. (*Science* 261: 745-756 (1993)) han expresado el gen nahG de *Pseudomonas* en plantas transgénicas bajo el control del promotor 35S de CaMV y el terminador tml de CaMV con éxito sin modificación de la secuencia codificadora y con los nucleótidos del gen de *Pseudomonas* después del

extremo 3' del ATG aún unidos y los nucleótidos antes del extremo 5' del codon PARADA aún unidos a la ORF de nahG. Específicamente, se deberá dejar una secuencia microbiana adyacente lo más pequeña posible después del extremo 3' del ATG y antes del extremo 5' del codon PARADA. En la práctica, esta construcción puede depender de la disponibilidad de sitios de restricción.

5

[0164] En otros casos, la expresión de genes derivados de fuentes microbianas puede proporcionar problemas de expresión. Estos problemas han sido bien caracterizados en la técnica y son especialmente frecuentes con genes derivados de determinadas fuentes como *Bacillus*. Estos problemas pueden aplicarse a la secuencia de nucleótidos de esta invención y la modificación de estos genes puede realizarse usando técnicas ahora bien conocidas en la técnica. Pueden aparecer los siguientes problemas:

10

1. Uso de codones.

[0165] El uso de codones específicos en plantas difiere del uso de codones específicos en determinados microorganismos. La comparación del uso de codones dentro de una ORF microbiana clonada con el uso en genes vegetales (y en particular, en los genes de la planta diana) permitirá una identificación de los codones dentro del ORF que deberían cambiarse específicamente. Típicamente, la evolución vegetal ha tendido hacia una fuerte preferencia de los nucleótidos C y G en la tercera posición de base de las monocotiledóneas, mientras que en las dicotiledóneas a menudo se usan los nucleótidos A o T en esta posición. Modificando un gen para incorporar un uso de codones específico para una especie transgénica diana en particular, podrían resolverse muchos de los problemas descritos a continuación sobre el contenido GC/AT y el ajuste ilegítimo.

20

2. Contenido en CG/AT.

[0166] Los genes de plantas típicamente tienen un contenido GC de más del 35%. Las secuencias ORF que son ricas en nucleótidos A y T pueden causar varios problemas en las plantas. En primer lugar, se cree que los motivos ATTTA causan desestabilización de los mensajeros y se encuentran en el extremo 3' de muchos ARNm de vida corta. En segundo lugar, se considera que la aparición de señales de poliadenilación como AATAAA en posiciones inapropiadas dentro del mensajero causa un truncamiento prematuro de la transcripción. Además, las monocotiledóneas pueden reconocer secuencias ricas en AT como sitios de ajuste (véase a continuación).

30

3. Secuencias adyacentes a la metionina de inicio

[0167] Las plantas difieren de los microorganismos en que sus mensajeros no poseen un sitio de unión a ribosomas definido. Más bien se considera que los ribosomas se unen al extremo 5' del mensajero y buscan el primer ATG disponible donde iniciar la traducción. No obstante, se considera que existe preferencia por determinados nucleótidos adyacentes al ATG y que la expresión de los genes microbianos puede potenciarse mediante la inclusión de un iniciador de la traducción eucariota consenso en el codon ATG. Clontech (catálogo 1993/1994, página 210) ha sugerido una secuencia como iniciador de traducción consenso para la expresión del gen uidA de *E. coli* en plantas. Además, Joshi (N.A.R. 15: 6643-6653 (1987)) ha comparado muchas secuencias de plantas adyacentes al ATG y sugiere otra secuencia consenso. En situaciones en las que se encuentran dificultades para la expresión de ORF microbianos en plantas, la inclusión de una de estas secuencias en el ATG de inicio puede mejorar la traducción. En estos casos los tres últimos nucleótidos de la secuencia consenso pueden no ser apropiados para su inclusión en la secuencia modificada debido a su modificación del segundo resto de AA. Las secuencias específicas adyacentes a la metionina de inicio pueden diferir entre las diferentes especies vegetales. Un sondeo de 14 genes de maíz localizados en la base de datos GenBank proporcionó los siguientes resultados: Posición antes de ATG de inicio en 14 genes del maíz:

45

[0168]

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
C	3	8	4	6	2	5	6	0	10	7
T	3	0	3	4	3	2	1	1	1	0
A	2	3	1	4	3	2	3	7	2	3
G	6	3	6	0	6	5	4	6	1	5

50

Este análisis puede hacerse para las especies vegetales deseadas en las que se ha incorporado la secuencia de nucleótidos y la secuencia adyacente al ATG se ha modificado para incorporar los nucleótidos específicos.

4. Eliminación de sitios de ajuste ilegítimo

55

[0169] Los genes clonados de fuentes no vegetales y no optimizados para su expresión en plantas también pueden contener motivos que pueden ser reconocidos en plantas como sitios de ajuste 5' o 3' y ser escindidos, generando de este modo mensajeros truncados o delecionados. Estos sitios pueden eliminarse usando las técnicas

bien conocidas en la técnica.

[0170] Las técnicas para la modificación de secuencias codificadoras y secuencias adyacentes son bien conocidas en la técnica. En los casos donde la expresión inicial de una ORF microbiana es baja y se considera apropiado hacer alteraciones en la secuencia como se describe arriba, a continuación la construcción de genes sintéticos puede conseguirse según procedimientos bien conocidos en la técnica. Estos se describen, por ejemplo, en las descripciones de las patentes publicadas EP 0 385 962 (de Monsanto), EP 0 359 472 (de Lubrizol) y en el documento WO 93/07278 (de Ciba-Geigy). En la mayoría de los casos es preferible probar la expresión de construcciones génicas usando protocolos de ensayo transitorio (que son bien conocidos en la técnica) antes de su transferencia a plantas transgénicas.

B. Construcción de casetes de expresión en plantas

[0171] Las secuencias codificadoras dirigidas a la expresión en plantas transgénicas se ensamblan primero en casetes de expresión bajo un promotor adecuado expresable en plantas. Los casetes de expresión también pueden comprender cualquier secuencia adicional requerida o seleccionada para la expresión del transgén. Entre estas secuencias se incluyen, pero sin limitaciones, terminadores de transcripción, secuencias extrañas para potenciar la expresión como intrones, secuencias vitales y secuencias orientadas a dirigir el producto génico a orgánulos específicos y compartimentos celulares. Estos casetes de expresión puede transferirse fácilmente a continuación a los vectores de transformación de plantas descritos más abajo. A continuación se recoge una descripción de diversos componentes de casetes de expresión típica.

1. Promotores

[0172] La selección del promotor utilizado en los casetes de expresión determinará el patrón espacial y temporal de expresión del transgén en la planta transgénica. Los promotores seleccionados expresarán transgenes en tipos específicos de células (como células epidérmicas de las hojas, células mesófilas o células de la corteza de la raíz) y en tejidos u órganos específicos (raíces, hojas o flores, por ejemplo) y la selección reflejará la ubicación deseada para la acumulación del producto génico. Alternativamente, el promotor seleccionado puede dirigir la expresión del gen en diversas condiciones de inducción. Los promotores varían en su potencia, es decir, en su capacidad para promover la transcripción. Dependiendo del sistema celular huésped utilizado, puede usarse cualquier de los diversos promotores adecuados, incluso el promotor nativo del gen. A continuación se recogen ejemplos no limitantes de promotores que pueden usarse en casetes de expresión.

35 a. Expresión constitutiva, el promotor de la ubiquitina:

[0173] La ubiquitina es un producto génico conocido por acumularse en muchos tipos de células y su promotor se ha clonado a partir de varias especies para su uso en plantas transgénicas [p. ej., girasol (Binet y col. *Plant Science* 79: 87-94 (1991)); maíz (Christensen y col. *Plant Molec. Biol.* 12: 619-632 (1989)) y *Arabidopsis* (Callis y col., *J. Biol. Chem.* 265:12486-12493 (1990) y Norris y col., *Plant Mol. Biol.* 21:895-906 (1993)). El promotor de la ubiquitina del maíz se ha desarrollado en sistemas monocotiledóneos transgénicos y su secuencia y vectores construidos para la transformación de monocotiledóneas se describen en la publicación de patente EP 0 342 926 (de Lubrizol). Taylor y col. (*Plant Cell Rep.* 12: 491-495 (1993)) describen un vector (pAHC25) que comprende el promotor de la ubiquitina del maíz y el primer intrón y su alta actividad en suspensiones celulares de numerosas monocotiledóneas cuando se introduce mediante bombardeo con microproyectiles. El promotor de la ubiquitina de *Arabidopsis* es ideal para su uso con las secuencias de nucleótidos de la presente invención. El promotor de la ubiquitina es adecuado para la expresión génica en plantas transgénicas, tanto monocotiledóneas como dicotiledóneas. Los vectores adecuados son derivados de pAHC25 y cualquiera de los vectores de transformación descritos en esta aplicación modificados mediante la introducción del promotor de ubiquitina y/o secuencias de intrones adecuados.

b. Expresión constitutiva, el promotor 35S de CaMV:

[0174] La construcción del plásmido pCGN1761 se describe en la solicitud de patente publicada EP 0 392 225 (ejemplo 23). pCGN1761 contiene el promotor 35S de CNMV «doble» y el terminador de transcripción tml con un único sitio para EcoRI entre el promotor y el terminador y tiene una estructura central de tipo pUC. Se construye un derivado de pCGN1761 de modo que tenga un polienlazador modificado que incluya sitios NotI o XhoI además del sitio EcoRI existente. Este derivado se designa pCGN1761ENX. pCGN1761ENX es útil para la clonación de secuencias de ADNc o secuencias codificadoras (incluyendo secuencias ORF microbianas) dentro de su polienlazador con el objetivo de su expresión bajo el control del promotor 35S en plantas transgénicas. El casete completo promotor 35S-secuencia codificadora-terminador tml de esta construcción puede escindirse en los sitios HindIII, SphI, Sall y XbaI en posición 5' del promotor y sitios XbaI, BamHI y BglI en posición 3' con respecto al terminador para su transferencia a vectores de transformación como los indicados a continuación. Además, el fragmento del promotor 35S doble puede eliminarse mediante escisión en posición 5' con HindIII, SphI, Sall, XbaI o PstI y escisión en posición 3' en cualquier de los sitios de restricción del polienlazador (EcoRI, NotI o XhoI) para su sustitución por otro promotor. Si se desea, pueden hacerse modificaciones alrededor de los sitios de clonación

mediante la introducción de secuencias que pueden potenciar la traducción. Esto es particularmente útil cuando se desea la sobreexpresión. Por ejemplo, puede modificarse pCGN1761ENX mediante la optimización del sitio de inicio de traducción como se describe en el ejemplo 37 de la patente de EE. UU. N.º 5.639.949.

5 c. Expresión constitutiva, el promotor de la actina:

[0175] Se conocen varias isoformas de actina para su expresión en la mayoría de los tipos celulares y, por consiguiente, el promotor de la actina es una buena elección de promotor constitutivo. En particular, se ha clonado y caracterizado el promotor del gen Act1 del arroz (McElroy y col. Plant Cell 2: 163-171 (1990)). Se encontró que un
10 fragmento de 1,3 kb del promotor contenía todos los elementos reguladores necesarios para la expresión en protoplastos de la planta del arroz. Adicionalmente, se han construido numerosos vectores de expresión en base al promotor de Act1 especialmente para su uso en monocotiledóneas (McElroy y col., Mol. Gen. Genet. 231: 150-160 (1991)). Estos incorporan el intrón 1 de Act1, la secuencia flanqueante Adh1 5' y el intrón 1 de Adh1 (a partir del gen de la alcohol deshidrogenasa de maíz) y la secuencia del promotor 35S de CaMV. Los vectores que mostraban la
15 mayor expresión eran fusiones del promotor 35S y el intrón de Act1 o la secuencia flanqueante 5' de Act1 y el intrón de Act1. La optimización de las secuencias alrededor del ATG de inicio (del gen indicador GUS) también potenciaba la expresión. Los casetes de expresión del promotor descritos por McElroy y col. (Mol. Gen. Genet. 231: 150-160 (1991)) pueden modificarse fácilmente para la expresión génica y son especialmente adecuados para su uso en huéspedes monocotiledóneos. Por ejemplo, los fragmentos que contienen el promotor se eliminaron de las
20 construcciones de McElroy y se usaron para sustituir el promotor 35S doble en pCGN1761ENX que entonces está disponible para la inserción de secuencias génicas específicas. Los genes de fusión construidos de este modo pueden transferirse a continuación a vectores de transformación apropiados. En una publicación aparte, se ha encontrado que el promotor Act1 del arroz con su primer intrón dirige una alta expresión en células de cebada cultivadas (Chibbar y col. Plant Cell Rep. 12: 506-509 (1993)).

25

d. Expresión inducible, promotores PR-1:

[0176] El promotor 35S doble en pCGN1761ENX puede ser sustituido por cualquier otro promotor de elección quede lugar a niveles de expresión adecuadamente altos. A modo de ejemplo, uno de los promotores químicamente regulables descritos en la patente de EE. UU. N.º 5.614.395, como el promotor PR-1a del tabaco, puede sustituir al
30 promotor 35S doble. Alternativamente, puede usarse el promotor PR-1 de *Arabidopsis* descrito en Lebel y col., Plant J. 16:223-233 (1998). El promotor de elección se escinde específicamente a partir de su fuente mediante enzimas de restricción, aunque puede amplificarse mediante PCR usando cebadores portadores de los sitios de restricción terminales apropiados. Si se realiza la amplificación mediante PCR, se deberá secuenciar de nuevo el promotor para
35 comprobar los errores de amplificación antes de la clonación del promotor amplificado en el vector diana. El promotor PR-1a del tabaco regulable químicamente o mediante patógeno se escinde a partir del plásmido pCIB1004 (para su construcción, véase el ejemplo 21 del documento EP 0 332 104) y se transfiere a un plásmido pCGN1761ENX (Uknes y col., Plant Cell 4: 645-656 (1992)). pCIB1004 se escinde con NcoI y el extremo 3' colgante resultantes del fragmento linearizado se convierte el romo mediante el tratamiento con la ADN polimerasa T4. A
40 continuación, el fragmento se escinde con HindIII y el fragmento que contiene el promotor PR-1a resultante se purifica en gel y se clona dentro de pCGN1761ENX en el que se ha eliminado el promotor 35S doble. Esto se consigue mediante la escisión con XhoI y el recorte con la polimerasa T4, seguida de la escisión con HindIII y el aislamiento del fragmento que contiene el vector-terminador más largo en el que se clonó el fragmento del promotor pCIB1004. Esto genera un derivado de pCGN1761ENX con el promotor PR-1a y el terminador tml y la intervención
45 de un polienlazador con sitios EcoRI y NotI exclusivos. La secuencia codificadora seleccionada puede insertarse en este vector y los productos de fusión (p. ej., promotor-gen-terminador) pueden transferirse posteriormente a cualquier vector de transformación seleccionado, incluyendo los descritos a continuación. Puede emplearse diversos reguladores químicos para inducir la expresión de la secuencia codificadora seleccionada en las plantas transformadas según la presente invención, incluyendo el benzotriazol, el ácido isonicotínico y compuestos del
50 ácido salicílico descritos en las patentes de EE. UU. N.º 5.523.311 y 5.614.395.

e. Expresión inducible, un promotor inducible por etanol:

[0177] También puede usarse un promotor inducible por determinados alcoholes o cetonas, como el etanol,
55 para conferir la expresión inducible de una secuencia codificadora de la presente invención. Este promotor es, por ejemplo, el promotor del gen alcA de *Aspergillus nidulans* (Caddick y col., (1998) Nat. Biotechnol 16:177-180). En *A. nidulans*, el gen alcA codifica una alcohol deshidrogenasa I, cuya expresión está regulada por los factores de transcripción AlcR en presencia del inductor químico. Para los fines de la presente invención, las secuencias que codifican CAT en el plásmido palcA:CAT que comprende una secuencia promotora del gen alcA fusionada con un
60 promotor 35S mínimo (Caddick y col. (1998) Nat. Biotechnol 16:177-180) se sustituyen por una secuencia codificadora de la presente invención para formar un casete de expresión que tiene la secuencia codificadora bajo en control del promotor del gen alcA. Esto se lleva a cabo usando procedimientos bien conocidos en la técnica.

f. Expresión inducible, un promotor inducible por glucocorticoides:

65

[0178] También se contempla la inducción de la expresión de una secuencia de ácido nucleico de la presente

invención usando sistemas a base de hormonas esteroides. Por ejemplo, se usa un sistema de inducción mediado por glucocorticoides (Aoyama y Chua (1997) *The Plant Journal* 11: 605-612) y se induce la expresión génica mediante la aplicación de un glucocorticoide, por ejemplo, un glucocorticoide sintético, específicamente dexametasona, específicamente a una concentración que oscila de 0,1 mM a 1 mM, más específicamente de 10 mM a 100 mM. Para los fines de la presente invención, las secuencias del gen de la luciferasa se sustituyen por una secuencia de ácido nucleico de la invención para formar un casete de expresión que tiene una secuencia de ácido nucleico de la invención bajo el control de seis copias de las secuencias activadoras antes del extremo 5' de GAL4 fusionadas con el promotor 35S mínimo. Esto se lleva a cabo usando procedimientos bien conocidos en la técnica. El factor de transactivación comprende el dominio de unión a ADN GAL4 (Keegan y col. (1986) *Science* 231: 699-704) unido al dominio de transactivación de la proteína VP16 del virus del herpes (Triezenberg y col. (1988) *Genes Dev.* 2: 781-729) unido al dominio de unión a hormonas del receptor de glucocorticoides de rata (Picard y col. (1988) *Cell* 54: 1073-1080). La expresión de la proteína de fusión está controlada por un promotor conocido en la técnica o descrito en este documento. Este casete de expresión también está incluido en la planta que comprende una secuencia de ácido nucleico de la invención unida a 6xGAL4/promotor mínimo. Por tanto, la especificidad de tejido u órgano de la proteína de fusión se logra llevando a una especificidad de tejido u órgano inducible de la toxina insecticida.

g. Expresión específica de raíces:

[0179] Otro patrón de expresión génica es la expresión en las raíces. Un promotor de raíces adecuado es el promotor de gen similar a metalotioneina (MTL) del maíz descrito por Framons (*FEBS* 290: 103-106 (1991)) y también en la patente de EE. UU. N.º 5.466.785. Este promotor «MTL» se transfiere a un vector adecuado, como pCGN1761ENX, para la inserción de un gen seleccionado y la posterior transferencia del casete promotor-gen-terminador completo a un vector de transformación de interés.

h. Promotores inducibles por heridas:

[0180] Los promotores inducibles por heridas también pueden ser adecuados para la expresión génica. Se ha descrito un gran número de estos promotores (p. ej., Xu y col. *Plant Molec. Biol.* 22: 573-588 (1993), Logemann y col. *Plant Cell* 1: 151-158 (1989), Rohrmeier y Lehle, *Plant Molec. Biol.* 22: 783-792 (1993), Firek y col. *Plant Molec. Biol.* 22: 129-142 (1993), Warner y col. *Plant J.* 3: 191-201 (1993)) y todos son adecuados para su uso en la presente invención. Logemann y col. describen las secuencias antes de 5' del gen *wun1* de la planta dicotiledónea de la patata. Xu y col. muestran que un promotor de la planta dicotiledónea de la patata (*pin2*) es activo en la planta monocotiledónea del arroz. Adicionalmente, Rohrmeier y Lehle describen la clonación del ADNc *Wipl* del maíz que es inducido por herida y puede usarse para aislar el promotor relacionado usando técnicas convencionales. De forma similar, Firek y col. y Warner y col. han descrito un gen inducido por heridas de la monocotiledónea *Asparagus officinalis*, que se expresa en una herida local y en sitios de invasión de patógenos. Usando técnicas de clonación bien conocidas en la materia, estos promotores pueden transferirse a vectores adecuados, fusionarse con los genes relacionados con esta invención y utilizarse para expresar estos genes en los sitios de heridas en las plantas.

i. Expresión específica de la médula:

[0181] En la solicitud de patente WO 93/07278, se describe el aislamiento del gen *trpA* del maíz, que se expresa preferiblemente en células de la médula. Están presentes la secuencia génica y el promotor que se extiende hasta -1.726 pb desde el inicio de la transcripción. Usando técnicas de biología molecular convencionales, este promotor, o partir del mismo, puede transferirse a un vector como pCGN1761 donde puede sustituirse al promotor 35S y usarse para dirigir la expresión de un gen extraño de manera específica de la médula. De hecho, los fragmentos que contienen el promotor específico de la médula o de parte del mismo pueden transferirse a cualquier vector o modificarse para su utilización en plantas transgénicas.

j. Expresión específica de hojas

[0182] Un gen del maíz que codifica la fosfoenol carboxilasa (PEPC) ha sido descrito por Hudspeth y Gula (*Plant Molec Biol* 12: 579-589 (1989)). Usando técnicas biológicas moleculares convencionales el promotor para este gen puede usarse para dirigir la expresión de cualquier gen de una forma específica de hojas en plantas transgénicas.

k. Expresión específica del polen

[0183] En el documento WO 93/07278 se describe el aislamiento del gen de la proteína quinasa dependiente de calcio (CDPK) de maíz que se expresa en células del polen. La secuencia génica y el promotor se extienden hasta los 1.400 pb desde el inicio de la transcripción. Usando técnicas convencionales de biología molecular, este promotor, o partes del mismo, puede transferirse a un vector como pCGN1761 donde puede sustituirse al promotor 35S y usarse para dirigir la expresión de una secuencia de ácido nucleico de la invención de forma dependiente del polen.

2. Terminadores de la transcripción

[0184] Se dispone de diversos terminadores de transcripción para su uso en casetes de expresión. Estos son responsables de la terminación de la transcripción más allá del transgén y de la poliadenilación correcta del ARNm. Los terminadores de transcripción apropiados son aquellos que son conocidos por funcionar en plantas e incluye el terminador 35S de CaMV, el terminador tml, el terminador de la nopalina sintetasa y el terminador E9 de rbcS del guisante. Estos pueden usarse tanto en monocotiledóneas como en dicotiledóneas. Además puede usarse un terminador de la transcripción nativo de un gen.

3. Secuencias para potenciación o regulación de la expresión

10 **[0185]** Se han encontrado numerosas secuencias para potenciar la expresión génica desde el interior de la unidad transcripción y estas secuencias pueden usarse junto con los genes de esta invención para aumentar su expresión en plantas transgénicas.

[0186] Se ha mostrado que diversas secuencias de intrones potencian la expresión, especialmente en células monocotiledóneas. Por ejemplo, se ha encontrado que los intrones del gen *Adhl* del maíz potencian de forma significativa la expresión del gen nativo con su promotor relacionado cuando se introducen en células de maíz. Se encontró que el intrón 1 era especialmente eficaz y potenciaba la expresión en construcciones de fusión con el gen de la cloranfenicol acetiltransferasa (Callis y col., *Genes Develop.* 1: 1183-1200 (1987)). En el mismo sistema experimental, el intrón del gen *bronze1* del maíz tenía un efecto similar a la hora de potenciar la expresión. Las secuencias de intrones se han incorporado de forma rutinaria a vectores de transformación de plantas, típicamente dentro de secuencias líderes no traducidas.

[0187] También es sabido que varias secuencias líder no traducidas derivadas de virus potencian la expresión y estas son especialmente eficaces en células dicotiledóneas. Específicamente, se ha demostrado que secuencias líderes de virus del mosaico del tabaco (TMV la «secuencia W»), del virus del moteado clorótico del maíz (MCMV) y el virus del mosaico de la alfalfa (AMV) son eficaces a la hora de potencia la expresión (p. ej., Gallie y col., *Nucl. Acids. Res.* 15: 8693-8711 (1987); Skuzeski y col. *Plant Molec. Biol.* 15: 65-79 (1990)). Entre otras secuencias líderes conocidas en la técnica se incluyen, pero sin limitaciones: secuencias líderes de picornavirus, por ejemplo, secuencia líder EMCV (región no codificadora 5' del virus de la encefalomiocarditis), (Elroy-Stein, O., Fuerst, T. R. y Moss, B. *PNAS USA* 86:6126-6130 (1989)); secuencias líderes de potyvirus, por ejemplo, secuencia líder TEV (virus del jaspeado del tabaco) (Allison y col., 1986); secuencia líder de MDMV (virus del mosaico del enanismo del maíz; *Virology* 154:9-20); secuencia líder de la proteína de unión a la cadena pesada de la inmunoglobulina humana (BiP), (Macejak, D. G. y Sarnow, P., *Nature* 353: 90-94 (1991); secuencia líder no traducida del ARNm de la proteína de la cubierta del virus del mosaico de la alfalfa (ARN 4 de AMV), (Jobling, S. A. y Gehrke, L., *Nature* 325:622-625 (1987)); secuencia líder del virus del mosaico del tabaco (TMV), (Gallie, D. R. y col., *Molecular Biology of RNA*, páginas 237-256 (1989)) y secuencia líder del virus del moteado clorótico del maíz (MCMV) (Lommel, S. A. y col., *Virology* 81:382-385 (1991)). Véase también, Della-Cioppa y col., *Plant Physiology* 84:965-968 (1987).

[0188] Además de la incorporación de uno o más de los elementos mencionados anteriormente en la región reguladora 5' de un casete de expresión diana de la invención, también pueden incorporarse otros elementos peculiares al casete de expresión diana. Entre estos elementos se incluye, pero sin limitaciones, un promotor mínimo. Por promotor mínimo se entiende que los elementos promotores basales están inactivos, o prácticamente inactivos, por lo que no inducen la activación antes del extremo 5'. Este promotor tiene baja actividad basal en las plantas en las que no está presente un transactivador o cuando el potenciador o los sitios de unión a elementos de respuesta están ausentes. Un promotor mínimo que es especialmente útil para los genes diana en plantas es el promotor mínimo Bz1 que se obtiene del gen *bronze1* del maíz. El núcleo del promotor Bz1 se obtiene de la construcción Bz1 mutante «myc»-luciferasa pBz1LucR98 mediante escisión en el sitio *NheI* localizado entre las posiciones -53 a -58. Roth y col., *Plant Cell* 3: 317 (1991). El fragmento del núcleo del promotor de Bz1 derivado se extiende, por tanto, desde la posición -53 a la +227 e incluye el intrón 1 de Bz1 en la región no traducida 5'. También es útil para la invención un promotor mínimo creado usando un elemento TATA sintético. El elemento TATA permite el reconocimiento del promotor por factores de la ARN polimerasa y confiere un nivel basal de expresión génica en ausencia de activación (véase en general, Mukumoto (1993) *Plant Mol Biol* 23: 995-1003; Green (2000) *Trends Biochem Sci* 25: 59-63).

55 4. Direccionamiento del producto génico dentro de la célula

[0189] Se conocen en las plantas diversos mecanismos para el direccionamiento de productos génicos y las secuencias que controlan el funcionamiento de estos mecanismos se han caracterizado con cierto detalle. Por ejemplo, el direccionamiento de productos génicos a los cloroplastos está controlado por una secuencia señal que se encuentra en el extremo amino terminal de diversas proteínas, el cual se escinde durante su importación al cloroplasto para dar lugar a la proteína madura (p. ej., Comai y col., *J. Biol. Chem.* 263: 15104-15109 (1988)). Estas secuencias señal pueden fusionarse a productos génicos heterólogos para llevar a cabo la importación de productos heterólogos dentro del cloroplasto (van den Broeck, y col. *Nature* 313: 358-363 (1985)). El ADN que codifica secuencias señal apropiadas puede aislarse del extremo 5' de los ADNc que codifican la proteína RUBISCO, la proteína CAB, la enzima sintetasa EPSP, la proteína GS2 y muchas otras proteínas que se sabe están localizadas en el cloroplasto. Véase también, la sección titulada «Expresión con direccionamiento a cloroplastos» del ejemplo 37

de la patente de EE. UU. N.º 5.639.949.

- [0190]** Otros productos génicos se localizan en otros orgánulos como la mitocondria y el peroxisoma (p. ej., Unger y col. *Plant Molec. Biol.* 13: 411-418 (1989)). Los ADNc que codifican estos productos también pueden manipularse para llevar a cabo el direccionamiento de productos génicos heterólogos a estos orgánulos. Son ejemplos de dichas secuencias las ATPasas codificadas en el genoma nuclear e isoformas específicas de la mitocondria de la aspartato aminotransferasa. El direccionamiento de cuerpos proteicos celulares se ha descrito en Rogers y col. (*Proc. Natl. Acad. Sci. USA* 82: 6512-6516 (1985)).
- 10 **[0191]** Además, se han caracterizado secuencias que causan el direccionamiento de productos génicos a otros compartimentos celulares. Las secuencias amino terminales son responsables del direccionamiento hacia el RE, el apoplasto y la secreción extracelular a partir de células de la aleurona (Koehler y Ho, *Plant Cell* 2: 769-783 (1990)). Adicionalmente, las secuencias amino terminales junto con las secuencias carboxilo terminales son responsables del direccionamiento vacuolar de los productos génicos (Shinshi y col. *Plant Molec. Biol.* 14: 357-368 (1990)).
- 15 **[0192]** Mediante la fusión de las secuencias de direccionamiento apropiadas descritas anteriormente con secuencias de transgenes de interés es posible dirigir el producto transgénico a cualquier orgánulo o compartimento celular. Para el direccionamiento a cloroplastos, por ejemplo, la secuencia señal de cloroplastos del gen RUBISCO, el gen CAB, el gen de sintetasa EPSP o el gen GS2 se fusiona en marco con la secuencia ATG amino terminal del transgén. La secuencia señal seleccionada debería incluir el sitio de escisión conocido y la fusión construida debería tener en cuenta cualquier aminoácido después del sitio de escisión requerido para la escisión. En algunos casos este requisito puede cumplirse mediante la adición de un número pequeño de aminoácidos entre el sitio de escisión y el transgén ATG o, alternativamente, sustitución de algunos aminoácidos dentro de la secuencia del transgén. La eficacia de la captación en cloroplastos de las fusiones construidas para importar a los cloroplastos puede probarse mediante la traducción de las construcciones transcritas *in vitro* seguido por la captación en cloroplastos usando técnicas descritas por Bartlett y col. En: Edlmann y col. (Eds) *Methods in Chloroplast Molecular Biology*, Elsevier pag. 1081-1091 (1982) y Wasmann y col. *Mol. Gen. Genet.* 205: 446-453 (1986). Estas técnicas de construcción son bien conocidas en la técnica y son igualmente aplicables a la mitocondria y a los peroxisomas.
- 20 **[0193]** Los mecanismos descritos anteriormente para el direccionamiento celular pueden utilizarse no solo en conjunto con sus promotores relacionados, sino también en conjunto con promotores heterólogos de modo que se lleve a cabo un objetivo de direccionamiento celular específico bajo la regulación transcripcional de un promotor que tiene un patrón de expresión diferente al del promotor del cual deriva la señal de direccionamiento.

35 C. Construcción de vectores de transformación vegetal

- [0194]** Los expertos en las técnicas de transformación de plantas conocen numerosos vectores de transformación disponibles para la transformación de plantas y los genes pertinentes para esta invención pueden usarse en conjunto con cualquiera de estos vectores. La selección del vector dependerá de la técnica de transformación específica y las especies diana para la transformación. Para determinadas especies diana, pueden ser específicos diferentes marcadores de selección de antibióticos o de herbicidas. Entre los marcadores de selección usados de forma rutinaria en la transformación se incluyen el gen nptII, que confiere resistencia a la kanamicina y a antibióticos relacionados (Messing y Vierra, *Gene* 19: 259-268 (1982); Bevan y col., *Nature* 304:184-187 (1983)), el gen bar, que confiere resistencia al herbicida fosfinotricina (White y col., *Nucl. Acids Res.* 18: 1062 (1990), Spencer y col. *Theor. Appl. Genet* 79: 625-631 (1990)), el gen hph, que confiere resistencia al antibiótico higromicina (Blochinger y Diggelmann, *Mol Cell Biol* 4: 2929-2931) y el gen dhfr, que confiere resistencia al metotrexato (Bourouis y col., *EMBO J.* 2(7): 1099-1104 (1983)), el gen EPSPS, que confiere resistencia al glifosato (patentes de EE. UU. N.º 4.940.935 y 5.188.642) y el gen de la manosa-6-fofato isomerasa que proporciona la capacidad para metabolizar la manosa (patentes de EE. UU. N.º 5.767.378 y 5.994.629).

50 1. Vectores adecuados para la transformación de *Agrobacterium*

- [0195]** Se dispone de muchos vectores para la transformación usando *Agrobacterium tumefaciens*. Estos típicamente llevan al menos una secuencia límite de ADN-T e incluye vectores como pBIN19 (Bevan. *Nucl. Acids Res.* (1984)). A continuación se describe la construcción de dos vectores típicos adecuados para la transformación de *Agrobacterium*.

a. pCIB200 y pCIB2001:

- 60 **[0196]** Los vectores binarios pCIB200 y pCIB2001 se usan para la construcción de vectores recombinantes para su uso con *Agrobacterium* y se construyen de la siguiente forma. Se crea pTJS75kan mediante la digestión NarI de pTJS75 (Schimidhauser y Helinski, *J. Bacteriol.* 164: 446-455 (1985)) lo que permite la escisión del gen de resistencia a tetraciclina, seguido de la inserción de un fragmento Accl a partir de pUC4K portador de un sitio NPTII (Messing y Vierra, *Gene* 19: 259-268 (1982); Bevan y col., *Nature* 304: 184-187 (1983); McBride y col., *Plant Molecular Biology* 14: 266-276 (1990)). Los enlazadores XhoI se ligan al fragmento EcoRV de pCIB7 que contiene las secuencias límites izquierdo y derecho del ADN-T, un gen quimérico nos/nptII vegetal seleccionable y el

polienlazador pUC (Rothstein y col., Gene 53: 153-161 (1987)) y el fragmento digerido con XhoI se clona en pTSJ75kan digerido con Sall para crear pCIB200 (véase también el documento EP 0 332 104, ejemplo 19). pCIB200 contiene los siguientes sitios de restricción polienlazadores exclusivos: EcoRI, SstI, KpnI, BglII, XbaI y Sall. pCIB2001 es un derivado de pCIB200 creado mediante la inserción dentro del polienlazador de sitios de restricción adicionales. Los sitios de restricción exclusivos en el polienlazador de pCIB2001 son EcoRI, SstI, KpnI, BglII, XbaI, Sall, MluI, BclI, AvrII, ApaI, HpaI y StuI. pCIB2001, además de contener estos sitios de restricción exclusivos también tienen selección de kanamicina vegetal y bacteriana, límites izquierdo y derecho del ADN-T para la transformación mediada por *Agrobacterium*, la función trfA derivada de RK2 para movilización entre *E. coli* y otros huéspedes y las funciones OriT y OriV también de RK2. El polienlazador de pCIB2001 es adecuado para la clonación de casetes de expresión en plantas que contienen sus propias señales reguladoras.

b. pCIB10 y derivados de selección de higromicina del mismo:

[0197] El vector binario pCIB10 contiene un gen que codifica resistencia a la kanamicina para la selección en plantas y secuencias límites derecha e izquierda del ADN-T e incorpora secuencias a partir del plásmido de amplia gama de huésped pRK252 permitiendo que se replique tanto en *E. coli* como en *Agrobacterium*. Su construcción se describe en Rothstein y col. (Gene 53: 153-161 (1987)). Se construyen diversos derivados de pCIB10 que incorporan el gen para la higromicina B fosfotransferasa descrito por Gritz y col. (Gene 25: 179-188 (1983)). Estos derivados permiten la selección de células vegetales transgénicas sólo para higromicina (pCIB743) o para higromicina y kanamicina (pCIB715, pCIB717).

2. Vectores adecuados para la transformación sin *Agrobacterium*

[0198] La transformación sin el uso de *Agrobacterium tumefaciens* elude el requisito de secuencias de ADN-T en el vector de transformación elegido y, por consiguiente, pueden usarse vectores carentes de estas secuencias además de vectores como los descritos anteriormente que contenían secuencia ADN-T. Las técnicas de transformación que no dependen de *Agrobacterium* incluyen la transformación mediante bombardeo de partículas, captación de protoplastos (p. ej., PEG y electroporación) y microinyección. La elección del vector depende en gran medida de la selección específica para la especie que se va a transformar. A continuación se describe la construcción de dos vectores típicos adecuados para la transformación de *Agrobacterium*.

a. pCIB3064:

[0199] pCIB3064 es un vector derivado de pUC adecuado para técnicas directas de transferencia de genes en combinación con la selección por el herbicida basta (o fosfinotricina). El plásmido pCIB246 comprende el promotor CaMV 35S en función operacional con el gen GUS de *E. coli* y el terminador de transcripción 35S de CaMV y se describe en la solicitud de PCT publicada WO 93/07278. El promotor 35S de este vector contiene dos secuencias ATG 5' del sitio de inicio. Estos sitios se mutan usando técnicas de PCR convencionales de manera que se eliminan los ATG y se generan los sitios de restricción SspI y PvuII. Los nuevos sitios de restricción están a 96 y 37 pb del único sitio Sall y a 101 y 42 pb del sitio de inicio real. El derivado resultante de pCIB246 se designa pCIB3025. El gen GUS se escinde a continuación a partir de pCIB3025 mediante digestión con Sall y SacI, se recortan los extremos para hacerlos romos y se ligan de nuevo para generar el plásmido pCIB3060. El plásmido pJIT82 se obtiene a partir del Centro Jonh Innes, Norwich y el fragmento Smal de 400 pb que contiene el gen bar de *Streptomyces viridochromogenes* se escinde e inserta en el sitio HpaI de pCIB3060 (Thompson y col. EMBO J 6: 2519-2523 (1987)). Esto generó pCIB3064, que comprende el gen bar bajo en control del promotor 35S de CaMV y el terminador para la selección de herbicidas, un gen de resistencia a ampicilina (para la selección en *E. coli*) y un polienlazador con los sitios únicos SphI, PstI, HindIII y BamHI. Este vector es adecuado para la clonación de casetes de expresión de plantas que contienen sus propias señales reguladoras.

b. pSOG19 y pSOG35:

[0200] pSOG35 es un vector de transformación que utiliza el gen de la dihidrofolato reductasa (DFR) de *E. coli* como marcador seleccionable que confiere resistencia a metotrexato. Se usa la PCR para amplificar el promotor 35S (-800 pb), el intrón 6 del gen Adh1 del maíz (-550 pb) y 18 pb de la secuencia líder no traducida GUS de pSOG10. Un fragmento de 250 pb que codifica el gen de la dihidrofolato reductasa de tipo II de *E. coli* también se amplifica mediante PCR y estos dos fragmentos de PCR se acoplan con un fragmento SacI-PstI procedentes de pB1221 (Clontech) que comprende la estructura del vector pUC19 y el terminador de la nopalina sintetasa. El ensamblaje de estos fragmentos genera pSOG19 que contiene el promotor 35S en fusión con la secuencia del intrón 6, el líder GUS, el gen DHFR y el terminador de la nopalina sintetasa. La sustitución del líder GUS en pSOG19 con la secuencia líder del virus del moteado clorótico del maíz (MCMV) genera el vector pSOG35. pSOG19 y pSOG35 llevan el gen pUC para resistencia a ampicilina y tiene sitios HindIII, SphI, PstI y EcoRI disponibles para la clonación de sustancias extrañas.

3. Vector adecuado para la transformación de cloroplastos

[0201] Para la expresión de una secuencia de nucleótidos de la presente invención en plástidos de plantas, se usa el vector de transformación de plástidos pPH143 (documento WO 97/32011, ejemplo 36). La secuencia de nucleótidos se inserta en pPH143 sustituyendo de este modo la secuencia codificadora PROTOX. A continuación, este vector se usa para la transformación del plástido y la selección de transformantes para la resistencia a la espectinomicina. Alternativamente, la secuencia de nucleótidos se inserta en pPH143 de modo que esta sustituye al gen *aadH*. En este caso, se seleccionan los transformantes para resistencia a inhibidores de PROTOX.

10 D. Transformación

[0202] Una vez que se ha clonado una secuencia de ácido nucleico en un sistema de expresión, se transforma dentro de una célula vegetal. Los casetes del receptor y de expresión dirigida de la presente invención pueden introducirse en la célula vegetal en diversas formas reconocidas por la técnica. Los procedimientos para la regeneración de plantas también son bien conocidos en la técnica. Por ejemplo, los vectores plasmídicos Ti se han utilizado para la administración de ADN extraño así como en la captación dirigida de ADN, así como en la captación de ADN dirigida, liposomas, electroporación, microinyección y microproyectiles. Además, las bacterias del género *Agrobacterium* pueden utilizarse para transformar células vegetales. A continuación se encuentran descripciones de las técnicas representativas para la transformación de plantas tanto dicotiledóneas como monocotiledóneas.

20

1. Transformación de dicotiledóneas

[0203] Las técnicas de transformación de dicotiledóneas son bien conocidas en la técnica e incluyen técnica basada en *Agrobacterium* y técnicas que no requieren *Agrobacterium*. Las técnicas sin *Agrobacterium* implican la captación de material genético exógeno directamente mediante protoplastos o células. Esto puede conseguirse mediante PEG o captación mediada por electroporación, administración mediada por bombardeo de partículas o microinyección. Ejemplos de estas técnicas se describen en Paszkowski y col., EMBO J 3: 2717-2722 (1984), Potrykus y col., Mol. Gen. Genet. 199: 169-177 (1985), Reich y col., Biotechnology 4: 1001-1004 (1986) y Klein y col., Nature 327: 70-73 (1987). En cada caso las células transformadas regeneraban las plantas completas usando técnicas conocidas en la técnica.

[0204] La transformación mediada por *Agrobacterium* es una técnica específica para la transformación de dicotiledóneas debido a su alta eficiencia de transformación y su amplia utilidad en muchas diferentes especies. La transformación de *Agrobacterium* típicamente implica la transferencia del vector binario portador del ADN extraño de interés (p. ej., pCIB200 o pCIB2001) a una cepa de *Agrobacterium* apropiada, lo que puede depender del complemento de genes vir portados por la cepa *Agrobacterium* huésped en un plásmido Ti corresidente o de forma cromosómica (p. ej., cepa CIB542 para pCIB200 y pCIB2001 (Uknes y col. Plant Cell 5: 159-169 (1993)). La transferencia del vector binario recombinante a *Agrobacterium* se consigue mediante un procedimiento de apareamiento triparental usando *E. coli* que porta el vector recombinante binario, una cepa de *E. coli* auxiliar portadora de un plásmido como pRK2013 y que es capaz de movilizar el vector binario recombinante hasta la cepa de *Agrobacterium* diana. Alternativamente, el vector binario recombinante puede transferirse a *Agrobacterium* mediante transformación de ADN (Höfgen y Willmitzer, Nucl. Acids Res. 16: 9877 (1988)).

[0205] La transformación de la especie vegetal diana mediante *Agrobacterium* recombinante normalmente implica el cultivo conjunto del *Agrobacterium* con explantes de la planta y siguiendo los protocolos bien conocidos en la técnica. El tejido transformado se regenera en un medio seleccionable que lleva el marcador de resistencia a antibióticos o herbicidas presente entre los límites del ADN-T del plásmido binario.

[0206] Otra técnica para transformar células vegetales con un gen implica introducir partículas inertes o biológicamente activas en los tejidos y células vegetales. Esta técnica se describe en las patentes de EE. UU. N.º 4.945.050, 5.036.006 y 5.100.792 todas de Sanford y col. Generalmente, este procedimiento implica introducir partículas inertes o biológicamente activas en las células en las condiciones eficaces para penetrar la superficie externa de la célula y permitir la incorporación al interior de la misma. Cuando se utilizan partículas inertes, el vector puede introducirse en la célula recubriendo las partículas con el vector que contiene el gen deseado. Alternativamente, la célula diana puede estar rodeada por el vector de modo que dicho vector se introduce en la célula mediante la estela de la partícula. Las partículas biológicamente activas (p. ej., células desecadas de levadura, bacterias secas o un bacteriófago, conteniendo cada uno el ADN que se desea introducir) también pueden introducirse en el tejido celular vegetal.

60 2. Transformación de monocotiledóneas

[0207] La transformación de la mayoría de las especies de monocotiledóneas se ha convertido también ahora en rutinaria. Entre las técnicas específicas se incluyen transferencia génica directa dentro de protoplastos usando PEG o técnicas de electroporación y bombardeo de partículas en el tejido del callo. Las transformaciones pueden llevarse a cabo con una única especie de ADN o con múltiples especies de ADN (es decir, transformación conjunta) y ambas técnicas son adecuadas para su uso con esta invención. La transformación conjunta puede tener la ventaja

de evitar la construcción completa del vector y generara las plantas transgénicas con locis no ligado para el gen de interés y el marcador seleccionable, lo que permite la retirada del marcador seleccionable en generaciones posteriores, esto debería contemplarse como deseable. Sin embargo, una desventaja del uso de transformación conjunta es la frecuencia menor del 100% con la que las especies de ADN independientes se integran en el genoma (Schocher y col. *Biotechnology* 4: 1093-1096 (1986)).

[0208] En las solicitudes de patente EP 0 292 435, EP 0 392 225 y WO 93/07278 se describen técnicas para la preparación de callos y protoplastos de una línea endogámica pura de maíz, la transformación de los protoplastos usando PEG o electroporación, y la regeneración de plantas de maíz a partir de protoplastos transformados. Gordon-Kamm y col. (*Plant Cell* 2: 603-618 (1990)) y Fromm y col. (*Biotechnology* 8: 833-839 (1990)) han publicado técnicas para la transformación de una línea de maíz derivado de A188 usando bombardeo de partículas. Adicionalmente, en el documento WO 93/07278 y en Koziel y col. (*Biotechnology* 11: 194-200 (1993)) se describen técnicas para la transformación de líneas endogámicas puras de maíz mediante bombardeo de partículas. Esta técnica utiliza embriones de maíz inmaduros de 1,5-2,5 mm de longitud escindidos a partir de una mazorca de maíz de 14-15 días tras la polinización y un dispositivo PDS-1000-He Biolistics para el bombardeo.

[0209] La transformación del arroz también puede llevarse a cabo mediante técnicas de transferencia génica directa utilizando protoplastos o bombardeo de partículas. La transferencia mediada por protoplasto ha sido descrita para tipos de *Japonica* y tipos de *Indica* (Zhanget y col. *Plant Cell Rep* 7: 379-384 (1988); Shimamoto y col. *Nature* 338: 274-277 (1989); Datta y col. *Biotechnology* 8: 736-740 (1990)). Ambos tipos también se transforman de forma rutinaria usando bombardeo de partículas (Christou y col. *Biotechnology* 9: 957-962 (1991)). Adicionalmente, en el documento WO 93/21335 se describen técnicas para la transformación del arroz mediante electroporación.

[0210] En la solicitud de patente EP 0 332 581 se describen técnicas para la generación, transformación y regeneración de protoplastos de *Pooideae*. Estas técnicas permiten la transformación de *Dactylis* y del trigo. Adicionalmente, la transformación del trigo ha sido descrita por Vasil y col. (*Biotechnology* 10: 667-674 (1992)) usando bombardeo de partículas dentro de células de callos regenerables a largo plazo de tipo C y también por Vasil y col. (*Biotechnology* 11: 1553-1558 (1993)) y Weeks y col. (*Plant Physiol.* 102: 1077-1084 (1993)) usando bombardeo de partículas de embriones inmaduros y callos derivados de embriones inmaduros. Sin embargo, una técnica específica para la transformación del trigo implica la transformación de este cereal mediante bombardeo de partículas de embriones inmaduros e incluye un paso de cultivo a alta concentración de sacarosa o de maltosa previo a la administración del gen. Antes del bombardeo, se coloca cualquier número de embriones (0,75-1 mm de longitud) en medio MS con sacarosa al 3% (Murashiga y Skoog, *Physiologia Plantarum* 15: 473-497 (1962)) y 3 mg/l de 2,4-D para la inducción de embriones somática, lo que permite proceder en oscuridad. El día elegido para el bombardeo, los embriones se sacan del medio de inducción y se colocan dentro del agente osmótico (es decir, medio de inducción con sacarosa o maltosa añadida a la concentración deseada, típicamente al 15%). Se deja que los embriones se plasmolizen durante 2-3 horas y, a continuación, se bombardean. Típicamente se disponen 20 embriones por placa diana, aunque esto no es crítico. Un plásmido portador del gen apropiado (como pCIB3064 o pSG35) se precipita sobre partículas de oro de tamaño micrométrico usando procedimientos estándar. Se dispara cada placa de embriones con el dispositivo de helio Biolistics® de DuPont usando una presión de estallido de ~1.000 psi con un filtro de 80 mesh estándar. Tras el bombardeo, los embriones se colocaron de nuevo en oscuridad para su recuperación durante aproximadamente 24 horas (todavía en el agente osmótico). Después de 24 horas, los embriones se retiraron del agente osmótico y se colocaron de nuevo en medio de inducción donde estuvieron durante aproximadamente un mes antes de su regeneración. Aproximadamente un mes después, los explantes de embrión que desarrollaron callo embriogénico se transfirieron a un medio de regeneración (MS + NAA a 1 mg/litro, GA a 5 mg/litro), que además contenía el agente de selección adecuado (basta a 10 mg/l en el caso de pCIB3064 y metotrexato a 2 mg/l en el caso de pSOG35). Después de aproximadamente un mes, los brotes desarrollados se transfirieron a recipientes estériles más grandes conocidos como «GA7» que contenían MS a la mitad de concentración, sacarosa al 2% y la misma concentración del agente de selección.

[0211] También se ha descrito la transformación de monocotiledóneas usando *Agrobacterium*. Véase el documento WO 94/00977 y la patente de EE. UU. N.º 5.591.616. Véase también, Negrotto y col., *Plant Cell Reports* 19: 798-803 (2000). Para este ejemplo, se usa arroz (*Oryza sativa*) para generar plantas transgénicas. Pueden usarse diversas variedades cultivadas de arroz (Hiei y col., 1994, *Plant Journal* 6:271-282, Dong y col., 1996, *Molecular Breeding* 2:267-276; Hiei y col., 1997, *Plant Molecular Biology*, 35:205-218). También, los diversos constituyentes de medios de cultivo descritos a continuación pueden variar en cantidad o ser sustituidos. Se inician las respuestas embriogénicas y/o se establecen los cultivos a partir de embriones maduros cultivando en medio MS-CIM (contenido inicial de sales de MS, 4,3 g/litro; vitaminas B5 (200x) a 5 ml/litro; sacarosa, 30 g/litros; prolina, 500 mg/litro; glutamina, 500 mg/litro; hidrolizado de caseína, 300 mg/litro; 2,4-D (1 mg/ml), 2 ml/litro, ajustando el pH a 5,8 con KOH 1 N, Phytigel, 3 g/litro). Los embriones maduros en los estadios iniciales de respuesta de cultivo o las líneas de cultivo establecidas se inoculan y cultivan conjuntamente con la cepa LBA4404 de *Agrobacterium tumefaciens* (*Agrobacterium*) que contiene la construcción de vector deseada. *Agrobacterium* se cultiva a partir de las soluciones madre de glicerol en medio YPC sólido (espectinomina a 100 mg/l y cualquier otro antibiótico apropiado) durante ~2 días a 28°C. *Agrobacterium* se resuspendió en medio MS-CIM líquido. El cultivo de *Agrobacterium* se diluye hasta una D.O. a 600 nm de 0,2-0,3 y se añade acetosiringona hasta una concentración final de 200 µM. Se añade acetosiringona antes de mezclar la solución con los cultivos de arroz para inducir al

Agrobacterium para la transferencia de ADN a las células vegetales. Para la inoculación, los cultivos vegetales se sumergen en la suspensión bacteriana. La suspensión bacteriana líquida se elimina y las células inoculadas se colocan sobre el medio de cocultivo y se incuban a 22°C durante dos días. A continuación, los cultivos se transfieren a medio MS-CIM con Ticarclina (400 mg/litro) para inhibir el crecimiento de *Agrobacterium*. Para las construcciones que utilizan el gen marcador selectivo PMI (Reed y col., In Vitro Cell. Dev. Biol. Plant 37:127-132), los cultivos se transfieren a un medio de selección que contiene manosa como fuente de hidratos de carbono (MS con manosa al 2% y Ticarclina a 300 mg/litro) después de 7 días y se cultivan durante 3-4 semanas en oscuridad. A continuación, las colonias resistente se transfieren a medio de inducción de regeneración (MS sin 2,4-D, IAA a 0,5 mg/litro, zeatina a 1 mg/litro, timentina a 200 mg/litro, manosa al 2% y sorbitol al 3%) y se crecen en oscuridad durante 14 días. A continuación, las colonias que proliferan se transfieren a otra ronda de medios de inducción de regeneración y se cambian a la habitación de crecimiento con luz. Los brotes regenerados se transfieren a recipientes GA7 con medio GA7-1 (MS sin hormonas ni sorbitol al 2%) durante 2 semanas y, a continuación, se cambian al invernadero cuando hayan crecido lo suficiente y tengan las raíces adecuadas. Las plantas se transplantan a suelo en el invernadero (generación T0) donde crecen hasta la madurez, y se recogen las semilla de T1.

3. Transformación de plástidos

[0212] Las semillas de *Nicotiana tabacum* c.v. «Xanthi nc» se germinaron a razón de siete por placa en una matriz circular de 0,0254 metros en un medio de agar T y se bombardearon 12-14 días después de la siembra con partículas de tungsteno de 1 µm (M10, Biorad, Hercules, Calif.) recubiertas con ADN a partir de plástidos pPH143 y pPH145 esencialmente como se ha descrito (Svab, Z. y Maliga, P. (1993) PNAS 90, 913-917). Las plantas de semillero bombardeadas se incuban en medio T durante dos días después de los cuales se cortan las hojas y se colocan abaxial hacia arriba con luz brillante (350-500 µmoles de fotones/m²/s) en placas de medio RMOP (Svab, Z., Hajdukiewicz, P. y Maliga, P. (1990) PNAS 87, 8526-8530) que contienen 500 µg/ml de diclorhidrato de espectinomicina (Sigma, St. Louis, Mo.). Los brotes resistentes que aparecen bajos las hojas descoloridas de tres a ocho semanas después del bombardeo se subclonaron en el mismo medio selectivo, lo que permitió formar el callo y se aislaron y subclonaron los brotes secundarios. La segregación completa de las copias genómicas del plástido transformado (homoplasmicidas) en subclones independientes se evalúa mediante técnicas estándar de transferencias de ADN (Sambrook y col., (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor). El ADN celular total digerido con BamHI/EcoRI (Mettler, I. J. (1987) Plant Mol Biol Reporter 5, 346-349) se separa en geles de agarosa con tampón Tris-borato (TBE) al 1%, se transfiere a membranas de nailon (Amersham) y se incuba con las secuencias de ADN cebadas aleatorias marcadas con ³²P que se corresponden con un fragmento de ADN BamHI/HindIII de 0,7 kb a partir de pC8 que contiene una porción de la secuencia de direccionamiento del plástido rps7/12. Los brotes homoplásmicos se enraizaron en condiciones asépticas en medio MS/IBA que contenía espectinomicina (McBride, K. E. y col. (1994) PNAS 91, 7301-7305) y se transfieren al invernadero.

V. Cultivo y producción de semillas

40 A. Cultivo

[0213] Las plantas obtenidas mediante la transformación con una secuencia de ácido nucleico de la presente invención pueden ser cualquier de la amplia variedad de especies vegetales, incluyendo monocotiledóneas y dicotiledóneas; sin embargo, las plantas utilizadas en el procedimiento de la invención se seleccionan específicamente entre la lista de cosechas objetivo agrónomicamente importantes descritas anteriormente. La expresión de un gen de la presente invención en combinación con otras características importantes para la producción y calidad puede incorporarse a las líneas vegetales mediante el cultivo. En la materia se conocen estrategias y técnicas de cultivo. Véase, por ejemplo, Welsh J. R., Fundamentals of Plant Genetics and Breeding, John Wiley e hijos, NY (1981); Crop Breeding, Wood D. R. (Ed.) American Society of Agronomy Madison, Wis. (1983); Mayo O., The Theory of Plant Breeding, Segunda Edición, Clarendon Press, Oxford (1987); Singh, D. P., Breeding for Resistance to Diseases and Insect Pests, Springer-Verlag, NY (1986) y Wricke y Weber, Quantitative Genetics and Selection Plant Breeding, Walter de Gruyter y Co., Berlin (1986).

[0214] Las propiedades genéticas modificadas mediante ingeniería genética dentro de las semillas y plantas transgénicas descritas anteriormente se pasan mediante reproducción sexual o crecimiento vegetativo y pueden además mantenerse y propagarse a las plantas de la progenie. Generalmente, dicho mantenimiento y propagación utiliza procedimientos agrícolas conocidos desarrollados para cumplir los objetivos específicos como arado, cultivo y recolección. También pueden aplicarse procesos especializados como tecnologías hidropónicas o de invernadero. Durante el crecimiento, puesto que la cosecha es vulnerable al ataque y a los daños causados por insectos o infecciones así como por la competición con malas hierbas, se toman medidas para controlar dichas malas hierbas, enfermedades de plantas, insectos, nematodos y otras condiciones adversas para mejorar el rendimiento. Entre estas se incluyen medidas mecánicas como la labranza del suelo o eliminación de malas hierbas y de plantas infectadas, así como la aplicación de agroquímicos como herbicidas, fungicidas, gametocidas, nematocidas, reguladores del crecimiento, agentes para la maduración e insecticidas.

[0215] El uso de las propiedades genéticas avanzadas de las plantas transgénicas y las semillas según la

invención pueden además hacerse en el cultivo de plantas, el cual tiene por objetivo el desarrollo de plantas que mejoren propiedades como la tolerancia a plagas, herbicidas o estrés, mejora del valor nutritivo, aumento del rendimiento o mejora de la estructura lo que produce menos pérdidas debido al encamado o desgranado. Los diversos pasos del cultivo se caracterizan por la intervención humana bien definida como la selección de las líneas que se van a cruzar, el direccionamiento de la polinización de las líneas parentales o la selección de las plantas de la progenie apropiadas. Dependiendo de las propiedades deseadas, se toman diferentes medidas de cultivo. Las técnicas relevantes son bien conocidas en la técnica e incluyen, pero sin limitaciones la hibridación, endogamia, cultivo retrocruzado, cultivo multilínea, mezcla de variedad, hibridación interespecífica, técnicas de aneuploidia, etc. Entre las técnicas de hibridación también se incluyen la esterilización de plantas para obtener plantas masculinas o femeninas estériles mediante métodos mecánicos, químicos o bioquímicos. La polinización cruzada de una planta masculina estéril con polen de una línea diferente asegura que el genoma de la planta masculina estéril pero femenina fértil obtendrá propiedades homogéneas de ambas líneas parentales. Por tanto, las semillas y plantas transgénicas según la invención pueden usarse para el cultivo de líneas vegetales mejoradas en las que, por ejemplo, aumente de la eficacia de procedimientos convencionales, como el tratamiento con herbicidas o pesticidas, o permitan su dispensación con dichos procedimientos debido a sus propiedades genéticas modificadas. Alternativamente, pueden obtenerse nuevas cosechas con mejora de la tolerancia al estrés, que debido a su «equipo» genético optimizado, rinde un producto recolectado de mejor calidad que los productos que no eran capaces de tolerar condiciones de desarrollo adversas comparables.

20 B. Producción de semillas

[0216] En la producción de semillas, la calidad de germinación y la uniformidad de semillas son características esenciales del producto. Como es difícil mantener una cosecha libre de semillas de otras cosechas y de malas hierbas, para controlar las enfermedades transmitidas por semillas y producir semillas con buena germinación, los productores de semillas, con experiencia en la técnica de crecimiento, acondicionamiento y comercialización de semillas puras, han desarrollado prácticas de producción de semillas bastante extensiva y bien definidas, que se están experimentando en la técnica de crecimiento, acondicionamiento y venta de semillas puras. Por tanto, es una práctica común para el agricultor comprar semillas que cumplan los estándares de calidad específicos en lugar de usar la recogida de las semillas de su propia cosecha. El material de propagación utilizado como semillas se trata habitualmente con un recubrimiento protector que comprende herbicidas, insecticidas, fungicidas, bactericidas, nematocidas, moluscicidas o mezclas de ambos. Entre los compuestos utilizados habitualmente como agentes de recubrimiento protector se encuentran captano, carboxina, thiram (TMTD[®]), metalaxilo (Apron[®]) y metil pirimifos (Actellic[®]). Si se desea, estos compuestos se formulan conjuntamente con vehículos, tensioactivos o adyuvantes que promueven su aplicación empleados habitualmente en la técnica de la formulación para proporcionar protección frente al daño causado por bacterias, hongos o plagas animales. Los agentes de recubrimiento protectores pueden aplicarse mediante la impregnación del material de propagación con una formulación líquida o mediante el recubrimiento con una formulación húmeda o seca combinada. También son posible otros procedimientos de aplicación como el tratamiento directo de las yemas o el fruto.

40 VI. Alteración de la expresión de las moléculas de ácido nucleico

[0217] La alteración en la expresión de las moléculas de ácido nucleico de la presente invención se alcanza de una de las siguientes formas:

45 A. Supresión «sentido»

[0218] La alteración de la expresión de una secuencia de nucleótidos de la presente invención, específicamente la reducción de su expresión, se obtiene mediante la supresión «sentido» (referenciado en, p. ej., Jorgensen y col. (1996) *Plant Mol. Biol.* 31, 957-973). En este caso, la secuencia de nucleótidos de la presente descripción completa, o una porción de la misma está comprendida en una molécula de ADN. La molécula de ADN está unidad de forma operativa específicamente a un promotor funcional en una célula que comprende el gen diana, específicamente una célula vegetal, y se introduce en la célula, en la que la secuencia de nucleótidos puede expresarse. La secuencia de nucleótidos se inserta en la molécula de ADN en la «orientación sentido», lo que significa que puede transcribirse la cadena codificadora de la secuencia de nucleótidos. En un ejemplo específico, la secuencia de nucleótidos es completamente traducible y todas la información genética comprendida en la secuencia de nucleótidos, o una porción de la misma, se traduce en un polipéptido. En otro ejemplo específico, la secuencia de nucleótidos es parcialmente traducible y se traduce un péptido corto. En un ejemplo específico, esto se consigue insertando al menos un codón de parada prematuro en la secuencia de nucleótidos, que hace que la traducción se detenga. En otro ejemplo más específico, la secuencia de nucleótidos se transcribe pero no se obtiene ningún producto de traducción. Esto se consigue normalmente eliminando el codón de inicio, por ejemplo, el «ATG», del polipéptido codificado por la secuencia de nucleótidos. En un ejemplo específico adicional, la molécula de ADN que comprende la secuencia de nucleótidos o una porción de la misma, se integra de forma estable en el genoma de la célula vegetal. En otro ejemplo específico, la molécula de ADN que comprende la secuencia de nucleótidos, o una porción de la misma, está comprendida en una molécula que se replica a nivel extracromosómico.

[0219] En las plantas transgénicas que contienen una de las moléculas de ADN descritas inmediatamente

antes, se reduce específicamente la expresión de la secuencia de nucleótidos correspondiente a la secuencia de nucleótidos comprendida en las moléculas de ADN. Específicamente, la secuencia de nucleótidos en la molécula de ADN es idéntica al menos al 70% con la secuencia de nucleótidos cuya expresión se reduce, más específicamente es idéntica al menos al 80%, aún más específicamente idéntica al menos al 90%, aún más específicamente, idéntica al menos al 95%, todavía más específicamente idéntica al menos al 99%.

B. Supresión «complementaria»

[0220] En otro ejemplo específico, la alteración de la expresión de una secuencia de nucleótidos de la presente invención, específicamente la reducción de su expresión, se obtiene mediante supresión «complementaria». La secuencia de nucleótidos de la presente invención completa, o una porción de la misma, está comprendida en una molécula de ADN. La molécula de ADN está unida de forma operativa específicamente a un promotor funcional en una célula vegetal, y se introduce en una célula vegetal, en la que la secuencia de nucleótidos puede expresarse. La secuencia de nucleótidos se inserta en la molécula de ADN en la «orientación complementaria», lo que significa que puede transcribirse la secuencia complementaria inversa (también denominada en ocasiones cadena no codificadora) de la secuencia de nucleótidos. En un ejemplo específico, la molécula de ADN que comprende la secuencia de nucleótidos, o una porción de la misma, se integra de forma estable en el genoma de la célula vegetal. En otro ejemplo específico, la molécula de ADN que comprende la secuencia de nucleótidos, o una porción del mismo, está comprendida en una molécula que se replica a nivel extracromosómico. Se citan varias publicaciones en las que se describe esta estrategia para su ilustración adicional (Green, P. J. y col., *Ann. Rev. Biochem.* 55:569-597 (1986); van der Krol, A. R. y col., *Antisense Nuc. Acids & Proteins*, pág. 125-141 (1991); Abel, P. P. y col., *PNAS Proc. Natl. Acad. Sci. USA* 86:6949-6952 (1989); Ecker, J. R. y col., *Proc. Natl. Acad. Sci. USA* 83:5372-5376 (agosto de 1986)).

[0221] En las plantas transgénicas que contienen una de las moléculas de ADN descritas inmediatamente antes, se reduce específicamente la expresión de la secuencia de nucleótidos correspondiente a la secuencia de nucleótidos comprendida en las moléculas de ADN. Específicamente, la secuencia de nucleótidos en la molécula de ADN es idéntica al menos al 70% con la secuencia de nucleótidos cuya expresión se reduce, más específicamente es idéntica al menos al 80%, aún más específicamente idéntica al menos al 90%, aún más específicamente, idéntica al menos al 95%, todavía más específicamente idéntica al menos al 99%.

C. Recombinación homóloga

[0222] En otra realización específica, al menos una copia genómica correspondiente con una secuencia de nucleótidos de la presente invención se modifica en el genoma de la planta mediante recombinación homóloga como se muestra adicionalmente en Paszkowski y col., *EMBO Journal* 7:4021-26 (1988). Esta técnica utiliza la propiedad de las secuencias homólogas para reconocerse entre sí e intercambiar las secuencias de nucleótidos entre ellas mediante un proceso conocidos en la técnica como recombinación homóloga. La recombinación homóloga puede producirse entre la copia cromosómica de una secuencia de nucleótidos en una célula y una copia entrante de la secuencia de nucleótidos introducida en la célula mediante transformación. Por tanto, se introducen modificaciones específicas con exactitud en la copia cromosómica de la secuencia de nucleótidos. En un ejemplo, los elementos reguladores de la secuencia de nucleótidos de la presente descripción están modificados. Estos elementos reguladores se obtienen fácilmente mediante análisis de una biblioteca genómica usando la secuencia de nucleótidos de la presente invención, o una porción de la misma, como sonda. Los elementos reguladores existentes se sustituyen por diferentes elementos reguladores alterando, de este modo, la expresión de la secuencia de nucleótidos, o estos se mutan o someten a delección para eliminar de este modo la expresión de la secuencia de nucleótidos. En otro ejemplo, la secuencia de nucleótidos se modifica mediante delección de una parte de la secuencia de nucleótidos o de la secuencia de nucleótidos completa, o mediante mutación. En la presente descripción también se contempla la expresión de un polipéptido mutado en una célula vegetal. Se han descrito refinamientos más recientes de esta técnica para alterar genes endógenos (Kempin y col., *Nature* 389:802-803 (1997) y Miao y Lam, *Plant J.*, 7:359-365 (1995)).

[0223] En otro ejemplo específico, se introduce una mutación en la copia cromosómica de una secuencia de nucleótidos transformando una célula con un oligonucleótido quimérico compuesto por una extensión contigua de restos de ARN y ADN en una conformación dúplex con caperuzas dobles en horquilla en los extremos. Una característica adicional de los oligonucleótidos es, por ejemplo, la presencia de 2'-O-metilación en los restos del ARN. La secuencia de ARN/ADN se diseña para que esté alineada con la secuencia de una copia cromosómica de una secuencia de nucleótidos de la presente descripción y para que contenga el cambio de nucleótido deseado. Por ejemplo, esta técnica se ilustra adicionalmente en la patente de EE. UU. N.º 5.501.967 y en Zhu y col. (1999) *Proc. Natl. Acad. Sci. USA* 96: 8768-8773.

D. Ribozimas

[0224] En un ejemplo adicional, el ARN que codifica un polipéptido de la presente descripción se escinde mediante ARN catalítico, o ribozima, específico de dicho ARN. La ribozima se expresa en plantas transgénicas y tiene como resultado la reducción de las cantidades de ARN que codifica el polipéptido de la presente descripción en

células vegetales llevando de este modo a la reducción de las cantidades de polipéptido acumulado en las células. Este procedimiento se ilustra adicionalmente en la patente de EE. UU. N.º 4.987.071.

E. Mutantes dominantes negativos

5

[0225] En otro ejemplo adicional, se cambia la actividad del polipéptido codificado por las secuencias de nucleótidos de esta descripción. Esto se consigue mediante la expresión de mutantes dominantes negativos de las proteínas en plantas transgénicas, lo que lleva a la pérdida de la actividad de la proteína endógena.

10 F. Aptámeros

[0226] En un ejemplo adicional, la actividad del polipéptido de la presente descripción se inhibe mediante la expresión en plantas transgénicas de ligandos de ácido nucleico denominados aptámeros, que se unen específicamente a la proteína. Los aptámeros se obtienen preferiblemente por el procedimiento de evolución sistemática de ligandos por enriquecimiento exponencial (SELEX). En el procedimiento SELEX, una mezcla candidata de ácidos nucleicos de cadena sencilla que tienen regiones de secuencia aleatorizada se pone en contacto con la proteína y aquellos ácidos nucleicos que presentan un aumento de la afinidad por la diana se separan del resto de la mezcla candidata. Los ácidos nucleicos separados se amplifican para obtener una mezcla enriquecida en el ligando. Tras varias repeticiones se obtiene un ácido nucleico con afinidad óptima por el polipéptido y se usa para su expresión en plantas transgénicas. Este procedimiento se ilustra adicionalmente en la patente de EE. UU. N.º 5.270.163.

G. Proteínas con dedos de cinc

25 **[0227]** Una proteína con dedos de cinc que se une a una secuencia de ácido nucleico de la presente descripción o a su región reguladora también se usa para alterar la expresión de la secuencia de nucleótidos. Específicamente, se reduce o incrementa la transcripción de la secuencia de nucleótidos. Las proteínas con dedos de cinc se describen, por ejemplo, en Beerli y col. (1998) PNAS 95:14628-14633 o en los documentos WO 95/19431, WO 98/54311 o WO 96/06166.

30

H. ARNbc

[0228] La alteración de la expresión de una secuencia de nucleótidos de la presente descripción también se obtiene mediante interferencia con ARNbc como se describe, por ejemplo, en los documentos WO 99/32619, WO 35 99/53050 o WO 99/61631. En otro ejemplo específico, la alteración de la expresión de una secuencia de nucleótidos de la presente descripción, específicamente la reducción de su expresión, se obtiene mediante un ARN bicatenario (ARNbc) de interferencia. La secuencia de nucleótidos de la presente invención completa, o una porción de la misma, está comprendida en una molécula de ADN. El tamaño de la molécula de ADN es específicamente de 100 a 1.000 nucleótidos o más, determinándose empíricamente el tamaño óptimo. Se unen dos copias de la molécula de ADN idénticas, separadas por una molécula de ADN espaciadora, de modo que las copias primera y segunda están en orientaciones opuestas. En el ejemplo específico, la primera copia de la molécula de ADN está en la cadena complementaria inversa (también conocida como cadena no codificadora) y la segunda copia es la cadena codificadora; en el ejemplo más específico, la primera copia es la cadena codificadora y la segunda copia es la cadena complementaria inversa. El tamaño de la molécula de ADN espaciadora es específicamente de entre 200 y 45 10.000 nucleótidos, más específicamente de 400 a 5.000 nucleótidos y, lo más específico, de 600 a 1.500 nucleótidos de longitud. El espaciador es, específicamente, un fragmento aleatorio de ADN, más específicamente un fragmento aleatorio de ADN sin homología con el organismo diana del ARNbc de interferencia y, más específicamente, un intrón funcional que el organismo diana ajusta de forma eficaz. Las dos copias de la molécula de ADN separadas por el espaciador están unidas de forma operativa a un promotor funcional en una célula vegetal, 50 y se introduce en una célula vegetal, en la que la secuencia de nucleótidos se puede expresar. En un ejemplo específico, la molécula de ADN que comprende la secuencia de nucleótidos, o una porción de la misma, se integra de forma estable en el genoma de la célula vegetal. En otro ejemplo específico, la molécula de ADN que comprende la secuencia de nucleótidos, o una porción de la misma, está comprendida en una molécula que se replica a nivel extracromosómico. Se citan varias publicaciones en las que se describe esta estrategia para obtener más información (Waterhouse y col. (1998) PNAS 95:13959-13964; Chuang y Meyerowitz (2000) PNAS 97:4985-4990; Smith y col. (2000) Nature 407:319-320). La alteración de la expresión de una secuencia de nucleótidos mediante ARNbc de interferencia también se describe, por ejemplo, en los documentos WO 99/32619, WO 99/53050 o WO 99/61631.

60 **[0229]** En las plantas transgénicas que contienen una de las moléculas de ADN descritas inmediatamente antes, se reduce específicamente la expresión de la secuencia de nucleótidos correspondiente a la secuencia de nucleótidos comprendida en las moléculas de ADN. Específicamente, la secuencia de nucleótidos en la molécula de ADN es idéntica al menos al 70% con la secuencia de nucleótidos cuya expresión se reduce, más específicamente es idéntica al menos al 80%, aún más específicamente idéntica al menos al 90%, aún más específicamente, idéntica 65 al menos al 95%, todavía más específicamente idéntica al menos al 99%.

I. Inserción de una molécula de ADN (mutagénesis insercional)

[0230] En otro ejemplo específico, se inserta una molécula de ADN dentro de una copia cromosómica de una secuencia de nucleótidos de la presente invención, o dentro de una región reguladora de la misma. Específicamente, esta molécula de ADN comprende un elemento transponible capaz de sufrir transposición en una célula vegetal, como por ejemplo, Ac/Ds, Em/Spm, mutador. Alternativamente, la molécula de ADN comprende un límite de ADN-T de un ADN-T de *Agrobacterium*. La molécula de ADN también puede comprender un sitio de reconocimiento de recombinasa o integrasa que puede usarse para eliminar parte de la molécula de ADN del cromosoma de la célula vegetal. También se incluyen procedimientos de mutagénesis insercional usando ADN-T, transposones, oligonucleótidos u otros procedimientos conocidos por los expertos en la materia. Los procedimientos de uso del ADN-T y del transposón para mutagénesis insercional se describen en Winkler y col. (1989) *Methods Mol. Biol.* 82:129-136 y Martienssen (1998) *PNAS* 95:2021-2026.

J. Mutagénesis de delección

[0231] Aún en otro ejemplo, se crea una mutación de una molécula de ácido nucleico de la presente descripción en la copia genómica de la secuencia en la célula o planta mediante delección de una porción de la secuencia de nucleótidos o secuencia reguladora. Los procedimientos de mutagénesis de delección son conocidos por los expertos en la materia. Véase, por ejemplo, Miao y col., (1995) *Plant J.* 7:359.

[0232] Aún en otro ejemplo esta delección se crea de forma aleatoria en una población mayor de plantas mediante mutagénesis química o irradiación y se aísla una planta con una delección en un gen de la presente descripción mediante procedimiento de genética directa o inversa. Es sabido que la irradiación con neutrones rápidos o rayos gamma causan mutaciones por delección en plantas (Silverstone y col, (1998) *Plant Cell*, 10:155-169; Bruggemann y col., (1996) *Plant J.*, 10:755-760; Redei y Koncz in *Methods in Arabidopsis Research*, World Scientific Press (1992), pág. 16-82). Las mutaciones de delección en un gen de la presente descripción pueden recuperarse mediante una estrategia de genética inversa usando PCR con conjuntos agrupados de ADN genómicos como se ha mostrado en *C. elegans* (Liu y col., (1999), *Genoma Research*, 9:859-867). Una estrategia de genética directa podría implicar mutagénesis de una línea que exprese PTGS seguido de una selección de la progenie M2 para determinar la ausencia de PTGS. Entre estos mutantes podría preverse que en algunos de ellos se haya alterado un gen de la presente descripción. Esto podría evaluarse mediante transferencia de ARN o PCR para un gen de la presente invención con el ADN genómico a partir de estos mutantes.

K. Sobreexpresión en una célula vegetal

[0233] Aún en otra realización específica, se sobreexpresa una secuencia de nucleótidos de la presente invención que codifica un polipéptido. Anteriormente se describen ejemplos de moléculas de ácido nucleico y casetes de expresión para la sobreexpresión de una molécula de ácido nucleico. La presente invención también abarca procedimientos conocidos por los expertos en la materia de sobreexpresión de moléculas de ácido nucleico.

[0234] En una realización específica, la expresión de la secuencia de nucleótidos de la presente invención se altera en cada célula de una planta. Esto se obtiene, por ejemplo, mediante recombinación homóloga o inserción en el cromosoma. Esto se obtiene también por ejemplo expresando un ARN sentido o complementario, una proteína con dedos de cinc o una ribozima bajo el control de un promotor capaz de expresar el ARN sentido o complementario, la proteína con dedos de cinc o la ribozima en cada célula de una planta. La expresión constitutiva, inducible, específica de tejido o la expresión regulada durante el desarrollo también están dentro del alcance de la presente invención y tiene como resultado una alteración constitutiva, inducible, específica de tejido o regulada durante el desarrollo de la expresión de una secuencia de nucleótidos de la presente invención en la célula vegetal. Las construcciones para expresión del ARN sentido o complementaria, proteína con dedos de cinc o ribozima, o para la sobreexpresión de una secuencia nucleotídica de la presente invención, se preparan y transforman en una célula vegetal según las enseñanzas de la presente invención, por ejemplo, como se describe a continuación.

VII. Polipéptidos

[0235] La presente invención además se refiere a polipéptidos aislados que comprenden la secuencia de aminoácidos de SEC ID N.º 2. En particular, los polipéptidos aislados que comprenden la secuencia de aminoácidos SEC ID N.º 2 y las variantes con modificaciones de aminoácidos conservadoras. Un experto en la materia reconocerá que las sustituciones, deleciones o adiciones individuales a una secuencia de ácido nucleico, péptido, polipéptido o proteína que altera, añade o elimina un único aminoácido o un pequeño porcentaje de aminoácidos en la secuencia codificada es una «modificación conservadora» en la que la modificación da lugar a la sustitución de un aminoácido por otro aminoácido químicamente similar. Las variantes con modificaciones conservadoras proporcionan una actividad biológica similar al polipéptido no modificado. En la técnica se conocen tablas de sustituciones conservadoras en las que se listan los aminoácidos con funcionalidad similar. Véase Crighton (1984) *Proteins*, W.H. Freeman and Company.

[0236] En un ejemplo específico un polipéptido que tiene similitud considerable con la secuencia polipeptídica

SEC ID N.º 2, o un exón o dominio de la misma, es una variante alélica de la secuencia polipeptídica listada en SEC ID N.º 2. En otro ejemplo específico, un polipéptido que tiene similitud considerable con una secuencia polipeptídica listada en SEC ID N.º 2, o un exón o dominio de la misma, es una variante natural de la secuencia polipeptídica listada en SEC ID N.º 2. En otro ejemplo específico, un polipéptido que tiene similitud considerable con la secuencia polipeptídica listada en SEC ID N.º 2, o un exón o dominio de la misma, es una variante polimórfica de la secuencia polipeptídica listada en SEC ID N.º 2.

[0237] En un ejemplo específico alternativo, la secuencia que tiene similitud considerable contiene una delección o inserción de al menos un aminoácido. En un ejemplo más específico la delección o inserción es de menos de aproximadamente diez aminoácidos. En un ejemplo más específico la delección o inserción es de menos de aproximadamente tres aminoácidos.

[0238] En un ejemplo específico la secuencia que tiene similitud considerable codifica una sustitución en al menos un aminoácido.

[0239] Los ejemplos de la presente invención también contemplan un polipéptido aislado que contiene una secuencia polipeptídica que incluye

a) una secuencia polipeptídica listada en SEC ID N.º 2 o un exón o dominio de la misma;

b) una secuencia polipeptídica que tiene similitud considerable con a);

c) una secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o que tiene una similitud considerable con la secuencia de nucleótidos listada en SEC ID N.º 1, o un exón o dominio de la misma, o una secuencia complementaria a la misma;

d) una secuencia polipeptídica codificada por una secuencia de nucleótidos capaz de hibridar en condiciones de rigurosidad media con una secuencia de nucleótidos listada en SEC ID N.º 1 o con una secuencia complementaria de la misma; o

e) un fragmento funcional de a), b), c) o d).

[0240] En otro ejemplo específico el polipéptido que tiene similitud considerable es una variante alélica de la secuencia polipeptídica listada en SEC ID N.º 2 o un fragmento, dominio, repetición o quimera de la misma. En otro ejemplo específico el ácido nucleico aislado incluye diversas regiones de la secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o con similitud considerable con la secuencia de nucleótidos listada en SEC ID N.º 1 o fragmento o dominio de la misma, o una secuencia complementaria a la misma.

[0241] En otro ejemplo específico el polipéptido es la secuencia polipeptídica listada en SEC ID N.º 2. En otro ejemplo específico, el polipéptido es un fragmento o dominio funcional. Aún en otro ejemplo específico el polipéptido es una quimera, donde la quimera puede incluir dominios proteicos funcionales, como dominios, repeticiones, sitios de modificación postraduccionales u otras características. En un ejemplo más específico el polipéptido es un polipéptido de una planta. En un ejemplo más específico la planta es una dicotiledónea. En un ejemplo más específico, la planta es una gimnosperma. En un ejemplo más específico, la planta es una monocotiledónea. En un ejemplo más específico la monocotiledónea es un cereal. En un ejemplo más específico, el cereal puede ser, por ejemplo, maíz, trigo, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef, milo, lino, grama, *Tripsacum* y teosinte. En otro ejemplo más específico el cereal es arroz.

[0242] En un ejemplo específico el polipéptido se expresa en una ubicación o tejido específico de una planta. En un ejemplo más específico la ubicación o tejido es, por ejemplo, pero sin limitaciones, epidermis, tejido vascular, meristema, cámbium, corteza o médula. En un ejemplo más específico la ubicación o tejido es la hoja o vaina, raíz, flor y óvulo o semilla en desarrollo. En un ejemplo más específico la ubicación o tejido puede ser, por ejemplo, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor. En el ejemplo más específico la ubicación o el tejido es una semilla.

[0243] En un ejemplo específico la secuencia polipeptídica codificada por una secuencia de nucleótidos que tiene similitud considerable con la secuencia de nucleótidos listada en SEC ID N.º 1 o un fragmento o dominio de la misma o una secuencia complementaria a la misma, incluye una delección o inserción de al menos un nucleótido. En un ejemplo más específico la delección o inserción es de menos de aproximadamente treinta nucleótidos. En el ejemplo más específico la delección o inserción es de menos de aproximadamente cinco nucleótidos.

[0244] En un ejemplo específico la secuencia polipeptídica codificada por una secuencia de nucleótidos que tiene similitud considerable con la secuencia de nucleótidos listada en SEC ID N.º 1, o un fragmento o dominio de la misma, o una secuencia complementaria a la misma, incluye una sustitución de al menos un codón. En un ejemplo más específico, la sustitución es conservadora.

[0245] En un ejemplo específico, las secuencias polipeptídicas que tienen similitud considerable con la secuencia polipeptídica listada en SEC ID N.º 2 o un fragmento, dominio, repetición o quimeras de la misma incluye una delección o inserción de al menos un aminoácido.

5 **[0246]** Los polipéptidos de la descripción, fragmentos o variantes de los mismos pueden comprender cualquier número de restos de aminoácidos contiguos de un polipéptido de la descripción, en el que el número de restos se selecciona entre el grupo de números enteros compuesto por desde 10 al número de restos de un polipéptido de longitud completa de la descripción. Específicamente, la porción o fragmento del polipéptido es una proteína funcional. La presente descripción incluye polipéptidos activos que tienen una actividad específica de al
10 menos el 20%, 30% o 40%, y específicamente, al menos el 50%, 60% o 70%, y más específicamente, al menos el 80%, 90% o 95% la del polipéptido endógeno nativo (no sintético). Adicionalmente, la especificidad de sustrato (kcat/Km) es, de forma opcional, sustancialmente similar al polipéptido endógeno nativo (no sintético). Típicamente, la Km será al menos del 30%, 40% o 50% la del polipéptido nativo endógeno y, más específicamente, al menos el
15 del sustrato son bien conocidas por los expertos en la materia.

[0247] Los polipéptidos aislados de la presente descripción inducirán la producción de un anticuerpo que específicamente reacciona con un polipéptido de la presente invención cuando se presente como inmunógeno. Por tanto, los polipéptidos de la presente invención pueden emplearse como inmunógenos para la construcción de
20 anticuerpos inmunorreactivos con una proteína de la presente descripción para los fines, pero sin limitaciones, de inmunoensayos o técnicas de purificación de proteínas. Los inmunoensayos para la determinación de la unión son bien conocidos por los expertos en la materia como por ejemplo, entre otros, ELISAS o inmunoensayos competitivos.

[0248] Los ejemplos de la presente descripción también se refieren a polipéptidos quiméricos codificados por
25 las moléculas de ácido nucleico aisladas de la presente descripción que incluyen un polipéptido quimérico que contiene una secuencia polipeptídica codificada por un ácido nucleico aislado que contenga una secuencia de nucleótidos que incluya:

a) una secuencia de nucleótidos listada en SEC ID N.º 1, o un exón o dominio de la misma;

30

b) una secuencia de nucleótidos que tiene similitud considerable con a);

c) una secuencia de nucleótidos capaz de hibridar con a);

35 d) una secuencia de nucleótidos que es complementaria a a), b) o c); y

e) una secuencia de nucleótidos que es la complementaria inversa de a), b) o c); o

f) un fragmento funcional de la misma.

40

[0249] Un polipéptido que contiene una secuencia polipeptídica codificada por un ácido nucleico aislado que contiene una secuencia de nucleótidos, su complementaria o su complementaria inversa, que codifica un polipéptido que incluye una secuencia polipeptídica que incluye:

45 a) una secuencia polipeptídica listada en SEC ID N.º 2 o un dominio, repetición o quimeras funcionales de la misma;

b) una secuencia polipeptídica que tiene similitud considerable con a);

c) una secuencia polipeptídica codificada por una secuencia de nucleótidos idéntica o que tiene una similitud
50 considerable con la secuencia de nucleótidos listada en SEC ID N.º 1, o un exón o dominio de la misma, o una secuencia complementaria a la misma;

d) una secuencia polipeptídica codificada por una secuencia de nucleótidos capaz de hibridar en condiciones de rigurosidad media con una secuencia de nucleótidos listada en SEC ID N.º 1 o con una secuencia complementaria
55 de la misma y un fragmento funcional de a), b), c) o d); o

e) un fragmento funcional de la misma.

[0250] Las moléculas aisladas de ácido nucleico de la presente descripción son útiles para la expresión de un
60 polipéptido de la presente descripción en una célula recombinantemente modificada por ingeniería genética, tal como una célula de bacteria, levadura, insecto, mamífero o vegetal. Las células producen el polipéptido en condiciones no naturales (p. ej., en cantidad, composición, ubicación y/o tiempo) debido a que se han alterado genéticamente para hacerlo. Los expertos en la materia son conocedores de los numerosos sistemas de expresión disponibles para la expresión de ácidos nucleicos que codifican una proteína de la presente descripción y no se
65 describirán en detalle.

[0251] Brevemente, la expresión de ácidos nucleicos aislados que codifican un polipéptido de la descripción se conseguirá típicamente, por ejemplo, uniendo de forma operativa el ácido nucleico o el ADNc a un promotor (constitutivo o regulable) seguido de su incorporación dentro de un vector de expresión. Los vectores son adecuados para su replicación y/o integración en procariotas o en eucariotas. Los vectores de expresión utilizados con mayor frecuencia comprende terminadores de transcripción y traducción, secuencias de inicio y promotores para la regulación de la expresión de la molécula de ácido nucleico que codifica el polipéptido. Para obtener altos niveles de expresión de la molécula de ácido nucleico clonada, es deseable usar vectores de expresión que comprendan un promotor potente para dirigir la transcripción, un sitio de unión a ribosomas para el inicio de la traducción y un terminador de transcripción/traducción. Un experto en la material reconocerá qué modificaciones pueden hacerse en el polipéptido de la presente descripción sin disminuir su actividad biológica. Pueden hacerse algunas modificaciones para facilitar la clonación, expresión o incorporación del polipéptido de la descripción dentro de una proteína de fusión. Estas modificaciones son bien conocidas en la técnica e incluyen, pero sin limitaciones, adición de una metionina al extremo amino terminal para proporcionar un sitio de inicio, o aminoácidos adicionales (p. ej., polihistidina) situados en cualquiera de los extremos para crear secuencias de purificación localizadas convenientemente. También pueden introducirse en el vector sitios de restricción o codones de terminación.

[0252] En una realización específica, el vector de expresión incluye uno o más elementos, como por ejemplo, pero sin limitaciones, una secuencia promotora-potenciadora, una secuencia marcadora de selección, un origen de replicación, una secuencia codificadora de epítipo etiquetado o una secuencia codificadora de purificación por afinidad etiquetada. En una realización más específica, la secuencia promotora-potenciadora puede ser, por ejemplo, el promotor 35S de CaMV, el promotor 19S de CaMV, el promotor PR-1a del tabaco, el promotor de la ubiquitina y el promotor de la faseolina. En otra realización, el promotor puede funcionar en plantas y, más específicamente, es un promotor constitutivo o inducible. En otra realización específica, la secuencia marcadora de selección codifica un gen de resistencia a antibiótico. En otra realización específica, la secuencia de epítipo etiquetado codifica V5, el péptido Phe-His-His-Thr-Thr, hemaglutinina o glutatión-S-transferasa. En otra realización específica, la secuencia de purificación por afinidad etiquetada codifica una secuencia de ácido poliamino o un polipéptido. En una realización más específica, la secuencia de ácido poliamino es polihistidina. En una realización más específica, el polipéptido es el dominio de unión a quitina o glutatión-S-transferasa. En una realización más específica, la secuencia de purificación por afinidad etiquetada comprende una secuencia codificadora de inteína.

[0253] Las células procariotas pueden usarse como células huésped, por ejemplo, pero sin limitaciones, *Escherichia coli* y otras cepas microbianas conocidas por los expertos en la materia. Los procedimientos para la expresión de proteínas en células procariotas son bien conocidos por los expertos en la materia y pueden encontrarse en muchos manuales de laboratorio como *Molecular Cloning: A Laboratory Manual*, de J. Sambrook y col. (1989, Cold Spring Harbor Laboratory Press). Existen a disposición de los expertos en la materia diversos promotores, sitios de unión a ribosomas y operadores para controlar la expresión así como marcadores seleccionables como genes de resistencia a antibióticos. El tipo de vector elegido es para permitir un crecimiento y una expresión óptimos en el tipo de célula seleccionado.

[0254] Están disponible diversos sistemas de expresión eucariotas como, pero sin limitaciones, levaduras, líneas celulares de insectos, células vegetales y células de mamíferos. La expresión y síntesis de proteínas heterólogas en levaduras son bien conocidas (véase Sherman y col., *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press, 1982). Las cepas de levaduras utilizadas normalmente para la producción de proteínas eucariotas son *Saccharomyces cerevisiae* y *Pichia pastoris* y vectores, cepas y protocolos para la expresión están disponible a partir de casas comerciales (p. ej., Invitrogen).

[0255] Los sistemas de células de mamíferos pueden transferirse con vectores de expresión para la producción de proteínas. Están disponibles muchas líneas celulares huésped adecuado para los expertos en la materia como, por ejemplo, pero sin limitaciones, las líneas celulares HEK293, BHK21 y CHO. Los vectores de expresión para estas células pueden incluir secuencias de control de la expresión como un origen de replicación, un promotor (p. ej., el promotor de CMV, un promotor tk de HSV o el promotor de la fosfoglicerato quinasa (pgk)), un potenciador y sitios para el procesamiento de la proteína como sitios de unión a ribosomas, sitio de ajuste del ARN, sitios de poliadenilación y secuencias de terminación de la transcripción. Otras líneas celulares animales útiles para la producción de proteínas están disponibles en el mercado o a partir de depositarios como la American Type Culture Collection.

[0256] Los vectores de expresión para que las proteínas se expresen en células de insectos normalmente derivan del baculovirus SF9 o de otros virus conocidos en la técnica. Están disponibles varias líneas celulares de insectos adecuadas entre las que se incluyen, pero sin limitaciones, líneas celulares de larvas de mosquito, gusano de seda, gusano soldado, polillas y de *Drosophila*.

[0257] Se conocen procedimientos de transfección de células animales y de eucariotas inferiores. Se usan numerosos procedimientos para hacer competentes a células eucariotas a la introducción de ADN como, pero sin limitaciones: precipitación con fosfato de calcio, fusión de la célula receptora con protoplastos de bacterias que contiene el ADN, tratamiento de las células receptoras con liposomas que contienen el ADN, dextrina DEAE, eletroporación, biobalística y microinyección del ADN directamente dentro de las células. Las células transfectadas

se cultivan usando sistemas bien conocidos en la técnica (véase, Kuchler, R. J., *Biochemical Methods in Cell Culture and Virology*, Dowden, Hutchison y Ross, Inc. 1997).

[0258] Una vez que un polipéptido de la presente descripción se haya expresado puede aislarse y purificarse a partir de las células usando procedimientos conocidos por los expertos en la materia. El proceso de purificación puede controlarse usando técnicas de inmunotransferencia u otras técnicas estándar de inmunoensayo. Las técnicas de purificación de proteínas son normalmente conocidas y utilizadas por los expertos en la materia (véase, R. Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag, Nueva York 1982; Deutscher, *Guide to Protein Purification*, Academic Press (1990). Los ejemplos de la presente descripción proporcionan un procedimiento de producción de una proteína recombinante en el que el vector de expresión incluye uno o más elementos que incluyen una secuencia promotora-potenciadora, una secuencia marcadora de selección, un origen de replicación, una secuencia codificadora de epítipo etiquetado y una secuencia codificadora de purificación por afinidad etiquetada. En una realización, la construcción de ácido nucleico incluye una secuencia codificadora de epítipo etiquetado y la etapa de aislamiento incluye el uso de un anticuerpo específico para el epítipo etiquetado. En otro ejemplo específico, la construcción de ácido nucleico contiene una secuencia codificadora de ácido poliamino y la etapa de aislamiento incluye el uso de una resina que comprende una sustancia que se une a ácido poliamino, específicamente cuando el ácido poliamino es polihistidina y la resina que une compuestos poliamino es una resina de agarosa cargada con níquel. Aún en otro ejemplo específico, la construcción de ácido nucleico contiene una secuencia codificadora de un polipéptido y la etapa de aislamiento incluye el uso de una resina que contiene una sustancia de unión a polipéptido, específicamente cuando el polipéptido es un dominio de unión a quitina y la resina contiene quitina-sefarosa.

[0259] Los polipéptidos de la presente invención pueden sintetizarse usando procedimientos sintéticos no celulares conocidos por los expertos en la materia. Las técnicas para la síntesis en fase sólida se describen en Barany y Mayfield, *Solid-Phase Peptide Synthesis*, pág. 3-284 en *Peptides: Analysis, Synthesis, Biology*, Vol. 2, *Special Methods in Peptide Synthesis, Parte A*; Merrifield, y col., *J. Am. Chem. Soc.* 85:2149-56 (1963) y Stewart y col., *Solid Phase Peptide Synthesis*, 2ª ed. Pierce Chem. Co., Rockford, Ill. (1984).

[0260] La presente invención además proporciona un procedimiento para la modificación (es decir, aumento) de la concentración o composición de los polipéptidos de la invención en una planta o parte de la misma. La modificación puede efectuarse aumentando la concentración y/o composición (es decir, la proporción de polipéptidos de la presente invención) en una planta. El procedimiento comprendía introducir en una célula vegetal un casete de expresión que comprende una molécula de ácido nucleico de la presente invención obtener una célula o tejido vegetal transformado y cultivar la célula o tejido vegetal transformado. La molécula de ácido nucleico puede estar bajo la regulación de un promotor constitutivo o inducible. El procedimiento pueden comprender además la inducción o represión de la expresión de una molécula de ácido nucleico de una secuencia en la planta durante tiempo suficiente para modificar la concentración y/o composición en la planta o parte de la planta.

[0261] Una planta o parte de la planta que tenga la expresión modificada de una molécula de ácido nucleico de la invención puede analizarse y seleccionarse usando los procedimientos conocidos por los expertos en la materia tales como, pero sin limitaciones, transferencia de ARN, secuenciación de ADN o análisis mediante PCR usando cebadores específicos para la molécula de ácido nucleico y detectando amplicones producidos a partir de estos.

[0262] En general, la concentración o composición aumenta o disminuye en al menos el 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% o 90% en relación con una planta nativa control, parte de la planta o célula que carece del casete de expresión.

[0263] Los azúcares son los reguladores principales de muchos procesos vitales en plantas fotosintéticas, como fotosíntesis y metabolismo del carbono y del nitrógeno, y su regulación se consigue regulando la expresión génica, activando o reprimiendo los genes implicados. No se conocen bien los mecanismos por los que los azúcares controlan la expresión génica. El factor de transcripción GATA descrito en este documento está implicado en la regulación de la detección de azúcares y la expresión del propio factor está influenciada por el cambio del estado del N. El aumento de la expresión de este gen puede producir plantas con aumento del rendimiento, especialmente porque la manipulación de la rutas de señalización de azúcares puede llevar a un aumento de la fotosíntesis y a un aumento de la utilización del nitrógeno, y a alterar las relaciones fuente-sumidero en semillas, vasos, raíces y otros órganos de almacenamiento.

[0264] La invención se describirá más detalladamente en referencia a los siguientes ejemplos detallados. Estos ejemplos se proporcionan con fines solo ilustrativos y no pretenden ser limitantes siempre que no se especifique otra cosa.

EJEMPLOS

[0265] Las técnicas estándar de ADN recombinante y clonación molecular utilizadas en este documento son bien conocidas en la técnica y se describen en J. Sambrook y col., *Molecular Cloning: A Laboratory Manual*, 3er Ed., Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press (2001); de T. J. Silhavy, M. L. Berman y L. W.

Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) y de Ausubel, F. M. y col., Current Protocols in Molecular Biology, Nueva York, John Wiley e hijos Inc., (1988), Reiter, y col., Methods in *Arabidopsis* Research, World Scientific Press (1992) y Schultz y col., Plant Molecular Biology Manual, Kluwer Academic Publishers (1998).

5

ANTECEDENTES EXPERIMENTALES Y PROCEDIMIENTOS

[0266] A. Determinación de las condiciones de crecimiento del arroz y el maíz en condiciones limitantes de nitrógeno

10

[0267] En pasados experimentos realizados para estudiar los genes implicados en la captación y asimilación de nitratos, los presentes inventores y otros grupos utilizaron condiciones de crecimiento en las que los nitratos estaban presentes, generalmente, en exceso o ausentes en su totalidad. En este último caso, el nitrato se añade típicamente a plantas en crecimiento en ausencia de este elemento para entender la regulación del nitrato de estos y otros genes. Mientras que este tipo de tratamiento extremo es útil para definir algunos aspectos de la regulación génica, no es adecuado para comprender mejor el efecto de la limitación del nitrógeno. Los inventores han definido condiciones para *Arabidopsis* en las que el nitrógeno limita el crecimiento. Esto supuso desarrollar un sistema usando Rockwool (Hirai y col., 1995 Plant Cell Physiol 36, 1331-1339) y definir tres condiciones: una donde el crecimiento es máximo; otra donde el nitrógeno limita el crecimiento al 70-75% de los niveles de crecimiento máximo; una tercera donde existe una limitación mas severa hasta del 30-35% de los niveles de crecimiento máximo. La limitación del nitrógeno actúa como elemento de «estrés» con la posibilidad de variar fácilmente la cantidad de «estrés» alterando la concentración de nitrato. Los inventores probaron el «estado del nitrógeno» fisiológico midiendo el nitrato, clorofila (que a menudo se utiliza como reflexión del estado del nitrógeno en condiciones de campo (véase, p. ej., Fox R H y col 2001 Agron J. 93, 590-597; Minotti P L y col 1994 Hort Science 29, 1497-1550)), niveles de aminoácidos y actividades nitrato reductasa y glutamina sintetasa para proporcionar un valor basal con respecto al cual evaluar los estudios sobre líneas mutantes.

B. Experimentos de obtención de perfiles de expresión en plantas de *Arabidopsis* con limitación de nitrógeno

[0268] El perfil de expresión del transcripto puede usarse para comprobar los niveles de ARN de un gran número de genes al mismo tiempo. En el pasado se han llevado a cabo un gran número de experimentos de este tipo y, si el sistema experimental es flexible, estos pueden usarse para precisar el «estado de expresión» de un organismo en condiciones diferentes y usar esta información para obtener una hipótesis sobre qué genes y vías están implicadas en los diversos procesos. Los inventores encontraron que cuanto más profundas eran las diferencias en las condiciones de crecimiento, mayores eran también las diferencias en los perfiles de transcripción entre las plantas creciendo en estas condiciones y más difícil era descifrar cuales eran los cambios más importantes. El único experimento de obtención de un perfil genómico completo publicado en este área es uno en *Arabidopsis* en el que se estudió una variación extrema en los niveles de nitrato (Wang R y col., 2003 Plant Physiol. 132, 556-67). En el caso de limitaciones de nitrógeno, los inventores estudiaron el efecto de crecer plantas en condiciones de estrés crónico de nitrógeno así como los cambios en el nivel de nitrógeno disponible. Los inventores habían determinado previamente el impacto sobre el crecimiento de los diferentes niveles de nitrógeno en *Arabidopsis*.

[0269] Se estudió el efecto de los diferentes niveles de nitrógeno sobre los perfiles de transcriptos: cuando el nitrógeno no limita el crecimiento. En el caso de *Arabidopsis*, los inventores recogieron brotes de 4 semanas crecidos con los diferentes regímenes de nitrógeno. Se recogieron tres muestras diferentes (triplicados biológicos) para obtener resultados estadísticamente significativos. El perfil de transcripción se realizó usando una matriz del genoma completo de *Arabidopsis* GenChip® (Affymetrix) para estudiar los niveles de transcripción en *Arabidopsis*. Se realizó el análisis bioinformático necesario para estudiar los datos considerables producidos por estos experimentos. Estudiando el efecto de la limitación del nitrógeno sobre los patrones de expresión, los inventores pueden precisar qué rutas están implicadas en su respuesta al estrés de nutrientes.

EJEMPLO 1

Clonación y secuenciación de At5g56860

55

[0270] Las predicciones genéticas derivaron de la base de datos de secuencias y se usaron para diseñar cebadores oligonucleótidos para amplificación por PCR de los clones de ADNc de longitud completa (inclusive de los codones de inicio y parada previstos, para la sobreexpresión de genes transgénicos) o parcial (para silenciamiento de genes transgénicos) a partir de la primera cadena del ADNc del arroz. En algunos ejemplos, estos cebadores de PCR incluían secuencias 5' adicionales para la clonación basada en recombinación del sistema Gateway™ (Invitrogen). La amplificación por PCR se realizó usando los kits de PCR HF Advantage II (Clontech) o EXPAND (Roche) según las instrucciones del fabricante. Los productos de PCR se clonaron en pCR2.1-TOPO o pDONR201 según las instrucciones del fabricante (Invitrogen).

[0271] Los ADN de 4-8 clones independientes se procesaron con Miniprep siguiendo las instrucciones del fabricante. El ADN se sometió a análisis mediante secuenciación usando el kit del terminador BigDye™ según las

instrucciones del fabricante (ABI). La secuenciación se hizo usando los cebadores diseñados para ambas cadenas del gen previsto. Todos los datos de secuenciación se analizaron y ensamblaron usando el paquete de software Phred/Phrap/Consed (Universidad de Washington) a una tasa de error igual o menor de 10^{-4} a nivel de secuencia consenso.

5

[0272] Las secuencias consenso se validaron como intactas y del gen correcto de varias formas. Inicialmente, se usó análisis de restricción para confirmar la presencia de inserciones de ADNc de los tamaños previstos en clones individuales. Para los clones de longitud completa, se comprobó la ausencia de interrupciones (presencia de codones de inicio y parada previstos, codones de parada no internos) en la región codificadora, secuenciando la inserción de ADNc. Para ambos clones de ADNc de longitud completa y parcial, se usó el alineamiento con el gen previsto y el análisis BLAST para confirmar que se había amplificado el gen diana pretendido. Todos los clones de ADNc generados mediante PCR se clonaron dentro de vectores binarios de destino obtenidos por el cliente usando el sistema de clonación a base de recombinación Gateway™ según las instrucciones del fabricante. Alternativamente, los productos de PCR se clonaron usando procedimientos de clonación a base de enzimas de restricción convencionales.

Vectores de expresión y transformación de plantas

[0273] Los vectores binarios de destino para la transformación de plantas constan de un esqueleto binario y una porción de ADN-T. El esqueleto binario contiene las secuencias necesarias para la selección y el crecimiento en *Escherichia coli* DH-5 α (Invitrogen) y *Agrobacterium tumefaciens* LBA4404, incluyendo el gen bacteriano *aaadA* de resistencia a espectinomicina del transposón Tn7 de *E. coli*, orígenes de replicación de *E. coli* (ColE1) y de *A. tumefaciens* (VS1), y el gen *virG* de *A. tumefaciens*. La porción de ADN-T estaba flanqueada por las secuencias límites derecha e izquierda e incluían el marcador seleccionable de plantas Positech™ (Syngenta) y un casete de expresión génica que varía dependiente de la solicitud. El marcador seleccionable de plantas Positech™ en este ejemplo consta de un promotor ACT1 (actina) del arroz que dirige la expresión del gen PMI (fosfomanosa isomerasa), seguido del terminador de transcripción del virus del mosaico de la coliflor y confiere resistencia a la manosa.

[0274] La porción del casete de expresión génica de los vectores binarios de destino varía dependiendo de la solicitud. En general, el casete consta de un promotor diseñado para expresar el gen en determinados tejidos de la planta, seguido por sitios de clonación (en algunos casos interrumpidos por un segmento de ADN espaciador) y finalmente por el terminador trascricional del extremo 3' nos de *A. tumefaciens*. Los promotores usados se diseñan para expresar el gen de interés en tejidos diana específicos (ej. endosperma: RS-4 del arroz, glutelina de trigo, ADPgpp o γ -zeína del maíz o α -tionina de la cebada, ej. de embrión: globulina u oleosina del maíz; ej. aleuronax - amilasa de la cebada; ej. raíz: MSR1 y MSR3 del maíz; ej. tejido verde: PEPC del maíz) o constitutivamente (ej., UBI más intrón del maíz), dependiendo del gen de interés. El sitio de clonación contiene sitios para enzimas de restricción exclusivos (para clonación convencional) y/o un casete de clonación basado en la recombinación con el sistema Gateway™ (Invitrogen), en la orientación sentido o inversa. En los casetes de expresión génica diseñados para la producción de ARN bicatenario (ARNbc) de interferencia, el sitio de clonación está dividido por una región espaciadora (p. ej., primer intrón del gen SH1 del arroz). El espaciador permite la clonación de dos fragmentos génicos, uno en orientación directa y el otro en orientación inversa. La tecnología complementaria (expresión de orientación inversa) es otra tecnología disponible para el silenciamiento de los genes de interés.

[0275] La transformación de las moléculas de ácido nucleico de la presente invención en plantas se realizó usando procedimientos descritos anteriormente en la descripción detallada. Los kits de prueba para la detección de plantas que contienen las moléculas de ácido nucleico de la presente invención pueden producirse usando técnicas generales para la producción de anticuerpos para el uso en inmunoensayos de tipo ELISA. Alternativamente, los kits para el análisis de tipo PCR del ADN de plantas pueden usarse para comprobar la presencia de las moléculas de ácido nucleico de la presente invención.

RESULTADOS

La pérdida de expresión del gen At5g56860 causa la reducción de los niveles de clorofila en la línea con inserción de ADN-T

[0276] Se identificaron plantas homocigotas para la mutación de inserción de ADN-T en el gen At5g56860 y se seleccionó cualquier cambio fenotípico. Las hojas de las plantas mutantes eran de un color verde pálido debido a una reducción del nivel total de clorofila. El gen At5g56860 tenía un único dedo de cinc con 18 restos en el lazo de cinc (C-X2-C-X18-C-X2-C). La secuencia de la proteína de longitud completa contenía 398 aminoácidos. La inserción de ADN-T (SALK_001778) estaba próxima al extremo del segundo exón, causando la delección del dominio GATA. Mediante RT-PCR no se detectó expresión del gen At5g56860 en la planta mutante.

[0277] Para determinar si el cambio de fenotipo de la línea SALK_001778 estaba causado por la única mutación génica, los inventores hicieron la construcción para sobreexpresar el gen At5g56860 y transformaron este gen de nuevo en la línea mutante. El fenotipo en la planta mutante estaba complementado en el transformante en el

que se detectó la expresión del gen At5g56860, lo que demostraba que el gen At5g56860 era responsable del cambio fenotípico de reducción del nivel de clorofila en la línea mutante SALK_001778.

- [0278]** La planta mutante se retrocruzó también con el fenotipo silvestre y se identificaron plantas heterocigotas mediante PCR que tenían tanto el fenotipo silvestre como el alelo de inserción. Se permitió que estas plantas se autopropagaran y las semillas de la progenie mostraron una segregación 3:1, estando presente el fenotipo mutante solo en aquellas plantas homocigotas para la inserción de ADN-T en el gen At5g56860, lo que indica que el fenotipo mutante estaba causado por la única mutación del gen At5g56860 y que era recesivo respecto al alelo silvestre.
- 10 La expresión del gen At5g56860 es específico de tejido y se regula por el estado de nitratos; la misma tendencia que los genes de asimilación de nitratos

- [0279]** Para determinar el patrón de expresión del gen At5g56860, se extrajo el ARN total a partir de yemas, hojas y raíces y se obtuvo el ADNc. Se usó la RT-PCR para determinar el patrón de expresión de este gen y se demostró que el gen At5g56860 se expresaba en yemas y hojas pero no en raíces. Para determinar la expresión de este gen durante el desarrollo, se recogieron plántulas germinadas en medio MS después de 3 y 5 días y se extrajo el ARN total. Se encontró que el gen At5g56860 se expresaba correctamente después de la germinación.

- [0280]** Puesto que los niveles de clorofila a menudo se usan como reflejo del estado del nitrógeno (29, 30 - faltan referencias), la expresión del gen se estudió en diferentes condiciones de nitrógeno para determinar si su nivel de expresión está afectado por la disponibilidad de nitrógeno. El nivel basal de expresión del gen At5g56860 era inferior cuando crecían en un cultivo hidropónico con muy baja concentración de nitrógeno (0,3 mM) en comparación con la concentración alta (3 mM). Para determinar si este gen está regulado por el nitrato, las plantas se transfirieron de unas condiciones de crecimiento con 1 mM a 3 mM de nitrato. Después de 2 h, la concentración aumentada de nitrato aumentaba la expresión del gen At5g56860, reduciéndose el nivel de expresión transcurridas 24 horas, aunque seguía estando por encima del nivel basal de expresión. Se encontró el mismo patrón de expresión para nia1, nia2 y nir, las dos primeras, enzimas en la reducción de nitratos, aunque la magnitud del cambio era mayor para nia1, nia2 y nir que para el gen At5g56860.

- 30 La planta sin At5g56860 acumula menos N total cuando el aporte de N está limitado, aunque parece que At5g56860 no regula directamente la expresión de genes de asimilación de nitrato

- [0281]** Puesto que la expresión de At5g56860 se ve afectada por el estado del nitrógeno, los inventores probaron si las diferentes condiciones de nitrógeno afectaría al crecimiento de la planta mutante. Las plantas mutantes y silvestres se crecieron en condiciones limitantes de N cuando las plantas se crecían en suelo (nitrato 3 mM) y en condiciones ideales de nitrógeno (nitrato 10 mM). Se recogieron los brotes de plantas mutantes y control de 4 semanas y se midieron la biomasa y el N total. No se encontraron cambios significativos en la biomasa. No obstante, el N total en la planta mutante era inferior que en la silvestre en condiciones limitantes de N y similar al de las plantas silvestre cuando había suficiente aporte de N. La expresión de nia1, nia2 y nir se analizó en las plantas mutantes usando la PCR en tiempo real, mostrando que la expresión de nia1, nia2 y nir en la planta mutante no era significativamente diferente de la de la planta silvestre. Si el gen At5g56860 es uno de los genes reguladores que controlan directamente la expresión de nia1, nia2 y nir, el nivel basal de expresión de nia1, nia2 y nir estaría alterado en el mutante con ADN-T que le falta la expresión de At5g56860. Sin embargo, este no fue el caso.

- 45 El gen At5g56860 regula la expresión de los genes implicados en las diferentes categorías funcionales

- [0282]** Se realizó un experimento de perfil transcripcional para comparar la expresión génica basal a una escala de genoma completo en las plantas mutante y silvestre para ver qué genes tienen la expresión alterada en la planta mutante. De nuevo, las plantas silvestres y las plantas mutantes SALK_001778 se crecieron en condiciones limitantes de N (nitrato 3 mM) o con suficiente N (nitrato 10 mM) y se recogieron los brotes de 4 semanas. Se extrajo el ARN y se hibridó en una matriz de genoma completo GeneChip[®] de *Arabidopsis*. Se analizaron los genes que tenían una expresión al menos 1,5 veces inferior en las plantas mutantes que en las silvestres crecidas en condiciones limitantes de N (nitrato 3 mM) y los genes que tenían una expresión al menos 1,5 veces inferior en las plantas mutantes que en las silvestres cuando se crecían con concentración de N suficiente (10 mM). Se usó un programa de clasificación funcional basado en un navegador (Provart, N. y Zhu, T. (2003) Currents in Computational Molecular Biology 271-272) para demostrar que entre los genes con expresión reducida estaban aquellos implicados en el metabolismo del nitrógeno y el azufre (At3g44300 y At4g27450), en la regulación de los compuestos de C y utilización de hidratos de carbono (At2g18700 y At1g70290), en la biosíntesis de lípidos, metabolismo secundario y glicósidos (At4g37150, At1g19500 y At3g19260), en la percepción de nutrientes y adaptación nutricional (At5g24160), en la captación y absorción de nutrientes (At1g13220 y At2g37770) y en el transporte de electrones y conservación de la energía asociada a membrana (At5g24160 y At5g20230). Los inventores verificaron algunos de esos genes mediante RT-PCR cuantitativa y la correlación entre los datos de la PCR y de micromatrices fue muy alta

La planta mutante At5g56860 es más sensible a la glucosa exógena

- [0283]** Mientras que At5g56860 parece regular la expresión de genes importantes para diferentes procesos

biológicos, muchos de estos genes están implicados en el metabolismo del C, como los transportadores de azúcar y hexosa. Para analizar si el crecimiento de la planta mutante sería diferente al de la planta silvestre con una fuente de C diferente, estas plantas se germinaron en diferente concentración de glucosa y sacarosa. El crecimiento promedio de las plántulas mutantes estaba atrofiado en comparación con las plantas silvestres en un medio con 6% de glucosa. Los fenotipos de las plántulas en placas con el 6% de glucosa son bioensayos bien documentados (Jang, J., Leon, P, Zhou, L. y Sheen, J. (1997) *Plant Cell* 9, 5-19) y los resultados mostraron que el mutante sin expresión de At5g56860 es hipersensible a la glucosa al 6%.

El gen At5g56860 parece regular la expresión de HXK1 pero no la de HXK2 y también regula el gen del transportador de hexosas

[0284] El gen At5g56860 afecta a la sensibilidad a la glucosa por lo que se esperaría que afectara a la expresión de genes implicados en el metabolismo de la glucosa. Se ha demostrado que existen dos tipos de genes que afectan a la detección de la glucosa. El primero es la hexoquinasa (HXK), puesto que la sobreexpresión de HXK en este gen conduce a un fenotipo muy similar al de la mutación de inserción en el gen del factor de transcripción GATA At5g56860. El segundo es el gen del transportador de hexosas, el cual se ha demostrado que afecta a la detección del azúcar en levaduras. Existen dos genes HXK de *Arabidopsis* que se identificaron mediante complementación genética de un doble mutante *hvk1hvk2* de levadura (Jang, J., Leon, P, Zhou, L. y Sheen, J. (1997) *Plant Cell* 9, 5-19). Los resultados mostraron que el nivel de expresión basal de HXK1 era superior en la planta mutante que en la silvestre, pero el nivel de HXK2 no estaba alterado en absoluto en la planta mutante, lo que sugiere que At5g56860 es un regulador negativo de HXK1 pero no de HXK2.

[0285] Se investigó la secuencia de 500 pb antes del extremo 5' de HXK1 y HXK2 y la secuencia consenso de unión a GATA, (T/A)GATA(G/A), o (T/C)TATC(T/A) en la cepa complementaria, se presentaba varias veces en la secuencia promotora de HXK1, mientras que estaba ausente en la secuencia promotora de HXK2 (secuencia no mostrada). El gen del transportador de hexosas estaba regulado por disminución en la línea mutante y también tenía presente el motivo GATA en su región promotora. Aunque la presencia de un motivo GATA no es definitivo en la demostración de la regulación de HXK1 y el transportador de hexosa por un factor GATA, su presencia al menos respalda la noción de que podría estar regulando directamente estos genes.

Las plantas transgénicas de ganancia de función son hiposensibles al azúcar

[0286] Se generaron plantas de *Arabidopsis* transgénicas que sobreexpresaban el gen At5g56860 mediante transformación mediada por *Agrobacterium* (Bechtold, N., Ellis, J. y Pelletier, G. (1993) *C R Acad Sci* 316, 1194-1199). Las plantas transgénicas se seleccionaron en medio con kanamicina y las líneas T2 mostraron una relación de segregación 3:1 en kanamicina, lo que indica una única inserción del transgén. Estas se seleccionaron para la autopolinización. Las líneas transgénicas de la generación T3 homocigota para el transgén se usaron para análisis posteriores. Los niveles de expresión de At5g56860 en las líneas transgénicas se determinaron mediante RT-PCR en tiempo real (datos no mostrados). Las semillas procedentes de líneas de sobreexpresión se germinaron en placas con 6% de glucosa, observándose un escaso efecto inhibitorio sobre el crecimiento en las plántulas transgénicas con sobreexpresión, mientras que las plantas silvestres mostraban una considerable inhibición, lo que indica que son mucho menos sensibles a la glucosa exógena. Es interesante destacar que la expresión del gen HXK1 no está reducida en estas líneas transgénicas (datos no mostrados), mientras que la expresión del gen del transportador de hexosas está aumentada en estas plantas.

DISCUSIÓN

El gen At5g56860 es necesario para la detección del azúcar

[0287] En el mutante con ADN-T sin expresión del gen At5g56860, los mutantes eran hipersensibles al azúcar, mientras que las plantas transgénicas que sobreexpresan el gen At5g56860 muestran hiposensibilidad al azúcar, lo que demuestra que el gen At5g56860 está implicado en la regulación de la detección del azúcar. Se ha descrito hipersensibilidad a la glucosa al 6% en plantas *Arabidopsis* transgénicas que sobreexpresan AtHXK1 y AtHXK2 (Jang, J., Leon, P, Zhou, L. y Sheen, J. (1997) *Plant Cell* 9, 5-19) e hiposensibilidad a la glucosa al 6% en plantas *Arabidopsis* transgénicas que expresan AtHXK1 y AtHXK2 complementarios (Jang, J., Leon, P, Zhou, L. y Sheen, J. (1997) *Plant Cell* 9, 5-19). Se ha descrito que las plántulas transgénicas que sobreexpresan AtHXK1 tienen menor contenido de clorofila en las hojas (Jang, J., Leon, P, Zhou, L. y Sheen, J. (1997) *Plant Cell* 9, 5-19; Dai, N., Schaffer, A., Petreikov, M., Shahak, Y., Giller, Y., Ratner, K., Levine, A. y Granot, D. (1999) *Plant Cell* 11, 1253-1266). En la planta mutante con ADN-T sin expresión de At5g56860, la expresión basal de AtHXK1 es superior que en la planta silvestre control y la planta es hipersensible al azúcar. Las plantas mutantes muestran un nivel reducido de clorofila. No se encontraron elementos GATA en la secuencia de 500 pb a la izquierda del extremo 5' del gen estructural de HXK2. Existe un elemento GATA en la región -737 del promotor de HXK2. Sin embargo, puesto que es necesaria una repetición en tándem o dos elementos GATA localizados en 30 pb o la secuencia central (T/A)GATA(G/A) para una unión eficaz (Chiang, T.Y. y Marzluf, G.A. (1994) *Biochemistry* 33, 576-582; Lin, Y., Hwang, C.F., Brown, J.B. y Cheng, C.L. (1994) *Plant Physiol* 106, 477-484), es poco probable que HXK2 esté directamente regulado por factores GATA. Por el contrario, la secuencia de 500 pb a la izquierda del extremo 5' del

gen de HXK1 tiene varios elementos GATA. El nivel de expresión de HXK1 está alterado en la línea mutante At5g56860, pero no el de HXK2, lo que apoya la idea de que este factor GATA regula la expresión de HXK1.

- 5 **[0288]** Las líneas que sobreexpresan At5g56860 son resistentes a la presencia de altos niveles de glucosa. La sobreexpresión de este factor GATA no provoca un cambio en la expresión de HXK1, pero tampoco provoca un aumento en la expresión del transgén del transportador de hexosas. Se ha encontrado que este gen está implicado en la detección de azúcares en levaduras.

La expresión de At5g56860 responde a un cambio en el estado del N

10

- [0289]** La expresión del gen At5g56860 está aparentemente influenciada por el estado del N de la planta. Cuando el nivel de nitrato es bajo, la expresión basal del gen At5g56860 es relativamente baja y cuando el nivel de nitrato es alto, el nivel basal de expresión del gen At5g56860 es relativamente alto. Cuando las plantas se cambian de condiciones con baja concentración de nitrato a condiciones con alta concentración de nitrato, la expresión del gen At5g56860 está regulada por incremento. Este patrón de expresión es el mismo que el de enzimas clave en las rutas de asimilación de nitrato, como NR y NiR, aunque estas últimas tienen tasas más altas de inducción de la expresión cuando se exponen a niveles altos de nitrato. No obstante, el gen At5g56860 no controla directamente la expresión de *nia1*, *nia2* y *nir*, ya que estos genes no están afectados en la línea mutante. La expresión del gen At5g56860 tiene efecto sobre el metabolismo del nitrógeno, ya que la pérdida de este gen da lugar a la reducción de la acumulación total de N. Es interesante destacar que el gen At5g56860 forma una relación paráloga con el gen At4g26150 a partir del árbol filogenético de los 30 genes del factor de transcripción GATA de *Arabidopsis* (Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K. y Yu, G. (2000) *Science* 290, 2105-2110; Reyes, J.C., Muro-Pastor, M.I. y Florencio, F.J. (2004) *Plant Physiol.* 134, 1718-1732). Ya se ha demostrado que At4g26150 es un gen inducible por nitrato (Wang, R., Okamoto, M., Xing, X. y Crawford, N. M. (2003) *Plant Physiol* 132, 556-567; Scheible, W., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, D., Schindelasch, D., Thimm, O., Udvardi, M. K. y Stitt, M. (2004) *Plant Physiol* 136, 2483-2499). Sin embargo, su función no parece ser redundante, puesto que las firmas de expresión no son completamente similares para los dos genes (Czechowski, T., Bari, R., Stitt, M., Scheible, W. y Udvardi, M. (2004) *Plant J.* 38, 366-379). En la planta mutante SALK_001778 en la que se ha perdido la expresión del gen At5g56860, la expresión del gen At4g26150 está aumentada pero esto no complementa el fenotipo mutante.
- 15
20
25
30

Intercomunicación entre la regulación de C y N

- 35 **[0290]** Es un reto entender cómo las rutas de señalización del C y el N interactúan entre sí, por no mencionar su interacción con otras rutas de señalización como las relacionadas con las hormonas, la luz y el estrés. Hasta donde llega el conocimiento de los presentes inventores, este es el primer factor de transactivación que se describe que está implicado en la regulación de la detección de azúcar y es importante destacar que la expresión del propio factor está influenciada por el cambio en el estado del N.

40

EJEMPLO 2

- [0291]** El ADNc de At5g56860 de longitud completa (GNC) se amplificó a partir del ADNc de hojas de *Arabidopsis* usando los cebadores 5'-GCTCTAGATTCTCTCTCTTTGTGTCTTCATTG-3' (SEC ID N.º 4) y 5'-gcgagctctcgggtgactaatgttcgtcc-3' (SEC ID N.º 5). El fragmento de ~1.500 pb resultante se verificó mediante secuenciación y luego se digirió con XbaI y SacI y se clonó en el vector de expresión digerido con XbaI-SacI, pROK2, que contenía el promotor 35S del virus del mosaico de la coliflor (CaMV) que dirige la expresión de GNC de forma constitutiva y a un alto nivel (Fig. 4). El vector binario p35S-GNC se transformó en la cepa EHA105 de *Agrobacterium tumefaciens* y la cepa de *Agrobacterium* resultante se usó para transformar las plantas silvestres (Col-0) y los transformantes se seleccionaron por su resistencia a kanamicina. Las plantas con una única inserción del transgén se seleccionaron para la autopolinización con la intención de generar líneas homocigotas T3 para su posterior análisis. La sobreexpresión de GNC en los transgénicos se confirmó mediante RT-PCR cuantitativa. Las plantas silvestres y las plantas que sobreexpresaban GNC se crecieron en condiciones de aporte suficiente de nitrógeno (nitrato 10 mM) y de aporte limitado de nitrógeno (nitrato 3 mM). No se observaron diferencias obvias en las etapas iniciales del crecimiento entre ambas condiciones de nitrógeno. Transcurridos aproximadamente 28 días, cuando las plantas empezaron a entrar en la etapa reproductiva, se observó senescencia de las hojas en las plantas silvestres crecidas en condiciones limitantes de nitrógeno, pero no en las plantas que sobreexpresaban GNC; esta diferencia en la senescencia se observaba claramente en el periodo posterior a esta etapa. Asimismo, el nivel total de clorofila era aproximadamente un 15% superior en las plantas que sobreexpresaban GNC. El nivel de clorofila es un indicador de la disponibilidad de nitrógeno y de la capacidad fotosintética, indicando que estas líneas de sobreexpresión son más sanas en un régimen con nitrógeno reducido. Ni las plantas silvestres ni las plantas con sobreexpresión de GNC de 4 semanas de edad mostraron estas diferencias en condiciones con suficiente nitrógeno. No se hallaron diferencias significativas en el nivel de clorofila entre las plantas silvestres y las que sobreexpresaban GNC en condiciones de nitrógeno suficiente.
- 45
50
55
60

65

- [0292]** La invención se establecerá ahora adicionalmente en las reivindicaciones siguientes.

LISTADO DE SECUENCIAS

[0293]

- 5 <110> UNIVERSIDAD DE GUELPH
ROTHSTEIN, Steven
BI, Yong-Mei
- 10 <120> GEN Y PROTEÍNA PARA LA DETECCIÓN DE AZÚCARES REGULADOS POR NITRÓGENO Y
MODULACIÓN DE LOS MISMOS
<130> 6580-339
<140>
<141>
<150> US 60/643.575
- 15 <151> 14/01/2005
<160> 5
<170> PatentIn versión 3.3
- <210> 1
- 20 <211> 1.197
<212> ADN
<213> *Arabidopsis thaliana*
<400> 1

ES 2 429 430 T3

atggattcaa attttcatta ctcgatagat cttaacgaag atcaaaacca tcacgaacaa	60
ccctttttct atcctcttgg atcctcttcc togcttcatc atcatcatca tcatcatcat	120
catcaagtcc ettetaattc ttcatcttct tottctgcca tttcatcget ctctctttac	180
ctccctttct tgatcaactc tcaagaagat caacatggtg cctacaacaa cacttatcac	240
gctgatcatc tccatcttcc tcaaccctc aaggccaaga tgtttggtgg taacgggtgga	300
tcatcagcat gcgatcacat ggtgccaaag aaggagacaa gactgaaact aacgataagg	360
aaaaaagatc acgaagacca accccatcct ctccatcaaa acccgacaaa acccgattca	420
gactccgaca agtggttgat gtccccaag atgcgggtga tcaagaaaac aatcaccaac	480
aataaacagc tcattgatca gactaataat aataatcata aagaaagtga tcaactacct	540
ttgaatcata agactaattt cgacgaggat caccatgaag atottaattt caagaacgtc	600
ttgaccagga agaccacggc cgcgaccacc gagaatogct acaatacaat caacgagaac	660
ggttatagta ataacaatgg cgtgattagg gtttgttcgg attgtaacac caccaagact	720
cctctttggc gaagtggacc tcgaggtccc aagtctcttt gtaacgcatg tggtatacgg	780
caaagaaagg caaggcgagc cgctatggcc gcgcccgctg cagccggcga ccaagaggtg	840
gcggtagcgc cccgagtgca acaattaccg ctgaaaaaga agttgcaaaa taaaaaaaaag	900
agatcaaacg gaggggaaaa atacaatcac tctcctcaa tgggtggccaa ggccaaaaag	960
tgcaagatca aagaggaaga ggagaaggaa atggaagcgg aaacggttgc cggagattca	1020
gagatcagca aatctacaac ttcttetaat tcttcgattt cgtcaaacaa attttgcttc	1080
gatgatttga caataatggt gagcaaaagc tcagcttatc aacaagtgtt cccacaagat	1140
gagaaggagg ctgctgtttt gctcatggct ctgtcgtatg gaatggttca cggttga	1197

<210> 2

<211> 398

5 <212> PROT

<213> *Arabidopsis thaliana*

<400> 2

ES 2 429 430 T3

Met Asp Ser Asn Phe His Tyr Ser Ile Asp Leu Asn Glu Asp Gln Asn
 1 5 10 15

His His Glu Gln Pro Phe Phe Tyr Pro Leu Gly Ser Ser Ser Ser Leu
 20 25 30

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

Ser Ser Ser Ser Ser Ile Ser Ser Leu Ser Ser Tyr Leu Pro Phe Leu
 50 55 60

Ile Asn Ser Gln Glu Asp Gln His Val Ala Tyr Asn Asn Thr Tyr His
 65 70 75 80

Ala Asp His Leu His Leu Ser Gln Pro Leu Lys Ala Lys Met Phe Val
 85 90 95

Ala Asn Gly Gly Ser Ser Ala Cys Asp His Met Val Pro Lys Lys Glu
 100 105 110

Thr Arg Leu Lys Leu Thr Ile Arg Lys Lys Asp His Glu Asp Gln Pro
 115 120 125

His Pro Leu His Gln Asn Pro Thr Lys Pro Asp Ser Asp Ser Asp Lys
 130 135 140

Trp Leu Met Ser Pro Lys Met Arg Leu Ile Lys Lys Thr Ile Thr Asn
 145 150 155 160

Asn Lys Gln Leu Ile Asp Gln Thr Asn Asn Asn Asn His Lys Glu Ser
 165 170 175

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

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

<210> 3
 <211> 1.044
 5 <212> ADN
 <213> *Oryza sativa*
 <400> 3

atgtctacca tctacatgag tcagctctca gctgctctcc ctctcatgga gggggagcac

60

caccatcacc accaggatca tcaccaaggc cacttccaag ccttctcoct gcagcctaag 120
 gatccccag tottattccc ctttgtgatc agtagaagaa gcagcagcag cagccctagc 180
 gacagcacca ctctaagcta tggttcagac catcacttga cacagcagca gcagcatcag 240
 catcaagcca tgcttgagcc ccaaaatatg attggaggat catccgctgg catctttgcg 300
 acgccgttcc cgaccgtcaa gagcatccgc gacgacatga tgcagcggtc gcagttcgat 360
 ccatacgata ccgagaagct gcaggcgagc tgcgggtag ccaaggctcgt cgcggcgggc 420
 aagtggagcg cggtgccagc ggccaagatg aagatcaoga ggaagatggg tgagccgctc 480
 tccgggtgtca ctggcggggc tgcgacgacg gtggcgccga agaagccgag gaggaggccg 540
 gcgcaggcgt acgaggatca cggccatggc ggcgccatgg gccaaagcttt tggcgtgatt 600
 aggggtgtgct ccgactgoaa caccaccaag actcccttgt ggaggagtgg cccgtgcggc 660
 cccaagtgcg tttgcaacgc gtgcggcatc aggcagagga aggcgcggcg ggcgatgatg 720
 gcctccggac taccagcgtc ccccaacgcc gccggcccca aggcggccgc acatagcggc 780
 gccacaaacg cagccgccgc agctgccatg gaggagacgg ccgagtccgc caccgtcgcc 840
 ccgcccccg cgccgacgac gagggggtgt actctcgtcg acagcatcgg gctcagctgg 900
 agcaagaccc atgccgccgc caccgcctcc tgcagcttcc ggccgtcacc ggtggctccc 960
 ggcttcgctg cggcgggtgca ggacgagatc actgacgccg ccatgctgct catgacgctg 1020
 tcctgcgggc ttgtccggag ctga 1044

<210> 4
 <211> 36
 5 <212> ADN
 <213> secuencia artificial
 <220>
 <223> cebador
 <400> 4

10 gctctagatt tctctctc tttgtctt catttg 36

<210> 5
 <211> 30
 15 <212> ADN
 <213> secuencia artificial
 <220>
 <223> cebador
 <400> 5

20 gcgagctctc gggtgactaa tgtctgtcc 30

REIVINDICACIONES

1. Un procedimiento de producción de una planta transgénica con una utilización de nitrógeno mejorada o aumentada que comprende:
- 5 a) transformar una célula vegetal con un ácido nucleico que codifica un factor de transcripción GATA funcional, y
- b) usar la célula vegetal en la producción de una planta que exprese dicho ácido nucleico, en la que la expresión del ácido nucleico tiene como resultado que la planta tenga un nivel elevado de expresión del factor de transcripción
- 10 GATA funcional en comparación con una planta que no ha sido transformada con el ácido nucleico;
- en la que el ácido nucleico que codifica el factor de transcripción GATA funcional comprende:
- i) la secuencia de nucleótidos de SEC ID N.º 3; o
- 15 ii) un ácido nucleico capaz de hibridar con una secuencia complementaria a una secuencia de nucleótidos de SEC ID N.º 3 en condiciones de hibridación que comprenden dodecil sulfato sódico (SDS) al 7%, NaPO₄ 0,5 M, EDTA 1 mM a 50°C con lavados en SSC 0,1X, SDS al 0,1% a 65°C.
- 20 2. El procedimiento de la reivindicación 1, en el que la planta es una dicotiledónea, gimnosperma o monocotiledónea.
3. El procedimiento de la reivindicación 2, en el que la monocotiledónea se selecciona a partir del grupo compuesto por maíz, trigo, arroz, cebada, avena, centeno, mijo, sorgo, triticale, secale, carraón, escanda, farro, tef,
- 25 milo, lino, grama, *Tripsacum* sp. y teosinte.
4. El procedimiento de la reivindicación 2, en el que la dicotiledónea se selecciona a partir del grupo formado por soja, tabaco y algodón.
- 30 5. El procedimiento de la reivindicación 1, en el que la molécula de ácido nucleico se expresa en una ubicación y tejido específico de la planta.
6. El procedimiento de la reivindicación 5, en el que la ubicación o tejido específico se selecciona a partir de uno o más de entre semilla, epidermis, raíz, tejido vascular, meristema, cámbium, corteza, médula, hoja y flor.
- 35 7. El procedimiento de la reivindicación 6, en el que la ubicación o tejido es una semilla.
8. El procedimiento de la reivindicación 1, en el que el ácido nucleico está en un casete de expresión que comprende una secuencia promotora unida de forma operativa al ácido nucleico.
- 40 9. El procedimiento de la reivindicación 8, en el que el ácido nucleico se introduce en la célula vegetal usando un procedimiento que se selecciona a partir del grupo compuesto por bombardeo de micropartículas, transformación mediada por *Agrobacterium* y transformación mediada por fibras.
- 45 10. Una planta, célula vegetal o semilla transgénica con utilización del nitrógeno mejorada o aumentada, que comprende un ácido nucleico que codifica un factor de transcripción GATA funcional, en la que el ácido nucleico que codifica un factor de transcripción GATA funcional comprende:
- a) la secuencia de nucleótidos SEC ID N.º 3; o
- 50 b) un ácido nucleico capaz de hibridar con una secuencia complementaria a una secuencia de nucleótidos SEC ID N.º 3 en condiciones de hibridación que comprenden dodecil sulfato sódico (SDS) al 7%, NaPO₄ 0,5 M, EDTA 1 mM a 50°C con lavados en SSC 0,1X, SDS al 0,1% a 65°C,
- 55 y en el que la utilización de nitrógeno mejorada o aumentada es atribuible al ácido nucleico.

CDS de At5g56860 de longitud completa (1.197 pb) SEC ID N.º 1

```

1 ATGGATTCAA ATTTTCATTA CTCGATAGAT CTTAACGAAG ATCAAAACCA
51 TCACGAACAA CCCTTTTCT ATCCTCTTGG ATCCTCTTCC TCGCTTCATC
101 ATCATCATCA TCATCATCAT CATCAAGTCC CTTCTAATC TTCATCTTCT
151 TCTTCGTCCA TTTCATCGCT CTCCTCTTAC CTCCCTTTCT TGATCAACTC
201 TCAAGAAGAT CAACATGTTG CCTACAACAA CACTTATCAC GCTGATCATC
251 FCCATCTTTC TCAACCCCTC AAGGCCAAGA TGTTTGTGGC TAACGGTGGA
301 TCATCAGCAT GCGATCACAT GGTGCCAAAG AAGGAGACAA GACTGAAACT
351 AACGATAAGG AAAAAAGATC ACGAAGACCA ACCCCATCCT CTTCATCAAA
401 ACCCGACAAA ACCCGATTCA GACTCCGACA AGTGGTTGAT GTCCCAAAG
451 ATGCGGTTGA TCAAGAAAAC AATCACCAAC AATAAACAGC TCATTGATCA
501 GACTAATAAT AATAATCATA AAGAAAGTGA TCACTACCCT TTGAATCATA
551 AGACTAATTT CGACGAGGAT CACCATGAAG ATCTTAATTT CAAGAACGTC
601 TTGACCAGGA AGACCACGGC CGCGACCACC GAGAATCGCT ACAATACAAT
651 CAACGAGAAC GGTTATAGTA ATAACAATGG CGTGATTAGG GTTTGTTCCG
701 ATTGTAACAC CACCAAGACT CCTCTTTGGC GAAGTGGACC TCGAGGTCCC
751 AAGTCTCTTT GTAACGCATG TGGTATACGG CAAAGAAAGG CAAGGCGAGC
801 CGCTATGGCC GCGGCCGCTG CAGCCGGCGA CCAAGAGGTG GCGGTAGCGC
851 CCCGAGTGCA ACAATTACCG CTGAAAAAGA AGTTGCAAAA TAAAAAAAAG
901 AGATCAAACG GAGGGGAAAA ATACAATCAC TCTCCTCCAA TGGTGGCCAA
951 GGCCAAAAAG TGCAAGATCA AAGAGGAAGA GGAGAAGGAA ATGGAAGCGG
1001 AAACGGTTGC CGGAGATTCA GAGATCAGCA AATCTACAAC TTCTTCTAAT
1051 TCTTCGATTT CGTCAAACAA ATTTTGCTTC GATGATTTGA CAATAATGTT
1101 GAGCAAAAGC TCAGCTTATC AACAAGTGTT CCCACAAGAT GAGAAGGAGG
1151 CTGCTGTTTT GTCATGGCT CTGTCTATG GAATGGTTCA CGGTTGA

```

FIGURA 1

Secuencia de la proteína de At5g56860 (398 aa); SEC ID N.º 2

```
1 MDSNFHYSID LNEDQNHHEQ PFFYPLGSSS SLHHHHHHHH HQVPSNSSSS
51 SSSISSLSSY LPFLINSQED QHVAYNNTYH ADHLHLSQPL KAKMFVANGG
101 SSACDHMVPK KETRLKLTIR KKDHEQDQPH LHQNPTKPDS DSDKWLMSPK
151 MRLIKKTITN NKQLIDQTNN NNHKESDHYP LNHKTNFEDE HHEDLNFKNV
201 LTRKTTAATT ENRYNTINEN GYSNNNGVIR VCSDCNTTKT PLWRSQPRGP
251 KSLCNACGIR QRKARRAAMA AAAAAGDQEV AVAPRVQQLP LKKKLQNKKK
301 RSNGGEKYNH SPPMVAKAKK CKIKEEEEEKE MEAETVAGDS EISKSTTSSN
351 SSISSNKFCF DDLTIMLSKS SAYQQVFPQD EKEAAVLLMA LSYGMVHG
```

FIGURA 2

Ortólogo del arroz SEC ID N.º 3

>OsGATA16_205599

ATGTCCTACCÁTCTACATGAGTCAGCTCTCAGCTGCTCTCCCTCTCATGGAGGGGGAG
CACCACCATCACCACCAGGATCATCACCAAGGCCACTTCCAAGCCTTCTCCCTGCAG
CCTAAGGATCCCCAGTCTTATTCCCCTTTGTGATCAGTAGAAGAAGCAGCAGCAGC
AGCCCTAGCGACAGCACCCTCTAAGCTATGGTTCAGACCATCACTTGACACAGCAG
CAGCAGCATCAGCATCAAGCCATGCTTGAGCCCCAAAATATGATTGGAGGATCATCC
GCTGGCATCTTTGCGACGCCGTTCCCGACCGTCAAGAGCATCCGCGACGACATGATC
GAGCGGTCGCAGTTCGATCCATACGATAACCGAGAAGCTGCAGGCGAGCTGCGGGTTA
GCCAAGGTCGTCGCCGGCGGCAAGTGGAGCGCGGTGCCAGCGGCCAAGATGAAGATC
ACGAGGAAGATGGGTGAGCCGTCGTCGGTGTCACTGGCGGGGCTGCGACGACGGTG
GCGCCGAAGAAGCCGAGGAGGAGGCCGGCGCAGGCGTACGAGGATCACGGCCATGGC
GGCGCCATGGGCCAAGCTTTTGGCGTGATTAGGGTGTGCTCCGACTGCAACACCACC
AAGACTCCCTTGTGGAGGAGTGGCCCGTGCGGCCCCAAGTCGCTTTGCAACGCGTGC
GGCATCAGGCAGAGGAAGGCGCGGGCGGCGATGATGGCCTCCGGACTACCAGCGTCC
CCCAACGCCCGCGGCCCAAGGCGGCCGCACATAGCGGCGCCACAAACGCAGCCGCC
GCAGCTGCCATGGAGGAGACGGCCGAGTCCGCCACCGTCGCCCCGCCCCGGCGCCG
ACGACGAGGGGTGGTACTCTCGTCGACAGCATCGGGCTCAGCTGGAGCAAGACCCAT
GCCGCCGCCACCGCCTCCTGCAGCTTCCGGCCGTCACCGGTGGCTCCCGGCTTCGCG
GCGGCGGTGCAGGACGAGATCACTGACGCCGCCATGCTGCTCATGACGCTGTCCTGC
GGGCTTGTCCGGAGCTGA

FIGURA 3

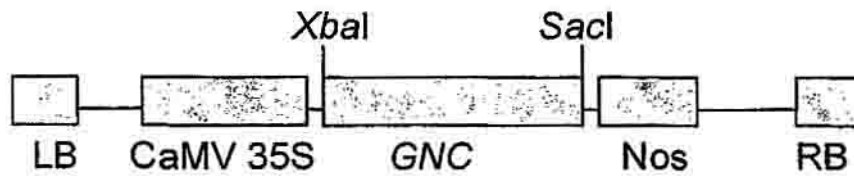


FIGURA 4