

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 402 309**

51 Int. Cl.:

C07K 14/00 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.05.2007** **E 07786718 (2)**

97 Fecha y número de publicación de la concesión europea: **09.01.2013** **EP 2029619**

54 Título: **Manipulación del metabolismo de nitrógeno usando transportador de amonio o glucosa 6-fosfato deshidrogenasas o farnesil fosfato sintetasas (FPP)**

30 Prioridad:

31.05.2006 EP 06114744

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

30.04.2013

73 Titular/es:

**METANOMICS GMBH (100.0%)
TEGELER WEG 33
10589 BERLIN, DE**

72 Inventor/es:

**PLESCH, GUNNAR;
PUZIO, PIOTR;
BLAU, ASTRID;
LOOSER, RALF;
WENDEL, BIRGIT;
KAMLAGE, BEATE y
SCHMITZ, OLIVER**

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 402 309 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

Manipulación del metabolismo de nitrógeno usando transportador de amonio o glucosa 6-fosfato deshidrogenasas o farnesil fosfato sintetetasas (FPP).

5 La presente invención se relaciona con la manipulación del metabolismo del nitrógeno en organismos fotosintéticos activos, preferiblemente en plantas. En particular, la presente invención se relaciona con un proceso para la asimilación, acumulación y/o utilización mejorada de nitrógeno y/o para el contenido incrementado de nitrógeno total y la eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento de biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno en un organismo fotosintético activo.

10 La asimilación en la nutrición de las plantas es esencial para el crecimiento y desarrollo de las plantas y por lo tanto también para la cantidad y calidad de los productos vegetales. Debido a la fuerte influencia de la eficiencia en la utilización de la nutrición sobre el rendimiento de las plantas y la calidad del producto, se vierte una gran cantidad de fertilizante en los campos para optimizar el crecimiento y calidad de las plantas. La productividad de las plantas ordinariamente está limitada por tres nutrientes primarios, fósforo, potasio y nitrógeno, el cual es usualmente el elemento limitante en proporción en el crecimiento de las plantas de estos tres. Por lo tanto, el elemento nutricional principal requerido para el crecimiento de las plantas es el nitrógeno (N). Es un constituyente de numerosos compuestos importantes encontrados en las células vivas, incluyendo los aminoácidos, proteínas (enzimas), ácidos nucleicos y clorofila. Del 1.5% al 2% del material seco de la planta es nitrógeno y aproximadamente el 16% de la proteína total de las plantas. Así, la disponibilidad del nitrógeno es un factor limitante principal para el crecimiento y producción de los cultivos vegetales (Frink et al., 1999), y tiene también un impacto principal en la acumulación de proteínas y en la composición de los aminoácidos.

15 Las plantas pueden utilizar un amplio rango de especies de nitrógeno incluyendo amoníaco volátil (NH₃), óxidos de nitrógeno (NO_x), nitrógeno mineral, como nitratos (NO₃⁻) y sales de amonio (NH₄⁺), urea y derivados de urea, y nitrógeno orgánico (aminoácidos, péptidos, etc.). Algunas plantas son capaces de utilizar el nitrógeno atmosférico mediante bacterias simbióticas o ciertos hongos. Sin embargo, en la mayor parte de los suelos agrícolas, el nitrato (NO₃⁻) es la fuente más importante de nitrógeno (Crawford and Glass, 1998; Hirsch and Sussman, 1999). No obstante el amonio NH₄⁺ juega también un papel importante probablemente subestimado, porque la mayor parte de las plantas preferencialmente toman NH₄⁺ cuando ambas formas están presentes - incluso si NH₄⁺ está presente a concentraciones más bajas que NO₃⁻ (Von Wieren et al., 2000).

20 Debido a los altos requerimientos de nitrógeno para las plantas de cultivo, la fertilización con nitrógeno es una inversión agrícola mundial principal, con 80 millones de toneladas métricas de fertilizantes de nitrógeno aplicadas anualmente (como nitrato y/o amonio) (Frink et al., 1999). Hay también consecuencias ambientales negativas por el uso extensivo de fertilizantes que contienen nitrógeno en la producción de cultivos porque los cultivos agrícolas solamente retienen aproximadamente dos tercios del nitrógeno aplicado. Por lo tanto aplicaciones altas de fertilizante son seguidas por escapes altos por filtración, pérdidas gaseosas y eliminación de cultivos. El nitrógeno no absorbido puede fugarse subsecuentemente hacia el suelo y contaminar los suministros de agua (Frink et al., 1999). Debido a las altas pérdidas por filtración del nitrógeno desde los ecosistemas agrícolas a las aguas subterráneas y aguas superficiales, el nitrógeno es reconocido actualmente como un contaminante importante. La fuga de nitrógeno, a saber cómo nitrato desde las tierras agrícolas, afecta la calidad del agua para bebida y produce eutroficación de lagos y áreas costeras. El uso abundante de fertilizantes que contienen nitrógeno puede llevar adicionalmente al deterioro final de la calidad del suelo, a la polución ambiental y a riesgos para la salud.

25 Debido a los altos costes de los fertilizantes de nitrógeno para la producción agrícola, y adicionalmente por su efecto nocivo sobre el ambiente, es deseable desarrollar estrategias para reducir la aplicación de nitrógeno y/o para optimizar la asimilación, acumulación y/o utilización del nitrógeno para una disponibilidad dada de nitrógeno mientras que simultáneamente se mantenga una productividad óptima y la calidad de los organismos fotosintéticos activos, preferiblemente plantas cultivadas, por ejemplo cultivos. Preferiblemente, las plantas cultivadas utilizadas como alimento y/o forraje deberían tener una calidad mejorada, especialmente en términos de acumulación y composición de sus proteínas.

30 Para una asimilación, acumulación y utilización eficiente del consumo de nitrógeno, se requieren procesos complejos asociados con la absorción, traslocalización, asimilación y redistribución del nitrógeno para operar con eficiencia. Las diferencias en la absorción y utilización del nitrógeno entre genotipos han sido demostradas para varias especies por diferentes investigadores (Chang & Robinson, 2001). Se ha reportado evidencia considerable sobre las diferencias genotípicas en el consumo de nitrógeno, por ejemplo en su acumulación para el maíz y la canola (Weisler et al., 2001; Gallais & Hirel, 2004).

35 El consumo de nitrógeno en las plantas está altamente regulado y coordinado con otras rutas de transporte y metabólicas (Crawford, 1995), y un cierto número de genes relacionados con el consumo y asimilación de nitratos han

sido identificados y caracterizados (Forde, 2002). Las plantas absorben el nitrato a través de transportadores localizados en la membrana de plasma celular epidérmica y cortical de las raíces a lo largo de un amplio rango de concentración de nitratos utilizando diferentes mecanismos de transporte, incluyendo sistemas de transporte de alta afinidad constitutivos e inducibles por nitrato, así como transportadores de baja afinidad inducibles por nitrato (Stitt, 1999). Una vez en el citoplasma de la célula de la raíz, el nitrato puede ser almacenado en las vacuolas para uso posterior, transportado hacia el xilema y traslocalizado hacia el brote para asimilación y/o almacenamiento, liberado de regreso a la rizosfera o reducido a nitrito y luego a amoníaco a través del nitrato reductasa (NR) y las nitrito reductasa (NiR). La reducción del nitrato a nitrito y luego a amoníaco genera nitrógeno en una forma que puede ser asimilada en aminoácidos a través de la ruta GOGAT (Stitt, 1999). Con el fin de ser incorporado en los aminoácidos, ácidos nucleicos, y otros compuestos, el NO_3^- debe ser reducido a NH_4^+ . La NR (nitrato reductasa) es la primera enzima en el proceso de reducción del NO_3^- a la forma amino. Es una enzima inducible por el sustrato y se cree que es la etapa más limitante en la asimilación del nitrógeno.

La rata *in situ* de la reducción del NO_3^- es controlada primariamente por la rata de consumo de NO_3^- , más que por cualquier alteración en la actividad de la nitrato reductasa (NRA) o limitaciones en el poder reductor. Así, el consumo de NO_3^- parece ser de importancia primaria en la asimilación de nitrógeno en plantas alimentadas con NO_3^- . La variación genética en NRA está bien documentada en varias especies. La NRA está afectada por factores tales como condiciones ambientales y etapa de desarrollo de la planta, así como la parte de la planta, tal como raíces y parte superior. Adicionalmente, pruebas *in vivo* e *in vitro* usualmente dan resultados diferentes. Resultados variables fueron encontrados por varios investigadores en sus esfuerzos para relacionar la NRA al rendimiento de granos y de efectos relacionados con N tales como N total reducido en las plantas, contenido de nitrógeno en granos, concentración de nitrógeno en los granos e índice de recolección de nitrógeno.

Con el fin de describir la eficiencia de la ruta completa del nitrógeno, comenzando con la toma desde el suelo, asimilación, acumulación y finalmente utilización del nitrógeno para el crecimiento hasta la madurez y para la maduración de frutos y semillas, se conocen diferentes metodologías. A la luz de la importancia de la adquisición y utilización óptima del nitrógeno, se han seguido diferentes estrategias para las optimizaciones en las plantas.

La Patente de los Estados Unidos 6,727,411 divulga un método para producir tomates transgénicos que tienen un contenido de aminoácidos libres incrementado en los fruto de tomate transformado un tomate con un constructo genético que contiene la secuencia antisentido de un gen que codifica la glutamato descarboxilasa.

En algunos casos las enzimas de la ruta de asimilación de nitrógeno fueron sobreexpresadas. Aunque inicialmente no hubo éxito como en la sobreexpresión del gen citosólico de la glutamina sintetasa en Lotus (Vincent et al., Planta. 201 (4):424-33, 1997), recientes documentos muestran al menos algún éxito. La WO95/09911 describe la sobreexpresión de la glutamina-sintetasa, asparagina-sintetasa y asparaginasa en plantas transgénicas para la aplicación en la fijación potenciada de nitrógeno y en rendimiento mejorado. Chichkova et al., J. Exp. Bot.;(2001) reportó que las plantas transgénicas de tabaco que sobreexpresan la NADH-glutamato-sintasa de alfalfa tenían un contenido mayor de carbono y nitrógeno, pero no un enriquecimiento específico de nitrógeno en comparación con el carbono. En otro caso, por ejemplo como el descrito en Long et al., Plant.Physiol.; (1996) 111, 2, Suppl., 48, la sobreexpresión de un gen de asimilación de nitrógeno, en este caso el Escherichia coli glutamato- deshidrogenasa, no lleva a un incremento relativo en el contenido de nitrógeno, sino más bien a un incremento significativo en el peso fresco y en el peso seco. En otro caso, la sobreexpresión del gen ASN1 potencia el estatus del nitrógeno en semillas de Arabidopsis (Lam et al., Plant Physiology, 2003, 321, 926-935. En semillas de estas líneas sobreexpresadas los autores observaron la elevación del contenido de proteínas solubles en semillas, la elevación de la proteína total, el contenido de las semillas hidrolizadas en ácido y una tolerancia más alta de los plantones jóvenes cuando crecen bajo condiciones con limitación de nitrógeno, demostrando que aquellos defectos están estrechamente interrelacionados.

La Patente de los Estados Unidos 6,969,782 divulga plantas que contienen aminoácidos libres acumulados en una gran cantidad por expresión excesiva de glutamato deshidrogenasa (GDH).

La solicitud de Patente de los Estados Unidos 20030115638 provee una planta transformada que tiene un contenido incrementado de aminoácidos libres introduciendo genes de fosfoenolpiruvato carboxilasa (PEPC).

Las plantas con niveles elevados de proteínas para utilización de nitrógeno en la raíz de estas plantas se divulgan en US 20050044585 por expresión de un gen de alanina aminotransferasa.

Una metodología diferente interesante fue seguida por Yanagisawa et al., PNAS (2004) 101, 20, 7833-7838. Los autores que identificaron y sobreexpresaron un factor regulador, el cual indujo la superregulación de genes que codifican enzimas para la producción de esqueletos de carbono, un incremento marcado en el contenido de aminoácidos, y una reducción en el nivel de glucosa en Arabidopsis transgénica. Los análisis elementales revelaron que el contenido de nitrógeno se incrementó en las plantas transgénicas (aproximadamente un 30%), indicando una promoción de la asimilación neta de nitrógeno. Más significativamente, las plantas transgénicas Dof 1 exhiben un crecimiento mejorado bajo condiciones de nitrógeno bajo, un desafío agrónomicamente importante. Aunque parece promisorio, esta metodología tiene probablemente la desventaja de que se basa en un factor de transcripción de la

- planta y la cascada de señalización correspondiente compleja que podrían ser ambos el sujeto de un mecanismo de regulación interna y retroalimentación diferente que modifique o incluso disminuya el efecto deseado al menos bajo ciertas condiciones. Además la función de un factor de transmisión en una planta se basa en su interacción con secuencias objetivo en diferentes promotores, haciendo que la transferencia de resultados entre diferentes especies de plantas sea compleja e impredecible.
- 5
- No obstante, hay una necesidad por organismos fotosintéticos activos que sean capaces de asimilar y acumular el nitrógeno más eficientemente. Además, los organismos fotosintéticos activos tienen que ser capaces de una utilización más eficiente de nitrógeno de tal manera que se requiera menos nitrógeno para el mismo rendimiento o que puedan obtenerse más rendimientos con los niveles actuales de uso del nitrógeno.
- 10
- Hay adicionalmente una necesidad por organismos fotosintéticos activos que muestren un rendimiento de biomasa incrementado, preferiblemente con una tasa de crecimiento más rápida, lo cual puede llevar a un rendimiento mayor en frutas o semillas.
- Los nuevos organismos fotosintéticos activos presentarán un contenido de proteínas mayor pero definido (con respecto a la proporción de los diferentes aminoácidos) en el fruto o semillas o en el organismo completo.
- 15
- Los nuevos organismos fotosintéticos activos mostrarán al menos una de estas características bajo condiciones de contenido reducido de nitrógeno en el medio circundante, suelo o ambiente.
- En una realización de la presente invención, estas características se logran mediante un proceso de asimilación, acumulación y/o utilización incrementada de nitrógeno en un organismo fotosintético activo que lleva a un contenido incrementado de nitrógeno total y a una eficiencia en el uso del nitrógeno incrementada (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno en el fruto o semilla o en el organismo completo.
- 20
- En una realización de la presente invención, esto se logra mediante una eficiencia incrementada en el uso de nitrógeno (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo los fertilizantes de nitrógeno.
- 25
- Las plantas pueden tomar el nitrógeno también en la forma de amonio. Aunque las concentraciones promedio de NH_4^+ en suelo frecuentemente son de 10 a 1000 veces inferiores a las de NO_3^- (Marschner HL, Mineral Nutrition in Higher Plants. London Academic Press; 1995), la diferencia en las concentraciones en el suelo no necesariamente refleja la relación de consumo de cada fuente de nitrógeno. Las plantas toman NH_4^+ preferencialmente cuando más formas de nitrógeno están disponibles, eventualmente porque su asimilación requiere menos energía porque el NO_3^- no tiene que ser reducido antes de la asimilación (Bloom et al., Plants Phys. 1992, 1294-1301).
- 30
- Se han caracterizado sistemas de consumo de amonio en diferentes organismos, incluyendo levaduras y plantas. La levadura *Saccharomyces cerevisiae* contiene tres genes MEP para transportadores de amonio, los cuales son controlados todos por el nitrógeno, siendo reprimidos en presencia de una fuente de nitrógeno que sea fácilmente metabolizada, tal como NH_4^+ (Marini et al., Mol Cell Biol 1997, 17:4282-4293). Los genes vegetales que codifican sistemas de transportes de amonio han sido clonados por complementación de un mutante levadura, búsquedas de homología en bases de datos e hibridaciones heterólogas (revisadas en van Wieren et al., Current Opinion in Plant Biology, 200, 3:254-261). La evidencia experimental de la función fisiológica de los transportadores de NH_4^+ se basa principalmente en correr acciones entre la expresión de los transportadores de amonio y el flujo de amonio marcado. La situación se complica por el hecho de que en *Arabidopsis* pero también en otras plantas los transportadores de amonio están presentes en familias de genes, cuyos miembros tienen diferentes patrones de expresión y características fisiológicas. Aunque la DE 4337597 reivindica secuencias para transportadores de amonio en plantas y su uso para la manipulación del metabolismo del nitrógeno y el crecimiento de las plantas bajo ciertas condiciones, falta cualquier evidencia con respecto a los efectos positivos de la asimilación de nitrógeno o el crecimiento de las plantas bajo ciertas condiciones a través de la expresión ectópica de los transportadores de amonio en las plantas.
- 35
- 40
- 45
- Por lo tanto la evidencia en la literatura para la manipulación de la asimilación del nitrógeno en plantas está aún limitada a unos pocos casos, sin incluir transportadores.
- La US 2003/0204867 divulga la expresión de transportadores de amonio en *Arabidopsis thaliana* para incrementar el consumo de nitrógeno. Marini et al., (1997, Molecular and cellular biology 17(8): 4282-4293) caracterizan Mep2p/Mep3p en levadura (D2: extracto). Adicionalmente, Ninnemann et al. (1994, The EMBO Journal 13(15):3464-3471) divulgan la identificación de un transportador de NH_4^+ de *A. thaliana*.
- 50
- Marini et al., (1994, EMBO Journal 13(15):3456-3463) divulgan el análisis de transportadores de amonio MEP 1 en levaduras. Arrillaga y colaboradores (1998, Plant Science 136:219-226) describen la expresión de los genes de

“halotolerancia” (HAL2) en tomate para potenciar la tolerancia a la sal (véase D5: resumen/introducción). La WO 2006/069610 divulga un contenido de proteína potenciado por expresión de SEQ ID NO: 689.

5 Es un objeto de la presente invención desarrollar un proceso no costoso para una asimilación, acumulación y/o utilización potenciadas de nitrógeno y/o para el contenido de nitrógeno total incrementado en un organismo fotosintético activo que lleva a un contenido de nitrógeno total incrementado en la fruta o semilla o en el organismo completo y a una eficiencia incrementada en el uso de nitrógeno (NUE).

Se ha encontrado ahora que este objetivo se logra proveyendo el proceso de acuerdo con la invención descrita aquí en las realizaciones caracterizadas en las reivindicaciones.

10 De acuerdo con lo anterior, en una primera realización, la invención se relaciona con un proceso para la asimilación, acumulación y/o utilización potenciada de nitrógeno en un organismo fotosintético activo.

De acuerdo con lo anterior, en otra realización, la invención se relaciona con un proceso para incrementar el contenido total de nitrógeno y la eficiencia incrementada del uso de nitrógeno (NUE) definida como rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno en un organismo fotosintético activo.

15 De acuerdo con lo anterior, en una realización esto se logra por producción (incrementada) de nitrógeno o compuestos que contienen nitrógeno, mediante lo cual el compuesto o los compuestos que contienen nitrógeno son compuestos que contienen nitrógeno (N). En una realización el término “nitrógeno o compuestos que contienen nitrógeno” tal como se utiliza aquí se refiere a “aminoácidos”, preferiblemente fenilalanina, prolina, ácido aspártico, 5-oxoprolina compuestos y/o derivados que contienen “complejo heme”, “purina” y/o “pirimidina”. Adicionalmente, en otra realización el término “nitrógeno o compuestos que contienen nitrógeno” tal como se utiliza aquí se refiere a composiciones de sustancias químicas finas que comprenden compuestos que contienen nitrógeno.

De acuerdo con lo anterior, la presente invención se relaciona con un proceso que comprende

(a) incrementar o generar la actividad de YPR138C por las secuencias de ácido nucleico como se muestra en SEQ ID NO: 689, en un organismo no humano o en una o más partes o compartimientos del mismo y

25 (b) cultivar el organismo bajo condiciones que permitan la producción de nitrógeno o compuestos que contienen nitrógeno, así, compuestos que contienen N y/o asimilación, acumulación y/o utilización incrementada de nitrógeno y/o contenido de nitrógeno total y eficiencia en el uso de nitrógeno (NUE) incrementadas, definida como el rendimiento de biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, en dicho organismo.

De acuerdo con lo anterior, la presente invención se relaciona con un proceso para la producción de un compuesto químico fino que comprende

30 (a) incrementar o generar la actividad de una o más proteínas que tengan la actividad de YPR138C, línea 1 o que tengan la secuencia de un polipéptido codificado por una molécula de ácido nucleico indicada en SEQ ID NO: 689, en un organismo no humano en una o más partes o compartimientos del mismo y

(b) cultivar el organismo bajo condiciones que permitan la producción de nitrógeno o compuestos que contienen nitrógeno, en particular un compuesto que contiene N.

35 Comprende/que comprende y variaciones gramaticales del mismo cuando se utilizan en esta especificación se entienden como la especificación de la presencia de las características establecidas, enteros, etapas o componentes establecidos o grupos de los mismos, pero no evitan la presencia o adición de una o más otras características, enteros, etapas, componentes o grupos de los mismos.

40 El término “Tabla I” utilizado en esta especificación se debe tomar como la especificación del contenido en la Tabla I A y la Tabla I B. El término “Tabla II” usado en esta especificación debe tomarse como especificación del contenido de la Tabla II A y Tabla II B. El término “Tabla I A” que se utiliza en esta especificación debe tomarse para especificar el contenido de la Tabla I A. El término “Tabla I B” utilizado en esta especificación debe tomarse para especificar el contenido de la Tabla I B. El término “Tabla II A” utilizado en esta especificación debe tomarse para especificar el contenido de la Tabla II A. El término “Tabla II B” utilizado en esta especificación debe tomarse para especificar el contenido de la Tabla II B. En una realización preferida, el término “Tabla I” significa Tabla I B. En una realización preferida, el término “Tabla II” significa Tabla II B.

45 Los términos “potenciados” o “incrementados” significan al menos una producción más alta en 10%, 20%, 30%, 40% a 50%, preferiblemente al menos 60%, 70%, 80%, 90% o 100%, más preferiblemente 150%, 200%, 300%, 400% o 500% de nitrógeno o compuestos que contienen nitrógeno en comparación con la referencia como se define más adelante, por ejemplo, que significa en comparación con un organismo sin la modificación antes mencionada de la actividad de una

- 5 proteína que tiene la actividad de YPR138C o codificada por la molécula de ácido nucleico identificada en SEQ ID NO: 689. El término compartimiento se refiere a todos los diferentes compartimientos subcelulares de una célula, incluyendo pero no limitándose a mitocondria, vacuola, el núcleo, todos los tipos de plástidos, tales como aminoplastos, cloroplastos, cromoplastos, el espacio extracelular, la mitocondria, el retículo endoplásmico, elaioplastos, peroxisomas, glicosomas y otros compartimientos.
- 10 Sorprendentemente, se ha encontrado que la expresión transgénica de la proteína YPR138C de *Saccharomyces cerevisiae* confiere un incremento en el contenido de compuestos que contienen N y/o confieren una asimilación, acumulación y/o utilización potenciada de nitrógeno y/o un contenido incrementado de nitrógeno total y una eficiencia incrementada del uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo los fertilizantes de nitrógeno del organismo transformado.
- 15 Sorprendentemente, se ha encontrado que la expresión transgénica de la proteína YPR138C de *Saccharomyces cerevisiae* en *Arabidopsis thaliana* confiere un incremento en el contenido de compuestos que contienen N y/o confieren una asimilación, acumulación y/o utilización potenciada de nitrógeno y/o un contenido incrementado de nitrógeno total y una eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo los fertilizantes de nitrógeno del organismo transformado, cuando se expresa en células anfitrionas, preferiblemente en el citosol, de las células vegetales.
- De acuerdo con la invención, el término "organismo" tal como se entiende aquí se relaciona siempre con un organismo no humano en particular con un organismo fotosintético activo, preferiblemente en un organismo vegetal o un microorganismo.
- 20 La secuencia de YPR138C de la *Saccharomyces cerevisiae* ha sido publicada en Goffeau et al., *Science* 274 (5287), 546-547, 1996 y Bussey et al., *Nature* 387 (6632 Suppl), 103-105 (1997) y su actividad está siendo definida como transportador de NH_4^{++} . De acuerdo con lo anterior, en una realización, el proceso de la presente invención comprende el uso de un producto genético con una actividad de proteína de transporte de amonio; una superfamilia nrgA de transportadores de amonio, preferiblemente una proteína con una actividad de transporte de NH_4^{++} , de *Saccharomyces cerevisiae* o su homólogo, por ejemplo, tal como se muestra aquí, para la producción de nitrógeno o compuestos que contienen nitrógeno, significando esto un compuesto que contienen N, y/o para conferir una asimilación, acumulación y/o utilización potenciada de nitrógeno y/o un contenido incrementado de nitrógeno total y/o una eficiencia incrementada en el uso de nitrógeno (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno, en particular para incrementar la cantidad de un compuesto que contienen N en un organismo o una parte del mismo, tal como se mencionó. De acuerdo con lo anterior, en una realización, el proceso de la presente invención comprende el uso de un producto genético con una actividad de proteína de transporte de amonio; una superfamilia nrgA transportadora de amonio, preferiblemente una proteína con una actividad de transporte de NH_4^{++} , de *Saccharomyces cerevisiae* o su homólogo, por ejemplo, como se muestra aquí, para el consumo y/o asimilación mejorados de nitrógeno.
- 25 De acuerdo con lo anterior, en una realización, el proceso de la presente invención comprende el uso de un producto genético con una actividad de proteína de transporte de amonio; una superfamilia nrgA de transporte de amonio, preferiblemente una proteína con una actividad de transporte de NH_4^{++} , de *Saccharomyces cerevisiae* o su homólogo, por ejemplo, como se muestra aquí, para el consumo y/o utilización y/o asimilación incrementados de nitrógeno bajo condiciones limitadas en nitrógeno.
- 30 Homólogos (= homólogos) de los presentes productos genéticos pueden ser derivados de cualquier organismo en tanto el homólogo confiere la actividad aquí mencionada, en particular, confiere un incremento en la cantidad o contenido de nitrógeno o compuestos que contienen nitrógeno. Adicionalmente, en la presente invención, el término "homólogo" se refiere a la secuencia de un organismo que tiene la homología en secuencia más alta a las secuencias mencionadas o listadas aquí de todas las secuencias expresadas de dicho organismo.
- 35 Sin embargo, la persona experimentada en la técnica sabe que, preferiblemente, el homólogo tiene dicho contenido de nitrógeno y esta eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno - actividad incrementada y, si se sabe, la misma función o actividad biológica en el organismo que la proteína YPR138C, por ejemplo, tiene la secuencia de un polipéptido codificado por la molécula de ácido nucleico que comprende la secuencia indicada en SEQ ID NO: 689.
- 40 En una realización, el polipéptido de la invención o el polipéptido utilizado en el método de la invención confieren dicha actividad, por ejemplo, el incremento de nitrógeno o compuestos que contienen nitrógeno y/o la asimilación, acumulación y/o utilización incrementada de nitrógeno y/o el contenido incrementado de nitrógeno total y la eficiencia incrementada en el uso de nitrógeno (NUC), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno en un organismo o una parte del mismo, si se deriva de un organismo, el cual está evolutivamente distante del organismo en el cual se expresa. Por ejemplo los organismos de origen y expresión son derivados de diferentes familias, ordenes, clases o filums.
- 55

- Los términos “incrementados”, “elevados”, “extendido”, “potenciado”, “mejorado” o “amplificados” se relacionan a un cambio correspondiente de una propiedad en un organismo, una parte de un organismo tal como un tejido, semilla, raíz, hoja, flor, etc. u en una célula y son intercambiables. Preferiblemente, la actividad global en el volumen se incrementa o potencia en casos si el incremento o potenciamiento está relacionado con el incremento o potenciamiento de una actividad de un producto genético, independientemente de si la cantidad de gen producida o la actividad específica del producto genético o ambos son incrementados o potenciados o si la cantidad, estabilidad o eficacia en la traducción de la secuencia de ácido nucleico o del gen que codifica para el producto genético son incrementados o potenciados. Los términos “reducción”, “disminución” o “eliminación” se refieren a un cambio correspondiente de la propiedad en un organismo, una parte de un organismo tal como un tejido, semilla, raíz, hoja, flor, etc. o en una célula.
- 5
- 10 Preferiblemente, la actividad global en el volumen es reducida, disminuida o eliminada en casos si la reducción, disminución o eliminación está relacionada con la reducción, disminución o eliminación de una actividad de un producto genético, independientemente de si la cantidad de producto genético o la actividad específica del producto genético o ambas son reducidas, disminuidas o eliminadas o si la cantidad, estabilidad y eficacia en la traducción de la secuencia de ácidos nucleicos o del gen que codifica el producto genético son producidas, disminuidas o eliminadas.
- 15 Los términos “incrementar” o “disminuir” se relacionan a un cambio correspondiente de una propiedad en un organismo o en una parte de un organismo, tal como un tejido, semilla, raíz, hoja, flor, etc. o en una célula. Preferiblemente la actividad global en el volumen es incrementada en casos en que el incremento se relacionan con el incremento de una actividad de un producto genético, independientemente de si la cantidad del producto genético o la actividad específica del producto genético o ambas se incrementan o generan o si la cantidad, estabilidad y eficacia en la traducción de la secuencia de ácidos nucleicos o del gen que codifica el producto genético se incrementan.
- 20
- Bajo “cambio de una propiedad” se entiende que la actividad, nivel de expresión o cantidad de un producto genético o el contenido en metabolitos o el contenido en elementos cambia en un volumen específico con relación a un volumen correspondiente de un control, referencia o tipo silvestre, incluyendo la creación de novo de la actividad o expresión.
- 25 Los términos “incremento” o “disminuye” incluyen el cambio de modulación de dicha propiedad en solamente partes del sujeto de la presente invención, por ejemplo, la modificación puede ser encontrada en un compartimiento de una célula, tal como un organelo, o en una parte de una planta, tales como tejido, semilla, raíz, hoja, flor, etc. pero no es detectable si se prueba el sujeto global, esto es, la célula o planta completa. Preferiblemente, el incremento o descenso es encontrado a nivel celular, así el término “incremento de una actividad” o “incremento de un metabolito o contenido de un elemento” se relaciona con el incremento celular comparado con la célula tipo silvestre.
- 30 Sin embargo, los términos incremento o disminución tal como se utilizan aquí también incluyen el cambio o modulación de una propiedad en el organismo completo tal como se mencionó.
- 35 De acuerdo con lo anterior, el término “incremento” o “disminución” significa que la actividad específica de una enzima, preferiblemente la cantidad de un compuesto o metabolito, por ejemplo de un polipéptido, una molécula de ácido nucleico o de nitrógeno o de compuestos que contiene nitrógeno de la invención o de un ARNm o ADN que codifican, pueden ser incrementado o disminuido en un volumen.
- 40 Los términos “tipo silvestre”, “control” o “referencia” son intercambiables y pueden ser una célula o parte de organismos tales como un organelo o un tejido, o un organismo, en particular un microorganismo o una planta, que no se ve modificada o tratada de acuerdo con el proceso descrito aquí de acuerdo con la invención. De acuerdo con lo anterior, la célula o una parte de organismos tal como un organelo o un tejido, o un organismo, en particular un microorganismo o una planta utilizados como tipo silvestre, control o referencia corresponden a la célula, organismo o parte de los mismos tanto como sea posible y está en cualquier otra propiedad excepto en el resultado del proceso de la invención idéntica al objeto de la invención tanto como sea posible. Así, el tipo silvestre, control o referencia se trata de manera idéntica o tan idéntica como sea posible, diciendo que solamente las condiciones o propiedades podrían ser diferentes sino influyen la calidad de la propiedad probada.
- 45 Preferiblemente, cualquier comparación se lleva a cabo bajo condiciones análogas. El término “condiciones análogas” significa que todas las condiciones tales como, por ejemplo, condiciones de cultivo o crecimiento, condiciones de ensayo (tales como una composición reguladora, temperatura, sustratos, cepa del patógeno, concentraciones y similares) se mantienen idénticas entre los experimentos que se van a comparar.
- 50 La “referencia”, “control” o “tipo silvestre” es preferiblemente un sujeto, por ejemplo un organelo, una célula, un tejido, un organismo, en particular una planta o un microorganismo, que no ha sido modificado tratada de acuerdo con el proceso aquí descrito de la invención y es en cualquier otra propiedad tan similar como sea posible al objeto de la invención. La referencia, control, o tipo silvestre es

en su genoma, transcriptoma, proteoma o metaboloma tan similar como sea posible al sujeto de la presente invención. Preferiblemente el término organelo, célula, tejido u organismo de “referencia”, “control” o “tipo silvestre”, en particular

5 un planta o un microorganismo se refiere a un organelo, célula, tejido u organismo, en particular planta o microorganismo el cual es cercanamente idéntico desde el punto de vista genético al organelo, célula, tejido u organismo, en particular microorganismo o planta, de la presente invención o una parte de la misma preferiblemente 95%, más preferiblemente son 98%, incluso más preferido son 99.00%, en particular 99.10%, 99.30%, 99.50%, 99.70%, 99.90%, 99.99% 99.999% o más. Lo más preferiblemente la "referencia", "control" o "tipo silvestre" es un sujeto, por ejemplo un organelo, una célula, un tejido, un organismo que es genéticamente idéntico al organismo, célula u organelo usado de acuerdo con el proceso de la invención excepto en que las moléculas de ácido nucleico responsables que confieren la actividad o el producto genético codificado por ellas están enmendados, manipulados e intercambiados o introducidos de acuerdo con el proceso de la invención.

10 Preferiblemente, la referencia, control o tipo silvestre difiere del sujeto de la presente invención solamente en la actividad celular del polipéptido de la invención o el polipéptido utilizado en el método de la invención, por ejemplo, como resultado de un incremento en el nivel de la molécula de ácido nucleico de la presente invención o un incremento en la actividad específica del polipéptido de la invención o del polipéptido utilizado en el método de la invención. Por ejemplo, difiere por o en el nivel de expresión de la actividad de una proteína que tiene la actividad del YPR138C, sus causas bioquímicas o genéticas y por lo tanto muestra la cantidad incrementada de nitrógeno o de compuestos que contienen nitrógeno, la asimilación, acumulación y/o utilización potenciada de nitrógeno y/o el contenido incrementado de nitrógeno total y de la eficiencia incrementada del uso de nitrógeno (NUE), definida como la producción de biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno.

20 En caso de que no pueda proveerse un control, referencia o tipo silvestre que difiera del sujeto de la presente invención solamente por no ser sujeto del proceso de la invención, un control, referencia o tipo silvestre puede ser un organismo en el cual la causa de la modulación de una actividad que confiere el incremento de nitrógeno o compuestos que contienen nitrógeno o la expresión de las moléculas de ácido nucleico tal como se describen aquí ha sido conmutada por ejemplo suprimiendo la expresión del producto genético responsable, por ejemplo, por inhibición antisentido, por inactivación de un activador o agonista, por activación de un inhibidor o antagonista, o por inhibición a través de la adición de anticuerpos inhibidores, por la adición de compuestos activos, por ejemplo hormonas, por la introducción de mutantes dominantes negativos, etc. Una producción genética puede ser interrumpida por ejemplo introduciendo mutaciones puntuales inactivadoras, lo cual lleva a una inhibición o desestabilización enzimática o de la actividad biológica o a una inhibición de la capacidad de enlazarse a los cofactores etc.

30 De acuerdo con lo anterior, sujetos de referencia preferidos son el sujeto de partida del presente proceso de la invención. Preferiblemente, la referencia en la materia objeto de la invención se comparan después de la estandarización y normalización, por ejemplo, en la cantidad de ARN, ADN o proteína totales o la actividad o expresión de genes de referencia, como genes de mantenimiento, tales como ubiquitina, actina o proteínas ribosómicas.

35 Existe una serie de mecanismos a través de los cuales una modificación de una proteína, por ejemplo el polipéptido de la invención o el polipéptido utilizado en el método de la invención pueden afectar directa o indirectamente el consumo o asimilación de nitrógeno o el rendimiento, producción y/o eficiencia en la producción de compuestos que contienen nitrógeno.

40 Por ejemplo, se pueden implementar el número de moléculas o la actividad específica del polipéptido o la molécula de ácido nucleico. Pueden asimilarse o consumirse cantidades más grandes de nitrógeno o en el caso de que los compuestos que contienen nitrógeno producidos si el polipéptido o el ácido nucleico de la invención están expresados de novo en un organismo que carezca de la actividad de dicha proteína. Sin embargo, también es posible incrementar la expresión del gen que está presente de manera natural en los organismos, por ejemplo amplificando el número de genes, modificando la regulación del gen, o incrementando la estabilidad del ARNm correspondiente o del producto genético correspondiente codificado por la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención, o introduciendo genes homólogos a partir de otros organismos que están regulados diferentemente, por ejemplo, que no son sensibles a la retroalimentación.

50 El incremento, disminución o modulación de acuerdo con esta invención puede ser constitutivo, por ejemplo, debido a una expresión transgénica permanente estable o a una mutación estable en el gen endógeno correspondiente que codifica la molécula de ácido nucleico de la invención o la molécula de ácido nucleico utilizada en el método de la invención o a una modulación de la expresión o del comportamiento de un gen que confiere la expresión del polipéptido de la invención o el polipéptido utilizado en el método de la invención, o transiente, por ejemplo debido a una transformación transiente o a una edición temporal de un modulador tal como un agonista o antagonista o inducible, por ejemplo después de la transformación con un constructo inducible que porta la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención bajo control de un promotor inducible y agregando el inductor, por ejemplo, tetraciclina o tal como se describe aquí más adelante. El incremento en la actividad de las cantidades de polipéptido en una célula, un tejido, un organelo, un órgano o un organismo o una parte del mismo preferiblemente hasta al menos 5%, preferiblemente al menos 20% o a al menos 50%, preferiblemente de manera especial a al menos 70%, 80%, 90% o más, muy especialmente de manera preferible van hasta al menos 200%, lo más preferiblemente van hasta al menos 500%, o más en comparación con el control, referencia o tipo silvestre.

5 La actividad específica de un polipéptido codificado por una molécula de ácido nucleico de la presente invención o el polipéptido de la presente invención puede probarse según se describe en los ejemplos. En particular, la expresión de una proteína en cuestión en una célula, por ejemplo una célula vegetal o un microorganismo y la detección de un incremento en los niveles de nitrógeno o de compuestos que contienen nitrógeno frente a un control es una prueba fácil y puede llevarse a cabo tal como se describe en el estado de la técnica.

El término "incremento" incluye, que un compuesto o una actividad es introducida en una célula de novo de tal forma que el compuesto o la actividad no ha sido detectable antes, en otras palabras es "generado".

10 De acuerdo con lo anterior, en lo que sigue, el término "incremento" también comprende el término "generación" o "estimulación". La actividad incrementada se manifiesta por sí misma en una cantidad incrementada de nitrógeno o de compuestos que contienen nitrógeno.

15 En una realización, en el caso de que la actividad de la proteína YPR138C de *Saccharomyces cerevisiae* o sus homólogos, por ejemplo, como se indica en SEQ ID NO: 689, se incrementa; preferiblemente, un incremento de nitrógeno o de los compuestos que contienen nitrógeno entre 17% y 24% o más se confiere, preferiblemente se confiere un incremento del contenido de proteína o aminoácidos en una planta entre 17% y 24% o más. Preferiblemente este incremento es conferido en semillas o frutos de plantas.

En una realización, en el caso de que la actividad de la proteína YPR138C de la *Saccharomyces cerevisiae* o sus homólogos se incrementa, preferiblemente se confiere un incremento de nitrógeno o compuestos que contienen nitrógeno y de coenzima Q10, ácido fumárico, ácido málico y/o ácido lignocélico en hojas y/o glicerol-3-fosfato, ácido benzoico, ácido hidroxil benzoico y/o dodecanol en semillas de una planta.

20 Obedeciendo la actividad biológica de las proteínas que se utilizan en el proceso de acuerdo con la invención y que son codificadas por las moléculas de ácido nucleico de acuerdo con la invención, es posible producir composiciones que comprenden nitrógeno o compuestos que contienen nitrógeno. Dependiendo de la selección del organismo utilizado para el proceso de acuerdo con la presente invención, por ejemplo un microorganismo o una planta, pueden producirse composiciones o mezclas de diversos compuestos que contienen nitrógeno, por ejemplo, que contienen aminoácidos, ácidos grasos, vitamina, hormonas, azúcares, lípidos, etc. adicionalmente distintivos.

25 El término "expresión" se refiere a la transcripción y/o traducción de un segmento de gen o gen codogénicos. Como regla, el producto resultante es un ARNm o una proteína. Sin embargo, los productos de expresión también pueden incluir ARN funcionales tales como, por ejemplo, antisentido, ácidos nucleicos, ARNt, ARNsn, ARNr, ARNi, ARNsi, ribozimas, etc. La expresión puede ser sistémica, local o temporal, por ejemplo limitada a ciertos tipos de células, tejidos, órganos o periodos de tiempo.

En una realización, el proceso de la presente invención comprende una o más de las siguientes etapas;

35 a) estabilizar una proteína que confiere la expresión incrementada de una proteína codificada por la molécula de ácido nucleico de la invención o del polipéptido de la invención o de la molécula de ácido nucleico o el polipéptido usado en el método de la invención, por ejemplo, de un polipéptido que tiene una actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690, que tiene la actividad incrementadora aquí mencionada;

b) estabilizar un ARNm que confiere la expresión incrementada de una proteína codificada por la molécula de ácido nucleico de la invención o la molécula de ácido nucleico utilizada en el método de la invención, por ejemplo de un polipéptido que tiene actividad de YPR138C, por ejemplo, como se indica en SEQ ID NO: 690, o de un ARNm que codifica el polipéptido de la presente invención que tiene la actividad de incremento aquí mencionado;

40 c) incrementar la actividad específica de una proteína que confiere la expresión incrementada de una proteína codificada por la molécula de ácido nucleico de la invención o del polipéptido de la presente invención o la molécula de ácido nucleico o polipéptido utilizado en el método de la invención, que tiene la actividad incrementada aquí mencionada, por ejemplo, de un polipéptido que tiene actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690, o disminuir la regulación inhibitoria del polipéptido de la invención o del polipéptido usado en el método de la invención;

45 d) generar o incrementar la expresión de un factor de transcripción endógeno o artificial que media en la expresión de una proteína que confiere la expresión incrementada de una proteína codificada por una molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención o el polipéptido de la invención o el polipéptido utilizado en el método de la invención que tiene el incremento de actividad aquí mencionado, por ejemplo, de un polipéptido que tiene actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690;

e) estimular la actividad de una proteína que confiere la expresión incrementada de una proteína codificada por la molécula de ácido nucleico de la presente invención o un polipéptido de la presente invención que tiene la actividad de

incremento mencionada aquí, por ejemplo, de un polipéptido que tiene actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690, agregando uno o más factores inductores exógenos al organismo o partes del mismo;

5 f) expresar un gen transgénico que codifica una proteína que confiere la expresión incrementada de un polipéptido codificado por la molécula de ácido nucleico de la presente invención o un polipéptido de la presente invención, que tiene la actividad de incremento aquí mencionada, por ejemplo, de un polipéptido que tiene actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690;

10 g) incrementar el número de copias de un gen que confiere la expresión incrementada de una molécula de ácido nucleico que codifica un polipéptido codificado por la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención o el polipéptido de la invención o el polipéptido usado en el método de la invención que tiene la actividad de incremento aquí mencionada, por ejemplo, de un polipéptido que tiene actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690;

15 h) incrementar la expresión del gen endógeno que codifica el polipéptido de la invención o el polipéptido usado en el método de la invención, por ejemplo un polipéptido que tiene actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690, agregando elementos de expresión positiva o eliminando los de expresión negativa, por ejemplo, la recombinación homóloga puede ser utilizada para introducir elementos reguladores positivos como para las plantas el potenciador 35S en el promotor o para eliminar elementos represores de las regiones reguladores. Pueden utilizarse métodos adicionales de conversión genética para perturbar los elementos represores o para potenciar la actividad de los elementos positivos. Los elementos positivos pueden ser introducidos aleatoriamente en plantas por ADN-T o mutagénesis de transposón y pueden identificarse líneas en las cuales los elementos positivos pueden ser integrados
20 cerca a un gen de la invención, cuya expresión es por lo tanto potenciada,

25 i) modular las condiciones de crecimiento de un organismo de tal manera que la expresión o actividad del gen que codifica la proteína de la invención o la proteína misma se potencia, por ejemplo, microorganismos o plantas pueden ser cultivados bajo un régimen de temperatura más alto llevando a una expresión potenciada de las proteínas de choque al calor, por ejemplo, la proteína de choque al calor de la invención, la cual puede llevar a una producción potenciada del producto químico fino; y/o

j) seleccionar los organismos con actividad especialmente alta de las proteínas de la invención de fuentes naturales o mutagenizadas y cultivarlos en los organismos objetivo, por ejemplo, los cultivos de élite.

30 Preferiblemente, dicho ARNm es la molécula de ácido nucleico de la presente invención y/o la proteína que confiere la expresión incrementada de una proteína codificada por la molécula de ácido nucleico de la presente invención o el polipéptido que tiene la actividad mencionada aquí es el polipéptido de la presente invención, por ejemplo, que confiere el incremento del compuesto con contenido de N después de incrementar la expresión o actividad del polipéptido codificado o tener la actividad de un polipéptido que tiene actividad de YPR138C, por ejemplo como se indica en SEQ ID NO: 690.

35 En general, la cantidad de ARNm o polipéptido en una célula o compartimiento de un organismo se correlaciona con la cantidad de proteína codificada y así con la actividad global de la proteína codificada en dicho volumen. Dicha correlación no siempre es lineal, la actividad en el volumen es dependiente de la estabilidad de las moléculas o la presencia de cofactores activadores o inhibidores. Además, las inhibiciones del producto y educto de las enzimas son bien conocidas y están descritas en libros de texto, por ejemplo, Stryer, Biochemistry.

40 En general, la cantidad de moléculas de ARNm, polinucleótidos o ácidos nucleicos en una célula o un compartimiento de un organismo se correlaciona con la cantidad de proteína codificada y así con la actividad global de la proteína codificada en dicho volumen. Dicha correlación no siempre es lineal, la actividad en el volumen es dependiente de la estabilidad de las moléculas, la degradación de las moléculas a la presencia de cofactores activadores o inhibidores. Adicionalmente, las inhibiciones de producto y educto de las enzimas son bien conocidas, por ejemplo, Zinser et al., "Enzyminhibitoren"/Enzyme inhibitors".

45 La actividad de las proteínas y/o polipéptidos antes mencionados codificados por la molécula de ácido nucleico de la presente invención puede incrementarse de diversas maneras. Por ejemplo, la actividad en un organismo o en una parte del mismo, como una célula o un organelo, se incrementa incrementando el número de producto de gen, por ejemplo, incrementando la tasa de expresión, tal como mediante la introducción de un promotor más fuerte, o incrementando la estabilidad del ARNm expresado incrementando así la tasa de traducción y/o incrementando la
50 estabilidad del producto genético, reduciendo así las proteínas destruidas. Adicionalmente, la actividad o rendimiento de las enzimas puede ser influenciada de tal manera que se alcanza una reducción o incremento de la tasa de reacción o una modificación (reducción o incremento) de la afinidad al sustrato. Una mutación en el centro catalítico de un polipéptido de la invención o el polipéptido utilizado en el método de la invención, por ejemplo como una enzima, puede modular la tasa de rendimiento de la enzima, por ejemplo una supresión de un aminoácido esencial puede llevar a una
55 supresión reducida o completa de la actividad de la enzima, o la eliminación o mutación de sitios de un lanzamiento de

reguladores pueden reducir una regulación negativa como una inhibición de retroalimentación (o una inhibición de sustrato, si el nivel de sustrato también se incrementa). La actividad específica de la enzima de la presente invención puede incrementarse de tal manera que la rata de rendimiento se incrementa o el enlazamiento de un cofactor se mejora. Mejorando la estabilidad del ARNm codificador o la proteína también puede incrementar la actividad del producto genético. La estimulación de la actividad también está bajo el alcance del término "actividad incrementada".

Además, la regulación de las secuencias de ácidos nucleicos antes mencionadas puede ser modificada de tal manera que la expresión genética se incremente. Esto puede ser logrado ventajosamente por medio de secuencias reguladoras heterólogas o modificando, por ejemplo mutando, las secuencias reguladoras naturales que están presente. Los métodos ventajosos también pueden ser combinados uno con otro.

En general, una actividad de un producto genético en un organismo o parte del mismo, en particular en una célula vegetal, una planta, o un tejido de una planta, una parte de la misma o un organelo o un microorganismo puede incrementarse incrementando la cantidad del ARNm específico codificador o la proteína correspondiente en dicho organismo o parte del mismo. "Cantidad de proteína o ARNm" se entiende con el significado del número de moléculas de polipéptidos o moléculas de ARNm en un organismo, un tejido, una célula, o un compartimiento de célula. "Incremento" en la cantidad de una proteína significa el incremento cuantitativo del número de moléculas de dicha proteína en un organismo, un tejido, una célula o un compartimiento de célula o una parte del mismo – por ejemplo uno de los métodos descritos aquí más adelante - en comparación con tipo silvestre, control o referencia.

El incremento del número de moléculas se eleva preferiblemente a al menos 1%, preferiblemente más de 10%, más preferiblemente a 30% o más, especialmente de manera preferible a 50%, 70% o más, muy especialmente de forma preferible al 100%, lo más preferible a 500% o más. Sin embargo, una expresión de novo también es vista como sujeto de la presente invención.

Una modificación, esto es un incremento o descenso, puede ser causado por factores endógenos o exógenos. Por ejemplo, un incremento en una actividad en un organismo o parte del mismo puede ser causado por la adición de un producto genético o un precursor o un activador o un agonista al medio de nutrición o puede ser causado por la introducción de dichos objetos en un organismo, transiente o estable.

En una realización el incremento de la cantidad de nitrógeno o compuestos que contienen nitrógeno en el organismo o parte del mismo, por ejemplo en una célula, un tejido, un órgano, un organelo, etc. se alcanza incrementando el nivel endógeno del polipéptido de la invención o el polipéptido según el método de la invención en el citosol o en un compartimiento como los plástidos. De acuerdo con lo anterior, en una realización de la presente invención, la presente invención se relaciona con un proceso en donde el número de copias de gen de un gen que codifica el polinucleótido o molécula de ácido nucleico de la invención o el nivel de los polipéptidos de la invención o el polipéptido utilizado en el método de la invención, tal como se describe por ejemplo, puede ser implementado modificando la regulación transcripcional o translacional del polipéptido.

En una realización la cantidad de nitrógeno o compuestos que contienen nitrógeno en el organismo o parte del mismo puede incrementarse por mutagénesis direccionada o aleatoria de los genes endógenos de la invención. Por ejemplo puede utilizarse la recombinación homóloga bien sea para introducir elementos reguladores positivos como para plantas el potenciador 35S en el promotor o para retirar elementos represores de las regiones reguladoras. Además, métodos similares a la conversión genética descritos por Kochevenko y Willmitzer (Plant Physiol. 2003 May; 132(1): 174-84) y sus citaciones pueden ser utilizadas para perturbar elementos represores o potenciar la actividad de elementos reguladores positivos.

Además los elementos positivos pueden ser introducidos aleatoriamente en genomas (de plantas) por ADN-T o mutagénesis de transposón y pueden seleccionarse líneas y por ejemplo pueden utilizarse para recombinación homóloga para introducir elementos reguladores positivos como para las plantas el potenciador 35S en el promotor o para eliminar elementos represores de las regiones moduladoras. Adicionalmente pueden introducirse aleatoriamente en genomas (de plantas) por ADN-T o mutagénesis por trasposón y pueden seleccionarse líneas, en las cuales los elementos positivos tienen que ser integrados cerca a un gen de la invención, cuya expresión es por lo tanto potenciada. La activación de los genes de plantas por integraciones aleatorias de elementos potenciadores han sido descritas por Hayashi et al., 1992 (Science 258: 1350-1353) o Weigel et al., 2000 (Plant Physiol. 122, 1003-1013) y otros citados en el mismo.

Estrategias genéticas reversas para identificar inserciones (las cuales eventualmente portan los elementos de activación) cerca en genes de interés han sido descritas para varios casos, por ejemplo Krysan et al., 1999 (Plant Cell 1999, 11, 2283-2290); Sessions et al., 2002 (Plant Cell 2002, 14, 2985-2994); Young et al., 2001, (Plant Physiol, 2001, 125, 513-518); Koprek et al., 2000 (Plant J. 2000, 24, 253-263); Jeon et al., 2000 (Plant J. 2000, 22, 561-570); Tissier et al., 1999 (Plant Cell 1999, 11, 1841-1852); Speulmann et al., 1999 (Plant Cell 1999, 11, 1853-1866). En resumen el material de todas las plantas de un ADN-T grande o una población de plantas mutagenizadas por transposon se recolecta y se prepara el ADN genómico. Luego el ADN genómico es reunido siguiendo arquitecturas específicas como se describe por ejemplo en Krysan et al., 1999 (Plant Cell 1999, 11, 2283-2290). Las combinaciones de ADN genómico

son seleccionadas luego por reacciones de PCR multiplex específicas que detectan la combinación del mutágeno insercional (por ejemplo ADN-T o transposon) y el gen de interés. Por lo tanto las reacciones de PCR son corridas en las reservas de ADN con combinaciones específicas de ADN-T o cebadores de frontera de transposon y cebadores específicos del gen. La regla general es para un primer diseño pueden tomarse de Krysan et al., 1999 (Plant Cell 1999, 11, 2283-2290). La reelección de reservas de ADN en niveles más bajos lleva a la identificación de plantas individuales en las cuales el gen de interés es perturbado por el mutágeno de inserción.

El potenciamiento de los elementos reguladores positivos o la perturbación o debilitamiento de los elementos reguladores negativos también puede lograrse a través de técnicas de mutagénesis comunes: La producción de poblaciones mutadas químicamente por radiación es una técnica común y conocida para el experimentado en la técnica. Los métodos para plantas se describen en Koorneef et al., 1982 and the citations therein and by Lightner and Caspar in "Methods in Molecular Biology" Vol 82. Estas técnicas usualmente introducen mutaciones puntuales que pueden ser identificadas en cualquier gen conocido utilizando métodos tales como el *tilling* (Colbert et al., 2001).

De acuerdo con lo anterior, el nivel de expresión pueden ser incrementados si los genes endógenos codifican un polipéptido que confiere una expresión incrementada del polipéptido de la presente invención, en particular genes que comprenden la molécula de ácido nucleico de la presente invención, son modificados a través de la recombinación homóloga, metodología de *tilling* o conversión genética.

Las secuencias reguladoras pueden ser enlazadas operativamente a la región codificadora de una proteína endógena y controlar su transcripción y traducción o la estabilidad o pérdida del ARNm o la proteína expresada. Con el fin de modificar y controlar la expresión, pueden ser cambiados también el promotor, los UTR, los sitios de división, las señales de procesamiento, los sitios de poliadenilación, los terminadores, los potenciadores, los represores, los sitios de modificación post-transcripcionales o post-traducción, agregados o enmendados por ejemplo, la activación de los genes de plantas por integraciones aleatorias de elementos potenciadores que has sido descritas por Hayashi et al., 1992 (Science 258: 1350-1353) o Weigel et al., 2000 (Plant Physiol. 122, 1003-1013) y otros citados allí. Por ejemplo, el nivel de expresión de la proteína endógena puede ser modulado reemplazando el promotor endógeno con un promotor transgénico más fuerte o reemplazando los endógenos 3'UTR con 3'UTR, lo cual provee más estabilidad sin enmendar una región de codificación. Adicionalmente, la regulación de la transcripción puede ser modulada por la introducción de un factor de transcripción artificial tal como el que se describió en los ejemplo. A continuación se describen promotores alternos, terminadores y UTR.

La activación de un polipéptido endógeno que tiene la actividad arriba mencionada, el polipéptido de la invención o el polipéptido utilizado en el método de la invención, por ejemplo, que confiere el incremento de nitrógeno o compuestos que contienen nitrógeno después de incrementar la expresión o actividad en el citosol y/o en un organelo como un plástido, puede ser incrementada también introduciendo un factor de transcripción sintético, el cual se enlaza cerca a la región de codificación de un polipéptido endógeno de la invención o el polipéptido usado en el método de la invención – o usado en el proceso de la invención o su homólogo endógeno - que codifica un gen mediante lo cual el factor de transcripción sintético activa su transcripción. Puede considerarse una proteína de dedo de zinc quimérica, la cual comprende un dominio enlazante a ADN específico y un dominio de activación tal como por ejemplo un dominio de VP16 del virus de Herpes Simplex. El dominio de enlazamiento específico puede enlazarse a la región reguladora de la región de codificación de proteína endógena. La expresión del factor de transición quimérico en un organismo, en particular en una planta, lleva a una expresión específica de un polipéptido endógeno de la invención o usado en el proceso de la invención, en particular un homólogo vegetal del mismo, véase por ejemplo en WO 01/52620, Oriz, Proc.Natl. Acad. Sci. USA, 2002, Vol. 99, 13290 o Guan, Proc. Natl. Acad. Sci. USA, 2002, Vol. 99, 13296.

En una realización adicional del proceso de acuerdo con la invención, se usan organismos en los cuales uno de los genes antes mencionados, o uno de los ácidos nucleicos antes mencionados se hacen mutar en una manera tal que la actividad en los productos genéticos codificados esta menos influenciada por factores celulares, o no lo están del todo, en comparación con las proteínas no mutantes. Por ejemplo, un mecanismo de regulación bien conocido de la actividad enzimática es la inhibición del sustrato o mecanismos de regulación de retroalimentación. Las formas y técnicas para la introducción de las sustituciones, eliminaciones y adiciones de una o más bases, nucleótidos o aminoácidos de una secuencia correspondiente se describen aquí más adelante en los párrafos correspondientes y en las referencias listadas aquí, por ejemplo, en Sambrook et al., Molecular Cloning, Cold Spring Harbour, NY, 1989. La persona experimentada en la técnica será capaz de identificar los dominios de regulación y los sitios de enlazamiento de los reguladores comparando la secuencia de la molécula de ácido nucleico de la presente invención o el producto de expresión de la misma con el estado del arte por medios de software de ordenador los cuales comprenden algoritmos para identificar sitios de enlazamiento y dominios de regulación o introduciendo en una molécula de ácido nucleico o en una proteína sistemáticamente mutaciones y probando estas mutaciones las cuales llevan a una actividad específica incrementada o a una actividad incrementada por volumen, en particular por célula.

Por lo tanto es ventajoso expresar en un organismo una molécula de ácido nucleico de la invención o la molécula de ácido nucleico usado en el método de la invención o un polipéptido de la invención o el polipéptido usado en el método de la invención derivado de un organismo evolutivamente relacionado de manera distante, por ejemplo, utilizando un

gen procariota en un anfitrión eucariota, como en estos casos el mecanismo de regulación de la célula anfitriona puede no debilitar la actividad (celular o específica) del gen o su producto de expresión.

Menor influencia sobre la regulación de un gen o su producto genético se entiende con el significado de una regulación reducida de la actividad enzimática biológica que lleva a una actividad incrementada específica o celular del gen o su producto. Un incremento en la actividad enzimática biológica se entiende con el significado de una actividad enzimática o biológica, la cual se incrementa en al menos 10%, ventajosamente al menos, 20, 30 o 40%, especialmente de manera ventajosa en al menos 50, 60 o 70% en comparación con el organismo de partida. Esto lleva a una productividad incrementada del nitrógeno o compuestos que contienen nitrógeno deseados.

Obedeciendo a la introducción de un gen o a una pluralidad de genes que confieren la expresión de la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención o el polipéptido de la invención o el polipéptido usado en el método de la invención tal como se describe más adelante, por ejemplo el constructo de ácido nucleico mencionado más abajo, en un organismo solo o en combinación con otros genes, es posible no solamente incrementar el flujo biosintético hacia el producto final, por ejemplo, con el significado de compuestos que contienen nitrógeno, sino que también incrementa, modifica o crea de novo una composición en metabolitos novedosa preferiblemente, ventajosa en el organismo, por ejemplo, una composición ventajosa de aminoácidos que comprende un contenido más alto de (desde un punto de vista de fisiología nutricional limitada) productos químicos fino respectivos, en particular aminoácidos, probablemente compuestos de nitrógeno que contienen nitrógeno.

Preferiblemente, la composición comprende adicionalmente cantidades más grandes de metabolitos que afectan positivamente o disminuyen las cantidades de metabolitos que afectan negativamente la nutrición o la salud de los animales o humanos provistos con dichas composiciones u organismos de la invención o partes de los mismos. De la misma forma, el número o actividad de genes adicionales que se requiere para importar o exportar los nutrientes o metabolitos, incluyendo aminoácidos o sus precursores, requerido para la biosíntesis celular de los aminoácidos puede incrementarse de tal manera que la concentración de precursores necesarios relevantes, de los factores o intermedios dentro de las células o dentro de los compartimientos de almacenamiento se incrementa. Obedeciendo a la actividad incrementada o recién generada del polipéptido de la invención o el polipéptido utilizado en el método de la invención u obedeciendo al número incrementado de secuencias de ácidos nucleicos de la invención y/o a la modulación de genes adicionales que están involucrados en la biosíntesis de los aminoácidos, por ejemplo incrementando la actividad de enzimas que sintetizan precursores o destruyendo la actividad de uno o más genes que están involucrados en la ruptura de los aminoácidos, es posible incrementar el rendimiento, producción y/o eficiencia de producción directamente de aminoácidos en el organismo anfitrión, tal como las plantas o microorganismos.

De acuerdo con lo anterior, en una realización, el proceso de acuerdo con la invención se relaciona con un proceso que comprende:

a) proveer un organismo fotosintético activo, preferiblemente un microorganismo, una planta o un tejido de una planta o una planta;

b) incrementar una actividad de un polipéptido de la invención o el polipéptido utilizado en el método de la invención o un homólogo del mismo, por ejemplo, como se indica en SEQ ID NO: 689, o de un polipéptido que está siendo codificado por la molécula de ácido nucleico de la presente invención y descrita más abajo, esto es, confiriendo un incremento de nitrógeno o de compuestos que contienen nitrógeno en el organismo e incrementando la eficiencia del uso de nitrógeno (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo incluyendo fertilizantes de nitrógeno, preferiblemente en un organismo fotosintético activo, preferiblemente un microorganismo, una planta o un tejido vegetal o una planta.

c) hacer crecer el organismo, preferiblemente un organismo activo fotosintético, preferiblemente un microorganismo, una planta o un tejido de planta o una planta, bajo condiciones que permiten la acumulación y/o producción de nitrógeno o compuestos que contienen nitrógeno eficiencia en el uso de nitrógeno incrementado (NUC) definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno respectivamente en el organismo, preferiblemente un organismo activo fotosintético, preferiblemente un microorganismo, una planta o tejido de planta o una planta.

d) después del incremento antes descrito (el cual tal como se define arriba también abarca la generación de una actividad en un organismo, por ejemplo actividad de novo), por ejemplo después de la introducción y la expresión de las moléculas de ácido nucleico de la invención o de describirlo en métodos y procesos de acuerdo con la invención, el organismo de acuerdo con la invención, ventajosamente, un organismo fotosintético activo, preferiblemente un microorganismo, una planta o un tejido de planta o una planta, se cultiva y subsecuentemente se recolecta.

Organismos adecuados u organismos anfitriones (organismos transgénicos) para la molécula de ácido nucleico usada de acuerdo con la invención y para el proceso de la invención, el constructo de ácido nucleico o el vector (ambos como

se describen más adelante) están en principio todos los organismos que son capaces de sintetizar nitrógeno o compuestos que contienen nitrógeno, los cuales son adecuados para la activación, introducción o estimulación de genes. Ejemplos que pueden ser mencionados son plantas, microorganismos tales como hongos, bacterias, levaduras, algas o diatomáceas, transgénicas obtenidas por mutagénesis dirigida al sitio o mutagénesis aleatoria combinada con procedimientos de selección específica. Organismos preferidos son aquellos que son naturalmente capaces de acumular y/o sintetizar nitrógeno o compuestos que contienen nitrógeno en cantidades sustanciales, tales como hongos, levaduras, bacterias o plantas. En principio, los animales transgénicos, por ejemplo *Caenorhabditis elegans*, son adecuados como organismos anfitriones.

En el evento de que el organismo transgénico sea un microorganismo, tal como un organismo eucariote, por ejemplo un hongo, un alga, una diatomácea o una levadura en particular un hongo, alga, diatomácea o levadura seleccionadas de las familias Chaetomiaceae, Choanephoraceae, Cryptococcaceae, Cunninghamellaceae, Dematiaceae, Moniliaceae, Mortierellaceae, Mucoraceae, Pythiaceae, Saccharomycetaceae, Saprolegniaceae, Schizosaccharomycetaceae, Sodariaceae, Sporobolomycetaceae Tuberculariaceae, Adelotheciaceae, Dinophyceae, Ditrichaceae o Prasinophyceae, o un organismo procarionte, por ejemplo una bacteria o un alga azul, en particular una bacteria de las familias Actinomycetaceae, Bacillaceae, Brevibacteriaceae, Corynebacteriaceae, Enterobacteriaceae, Gordoniaceae, Nocardiaceae, Micrococcaceae, Mycobacteriaceae, Pseudomonaceae, Rhizobiaceae o Streptomycetaceae, este microorganismo se cultiva en un medio sólido o líquido el cual es conocido por la persona experimentada y es adecuado para el organismo. Después de la fase de crecimiento, los organismos pueden ser recolectados.

Los microorganismos, o respectivamente el nitrógeno o los compuestos que contienen nitrógeno recuperados, y si se desea aislados, como aminoácidos, pueden ser procesados adicionalmente de forma directa en materiales alimenticios o forrajes para animales u otras aplicaciones, por ejemplo de acuerdo con las divulgaciones hechas en EP-B-0 533 039 o EP-A-0 615 693, los cuales se incorporan aquí expresamente como referencia. El caldo de fermentación o los productos de fermentación pueden ser purificados de la forma habitual por extracción y precipitación o a través de intercambiadores de iones u otros métodos conocidos para las personas experimentadas en la técnica descritos aquí más adelante. Los productos de estos diferentes procedimientos de manipulación son aminoácidos o composiciones de aminoácidos que aún comprenden caldos de fermentación y componentes celulares en diferentes cantidades, ventajosamente en el rango de 0 a 99% en peso, preferiblemente por debajo de 80% en peso, preferiblemente de manera especial por debajo de 50% en peso.

Cepas preferidas son cepas seleccionadas del grupo consistente de Bacillaceae, Brevibacteriaceae, Corynebacteriaceae, Nocardiaceae, Mycobacteriaceae, Streptomycetaceae, Enterobacteriaceae tales como *Bacillus circulans*, *Bacillus subtilis*, *Bacillus sp.*, *Brevibacterium albidum*, *Brevibacterium album*, *Brevibacterium cerinum*, *Brevibacterium flavum*, *Brevibacterium glutamigenes*, *Brevibacterium iodinum*, *Brevibacterium ketoglutamicum*, *Brevibacterium lactofermentum*, *Brevibacterium linens*, *Brevibacterium roseum*, *Brevibacterium saccharolyticum*, *Brevibacterium sp.*, *Corynebacterium acetoacidophilum*, *Corynebacterium acetoglutamicum*, *Corynebacterium ammoniagenes*, *Corynebacterium glutamicum* (= *Micrococcus glutamicum*), *Corynebacterium melassecola*, *Corynebacterium sp.*, *Nocardia rhodochrous* (*Rhodococcus rhodochrous*), *Mycobacterium rhodochrous*, *Streptomyces lividans* y *Escherichia coli* especialmente *Escherichia coli* K12.

Además cepas particularmente preferidas son cepas seleccionadas del grupo consistente de Cryptococcaceae, Saccharomycetaceae, Schizosaccharomycetaceae tales como los géneros *Candida*, *Hansenula*, *Pichia*, *Saccharomyces* y *Schizosaccharomyces* preferidos son cepas seleccionadas del grupo consistente de las especies *Rhodotorula rubra*, *Rhodotorula glutinis*, *Rhodotorula graminis*, *Yarrowia lipolytica*, *Sporobolomyces salmonicolor*, *Sporobolomyces shibatanus*, *Saccharomyces cerevisiae*, *Candida boidinii*, *Candida bombicola*, *Candida cylindracea*, *Candida parapsilosis*, *Candida rugosa*, *Candida tropicalis*, *Pichia methanolica* y *Pichia pastoris*.

Anacardiaceae tales como los géneros *Pistacia*, *Mangifera*, *Anacardium* por ejemplo las especies *Pistacia vera* [pistachios, Pistazie], *Mangiferindica* [Mango] o *Anacardium occidentale* [castaña]; Asteraceae tales como los géneros *Calendula*, *Carthamus*, *Centaurea*, *Cichorium*, *Cynara*, *Helianthus*, *Lactuca*, *Locusta*, *Tagetes*, *Valeriana* e.g. la especie *Calendula officinalis* [caléndula], *Carthamus tinctorius* [cártamo], *Centaurea cyanus* [flor de maíz], *Cichorium intybus* [margarita azul], *Cynara scolymus* [alcachofa], *Helianthus annuus* [girasol], *Lactuca sativa*, *Lactuca crispata*, *Lactuca esculenta*, *Lactuca scariola* L. ssp. *sativa*, *Lactuca scariola* L. var. *integrata*, *Lactuca scariola* L. var. *integrifolia*, *Lactuca sativa* subsp. *romana*, *Locusta communis*, *Valeriana locusta* [lechuga], *Tagetes lucida*, *Tagetes erecta* o *Tagetes tenuifolia* [Caléndula]; Apiaceae tales como los géneros *Daucus* e.g. la especie *Daucus carota* [zanahoria]; Betulaceae tales como los géneros *Corylus* e.g. la especie *Corylus avellana* o *Corylus colurna* [avellana]; Boraginaceae tales como los géneros *Borago* e.g. la especie *Borago officinalis* [borage]; Brassicaceae tales como los géneros *Brassica*, *Melanosinapis*, *Sinapis*, *Arabadopsis* e.g. la especie *Brassica napus*, *Brassica rapa* ssp. [canola, colza, nabo], *Sinapis arvensis*, *Brassica juncea*, *Brassica juncea* var. *juncea*, *Brassica juncea* var. *crispifolia*, *Brassica juncea* var. *foliosa*, *Brassica nigra*, *Brassica sinapioides*, *Melanosinapis communis* [mostaza], *Brassica oleracea* [remolacha para forraje] o *Arabadopsis thaliana*; Bromeliaceae tales como los géneros *Ananas*, *Bromelia* e.g. la especie *Ananas comosus*, *Ananas ananas* or *Bromelia comosa* [pineapple]; Caricaceae tales como los géneros *Carica* e.g. la especie *Carica papaya* [papaya]; Cannabaceae tales como los géneros *Cannabis* e.g. la especie *Cannabis sativa* [hemp], Convolvulaceae tales como los géneros *Ipomea*, *Convolvulus* e.g. la especie *Ipomea batatas*, *Ipomea pandurata*, *Convolvulus batatas*,

Convolvulus tiliaceus, *Ipomoea fastigiata*, *Ipomoea tiliacea*, *Ipomoea triloba* o *Convolvulus panduratus* [patata dulce, Hombre de la Tierra, patata silvestre]; Chenopodiaceae tales como los géneros Beta, i.e. la especie *Beta vulgaris*, *Beta vulgaris* var. *altissima*, *Beta vulgaris* var. *Vulgaris*, *Beta maritima*, *Beta vulgaris* var. *perennis*, *Beta vulgaris* var. *conditiva* o *Beta vulgaris* var. *esculenta* [remolacha de azúcar]; Cucurbitaceae tales como los géneros Cucurbita e.g. la especie *Cucurbita maxima*, *Cucurbita mixta*, *Cucurbita pepo* o *Cucurbita moschata* [calabaza, calabacín]; Elaeagnaceae tales como los géneros Elaeagnus e.g. la especie *Olea europaea* [oliva]; Ericaceae tales como los géneros Kalmia e.g. la especie *Kalmia latifolia*, *Kalmia angustifolia*, *Kalmia microphylla*, *Kalmia polifolia*, *Kalmia occidentalis*, *Cistus chamaerhodendros* o *Kalmia lucida* [laurel americano, laurel de hoja ancha, arbusto calico, madera para cuchara, laurel ovejero, laurel alpino, laurel de ciénaga, laurel de ciénaga occidental, laurel de pantano]; Euphorbiaceae tales como los géneros Manihot, Janipha, Jatropha, Ricinus e.g. la especie *Manihot utilissima*, *Janipha manihot* " *Jatropha manihot*, *Manihot aipil*, *Manihot dulcis*, *Manihot manihot*, *Manihot melanobasis*, *Manihot esculenta* [mandioca, raíz flecha, tapioca, cassava] o *Ricinus communis* [judía castor, arbusto de aceite de castor, planta de aceite de castor, Palma Christi, árbol maravilla]; Fabaceae tales como los géneros Pisum, Albizia, Cathormion, Feuillea, Inga, Pithecolobium, Acacia, Mimosa, Medicago, Glycine, Dolichos, Phaseolus, Soja e.g. la especie *Pisum sativum*, *Pisum arvense*, *Pisum humile* [guisante], *Albizia berteriana*, *Albizia julibrissin*, *Albizia lebbek*, *Acacia berteriana*, *Acacia littoralis*, *Albizia berteriana*, *Albizia berteriana*, *Cathormion berteriana*, *Feuillea berteriana*, *Inga fragrans*, *Pithecolobium berterianum*, *Pithecolobium fragrans*, *Pithecolobium berterianum*, *Pseudalbizia berteriana*, *Acacia julibrissin*, *Acacia nemu*, *Albizia nemu*, *Feuillea julibrissin*, *Mimosa julibrissin*, *Mimosa speciosa*, *Sericanrda julibrissin*, *Acacia lebbeck*, *Acacia macrophylla*, *Albizia lebbeck*, *Feuillea lebbeck*, *Mimosa lebbeck*, *Mimosa speciosa* [leño bastardo, árbol de seda, nuez de las Indias Orientales], *Medicago sativa*, *Medicago falcata*, *Medicago varia* [alfalfa] *Glycine max* *Dolichos soja*, *Glycine gracilis*, *Glycine hispida*, *Phaseolus max*, *Soja hispida* or *Soja max* [soja]; Geraniaceae tales como los géneros Pelargonium, Cocos, Oleum e.g. la especie *Cocos nucifera*, *Pelargonium grossularioides* o *Oleum cocois* [coco]; Gramineae tales como los géneros Saccharum e.g. la especie *Saccharum officinarum*; Juglandaceae tales como los géneros Juglans, Wallia e.g. la especie *Juglans regia*, *Juglans ailanthifolia*, *Juglans sieboldiana*, *Juglans cinerea*, *Wallia cinerea*, *Juglans bixbyi*, *Juglans californica*, *Juglans hindsii*, *Juglans intermedia*, *Juglans jamaicensis*, *Juglans major*, *Juglans microcarpa*, *Juglans nigra* o *Wallia nigra* [nuez, nuez negra, nuez común, nuez persa, nuez blanc, nuez de manteca, nuez negra.]; Lauraceae tales como los géneros Persea, Laurus e.g. la especie laurel *Laurus nobilis* [diversos laureles, laurel dulce], *Persea americana* *Persea americana*, *Persea gratissima* o *Persea persea* [aguacate]; Leguminosae tales como los géneros Arachis e.g. la especie *Arachis hypogaea* [peanut]; Linaceae tales como los géneros Linum, Adenolinum e.g. la especie *Linum usitatissimum*, *Linum humile*, *Linum austriacum*, *Linum bienne*, *Linum angustifolium*, *Linum catharticum*, *Linum flavum*, *Linum grandiflorum*, *Adenolinum grandiflorum*, *Linum lewisii*, *Linum narbonense*, *Linum perenne*, *Linum perenne* var. *lewisii*, *Linum pratense* or *Linum trigynum* [lino, linaza]; Lytharieae tales como los géneros Punica e.g. la especie *Punica granatum* [granada]; Malvaceae tales como los géneros Gossypium e.g. la especie *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, *Gossypium herbaceum* o *Gossypium thurberi* [algodón]; Musaceae tales como los géneros Musa e.g. la especie *Musa nana*, *Musa acuminata*, *Musa paradisiaca*, *Musa* spp. [plátano]; Onagraceae tales como los géneros Camissonia, Oenothera e.g. la especie *Oenothera biennis* o *Camissonia brevipes* [prímula, prímula de la tarde]; Palmae tales como los géneros Elacis e.g. la especie *Elaeis guineensis* [oil plam]; Papaveraceae tales como los géneros Papaver e.g. la especie *Papaver orientale*, *Papaver rhoeas*, *Papaver dubium* [amapola, amapola oriental, amapola florida, amapola de campo, amapola Shirley, amapola de campo, amapola de cabeza larga, amapola de vaina larga]; Pedaliaceae tales como los géneros Sesamum e.g. la especie *Sesamum indicum* [sésamo]; Piperaceae tales como los géneros Piper, Artanthe, Peperomia, Steffensia e.g. la especie *Piper aduncum*, *Piper amalago*, *Piper angustifolium*, *Piper auritum*, *Piper betel*, *Piper cubeba*, *Piper longum*, *Piper nigrum*, *Piper retrofractum*, *Artanthe adunca*, *Artanthe elongata*, *Peperomia elongata*, *Piper elongatum*, *Steffensia elongata*. [pimienta de Cayena, pimienta silvestre]; Poaceae tales como los géneros Hordeum, Secale, Avena, Sorghum, Andropogon, Holcus, Panicum, Oryza, Zea, Triticum e.g. la especie *Hordeum vulgare*, *Hordeum jubatum*, *Hordeum murinum*, *Hordeum secalinum*, *Hordeum distichon* *Hordeum aegiceras*, *Hordeum hexastichon.*, *Hordeum hexastichum*, *Hordeum irregulare*, *Hordeum sativum*, *Hordeum secalinum* [cebada, cebada perla, cebada cola de zorro, cebada de pared, cebada de prado], *Secale cereale* [centeno], *Avena sativa*, *Avena fatua*, *Avena byzantina*, *Avena fatua* var. *sativa*, *Avena hybrida* [avena], *Sorghum bicolor*, *Sorghum halepense*, *Sorghum saccharatum*, *Sorghum vulgare*, *Andropogon drummondii*, *Holcus bicolor* o *Holcus sorghum*, *Holcus sorghum*, *Sorghum aethiopicum*, *Sorghum arundinaceum*, *Sorghum caffrorum*, *Sorghum cernuum*, *Sorghum dochna*, *Sorghum drummondii*, *Sorghum durra*, *Sorghum guineense*, *Sorghum lanceolatum*, *Sorghum nervosum*, *Sorghum saccharatum*, *Sorghum subglabrescens*, *Sorghum verticilliflorum*, *Sorghum vulgare*, *Holcus halepensis*, *Sorghum miliaceum* millo, *Panicum militaceum* [sorgo, millo], *Oryza sativa*, *Oryza latifolia* [arroz], *Zea mays* [maíz] *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum* o *Triticum vulgare* [trigo, trigo de pan, trigo común], Proteaceae tales como los géneros Macadamia e.g. la especie *Macadamia intergrifolia* [macadamia]; Rubiaceae tales como los géneros Coffea e.g. la especie *Coffea* spp., *Coffea arabica*, *Coffea canephora* o *Coffea liberica* [café]; Scrophulariaceae tales como los géneros Verbascum e.g. la especie *Verbascum blattaria*, *Verbascum chaixii*, *Verbascum densiflorum*, *Verbascum lagurus*, *Verbascum longifolium*, *Verbascum lychnitis*, *Verbascum nigrum*, *Verbascum olympicum*, *Verbascum phlomoides*, *Verbascum phoenicum*, *Verbascum pulverulentum* o *Verbascum thapsus* [gordolobo, gordolobo de boca blanca, gordolobo con hojas de ortiga, gordolobo de flores densas, gordolobo plateada, gordolobo de hojas largas, gordolobo blanca, gordolobo oscura, gordolobo verde, gordolobo naranja, gordolobo púrpura, gordolobo cano, gran gordolobo]; Solanaceae tales como los géneros Capsicum, Nicotiana, Solanum, Lycopersicon e.g. la especie *Capsicum annuum*, *Capsicum annuum* var. *glabrusculum*, *Capsicum frutescens* [pimiento], *Capsicum annuum* [paprika], *Nicotiana tabacum*, *Nicotiana glauca*, *Nicotiana glauca*, *Nicotiana glauca*, *Nicotiana langsdorffii*, *Nicotiana obtusifolia*, *Nicotiana quadrivalvis*, *Nicotiana repanda*, *Nicotiana rustica*, *Nicotiana sylvestris* [tabaco], *Solanum tuberosum* [patata], *Solanum*

5 *melongena* [berenjena] (*Lycopersicon esculentum*, *Lycopersicon lycopersicum.*, *Lycopersicon pyriforme*, *Solanum integrifolium* o *Solanum lycopersicum* [tomate]; Sterculiaceae tales como los géneros *Theobroma* e.g. la especie *Theobroma cacao* [cacao]; Theaceae tales como los géneros *Camellia* e.g. la especie *Camellia sinensis*) [té] pueden ser bien sea donadores de organismos para los ácidos nucleicos o polipéptidos de la invención o utilizados en la presente invención o representan organismos anfitriones preferidos.

10 Plantas anfitrionas preferidas en particular son plantas seleccionadas del grupo consistente de Asteraceae tales como los géneros *Helianthus*, *Tagetes* e.g. la especie *Helianthus annuus* [girasol], *Tagetes lucida*, *Tagetes erecta* o *Tagetes tenuifolia* [Caléndula], Brassicaceae tales como los géneros *Brassica*, *Arabidopsis* e.g. la especie *Brassica napus*, *Brassica rapa* ssp. [canola, colza, nabo] or *Arabidopsis thaliana*. Fabaceae tales como los géneros *Glycine* e.g. la especie *Glycine max*, *Soja hispida* o *Soja max* [soybean]. Linaceae tales como los géneros *Linum* e.g. la especie *Linum usitatissimum*, [flax, linseed]; Poaceae tales como los géneros *Hordeum*, *Secale*, *Avena*, *Sorghum*, *Oryza*, *Zea*, *Triticum* e.g. la especie *Hordeum vulgare* [cebada]; *Secale cereale* [centeno], *Avena sativa*, *Avena fatua*, *Avena byzantina*, *Avena fatua* var. *sativa*, *Avena hybrida* [avena], *Sorghum bicolor* [sorgo, millo], *Oryza sativa*, *Oryza latifolia* [arroz], *Zea mays* [maíz] *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum* o *Triticum vulgare* [trigo, trigo blanco, tigo común]; Solanaceae tales como los géneros *Solanum*, *Lycopersicon* e.g. la especie *Solanum tuberosum* [patata], *Lycopersicon esculentum*, *Lycopersicon lycopersicum.*, *Lycopersicon pyriforme*, *Solanum integrifolium* or *Solanum lycopersicum* [tomate]. Aun organismo anfitrión preferido adicional es algodón, por ejemplo *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, *Gossypium herbaceum* o *Gossypium thurberi*.

20 Todos los organismos antes mencionados pueden en principio funcionar también como organismos donantes.

Con respecto a la secuencia de ácido nucleico puesto que representa un constructo de ácido nucleico que contiene una secuencia de ácidos nucleicos mencionada aquí o un organismo (igual organismo transgénico) el cual es transformado con dicha secuencia de ácidos nucleicos o dicho constructo de ácidos nucleicos, "transgen" significa todos aquellos constructos que han sido preparados por métodos de manipulación genética, preferiblemente en los cuales

- 25 a) una secuencia de ácidos nucleicos seleccionada del grupo tal como se indica en SEQ ID NO: 689, o un derivado del mismo, o
- b) un elemento genético regulador, por ejemplo, un promotor, el cual está enlazado funcionalmente a la secuencia de ácidos nucleicos como se indica en SEQ ID NO: 689, o un derivado del mismo, o
- c) (a) y (b)

30 No es/están presentes en su ambiente genético natural o ha/han sido modificados por medios de métodos de manipulación genética, siendo posible para que la modificación sea, a manera de ejemplo, una sustitución, adición, eliminación, inversión o inserción de uno o más nucleótidos. "Ambiente genético natural" significa la localización cromosómica natural en el organismo de origen o la presencia en una biblioteca genómica. En el caso de una biblioteca genómica, el ambiente genético natural de la secuencia de ácidos nucleicos es preservado preferiblemente al menos de forma parcial. El ambiente flanquea la secuencia de ácidos nucleicos al menos en un lado y tiene una longitud de secuencia de al menos 50bp, preferiblemente al menos 500 bp, particularmente de manera preferible al menos 1000 bp, muy particularmente de manera preferible al menos 5000 bp.

El uso de la secuencia de ácidos nucleicos de acuerdo con la invención o el constructo de ácidos nucleicos de acuerdo con la invención para la generación de plantas transgénicas por lo tanto objetivo de la invención.

40 En una realización ventajosa de la invención, el organismo toma la forma de una planta cuyo contenido de nitrógeno o compuestos que contienen nitrógeno es modificado ventajosamente obedeciendo a la molécula de ácido nucleico de la presente invención expresada. Esto es importante para cultivadores de plantas por varias razones:

45 a) la inmensa mayoría de nitrógeno está presente en las células en forma de aminoácidos enlazados en proteínas. Por lo tanto un contenido incrementado de nitrógeno o de compuestos que contienen nitrógeno refleja un contenido incrementado de proteína y por lo tanto un valor nutricional adicional para la industria alimentaria.

b) un método para consumo incrementado de nitrógeno y/o acumulación de nitrógeno o de compuestos que contienen nitrógeno podría permitir reducir la aplicación de fertilizantes de nitrógeno, lo cual a su vez lleva a costos reducidos y beneficios ambientales.

50 c) un método para consumo y/o acumulación incrementados de nitrógeno podría soportar el crecimiento de la planta, su salud y productividad, preferiblemente bajo condiciones con limitaciones de nitrógeno.

5 En una realización, después de que una actividad de un polipéptido de la presente invención o usado en el proceso de la presente invención haya sido incrementada o generada, o después de que la expresión de una molécula de ácido nucleico o polipéptido de acuerdo con la invención ha sido generada incrementada, la planta transgénica generada puede cultivarse sobre o en un medio nutritivo o aún más en el suelo y subsecuentemente recolectada. En una realización la planta transgénica generada puede ser cultivada bajo condiciones limitantes de nitrógeno.

10 Las plantas o partes de las mismas, por ejemplo, las hojas, raíces, flores y/o tallos y/o otro material recolectable tal como se describe anteriormente, pueden ser utilizados entonces directamente como material alimenticio o forraje para animales o incluso pueden ser procesados. De nuevo, los aminoácidos pueden ser purificados adicionalmente en la forma habitual a través de la extracción y precipitación o a través de intercambiadores de iones y otros métodos conocidos para la persona experimentada en la técnica y descritos aquí más adelante. Los productos que son adecuados para diversas aplicaciones y que resultan de estos diferentes procedimientos de procesamiento son aminoácidos o composiciones de aminoácidos los cuales aún comprenden componentes adicionales de las plantas en cantidades diferentes, ventajosamente en el rango de 0 a 99% en peso, preferiblemente por debajo de 90% en peso, especialmente de manera preferible por debajo de 80% en peso. Las plantas también pueden ser utilizadas ventajosamente de manera directa sin procesamiento adicional, por ejemplo, como alimento o para extracción.

15 Los compuestos que contienen nitrógeno químicamente puro o composiciones químicamente puras que comprenden nitrógeno o compuestos que contienen nitrógeno pueden ser producidos también por el proceso descrito anteriormente. Con este fin, el nitrógeno o los compuestos que contienen nitrógeno o las composiciones se aíslan de la manera conocida a partir de un organismo de acuerdo con la invención, tales como microorganismos, animales no humanos o plantas, y/o su medio de cultivo sobre los cuales los organismos han sido cultivados. Estos compuestos que contienen nitrógeno químicamente puro o dichas composiciones son ventajosos para aplicaciones en el campo de la industria de forrajes o de alimentos.

20 En una realización preferida, la presente invención se relaciona con un proceso para la asimilación, acumulación y/o utilización incrementada de nitrógeno y para el contenido incrementado de nitrógeno total y la eficiencia en el uso de nitrógeno incrementada (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo el nitrógeno de fertilizantes en organismos fotosintéticos activos, que comprende incrementar o generar en un organismo o en una parte o en un compartimento del mismo la expresión de al menos una molécula de ácido nucleico que comprende una molécula de ácido nucleico seleccionada del grupo consistente de:

25 a) molécula de ácido nucleico que codifica un polipéptido como se muestra en SEQ ID NO: 690 o un fragmento de la misma con una actividad de proteína de transporte de amonio, lo cual confiere asimilación, acumulación y/o utilización potenciada de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno,

30 b) molécula de ácido nucleico que comprende una molécula de ácido nucleico como se muestra en SEQ ID NO: 689 la cual confiere asimilación, acumulación y/o utilización incrementadas de nitrógeno y contenido de nitrógeno total incrementado y eficiencia incrementada en el uso de nitrógeno;

35 c) molécula de ácido nucleico cuya secuencia puede ser deducida a partir de una secuencia de polipéptidos codificada por una molécula de ácido nucleico de (a) o (b) como resultado de la degeneración del código genético y que confiere asimilación, acumulación y/o utilización potenciada de nitrógeno y contenido de nitrógeno total incrementado y eficiencia incrementada en el uso de nitrógeno;

40 d) molécula de ácido nucleico que codifica un polipéptido con una actividad de proteína de transporte de amonio que tiene al menos 70% de identidad con la secuencia de aminoácidos del polipéptido codificado por la molécula de ácido nucleico de (a) a (c) y que confiere asimilación, acumulación y/o utilización potenciada de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno.

45 A menos que se especifique otra cosa, los términos "polinucleótidos", "ácido nucleico" y "molécula de ácido nucleico" son intercambiables en el presente contexto. A menos que se especifique otra cosa, los términos "péptidos", "polipéptidos" y "proteínas" son intercambiables en el presente contexto. El término "secuencia" puede relacionarse con polinucleótidos, ácidos nucleicos, moléculas de ácidos nucleicos, péptidos, polipéptidos y proteínas, dependiendo del contexto en el cual se utilice el término "secuencia". Los términos "gen(es)", "polinucleótido", "secuencia de ácidos nucleicos", "secuencia de nucleótidos", o "moléculas de ácidos nucleicos" tal como se utilizan aquí se refieren a una forma polimérica de nucleótidos de cualquier longitud, bien sean ribonucleótidos o desoxirribonucleótidos. Los términos se refieren solamente a la estructura primaria de la molécula.

55 Así, los términos "gen(es)", "polinucleótidos", "secuencia de ácidos nucleicos", "secuencia de nucleótidos", o "moléculas de ácidos nucleicos", tal como se utilizan aquí incluyen ADN y ARN de cadena doble y sencilla. También incluyen tipos conocidos de modificaciones, por ejemplo, metilación, "tapas", sustituciones de uno o más de los nucleótidos de origen natural con un análogo. Preferiblemente, la secuencia de ADN o ARN de la invención comprende una secuencia de codificación que codifica el polipéptido aquí definido.

- Una "secuencia de codificación" es una secuencia de nucleótidos, la cual es transcrita en ARNm y/o traducida en un polipéptido cuando se coloca bajo control de secuencias reguladoras apropiadas. Los límites de la secuencia de codificación son determinados por un codón de inicio de traducción en el terminal 5' y un codón de detención de la traducción en el terminal 3'. Una secuencia de codificación puede incluir, pero no se limita a ARNm, ADNc, secuencia de nucleótidos recombinantes o ADN genómico, mientras que los intrones pueden estar presentes también bajo ciertas circunstancias.
- En una realización, dichas secuencias son clonadas en constructos de ácidos nucleicos, bien sea individualmente o en combinación. Estos constructos de ácidos nucleicos permiten una acumulación y/o síntesis óptimas de nitrógeno o de compuestos que contienen nitrógeno producidos respectivamente en el proceso de acuerdo con la invención.
- Las moléculas del ácido nucleico, los cuales son ventajosas para el proceso de acuerdo con la invención y que codifican polipéptidos con una actividad de un polipéptido de la invención o el polipéptido usado en el método de la invención o usado en el proceso de la invención, por ejemplo, de una proteína como se indica en SEQ ID NO: 690, o que son codificadas por una molécula de ácido nucleico en SEQ ID NO: 689, puede determinarse a partir de bases de datos en general accesibles.
- Aquellas, las cuales deben ser mencionadas en particular en este contexto son bases de datos generales tales como la base de datos EMBL (Stoesser G. et al., *Nucleic Acids Res* 2001, Vol. 29, 17-21), la base de datos GenBank (Benson D.A. et al., *Nucleic Acids Res* 2000, Vol. 28, 15-18) o la base de datos PIR (Baker W.C. et al., *Nucleic Acids Res.* 1999, Vol. 27, 39-43). Además es posible usar bases de datos de genes específicos para organismos para determinar secuencias ventajosas, en el caso de levaduras por ejemplo ventajosamente la base de datos SGD (Cherry J.M. et al., *Nucleic Acids Res.* 1998, Vol. 26, 73-80) o la base de datos MIPS (Mewes H.W. et al., *Nucleic Acids Res.* 1999, Vol. 27, 44-48), en el caso de *E. coli* la base de datos GenProtEC (<http://web.bham.ac.uk/bcm4ght6/res.html>), y en el caso de *Arabidopsis* la base de datos TAIR (Huala, E. et al., *Nucleic Acids Res.* 2001 Vol. 29(1), 102-5) o la base de datos MIPS.
- Las moléculas de ácidos nucleicos usadas en el proceso de acuerdo con la invención toman la forma de secuencias de ácidos nucleicos aisladas, las cuales codifican polipéptidos con una actividad de un polipéptido seleccionado del grupo como se indica en SEQ ID NO: 689 o que tiene la secuencia de un polipéptido como se indica en SEQ ID NO: 690 y confiere un incremento de nitrógeno o de compuestos que contienen nitrógeno.
- Las secuencias de ácidos nucleicos usadas en el proceso para la producción de nitrógeno o compuestos que contienen nitrógeno en organismos transgénicos se originan ventajosamente de un eucariote pero también pueden originarse de un procarionte o de una arqueobacteria, así puede ser derivada de por ejemplo un microorganismo, un animal o una planta.
- Para los propósitos de la invención, como regla se entiende que el plural abarca el singular y viceversa.
- Con el fin de mejorar la introducción de las secuencias de ácido nucleico y la expresión de la secuencia en los organismos transgénicos, que se utilizan en el proceso, las secuencias de ácidos nucleicos se incorporan en un constructo de ácido nucleico y/o un vector. Además de las secuencias aquí descritas que se utilizan en el proceso de acuerdo con la invención, pueden estar presentes también secuencias adicionales de ácidos nucleicos, que ventajosamente codifican la asimilación del nitrógeno o la biosíntesis de aminoácidos que generan proteínas de almacenamiento de valor nutricional, en el constructo de ácido nucleico o en el vector y pueden ser introducidas en el organismo juntas. Sin embargo, estas secuencias adicionales pueden ser introducidas también en los organismos a través de otros constructos de ácidos nucleicos o vectores separados.
- Usando los vectores de clonación y los métodos de transformación aquí mencionados tales como los que se publican y se citan en: *Plant Molecular Biology and Biotechnology* (CRC Press, Boca Raton, Florida), chapter 6/7, pp. 71-119 (1993); F.F. White, *Vectors for Gene Transfer in Higher Plants*; in: *Transgenic Plants*, vol. 1, Engineering and Utilization, Ed.: Kung and R. Wu, Academic Press, 1993, 15-38; B. Jenes et al., *Techniques for Gene Transfer*, in: *Transgenic Plants*, vol. 1, Engineering and Utilization, Ed.: Kung and R. Wu, Academic Press (1993), 128-143; Potrykus, *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 42 (1991), 205-225) y citados posteriormente aquí, los ácidos nucleicos pueden ser utilizados para la modificación por recombinación de un amplio rango de organismos, en particular microorganismos procariontes o eucariotes o plantas de tal manera que se convierten en un acumulador y/o productor mejor o más eficiente de nitrógeno o de compuestos que contienen nitrógeno respectivamente de acuerdo con la invención. Esta acumulación mejorada de nitrógeno o producción de compuestos que contienen nitrógeno o productos derivados de los mismos, tales como proteínas, puede obtenerse por un efecto directo de la manipulación o por un efecto indirecto de esta manipulación.
- En una realización, la molécula de ácido nucleico de acuerdo con la invención se origina de una planta, tal como una planta seleccionada de las familias Aceraceae, Anacardiaceae, Apiaceae, Asteraceae, Brassicaceae, Cactaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, Malvaceae, Nymphaeaceae, Papaveraceae, Rosaceae, Salicaceae, Solanaceae, Arecaceae, Bromeliaceae, Cyperaceae, Iridaceae, Liliaceae, Orchidaceae, Gentianaceae, Labiaceae, Magnoliaceae, Ranunculaceae, Carifolaceae, Rubiaceae, Scrophulariaceae, Caryophyllaceae, Ericaceae,

Polygonaceae, Violaceae, Juncaceae las Poaceae y preferiblemente de una planta seleccionada del grupo de las familias Apiaceae, Asteraceae, Brassicaceae, Cucurbitaceae, Fabaceae, Papaveraceae, Rosaceae, Solanaceae, Liliaceae o Poaceae. Se prefieren plantas de cultivo y en particular plantas mencionadas aquí anteriormente como plantas anfitrionas tales como las familias y géneros mencionados anteriormente por ejemplo se prefieren las especies

5 *Anacardium occidentale*, *Calendula officinalis*, *Carthamus tinctorius*, *Cichorium intybus*, *Cynara scolymus*, *Helianthus annuus*, *Tagetes lucida*, *Tagetes erecta*, *Tagetes tenuifolia*; *Daucus carota*; *Corylus avellana*, *Corylus colurna*, *Borago officinalis*; *Brassica napus*, *Brassica rapa* ssp., *Sinapis arvensis* *Brassica juncea*, *Brassica juncea* var. *juncea*, *Brassica juncea* var. *crispifolia*, *Brassica juncea* var. *foliosa*, *Brassica nigra*, *Brassica sinapioides*, *Melanosinapis communis*, *Brassica oleracea*, *Arabidopsis thaliana*, *Anana comosus*, *Ananas ananas*, *Bromelia comosa*, *Carica papaya*, *Cannabis sativa*, *Ipomoea batatas*, *Ipomoea pandurata*, *Convolvulus batatas*, *Convolvulus tiliaceus*, *Ipomoea fastigiata*, *Ipomoea tiliacea*, *Ipomoea triloba*, *Convolvulus panduratus*, *Beta vulgaris*, *Beta vulgaris* var. *altissima*, *Beta vulgaris* var. *vulgaris*, *Beta maritima*, *Beta vulgaris* var. *perennis*, *Beta vulgaris* var. *conditiva*, *Beta vulgaris* var. *esculenta*, *Cucurbita maxima*, *Cucurbita mixta*, *Cucurbita pepo*, *Cucurbita moschata*, *Olea europaea*, *Manihot utilissima*, *Janipha manihot*" *Jatropha manihot*., *Manihot aipil*, *Manihot dulcis*, *Manihot manihot*, *Manihot melanobasis*, *Manihot esculenta*, *Ricinus communis*,

15 *Pisum sativum*, *Pisum arvense*, *Pisum humile*, *Medicago sativa*, *Medicago falcata*, *Medicago varia*, *Glycine max* *Dolichos soja*, *Glycine gracilis*, *Glycine hispida*, *Phaseolus max*, *Soja hispida*, *Soja max*, *Cocos nucifera*, *Pelargonium grossularioides*, *Oleum cocoas*, *Laurus nobilis*, *Persea americana*, *Arachis hypogaea*, *Linum usitatissimum*, *Linum humile*, *Linum austriacum*, *Linum bienne*, *Linum angustifolium*, *Linum catharticum*, *Linum flavum*, *Linum grandiflorum*, *Adenolinum grandiflorum*, *Linum lewisii*, *Linum narbonense*, *Linum perenne*, *Linum perenne* var. *lewisii*, *Linum pratense*,

20 *Linum trigynum*, *Punica granatum*, *Gossypium hirsutum*, *Gossypium arboreum*, *Gossypium barbadense*, *Gossypium herbaceum*, *Gossypium thurberi*, *Musa nana*, *Musa acuminata*, *Musa paradisiaca*, *Musa* spp., *Elaeis guineensis*, *Papaver orientale*, *Papaver rhoeas*, *Papaver dubium*, *Sesamum indicum*, *Piper aduncum*, *Piper amalago*, *Piper angustifolium*, *Piper auritum*, *Piper betel*, *Piper cubeba*, *Piper longum*, *Piper nigrum*, *Piper retrofractum*, *Artanthe adunca*, *Artanthe elongata*, *Peperomia elongata*, *Piper elongatum*, *Steffensia elongata*, , *Hordeum vulgare*, *Hordeum jubatum*, *Hordeum murinum*, *Hordeum secalinum*, *Hordeum distichon* *Hordeum aegiceras*, *Hordeum hexastichon.*, *Hordeum hexastichum*, *Hordeum irregulare*, *Hordeum sativum*, *Hordeum secalinum*, *Avena sativa*, *Avena fatua*, *Avena byzantina*, *Avena fatua* var. *sativa*, *Avena hybrida*, *Sorghum bicolor*, *Sorghum halepense*, *Sorghum saccharatum*, *Sorghum vulgare*, *Andropogon drummondii*, *Holcus bicolor*, *Holcus sorghum*, *Sorghum aethiopicum*, *Sorghum arundinaceum*, *Sorghum caffrorum*, *Sorghum cernuum*, *Sorghum dochna*, *Sorghum drummondii*, *Sorghum durra*,

30 *Sorghum guineense*, *Sorghum lanceolatum*, *Sorghum nervosum*, *Sorghum saccharatum*, *Sorghum subglabrescens*, *Sorghum verticilliflorum*, *Sorghum vulgare*, *Holcus halepensis*, *Sorghum miliaceum* millet, *Panicum militaceum*, *Zea mays*, *Triticum aestivum*, *Triticum durum*, *Triticum turgidum*, *Triticum hybernum*, *Triticum macha*, *Triticum sativum* or *Triticum vulgare*, *Cofea* spp., *Cofea arabica*, *Cofea canephora*, *Cofea liberica*, *Capsicum annuum*, *Capsicum annuum* var. *glabriusculum*, *Capsicum frutescens*, *Capsicum annuum*, *Nicotiana tabacum*, *Solanum tuberosum*, *Solanum melongena*, *Lycopersicon esculentum*, *Lycopersicon lycopersicum.*, *Lycopersicon pyriforme*, *Solanum integrifolium*, *Solanum lycopersicum* *Theobroma cacao* o *Camellia sinensis*.

En una realización, la secuencia de moléculas de ácidos nucleicos para la acumulación y/o producción de nitrógeno o compuestos que contienen nitrógeno se originan ventajosamente de un microorganismo tal como se mencionó anteriormente bajo organismos anfitriones tales como hongos por ejemplo los géneros *Aspergillus*, *Penicillium* o *Claviceps* o de levaduras tales como los géneros *Pichia*, *Torulopsis*, *Hansenula*, *Schizosaccharomyces*, *Candida*, *Rhodotorula* o *Saccharomyces*, muy especialmente ventajosamente de las levaduras de la familia *Saccharomycetaceae*, tales como el género ventajoso *Saccharomyces* y el género y especie muy ventajoso *Saccharomyces cerevisiae*.

40

La persona experimentada conoce otras fuentes adecuadas de ácidos nucleicos que pueden ser utilizados para la acumulación y/o producción de nitrógeno o compuestos que contienen nitrógeno respectivamente. Incluyen en general todas las células procariones o eucariones, preferiblemente microorganismos unicelulares, tales como hongos como el género *Claviceps* o *Aspergillus* o bacterias gram positivas tales como los géneros *Corynebacterium*, *Micrococcus*, *Brevibacterium*, *Rhodococcus*, *Nocardia*, *Caseobacter* o *Arthrobacter* o bacterias gram negativas tales como los géneros *Escherichia*, *Flavobacterium* o *Salmonella*, o levaduras tales como los géneros *Rhodotorula*, *Hansenula* o *Candida*.

50

Sin embargo, también es posible utilizar secuencias artificiales que difieren en una o más bases de las secuencias de ácidos nucleicos encontradas en los organismos, o en una o más moléculas de aminoácidos de secuencias de polipéptidos encontradas en organismos, en particular de las secuencias de polipéptidos indicadas en SEQ ID NO: 690, o los homólogos funcionales de los mismos tales como los descritos aquí, preferiblemente que confieren la actividad antes mencionada, esto es, confieren un incremento de nitrógeno o compuestos que contienen nitrógeno después de incrementar su actividad. En realizaciones preferidas de la presente invención, las secuencias de polipéptidos indicadas en SEQ ID NO: 690 o los homólogos funcionales de los mismos tal como se describen aquí, que confieren preferiblemente la actividad antes mencionada, esto es confieren un incremento de nitrógeno o de compuestos que contienen nitrógeno después de incrementar su actividad, son expresados a partir de secuencias de ácidos nucleicos que han sido optimizadas de acuerdo con el uso del codón en el organismo anfitrión seleccionado u organelo. La persona experimentada en la técnica está familiarizada con fuentes para tablas que muestran el uso del codón preferido para los diferentes aminoácidos en los organismos seleccionados.

60

5 En el proceso de acuerdo con la invención pueden usarse las secuencias de ácidos nucleicos que, si es apropiado, contienen bases de nucleótidos sintéticas, no naturales o modificadas, las cuales pueden ser incorporadas en ADN o ARN. Dichas bases sintéticas, no naturales o modificadas pueden incrementar por ejemplo la estabilidad de la molécula de ácido nucleico por fuera o dentro de una célula. Las moléculas de ácido nucleico de la invención pueden contener las mismas modificaciones mencionadas anteriormente.

10 Tal como se utiliza en el presente contexto el término "molécula de ácido nucleico" puede abarcar también la secuencia no traducida localizada en los extremos 3' y 5' de la región del gen codificante, por ejemplo al menos 500, preferiblemente 200, preferiblemente de manera especial 100, nucleótidos de la secuencia corriente arriba del extremo 5' de la región codificadora y al menos 100, preferiblemente 50, preferible de manera especial 20, nucleótidos de la secuencia corriente abajo del extremo 3' de la región del gen codificador. Frecuentemente es ventajoso escoger solamente la región codificadora para propósitos de clonación y expresión.

Preferiblemente, la molécula de ácidos nucleicos utilizada en el proceso de acuerdo con la invención o la molécula de ácido nucleicos de la invención es una molécula de ácido nucleico aislada.

15 Un polinucleótido o molécula de ácido nucleico "aislados" se separa de otros polinucleótidos o moléculas de ácidos nucleicos que están presentes en la fuente natural de la molécula de ácido nucleico. Una molécula de ácido nucleico aislada puede ser un fragmento cromosómico de varios kb, o preferiblemente, una molécula que comprende solamente la región de codificación del gen.

20 De acuerdo con lo anterior, una molécula de ácido nucleico aislada de la invención puede comprender regiones cromosómicas, las cuales son adyacentes a 5' y 3' o a regiones cromosómicas adyacentes adicionales, pero preferiblemente no comprende tales secuencias las cuales flanquean de manera natural la secuencia de la molécula de ácido nucleico en el contexto genómico o cromosómico en el organismo en el cual se origina la molécula de ácido nucleico (por ejemplo secuencias que son adyacentes a las regiones que codifican el UTR de 5' y 3' de la molécula de ácido nucleico). En diversas realizaciones, la molécula de ácido nucleico aislada usada en el proceso de acuerdo con la invención pueden comprender por ejemplo menos de aproximadamente 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb o 0.1 kb
25 secuencias de nucleótidos que flanquean de manera natural la molécula de ácido nucleico en el ADN genómico de la célula en la cual se origina la molécula de ácido nucleico.

30 Las moléculas de ácido nucleico utilizadas en el proceso, por ejemplo los nucleótidos de la invención o de una parte de los mismos puede aislarse utilizando técnicas estándar de biología molecular, y la información de secuencias provista aquí. También, por ejemplo, pueden identificarse unas secuencias homólogas u homólogas regiones de secuencia conservadas en el nivel de ADN o aminoácidos con la ayuda de algoritmos de comparación. Las primeras pueden utilizarse como sondas de hibridación bajo técnicas de hibridación estándar (por ejemplo las descritas en Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) para aislar secuencias adicionales de ácidos nucleicos útiles en este proceso.

35 La molécula de ácido nucleico que abarca una secuencia completa de las moléculas de ácidos nucleicos usadas en el proceso, por ejemplo el polinucleótido de la invención, o una parte del mismo puede ser aislada adicionalmente por reacción en cadena de polimerasa, cebadores de oligonucleótidos basados en esta secuencia o partes de los mismos que están siendo usados. Por ejemplo, una molécula de ácido nucleico que comprende la secuencia completa o parte de la misma puede ser aislada por la reacción en cadena de polimerasa utilizando cebadores de oligonucleótidos que han sido generados sobre la base de esta secuencia, por ejemplo, puede aislarse ARNm a partir de células (por ejemplo por medio del método de extracción con tiosinato de guanidinio de Chirgwin et al., (1979) *Biochemistry* 18: 5294-5299), y el ADNc puede ser generado por medio de transcriptasa reversa (por ejemplo Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD, o AMV reverse transcriptase, obtainable from Seikagaku America, Inc. St. Petersburg, FL).

45 Los cebadores de oligonucleótidos sintéticos para la amplificación, por ejemplo como los pares indicados en SEQ ID NO: 949 y SEQ ID NO: 950, por medio de la reacción en cadena de polimerasa pueden ser generados sobre la base de una secuencia mostrada aquí, por ejemplo, la secuencia tal como se indica en SEQ ID NO: 689 o las secuencias de ácidos nucleicos derivadas de secuencias de polipéptidos como se indica en SEQ ID NO: 690.

50 Además, es posible identificar regiones conservadas a partir de diversos organismos llevando a cabo alineamientos de secuencias de proteínas con el polipéptido codificado por las moléculas de ácido nucleico de la presente invención, en particular con la secuencias seleccionadas del grupo como se muestra en SEQ ID NO: 690, de la cual pueden derivarse regiones conservadas, y a su vez, cebadores degenerados.

55 Las regiones conservadas son aquellas que muestran una variación muy pequeña en los aminoácidos en una posición particular de varios homólogos de origen diferente. La secuencia de consenso y los motivos de polipéptido mostrados en SEQ ID NO: 951, SEQ ID NO: 952, SEQ ID NO: 953 y SEQ ID NO: 954 se derivan de dichos alineamiento. Además, es posible identificar regiones conservadas de diversos organismos llevando a cabo alineamientos de secuencia de

proteínas con el polipéptido codificado o del ácido nucleico de la presente invención, en particular con las moléculas de polipéptidos mostradas en SEQ ID NO: 690, de las cuales pueden derivarse las regiones conservadas, y a su vez, los cebadores degenerados.

5 En una realización ventajosa, en el método de la presente invención la actividad de un polipéptido se incrementa comprendiendo o consistiendo una secuencia de consenso o un motivo de polipéptido mostrado en la Tabla IV columna 7 y en otra realización, la presente invención se relaciona con un polipéptido que comprende o consiste de una secuencia de consenso o un motivo de polipéptido mostrado en la Tabla IV, columna 7, por lo cual, pueden reemplazarse 20 o menos, preferiblemente 15 o 10, preferiblemente 9, 8, 7 o 6, más preferiblemente 5 o 4, aún más preferiblemente 3, aún más preferiblemente 2, incluso más preferiblemente 1, lo más preferido 0 de las posiciones de aminoácidos indicadas por un aminoácido. En una realización se reemplazan no más de 15%, preferiblemente 10%, aún más preferiblemente 5%, 4%, 3% o 2%, lo más preferido 1% o 0% de la posición de aminoácidos indicada por una letra con otro aminoácido. En una realización se insertan, 20 o menos, preferiblemente 15 o 10, preferiblemente 9, 8, 7, 6, más preferiblemente 5 o 4, aún más preferiblemente 3, incluso más preferiblemente 2, incluso más preferiblemente 1, lo más preferiblemente 0 aminoácidos en una secuencia de consenso o motivo de proteína.

15 La secuencia de consenso fue derivada a partir de un alineamiento múltiple de las secuencias, como aparecen en la lista en las columnas 5 y 7 de la Tabla II. Las letras representan el código de aminoácidos de una letra e indican que los aminoácidos se conservan en todas las proteínas alineadas. La letra X representa los aminoácidos, que no son conservados en todas las secuencias. En un ejemplo, en el caso donde solamente un subgrupo pequeño seleccionado de aminoácidos son posibles en una cierta posición estos aminoácidos se dan entre paréntesis. El número de X dados indican las distancias entre los residuos de aminoácidos conservados, por ejemplo Y – x (21,23)-F significa que los residuos de tirosina y fenilalanina conservados son separados unos de otros por un mínimo de 21 y un máximo de 23 residuos de aminoácidos en todas secuencias investigadas.

25 Los dominios conservados fueron identificados a partir de todas las secuencias y se describen utilizando un subconjunto de la notación estándar Prosite, por ejemplo, el patrón es Y-x (21-23)-(FW) significa que una tirosina conservada separada por mínimo 21 y máximo 23 residuos de aminoácidos bien sea de fenilalanina o triptófano.

30 Los patrones conservados fueron identificados con la herramienta de software MEME versión 3.5.1 o manualmente. El MEME fue desarrollado por Timothy L. Bailey y Charles Elkan, Dept. of Computer Science and Engineering, University of California, San Diego, USA y está descrita por Timothy L. Bailey y Charles Elkan [Fitting a mixture model by expectation maximization to Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994]. El código de fuente para el programa independiente es público disponible en el centro de San Diego Supercomputer (<http://meme.sdsc.edu>).

Para identificar motivos comunes en todas las secuencias con la herramienta del software MEME, se utilizaron los siguientes parámetros: -maxsize 500000, -nmotifs 15, -evt 0.001, -maxw 60, -distance 1e-3, -minsites número de secuencias utilizadas para el análisis. Las secuencias de entrada para MEME no fueron secuencias alineadas en formato Fasta. Se utilizaron otros parámetros en la definición de sistema de esta versión del software.

35 Los patrones Prosite para dominios conservados fueron generados con la herramienta de software Pratt versión 2.1 o manualmente. El Pratt fue desarrollado por Inge Jonassen, Dept. of Informatics, University of Bergen, Norway and is described by Jonassen et al., [I. Jonassen, J.F.Collins and D.G.Higgins, Finding flexible patterns in unaligned protein sequences, Protein Science 4 (1995), pp. 1587-1595; I.Jonassen, Efficient discovery of conserved patterns using a pattern graph, Submitted to CABIOS Febr. 1997]. El código fuente (ANSI C) para el programa individual está disponible públicamente, por ejemplo, en centros bioinformáticos establecidos tales como established Bioinformatic centers like EBI (European Bioinformatics Institute).

45 Para generar patrones con la herramienta de software Pratt, se utilizaron los siguientes parámetros: PL (max Pattern Length): 100, PN (max Nr of Pattern Symbols): 100, PX (max Nr of consecutive x's): 30, FN (max Nr of flexible spacers): 5, FL (max Flexibility): 30, FP (max Flex.Product): 10, ON (max number patterns): 50. Las secuencias de introducción para Pratt fueron regiones distintivas de las secuencias de proteínas que exhiben alta similitud tal como fueron identificadas a partir de la herramienta de software MEME. El número mínimo de secuencias, que tiene que coincidir con los patrones generados (CM, min Nr of Seqs to Match), fue definido en al menos 80% de las secuencias provistas. Los parámetros no mencionados aquí fueron utilizados con sus valores por sistema.

50 Los patrones Prosite de los dominios conservados pueden utilizarse para investigar secuencias de proteínas que coinciden con este patrón. Diversos centros bioinformáticos establecidos proveen portales de internet públicos para utilizar estos patrones en búsquedas en bases de datos (por ejemplo PIR [Protein Information Resource, located at Georgetown University Medical Center] or ExpASy [Expert Protein Analysis System]). Alternativamente, hay disponible softwares individuales, como el programa Fuzzpro el cual es parte del paquete de software EMBOSS. Por ejemplo, el programa Fuzzpro no solamente permite buscar una coincidencia de proteína exacta al patrón sino que también permite definir ambigüedades diversas en la búsqueda ejecutada.

5 El alineamiento fue llevado a cabo con el software ClustalW (versión 1.83) y fue descrito por Thompson et al., [Thompson, J.D., Higgins, D.G. y Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680]. El código fuente para el programa individual está disponible públicamente en la European Molecular Biology Laboratory; Heidelberg, Alemania. El análisis fue llevado a cabo utilizando los parámetros por sistema de ClustalW v1.83 (gap open penalty: 10.0; gap extension penalty: 0.2; protein matrix: Gonnet; pprotein/DNA endgap: -1; protein/DNA gapdist: 4).

10 Los cebadores degenerados pueden ser utilizados entonces por PCR para la amplificación de fragmentos de proteínas novedosas que tiene actividad antes mencionada, por ejemplo, confiriendo el incremento de nitrógeno o compuestos que contienen nitrógeno después de incrementar la expresión o actividad u homólogos funcionales adicionales del polipéptido de la invención o el polipéptido utilizado en el método de la invención de otros organismos.

15 Estos fragmentos pueden ser utilizados entonces como sondas de hibridación para aislar las secuencias genéticas completas. Como alternativa, las secuencias faltantes 5' y 3' pueden ser aisladas por medio de RACE-PCR (amplificación rápida de extremos de ADNc). Una molécula de ácido nucleico de acuerdo con la invención puede ser amplificada utilizando ADNc o, como alternativa, ADN genómico como formato y cebadores de oligonucleótidos adecuados, siguiendo técnicas de amplificación por PCR estándar. La molécula de ácido nucleico amplificada así puede ser clonada en un vector adecuado y caracterizada por medio de análisis de secuencia de ADN. Los oligonucleótidos, los cuales corresponden a una de las moléculas de ácido nucleico utilizadas en el proceso, pueden ser generados por métodos de síntesis estándar, por ejemplo utilizando un sintetizador de ADN automático.

20 Las moléculas de ácido nucleico que son ventajosas para el proceso de acuerdo con la invención pueden aislarse con base en su homología con las moléculas de ácido nucleico divulgadas aquí utilizando las secuencias o parte de las mismas como sonda de hibridación y siguiendo técnicas de hibridación estándar bajo condiciones de hibridación respectivas. En este contexto, es posible usar, por ejemplo, moléculas de ácido nucleico aisladas de al menos 15, 20, 25, 30, 35, 40, 50, 60 o más nucleótidos, preferiblemente de al menos 15, 20 o 25 nucleótidos de longitud las cuales
25 hibridizan bajo condiciones restrictivas con las moléculas de ácido nucleico descritas anteriormente, en particular con aquellas que abarcan una secuencia de nucleótidos de la molécula de ácido nucleico usada en el proceso de la invención o codificar una proteína utilizada en la invención o de la molécula de ácido nucleico de la invención. También pueden usarse moléculas de ácido nucleico con 30, 50, 100, 250 o más nucleótidos.

30 El término "homología" significa que las moléculas de ácido nucleico o proteínas codificadas respectivas son funcional y/o estructuralmente equivalentes. Las moléculas de ácido nucleico que son homólogas a las moléculas de ácido nucleico descritas anteriormente que son derivados de dichas moléculas de ácido nucleico son, por ejemplo, variaciones de dichas moléculas de ácido nucleico que representan modificaciones que tienen la misma función biológica, en particular la codificación de proteínas con la misma o sustancialmente la misma función biológica. Pueden ser variaciones de origen natural, tales como las secuencias de otras variedades o especies de plantas, o mutaciones.
35 Estas mutaciones pueden ocurrir de forma natural o pueden ser obtenidas por técnicas de mutagénesis. Las variaciones alélicas pueden ser variantes alélicas de origen natural así como variantes producidas sintéticamente o manipuladas genéticamente. Los equivalentes estructurales pueden ser identificados, por ejemplo, probando el enlazamiento de dicho polipéptido a anticuerpos o predicciones basadas en ordenador. Las equivalentes estructurales tienen características inmunológicas similares, por ejemplo, comprenden epitopos similares.

40 Por "hibridar" se entiende que tales moléculas de ácido nucleico hibridan bajo condiciones de hibridación convencionales, preferiblemente bajo condiciones restrictivas tales como las descritas por, por ejemplo, Sambrook (*Molecular Cloning; A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)) o en *Current Protocols in Molecular Biology*, John Wiley & Sons, N. Y. (1989), 6.3.1-6.3.6.

45 De acuerdo con la invención, las moléculas de ADN así como de ARN del ácido nucleico de la invención pueden usarse como sondas. Además, como patrón para la identificación de homólogos funcionales pueden llevarse a cabo ensayos de inmunoprecipitación Northern así como Southern. La prueba de inmunoprecipitación Northern proveen ventajosamente información adicional acerca del producto genético expresado: por ejemplo, patrón de expresión, presencia de etapas del proceso, tales como división y taponamiento, etc. La prueba de inmunoprecipitación Southern provee información adicional acerca de la localización cromosómica y organización del gen que codifica la molécula de
50 ácido nucleico de la invención.

Un ejemplo preferido no limitante de condiciones de hibridación restrictivas son la hibridación en cloruro de sodio/citrato de sodio 6 x (= SSC) a aproximadamente 45°C, seguida por una o más etapas de lavado en 0.2 x SSC, SDS al 0.1% a 50 hasta 65°C, por ejemplo a 50°C, 55°C o 60°C. La persona experimentada sabe que estas condiciones de hibridación difieren como función del tipo de ácido nucleico y, por ejemplo, cuando hay presentes solventes orgánicos, con respecto
55 a la temperatura y concentración del regulador. La temperatura bajo "condiciones de hibridación estándar" difiere por ejemplo como función del tipo de ácido nucleico entre 42°C y 58°C, preferiblemente entre 45°C y 50°C en un regulador acuoso con una concentración de 0.1 x, 0.5 x, 1x, 2x, 3x, 4x o 5x SSC (pH 7.2). Si hay solventes orgánicos presentes en el regulador antes mencionado, por ejemplo formamida al 50%, la temperatura bajo condiciones estándar es de

aproximadamente 40°C, 42°C o 45°C. Las condiciones de hibridación para híbridos ADN: ADN son preferiblemente por ejemplo 0.1 x SSC y 20°C, 25°C, 30°C, 35°C, 40°C o 45°C, preferiblemente entre 30°C y 45°C. Las condiciones de hibridación para híbridos de ADN: ARN son preferiblemente por ejemplo 0.1 x SSC y 30°C, 35°C, 40°C, 45°C, 50°C o 55°C, preferiblemente entre 45°C y 55°C. Las temperaturas de hibridación antes mencionadas son determinadas por ejemplo para un ácido nucleico de aproximadamente 100 bp (= pares de bases) de longitud y un contenido G + C de 50% en la ausencia de formamida. La persona experimentada sabe determinar las conexiones de hibridación requeridas con la ayuda de manuales de texto, por ejemplo los mencionados anteriormente, o de los siguientes libros de texto: Sambrook et al., "Molecular Cloning", Cold Spring Harbor Laboratory, 1989; Hames and Higgins (Ed.) 1985, "Nucleic Acids Hybridization: A Practical Approach", IRL Press at Oxford University Press, Oxford; Brown (Ed.) 1991, "Essential Molecular Biology: A Practical Approach", IRL Press at Oxford University Press, Oxford.

Un ejemplo adicional de una tal condición de hibridación restrictiva es la hibridación a 4XSSC a 65°C, seguida por un lavado en 0.1 XSSC a 65°C durante una hora. Alternativamente, una condición de hibridación restrictiva de ejemplo es en formamida al 50%, 4XSSC a 42°C. Adicionalmente, las condiciones durante la etapa de lavado pueden ser seleccionadas del rango de condiciones delimitadas por las condiciones de baja restricción (aproximadamente 2X SSC a 50°C) y condiciones de alta restricción (aproximadamente 0.2X SSC a 50°C, preferiblemente a 65°C) (20X SSC: citrato de sodio 0.3M, NaCl 3M, pH 7.0). Además, la temperatura durante la etapa de lavado puede ser elevada desde condiciones de baja restricción a temperatura ambiente, aproximadamente 22°C, a condiciones de alta restricción a aproximadamente 65°C. Los parámetros de concentración de sal y temperatura pueden variarse simultáneamente, o uno de los dos parámetros puede ser mantenido constante mientras que el otro se varía. Los desnaturizantes, por ejemplo formamida o SDS también, pueden ser empleados durante la hibridación. En la presencia de formamida al 50%, la hibridación se efectúa preferiblemente a 42°C. Factores relevantes tales como i) longitud del tratamiento, ii) condiciones de sal, iii) condiciones de detergente, iv) ADNs competidores, v) temperatura y vi) selección de sonda pueden ser combinados caso a caso de tal manera que no todas las posibilidades pueden ser mencionadas aquí.

Así, en una realización preferida, las inmunoprecipitaciones Northern son prehibridadas con el regulador Rothi-Hybr-Quick (Roth, Karlsruhe) a 68°C durante 2 horas. La hibridación con una sonda radioactiva marcada se hace durante la noche a 68°C. Las etapas de lavado subsecuentes se llevan a cabo a 68°C con 1xSSC. Para las pruebas de inmunoprecipitación Southern la membrana se prehibridiza con regulador Rothi-Hybr-Quick (Roth, Karlsruhe) a 68°C durante 2 horas. La hibridación con la sonda marcada radioactiva se lleva a cabo durante la noche a 68°C. Subsecuentemente se descarta el regulador de hibridación y el filtro se lava rápidamente utilizando 2xSSC; SDS al 0.1%. Después de descartar el regulador de lavado se agrega nuevo regulador 2xSSC; SDS al 0.1% y se incuba a 68°C durante 15 minutos. Esta etapa de lavado se lleva a cabo dos veces seguida por una etapa de lavado adicional utilizando 1xSSC; SDS al 0.1% a 68°C durante 10 minutos.

Algunos ejemplos adicionales de condiciones para hibridación de ADN (pruebas de inmunoprecipitación Southern) y etapas de lavado se muestran aquí a continuación:

1. Las condiciones de hibridación pueden ser seleccionadas, por ejemplo, a partir de las siguientes condiciones:

a) 4X SSC a 65°C,

b) 6X SSC a 45°C,

c) 6X SSC, 100 mg/ml fragmento desnaturizado de ADN de esperma de pez a 68°C,

d) 6X SSC, 0.5% SDS, 100 mg/ml ADN de esperma desnaturizado de salmón a 68°C,

e) 6X SSC, 0.5% SDS, 100 mg/ml fragmento desnaturizado de ADN de esperma de salmón, 50% formamida a 42°C,

f) 50% formamida, 4X SSC at 42°C,

g) 50% (vol/vol) formamida, albúmina de suero bovino al 0.1%, Ficoll al 0.1%, polivinilpirrolidona al 0.1%, regulador de fosfato de sodio 50 mM pH 6.5, NaCl 750 mM, citrato de sodio 75 mM a 42°C,

h) 2X o 4X SSC a 50°C (condiciones de baja restricción), o

i) formamida al 30 a 40%, 2X o 4X SSC a 42°C (condiciones de baja restricción).

2. Las etapas de lavado pueden seleccionarse, por ejemplo, entre las siguientes condiciones:

a) 0.015 M NaCl/0.0015 M citrato de sodio/0.1 % SDS at 50°C.

b) 0.1X SSC a 65°C.

c) 0.1X SSC, 0.5 % SDS a 68°C.

d) 0.1X SSC, 0.5% SDS, 50% formamida a 42°C.

5 e) 0.2X SSC, 0.1% SDS a 42°C.

f) 2X SSC a 65°C (condiciones de baja restricción).

Además, algunas aplicaciones tienen que llevarse a cabo en condiciones de hibridación con baja restricción, sin consecuencias para la especificidad de la hibridación. Por ejemplo un análisis de inmunoprecipitación Southern de ADN total podría ser sondeado con una molécula de ácido nucleico de la presente invención y lavado con baja restricción (55°C in 2xSSPE 0.1% SDS). El análisis de hibridación podría revelar un patrón simple de genes que solamente codifican polipéptidos de la presente invención o se utilizan en el proceso de la invención, por ejemplo que tienen la actividad aquí mencionada de incrementar el nitrógeno o los compuestos que contienen nitrógeno. Un ejemplo adicional de tales condiciones de hibridación de baja restricción es 4XSSC a 50°C o hibridación con 30 a 40% de formamida a 42°C. Tales moléculas comprenden aquellas que son fragmentos, análogos o derivados del polipéptido de la invención o se utilizan en el proceso de la invención y difieren, por ejemplo, por la forma de la eliminación, inserción, sustitución, adición y/o recombinación de aminoácidos o cualquier otra modificación conocida en la técnica sola o en combinación con las secuencias de aminoácidos descritos anteriormente o su secuencia de nucleótido subyacente. Sin embargo, se prefiere utilizar condiciones de hibridación de alta astringencia.

La hibridación debería ser llevada a cabo ventajosamente con fragmentos de al menos 5, 10, 15, 20, 25, 30, 35, o 40 bp, ventajosamente al menos 50, 60, 70 u 80 bp, preferiblemente al menos 90, 100 o 110 bp. Lo más preferiblemente son fragmentos de al menos 15, 20, 25 o 30 bp. Preferiblemente también son hibridaciones con al menos 100 bp o 200, muy especialmente preferiblemente al menos 400 bp de longitud. En una realización especialmente preferida, la hibridación debería ser llevada a cabo con la secuencia completa de ácido nucleico con las condiciones descritas anteriormente.

Los términos "fragmento", "fragmento de una secuencia" o "parte de una secuencia" significan una secuencia truncada de la secuencia original a la que se ha hecho referencia. La secuencia truncada (secuencia de ácidos nucleicos o proteínas) puede variar ampliamente en longitud; el tamaño mínimo para hacer una secuencia de tamaño suficiente para proveer una secuencia con al menos una función comparable y/o actividad de la secuencia original referida o para hibridar con la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención o usada en el proceso de la invención bajo condiciones restrictivas, mientras que el máximo tamaño no es crítico. En algunas aplicaciones, el tamaño máximo usualmente no es sustancialmente superior que el requerido para proveer la actividad y/o funciones deseadas de la secuencia original.

Típicamente, la secuencia de aminoácidos truncada variará de aproximadamente 5 hasta aproximadamente 310 aminoácidos de longitud. Más típicamente, sin embargo, la secuencia será un máximo de aproximadamente 250 aminoácidos de longitud, preferiblemente un máximo de aproximadamente 200 o 100 aminoácidos. Usualmente es deseable seleccionar secuencia de al menos aproximadamente 10, 12 o 15 aminoácidos, hasta un máximo de aproximadamente 20 a 25 aminoácidos.

El término "epítipo" se relaciona con sitios inmunorreactivos específicos con un antígeno, también conocidos como determinados antigénicos. Estos epítipos pueden ser una disposición lineal de monómeros en una composición polimérica – tal como aminoácidos en una proteína - o consisten de o comprenden una estructura secundaria o terciaria más compleja. Las personas experimentadas reconocerán que los inmunógenos (esto es, sustancias capaces de disparar una respuesta inmune) son antígenos; sin embargo, algunos antígenos, tales como los haptenos, no son inmunógenos pero pueden ser hechos inmunogénicos acoplándolos con una molécula portadora. El término "antígeno" incluye referencias a una sustancia a la cual pueden generarse un anticuerpo y/o hacia la cual el anticuerpo es específicamente inmunorreactivo.

En una realización la presente invención se relaciona con un epítipo del polipéptido de la presente invención o usado en el proceso de la presente invención y que confiere la actividad antes mencionada, preferiblemente que confiere un incremento en nitrógeno o en compuestos que contienen nitrógeno.

El término "uno o varios aminoácidos" se relaciona con al menos un aminoácido pero no más que el número de aminoácidos, los cuales podrían resultar en una homología de menos de 50% de identidad. Preferiblemente, la identidad es más de 70% u 80%, más preferida 85%, 90%, 91%, 92%, 93%, 94% o 95%, incluso más preferida 96%, 97%, 98% o 99% de identidad.

Adicionalmente, la molécula de ácido nucleico de la invención comprende una molécula de ácido nucleico que es un complemento de una de las secuencias de nucleótidos de las moléculas de ácido nucleico antes mencionadas o una porción de las mismas. Una molécula de ácido nucleico que es complementaria a una de las secuencias de nucleótidos indicadas en SEQ ID NO: 689, es una que es suficientemente complementaria a una de dichas secuencias de nucleótidos de tal forma que pueden hibridar a una de dichas secuencias de nucleótidos formando por lo tanto un dúplex estable.

Preferiblemente, la hibridación se lleva a cabo bajo condiciones de hibridación restrictivas. Sin embargo, un complemento de una de las secuencias aquí divulgadas es preferiblemente un complemento de secuencia de la misma de acuerdo con el apareamiento de bases de las moléculas de ácido nucleico bien conocidos para una persona experimentada en la técnica. Por ejemplo, las bases A y G sufren apareamiento de bases con la bases T y U o C, respectivamente, y viceversa. Las modificaciones de las bases pueden influir en el asociado de apareamiento de bases.

La molécula de ácido nucleico de la invención comprende una secuencia de nucleótidos que es al menos 70% o más homóloga a una secuencia de nucleótidos indicada en SEQ ID NO: 689 y preferiblemente tiene la actividad antes mencionada, en particular tiene la actividad de incrementar el nitrógeno o los compuestos que contienen nitrógeno después de incrementar su actividad o una actividad de un producto de un gen que codifica dicha secuencia o sus homólogos.

Las moléculas de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención comprende una secuencia de nucleótidos que hibrida, preferiblemente hibrida bajo condiciones restrictivas como se define aquí, a una de las secuencias de nucleótidos indicadas en SEQ ID NO: 689, o una porción de la misma y codifica una proteína que tiene la actividad antes mencionada y como se indica en SEQ ID NO: 690, confiriendo un incremento de nitrógeno o compuestos que contienen nitrógeno.

Opcionalmente, la secuencia de nucleótidos, la cual hibrida a una de las secuencias de nucleótidos indicadas en SEQ ID NO: 689 tiene adicionalmente una o más de las actividades anotadas o conocidas para la proteína YPR138C.

Además, la molécula de ácido nucleico de la invención o usada en el proceso de la invención puede comprender solamente una porción de la región de codificación de una de las secuencias indicadas en SEQ ID NO: 689, por ejemplo un fragmento que puede ser usado como sonda o cebador o un fragmento que codifica una porción biológicamente activa del polipéptido de la presente invención o de un polipéptido usado en el proceso de la presente invención, esto es que tiene la actividad antes mencionada, por ejemplo el conferir un contenido incrementado de nitrógeno o de compuestos que contienen nitrógeno si se incrementa su actividad. Las secuencias de nucleótidos determinadas a partir de la clonación del gen de la proteína presente de acuerdo con la invención permite la generación de sondas y cebadores diseñados para uso en la identificación y/o clonación de sus homólogos en otros tipos de células y organismos. La sonda/cebador comprende típicamente oligonucleótidos purificados sustancialmente. El oligonucleótido comprende sustancialmente una región de secuencia de nucleótido que hibridiza bajo condiciones de restricción a al menos aproximadamente 12, 15, preferiblemente de forma aproximadamente 20 o 25, más preferiblemente de forma aproximada 40, 50, o 75 nucleótidos consecutivos de una cadena sentido de una de las secuencias indicadas en la Tabla I, columna 5 o 7 una secuencia antisentido de una de las secuencias indicadas en Tabla I, columnas 5 o 7, o de mutantes de los mismos de origen natural.

Los cebadores basados en una secuencia de nucleótidos de la invención pueden ser utilizados en reacciones de PCR para clonar homólogos del polipéptido de la invención o del polipéptido usado en el proceso de la invención, por ejemplo, como los cebadores descritos en los ejemplos de la presente invención, por ejemplo como se muestra en los ejemplos. Un PCR con los pares del cebador indicados en SEQ ID NO: 949 y SEQ ID NO: 950, da como resultado un fragmento de una secuencia de polinucleótidos como se indica en SEQ ID NO: 689.

Los conjuntos de cebadores son intercambiables. La persona experimentada en la técnica sabe combinar dichos cebadores para dar como resultado el producto deseado, por ejemplo, en un clon de longitud completa o una secuencia parcial. Las sondas basadas en las secuencias de la molécula de ácido nucleico de la invención o usada en el proceso de la presente invención pueden ser utilizadas para detectar transcritos o secuencias genómicas que codifican las mismas proteína u homólogas. La sonda puede comprender adicionalmente un grupo de marcación unido a la misma, por ejemplo, el grupo de marcación puede ser un radioisótopo como un compuesto fluorescente, una enzima o un cofactor de enzima. Tales sondas pueden ser utilizadas como parte de un kit de prueba de marcador genómico para identificar células que expresan un polipéptido de la invención o se utilizan en el proceso de la presente invención, tal como mediante la medición de un nivel de una molécula de ácido nucleico codificadora en una muestra de células, por ejemplo, detectando niveles de ARNm o determinando, si un gen genómico que comprende la secuencia del polinucleótido de la invención utilizado en el proceso de la presente invención ha sido sometido a mutación o eliminado.

La molécula de ácido nucleico de la invención o la molécula de ácido nucleico usado en el método de la invención codifica un polipéptido o porción del mismo que incluye una secuencia de aminoácidos que es suficientemente homóloga a una secuencia de aminoácidos tal como se indica en SEQ ID NO: 690 de tal manera que la proteína o porción de la misma mantiene la capacidad de participar la acumulación y/o producción de nitrógeno o compuestos que

contienen nitrógeno respectivamente, en particular comprende un contenido de proteína que incrementa la actividad como se mencionó anteriormente o como se describe en los ejemplos en plantas o microorganismos.

5 Tal como se utiliza aquí, la expresión “suficientemente homologas” se refiere a proteínas o porciones de las mismas que tienen secuencias de aminoácidos que incluyen un número mínimo de residuos de aminoácidos idénticos o equivalentes (por ejemplo, un residuo de aminoácidos que tiene una cadena lateral similar que un residuo de aminoácidos en una de las secuencias del polipéptido de la presente invención) a una secuencia de aminoácidos se indica en SEQ ID NO: 690, de tal manera que la proteína o porciones de la misma son capaces de participar en la acumulación y/o producción de nitrógeno o compuestos que contienen nitrógeno respectivamente. En una realización, una proteína o porción de la misma tal como la que se indica en SEQ ID NO: 690, tiene por ejemplo una actividad de un polipéptido indicado en
10 YPR138C.

En una realización, la molécula de ácido nucleico de la presente invención comprende un ácido nucleico que codifica una porción de la proteína de la presente invención. La proteína es al menos 70% o más idéntica a una secuencia de aminoácidos completa como se indica en la Tabla II, columnas 5 o 7, y tiene la actividad antes mencionada, por ejemplo el conferir preferiblemente el incremento de nitrógeno o compuestos que contienen nitrógeno.

15 Las porciones de proteína codificadas por la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención son preferiblemente activas biológicamente, preferiblemente tienen la actividad antes anotada, por ejemplo conferir un incremento en nitrógeno o en compuestos que contienen nitrógeno después de un incremento en la actividad.

20 Como se menciona aquí, el término “porción biológicamente activa” pretende incluir una porción, por ejemplo un dominio/motivo que confiere el incremento de nitrógeno o compuestos que contienen nitrógeno o tiene una actividad inmunológica tal que se enlaza a un enlazamiento de un anticuerpo específicamente para el polipéptido de la presente invención o un polipéptido utilizado en el proceso de la presente invención para producir nitrógeno o compuestos que contienen nitrógeno.

25 Además, será evidente para las personas experimentadas en la técnica que el polimorfismo de la secuencia de ADN que lleva cambios en las secuencias de aminoácido puede existir dentro de una población. Tal polimorfismo genético en el gen de codifica el polipéptido de la invención o el polipéptido utilizado en el método de la invención o que comprende la molécula de ácido nucleico de la invención o la molécula de ácido nucleico utilizado en el método de la invención pueden existir entre individuos de una población debido a una variación natural.

30 Tal como se utilizan aquí, los términos “gen” y “gen recombinante” se refieren a moléculas de ácidos nucleicos que comprenden un marco de lectura abierto que codifica el polipéptido de la invención o el polipéptido utilizado en el método de la invención o que comprende la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención o codifican el polipéptido usado en el proceso de la presente invención, preferiblemente de una planta de cultivo o de un microorganismo útil para la acumulación y/o producción de nitrógeno o compuestos que contienen nitrógeno, respectivamente, en particular para la producción de proteínas. Tales variaciones naturales pueden resultar lógicamente en variación del 1 - 5% en la secuencia del nucleótido del gen. Cualquier otra
35 otra variación en los nucleótidos y polímeros resultantes en aminoácidos en los genes que codifican un polipéptido de la invención o el polipéptido usado en el método de la invención o que comprende una molécula de ácido nucleico de la invención o de la molécula de ácido nucleico usada en el método de la invención que son el resultado de la variación natural y que no alteran su actividad funcional se describen con la intención de que se sitúen dentro del alcance de la
40 invención.

Las moléculas de ácido nucleico correspondientes a variantes homólogas de una molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención, que también puede ser un ADNc, pueden aislarse con base en su homología a las moléculas de ácido nucleico divulgadas aquí utilizando la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención, o una porción de las
45 mismas, tal como una sonda de hibridación de acuerdo con técnicas de hibridación estándar bajo condiciones de hibridación estándar.

De acuerdo con lo anterior, en otro ambiente, una molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención tiene al menos 15, 20, 25 o 30 nucleótidos de longitud preferiblemente, hibridiza bajo condiciones astringentes restrictivas a una molécula de ácido nucleico que comprende una secuencia de nucleótidos de la molécula de ácido nucleico de la presente invención o se utilizan en el proceso de la presente invención, por ejemplo, comprendiendo una secuencia tal como se indica en SEQ ID NO: 689. La molécula de ácido nucleico es preferiblemente al menos 20, 30, 50, 100, 250 o más nucleótidos de longitud.

El término “hibrido bajo condiciones restrictivas” se define anteriormente. En una realización, el término “hibrido bajo condiciones restrictivas” pretende describir condiciones para hibridación y lavado a las cuales las secuencias de nucleótidos a 70°C al menos, o más idéntico típicamente permanecen hibridizados entre sí.
55

Preferiblemente, la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usado en los métodos de la invención que hibridiza bajo condiciones restrictivas a una secuencia tal como la que se indica en SEQ ID NO: 689 corresponde a la molécula de ácido nucleico de origen natural de la invención. Tal como se utiliza aquí, una molécula de ácido nucleico “que se presenta de manera natural” se refiere a una molécula de ARN o ADN que tiene una secuencia de nucleótidos que se presenta en la naturaleza (por ejemplo codifica una proteína natural). Preferiblemente la molécula de ácido nucleico codifica una proteína natural que tiene la actividad antes mencionada esto es conferir nitrógeno o compuestos que contienen nitrógeno e incrementar la eficiencia del uso de nitrógeno (NUE), definida como el rendimiento de biomasa por unidad de nitrógeno disponible del suelo, incluyendo el incremento en fertilizantes después de incrementar la expresión o actividad del mismo o la actividad de una rutina de la invención o se usa en el proceso de la invención.

Además de las variantes de origen natural de las secuencias de los polipéptidos o moléculas de ácido nucleico de la invención así como del polipéptido o de la molécula de ácido nucleico utilizado en el proceso de la invención que puede existir en la población, el técnico experimentado apreciará adicionalmente que pueden introducirse cambios por mutación en una secuencia de nucleótidos de la molécula de ácido nucleico que codifica el polipéptido de la invención usadas en el proceso de la presente invención, llevando por lo tanto a cambios en la secuencia de aminoácidos de dicho polipéptido codificado, sin alterar la capacidad funcional del polipéptido, preferiblemente sin disminuir dicha actividad.

Por ejemplo, las sustituciones de nucleótidos que llevan a sustituciones de aminoácidos en residuos de aminoácidos “no esenciales” pueden hacerse en una secuencia de la molécula de ácido nucleico de la invención a usarse en el proceso de la invención, por ejemplo, como se indica en SEQ ID NO: 689.

Un residuo de aminoácido “no esencial” es un residuo que puede ser alterado de la secuencia tipo silvestre sin alterar la actividad de dicho polipéptido, mientras que un residuo de aminoácido “esencial” se requiere para una actividad tal como se mencionó anteriormente, por ejemplo, llevar a un incremento en nitrógeno o en compuestos que contienen nitrógeno y eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno en un organismo después de un incremento en la actividad del polipéptido. Otros residuos de aminoácidos, sin embargo (por ejemplo aquellos que no son conservados o solamente semiconservados en el dominio que tiene dicha actividad) pueden no ser esenciales para la actividad y así probablemente serán adecuados para alteración sin alterar dicha actividad.

Adicionalmente, una persona experimentada en la técnica sabe que el uso de codón entre organismo puede diferir. Por ejemplo, puede adaptar el uso de codón en la molécula de ácido nucleico de la presente invención al uso en el organismo en el cual el polinucleótido o polipéptido se expresan.

De acuerdo con lo anterior, la invención se relaciona con moléculas de ácido nucleico que codifican un polipéptido que tiene la actividad antes mencionada, por ejemplo, conferir un incremento en nitrógeno o en compuestos que contienen nitrógeno y una eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno en un organismo o partes del mismo que contienen cambios en residuos de aminoácidos que no son esenciales para dicha actividad. Tales polipéptidos difieren en la secuencia de aminoácidos de una secuencia contenida en una secuencia según se indica en SEQ ID NO: 690, ya retenían dicha actividad descrita aquí. La molécula de ácido nucleico puede comprender una secuencia de nucleótido que codifica un polipéptido, en donde el polipéptido comprende una secuencia de aminoácidos de al menos 70% o más idéntica a una secuencia de aminoácidos como la indicada en SEQ ID NO: 690, y es capaz de participar en la acumulación y/o producción incrementadas de nitrógeno o compuestos que contienen nitrógeno y eficiencia incrementada en el uso del nitrógeno (NUE), definida como el rendimiento de biomasa por unidad de nitrógeno que tiene compuestos y eficiencia incrementada del uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno respectivamente después de incrementar su actividad, por ejemplo su expresión.

Para determinar el porcentaje de homología (= en identidad) de dos secuencias de aminoácidos o de dos moléculas de ácido nucleico, las secuencias se escriben una bajo la otra para una comparación óptima (por ejemplo pueden insertarse brechas en la secuencia de una proteína o de un ácido nucleico con el fin de generar un alineamiento óptico con la otra proteína o el otro ácido nucleico).

Los residuos de aminoácidos o moléculas de ácido nucleico en las posiciones de aminoácidos o posiciones de nucleótidos correspondientes se comparan entonces. Si una posición en una secuencia es ocupada por el mismo residuo de aminoácido en la misma molécula de ácido nucleico en la posición correspondiente en la otra secuencia, las moléculas son homólogas en esta posición (esto es, la “homología” de aminoácidos o ácidos nucleicos tal como se utiliza en el presente contexto corresponden a “identidad” de aminoácidos o ácidos nucleicos. El porcentaje de homología entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (esto es, % de homología = al número de posiciones idénticas/número total de posiciones x 100). Los términos “homología” y “identidad” son así considerados como sinónimos.

Para la determinación del porcentaje de homología (= identidad) de dos o más aminoácidos o de dos o más secuencias de nucleótidos se han desarrollado varios programas de software para ordenador. La homología de dos o más secuencias pueden calcularse con, por ejemplo, el software Fasta el cual actualmente ha sido utilizado en la versión Fasta 3 (W.R. Pearson and D. J. Lipman (1988). Improved Tools for Biological Sequence Comparison. PNAS 85: 2444-2448; W.R. Pearson (1990) Rapid and Sensitive Sequence Comparison with FASTP and FASTA, Methods in Enzymology 183:63 - 98. Otro programa útil para el cálculo de homologías de diferentes secuencias es el programa estándar blast, el cual está incluido en el software Biomax pedant (Biomax, Munich, República Federal de Alemania). Esto lleva algunas veces infortunadamente a resultados no óptimos puesto que el blast no siempre incluye secuencias completas del sujeto y la pregunta. No obstante puesto que este programa es muy eficiente puede ser utilizado para la comparación de un enorme número de secuencias. Se utilizan los siguientes parámetros típicamente para tales comparaciones de secuencias:

-p Program Name [String]; -d Database [String]; default = nr; -i Query File [File In]; default = stdin; -e Expectation value (E) [Real]; default = 10.0; -m alignment view options: 0 = pairwise; 1 = query-anchored showing identities; 2 = query-anchored no identities; 3 = flat query-anchored, show identities; 4 = flat query-anchored, no identities; 5 = query-anchored no identities and blunt ends; 6 = flat query-anchored, no identities and blunt ends; 7 = XML Blast output; 8 = tabular; 9 tabular with comment lines [Integer]; default = 0; -o BLAST report Output File [File Out] Optional; default = stdout; -F Filter query sequence (DUST with blastn, SEG with others) [String]; default = T; -G Cost to open a gap (zero invokes default behavior) [Integer]; default = 0; -E Cost to extend a gap (zero invokes default behavior) [Integer]; default = 0; -X X dropoff value for gapped alignment (in bits) (zero invokes default behavior); blastn 30, megablast 20, tblastx 0, all others 15 [Integer]; default = 0; -I Show GI's in defines [T/F]; default = F; -q Penalty for a nucleotide mismatch (blastn only) [Integer]; default = -3; -r Reward for a nucleotide match (blastn only) [Integer]; default = 1; -v Number of database sequences to show one-line descriptions for (V) [Integer]; default = 500; -b Number of database sequence to show alignments for (B) [Integer]; default = 250; -f Threshold for extending hits, default if zero; blastp 11, blastn 0, blastx 12, tblastn 13; tblastx 13, megablast 0 [Integer]; default = 0; -g Perform gapped alignment (not available with tblastx) [T/F]; default = T; -Q Query Genetic code to use [Integer]; default = 1; -D DB Genetic code (for tblast[nx] only) [Integer]; default = 1; -a Number of processors to use [Integer]; default = 1; -O SeqAlign file [File Out] Optional; -J Believe the query define [T/F]; default = F; -M Matrix [String]; default = BLOSUM62; -W Word size, default if zero (blastn 11, megablast 28, all others 3) [Integer]; default = 0; -z Effective length of the database (use zero for the real size) [Real]; default = 0; -K Number of best hits from a region to keep (off by default, if used a value of 100 is recommended) [Integer]; default = 0; -P 0 for multiple hit, 1 for single hit [Integer]; default = 0; -Y Effective length of the search space (use zero for the real size) [Real]; default = 0; -S Query strands to search against database (for blast[nx], and tblastx); 3 is both, 1 is top, 2 is bottom [Integer]; default = 3; -T Produce HTML output [T/F]; default = F; -I Restrict search of database to list of GI's [String] Optional; -U Use lower case filtering of FASTA sequence [T/F] Optional; default = F; -y X dropoff value for ungapped extensions in bits (0.0 invokes default behavior); blastn 20, megablast 10, all others 7 [Real]; default = 0.0; -Z X dropoff value for final gapped alignment in bits (0.0 invokes default behavior); blastn/megablast 50, tblastx 0, all others 25 [Integer]; default = 0; -R PSI-TBLASTN checkpoint file [File In] Optional; -n MegaBlast search [T/F]; default = F; -L Location on query sequence [String] Optional; -A Multiple Hits window size, default if zero (blastn/megablast 0, all others 40 [Integer]; default = 0; -w Frame shift penalty (OOF algorithm for blastx) [Integer]; default = 0; -t Length of the largest intron allowed in tblastn for linking HSPs (0 disables linking) [Integer]; default = 0.

Los resultados de alta calidad se alcanzan utilizando el algoritmo de Needleman y Wunsch o Smith y Waterman. Por lo tanto se prefieren los programas basados en dichos algoritmos. Ventajosamente, las comparaciones de secuencias pueden hacerse con el programa PileUp (J. Mol. Evolution., 25, 351-360, 1987, Higgins et al., CABIOS, 5 1989: 151-153) o preferiblemente con los programas Gap y BestFit, los cuales se basan respectivamente en los algoritmos de Needleman y Wunsch [J. Mol. Biol. 48; 443-453 (1970)] y Smith y Waterman [Adv. Appl. Math. 2; 482-489 (1981)]. Ambos programas son parte del paquete de software GCG [Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1991); Altschul et al. (1997) Nucleic Acids Res. 25:3389 et seq.]. Por lo tanto preferiblemente los cálculos para determinar los porcentajes de homología de secuencia se hacen con el programa Gap sobre el rango completo de secuencias. Se usan los siguientes ajustes estándar para la comparación de las secuencias de ácidos nucleicos: peso de la brecha: 50, peso de longitud: 3, coincidencia promedio: 10.000, diferencia promedio: 0.000.

Por ejemplo una secuencia que tiene un 80% de homología con la secuencia SEQ ID NO: 1 a nivel de ácido nucleico se entiende con el significado de una secuencia que, por comparación con la secuencia SEQ ID NO: 1 con el algoritmo del programa Gap con el conjunto de parámetros anteriores, tiene un 80% de homología.

En el estado de la técnica, la homología entre dos polipéptidos se entiende también con el significado de la identidad de la secuencia de aminoácidos sobre la cual en cada caso la longitud de la secuencia completa se calcula por comparación con ayuda del algoritmo del programa GAP ((Wisconsin Package Version 10.0, University of Wisconsin, Genetics Computer Group (GCG), Madison, Estados Unidos de América), definiendo los siguientes parámetros:

Peso de brecha:	8	Peso de longitud:	2
Coincidencia promedio:	2,912	Diferencia promedio:	-2,003

Por ejemplo una secuencia que tiene 80% de homología con la secuencia SEQ ID NO: 2 a nivel de proteína se entiende con el significado de una secuencia que, por comparación con la secuencia SEQ ID NO: 2 mediante el algoritmo de programa anterior con el conjunto de parámetros anterior, tiene un 80% de homología.

5 Los equivalentes funcionales derivados de los polipéptidos como se indica en SEQ ID NO: 690, de acuerdo con la invención por sustitución, inserción o eliminación tienen al menos 70% de identidad con uno de los polipéptidos como se indica en SEQ ID NO: 690, de acuerdo con la invención y se distinguen por esencialmente las mismas propiedades como polipéptidos tal como se indica en SEQ ID NO: 690.

10 Los equivalentes funcionales derivados de una secuencia de ácidos nucleicos tal como se indican en SEQ ID NO: 689 de acuerdo con la invención por sustitución, inserción o eliminación tienen al menos 70% de identidad con una del polipéptido tal como se indica en SEQ ID NO: 690, de acuerdo con la invención y codifican polipéptidos que tienen esencialmente las mismas propiedades que el polipéptido tal como se indica en SEQ ID NO: 690.

15 “Esencialmente las mismas propiedades” de un equivalente funcional se entiende todo anteriormente con el significado de que el equivalente funcional tiene la actividad mencionada anteriormente, por ejemplo, conferir un incremento de nitrógeno o en compuestos que contienen nitrógeno y una eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo la cantidad de fertilizante de nitrógeno a la vez que se incrementara la cantidad de proteína, actividad o función de dicho equivalente funcional en un organismo, por ejemplo un microorganismo, una planta o tejido vegetal o animal, células vegetales o animales o una parte de las mismas, por ejemplo, los plástidos.

20 Una molécula de ácido nucleico que codifica un homólogo a una secuencia de proteína tal como se indica en SEQ ID NO: 690, puede ser creada introduciendo una o más sustituciones, adiciones o eliminaciones de nucleótidos en una secuencia de nucleótidos de SEQ ID NO: 689, de tal forma que una o más sustituciones, adiciones o eliminaciones de aminoácidos se introducen en la proteína codificada. Pueden introducirse mutaciones en las secuencias de codificación por ejemplo en las secuencias tal como se indica en SEQ ID NO: 689, mediante técnicas estándar, tales como la mutagénesis dirigida al sitio y la mutagénesis mediada por PCR.

25 Preferiblemente, las sustituciones de aminoácidos conservadoras se hacen en uno o más residuos de aminoácidos no esenciales predichos. Una “sustitución de aminoácidos conservadora” es una en la cual el residuo de aminoácidos es reemplazado con un residuo de aminoácido que tiene una cadena lateral similar. Las familias de residuos de aminoácidos que tienen cadenas laterales similares han sido definidas en la técnica. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares no cargadas (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales con ramificación beta (por ejemplo treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo tirosina, fenilalanina, triptófano, histidina).

35 Así, un residuo de aminoácidos no esenciales en un polipéptido de la invención o un polipéptido usado en el proceso de la invención se reemplaza preferiblemente con otro residuo de aminoácido de la misma familia. Alternativamente, en otra realización, las mutaciones pueden ser introducidas aleatoriamente a lo largo de toda o parte de la secuencia de codificación de una molécula de ácido nucleico de la invención o usada en el proceso de la invención, tal como mediante mutagénesis por saturación, y los mutantes resultantes pueden ser seleccionados para la actividad descrita aquí para identificar mutantes que retienen o incluso incrementan por encima la actividad mencionada, por ejemplo confiriendo un incremento en el contenido de nitrógeno o compuestos que contienen nitrógeno y eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno.

45 Después de la mutagénesis de una de las secuencias mostradas aquí, la proteína codificada puede ser expresada de forma recombinante y puede determinarse la actividad de la proteína utilizando, por ejemplo, ensayos descritos aquí (véanse Ejemplos).

La homología más alta de la molécula de ácido nucleico usada en el proceso de acuerdo con la invención puede encontrarse para bases de datos accesibles de forma general mediante una búsqueda Gap.

50 Estas bases de datos, las cuales deben ser mencionadas, en particular en este contexto son bases de datos genéticas generales tales como la base de datos EMBL (Stoesser G. et al., *Nucleic Acids Res* 2001, Vol. 29, 17-21), la base de datos GenBank (Benson D.A. et al., *Nucleic Acids Res* 2000, Vol. 28,15-18), o la base de datos PIR (Barker W. C. et al., *Nucleic Acids Res.* 1999, Vol. 27, 39-43). Además es posible utilizar bases de datos genéticas específicas para un organismo para determinar secuencias ventajosas, en el caso de las levaduras por ejemplo ventajosamente la base de datos SGD (Cherry J. M. et al., *Nucleic Acids Res.* 1998, Vol. 26, 73-80) o la base de datos MIPS (Mewes H.W. et al., *Nucleic Acids Res.* 1999, Vol. 27, 44-48), en el caso de E. coli la base de datos GenProtEC

(<http://web.bham.ac.uk/bcm4ght6/res.html>), y en el caso de Arabidopsis la base de datos TAIR (Huala, E. et al., Nucleic Acids Res. 2001 Vol. 29(1), 102-5) o la base de datos MIPS.

5 Homólogos de las secuencias de ácido nucleico usadas, con una secuencia como la indicada en SEQ ID NO: 689 o de las secuencias de ácido nucleico derivadas de SEQ ID NO: 690, comprenden también variantes alélicas con al menos aproximadamente 70% o más de homología con una de las secuencias de nucleótidos mostradas o las secuencias de ácidos nucleicos derivadas mencionadas anteriormente o sus homólogos, derivados o análogos o partes de estos. Las variantes alélicas abarcan en particular variantes funcionales las cuales pueden ser obtenidas por eliminación, inserción o sustitución de nucleótidos de las secuencias mostradas, preferiblemente de una secuencia como se indica en SEQ ID NO: 689, o de las secuencias de ácidos nucleicos derivadas, siendo la intención, sin embargo, que la actividad enzimática o la actividad biológica de las proteínas sintetizadas resultantes sea retenida o incrementada ventajosamente.

15 Los polipéptidos (= proteína), que tienen aún la actividad enzimática esencial del polipéptido de la presente invención que confiere un incremento de nitrógeno o de compuestos que contienen nitrógeno y eficiencia incrementada del uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, esto es, cuya actividad es esencialmente no reducida, son polipéptidos con al menos 10% o 20%, de preferencia 30% o 40%, preferiblemente de manera especial 50% o 60%, preferiblemente de manera muy especial 80% o 90% o más de la actividad biológica o actividad enzimática tipo silvestre, ventajosamente, la actividad no se reduce esencialmente en comparación con la actividad de un polipéptido como se indica en SEQ ID NO: 690 y se expresa bajo condiciones idénticas. Los homólogos de una secuencia como se indica en SEQ ID NO: 689 o de una secuencia derivada como se indica en la Tabla II columnas 5 o 7, también significa secuencias truncadas, ADNc, ADN o ARN de cadena sencilla de la secuencia de ADN de codificación y no codificación. Homólogos de dichas secuencias también se entienden con el significado de derivados, los cuales comprenden regiones no codificadoras tales como UTR, terminadores, potenciadores o variantes promotoras. Los promotores corriente arriba de las secuencias de nucleótidos establecidas pueden ser modificados por una o más sustituciones, inserciones y/o eliminaciones de nucleótidos sin, sin embargo, interferir con la funcionalidad o actividad bien sea de los promotores, el marco de lectura abierta (= ORF) o con la región reguladora 3' tal como los terminadores u otras regiones reguladoras 3', los cuales están lejos del ORF. Es además posible que la actividad de los promotores se incremente por las modificaciones de su secuencia o que sean reemplazados completamente por promotores más activos, incluso promotores de organismos heterólogos. Promotores apropiados son conocidos por la persona experimentada en la técnica y se mencionan aquí más adelante.

En una realización adicional, el proceso de acuerdo con la presente invención comprende las siguientes etapas:

- (a) seleccionar un organismo o una parte del mismo que expresa el polipéptido de esta invención en el citosol y/o en un organelo tal como un plástido o mitocondria;
 - (b) mutagenizar el organismo seleccionado o la parte del mismo;
 - 35 (c) comparar la actividad del nivel de expresión de dicho polipéptido en el organismo mutagenizado o la parte del mismo con la actividad o la expresión de dicho polipéptido en el organismo seleccionado o la parte del mismo;
 - (d) seleccionar los organismos mutagenizados a partir de los mismos que comprenden una actividad o nivel de expresión incrementados de dicho polipéptido en comparación con el organismo seleccionado o la parte del mismo;
 - (e) opcionalmente, hacer crecer y cultivar los organismos de las partes de los mismos.
- 40 Ventajosamente los organismos seleccionados son mutagenizados de acuerdo con la invención. De acuerdo con la invención la mutagénesis es cualquier cambio de la información genética en el genoma de un organismo, que significa cualquier cambio estructural o de composición en el ácido nucleico preferiblemente ADN de un organismo que no es causada por segregación normal o procesos de recombinación genéticos. Tales mutaciones pueden ocurrir espontáneamente, o pueden ser inducidas por mutágenos como se describe más adelante. Tal cambio puede ser inducido bien sea aleatoriamente o selectivamente. En ambos casos la información genética del organismo es modificada. En general esto lleva a la situación de que la actividad del producto genético de los genes relevantes dentro de las células o dentro del organismo se incrementa.

50 En caso de la mutagénesis específica o así llamada dirigida al sitio se muta un gen distintivo y por lo tanto su actividad y/o la actividad del producto genético codificado se reprime, reduce o incrementa, preferiblemente se incrementa. En el evento de una mutagénesis aleatoria se mutan uno o más genes por cambio en sus actividades y/o las actividades de sus productos genéticos son reprimidas, reducidas o incrementadas, preferiblemente incrementadas.

Para el propósito de una mutagénesis de una población enorme de organismos, tal población puede ser transformada con un constructo de ADN, el cual es útil para la activación de tantos genes como sean posibles de un organismo,

preferiblemente todos los genes. Por ejemplo el constructo puede contener un promotor fuerte o uno o más potenciadores, los cuales son capaces de activar de manera transcripcional los genes en la vecindad de su lado de integración. Con este método es posible mutagenizar estadísticamente, por ejemplo activar casi todos los genes de un organismo por la integración aleatoria de un constructo de activación. Después de esto el técnico experimentado puede identificar aquellas líneas mutagenizadas en las cuales un gen de la invención ha sido activado, lo cual a su vez lleva al incremento deseado de nitrógeno o en compuestos que contienen nitrógeno y a una eficiencia incrementada en el uso del nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno.

Los genes de la invención también pueden ser activados por mutagénesis, bien sea de regiones reguladoras o de codificación. En el evento de una mutagénesis aleatoria se trata un número enorme de organismos con un agente mutagénico. La cantidad de dicho agente y la intensidad del tratamiento se escogerán de manera tal que estadísticamente casi cada gen muta una vez. El proceso para la mutagénesis aleatoria así como los agentes respectivos son bien conocidos por las personas experimentadas. Tales métodos se divulgan por ejemplo en A.M. van Harten (1998), "Mutation breeding: theory and practical applications", Cambridge University Press, Cambridge, UK], E Friedberg, G Walker, W Siede [(1995), "DNA Repair and Mutagenesis", Blackwell Publishing], or K. Sankaranarayanan, J. M. Gentile, L. R. Ferguson [(2000) "Protocols in Mutagenesis", Elsevier Health Sciences]. El operario experimentado sabe que la rata de mutación espontánea en las células de un organismo es muy baja y que hay disponible un gran número de agentes químicos, físicos o biológicos para la mutagénesis de los organismos. Estos agentes se denominan mutagénos o agentes mutagénicos. Como se mencionó anteriormente hay disponibles tres clases diferentes de mutagénos (agentes químicos, físicos o biológicos).

Hay casos diferentes de mutagénos químicos, los cuales pueden ser separados por su modo de acción. Por ejemplo los análogos de bases tales como 5-bromuroacilo, 2-aminopurina. Otros mutagénos químicos interactúan con el ADN tal como ácido sulfúrico, ácido nitroso, hidroxilamina; u otros agentes aspirantes tales como agentes monofuncionales como metasulfonato de metilo, dimetilsulfato, metanosulfonato de metilo), bifuncionales como sulfuro de dicloroetilo, mitomicina, nitrosoguanidina-dialquilnitrosamina, derivados de n-nitroso guanidina, n-alkil-n-nitro-n-nitroso-guanidina), colorantes intercaladores como acridina, y bromuro de etidio).

Los mutagénos físicos son por ejemplo irradiación ionizante (rayos X), irradiación UV. Están disponibles diferentes formas de irradiación y son fuertes mutágenos. Pueden distinguirse dos clases principales de irradiación: a) irradiación no ionizante tal como la luz UV o irradiación ionizante tal como rayos X. Los mutagénos biológicos son por ejemplo elementos transponibles por ejemplo elementos IS tales como IS 100, transposones tales como Tn5, Tn10, Tn916 o Tn1000 o fagos tales como Mu^{amplac}, P1, T5, λplac, etc. Los métodos para introducir este ADN fago en el microorganismo apropiado son bien conocidos por los experimentados (véase Microbiology, Third Edition, Eds. Davis, B.D., Dulbecco, R., Eisen, H.N. and Ginsberg, H.S., Harper International Edition, 1980). El procedimiento común de una mutagénesis por trasposón es la inserción de un elemento transponible con un gen, o cercano por ejemplo en la región promotora o terminadora y que por lo tanto lleva a una pérdida de la función genética. Los procedimientos para localizar el transposon dentro del genoma de los organismos son bien conocidos por una persona experimentada en la técnica.

Preferiblemente se utiliza un procedimiento químico o bioquímico para la mutagénesis de los organismos. Un método químico es la mutagénesis con N-metil-N-nitro-nitrosoguanidina.

Otros métodos biológicos son divulgados por Spee et al. (Nucleic Acids Research, Vol. 21, No. 3, 1993: 777 - 778). Speed et al., enseña un método de PCR utilizando dITP para la mutagénesis aleatoria. Este método descrito por Speed et al., fue mejorado posteriormente por Rellos et al., (Protein Expr. Purif., 5, 1994 : 270 - 277). El uso de una técnica de recombinación in vitro para mutagénesis moleculares descrito por Stemmer (Proc. Natl. Acad. Sci. USA, Vol. 91, 1994: 10747 - 10751). Moore et al. (Nature Biotechnology Vol. 14, 1996: 458 - 467) describe la combinación de los métodos de PCR y recombinación para incrementar la actividad enzimática de una esterasa hacia un para-nitrobencil éster. Otra ruta a la mutagénesis de enzimas está descrita por Greener et al., in Methods in Molecular Biology (Vol. 57, 1996: 375 - 385). Greener et al. usan la cepa específica de Escherichia coli XL1-Red para generar mutantes de Escherichia coli que tienen resistencia antibiótica incrementada.

En una realización, la proteína de acuerdo con la invención o la molécula de ácido nucleico caracterizada aquí se origina de un organismo eucariote o procariote tal como un animal no humano, una planta, un microorganismo tal como un hongo, una levadura, un alga, una diatomea o una bacteria. Las moléculas de ácido nucleico, las cuales ventajosamente pueden ser utilizadas en el proceso de la invención se originan a partir de levaduras, por ejemplo de la familia Saccharomycetaceae, en particular el género Saccharomyces, o géneros de levaduras tales como Candida, Hansenula, Pichia, Yarrowia, Rhodotorula o Schizosaccharomyces y especialmente de forma ventajosa de la especie Saccharomyces cerevisiae.

En una realización, las moléculas de ácido nucleico, las cuales ventajosamente pueden ser utilizadas en el proceso de la invención se originan a partir de bacterias, por ejemplo de Proteobacterias, en particular Gammaproteobacterias, más preferidas de Enterobacteriales, por ejemplo de la familia Enterobacteriaceae, particularmente de los géneros Escherichia, Salmonella, Klebsiella, ventajosamente de la especie Escherichia coli K12.

5 Sí, en el proceso de acuerdo con la invención, se seleccionan plantas como organismo donante, esta planta puede, en principio, estar en cualquier relación filogenética de la planta receptora. Las plantas donantes y receptoras pueden pertenecer a la misma familia, género, especie, variedad o línea, dando como resultado una homología creciente entre los ácidos nucleicos que se van a integrar y las partes correspondientes del genoma de la planta receptora. Esto también se aplica análogamente a microorganismos como organismo donante y receptor.

También podría ser ventajoso utilizar moléculas de ácido nucleico para especies muy distintas, puesto que estas pueden exhibir sensibilidad reducida contra mecanismos reguladores endógenos y tales secuencias podrían no ser reconocidas por mecanismos silenciadores endógenos.

10 De acuerdo con lo anterior, una realización de la solicitud se relaciona con el uso de moléculas de ácido nucleico en el proceso de la invención desde plantas, por ejemplo plantas de cultivo, por ejemplo de *B. napus*; glicine max; girasol, arroz, algodón, linaza o maíz y sus homólogos.

15 Las secuencias de ácidos nucleicos usadas en el proceso son introducidas ventajosamente en un constructo de ácido nucleico, preferiblemente un casete de expresión que hace posible la expresión de las moléculas de ácido nucleico en el organismo, ventajosamente una planta o un microorganismo. Las secuencias de ácidos nucleicos mostradas en SEQ ID NO: 689 son preferiblemente introducidas en un constructo de ácido nucleico, el cual contiene adicionalmente una secuencia de ácido nucleico que codifican una secuencia objetivo plastídica, de tal manera que se codifica una proteína de la cual por traducción dirige el polipéptido codificado mediante secuencias de ácidos nucleicos mostradas en SEQ IDE NO: 689 hacia el compartimiento plastídico. La persona experimentada en la técnica está familiarizada con el diseño de tales constructos de ácidos nucleicos. De acuerdo con lo anterior, la invención también se relaciona con un constructo de ácido nucleico, preferiblemente con un constructo de expresión, que comprende la molécula de ácido nucleico de la presente invención enlazada funcionalmente a uno o más elementos o señales reguladoras. De acuerdo con lo anterior, la invención también se relaciona con un constructo de ácido nucleico, preferiblemente con un constructo de expresión, que comprende la molécula de ácido nucleico de la presente invención, preferiblemente una secuencia como se muestra en SEQ ID NO: 689 enlazada funcionalmente con una secuencia objetivo plastídica.

25 Tal como se describe aquí, el constructo de ácido nucleico puede comprender también adicionalmente genes, que van a ser introducidos en los organismos o células. Es posible y ventajoso introducir en, y expresar en, los organismos anfitriones genes reguladores tales como genes para inductores, represores o enzimas, los cuales, obedeciendo a su actividad enzimática, encajan en la regulación de uno o más genes de una ruta biosintética. Estos genes pueden ser de origen heterólogo u homólogo. Además, pueden estar presentes ventajosamente genes de biosíntesis adicionales, o aún más estos genes pueden ser localizados en uno o más constructos de ácidos nucleicos adicionales. Los genes, los cuales se emplean ventajosamente como genes para biosíntesis son genes del metabolismo de aminoácidos, de glicólisis, del metabolismo del ácido tricarbóxico u otras combinaciones. Tal como se describen aquí, las secuencias o factores reguladores pueden tener un efecto positivo sobre preferiblemente la expresión genética de los genes introducidos, incrementándola así. De esta forma, un potenciamiento de los efectos reguladores puede tener lugar ventajosamente a nivel de transcripción utilizando señales de transcripción fuertes tales como promotores y/o potenciadores. Además, sin embargo, también es posible un potenciamiento de la traducción, por ejemplo incrementando la estabilidad de ARNm o insertando una secuencia potenciadora de la traducción.

40 En principio, el constructo de ácido nucleico puede comprender las secuencias reguladoras aquí descritas y secuencias adicionales relevantes para la expresión de los genes comprendidos. Así, el constructo de ácido nucleico de la invención puede ser utilizado como casete de expresión y puede ser usado así directamente para la introducción en la planta, o además puede ser introducido en un vector. De acuerdo con lo anterior en una realización el constructo de ácido nucleico es un casete de expresión que comprende un microorganismo promotor o un microorganismo terminador o ambos. En otra realización el casete de expresión abarca un promotor de planta o un terminador de planta o ambos. En otra realización el casete de expresión abarca una secuencia objetivo plastídica de planta.

45 De acuerdo con lo anterior, en una realización, el proceso de acuerdo con la invención comprende las siguientes etapas:

(a) introducir un constructo del ácido nucleico que comprende la molécula de ácido nucleico usado en el proceso de la invención o que codifica el polipéptido de la presente invención o se usa en el proceso de la invención; o

50 (b) introducir una molécula de ácido nucleico, que incluye secuencias o factores reguladores, cuya expresión incrementa la expresión de la molécula de ácido nucleico de la invención o se usa en el proceso de la invención o codifica el polipéptido de la presente invención o se usa en el proceso de la invención; en una célula, o un organismo o una parte de los mismos, preferiblemente en una planta, célula vegetal o un microorganismo; y

(c) expresar el producto genético codificado por el constructo de ácido nucleico o la molécula de ácido nucleico mencionada abajo (a) o (b) en la célula, en el citosol o en un organelo tal como un plástido o una mitocondria o ambos, preferiblemente en plástidos o preferiblemente en el citosol o el organismo.

Después de la introducción y expresión del constructo de ácido nucleico el organismo o células transgénicos se cultivan ventajosamente y se recolectan posteriormente. El organismo o célula transgénicos puede ser un organismo procarionte o eucariote tal como un microorganismo, una célula vegetal, un tejido vegetal, preferiblemente una planta de cultivo, o una parte de los mismos.

- 5 Para introducción de una molécula de ácido nucleico en un constructo de ácidos nucleicos, por ejemplo, como parte de un casete de expresión, el segmento del gen codogénico está sometido ventajosamente a una reacción de amplificación y ligación en la manera conocida por una persona experimentada. Preferiblemente se sigue un procedimiento similar o protocolo para la Pfu ADN polimerasa o una mezcla de Pfu/Taq ADN polimerasa. Los cebadores se seleccionan de acuerdo con la secuencia que se va a amplificar.
- 10 Los cebadores deberían ser escogidos de manera expeditiva de manera tal que el amplificado comprenda la secuencia codogénica desde el inicio hasta el codón de terminación. Después de la amplificación, el amplificado se analiza de manera expeditiva. Por ejemplo, los análisis pueden considerar calidad y cantidad y pueden llevarse a cabo siguiendo la separación por electroforesis por gel. Después de esto, el amplificado puede ser purificado siguiendo un protocolo estándar (por ejemplo Quiagen). Un alícuota del amplificado purificado está disponible entonces para la etapa subsiguiente de clonación. Los vectores de clonación adecuados son generalmente conocidos por la persona experimentada.

15 Incluyen, en particular, vectores que son capaces de replicación en sistemas de clonación fáciles de manejar tales como sistemas basados en bacterias, levaduras o células de insectos (por ejemplo expresión de baculovirus), es decir especialmente vectores que aseguran una eficiente clonación en *E. coli* y que hacen posible la transformación estable de plantas. Los vectores, los cuales deben ser mencionados en particular son diversos sistemas de vectores binarios y cointegrados los cuales son adecuados para transformación mediada por T-ADN. Tales sistemas de vectores son caracterizados generalmente porque contienen al menos los genes *vir*, los cuales son requeridos para la transformación mediada por *Agrobacterium*, y las secuencias de frontera de T-ADN.

20 En general, los sistemas de vectores preferiblemente comprenden también regiones adicionales reguladoras *cis* tales como promotores y terminadores y/o marcadores de selección por medio de los cuales pueden identificarse organismos transformados de manera adecuada. Mientras que los genes *vir* y las secuencias T-ADN están localizados en el mismo vector en el caso de sistemas de vectores cointegrados, los sistemas binarios están basados en al menos dos vectores, uno de los cuales porta los genes *vir*, pero no T-ADN, mientras que un segundo porta T-ADN pero no el gen *vir*. Obedeciendo a este hecho, los últimos vectores mencionados son relativamente pequeños, fáciles de manipular y capaces de replicarse en *E. coli* y en *Agrobacterium*. Estos vectores binarios incluyen vectores de las series pBIB-HYG, pPZP, pBecks, pGreen. Aquellos que se utilizan preferiblemente de acuerdo con la invención son Bin 19, pB1101, pBinAR, pGPTV y pCAMBIA. Una revisión de vectores binarios y su uso está dada por Hellens et al, Trends in Plant Science (2000) 5, 446-451.

25 Para la preparación de un vector, los vectores pueden ser primero linealizados utilizando endonucleasas de restricción y luego modificados enzimáticamente de manera adecuada. Después de esto, se purifica el vector, y se emplea una alícuota en la etapa de clonación. En la etapa de clonación, la enzima se escinde y si se requiere, el amplificado purificado se clona junto con fragmentos de vector preparados de manera similar, utilizando ligasa. En este contexto, un constructo de ácido nucleico específico o un constructo vector o plásmido, pueden tener uno o más segmentos genéticos codogénicos. Los segmentos genéticos codogénicos en estos constructos están enlazados preferiblemente de manera operativa a secuencias reguladoras. Las secuencias reguladoras incluyen, en particular, secuencias de plantas tales como los promotores y terminadores antes descritos y eventualmente en una secuencia objetivo entre el segmento codogénico y el promotor. Los constructos pueden ser propagados ventajosamente de manera estable en microorganismos, en particular *Escherichia coli* y/o *Agrobacterium tumefaciens*, bajo condiciones selectivas y permiten la transferencia del ADN heterólogo en plantas y otros microorganismos. De acuerdo con una realización particular, los constructos están basados en vectores binarios (la revisión de un vector binario: Hellens et al., 2000). Como regla, contienen secuencias reguladoras procarionte, tales como origen de replicación y marcadores de selección, para la multiplicación en microorganismos tales como *Escherichia coli* y *Agrobacterium tumefaciens*. Los vectores pueden contener adicionalmente secuencias agrobacterianas de T-ADN para la transferencia de ADN en genomas vegetales u

35 otras secuencias reguladoras eucariotes para transferencia en otras células eucariotes por ejemplo *Saccharomyces sp* u otras secuencias reguladoras procarionte para la transferencia en otras células procariontes, por ejemplo, *Corynebacterium sp* o *Bacillus sp*. Para la transformación de plantas, la secuencia límite derecha, que comprende aproximadamente 25 pares de bases, de la secuencia total agrobacteriana de ADN-T se incluye de manera ventajosa. Usualmente, los constructos del vector de transformación en plantas de acuerdo con la invención, contienen secuencias de T-ADN tanto de la región frontera derecha como de la izquierda, las cuales contienen sitios de reconocimiento

40 expeditivos para enzimas que actúan específicamente en los sitios, las cuales a su vez, son codificadas por algunos de los genes *vir*.

Organismo anfitriones adecuados son conocidos por la persona experimentada. Los organismos ventajosos son descritos adicionalmente más arriba en la presente solicitud. Incluye en particular eucariotes o eubacterias, por ejemplo procariontes o bacterias arcae. Ventajosamente los organismos anfitriones son microorganismos seleccionados del grupo

consistente de Actinomycetaceae, Bacillaceae, Brevibacteriaceae, Corynebacteriaceae, Enterobacteriaceae, Gordoniaceae, Micrococcaceae, Mycobacteriaceae, Nocardiaceae, Pseudomonaceae, Rhizobiaceae, Streptomyetaceae, Chaetomiaceae, Choanephoraceae, Cryptococcaceae, Cunninghamellaceae, Demetiaceae, Moniliaceae, Mortierellaceae, Mucoraceae, Pythiaceae, Sacharomycetaceae, Saprolegniaceae, Schizosacharomycetaceae, Sodariaceae, Sporobolomycetaceae, Tuberculariaceae, Adelotheciaceae, Dinophyceae, Ditrichaceae y Prasinophyceae. Preferiblemente los microorganismos unicelulares, por ejemplo, hongos, bacterias o protozoos, tales como los hongos del género *Claviceps* o *Aspergillus* o bacterias gram positivas tales como los géneros *Bacillus*, *Corynebacterium*, *Micrococcus*, *Brevibacterium*, *Rhodococcus*, *Nocardia*, *Caseobacter* o *Arthrobacter* o bacterias gram negativas tales como los géneros *Escherichia*, *Flavobacterium* or *Salmonella*, o levaduras tales como los géneros *Rhodotorula*, *Hansenula*, *Pichia*, *Yarrowia*, *Saccharomyces*, *Schizosaccharomyces* o *Candida*.

Los organismos anfitriones que son seleccionados de manera especialmente ventajosa en el proceso de acuerdo con la invención son microorganismos seleccionados del grupo de los géneros y especies consistentes de *Hansenula anomala*, *Candida utilis*, *Claviceps purpurea*, *Bacillus circulans*, *Bacillus subtilis*, *Bacillus sp.*, *Brevibacterium albidum*, *Brevibacterium album*, *Brevibacterium cerinum*, *Brevibacterium flavum*, *Brevibacterium glutamigenes*, *Brevibacterium iodinum*, *Brevibacterium ketoglutamicum*, *Brevibacterium lactofermentum*, *Brevibacterium linens*, *Brevibacterium roseum*, *Brevibacterium saccharolyticum*, *Brevibacterium sp.*, *Corynebacterium acetoacidophilum*, *Corynebacterium acetoglutamicum*, *Corynebacterium ammoniagenes*, *Corynebacterium glutamicum* (= *Micrococcus glutamicum*), *Corynebacterium melassecola*, *Corynebacterium sp.* o *Escherichia coli*, específicamente *Escherichia coli* K12 y sus cepas descritas.

Preferidos de manera ventajosa de acuerdo con la invención son organismos anfitriones de los del género *Agrobacterium tumefaciens* o plantas. Plantas preferidas son seleccionadas entre las familias *Aceraceae*, *Anacardiaceae*, *Apiaceae*, *Asteraceae*, *Betulaceae*, *Boraginaceae*, *Brassicaceae*, *Bromeliaceae*, *Cactaceae*, *Caricaceae*, *Caryophyllaceae*, *Cannabaceae*, *Convolvulaceae*, *Chenopodiaceae*, *Elaeagnaceae*, *Geraniaceae*, *Gramineae*, *Juglandaceae*, *Lauraceae*, *Leguminosae*, *Linaceae*, *Cucurbitaceae*, *Cyperaceae*, *Euphorbiaceae*, *Fabaceae*, *Malvaceae*, *Nymphaeaceae*, *Papaveraceae*, *Rosaceae*, *Salicaceae*, *Solanaceae*, *Arecaceae*, *Iridaceae*, *Liliaceae*, *Orchidaceae*, *Gentianaceae*, *Labiaceae*, *Magnoliaceae*, *Ranunculaceae*, *Carifolaceae*, *Rubiaceae*, *Scrophulariaceae*, *Ericaceae*, *Polygonaceae*, *Violaceae*, *Juncaceae*, *Poaceae*, pastos perennes, cultivos para forraje, vegetales y ornamentales.

Se prefieren especialmente plantas seleccionadas de los grupos de las familias *Apiaceae*, *Asteraceae*, *Brassicaceae*, *Cucurbitaceae*, *Fabaceae*, *Papaveraceae*, *Rosaceae*, *Solanaceae*, *Liliaceae* o *Poaceae*. Especialmente ventajosas son en particular, plantas de cultivo. De acuerdo con lo anterior, una planta ventajosa pertenece preferiblemente al grupo del género cacahuete, colza para aceite, canola, girasol, cártamo, oliva, sésamo, avellana, almendra, aguacate, laurel, calabaza/calabacin, linaza, soja, pistacho, borraja, maíz, trigo, centeno, avena, sorgo y millo, tritical, arroz, algodón, cebada, cassava, patata, remolacha, remolacha para forraje, berenjena y pastos perennes y plantas para forraje, palma de aceite, vegetales (brassicas, vegetales de raíz, vegetales tuberosos, vegetales de vaina, vegetales frutales, vegetales tipo cebolla, vegetales de hojas y vegetales de tallos), trigo sarraceno, alcachofa de Jerusalen, judía ancha, algarroba, lenteja, alfalfa, judía enana, lupino, ajo y lucerna.

Con el fin de introducir, en una planta, la molécula de ácido nucleico de la invención o usado en el proceso de acuerdo con la invención, se ha demostrado ventajoso transferirla primero en un anfitrión intermedio, por ejemplo una bacteria o una célula eucariote unicelular. La transformación en *E. coli* que puede ser llevada cabo de una manera conocida per se, por ejemplo por medio de un choque de calor o electroporación, se ha mostrado expeditiva en este contexto. Así, las colonias de *E. coli* transformadas pueden ser analizadas en cuanto a su eficiencia de clonación. Esto puede llevarse a cabo con la ayuda de un PCR. Aquí, no solamente la identidad, sino también la integridad, del constructo del plásmido pueden ser verificadas con la ayuda de un número de colonias definido sometiendo a una alícuota de las colonias a dicho PCR. Como regla, los cebadores universales que son derivados de las secuencias vectores se utilizan para este propósito, siendo posible, por ejemplo, para un cebador de avance estar dispuesto corriente arriba del ATG de inicio y un cebador reverso estar dispuesto corriente abajo del codón de detención del segmento genético codogénico. Los amplificados son separados por electroforesis y definidos con respecto a cantidad y calidad.

Los constructos de ácidos nucleicos, que se verifican opcionalmente, se utilizan subsecuentemente para la transformación de las plantas u otros huéspedes, por ejemplo, otras células eucariotes u otras células procariotes. Con este fin, puede ser necesario obtener primero los constructos a partir del anfitrión intermedio. Por ejemplo, los constructos pueden ser obtenidos como plásmidos a partir de anfitriones bacterianos por los métodos similares al aislamiento convencional de plásmidos.

La molécula de ácido nucleico de la invención o usada en el proceso de acuerdo con la invención también puede ser introducida en vectores virales modificados tales como vectores de baculovirus para expresión en células de insectos o vectores virales de plantas tales como el virus mosaico del tabaco o los vectores basados en el virus X de la patata. Las metodologías que llevan a la expresión de las proteínas a partir del genoma viral modificado incluyendo la molécula del ácido nucleico de la invención o usada en el proceso de acuerdo con la invención involucran por ejemplo la inoculación de plantas de tabaco con ARN infeccioso transcrito in vitro a partir de una copia de ADNc del genoma viral

recombinante. Otras metodologías utilizan la transfección de plantas enteras a partir de heridas inoculadas con *Agrobacterium tumefaciens* que contienen copias de ADNc de virus recombinantes más ARN sentido. Diferentes vectores y virus son conocidos por las personas experimentadas en la técnica para expresión en diferentes objetivos, por ejemplo plantas de producción.

5 Se conoce un gran número de métodos para la transformación de plantas. Puesto que, de acuerdo con la invención, una integración estable del ADN heterólogo en el genoma de las plantas es ventajosa, la transformación mediada por T-ADN se ha demostrado expeditiva en particular. Para este propósito, es necesario primero transformar vehículos adecuados, en particular *Agrobacterias*, con un segmento genético codogénico con el correspondiente constructo de plásmido que comprende la molécula de ácido nucleico de la invención. Esto puede llevarse a cabo de una manera conocida per se. Por ejemplo, dicho constructo de ácido nucleico de la invención, o dicho constructo de expresión o dicho constructo de plásmido, el cual ha sido generado de acuerdo con lo que se ha detallado anteriormente, puede ser transformado en *agrobacterias* competentes por medio de electroporación o choque de calor. En principio, se debe diferenciar entre la formación de vectores cointegrados por un lado y la transformación con vectores binarios por otro lado. En el caso de la primera alternativa, los constructos, los cuales comprenden el segmento genético codogénico o la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención no tienen secuencias de T-ADN, pero la formación de los vectores cointegrados o los constructos tienen lugar en la *agrobacteria* por recombinación homóloga del constructo con T-ADN. El T-ADN está presente en las *agrobacterias* en la forma de plásmidos Ti o Ri en los cuales el ADN exógeno ha reemplazado de manera expeditiva los oncógenos. Si se utilizan vectores binarios, pueden ser transferidos a *agrobacterias* bien sea por conjugación bacteriana o por transferencia directa. Estas *agrobacterias* comprenden ya de manera expeditiva el vector que porta los genes vir (denominados corrientemente como plásmido auxiliar Ti (Ri)).

Uno o más marcadores pueden ser utilizados también de manera expeditiva junto con el constructo de ácido nucleico, o el vector de la invención y, si se van a transformar plantas o células de plantas junto con el T-ADN, con la ayuda del cual es posible el aislamiento o selección de los organismos transformados, tales como *agrobacterias* o células vegetales transformadas. Estos genes marcadores posibilitan la identificación de una transferencia exitosa a las moléculas de ácido nucleico de acuerdo con la invención a través de una serie de principios diferentes, por ejemplo a través de la identificación visual con la ayuda de fluorescencia, luminiscencia o en el rango de longitud de onda de luz que es discernible para el ojo humano, mediante una resistencia a herbicidas o antibióticos, a través de las que son conocidas como marcadores nutritivos (marcadores de auxotrofismo) o marcadores antinutritivos, a través de ensayos enzimáticos o a través de fitohormonas. Ejemplos de tales marcadores que pueden ser mencionados son GFP (= proteína fluorescente verde); los sistemas luciferina/luciferasa, la galactosidasa con sus sustratos coloreados, por ejemplo X-Gal, la resistencia a herbicidas a, por ejemplo, imidazolinona, glifosfato, fosfotrisina o sulfonilurea, las resistencias a antibióticos a, por ejemplo, bleomicina, higromicina, estreptomycin, kanamicina, tetraciclina, cloranfenicol, ampicilina, gentamicina, geneticina (G418), espectinomicina o blasticidina, para mencionar solo unos pocos, marcadores nutritivos tales como la utilización de manosa o xilosa, o marcadores antinutritivos tales como la resistencia a 2-desoxiglucosa o D-aminoácidos. Estas listas son un número pequeño de marcadores posibles. La persona experimentada está muy familiarizada con tales marcadores. Los marcadores diferentes son los preferidos dependiendo del organismo y del método de selección.

Como regla, se desea que los constructos de ácidos nucleicos de plantas estén flanqueados por T-ADN en uno o en ambos lados del segmento genético codogénico. Esto es particularmente útil cuando las bacterias de las especies *Agrobacterium tumefaciens* o *Agrobacterium rhizogenes* se utilizan para la transformación. Un método, que es preferido de acuerdo con la invención, es la transformación con la ayuda de *Agrobacterium tumefaciens*. Sin embargo, también pueden utilizarse métodos biolísticos ventajosamente para introducir las secuencias en el proceso de acuerdo con la invención, y la introducción por medio de PEG también es posible. Las *agrobacterias* transformadas pueden ser cultivadas de manera conocida per se y por lo tanto están disponibles para la transformación expeditiva de las plantas. Las plantas o partes de plantas que se van a transformar se cultivan o proveen de la manera habitual. Se permite que las *agrobacterias* transformadas subsecuentemente actúen sobre las plantas o partes de plantas hasta que se alcance una tasa de transformación suficiente. Al permitir que la *agrobacteria* actúe sobre las plantas o partes de plantas, pueden tener lugar diferentes formas. Por ejemplo, puede utilizarse un cultivo de células o tejidos vegetales morfológicos. Después de la transferencia de T-ADN, las bacterias son, como regla, eliminadas por antibióticos, y se inducen a regeneración del tejido vegetal. Esto se hace en particular usando hormonas vegetales adecuadas con el fin de inducir inicialmente la formación de callos y promover entonces el desarrollo de brotes.

La transferencia de genes foráneos en el genoma de una planta se denomina transformación. Al hacer esto los métodos descritos para la transformación y regeneración de plantas a partir de tejidos vegetales o células de plantas se utilizan para transformación transiente o estable. Un método de transformación ventajoso es la transformación *in planta*. Con este fin, es posible, por ejemplo, permitir que las *agrobacterias* actúen sobre semillas de plantas o inocular el meristema de la planta con *agrobacterias*. Se ha demostrado particularmente expeditivo de acuerdo con la invención permitir que una suspensión de *agrobacterias* transformadas actúe sobre la planta intacta o al menos sobre las flores primordiales. La planta es cultivada subsecuentemente hasta obtener las semillas de la planta tratada (Clough y Bend, Plant J. (1998) 16, 735-743). Para seleccionar plantas transformadas, el material vegetal obtenido en la transformación es, como regla, sometido a condiciones selectivas de tal forma que las plantas transformadas puedan ser extinguidas de las plantas no

transformadas. Por ejemplo, las semillas obtenidas en la forma antes descrita pueden ser plantadas y, después de un crecimiento inicial, sometidas a una selección adecuada por aspersión. Una posibilidad adicional consiste en cultivar las células, si es apropiado después de la esterilización, sobre placas de agar utilizando un agente de selección adecuado de tal manera que solamente las semillas transformadas puedan crecer hasta ser plantas. Métodos de transformación ventajosos adicionales, en particular para plantas, son conocidos por las personas experimentadas en la técnica y se describen más adelante.

Métodos ventajosos y adecuados adicionales son transformación de protoplastos por consumo de ADN inducido por poli (etilen glicol), el método "biolístico" utilizando el canon genético – denominado como método de bombardeo de partículas, electroporación, incubación de embriones secos en solución de ADN, microinyección y transferencia de genes mediada por *Agrobacterium*. Dichos métodos son descritos a manera de ejemplo en B. Jenet et al., *Techniques for Gene Transfer*, in: *Transgenic Plants*, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press (1993) 128-143 and in Potrykus *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 42 (1991) 205-225). Los ácidos nucleicos o el constructo que se van a expresar se clonan preferiblemente en un vector, el cual es adecuado para la transformación de *Agrobacterium tumefaciens*, por ejemplo pBin19 (Bevan et al., *Nucl. Acids Res.* 12 (1984) 8711). Las agrobacterias transformadas o por tal vector pueden ser utilizadas entonces de manera conocida para la transformación de plantas, en particular de plantas de cultivo tales como a manera de ejemplo, plantas de tabaco, por ejemplo bañando las hojas raspadas u hojas picadas en una solución de agrobacterias y luego cultivándolas en medio adecuado. La transformación de las plantas por medio de *Agrobacterium tumefaciens* está descrita, por ejemplo, por Höfgen and Willmitzer in *Nucl. Acid Res.* (1988) 16, 9877 or is known inter alia from F.F. White, *Vectors for Gene Transfer in Higher Plants*, in *Transgenic Plants*, Vol. 1, Engineering and Utilization, eds. S.D. Kung and R. Wu, Academic Press, 1993, pp. 15-38.

Las moléculas de ácido nucleico antes mencionadas pueden ser clonadas en los constructos o vectores de ácidos nucleicos de acuerdo con la invención en combinación junto con genes adicionales, o además genes diferentes que son introducidos transformando diferentes constructos o vectores de ácidos nucleicos (incluyendo plásmidos) en una célula anfitriona, ventajosamente en una célula vegetal o en microorganismos.

Además de una secuencia indicada en SEQ ID NO: 689, es ventajoso expresar y/o mutar adicionalmente genes adicionales en los organismos. De manera especialmente ventajosa, al menos un gen adicional de la ruta biosintética de los aminoácidos tales como L-lisina, L-treonina y/o L-metionina se expresan los organismos tales como planta o microorganismos. También es posible que la regulación de los genes naturales haya sido modificada ventajosamente de tal manera que el gen y/o su producto genético no esté sometido más a los mecanismos reguladores que existen en los organismos. Esto lleva a una síntesis incrementada de los aminoácidos deseada puesto que, por ejemplo, no existen más las regulaciones de retroalimentación al mismo grado o totalmente. Además, puede ser ventajoso conminar una secuencia indicada en SEQ ID NO: 689, con genes que generalmente soportan o potencian el rendimiento del organismo objetivo, por ejemplo, genes que llevan a una rata de crecimiento más rápida de microorganismos o genes que producen plantas resistentes al estrés, patógenos o herbicidas.

En una realización ventajosa adicional del proceso de la invención, los organismos usados en los procesos son aquellos en los cuales simultáneamente al menos uno de los genes antes mencionados o uno de los ácidos nucleicos antes mencionados se hacen mutar de tal manera que la actividad de las proteínas correspondientes está influenciada por metabolitos hasta un nivel más pequeño en comparación con las proteínas no mutadas, o no del todo, y que en particular la acumulación o producción de acuerdo con la invención de nitrógeno o compuestos que contienen nitrógeno respectivamente y una eficiencia incrementada del uso de nitrógeno (NUE), definida como el rendimiento de biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno, no está impedida, o de tal forma que se incrementa su actividad enzimática específica. Menos influencia significa en este contexto que la regulación de la actividad enzimática es menor en al menos 10%, ventajosamente al menos 20, 30 o 40%, particularmente ventajosa en al menos 50, 60, 70, 80 o 90%, en comparación con el organismo de partida, y así la actividad de la enzima se incrementan estas cifras mencionadas en comparación con el organismo estándar. Un incremento en la actividad enzimática significa una actividad enzimática que se incrementa en al menos 10%, ventajosamente al menos 20, 30, 40 o 50%, particularmente de forma ventajosa en al menos 60, 70, 80, 90, 100, 200, 300, 500 o 1000% en comparación con el organismo de partida. Esto lleva a una asimilación, acumulación y/o utilización potenciada de nitrógeno en el organismo.

En una realización adicional ventajosa del proceso de la invención, los organismos usados en el proceso son aquellos en los cuales se atenúa una proteína que degrada un compuesto que contiene N, en particular reduciendo la rata de expresión del gen correspondiente.

En otra realización del proceso de la invención, los organismos usados en el proceso y aquellos en los cuales simultáneamente al menos uno de los ácidos nucleicos antes mencionados o los genes antes mencionados se hace mutar de tal manera que la actividad enzimática biológica de la proteína correspondiente se reduce parcialmente o se bloquea completamente. Una reducción en la actividad enzimática o biológica significa una actividad enzimática, en la cual es reducida en al menos 10%, ventajosamente al menos 20, 30, o 40%, particularmente en forma ventajosa en al menos 50, 60 o 70%, preferiblemente más, en comparación con el organismo de partida.

Si se pretende transformar la célula anfitriona, en particular la célula vegetal, con varios constructos o vectores, el marcador de una transformación precedente debe ser eliminado o emplearse un marcador adicional en una transformación siguiente. Los marcadores pueden ser retirados de la célula anfitriona, en particular la célula vegetal, tal como se describe aquí más adelante a través de métodos con los cuales está familiarizado el técnico experimentado.

5 En plantas particulares sin un marcador, en particular sin resistencia a los antibióticos, son una realización especialmente preferida de la presente invención.

En el proceso de acuerdo con la invención, las secuencias de ácidos nucleicos utilizadas en el proceso de acuerdo con la invención están enlazadas de manera ventajosa operativamente a una o más señales reguladoras con el fin de incrementar la expresión genética. Estas secuencias reguladoras pretenden permitir la expresión genética de los genes y la expresión de la proteína. Dependiendo del organismo anfitrión por ejemplo una planta o microorganismo, esto puede significar, por ejemplo, que el gen está expresado y/o sobreexpresado después de la inducción solamente, o que está expresado y/o sobreexpresado constitutivamente. Estas secuencias reguladoras son, por ejemplo, secuencia en los cuales se enlazan los inductores o represores y los cuales regulan así la expresión del ácido nucleico. Además de estas secuencias reguladoras novedosas, o en vez de estas secuencias, la regulación natural de estas secuencias puede estar presente aún antes de los genes estructurales reales y, si es apropiado, pueden haber sido modificados genéticamente de tal manera que la regulación natural haya sido anulada y la expresión genética haya sido incrementada. Sin embargo, el constructo de ácido nucleico de la invención adecuado como casete de expresión (= constructo de expresión = constructo genético) puede ser también más simple en construcción, es decir no se han insertado señales reguladoras adicionales antes de la secuencia de ácidos nucleicos o sus derivados, y el promotor natural junto con su regulación no ha sido eliminado. En vez de esto, la secuencia reguladora natural ha sido mutada de tal manera que la regulación no tiene lugar más y/o se incrementa la expresión genética. Estos promotores modificados pueden ser introducidos también por sí mismos antes del gen natural en la forma de secuencias parciales (= promotor con partes de secuencias de ácido nucleicos de acuerdo con la invención) con el fin de incrementar la actividad. Además, el constructo del gen puede comprender también ventajosamente uno o más de lo que se conoce como secuencias potenciadoras en el lanzamiento operable con el promotor, y estos permiten una expresión incrementada de la secuencia de ácidos nucleicos. También, es posible insertar secuencias adicionales ventajosas en el extremo 3' de las secuencias de ADN, tal como, por ejemplo, elementos o terminadores reguladores adicionales.

Las moléculas de ácidos nucleicos, que codifican proteínas de acuerdo con la invención y moléculas de ácidos nucleicos, que codifican otros polipéptidos pueden estar presentes en un constructo o vector de ácido nucleico o en varios. Ventajosamente, solamente una copia de la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usado en el método de la invención o sus genes de codificación está presente en el constructo o vector de ácido nucleico. Varios constructos de vectores o ácidos nucleicos pueden ser expresados juntos en el organismo anfitrión. La molécula de ácido nucleico o el constructo o vector de ácido nucleico de acuerdo con la invención pueden ser insertados en un vector y estar presentes en la célula en forma libre. Si se prefiere una transformación estable, se utiliza un vector, el cual se duplica de manera estable a lo largo de varias generaciones o el cual es insertado además en el genoma. En el caso de las plantas, puede haber tenido lugar la integración en el genoma de plástitos o, en particular, en el genoma nuclear. Para la inserción de más de un gen en el genoma anfitrión los genes que se van a expresar están presentes juntos en un constructo genético, por ejemplo en los vectores antes descritos portando una pluralidad de genes.

40 Como regla, las secuencias reguladoras para la tasa de expresión de un gen están localizadas corriente arriba (5'), dentro de y/o corriente abajo (3') con respecto a la secuencia de codificación de la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención u otros segmento genético codogénico. Controla en particular la estabilidad de la transcripción y/o traducción y/o el transcripto. El nivel de expresión depende de la conjunción del sistema de reguladores celulares adicionales tales como la biosíntesis de proteína y los sistemas de degradación de la célula.

Las secuencias reguladoras incluyen secuencias o señales reguladoras de la transcripción y traducción por ejemplo, secuencias localizadas corriente arriba (5'), las cuales son importantes en particular para la regulación de la transcripción o la iniciación de la traducción, tales como promotores o codones de inicio y secuencias localizadas corriente abajo (3'), las cuales conciernen en particular a la regulación de la transcripción o terminación de la traducción y la estabilidad del transcripto, tales como las señales de poliadenilación o codones de detención. Las secuencias reguladoras también pueden estar presentes en regiones de codificación transcritas así como en regiones transcritas sin codificación, por ejemplo, en intrones, como por ejemplo sitios de división. Los promotores para la regulación de la expresión de la molécula de ácido nucleico de acuerdo con la invención en una célula y que pueden ser empleados, son, en principio, todos aquellos que son capaces de estimular la transcripción de genes en los organismos en cuestión, tales como microorganismos o plantas. Promotores adecuados, que son funcionales en estos organismos son conocidos en general. Pueden tomar la forma de promotores constitutivos o inducibles. Promotores adecuados pueden permitir el desarrollo – y/o expresión específica para el tejido en eucariotes de células múltiples; así, pueden utilizarse ventajosamente en plantas promotores específicos para hojas, raíces, flores, semillas, estomas, tubérculos o frutas.

Las secuencias o factores reguladores pueden tener, como se describe anteriormente, un efecto positivo sobre la expresión de los genes introducidos, implementando así su expresión. Así, un potenciamiento de la expresión puede

tener lugar ventajosamente a nivel de transcripción utilizando señales de transcripción fuertes tales como promotores fuertes y/o potenciadores fuertes. Además, el potenciamiento de la expresión a nivel de traducción también es posible, por ejemplo, introduciendo secuencias potenciadoras de la traducción, por ejemplo, el potenciador, que por ejemplo mejora el enlazamiento ribosómico del transcrito, o incrementando la estabilidad del ARNm, por ejemplo reemplazando la región de codificación 3' UTR por una región que codifica un 3'UTR conocido confiriendo una alta estabilidad del transcrito o por estabilización del transcrito a través de la eliminación de la inestabilidad del transcrito, de tal forma que la molécula de ARNm es traducida más frecuentemente que el tipo silvestre. Por ejemplo en plantas los elementos ricos en AU (ARE) y elementos DST (corriente abajo) desestabilizan los transcritos. Los estudios de mutagénesis han demostrado que los residuos dentro de dos de los dominios conservados, las regiones ATAGAT y GTA, son necesarios para la función de inestabilidad. Por lo tanto la eliminación o mutación de tales elemento obviamente llevaría a transcritos más estables, tasas de transcripción más altas y actividad proteínica más alta. Los potenciadores de la traducción son también la "secuencia de superimpulso" la cual comprende la secuencia líder no traducida del 5' del virus mosaico del tabaco y la cual incrementa la relación proteína/ARN (Gallie et al., 1987, Nucl. Acids Research 15:8693-8711).

Los potenciadores se definen en general como elementos activos cis los cuales pueden estimular la traducción genética independiente de la posición y la orientación. Se han identificado diferentes potenciadores en plantas, los cuales pueden estimular la transcripción constitutivamente o específica para tejidos o estímulos. Ejemplos bien conocidos de potenciadores constitutivos son el potenciador del promotor 35S (Odell et al., 1985, Nature 313: 810-812) o el potenciador ocs (Fromm et al., 1989, Plant Cell 1: 977-984). Otros ejemplos son el tetrámero del motivo G-Box el cual confiere expresión constitutiva a alto nivel en plantas dicotiledóneas y monocotiledóneas (Ishige et al., 1999, Plant Journal, 18, 443-448) o el petE, una secuencia rica en A/T que actúa como potenciador cuantitativo de la expresión genética en plantas de tabaco y patatas transgénicas (Sandhu et al., 1998; Plant Mol Biol. 37(5): 885-96). Además de esto, se ha descrito una gran variedad de elementos activos en cis los cuales contribuyen a un patrón de expresión específico, como la expresión específica en órganos o expresión inducida en respuesta a atención biótica o abiótica. Ejemplos son elemento que proveen expresión inducida por patógenos o lesiones (Rushton, 2002, Plant Cell, 14, 749-762) o expresión específica de células guardián (Plesch, 2001, Plant Journal 28, 455-464).

Secuencias reguladoras ventajosas para la expresión de la molécula de ácido nucleico de acuerdo con la invención en microorganismos están presentes por ejemplo en promotores tales como los promotores cos, tac, rha, trp, tet, trp-tet, lpp, lac, lpp-lac, lacI^q, T7, T5, T3, gal, trc, ara, SP6, λ-PR or λ-PL, los cuales se utilizan ventajosamente en bacterias gram negativas. Secuencias reguladoras ventajosa adicionales están presentes por ejemplo en los promotores gram positivos amy, dnaK, xylS and SPO2, en los promotores de levaduras u hongos ADC1, MFα, AC, P-60, UASH, MCB, PHO, CYC1, GAPDH, TEF, rp28, ADH. Los promotores que son particularmente ventajosos, son promotores constitutivos, específicos de tejidos o compartimientos e inducibles. En general, "promotor" se entiende con el significado, en el presente contexto, de una secuencia reguladora en una molécula de ácido nucleico, la cual media en la expresión de un segmento de secuencia de codificación de una molécula de ácido nucleico. En general, el promotor está localizado corriente arriba del segmento de secuencia de codificación. Algunos elementos, por ejemplo elementos de potenciamiento de la expresión tales como un potenciador, pueden, sin embargo, estar localizados también corriente abajo o incluso en la región transcrita.

En principio, es posible utilizar promotores naturales junto con sus secuencias reguladoras, tales como las mencionadas anteriormente, para el proceso novedoso. También es posible ventajosamente utilizar promotores sintéticos, bien sea adicionalmente o solos, en particular cuando median una expresión específica para semillas tales como la descrita en, por ejemplo, WO 99/16890.

La expresión de las moléculas de ácido nucleico usadas en el proceso pueden ser deseada sola o en combinación con otros genes o ácidos nucleicos. Las moléculas de ácido nucleico múltiples que confieren la expresión de genes ventajosos pueden ser introducidas a través de la transformación simultánea de varios constructos de ácidos nucleicos adecuados individuales, esto es, constructos de expresión, o, preferiblemente, combinando varios casetes de expresión en un constructo. También es posible transformar varios vectores con casetes de expresión paso a paso en cada caso en los organismos receptores.

Como se describió anteriormente la transcripción de los genes introducidos debería ser terminada ventajosamente por terminadores adecuados en el extremo 3' de los genes de biosíntesis introducidos (tras el codón de detención). Un terminador, el cual puede ser utilizado para este propósito es, por ejemplo, el terminador OCS1, el terminador nos3 o el terminador 35S. Como sucede en el caso con los promotores, las secuencias determinadoras diferentes deberían ser utilizadas para cada gen. Los terminadores, que son útiles en microorganismos son por ejemplo el terminador fimA, el terminador txn o el terminador trp. Tales terminadores pueden ser dependientes de rho o independientes de rho.

Promotores vegetales diferentes tales como por ejemplo, el promotor USP, el LegB4- y DC3 o el promotor ubiquitina de perejil u otros promotores aquí mencionados son terminadores diferentes que pueden ser utilizados ventajosamente en el constructo de ácido nucleico.

5 Con el fin de asegurar la integración estable, en la planta transgénica, de las moléculas de ácido nucleico usadas en el proceso de acuerdo con la invención en combinación con genes de biosíntesis adicionales sobre una pluralidad de generaciones, cada una de las regiones de codificación usadas en el proceso debería ser expresado bajo el control de su propio promotor, preferiblemente único, puesto que la repetición de motivos de secuencia puede llevar a eventos de recombinación o al silenciamiento o, en plantas, a la inestabilidad del T-ADN.

10 El constructo de ácido nucleico es construido ventajosamente de tal manera que un promotor es seguido por un sitio de escisión adecuado para inserción del ácido nucleico que se va a expresar, ventajosamente en un polienlazante, seguido, si es apropiado, por un terminador localizado tras el polienlazante. Si es apropiado, este orden se repite varias veces de tal manera que se combinan varios genes en un constructo y así pueden ser introducidos en la planta transgénica con el fin de ser expresados. La secuencia es repetida ventajosamente hasta tres veces. Para la expresión, las secuencias de ácidos nucleicos se insertan a través de sitios de escisión adecuado, por ejemplo en el polienlazante tras el promotor. Es ventajoso para cada secuencia de ácido nucleico tener su propio promotor y, si es apropiado, su propio terminador, como se mencionó anteriormente. Sin embargo, también es posible insertar varias secuencias de ácidos nucleicos tras un promotor y, si es apropiado, antes de un terminador si es posible una transcripción policistrónica en las células anfitriónas u objetivo. En este contexto, el sitio de inserción, o la secuencia de las moléculas del ácido nucleico insertadas, en el constructo de ácido nucleico no es decisiva, es decir una molécula de ácido nucleico puede ser insertada en la primera o última posición en el casete sin que esto tenga un efecto sustancial sobre la expresión. Sin embargo, también es posible utilizar solamente un tipo de promotor en el constructo. Sin embargo, esto puede llevar a eventos de recombinación o efectos de silenciamiento indeseados, como se dijo.

20 De acuerdo con lo anterior, en una realización preferida, el constructo de ácido nucleico de acuerdo con la invención confiere la expresión de la molécula de ácido nucleico de la invención o de la molécula de ácido nucleico usada en el método de la invención, y, opcionalmente genes adicionales, en una planta y comprende uno o más elementos reguladores de la planta. Dicho constructo de ácido nucleico de acuerdo con la invención abarca ventajosamente un promotor de planta o un terminador de planta o un promotor de planta y un terminador de planta.

25 Un promotor "de planta" comprende elementos reguladores, los cuales median la expresión de un segmento de una secuencia de codificación en células vegetales. De acuerdo con lo anterior, un promotor vegetal no necesita tener origen vegetal, sino que puede originarse a partir de virus o microorganismos, en particular por ejemplo de virus que atacan las células vegetales.

30 El promotor vegetal también puede originarse a partir de una célula vegetal, como por ejemplo a partir de la planta, la cual es transformada por el constructo o vector de ácido nucleico como se describe aquí.

Esto también se aplica a otras señales reguladoras "vegetales", por ejemplo en terminadores "vegetales".

35 Un constructo de ácido nucleico adecuado para la expresión vegetal comprende preferiblemente elementos reguladores que son capaces de controlar la expresión de los genes en células vegetales y que están enlazados de manera operativa de tal manera que cada secuencia puede satisfacer su función. De acuerdo con lo anterior, el constructo de ácido nucleico también puede comprender terminadores de transcripción. Ejemplos para terminación de transcripción son señales de poliadenilación. Señales de poliadenilación preferidas son aquellas que se originan a partir de T-ADN de *Agrobacterium tumefaciens*, tales como el gen 3 del plásmido Ti pTiACH5, el cual es conocido como octopina sintasa (Gielen et al., EMBO J. 3 (1984) 835 et seq.) o equivalentes funcionales de la misma, pero todos los otros terminales que son funcionalmente activos en plantas también son adecuados.

40 El constructo de ácido nucleico adecuado para la expresión vegetal preferiblemente comprende también otros elementos reguladores enlazados de manera operativa tales como potenciadores de la traducción, por ejemplo la secuencia de sobre impulso, la cual comprende la secuencia líder no traducida en 5' del virus mosaico del tabaco, la cual incrementa la relación proteína/ARN (Gallie et al., 1987, Nucl. Acids Research 15: 8693-8711).

45 Otras secuencias preferidas para el uso en el enlazamiento operativo en constructo de expresión genética son secuencias objetivo, las cuales se requieren para direccionar el producto genético en compartimientos específicos de la célula (para una revisión, véase Kemode, Crit. Rev. Plant Sci. 15, 4 (1996) 285-423 y referencias citadas aquí), por ejemplo en la vacuola, el núcleo, todos los tipos de plástidos tales como amiloplastos, cloroplastos, cromoplastos, el espacio extracelular, las mitocondrias, el retículo endoplásmico, elaioplastos, peroxisomas, glicosomas, y otros compartimientos celulares o extracelulares. Las secuencias, que deben ser mencionadas en este contexto son, en particular, las secuencias que codifican el péptido de señal o el péptido de tránsito que son conocidas por sí mismas. Por ejemplo, la secuencia de codificación del péptido del plástido de tránsito permite el direccionamiento del producto de expresión en los plástidos de una célula vegetal. Las secuencias de direccionamiento también son conocidas para eucariotes y hasta un nivel más bajo para organismos procariotes y pueden ser operables ventajosamente enlazadas con la molécula de ácido nucleico de la presente invención para alcanzar una expresión en uno de dicho compartimientos o extracelularmente. Para expresión en plantas, la molécula de ácido nucleico debe, tal como se describió anteriormente, estar enlazada operativamente a o comprender un promotor adecuado que exprese el gen en el punto del hecho en el tiempo y en una célula en una forma específica para una célula o un tejido. Promotores

utilizable son promotores constitutivos (Benfey et al., EMBO J. 8 (1989) 2195-2202), tales como los que se originan a partir de virus de plantas tales como 35S CAMV (Franck et al., Cell 21 (1980) 285-294), 19S CaMV (see also US 5352605 and WO 84/02913), 34S FMV (Sanger et al., Plant. Mol. Biol., 14, 1990: 433-443), el promotor de ubiquitina de perejil, o promotores vegetales tales como el promotor de subunidad pequeña Rubisco descrito en US 4,962,028 o los promotores vegetales PRP1 [Ward et al., Plant. Mol. Biol. 22 (1993)], SSU, PGEL1, OCS [Leisner (1988) Proc Natl Acad Sci USA 85(5):2553-2557], lib4, usp, mas [Comai (1990) Plant Mol Biol 15 (3):373-381], STLS1, ScBV (Schenk (1999) Plant Mol Biol 39(6):1221-1230), B33, SAD1 or SAD2 (promotores de linaza, Jain et al., Crop Science, 39 (6), 1999: 1696-1701) or nos [Shaw et al. (1984) Nucleic Acids Res. 12(20):7831-7846]. La expresión constitutiva estable de las proteínas de acuerdo con la invención en una planta pueden ser ventajosas. Sin embargo, la expresión inducible del polipéptido de la invención o del polipéptido usado en el método de la invención es ventajosa, si una expresión tardía antes de la recolección es ventajosa, como manipulación metabólica que puede llevar a un retardo en el crecimiento de la planta.

La expresión de los genes vegetales también puede ser facilitada como se describió anteriormente a través de un promotor inducible químico (para una revisión, véase Gatz 1997, Annu Rev. Plant Physiol. Plant Mol. Biol., 48: 89-108). Los promotores inducibles químicamente son particularmente adecuados cuando se desea expresar el gen en una forma específica con el tiempo. Ejemplos de tales promotores son un promotor inducible por ácido salicílico (WO 95/19443) y un promotor inducible por ácido abscísico (EP 335 528) como un promotor inducible por tetraciclina (Gatz et al., (1992) Plant J. 2, 397-404), un promotor inducible por ciclohexanol o etanol (WO 93/21334) u otros tal como se describe aquí.

Otros promotores adecuados son aquellos que reaccionan a condiciones de estrés biótico o abiótico, por ejemplo, el promotor del gen PRP1 inducido por patógenos (Ward et al., Plant. Mol. Biol. 22 (1993) 361-366), el promotor hsp80 de tomate inducible por calor (US 5, 187, 267), el promotor de alfa amilasa en patata inducible por frío (WO 96/12814) o el promotor pinII inducible por lesiones (EP-A-0 375 091) u otros tal como se describen aquí.

Promotores preferidos son en particular aquellos que aportan expresión genética en tejidos y órganos en los cuales la biosíntesis de nitrógeno o compuestos que contienen nitrógeno tales como aminoácidos o proteínas tiene lugar, en células de semillas, tales como células endospermas y células del embrión en desarrollo. Promotores adecuados son el promotor del gen napina de colza (US 5,608,152), el promotor USP de Vicia Faba (Baeumlein et al., Mol Gen Genet, 1991, 225 (3): 459-67), el promotor de oleosina en Arabidopsis (WO 98/45461), el promotor de faseolina en Phaseolus vulgaris (US 5,504,200), el promotor de Bce4 de Brassica (WO 91/13980), el promotor arc5 de judías, el promotor de DcG3 de zanahoria, o el promotor B4 de leguminosas (LeB4; Baeumlein et al., 1992, Plant Journal, 2 (2): 233-9), y promotores que aportan la expresión específica para semillas en plantas monocotiledóneas tales como maíz, cebada, trigo, centeno, arroz y similares. Promotores específicos de semillas ventajosos son el promotor de proteína de enlazamiento de la sacarosa (WO 00/26388), el promotor de faseolina y el promotor de napina. Promotores adecuados que deben ser considerados son el promotor genético lpt2 o lpt1 de cebada (WO 95/15389 y WO 95/23230), y los promotores descritos en WO 99/16890 (promotores del gen de cebada hordeina, el gen de arroz glutelina, el gen de arroz orizina, el gen de arroz prolamina, el gen de trigo gliadina, el gen de trigo glutelina, el gen de maíz zeina, el gen de avena glutelina, el gen de sorgo kasirina y el gen de centeno cecalina). Promotores adecuados adicionales son Amy32b, Amy 6-6 y Aleurain [US 5,677,474], Bce4 (colza) [US 5,530,149], glicinia (soja) [EP 571 741], fosfoenolpiruvato carboxilasa (soja) [JP 06/62870], ADR12-2 (soja) [WO 98/08962], isocitrato iliasa (colza) [US 5,689,040] o α -amilasa (cebada) [EP 781 849]. Otros promotores que son adecuados para la expresión de genes en plantas son promotores específicos para hojas tales como los descritos en DE-A 19644478 o promotores regulados por la luz tales como, por ejemplo, el promotor de guisante petE.

Promotores vegetales adecuados adicionales son el promotor citosólico FBPassa o el promotor de patata ST-LSI (Stockhaus et al., EMBO J. 8, 1989, 2445), el promotor de fosforibosilpirofosfato amidotransferasa de Glicine max (GenBank Accession No. U87999) o el promotor específico para nódulos descrito en EP-A-0 249 676.

Otros promotores, que son particularmente adecuados, son aquellos que aportan expresión específica para plástidos. Promotores adecuados tales como el promotor de polimerasa de ARN viral están descritos en WO 95/16783 y WO 97/06250, y el promotor clpP de Arabidopsis el cual está descrito en WO 99/46394.

Otros promotores, los cuales se usan para la expresión fuerte de secuencias heterólogas en tantos tejidos como sea posible, en particular también en hojas, son, además de varios de los promotores virales y bacteriales antes mencionados, preferiblemente, promotores vegetales de genes de actina o ubiquitina tales como, por ejemplo, el promotor de actina 1 del arroz. Ejemplos adicionales de promotores de plantas constitutivos son los promotores de V-ATPassa de remolacha de azúcar (WO 01/14572). Ejemplos de promotores constitutivos sintéticos son el Super promotor (WO 95/14098) y promotores derivados de cajas G (WO 94/12015). Si es apropiado, puede utilizarse además adicionalmente promotores inducibles químicamente comparando EP-A 388186, EP-A 335528, WO 97/06268.

Como ya se mencionó aquí, las secuencias reguladoras adicionales, que pueden ser expeditivas, si es apropiado, incluyen también secuencias, que apuntan al transporte y/o localización de los productos de expresión. Las secuencias, las cuales deben ser mencionadas en este contexto son, en particular, las secuencias que codifican el péptido de señal

o el péptido de tránsito que son conocidos por sí mismos. Por ejemplo, las secuencias que modifican el péptido de tránsito de plástido permiten el direccionamiento del producto de expresión hacia los plástidos de una célula vegetal.

Plantas receptoras preferidas son, como se describió anteriormente, en particular aquellas plantas que pueden ser transformadas de una manera adecuada. Estas incluyen plantas monocotiledóneas y dicotiledóneas. Las plantas que deben ser mencionadas en particular son plantas útiles en la agricultura tales como ciertos cereales y pastos, por ejemplo *Triticum* spp., *Zea mays*, *Hordeum vulgare*, oats, *Secale cereale*, *Oryza sativa*, *Pennisetum glaucum*, *Sorghum bicolor*, *Triticale*, *Agrostis* spp., *Cenchrus ciliaris*, *Dactylis glomerata*, *Festuca arundinacea*, *Lolium* spp., *Medicago* spp. and *Saccharum* spp., legumbres y cultivos para aceite, por ejemplo *Brassica juncea*, *Brassica napus*, *Glycine max*, *Arachis hypogaea*, *Gossypium hirsutum*, *Cicer arietinum*, *Helianthus annuus*, *Lens culinaris*, *Linum usitatissimum*, *Sinapis alba*, *Trifolium repens* and *Vicia narbonensis*, vegetales y frutos, por ejemplo bananos, uvas, *Lycopersicon esculentum*, espárragos, repollo, melones, fruta de kiwi, *Solanum tuberosum*, *Beta vulgaris*, cassava y chicory, árboles por ejemplo especies de *Coffea*, *Citrus* spp., *Eucalyptus* spp., *Picea* spp., *Pinus* spp. y *Populus* spp., plantas y árboles medicinales y flores.

Una realización de la presente invención también se relaciona con un método para generar un vector, el cual comprende la inserción, en un vector, de la molécula de ácido nucleico caracterizada aquí, la molécula de ácido nucleico de acuerdo con la invención o el casete de expresión de acuerdo con la invención. El vector, por ejemplo, puede ser introducido en una célula, por ejemplo un microorganismo o una célula vegetal, tal como se describe aquí para el constructo de ácido nucleico, o más abajo bajo la transformación o transfección o como se muestra en los ejemplos. Es posible una transformación transiente o estable de la célula anfitrión u objetivo, sin embargo, se prefiere una transformación estable. El vector de acuerdo con la invención es preferiblemente un vector, el cual es adecuado para expresar el polipéptido de acuerdo con la invención en una planta. El método puede así abarcar una o más etapas para integrar las señales reguladoras en el vector, en particular señales que median la expresión en microorganismos o plantas.

De acuerdo con lo anterior, la presente invención también se relaciona con un vector que comprende la molécula de ácido nucleico caracterizada aquí como parte de un constructo de ácido nucleico adecuado para la expresión en plantas o la molécula de ácido nucleico de acuerdo con la invención. Los vectores ventajosos de la invención comprenden las moléculas de ácido nucleico que codifican proteínas de acuerdo con la invención, moléculas de ácido nucleico que se utilizan en el proceso, o constructos de ácido nucleico adecuados para la expresión de las plantas que comprende las moléculas de ácido nucleico usadas, bien sea solas o en combinación con genes adicionales tales como los genes de biosíntesis o reguladores del metabolismo de aminoácidos, por ejemplo con los genes mencionados aquí anteriormente. De acuerdo con la invención, el término "vector" se refiere a una molécula de ácido nucleico, la cual es capaz de transportar otro ácido nucleico al cual está enlazado. Un tipo de vector es un "plásmido", lo cual significa un bucle de ADN de doble cadena circular al cual pueden ligarse segmentos de ADN adicionales. Un tipo adicional de vectores es un vector viral, siendo posible ligar segmentos adicionales de ácidos nucleicos en el genoma viral. Ciertos vectores son capaces de replicación autónoma en una célula anfitriona en la cual han sido introducidos (por ejemplo vectores bacterianos con origen en replicación bacteriana). Otros vectores preferidos están ventajosamente de manera completa o parcial integrados en el genoma de una célula anfitriona cuando son introducidos en la célula anfitriona y pueden así replicarse junto con el genoma anfitrión. Además, ciertos vectores son capaces de controlar la expresión de los genes con los cuales están en enlazamiento operativo. En el presente contexto, estos vectores se denominan como "vectores de expresión". Como se mencionó anteriormente, son capaces de replicación autónoma o puede ser parcial o completamente integrados en el genoma anfitrión. Los vectores de expresión, que son adecuados para técnicas de recombinación de ADN usualmente toman la forma de plásmidos. En la presente descripción, "plásmidos" y "vector" pueden ser utilizados de manera intercambiable puesto que el plásmido está en la forma usada más frecuentemente de un vector. Sin embargo, la invención también pretende abarcar estas otras formas de vectores de expresión, tales como vectores virales, los cuales ejercen funciones similares. El término vector adicionalmente también abarca otros vectores que son conocidos para la persona experimentada, tales como fagos, virus tales como SV40, CMV, TMV, transposones, elementos IS, fasmidos, fagémidos, cósmidos y ADN lineal o circular.

Los vectores de expresión recombinantes que se utilizan ventajosamente en el proceso comprenden las moléculas de ácido nucleico de acuerdo con la invención o el constructo de ácido nucleico de acuerdo con la invención en una forma que es adecuada para la expresión, en una célula anfitriona, de las moléculas de ácido nucleico de acuerdo con la invención o descritas aquí. De acuerdo con lo anterior, los vectores de expresión recombinantes comprenden una o más señales reguladoras seleccionadas sobre la base de las células anfitrionas para ser usadas para la expresión, el enlazamiento operativo con la secuencia de ácido nucleico que se va a expresar.

En un vector de expresión recombinante, "enlazamiento operativo" significa que la molécula de ácido nucleico de interés está enlazada a las señales reguladoras de manera tal que la expresión de la molécula de ácido nucleico es posible: están enlazados uno a otro de tal forma que las dos secuencias satisfacen la función predicha asignada a la secuencia (por ejemplo en un sistema de transcripción/traducción *in vitro*, o en una célula anfitriona si el interés es introducirla en una célula anfitriona).

El término “secuencia reguladora” pretende abarcar promotores, potenciadores y otros elementos de control de expresión (por ejemplo señales de poliadenilación). Estas secuencias reguladoras están descritas, por ejemplo, en Goeddel: Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990), or see: Gruber and Crosby, in: Methods in Plant Molecular Biology and Biotechnology, CRC Press, Boca Raton, Florida, Ed.: Glick and Thompson, chapter 7, 89-108, incluyendo las referencias citadas aquí. Las secuencias reguladoras abarcan aquellas, que controlan la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células anfitrionas y aquellas que controlan la expresión directa de la secuencia de nucleótidos en células anfitrionas específicas solamente, y bajo condiciones específicas. La persona experimentada sabe el diseño del vector de expresión puede depender de factores tales como la selección de la célula anfitriona que va a ser transformada, el grado al cual se expresa la proteína deseada, y similares. Arriba se describe una selección preferida de secuencias reguladoras, por ejemplo, promotores, terminadores, potenciadores y similares. El término secuencia reguladora debe considerarse como abarcado por el término señal reguladora. Varias secuencias reguladoras ventajosas en particular promotores y terminadores se describen más arriba. En general, las secuencias reguladoras descritas como ventajosas para el constructo de ácido nucleico adecuado para la expresión también son aplicables para vectores.

Los vectores de expresión recombinantes usados pueden ser diseñados específicamente para la expresión, en células procariotes y/o eucariotes, de moléculas de ácidos nucleicos usados en el proceso. Esto es ventajoso puesto que las etapas intermedias de la construcción del vector se llevan a cabo frecuentemente en microorganismos en pro de la simplicidad. Por ejemplo, los genes de acuerdo con la invención y otros genes pueden ser expresados en células bacterianas, células de insectos (utilizando vectores de expresión de baculovirus), células de levaduras y otras células fúngicas ([Romanos (1992), *Yeast* 8:423-488; van den Hondel, (1991), in: *More Gene Manipulations in Fungi*, J.W. Bennet & L.L. Lasure, Ed., pp. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. (1991), in: *Applied Molecular Genetics of Fungi*, Peberdy, J.F., et al., Ed., pp. 1-28, Cambridge University Press: Cambridge], algas [Falcitore et al., 1999, *Marine Biotechnology*, 1, 3:239-251] utilizando vectores y siguiendo un método de transformación tal como el descrito en WO 98/01572, y preferiblemente en células de plantas multicelulares [véase Schmidt, R. and Willmitzer, L. (1988) *Plant Cell Rep.*:583-586; *Plant Molecular Biology and Biotechnology*, C Press, Boca Raton, Florida, chapter 6/7, pp.71-119 (1993); F.F. White, in: *Transgenic Plants*, Bd. 1, Engineering and Utilization, Ed.: Kung and R. Wu, Academic Press (1993), 128-43; Potrykus, *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 42 (1991), 205-225 (y referencias citadas aquí)]. Células anfitrionas adecuadas son discutidas adicionalmente en Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Como alternativa, la secuencia del vector de expresión recombinante puede ser transcrita y traducida in vitro, por ejemplo utilizando secuencias reguladoras promotoras T7 y polimerasa T7.

Las proteínas pueden ser expresadas en procariotes utilizando vectores que comprenden promotores constitutivos o inducibles, los cuales controlan la expresión de proteínas de fusión o proteínas de no fusión. Los vectores de expresión de fusión típicos son, entre otros, pGEX (Pharmacia Biotech Inc; Smith, D.B., and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ), en las cuales la glutatona-S-transferasa (GST), la proteína enlazante de la maltosaE o proteína A es fusionada con la proteína objetivo recombinante. Ejemplos de vectores de expresión de *E. coli* de no fusión inducibles adecuados son, entre otros pTrc (Amann et al. (1988) *Gene* 69:301-315) y pET 11 d [Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89]. El gen de expresión objetivo del vector pTrc está basado en la transcripción de un promotor de fusión híbrido trp-lac por la ARN polimerasa anfitriona. La expresión genética objetivo del vector pET 11 d está basada en la transcripción de un promotor de fusión T7-gn10-lac, el cual es mediado por una ARN polimerasa viral coexpresada (T7 gn1). Está polimerasa viral está provista por las cepas anfitrionas BL21 (DE3) o HMS174 (DE3) por un profago λ residente el cual acoge un gen T7 gn1 bajo el control transcripcional del promotor lacUV 5.

Otros vectores que son adecuados en organismos procariotes son conocidos para la persona experimentada en la técnica; estos vectores son por ejemplo en *E. coli* pLG338, pACYC184, the pBR series, tales como pBR322, the pUC series tales como pUC18 or pUC19, the M113mp series, pKC30, pRep4, pHS1, pHS2, pPLc236, pMBL24, pLG200, pUR290, pIN-III113-B1, λ gt11 or pBdCl, in *Streptomyces* pIJ101, pIJ364, pIJ702 or pIJ361, in *Bacillus* pUB110, pC194 or pBD214, in *Corynebacterium* pSA77 or pAJ667.

En una realización adicional, el vector de expresión es un vector de expresión de levadura. Ejemplos de vectores para expresión en las levaduras *S. cerevisiae* abarcan pYeDesaturasec1 (Baldari et al. (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123) and pYES2 (Invitrogen Corporation, San Diego, CA). Los vectores y métodos para la construcción de vectores que son adecuados para uso en otros hongos, tales como los hongos filamentosos, abarcan aquellos que están descritos en detalle en: van den Hondel, C.A.M.J.J. [(1991), J.F. Peberdy, Ed., pp. 1-28, Cambridge University Press: Cambridge; or in: *More Gene Manipulations in Fungi*; J.W. Bennet & L.L. Lasure, Ed., pp. 396-428: Academic Press: San Diego]. Ejemplos de otros vectores adecuados para levaduras son 2mM, pAG-1, YEp6, YEp13 or pEMBLE23.

Vectores adicionales, los cuales pueden ser mencionados a manera de ejemplo, son pALS1, pIL2 o pBB116 en hongos o pLGV23, pGHlac+, pBIN19, pAK2004 or pDH51 en plantas.

Como alternativa, las secuencias de ácido nucleico pueden expresarse en células de insectos utilizando vectores de expresión de baculovirus. Los vectores de baculovirus, que están disponibles para expresar proteínas en células de insectos cultivadas (por ejemplo células Sf9) abarcan la serie pAc (Smith et al. (1983) Mol. Cell Biol. 3:2156-2165) y la serie pVL (Lucklow and Summers (1989) Virology 170:31-39).

- 5 Los vectores antes mencionados son solamente una pequeña revisión de vectores potencialmente adecuados. Plásmidos adicionales son conocidos por las personas experimentadas en la técnica y están descritos, por ejemplo, en: Cloning Vectors (Ed. Pouwels, P.H., et al., Elsevier, Amsterdam-New York-Oxford, 1985, ISBN 0 444 904018). Sistemas de expresión adecuados adicionales para células procariotes y eucariotes, véanse en los capítulos 16 y 17 de Sambrook, J., Fritsch, E.F., and Maniatis, T., Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

De acuerdo con lo anterior, una realización de la invención se relaciona con un vector en donde la molécula de ácido nucleico de acuerdo con la invención es enlazada de manera operativa a secuencias reguladoras que permiten la expresión en una célula procariote o eucariote o en un anfitrión procariote y eucariote.

- 15 De acuerdo con lo anterior, una realización de la invención se relaciona con una célula anfitriona, la cual ha sido transformada de manera estable o transiente con el efector de acuerdo con la invención o la molécula de ácido nucleico de acuerdo con la invención o el constructo de ácido nucleico de acuerdo con la invención.

La presente invención también se relaciona con un proceso para la producción de un polipéptido de acuerdo con la presente invención, siendo expresado el polipéptido en una célula anfitriona de acuerdo con la invención, preferiblemente en un microorganismo o una planta vegetal transgénica.

- 20 En una realización, la molécula de ácido nucleico usada en el proceso para la producción de polipéptido es derivada de un microorganismo, preferiblemente de una célula procariote o protozoica con un organismo eucariote como célula anfitriona. Por ejemplo, en una realización el polipéptido es producido en una célula vegetal o planta con una molécula de ácido nucleico derivada de un procariote o de un hongo o de un alga o de otro microorganismo pero no de otra planta.

- 25 Los experimentados en la técnica saben que la proteína y el ADN expresados en diferentes organismos difieren en muchos aspectos y propiedades, por ejemplo modulación e impresión de ADN, tal como metilación o modificación postraslacional, como por ejemplo glucosilación, fosforilación, acetilación, miristoilación, ADP-ribosilación, farnesilación, carboxilación, sulfatación, ubiquinación, etc., aunque tienen la misma secuencia de codificación. Preferiblemente, el control de expresión celular de la proteína correspondiente difiere de acuerdo con lo anterior en los mecanismos de control que controlan la actividad y expresión de una proteína endógena u otra proteína eucariote. Una diferencia principal entre las proteínas expresadas en organismos procariotes o eucariotes es la cantidad y el patrón de glicosilación. Por ejemplo en la E. coli no hay proteínas glicosiladas. Las proteínas expresadas en levaduras tienen alto contenido de manosa en las proteínas glicosiladas, mientras que en plantas el patrón de glicosilación es complejo.

- 35 El polipéptido de la presente invención se produce preferiblemente mediante técnicas de ADN recombinante. Por ejemplo, una molécula de ácido nucleico que codifica la proteína es clonada en un vector (como se describe más arriba), el vector es introducido en una célula anfitriona (como se describe anteriormente) y dicho polipéptido se expresa en la célula anfitriona. Dicho polipéptido puede ser aislado entonces a partir de las células mediante un esquema de purificación apropiado utilizando técnicas estándar de purificación de proteínas. Como alternativa a la expresión recombinante, el polipéptido o péptido de la presente invención puede ser sintetizado químicamente utilizando técnicas estándar de síntesis de proteínas.

- 40 En una realización, la presente invención se relaciona con un polipéptido que tiene la secuencia de aminoácidos codificada por una molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención u obtenible mediante un proceso de la invención. Dicho polipéptido confiere preferiblemente la actividad antes mencionada, en particular, el polipéptido confiere el incremento en nitrógeno o compuestos que contienen nitrógeno y eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno en una celda o en un organismo o en una parte de los mismos después de incrementar la actividad celular, por ejemplo, incrementando la expresión o la actividad específica del polipéptido.

- 45 En una realización, la presente invención se relaciona con un polipéptido que tiene una secuencia seleccionada del grupo como se indica en SEQ IDE NO: 690.

En una realización ventajosa, en el método de la presente invención se incrementa la actividad de un polipéptido el cual comprende o consiste de una secuencia de consenso seleccionada de grupo como se indica en SEQ ID NO: 951, SEQ ID NO: 952, SEQ ID NO: 953 o SEQ ID NO: 954 y en otra realización, la presente invención se relaciona con un polipéptido que comprende o consiste de una secuencia de consenso como se indica en SEQ ID NO: 951, SEQ ID NO:

952, SEQ ID NO: 953 o SEQ ID NO: 954 con lo cual 20 o menos, preferiblemente 15 o 10, preferiblemente 9, 8, 7 o 6, más preferido 5 o 4, aún más preferido 3, incluso más preferido 2, incluso más preferido 1, lo más preferido 0 de las posiciones de los aminoácidos indicados puede ser reemplazadas por un aminoácido o, en una realización adicional, pueden ser reemplazadas y/o estar ausentes. En una realización, la presente invención se relaciona con el método de la presente invención que comprende un polipéptido o con un polipéptido que comprende más de una secuencia de consenso (de una línea individual) como se indica en SEQ ID NO: 951, SEQ ID NO: 952, SEQ ID NO: 953 o SEQ ID NO: 954.

En una realización no más del 15%, preferiblemente 10%, incluso más preferiblemente 5%, 4%, 3% o 2%, lo más preferiblemente 1% o 0% de la posición de aminoácidos indicados por una letra es reemplazada por otro aminoácido o, en otra realización, está ausente y/o reemplazada. En otra realización los estiramientos de aminoácidos no conservados, indicados por (X), varían en su longitud en 20%, preferiblemente 15% o 10%, incluso más preferiblemente 5%, 4%, 3%, 2% o lo más preferido por solamente 1%.

En una realización, 20 o menos, preferiblemente 15 o 10, preferiblemente 9, 8, 7 o 6, más preferiblemente 5 o 4, incluso aún más preferiblemente 3, incluso aún más preferiblemente 2, incluso aún más preferiblemente 1, lo más preferido 0 aminoácidos son insertados en la secuencia de consenso o, en otra realización, están ausentes y/o son reemplazados.

Las secuencias de consenso fueron derivadas a partir de múltiples alineamientos de las secuencias como se muestra en lista en la Tabla II. Las letras representan el código de aminoácido de una letra e indican que los aminoácidos están conservados en todas las proteínas alineadas. La letra X representa aminoácidos, los cuales no están conservados en todas las secuencias. En un ejemplo, en los casos en donde solamente es posible un pequeño subconjunto seleccionado de aminoácidos en una cierta posición estos aminoácidos están dados entre corchetes. El número de X dados indica las distancias entre los residuos de aminoácidos conservados, por ejemplo Y-x (21-23)-F significa que los residuos de tirosina y fenilalanina conservados están separados uno de otro por un mínimo de 21 y un máximo de 23 residuos de aminoácidos en todas las secuencias investigadas.

Los dominios conservados fueron identificados a partir de todas las secuencias y están descritos utilizando un subconjunto de la anotación Prosite estándar, por ejemplo el patrón Y-x (21-23)-[FW] significa que una tirosina conservada está separada por mínimo 21 y máximo 23 residuos de aminoácidos bien sea de una fenilalanina o un triptófano.

Los patrones conservados son identificados con la herramienta de software MEME versión 3.5.1 o manualmente. La MEME fue desarrollada por Timothy L. Bailey and Charles Elkan, Dept. of Computer Science and Engeneering, University of California, San Diego, USA, y está descrita por Timothy L. Bailey and Charles Elkan [Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994]. El código fuente para el programa individual está disponible públicamente en el centro San Diego Supercomputer (<http://meme.sdsc.edu>).

Para identificar motivos comunes en todas las secuencias con la herramienta de software MEME, se utilizaron los siguientes parámetros: -maxsize 500000, -nmotifs 15, - evt 0.001, -maxw 60, -distance 1e-3, -minsites number de secuencias usada para el análisis. Las secuencias introducidas para MEME no fueron secuencias alineadas en formato Fasta. Se utilizaron otros parámetros en los valores por sistema en esta versión del software.

Para dominios conservados fueron generados con la herramienta de software Pratt versión 2.1 o manualmente. Pratt fue desarrollado por Inge Jonassen, Dept. of Informatics, University of Bergen, Noruega y está descrita por Jonassen et al. [I.Jonassen, J.F.Collins and D.G.Higgins, Finding flexible patterns in unaligned protein sequences, Protein Science 4 (1995), pp. 1587-1595; I.Jonassen, Efficient discovery of conserved patterns using a pattern graph, Submitted to CABIOS Febr. 1997]. El código fuente (ANSI C) para el programa individual está disponible públicamente, por ejemplo, en centros de bioinformática establecidos tales como EBI (European Bioinformatics Institute).

Para generación de patrones con la herramienta de software Pratt, se utilizaron los siguientes parámetros: PL (max Pattern Length): 100, PN (max Nr of Pattern Symbols): 100, PX (max Nr of consecutive x's): 30, FN (max Nr of flexible spacers): 5, FL (max Flexibility): 30, FP (max Flex.Product): 10, ON (max number patterns): 50. Las secuencias de entrada para Pratt fueron regiones distintas de las secuencias de proteínas que exhiben alta similitud según lo identifica la herramienta de software MEME. El número mínimo de secuencias, que tienen que coincidir con los patrones generados (CM, min Nr of Seqs to Match) fue definido en al menos 80% de las secuencias provistas. Los parámetros no mencionados aquí fueron utilizados en sus valores por sistema.

Los patrones Prosite de los dominios conservados pueden ser utilizados para buscar secuencias de proteínas que coinciden con este patrón. Diversos centros de bioinformática establecidos proveen portales de interés públicos para utilizar estos patrones en búsqueda de bases de datos (por ejemplo PIR [Protein Information Resource, located at Georgetown University Medical Center] or ExpPASy [Expert Protein Analysis System]). Alternativamente, hay disponibles

software individual, tal como el programa Fuzzpro, el cual es parte del paquete de software EMBOSS. Por ejemplo, el programa Fuzzpro no solamente permite buscar una coincidencia exacta patrón-proteína sino que también permite definir diversas ambigüedades en la búsqueda ejecutada.

5 El alineamiento fue llevado a cabo con el software ClustalW (versión 1.83) y está descrito por Thompson et al. [Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680]. El código fuente para el programa individual está disponible públicamente en la European Molecular Biology Laboratory; Heidelberg, Alemania. El análisis fue llevado a cabo utilizando los parámetros por sistema de ClustalW v1.83 (gap open penalty: 10.0; gap extension penalty: 0.2; protein matrix: Gonnet; pprotein/DNA endgap: -1; protein/DNA gapdist: 4).
10

En una realización ventajosa, el método de la presente invención comprende el incremento de un polipéptido que comprende o consiste de secuencias consenso específicas de plantas o microorganismos. De acuerdo con lo anterior, en una realización, la presente invención se relaciona con un polipéptido que comprende o consiste de secuencias consenso específicas de plantas o microorganismos.

15 El término "proteína" y "polipéptido" utilizado en esta solicitud son intercambiables. "Polipéptido" se refiere a un polímero de aminoácidos (secuencia de aminoácidos) y no se refiere a una longitud específica de la molécula. Así los péptidos y polipéptidos están incluidos dentro de la definición de polipéptido. Este término también se refiere o incluye modificaciones postraducción del polipéptido, por ejemplo, glicosilaciones, acetilaciones, fosforilaciones y similares. Incluidos dentro de la definición están, por ejemplo, polipéptidos que contienen uno o más análogos de un aminoácido (incluyendo, por ejemplo, aminoácidos no naturales, etc.), polipéptidos con enlaces sustituidos, así como otras modificaciones conocidas en la técnica, ambas de origen natural y de origen no natural.
20

Preferiblemente, el péptido es aislado. Una proteína o molécula de ácido nucleico o porción biológicamente activa de la misma "aislada" o "purificada" está sustancialmente libre de material celular cuando se produce por técnicas de ADN recombinante o precursores químicos u otros agentes químicos cuando se sintetiza químicamente.

25 La expresión "sustancialmente libre de material celular" incluye preparaciones del polipéptido de la invención en el cual la proteína es separada de componentes celulares de las células en las cuales es producida de manera natural o por recombinación. En una realización, la expresión "sustancialmente libre de material celular" incluye preparaciones que tienen menos de aproximadamente 30% (en peso seco) de "proteína contaminante", más preferiblemente menos de aproximadamente 20% de "proteína contaminante", aún más preferiblemente menos de aproximadamente 10% de "proteína contaminante" y lo más preferiblemente menos de aproximadamente 5% de "proteína contaminante". El término "proteína contaminante" se relaciona con polipéptidos, los cuales no son polipéptidos de la presente invención. Cuando el polipéptido de la presente invención o una porción biológicamente activa del mismo es producida por vía recombinante, también está preferiblemente libre de manera sustancial de medio de cultivo, esto es el medio de cultivo representa menos de aproximadamente 20%, más preferiblemente menos de aproximadamente 10%, y lo más preferiblemente menos de aproximadamente 5% del volumen de la preparación proteína. La expresión "sustancialmente libre de precursores químicos u otros agentes químicos" incluye preparaciones en las cuales el polipéptido de la presente invención es separado de precursores químicos u otros agentes químicos, los cuales están involucrados en la síntesis de la proteína. La expresión "sustancialmente libre de precursores químicos u otros agentes químicos" incluye preparaciones que tienen menos de aproximadamente 30% (en peso seco) de precursores químicos o no polipéptidos de la invención – agentes químicos, más preferiblemente menos de aproximadamente 20% de precursores químicos o no polipéptidos de la invención - agentes químicos, aún más preferiblemente menos de aproximadamente 10% de precursores químicos o no polipéptidos de la invención – agentes químicos, y lo más preferiblemente menos de aproximadamente 5% de precursores químicos o no polipéptidos de la invención - agentes químicos. En realizaciones preferidas, las proteínas aisladas o porciones biológicamente activas de las mismas carecen de proteínas contaminantes del mismo organismo del cual se deriva el polipéptido de la presente invención. Típicamente tales proteínas se producen por técnicas recombinantes.
30
35
40
45

Un polipéptido de la invención puede participar en el proceso de la presente invención. El polipéptido o una porción del mismo comprende preferiblemente una secuencia de aminoácidos que es suficientemente homóloga a la secuencia de aminoácidos tal como se indica en SEQ ID NO: 690 de tal manera que la proteína o porción de la misma mantiene la capacidad de conferir la actividad de la presente invención. La porción de la proteína es preferiblemente una porción biológicamente activa tal como se describe aquí. Preferiblemente, el polipéptido usado en el proceso de la invención tiene una secuencia de aminoácidos idéntica a una secuencia tal como se indica en SEQ ID NO: 690.
50

Adicionalmente, el polipéptido puede tener una secuencia de aminoácidos que es codificada por una secuencia de nucleótidos la cual hibrida, preferiblemente hibrida bajo condiciones restrictivas tal como se describió anteriormente, a una secuencia de nucleótidos de la molécula de ácido nucleico de la presente invención. De acuerdo con lo anterior, el polipéptido tiene una secuencia de aminoácidos la cual es codificada por una secuencia de nucleótidos que es al menos 70% o más idéntica a una de las secuencias de nucleótidos tal como se indica en SEQ ID NO: 689. El polipéptido preferido de la presente invención posee preferiblemente al menos una de las actividades de acuerdo con la invención y
55

descritas aquí. Un polipéptido preferido de la presente invención incluye una secuencia de aminoácidos codificada por una secuencia de nucleótidos que hibrida, preferiblemente hibrida bajo condiciones restrictivas, a una secuencia de nucleótidos como se indica en SEQ ID NO: 689 o la cual es homóloga a la misma, como se definió anteriormente.

5 De acuerdo con lo anterior el polipéptido de la presente invención puede variar en una secuencia como se indica en SEQ ID NO: 690 en secuencia de aminoácidos debida a variación natural o mutagénesis, tal como se describe en detalle aquí. De acuerdo con lo anterior, el polipéptido comprende una secuencia de aminoácidos la cual es al menos 70% o más idéntica a la secuencia de aminoácidos completa de una secuencia tal como se indica en SEQ ID NO: 690.

10 Para la comparación de secuencias de aminoácidos pueden utilizarse los mismos algoritmos o secuencias de ácidos nucléicos como se describieron anteriormente. Se alcanzan resultados de alta calidad utilizando el algoritmo de Needleman y Wunsch o Smith y Waterman. Por lo tanto se prefieren programas basados en dichos algoritmos. Ventajosamente las comparaciones de secuencias pueden hacerse con el programa PileUp (J. Mol. Evolution., 25, 351-360, 1987, Higgins et al., CABIOS, 5 1989: 151-153) o preferiblemente con el programa Gap and BestFit, los cuales están basados respectivamente en los algoritmos de Needleman y Wunsch [J. Mol. Biol. 48; 443-453 (1970)] y Smith y Waterman [Adv. Appl. Math. 2; 482-489 (1981)]. Ambos programas son parte del paquete de software GCG [Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1991); Altschul et al. (1997) Nucleic Acids Res. 25:3389 et seq.]. Por lo tanto preferiblemente los cálculos para determinar los porcentajes de homología en la secuencia se hacen con el programa Gap sobre el rango completo de las secuencias. Se utilizaron los siguientes ajustes estándar para la comparación de las secuencias de aminoácidos: peso de la brecha: 8, peso de longitud: 2, coincidencia promedio: 2.912, no coincidencia promedio: - 2.003.

20 Porciones biológicamente activas de un polipéptido de la presente invención incluyen péptidos que comprenden secuencias de aminoácidos derivadas de las secuencias de aminoácidos del polipéptido de la presente invención o usado en el proceso de la presente invención, por ejemplo, una secuencia de aminoácidos como la indicada en SE ID NO: 690 o la secuencia de aminoácidos de una proteína homóloga de la misma, la cual incluye menos aminoácidos que un polipéptido de longitud completa de la presente invención o usada en el proceso de la presente invención o la proteína de longitud completa que es homóloga a un polipéptido de la presente invención o usada en el proceso de la presente invención, representada aquí, y exhibe al menos una actividad de polipéptido de la presente invención o usada en el proceso de la presente invención.

30 Típicamente, las porciones biológicamente (o inmunológicamente) activas, esto es, péptidos, por ejemplo péptidos que tienen, por ejemplo, 5, 10, 20, 30, 35, 36, 37, 38, 39, 40, 50, o 100 más aminoácidos de longitud comprenden un dominio o motivo con al menos con una actividad o epítipo de un polipéptido de la presente invención o usado en el proceso de la presente invención. Además, otras porciones biológicamente activas, en las cuales se eliminan otras regiones del polipéptido, pueden ser preparadas por técnicas recombinantes y evaluarse en cuanto a una o más de las actividades descritas aquí.

35 La manipulación de la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención puede dar como resultado la producción de una proteína que tiene esencialmente la actividad de los polipéptidos tal como se indica en YPR138C pero que tiene diferencias en las secuencia de dicha proteína tipo silvestre. Estas proteínas pueden ser mejoradas en eficiencia o actividad, pueden estar presentes en números más grandes en la célula de lo usual, o pueden disminuir en eficiencia o actividad en relación con la proteína tipo silvestre.

40 Cualquier estrategia de mutagénesis para el polipéptido de la presente invención o el polipéptido usado en el proceso de la presente invención para dar como resultado una dicha actividad creciente no implica ser limitante; variaciones en estas estrategias pueden ser fácilmente evidentes para una persona experimentada en la técnica. Al utilizar tales estrategias, e incorporar los mecanismos divulgados aquí, la molécula de ácido nucleico y el polipéptido de la invención o el polipéptido usado en el método de la invención pueden ser utilizado para generar plantas o partes de las mismas, expresando una o más proteínas tipo silvestre o una o más proteínas con mutación con mutación que codifica moléculas de ácido nucleico o moléculas de polipéptidos de la invención de tal manera que se mejora el rendimiento, producción y/o eficiencia de producción de un compuesto deseado.

50 Este compuesto deseado puede ser cualquier producto natural de las plantas, que incluye los productos finales de las rutas de biosíntesis e intermedios de las rutas metabólicas de origen natural, así como moléculas que no se presentan de forma natural en el metabolismo de dichas células, pero que son producidas por dichas células de la invención. Preferiblemente, el compuesto es una composición que comprende nitrógeno o compuestos que contienen nitrógeno.

Tal como se utiliza aquí, una "proteína quimérica" o "proteína de fusión" comprende un polipéptido enlazado de manera operativa a un polipéptido que no confiere la actividad antes mencionada, en particular, que no confieren un incremento del contenido de nitrógeno o de compuestos que contienen nitrógeno en una célula u organismo o en una parte del mismo, si su actividad se incrementa.

En una realización, una referencia a "proteína (=polipéptido) de la invención" o como se indica en SEQ ID NO: 690 se refiere a un polipéptido que tiene una secuencia de aminoácidos que corresponde al polipéptido de la invención o se usa en el proceso de la invención, mientras que un "no polipéptido de la invención" u "otro polipéptido" que no está siendo indicado en SEQ ID NO: 690 se refiere a un polipéptido que tiene una secuencia de aminoácidos correspondiente a una proteína que no es sustancialmente homóloga a un polipéptido de la invención, preferiblemente que no es sustancialmente homóloga a un polipéptido como se indica en SE ID NO: 690, por ejemplo una proteína que no confiere la actividad descrita aquí o anotada o conocida como indicada en YPR138C y la cual es derivada del mismo o un diferente organismo. En una realización, un "no polipéptido de la invención" u "otros polipéptidos" que no están siendo indicados en SEQ ID NO: 690 no confiere un incremento de nitrógeno o de compuestos que contienen nitrógeno ni eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno en un organismo o parte del mismo.

Dentro de la proteína de fusión, el término "enlazado operativamente" pretende indicar que el polipéptido de la invención o un polipéptido utilizado en el proceso de la invención y el "otro polipéptido" o una parte del mismo se fusionan uno a otro de tal manera que ambas secuencias satisfacen la función propuesta correspondiente a la secuencia usada. El "otro polipéptido" puede ser fusionado con el terminal N o el terminal C del polipéptido de la invención o utilizado en el proceso de la invención. Por ejemplo, en una realización la proteína de fusión es una proteína de fusión GST-LMRP en la cual las secuencias del polipéptido de la invención o el polipéptido usado en el proceso de la invención se fusionan con el terminal C de las secuencias GST. Tales proteínas de fusión pueden facilitar la purificación de los polipéptidos recombinantes de la invención o de un polipéptido útil en el proceso de la invención.

En otra realización, la proteína de fusión es un polipéptido de la invención o un polipéptido utilizado en el proceso de la invención que contiene una secuencia de señal heteróloga en su terminal N. En ciertas células anfitrionas (por ejemplo, células anfitrionas de mamíferos), la expresión y/o secreción de un polipéptido de la invención o de un polipéptido utilizado en el proceso de la invención puede incrementarse a través del uso de una secuencia de señal heteróloga. Como ya se mencionó anteriormente, las secuencias objetivo son requeridas para direccionar el producto genético hacia compartimientos específicos de la células (para una revisión, véase Kermode, Crit. Rev. Plant Sci. 15, 4 (1996) 285-423 y referencias citadas aquí), por ejemplo en la vacuola, el núcleo, todos los tipos de plástidos, tales como amiloplastos, cloroplastos, cromoplastos, el espacio extracelular, las mitocondrias, el retículo endoplásmico, elaioplastos, peroxisomas, glicosomas y otros compartimientos de células o extracelulares. Las secuencias, las cuales pueden ser mencionadas en este contexto son, en particular, las secuencias que codifican péptidos de señal o péptidos de tránsito las cuales son bien conocidas por sí mismas. Por ejemplo, las secuencias que codifican plástido- tránsito, péptido permiten el direccionamiento del producto de excreción hacia los plástidos de una célula vegetal. Las secuencias objetivo son también conocidas para organismos eucariotes y en menor grado procarotes y pueden ventajosamente ser enlazadas de manera operativa con la molécula de ácido nucleico de la presente invención para alcanzar una expresión en dicho compartimiento o extracelular. Por lo tanto en una realización preferida del polipéptido de la invención, específicamente el polipéptido que abarca una secuencia como se muestra en SEQ ID NO: 690 están "enlazamiento operativo" con una secuencia de direccionamiento plástidica, que da como resultado una proteína de fusión funcional, la cual es capaz de dirigir la proteína de fusión al compartimiento plástidico y la cual media en la importación del polipéptido de la invención, específicamente el polipéptido que abarca una secuencia como se muestra en SEQ ID NO: 690, en el compartimiento plástidico.

Preferiblemente, una proteína quimérica o de fusión de la invención es producida por técnicas de ADN recombinante estándar. Por ejemplo, fragmentos de ADN que codifica para las diferentes secuencias de polipéptido son ligadas entre sí en marco de acuerdo con técnicas convencionales, por ejemplo, empleando términos con terminación roma o terminación aguda para ligación, restricción, tratamiento para evitar uniones indeseables, y ligación enzimática. El gen de fusión puede ser sintetizado por técnicas convencionales incluyendo sintetizadores automatizados de ADN. Alternativamente, la amplificación por PCR de fragmentos de genes puede ser llevada a cabo utilizando cebadores ancla, los cuales dan lugar a sobreporciones complementarias entre dos fragmentos genéticos consecutivos los cuales pueden ser fusionados subsecuentemente y reamplificados para generar una secuencia genética quimérica (véase, por ejemplo, Current Protocols in Molecular Biology, eds. Ausubel et al. John Wiley & Sons: 1992). Además, hay disponibles comercialmente muchos vectores de expresión que ya codifican una unidad estructural de fusión (por ejemplo un polipéptido GST). La molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención pueden ser clonadas en tal vector de expresión de tal manera que la unidad estructural de fusión está enlazada en marco a la proteína codificada.

Adicionalmente, las simulaciones de plegamiento y el rediseño por ordenador de los motivos estructurales de la proteína de la invención pueden ser ejecutados utilizando programas de ordenador apropiados (Olszewski, Proteins 25 (1996), 286-299; Hoffman, Comput. Appl. Biosci. 11 (1995), 675-679). La modelación por ordenador del plegamiento de proteínas puede ser utilizada para los análisis conformacional y energético de los modelos de péptido y proteína detallados (Monge, J. Mol. Biol. 247 (1995), 995-1012; Renouf, Adv. Exp. Med. Biol. 376 (1995), 37-45). Pueden utilizarse los programas apropiados para identificación de sitios interactivos del polipéptido de la invención o polipéptidos usados en el proceso de la invención y sus sustratos o factores de enlazamiento u otras proteínas interactuantes mediante búsquedas asistidas por ordenador para secuencias de péptidos complementarios (Fassina, Immunomethods (1994), 114-120). Se describen sistemas adicionales apropiados de ordenador para el diseño de

- proteínas y péptidos en la técnica anterior, por ejemplo en Berry, *Biochem. Soc. Trans.* 22 (1994), 1033-1036; Wodak, *Ann. N. Y. Acad. Sci.* 501 (1987), 1-13; Pabo, *Biochemistry* 25 (1986), 5987-5991. Los resultados obtenidos de los análisis por ordenador antes descritos pueden ser utilizados por ejemplo, para la preparación de peptidomiméticos de la proteína de la invención o fragmentos de los mismos. Tales análogos de seudopéptidos de la secuencia de aminoácidos natural de la proteína pueden imitar muy eficientemente la proteína original (Benkirane, *J. Biol. Chem.* 271 (1996), 33218-33224). Por ejemplo, la incorporación de residuos de aminoácidos Q aquirales fácilmente obtenibles en una proteína de la invención o un fragmento de la misma da como resultado la sustitución de enlaces amida por unidades polimetileno de una cadena alifática, proveyendo por lo tanto una estrategia conveniente para construir un péptidomimético (Banerjee, *Biopolymers* 39 (1996), 769-777).
- 5
- 10 Análogos peptidomiméticos superactivos de hormonas peptídicas pequeñas en otros sistemas están descritos en la técnica anterior (Zhang, *Biochem. Biophys. Res. Commun.* 224 (1996), 327-331). Peptidomiméticos apropiados de la proteína de la presente invención también pueden ser identificados por la síntesis de bibliotecas combinacionales peptidomiméticas a través de alquilación sucesiva de amidas y prueba de los compuestos resultantes, por ejemplo, en cuanto a su enlazamiento y propiedades inmunológicas. Los métodos para la generación y uso de bibliotecas combinacionales peptidomiméticas están descritos en la técnica anterior, por ejemplo en Ostresh, *Methods in Enzymology* 267 (1996), 220-234 and Dorner, *Bioorg. Med. Chem.* 4 (1996), 709-715.
- 15

Adicionalmente, puede utilizarse una estructura tridimensional y/o cristalográfica de la proteína de la invención para el diseño de inhibidores peptidomiméticos de la actividad biológica de la proteína de la invención (Rose, *Biochemistry* 35 (1996), 12933-12944; Rutenber, *Bioorg. Med. Chem.* 4 (1996), 1545-1558).

- 20 Además, una estructura tridimensional y/o cristalina de la proteína de la invención y la identificación de sitios interactivos del polipéptido de la invención o el polipéptido utilizado en el método de la invención y sus sustratos o factores de enlazamiento pueden utilizarse para identificar o diseñar mutantes de enlazamiento modulado o girar sobre otras actividades. Por ejemplo, el centro activo del polipéptido de la presente invención puede ser modelado y los residuos de aminoácidos que participan en la región catalítica pueden ser modulados para incrementar o hacer disminuir el
- 25 enlazamiento del sustrato para activar o mejorar el polipéptido. La identificación del centro activo y los aminoácidos involucrados en la reacción catalítica facilita la selección de mutantes que tienen una actividad incrementada.

Las secuencias mostradas aquí también han sido descritas bajo su nombre de proteína tal como se describe en la Tabla I, II, III o IV, columna 3.

- 30 Los métodos para modificar los niveles de expresión y/o la actividad son conocidos por personas experimentadas en la técnica e incluyen por ejemplo sobreexpresión, cosupresión, el uso de ribozimas, estrategias sentido y antisentido u otras metodologías de silenciamiento genético como interferencia de ARN (iARN) o promotores de la metilación. "Una hebra en sentido" se refiere a una hebra de una molécula de ADN de cadena doble que es homóloga con un transcrito de ARNm de la misma. La "hebra antisentido" contiene una secuencia invertida, la cual es complementaria a la de la "hebra en sentido".
- 35 Además los niveles de expresión y/o la actividad pueden ser modificadas por la introducción de mutaciones en las regiones reguladoras o codificadoras de los ácidos nucleicos de la invención. Adicionalmente pueden expresarse anticuerpos los cuales se enlazan específicamente a un polipéptido de interés y por lo tanto bloquean su actividad. Los factores de enlazamiento a proteína pueden también ser, por ejemplo, aptámeros [Famulok M and Mayer G (1999) *Curr. Top Microbiol. Immunol.* 243: 123-36] o anticuerpos o fragmentos de anticuerpos o anticuerpos de cadena sencilla. Se ha descrito la obtención de estos factores, y la persona experimentada está familiarizada con los mismos. Por ejemplo, se ha empleado un anticuerpo citoplásmico scFv para modular la actividad de la proteína fitocromo A en plantas de tabaco modificadas genéticamente [Owen M et al. (1992) *Biotechnology (NY)* 10(7): 790-794; Franken E et al. (1997) *Curr. Opin. Biotechnol.* 8(4): 411-416; Whitelam (1996) *Trend Plant Sci.* 1: 286-272]. En una realización adicional preferida de la invención el nivel de expresión y/o la actividad pueden cambiarse a través de modificaciones de los reguladores internos. La persona experimentada en la técnica está familiarizada con las diferentes opciones de reguladores internos que pueden ser utilizados para modificar el nivel de expresión de los genes o proteínas de la invención. En una realización de ejemplo un regulador de transcripción negativo, por ejemplo un represor de un ácido nucleico de la invención está inhibido o anulado, por ejemplo a través de inhibición antisentido o ARNi, llevando a una expresión potenciada del ácido nucleico de la invención. De la misma forma un inhibidor alostérico de una proteína de la invención puede estar sometido a subregulación por mutación, o inhibición antisentido o ARNi.
- 40
- 45
- 50

- Una molécula de ácido nucleico "antisentido" comprende una secuencia de nucleótidos, la cual es complementaria de una molécula de ácido nucleico "sentido" que codifica una proteína, por ejemplo, complementaria a la cadena de codificación de una molécula de ADNc de doble cadena o complementaria a una secuencia de ARNm de codificación. De acuerdo con lo anterior, una molécula de ácido nucleico antisentido puede enlazarse a través de puentes de hidrógeno a una molécula de ácido nucleico sentido. La molécula de ácido nucleico antisentido puede ser complementaria a una cadena de codificación completa de una molécula de ácido nucleico que confiere la regulación o expresión del polipéptido de la invención o usado en el proceso de la presente invención, así como en la cadena de
- 55

codificación de la molécula de ácido nucleico de la invención o de la molécula de ácido nucleico usada en la invención, o solamente a una porción de los mismos.

5 Una realización adicional de la invención también se relaciona con un método para la generación de un anfitrión transgénico o célula anfitriona, por ejemplo una célula eucariote o procariote, preferiblemente un microorganismo transgénico, una célula vegetal transgénica o un tejido vegetal transgénico o una planta transgénica, lo cual comprende la introducción, en la planta, la célula vegetal o el tejido vegetal, del constructo de ácido nucleico de acuerdo con la invención, del vector de acuerdo con la invención, o de la molécula de ácido nucleico de acuerdo con la invención.

10 Una realización adicional de la invención también se relaciona con un método para la generación transiente de un anfitrión o una célula anfitriona, célula eucariote o procariote, preferiblemente un microorganismo transgénico, una célula vegetal transgénica o un tejido vegetal transgénico o una planta transgénica, que comprende introducir, en la planta, la célula vegetal o el tejido vegetal, el constructo de ácido nucleico de acuerdo con la invención, el vector de acuerdo con la invención, la molécula de ácido nucleico caracterizada aquí como contenida en el constructo de ácido nucleico de la invención o la molécula de ácido nucleico de acuerdo con la invención, mediante lo cual las moléculas de ácido nucleico, el constructo de ácido nucleico y/o el vector introducidos no se integran en el genoma del anfitrión o la célula anfitriona. Por lo tanto, los transformantes no son estables durante la propagación del anfitrión con respecto a las moléculas de ácido nucleico, constructo de ácido nucleico y/o vector introducidos.

20 En el proceso de acuerdo con la invención, los organismos transgénicos también deben entenderse con el significado de que –si toman la forma de plantas- células vegetales- tejidos vegetales, órganos de plantas tales como raíces, brotes, tallos, semillas, flores, tubérculos u hojas, o plantas intactas que se cultivan para asimilación, acumulación y/o potenciada de nitrógeno y eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno.

25 El conocimiento debe entenderse con el significado de cultivar por ejemplo las células vegetales, tejidos vegetales u órganos vegetales transgénicos sobre o en un medio nutritivo o la planta intacta sobre o en un sustrato, por ejemplo en un cultivo hidropónico, compost para macetas o en un suelo en el campo. En una realización específica el crecimiento se relaciona con el crecimiento de las plantas, células vegetales, tejidos vegetal transgénicos bajo condiciones limitadas de nitrógeno.

30 En una realización ventajosa adicional del proceso, las moléculas de ácido nucleico pueden ser procesadas en células vegetales de célula individual (tales como algas), véase Falciatore et al., 1999, *Marine Biotechnology* 1 (3): 239-251 y referencias citadas allí, y células vegetales de plantas superiores (por ejemplo espermatofitos tales como cultivos). Ejemplos de vectores de expresión vegetal abarcan aquellos que se describen en detalle aquí o en: Becker, D. [(1992) *Plant Mol. Biol.* 20:1195-1197] y Bevan, M.W. [(1984), *Nucl. Acids Res.* 12:8711-8721; *Vectors for Gene Transfer in Higher Plants*; in: *Transgenic Plants, Vol. 1, Engineering and Utilization*, Ed.: Kung and R. Wu, Academic Press, 1993, pp. 15-38]. An overview of binary vectors and their use is also found in Hellens, R. [(2000), *Trends in Plant Science*, Vol. 5 No. 10, 446-451.

35 En una realización de la invención, los vectores de expresión vegetal abarcan aquellos que se describen en las figuras: Figura 1 y/o Figura 2.

40 El vector de ADN puede ser introducido en células procariotes o eucariotes a través de técnicas convencionales de transformación o transfección. Los términos “transformación” y “transfección” incluyen conjugación y transducción y, tal como se utilizan en el presente contexto, pretenden abarcar una multiplicidad de métodos de la técnica anterior para introducir moléculas de ácido nucleico foráneas (por ejemplo ADN) en una célula anfitriona, incluyendo coprecipitación con fosfato de calcio o coprecipitación con cloruro de calcio, transfección mediada por DEAE-dextrano, transfección mediada por PEG, lipofección, competencia natural, transferencia mediada químicamente, electroporación o bombardeo con partículas. Métodos adecuado para la transformación o transfección de células anfitrionas, incluyendo células vegetales, pueden encontrarse en Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) y en otros manuales de laboratorio tales como *Methods in Molecular Biology*, 1995, Vol. 44, *Agrobacterium protocols*, Ed.: Gartland and Davey, Humana Press, Totowa, New Jersey.

45 Los métodos antes descritos para la transformación y regeneración de plantas a partir de tejidos vegetales o células vegetales son explotados para la transformación transiente o estable de plantas. Métodos adecuados son la transformación de protoplastos por consumo de ADN inducido por polietileno glicol, el método biolístico con la pistola de genes –conocida como método de bombardeo con partículas-, electroporación, la incubación de embriones secos en una solución que contiene ADN, microinyección y la transferencia genética mediada por *Agrobacterium*. Los métodos antes mencionados están descritos, por ejemplo en B. Jené, *Techniques for Gene Transfer*, in: *Transgenic Plants, Vol. 1, Engineering and Utilization*, edited by S.D. Kung and R. Wu, Academic Press (1993) 128-143 and in Potrykus *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 42 (1991) 205-225. El constructo que se va a expresar se clona preferiblemente en un vector, el cual es adecuado para transformar *Agrobacterium tumefaciens*, por ejemplo pBin19 (Bevan, *Nucl. Acids Res.* 12 (1984) 8711). Las agrobacterias transformadas con tal vector pueden ser utilizadas entonces de la manera

conocida para la transformación de plantas, en particular plantas de cultivo, tales como, por ejemplo, plantas de tabaco, por ejemplo bañando hojas escarificadas o segmentos de hojas en una solución agrobacteriana y subsecuentemente cultivándolas en medios adecuados. La transformación de las plantas con *Agrobacterium tumefaciens* es descrita por ejemplo por Höfgen and Willmitzer in *Nucl. Acid Res.* (1988) 16, 9877 o conocida de, entre otros, F.F. White, *Vectors for Gene Transfer in Higher Plants*; in *Transgenic Plants*, Vol. 1, Engineering and Utilization, edited by S.D. Kung and R. Wu, Academic Press, 1993, pp. 15-38.

Para seleccionar la transferencia exitosa de la molécula de ácido nucleico, vector o constructo de ácido nucleico de la invención de acuerdo con la invención hacia un organismo anfitrión, es ventajoso utilizar genes marcadores como ya han sido descritos anteriormente en detalle. Se sabe de la integración estable o transiente de ácidos nucleicos en células vegetales que solamente una minoría de las células consumen el ADN foráneo y, si se desea, lo integran en su genoma, dependiendo del vector de expresión utilizado y de la técnica de transfección usada. Para identificar y seleccionar estos integrantes, se introduce usualmente un gen que codifica para un marcador seleccionable (como se describió anteriormente, por ejemplo resistencia a antibióticos) en las células anfitrionas junto con el gen de interés. Marcadores seleccionables preferidos en plantas comprenden aquellos que confieren resistencia a un herbicida tal como glifosato o glufosinato. Otros marcadores adecuados son, por ejemplo, marcadores que codifican genes involucrados en rutas biosintéticas, por ejemplo, azúcares o aminoácidos, tales como β -galactosidasa, *ura3*, *ilv2* o un gen de acetohidroxiácidos sintasa mutado (AHAS), también conocido como gen de acetolactato sintasa (ALS) o un gen para una enzima metabolizante de un D-aminoácido. Los marcadores que codifican genes tales como la luciferasa, *gfp* u otros genes de florecencia, son probablemente adecuados. Marcadores adicionales nombrados en la literatura algunas veces como marcadores secundarios, genes que codifican para la resistencia contra herbicidas tales como la fosfotrisina (=glufosinato, BASTA™, Liberty™, codificados por el gen de barra), glifosato (= N-(fosfonometil)glicina, Roundup Ready™, codificado por el gen 5-enolpiruvilshikimato-3-fosfato sintasa = *epsps*), sulfonilurea (= Staple™, codificado por el gen acetolactato sintasa), imidazolinona [= IMI, imazethapyr, imazamox, Clearfield™, codificado por el gen acetohidroxiácido sintasa (AHAS), también conocido como el gen acetolactato sintasa (ALS)] o bromoxinil (= Buctril™, codificado por el gen *oxy*) o genes que codifican para antibióticos tales como higromicina o G418 son útiles para la selección. Además marcadores de selección negativos tales como la citosina desaminasa bacteriana (codificada por el gen *codA*) también son útiles para la transformación de plástidos.

Estos marcadores y los marcadores antes mencionados pueden ser utilizados en mutantes en los cuales estos genes no son funcionales puesto que, por ejemplo, han sido eliminados por métodos convencionales. Adicionalmente, las moléculas de ácido nucleico, que codifican un marcador seleccionable, pueden ser introducidas en una célula huésped sobre el mismo vector como aquellos que codifican los polipéptidos de la invención o se utilizan en el proceso o aún más en un vector separado. Las células que han sido transfectadas de manera estable con el ácido nucleico introducido pueden ser identificadas por ejemplo por selección (por ejemplo, células que han integrado el marcador seleccionable sobreviven mientras que las otras células mueren).

Puesto que los genes marcadores, como regla específicamente el gen para resistencia a antibióticos y herbicidas, no son ya requeridos o son indeseables en la célula anfitriona transgénica una vez que los ácidos nucleicos han sido introducidos exitosamente, el proceso de acuerdo con la invención para introducir los ácidos nucleicos emplea ventajosamente técnicas que permiten la eliminación, o escisión de estos genes marcadores. Tal método es conocido como cotransformación. El método de cotransformación emplea dos vectores simultáneamente para la transformación, un vector que porta el ácido nucleico de acuerdo con la invención y un segundo que porta el gen marcador. Una proporción grande de transformantes recibe o, en el caso de plantas, comprende (hasta el 40% de los transformantes y más), ambos vectores. En el caso de transformación con agrobacteria, los transformantes usualmente reciben solamente una parte del vector, la secuencia flanqueada por el T-ADN, el cual representa usualmente el casete de expresión. Los genes marcadores pueden ser eliminados subsecuentemente de la planta transformada llevando a cabo cruzamientos. En otro método, los genes marcadores integrados en un transposon se utilizan para la transformación junto con el ácido nucleico deseado (conocido como la tecnología Ac/Ds). Los transformantes pueden ser cruzados con un recurso de transposasa o los transformantes se transforman con un constructo de ácido nucleico que confiere expresión de una transposasa, de forma transiente o estable. En algunos casos (aproximadamente el 10%), el trasposón salta fuera del genoma de la célula anfitriona una vez que la transformación ha tenido lugar exitosamente y se pierde. En un número adicional de casos, el trasposon salta a una diferente localización. En estos casos, el gen marcador debe ser eliminado llevando a cabo cruces. En microbiología, se desarrollaron técnicas que hacen posible, o facilitan, la detección de tales eventos. Un método ventajoso adicional se basa en lo que se conoce como sistemas de recombinación, cuya ventaja es que la eliminación por cruzamiento puede ser obviada. El sistema mejor conocido de este tipo es el denominado sistema Cre/lox. La Cre 1 es una recombinasa, la cual elimina la secuencia localizadas entre las secuencias loxP. Si el gen marcado está integrado entre las secuencias loxP, se retira, una vez que la transformación ha tenido lugar exitosamente, por expresión de la recombinasa. Sistemas de recombinaciones adicionales son los sistemas HIN/HIX, FLP/FRT y REP/STB (Tribble et al., *J. Biol. Chem.*, 275, 2000: 22255-22267; Velmurugan et al., *J. Cell Biol.*, 149, 2000: 553-566). Es posible una intervención específica para el sitio en el genoma vegetal de las secuencias de ácido nucleico de acuerdo con la invención. Naturalmente, estos métodos también pueden ser aplicados a microorganismos tales como levaduras, hongos o bacterias.

Las agrobacterias transformadas con un vector de expresión de acuerdo con la invención también pueden ser utilizadas en la manera conocida por sí misma para la transformación de plantas tales como plantas experimentales como Arabidopsis o plantas de cultivo, tales como, por ejemplo, cereales, maíz, avena, centeno, cebada, trigo, soja, arroz, algodón, remolacha de azúcar, canola, girasol, linaza, cáñamo, patata, tabaco, tomate, zanahoria, pimientos, colza, tapioca, cassava, raíz flecha, tagetes, alfalfa, lechuga y las diversas especies de árboles, nueces y vides, en particular plantas de cultivo que contienen aceite tales como soja, cacahuete, planta del aceite de castor, girasol, maíz, algodón, linaza, colza, algodón, coco, palma de aceite, cártamo (*Carthamus tinctorius*) o de granos de cacao, por ejemplo, bañando hojas escarificadas o segmentos de hojas en una solución agrobacteriana y subsecuentemente cultivándolos en medios adecuados.

Además de la transformación de las células somáticas, las cuales entonces tienen que ser regeneradas en plantas intactas, también es posible transformar las células de meristemas vegetales y en particular aquellas células que se desarrollan en gametos. En este caso, los gametos transformados siguen el desarrollo natural de la planta, dando lugar a plantas transgénicas. Así, por ejemplo, las semillas de Arabidopsis son tratadas con agrobacterias y se obtienen semillas de las plantas en desarrollo y de las cuales una cierta proporción es transformada y por lo tanto transgénica (Feldman, KA and Marks MD (1987). *Mol Gen Genet* 208:274-289; Feldmann K (1992). In: C Koncz, N-H Chua and J Shell, eds, *Methods in Arabidopsis Research*. World Scientific, Singapore, pp. 274-289). Los métodos alternativos se basan en la eliminación repetida de las inflorescencias y la incubación del sitio de escisión en el centro de la roseta con agrobacterias transformadas, por lo cual las semillas transformadas pueden probablemente ser obtenidas en un punto posterior en el tiempo (Chang (1994). *Plant J.* 5: 551-558; Katavic (1994). *Mol Gen Genet*, 245: 363-370). Sin embargo, un método especialmente efectivo es el método de infiltración al vacío con sus modificaciones tales como el método de "inmersión floral". En el caso de infiltración al vacío de Arabidopsis, se tratan plantas intactas bajo presión reducida con una suspensión bacteriana. (Bechthold, N (1993). *C R Acad Sci Paris Life Sci*, 316: 1194-1199), mientras que en el caso del método de "inmersión floral" el tejido floral en desarrollo es incubado brevemente con una suspensión agrobacteriana tratada con surfactante (Clough, SJ und Bent, AF (1998). *The Plant J.* 16, 735-743). Se recolecta cierta proporción de semillas transgénicas en ambos casos, y estas semillas pueden ser distinguidas de semillas no transgénicas cultivándolas bajo las condiciones selectivas antes descritas. Además la transformación estable de los plástidos es ventajosa porque los plástidos son heredados por vía materna en la mayoría de los cultivos reduciendo o eliminando el riesgo del flujo transgénico a través del polen. La transformación del genoma de cloroplastos se logra generalmente por un proceso que ha sido desplegado esquemáticamente en Klaus et al., 2004 (*Nature Biotechnology* 22(2), 225-229). En resumen las secuencias que se van a transformar son clonadas junto con un gen marcador seleccionable entre secuencias flanqueantes homólogas al genoma de cloroplasto. Estas secuencias homólogas flanqueantes dirigen la integración específica en sitio hacia el plastoma. La transformación plástidica ha sido descrita para muchas diferentes especies de plantas y puede encontrarse una revisión en Bock (2001) *Transgenic plastids in basic research and plant biotechnology*. *J Mol Biol.* 2001 Sep 21; 312 (3): 425-38 or Maliga, P (2003) *Progress towards commercialization of plastid transformation technology*. *Trends Biotechnol.* 21, 20-28. Se ha reportado recientemente un progreso biotecnológico adicional en la forma de transformantes de plástidos libres de marcadores, los cuales pueden ser producidos por genes marcadores cointegrados transientes (Klaus et al., 2004, *Nature Biotechnology* 22 (2), 225-229).

Las células vegetales modificadas genéticamente pueden ser regeneradas a través de todos los métodos con los cuales la persona experimentada está familiarizada. Métodos adecuados pueden encontrarse en las publicaciones antes mencionadas por S.D. Kung and R. Wu, Potrykus or Höfgen and Willmitzer.

De acuerdo con lo anterior, la presente invención también se relaciona así con una célula vegetal que comprende el constructo de ácido nucleico de acuerdo con la invención, la molécula de ácido nucleico de acuerdo con la invención o el vector de acuerdo con la invención.

"Transgénico" como por ejemplo con respecto a una molécula de ácido nucleico, un constructo de ácido nucleico o un vector que comprende dicha molécula de ácido nucleico o un organismo transformado con dicha molécula de ácido nucleico, constructo de ácido nucleico o vector, se refiere a todos los objetos originados por métodos recombinantes en los cuales

a) la secuencia de ácido nucleico, o

b) una secuencia de control genético enlazado operativamente a la secuencia de ácido nucleico, por ejemplo un promotor, o

c) (a) y (b)

No están localizadas en su ambiente genético natural o ha sido modificada por métodos recombinantes, siendo un ejemplo de modificación una sustitución, adición, eliminación, inversión, inserción de uno o más residuos de nucleótidos. Ambiente genético natural se refiere al locus cromosómico natural en el organismo de origen, o a la presencia en una biblioteca genómica. En el caso de una biblioteca genómica, el ambiente genético natural de la secuencia de ácido nucleico es retenido preferiblemente, al menos en parte. El ambiente flanquea la secuencia de ácido nucleico al menos

a un lado y tiene una secuencia de al menos 50 bp, preferiblemente al menos 500 bp, especialmente de manera preferible al menos 1000 bp, muy especialmente de manera preferible al menos 5000 bp, de longitud.

Un casete de expresión de origen natural –por ejemplo la combinación de origen natural de un promotor de un polipéptido de la invención con la correspondiente secuencia codificadora de proteína- se convierte en un casete de expresión genética cuando es modificado por métodos no naturales, sintéticos “artificiales” tales como, por ejemplo, mutagenización. Tales métodos han sido descritos (US 5,565,350; WO 00/15815; también véase más arriba).

Adicionalmente, la célula vegetal, tejido vegetal o planta también pueden ser transformados de tal manera que se (sobre)expresen enzimas y proteínas adicionales cuya expresión soporta un incremento de nitrógeno o compuestos que contienen nitrógeno y una eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno.

Sin embargo, transgénico también significa que los ácidos nucleicos de acuerdo con la invención están localizados en su posición natural en el genoma de un organismo, pero que la secuencia ha sido modificada en comparación con la secuencia natural y/o que las secuencias reguladoras de las secuencias naturales han sido modificadas. Preferiblemente, transgénico/recombinante debe entenderse con el significado de que la transcripción de los ácidos nucleicos usados en el proceso de acuerdo con la invención ocurren en una posición no natural en el genoma, esto es la expresión de los ácidos nucleicos es homóloga o, preferiblemente heteróloga. Está expresión puede ser transiente o de una secuencia integrada de manera estable en el genoma.

El término “plantas transgénicas” utilizadas de acuerdo con la invención, también se refiere a la progenie de una planta transgénica, por ejemplo, las T₁, T₂, T₃ y subsecuentes generaciones o las BC₁, BC₂, BC₃ y subsecuentes generaciones de plantas. Así, las plantas transgénicas de acuerdo con la invención pueden ser cultivadas e individualizadas o cruzadas con otros individuos con el fin de obtener plantas transgénicas adicionales de acuerdo con la invención. Las plantas transgénicas también pueden ser obtenidas por la propagación vegetativa de células de plantas transgénicas. La presente invención también se relaciona con material de plantas transgénicas, lo cual puede ser derivado de una población de plantas transgénicas de acuerdo con la invención. Tal material incluye células vegetales y ciertos tejidos, órganos y partes de las plantas en todas sus manifestaciones, tales como semillas, hojas, anteros, fibras, tubérculos, raíces, raíces vellosas, tallos, embriones, callos, cotiledones, peciolo, material recolectado, tejido vegetal, tejido reproductivo y cultivos celulares, los cuales se derivan de la planta transgénica real y/o pueden ser utilizados para configurar la planta transgénica.

Cualquier planta transformada obtenida de acuerdo con la invención puede ser utilizada en un esquema de cultivo convencional o de propagación de plantas in vitro para producir más plantas transformadas con las mismas características y/o puede ser utilizado para introducir la misma característica en otras variedades de la misma especie o relacionadas. Tales plantas son también parte de la invención. Las semillas obtenidas a partir de las plantas transformadas genéticamente también contienen la misma característica y son parte de la invención. Como se mencionó anteriormente, la presente invención es en principio aplicable a cualquier planta o cultivo que pueda ser transformado por cualquiera de los métodos de transformación conocidos para los experimentados en el arte.

En una realización especialmente preferida, el organismo, la célula anfitriona, la célula vegetal, planta y microorganismo o tejido vegetal de acuerdo con la invención son transgénicos.

De acuerdo con lo anterior, la invención se refiere por lo tanto a organismos transgénicos transformados por al menos una molécula de ácido nucleico, constructo de ácido nucleico o vector de acuerdo con la invención, y con células, cultivos celulares, tejidos, partes –tales como, por ejemplo, en el caso de organismos vegetales, tejido vegetal, por ejemplo hojas, raíces y similares- o material de propagación derivado de tales organismos, o plantas intactas. Los términos “(anfitrión) recombinante”, y “(anfitrión) transgénico” se utilizan de manera intercambiable en este contexto. Naturalmente, estos términos se refieren no solamente al organismo anfitrión o a la célula objetivo en cuestión, sino también a la progenie, o progenie potencial, de estos organismos o células. Puesto que pueden ocurrir ciertas modificaciones en generaciones subsecuentes obedeciendo la mutación o a los efectos ambientales, tal progenie no es necesariamente idéntica a la célula de origen, pero aún viene con el alcance del término tal como se utiliza aquí.

Organismos adecuados para el proceso de acuerdo con la invención o como anfitriones son todos aquellos organismos eucariote o procariote, los cuales son capaces de acumular o asimilar nitrógeno o compuestos que contienen nitrógeno respectivamente. Los organismos utilizados como anfitriones son microorganismos, tales como bacterias, hongos, levaduras o algas, animales no humanos, o plantas tales como plantas dicotiledóneas o monocotiledóneas.

En principio todas las plantas pueden ser utilizadas como organismo anfitrión, especialmente las plantas mencionadas anteriormente como organismos fuente. Plantas transgénicas preferidas son, por ejemplo, seleccionadas de las familias Aceraceae, Anacardiaceae, Apiaceae, Asteraceae, Brassicaceae, Cactaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, Malvaceae, Nymphaeaceae, Papaveraceae, Rosaceae, Salicaceae, Solanaceae, Arecaceae, Bromeliaceae, Cyperaceae, Iridaceae, Liliaceae, Orchidaceae, Gentianaceae, Labiaceae, Magnoliaceae, Ranunculaceae, Carifolaceae,

Rubiaceae, Scrophulariaceae, Caryophyllaceae, Ericaceae, Polygonaceae, Violaceae, Juncaceae or Poaceae and preferably from a plant selected from the group of the families Apiaceae, Asteraceae, Brassicaceae, Cucurbitaceae, Fabaceae, Papaveraceae, Rosaceae, Solanaceae, Liliaceae or Poaceae. Se prefieren plantas de cultivo tales como plantas seleccionadas ventajosamente del grupo del género cacahuete, colza, canola, girasol, cártamo, oliva, sésamo, avellana, almendra, aguacate, laurel, calabacín/calabaza, linaza, soja, pistacho, borraja, maíz, trigo, centeno, avena, sorgo y mijo, triticale, arroz, cebada, cassava, patata, remolacha de azúcar, berenjena, alfalfa y pastos perennes y plantas para forraje, palma de aceite, vegetales (brassicáceas), vegetales de raíz, vegetales tuberosos, vegetales de vaina, vegetales de frutales, vegetales de cebolla, vegetales para hojas y vegetales de tallo), trigo sarraceno, alcachofa de Jerusalén, judía ancha, algarrobas, lenteja, judía enana, lupino, ajo y lucerna para mencionar solamente algunos de ellos.

Células vegetales, órganos vegetales, tejidos vegetales o partes de plantas preferidos se originan de las familias mencionadas de organismos fuente, preferiblemente de los géneros de plantas antes mencionados, más preferidos de las especies de plantas antes mencionadas.

Las plantas transgénicas que comprenden el nitrógeno o los compuestos que contienen nitrógeno y una eficiencia incrementada en el uso del nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, por ejemplo los aminoácidos o proteínas sintetizadas en el proceso de acuerdo con la invención pueden ser comercializados directamente sin aislamiento de los compuestos sintetizados. En el proceso de acuerdo con la invención, las plantas se definen con el significado de todas las partes de las plantas, órganos de plantas tales como hojas, tallo, raíces, tubérculos o semillas o material de propagación o material recolectado o la planta intacta. En este contexto, las semillas abarcan todas las partes de las semillas tales como los recubrimientos de semillas, células epidérmicas, células de semilla, endospermos o tejido embrionario. El nitrógeno acumulado o asimilado en el proceso de acuerdo con la invención sin embargo puede ser aislado también de la planta en la forma de sus aminoácidos libres o enlazados en proteínas. Los aminoácidos producidos por este proceso pueden ser recolectados recolectando los organismos bien sea de un cultivo en el cual crecen o desde el campo. Esto puede hacerse a través de expresión, trituración y/o extracción, precipitación con sal y/o cromatografía de intercambio iónico de las partes de la planta, preferiblemente de las semillas de la planta, frutos de la planta, tubérculos de la planta y similares.

En aún otro aspecto, la invención también se relaciona con partes recolectables y con material de propagación de las plantas transgénicas de acuerdo con la invención las cuales contienen células de la planta transgénica que expresan una molécula de ácido nucleico de acuerdo con la invención o que contienen células que muestran una actividad celular incrementada del polipéptido de la invención o del polipéptido utilizado en el método de la invención, por ejemplo, un nivel de expresión incrementado o una actividad más alta de la proteína descrita.

Las partes recolectables pueden en principio ser cualquier parte útil de una planta, por ejemplo, flores, polen, plantones, tubérculos, hojas, tallos, frutas, semillas, raíces, etc. El material de propagación incluya, por ejemplo, semillas, frutos, cortes, plantones, tubérculos, raíces, etc. Se prefieren semillas, frutas, plantones o tubérculos como material recolectable o de propagación.

La invención se relaciona adicionalmente con el uso de organismos transgénicos de acuerdo con la invención y de las células, cultivos de células, partes –tales como por ejemplo, raíces, hojas y similares tales como se mencionaron anteriormente en el caso de organismos vegetales transgénicos- derivados de ellas, y con material de propagación transgénico tal como semillas o frutos y similares como se mencionó anteriormente, para la producción de materiales alimenticios o materiales para forraje, productos farmacéuticos o productos químicos finos, preferiblemente materiales para forraje.

De acuerdo con otra realización, la presente invención se relaciona con el uso de la molécula de ácido nucleico, el organismo, por ejemplo el microorganismo, la planta, células de la planta o tejidos de la planta, el vector, o el polipéptido de la presente invención para hacer ácidos grasos, carotinoides, isoprenoides, vitaminas, lípidos, ésteres cerosos (poli) sacáridos y/o polihidroxicanoatos, y/o sus productos del metabolismo, en particular, hormonas esteroides, colesterol, prostaglandina, triacilglicerol, ácidos biliares y/o cuerpos cetónicos que producen células, tejidos y/o plantas. Hay un cierto número de mecanismos mediante los cuales se pueden afectar el rendimiento, producción y/o eficiencia de la producción de ácidos grasos, carotenoides, isoprenoides, vitaminas, ésteres cerosos, lípidos, (poli) sacáridos y/o polihidroxicanoatos, y/o sus productos de metabolismo, en particular, hormonas esteroides, colesterol, triacilglicerol, prostaglandina, ácidos biliares y cuerpos cetónicos o adicionalmente a los productos químicos finos definidos anteriormente que incorporan tal proteína alterada. En el caso de plantas, por ejemplo por incremento de la expresión del acetil CoA la cual es la base para muchos productos, por ejemplo, ácidos grasos, carotenoides, isoprenoides, vitaminas, lípidos, (poli) sacáridos, ésteres cerosos y/o polihidroxicanoatos, y/o subproductos del metabolismo, en particular, prostaglandina, hormonas esteroidales, colesterol, triacilglicerol, ácidos biliares y/o cuerpos cetónicos en una células, puede ser posible incrementar la cantidad de dichos compuestos producidos permitiendo así una facilidad mayor para recolectar y purificar o en caso de plantas más eficientes para la partición. Adicionalmente, se pueden requerir uno o más de dichos productos del metabolismo, en cantidades incrementadas de los cofactores, moléculas precursoras y compuestos intermediarios para las rutas biocientíficas apropiadas. Por lo tanto, al incrementar el número

- 5 y/o actividad de las proteínas transportadoras involucradas en la importación de nutrientes, tales como fuentes de carbono (azúcares) fuentes de nitrógeno (esto es aminoácidos o sales de amonio), fosfato y azufre, puede ser posible mejorar la producción de acetil CoA y sus productos de metabolismo como se mencionó anteriormente, debido a la eliminación de cualquier limitación en el suministro de nutrientes para el proceso biosintético. En particular, puede ser posible incrementar el rendimiento, producción y/o eficiencia de la producción de dichos compuestos, por ejemplo, ácidos grasos, carotenoides, isoprenoides, vitaminas, ésteres cerosos, lípidos, (poli) sacáridos, y/o polihidroxicarboxilatos, y/o sus productos de metabolismo, en particular hormonas esteroidales, colesterol, prostaglandina, triacigliceroles, ácidos biliares y/o moléculas de cuerpos cetónicos, etc. en plantas.
- 10 Los organismos, preferiblemente las plantas de la presente invención muestran asimilación, acumulación y/o utilización potenciada de nitrógeno y eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno incluso bajo condiciones de nitrógeno limitado. En consecuencia los organismos, preferiblemente las plantas de la presente invención muestran también crecimiento, biomasa y/o rendimiento potenciados, preferiblemente bajo condiciones de nitrógeno limitado.
- 15 En una realización, la presente invención se relaciona con un método para la identificación de un producto genético que confiere una asimilación, acumulación y/o utilización potenciada de nitrógeno en un organismo y una eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes de nitrógeno, que comprende las siguientes etapas:
- 20 Poner en contacto, por ejemplo hibridando, las moléculas de ácido nucleico de una muestra, como por ejemplo células, tejidos, plantas o microorganismos o una biblioteca de ácidos nucleicos, la cual puede contener un gen candidato que codifique un producto genético que confiere una asimilación potenciada de nitrógeno, asimilación y/o utilización después de la expresión, con la molécula de ácido nucleico de la presente invención.
- Identificar las moléculas de ácido nucleico, que hibridan bajo condiciones relajadas con la molécula de ácido nucleico de la invención, acumulación y/o utilización de la expresión, con la molécula de ácido nucleico de la presente invención;
- 25 Identificar las moléculas de ácido nucleico que hibridan bajo condiciones de restricción relajadas con la molécula de ácido nucleico de la presente invención en particular con una secuencia de molécula de ácido nucleico tal como se indica en SEQ ID NO: 689 y opcionalmente, aislar el ADNc de longitud completa como clon o como clon genómico completo;
- Introducir las moléculas de ácidos nucleicos candidatos a células anfitrionas, preferiblemente en una célula vegetal o un microorganismo, apropiados para la asimilación, acumulación y/o utilización potenciadas de nitrógeno;
- 30 a) expresar las moléculas de ácido nucleico identificadas en las células anfitrionas;
- b) probar los niveles de asimilación, acumulación y/o utilización de nitrógeno en las células anfitrionas;
- c) identificar la molécula de ácido nucleico y su producto genético cuya expresión confiere una asimilación, acumulación y/o utilización incrementada de nitrógeno en la célula anfitriona después de la expresión en comparación con el tipo silvestre.
- 35 Las condiciones de hibridación relajadas son: después de los procedimientos de hibridación estándar puede llevarse a cabo etapas de lavado con condiciones de restricción baja a medias usualmente con condiciones de lavado de 40° - 55°C y condiciones salinas entre 2xSSC y 0.2xSSC, con 0.1% de SDS en comparación con condiciones de lavado restrictivas tales como por ejemplo 60°-68°C, con 0.1% de SDS. Ejemplos adicionales pueden encontrarse en las referencias que se presentan más adelante en listas para las condiciones de hibridación restrictiva. Usualmente las etapas de lavado son repetidas con restricción y longitud incrementadas hasta que se detecte una señal útil con relación al ruido y depende en muchos factores del objetivo, por ejemplo de su pureza, contenido de GC, tamaño, la sonda, su longitud, su sonda de ARN o ADN, condiciones de sal, temperatura de lavado o hibridación, tiempo de lavado de hibridación, etc.
- 40 En otra realización, la presente invención se relaciona con un método para la identificación de un producto genético que confiere una asimilación, acumulación y/o utilización potenciada de nitrógeno en un organismo, comprendiendo las siguientes etapas:
- 45 Identificar las moléculas de ácido nucleico de un organismo, que pueden contener un gen candidato que codifica un producto genético que confiere una asimilación, acumulación y/o utilización potenciada de nitrógeno después de la expresión, las cuales son al menos 20%, preferiblemente 25%, más preferiblemente 30%, aún más preferiblemente 35%, 40% o 50%, incluso aún más preferiblemente 60%, 70% u 80%, lo más preferido 90% y 95% o más homólogas a la molécula de ácido nucleico de la presente invención, por ejemplo a través de búsqueda de homología en un banco de datos;
- 50

introducir las moléculas de ácido nucleico candidatas en células anfitrionas, preferiblemente en células vegetales o microorganismos, mejora apropiada de la asimilación, acumulación y/o utilización de nitrógeno y eficiencia incrementada en el uso de nitrógeno (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible del suelo, incluyendo fertilizantes;

- 5 a) expresar las moléculas de ácido nucleico identificadas en las células anfitrionas,
- b) probar el nivel potenciado de asimilación, acumulación y/o utilización de nitrógeno en las células anfitrionas; y
- c) identificar las moléculas de ácido nucleico y su producto genético cuya expresión confiere una asimilación, acumulación y/o utilización potenciada de nitrógeno en la célula huésped después de la expresión comparada con el tipo silvestre.
- 10 Eventualmente los productos genéticos confieren el incremento en la asimilación potenciada de nitrógeno, acumulación y/o utilización pueden también identificar de acuerdo con una estructura idéntica o similar en 3D en la etapa (a) y por el método antes descrito.

Las moléculas de ácido nucleico identificadas pueden ser utilizadas entonces para la producción de organismos con asimilación, acumulación y/o utilización potenciada de nitrógeno en la misma forma que la molécula de ácido nucleico de la presente invención.

De acuerdo con lo anterior, en una realización, la presente invención se relaciona con un proceso para la acumulación o producción de nitrógeno o compuestos que contienen nitrógeno respectivamente, que comprende (a) identificar una molécula de ácido nucleico de acuerdo con las etapas antes mencionadas (a) a (f) o (a) a (e) y recuperar en los compuestos libres o enlazados que contienen nitrógeno, especialmente proteínas, de un organismo que tiene una actividad celular incrementada de un polipéptido codificado por la molécula de ácido nucleico aislada en comparación con un tipo silvestre.

Adicionalmente, en una realización, la presente invención se relaciona con un método para la identificación de un compuesto que estimula la asimilación, acumulación y/o utilización potenciada estimulante de nitrógeno y la eficiencia incrementada del uso de nitrógeno (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno de dicha planta que comprende:

- a) poner en contacto células que expresan el polipéptido de la presente invención o su ARNm con un compuesto candidato bajo condiciones de cultivo de células;
- b) probar un incremento en la expresión de dicho polipéptido de dicho ARNm;
- c) comparar el nivel de expresión con una respuesta estándar hecha en ausencia de dicho compuesto candidato; por lo tanto, una expresión incrementada sobre el estándar indica que el compuesto está estimulando una asimilación, acumulación y/o utilización incrementada de nitrógeno y una eficiencia incrementada en el uso de nitrógeno (NUE) definido como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno.

Adicionalmente, en una realización, la presente invención se relaciona con un método para la selección de agonistas o un antagonista de la actividad del polipéptido de la presente invención o usado en el proceso de la presente invención que comprende:

- a) poner en contacto células, tejidos, planta o microorganismos que expresan el polipéptido de acuerdo con la invención con un compuesto candidato o una muestra que comprende una pluralidad de compuestos bajo condiciones que permiten la expresión del polipéptido de la presente invención o usado en el proceso de la presente invención;
- b) probar el nivel de asimilación, acumulación y/o utilización potenciada de nitrógeno o el nivel de expresión del polipéptido en la célula, tejido, planta o microorganismo o el medio en el cual la célula, tejido, planta o microorganismo es cultivado o mantenido;
- c) identificar un agonista o antagonista comparando el nivel medido de asimilación, acumulación y/o utilización de nitrógeno o el polipéptido de la invención o el nivel de expresión usado en la invención con una asimilación, acumulación y/o utilización estándar de nitrógeno o nivel de expresión del polipéptido medido en ausencia de dicho compuesto candidato o una muestra que comprende dicha pluralidad de compuestos, mediante lo cual un nivel incrementado sobre el estándar indica que el compuesto o la muestra que comprende dicha pluralidad de compuestos es un agonista y un nivel disminuido contra el estándar indica que el compuesto o la muestra que comprende dicha pluralidad de compuestos es un antagonista.

Adicionalmente, en una realización, la presente invención se relaciona con un proceso para la identificación de un compuesto que confiere asimilación, acumulación y/o utilización potenciadas de nitrógeno en una planta o microorganismo, que comprende las etapas:

- 5 a) cultivar una célula o tejido o microorganismo o mantener una planta que expresa el polipéptido de acuerdo con la invención o una molécula de ácido nucleico que codifica dicho polipéptido y un sistema de lectura capaz de interactuar con el polipéptido bajo condiciones adecuadas lo cual permite la interacción del polipéptido con dicho sistema de lectura en presencia de un compuesto o una muestra que comprende una pluralidad de compuestos y capaz de proveer una señal detectable en respuesta al enlazamiento de un compuesto a dicho polipéptido bajo condiciones que permiten la expresión de dicho sistema de lectura y el polipéptido de la presente invención o usado en el proceso de la invención; y
- 10 b) identificar si el compuesto es un agonista efectivo detectando la presencia o ausencia o incremento de una señal producida por dicho sistema de lectura.

La selección de un productor de genes o un agonista que confiere un incremento en la producción de nitrógeno o de compuestos que contienen nitrógeno y un incremento en la eficiencia del uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno puede 15 ejecutarse cultivando un organismo por ejemplo un microorganismo en presencia de cantidades que reducen el crecimiento de un inhibidor de la acumulación o síntesis de nitrógeno o compuestos que contienen nitrógeno respectivamente. Un mejor crecimiento, por ejemplo, una tasa de división más alta o una masa seca alta en comparación con el control bajo tales condiciones identificaría un gen o producto genético o un agonista que confiere una asimilación, acumulación y/o utilización potenciada de nitrógeno.

20 Se puede pensar en seleccionar una asimilación, acumulación y/o utilización potenciadas de nitrógeno buscando por ejemplo una resistencia a un fármaco que bloquea la síntesis de nitrógeno o compuestos que contienen nitrógeno y buscar que este efecto esté pendiente de la actividad o expresión de un polipéptido como se indica en SEQ ID NO: 690, por ejemplo, comparando el fenotipo de organismos casi idénticos con actividad baja y alta de una proteína como se indica en SEQ ID NO: 690 después de incubación con el fármaco.

25 Dicho compuesto puede ser sintetizado químicamente o producido microbiológicamente y/o estar comprendido en, por ejemplo, muestras, por ejemplo extractos celulares de, por ejemplo, plantas, animales o microorganismos, por ejemplo patógenos. Además, dicho compuesto puede ser conocido en la técnica pero hasta ahora no conocido por ser capaz de suprimir o activar el polipéptido de la presente invención. La mezcla de reacción puede ser un extracto libre de células o puede comprender un cultivo de células o tejidos. Parámetros adecuados para el método de la invención son conocidos 30 por la persona experimentada en la técnica y están, por ejemplo, descritos en general en Alberts et al., Molecular Biology of the Cell, third edition (1994), en particular capítulo 17. Los compuestos pueden ser añadidos por ejemplo, a la mezcla de reacción, medio de cultivo, inyectados en la célula o asperjados sobre la planta.

Si una muestra que contienen un compuesto es identificada en el método de la invención, entonces es posible aislar el compuesto de la muestra original identificado por contener el compuesto capaz de activar o incrementar el contenido de 35 nitrógeno o de compuestos que contienen nitrógeno en un organismo o parte del mismo, o puede adicionalmente subdividir la muestra original, por ejemplo, si consiste de una pluralidad de compuestos diferentes, de tal manera que se reduzca el número de sustancias diferentes por muestra y se repite el método con las subdivisiones de la muestra original. Dependiendo de la complejidad de las muestras, las etapas descritas anteriormente pueden ser ejecutadas varias veces, preferiblemente hasta que la muestra identificada de acuerdo con el método de la invención comprenda 40 solamente un número limitado de o solamente una sustancia. Preferiblemente dicha muestra comprende sustancias de propiedades químicas y/o físicas similares, y lo más preferiblemente dichas sustancias son idénticas. Preferiblemente el compuesto identificado de acuerdo con el método antes descrito o su derivado se formula adicionalmente en una forma adecuada para la aplicación en cruce de plantas o cultivo de células y tejidos vegetales.

45 Los compuestos que pueden ser probados e identificados de acuerdo con un método de la invención, pueden ser bibliotecas de expresión, por ejemplo, bibliotecas de expresión de ADNc, péptidos, proteínas, ácidos nucleicos, anticuerpos, compuestos orgánicos pequeños, hormonas, peptidomiméticos, PNA o similares (Milner, Nature Medicine 1 (1995), 879-880; Hupp, Cell 83 (1995), 237-245; Gibbs, Cell 79 (1994), 193-198 y referencias citadas supra). Tales compuestos pueden ser derivados funcionales o análogos de inhibidores o activadores conocidos. Los métodos para preparación de derivados químicos o análogos son bien conocidos por las personas experimentadas en la técnica y 50 están descritos, por ejemplo, en Beilstein, Handbook of Organic Chemistry, Springer edition New York Inc., 175 Fifth Avenue, New York, N.Y. 10010 U.S.A. and Organic Synthesis, Wiley, New York, Estados Unidos. Adicionalmente, dichos derivados y análogos pueden ser probados en cuanto a sus efectos de acuerdo con métodos conocidos en la técnica. Adicionalmente, los peptidomiméticos y/o el diseño auxiliado por ordenador de derivados y análogos apropiados pueden utilizarse, por ejemplo, de acuerdo con los métodos descritos anteriormente. La célula o tejido que puede ser empleada en el método de la invención preferiblemente es una célula anfitriona, célula vegetal o tejido vegetal de la 55 invención descrita en las realizaciones anteriores.

Así, en una realización adicional la invención se relaciona con un compuesto obtenido o identificado de acuerdo con el método para identificar un agonista de la invención siendo dicho compuesto un agonista del polipéptido de la presente invención o usado en el proceso de la presente invención.

5 De acuerdo con lo anterior, en una realización, la presente invención se relaciona adicionalmente con un compuesto identificado por el método para identificar un compuesto de la presente invención.

Dicho compuesto es, por ejemplo, un homólogo del polipéptido de la presente invención. Los homólogos del polipéptido de la presente invención pueden ser generados por mutagénesis, por ejemplo, mutación o truncación en puntos discretos y del polipéptido de la presente invención. Tal como se utiliza aquí, el término "homólogo" se refiere a una variante de la proteína, el cual actúa como un agonista de la actividad del polipéptido de la presente invención. Un
10 agonista de dicha proteína puede retener sustancialmente las mismas, o un subconjunto, de las actividades biológicas del polipéptido de la presente invención. En particular, dicho agonista confiere el incremento del nivel de expresión del polipéptido de la presente invención y/o la expresión de dicho agonista en un organismo o parte del mismo confiere la asimilación, acumulación y/o utilización potenciada de nitrógeno en el organismo o en parte del mismo.

15 En una realización, la invención se relaciona con un anticuerpo que específicamente reconoce el compuesto o agonista de la presente invención.

La invención también se relaciona con una composición de diagnóstico que comprende al menos una de las moléculas de ácidos nucleicos, vectores, proteínas, anticuerpos o compuestos antes mencionados de la invención y medios opcionalmente adecuados para la detección.

20 La composición de diagnóstico de la presente invención es adecuada para el aislamiento de ARNm de una célula y poner en contacto el ARNm así obtenido con una sonda que comprende una sonda de ácido nucleico tal como se describió anteriormente bajo las condiciones de hibridación, detectar la presencia de ARNm hibridado a la sonda, y como consecuencia detectar la expresión de la proteína en la célula. Métodos adicionales para la detección de la presencia de una proteína de acuerdo con la presente invención, comprenden inmunotécnicas bien conocidas en el arte, por ejemplo el ensayo inmunosorbente enlazado a enzima. Adicionalmente, es posible utilizar las moléculas de
25 ácido nucleico de acuerdo con la invención como marcadores moleculares o cebadores en el cruce de plantas. Los medios adecuados para la detección son bien conocidos para una persona experimentada en la técnica, por ejemplo, reguladores y soluciones para pruebas de hibridación, por ejemplo, las soluciones y reguladores, siendo conocidos también medios para inmunoprecipitación Southern-, Western-, Northern-, etc descritos en Sambrook et al.

30 En otra realización, la presente invención se relaciona con un kit que comprende la molécula de ácido nucleico, el vector, la célula anfitriona, el polipéptido, el ácido nucleico antisentido, el anticuerpo, la célula vegetal, la planta o tejido de planta, la parte recolectable, el material de propagación y/o el compuesto agonista o antagonista identificados de acuerdo con el método de la invención.

35 Los compuestos del kit de la presente invención pueden ser empacados en contenedores tales como viales, opcionalmente con/en reguladores y/o solución. Si es apropiado, uno o más de dichos componente puede ser empacado en uno y en uno y mismo contenedor. Adicional o alternativamente, uno o más de dichos componentes podrían ser absorbidos en un soporte sólido, por ejemplo, un filtro de nitrocelulosa, una placa de vidrio, un chip, o una membrana de nylon o en el pozo de una placa de microtitulación. El kit puede ser utilizado para cualquiera de los métodos y realizaciones descritos aquí, por ejemplo, para la producción de células anfitrionas, plantas transgénicas, composiciones farmacéuticas, detección de secuencias homólogas, identificación de antagonistas o agonistas, como
40 alimento o forraje o como suplemento de los mismos, como suplemento para el tratamiento de plantas, etc.

Adicionalmente, el kit puede comprender instrucciones para el uso del kit para cualquiera de dichas realizaciones, en particular para el uso de la producción de organismos o parte de los mismos que tienen asimilación, acumulación y/o utilización potenciada de nitrógeno.

45 En una realización dicho kit comprende adicionalmente una molécula de ácido nucleico que codifica una o más de las proteínas antes mencionadas, y/o un anticuerpo, un vector, una célula anfitriona, un ácido nucleico antisentido, una célula vegetal o un tejido vegetal o una planta.

50 En una realización adicional, la presente invención se relaciona con un método para la producción de una composición para la agricultura que provee la molécula de ácido nucleico, el vector o el polipéptido de la invención o el polipéptido utilizado en el método de la invención o que comprende las etapas del método de acuerdo con la invención para la identificación de dicho compuesto, agonista o antagonista; y la formulación de la molécula de ácido nucleico, el vector o el polipéptido de la invención o el polipéptido usado en el método de la invención o el agonista, o el compuesto identificado de acuerdo con los métodos o procesos de la presente invención o con el uso de los objetivos de la presente invención en una forma aplicable como composición agrícola para plantas.

En otra realización, la presente realización se relaciona con un método para la producción de una composición para el cultivo de plantas que soportan la producción “de compuestos que contienen nitrógeno” que comprende las etapas del método para la presente invención; y la formulación del compuesto identificado en una forma aceptable como composición de uso agrícola.

- 5 Bajo “aceptable como composición de uso agrícola” se entiende que tal composición está de acuerdo con las leyes que regulan el contenido de fungicidas, nutrientes para plantas, herbicidas, etc. Preferiblemente tal composición no representa ningún riesgo para las plantas protegidas y los animales (incluidos humanos) alimentados con las mismas.

10 La presente invención también es pertinente a varias realizaciones relativas a usos y métodos adicionales. La molécula de ácido nucleico, polipéptido, homólogos de proteína, proteínas de fusión, cebadores, vectores, células anfitrionas, descritas aquí pueden ser utilizados en uno o más de los siguientes métodos: identificación de plantas útiles para la producción de nitrógeno o compuestos que contienen nitrógeno como se menciona y organismos relacionados; mapeo de genomas; identificación y localización de secuencias de interés; estados evolutivos; determinación de regiones requeridas para la función; modulación de una actividad.

15 Las moléculas de ácido nucleico de la presente invención tienen una variedad de usos. Primero, pueden ser utilizadas para identificar un organismo o un relacionado cercano del mismo. También, pueden ser utilizados para identificar la presencia del mismo o de un relacionado del mismo de una población mixta de microorganismos o plantas. Sondeando el ADN genómico extraído de un cultivo de una población única o mixta de plantas bajo condiciones restrictivas con una sonda que cubre una región del gen de la presente invención la cual es única para éste, se puede establecer si la presente invención ha sido utilizada o si está presente un relacionado cercano.

20 Adicionalmente, la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención puede ser suficientemente homóloga a las secuencias de especies relacionadas de tal forma que estas moléculas de ácido nucleico pueden servir como marcadores para la construcción de un mapa genómico en un organismo relacionado.

25 De acuerdo con lo anterior, la presente invención se relaciona con un método para cultivar plantas con asimilación, acumulación y/o utilización de nitrógeno potenciada, que comprende

a) proveer una primera variedad de plantas producidas de acuerdo con el proceso de la invención preferiblemente (sobre)expresando la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención,

b) cruzar la primera variedad de planta con una segunda variedad de planta; y

30 seleccionar las plantas germinantes que (sobre)acumulan nitrógeno o (sobre)producen compuestos que contienen nitrógeno por medio del análisis de la distribución de un marcador molecular en los brotes que representan la primera variedad de plantas y su capacidad para (sobre)acumular nitrógeno o (sobre)producir compuestos que contienen nitrógeno.

35 Las moléculas de ácido nucleico de la invención también son útiles para estudios evolutivos y estructurales de proteínas. Al comparar las secuencias de la invención o usadas en el proceso de la invención con aquellas que codifican enzimas similares de otros organismos, puede establecerse la relación evolutiva de los organismos. De la misma forma, tal comparación permite un establecimiento de cuáles regiones de las secuencias son conservadas y cuáles no, lo cual puede ayudar en la determinación de aquellas regiones de la proteína que sean esenciales para el funcionamiento de la enzima. Este tipo de determinaciones de valor para los estudios de ingeniería de proteína y
40 pueden dar una indicación de cual proteína puede tolerar en términos de mutagénesis sin perder la función.

De acuerdo con lo anterior, la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención puede ser utilizada para la identificación de otros ácidos nucleicos que confieren un incremento de nitrógeno o compuestos que contienen nitrógeno después de la expresión.

45 Adicionalmente, la molécula de ácido nucleico de la invención o la molécula de ácido nucleico usada en el método de la invención o un fragmento de un gen que confiere la expresión del polipéptido de la invención o el polipéptido usado en el método de la invención, que comprenden preferiblemente la molécula de ácido nucleico de la invención, pueden ser usados para cultivo asistido por marcador o mapeo de asociación de características derivadas de nitrógeno o compuestos que contienen nitrógeno.

50 De acuerdo con lo anterior, el ácido nucleico de la invención, el polipéptido de la invención y el polipéptido usado en el método de la invención, el constructo de ácido nucleico de la invención, los organismos, las célula anfitriona, los microorganismos, la planta, el tejido vegetal, la célula vegetal, o partes de las mismas de la invención, el vector de la invención, el agonista definido con el método de la invención, la molécula de ácido nucleico identificada con el método

de la invención, pueden ser utilizados en la producción de compuestos que contienen N o del compuesto que contiene N y uno o más otros productos químicos finos. De acuerdo con lo anterior, el ácido nucleico de la invención, o la molécula de ácido nucleico identificada con el método de la presente invención o las secuencias complementarias de la misma, el polipéptido de la invención o el polipéptido utilizado en el método de la invención, el constructo de ácido nucleico de la invención, los organismos, la células huésped, los microorganismos, la planta, el tejido vegetal, la célula vegetal, o las partes de los mismos de la invención, el vector de la invención, el antagonista y identificado con el método de la invención, el anticuerpo de la presente invención, la molécula antisentido de la presente invención, pueden ser utilizados para la reducción de nitrógeno o compuestos que contienen nitrógeno en un organismo o partes del mismo, por ejemplo, en una célula.

Adicionalmente, el ácido nucleico de la invención, el polipéptido de la invención o el polipéptido usado en el método de la invención, el constructo de ácido nucleico de la invención, los organismos, la célula anfitriona, los microorganismos, la planta, el tejido vegetal, la célula vegetal o la parte de los mismos de la invención, el vector de la invención, el antagonista o el agonista identificado con el método de la invención, el anticuerpo de la presente invención, la molécula antisentido de la presente invención o la molécula de ácido nucleico identificada con el método de la presente invención, pueden ser utilizados para la preparación de una composición de uso agrícola.

Adicionalmente, el ácido nucleico de la invención, el polipéptido de la invención o el polipéptido usado en el método de la invención, el constructo de ácido nucleico de la invención, los organismos, la célula anfitriona, los microorganismos, la planta, el tejido vegetal, la célula vegetal o las parte de los mismos de la invención, el vector de la invención, el antagonista o el agonista identificado con el método de la invención, el anticuerpo de la presente invención, la molécula antisentido de la presente invención o la molécula de ácido nucleico identificada con el método de la presente invención, pueden ser usados para la identificación y producción de compuestos capaces de conferir una modulación de los nivel de nitrógeno o compuestos que contienen nitrógeno en un organismo o partes del mismo, preferiblemente para identificar y producir compuestos que confieren un incremento de los niveles de nitrógeno o de compuestos que contienen nitrógeno en un organismo o partes del mismo, si dicho compuesto identificado es aplicado al organismo o parte del mismo, esto es, como parte de su alimentación, o en los medios de crecimiento o cultivo.

Estas y otras realizaciones son divulgadas y abarcadas por la descripción y ejemplos de la presente invención. Puede localizarse literatura adicional concerniente a cualquiera de los métodos, usos y compuestos que se van a emplear de acuerdo con la presente invención en bibliotecas públicas, utilizando por ejemplo dispositivos electrónicos. Por ejemplo, la base de datos pública "Medline" puede ser utilizada estando disponible en internet, por ejemplo, bajo <http://www.ncbi.nlm.nih.gov/PubMed/medline.html>. Bases de datos y direcciones adicionales, tales como <http://www.ncbi.nlm.nih.gov/>, <http://www.infobiogen.fr/>, <http://www.fmi.ch/biology/research-tools.html>, <http://www.tigr.org/>, son conocidas para las personas experimentadas en la técnica y también pueden ser obtenidas utilizando, por ejemplo <http://www.lycos.com>. Una revisión de la información en patentes en biotecnología y una búsqueda de fuentes relevantes de información sobre patentes útil para una búsqueda retrospectiva y para alertas presentes se dan en Berks, TIBTECH 12 (1994), 352-364.

La presente invención es ilustrada por los ejemplos que sigue. Los presentes ejemplos ilustran la invención básica sin pretender limitar el objeto de la invención. El contenido de todas las referencias, solicitudes de patentes y solicitudes de patentes publicadas citadas en la presente solicitud de patentes se incorporan aquí como referencia.

Ejemplo 1: Clonación de SEQ ID NO: 689 de *Saccharomyces cerevisiae* para la expresión en plantas.

Al menos que se especifique otra cosa, se usan métodos estándar como los descritos en Sambrook et al., Molecular Cloning: A laboratory manual, Cold Spring Harbor 1989, Cold Spring Harbor Laboratory Press.

La SEQ ID NO: 689 es amplificada por PCR como se describe en el protocolo de Pfu Turbo o DNA Herculase polymerase (Stratagene).

La composición para el protocolo del Pfu Turbo DNA polymerase fue como sigue: 1x PCR de regulador (Stratagene), 0.2 mM de cada dNTP, 100 ng de ADN genómico de *Saccharomyces cerevisiae* (cepa S288C; Research Genetics, Inc., ahora Invitrogen), 50 pmol de cebador de avance, 50 pmol de cebador reverso, 2.5 u de Pfu Turbo DNA polymerase. Los ciclos de amplificación fueron como sigue:

1 ciclo de 3 minutos a 94-95°C, seguido por 25-36 ciclos de en cada caso 1 minuto a 95°C o 30 segundos a 94°C, 45 segundos a 50°C, 30 segundos a 50°C o 30 segundos a 55°C y 210-480 segundos a 72°C, seguidos por 1 ciclo de 8 minutos a 72°C, luego 4°C. La composición para el protocolo de la Herculase polimerasa fue como sigue: 1x regulador PCR (Stratagene), 0.2 mM de cada dNTP, 100 ng de ADN genómico de *Saccharomyces cerevisiae* (cepa S288C; Research Genetics, Inc., ahora Invitrogen), 50 pmol de cebador de avance, 50 pmol de cebador reverso, 2.5 u de Herculase polimerasa. Los ciclos de amplificación son como sigue:

ES 2 402 309 T3

1 ciclo de 2-3 minutos a 94°C, seguido por 25-30 ciclos en cada caso 30 segundos a 94°C, 30 segundos a 55-60°C y 5-10 minutos a 72°C, seguido por 1 ciclo de 10 minutos a 72°C, luego a 4°C.

Se seleccionan las siguientes secuencias de cebadores para el gen SEQ ID NO: 689

i) cebador de avance (SEQ ID NO: 949)

5 atggctcggg gtgacggaca t

ii) cebador reverso (SEQ ID NO: 950)

tcatgcttct tttgcgtgat gcaat

Después de esto, amplificado se purificó sobre columnas de QIAquick siguiendo el protocolo estándar (Quiagen).

10 Para la clonación de productos de PCR, producidos por la Pfu Turbo DNA polimerasa, el vector de ADN (30 ng) fue restringido con SmaI siguiendo el protocolo estándar (MBI Fermentas) y detenido mediante la adición de regulador con salinidad alta. Los fragmentos de vector restringidos son purificados a través de columnas Nucleobond usando el protocolo estándar (Macherey-Nagel). Después de esto; el vector linealizado fue desfosforilado utilizando el protocolo estándar (MBI Fermentas).

15 Los productos de PCR, producidos por Pfu Turbo ADN polimerasa, fueron fosforilados utilizando una T4 ADN polimerasa que utiliza un protocolo estándar (por ejemplo MBI Fermentas) y fueron clonados en el vector binario procesado.

20 Los términos de ADN de los productos de PCR producidos por Herculasa ADN polimerasa, fueron amolados en una segunda reacción de síntesis utilizando Pfu Turbo ADN polimerasa. La composición para el protocolo del amolamiento de los términos de ADN fue como sigue: dTTP 0.2 mM con amolamiento y Pfu Turbo ADN polimerasa 1.25 u. La reacción fue incubada a 72°C durante 30 minutos. Luego los productos de PCR fueron fosforilados utilizando una T4 DNA polimerasa que utiliza un protocolo estándar (por ejemplo MBI Fermentas) y se clona en el vector procesado también.

25 Se utilizaron un vector binario que comprende un casete de selección (promotor, marcador de selección, terminador) y un casete de expresión con promotor, casete de clonación y secuencia terminadora entre las secuencias de frontera de T-ADN. Además de estos dentro del casete de clonación, el vector binario no tenía sitio de escisión SmaI. Los vectores binarios que pueden ser utilizados son conocidos por las personas experimentadas en la técnica; puede encontrarse una revisión de vectores binarios y su uso en Hellens, R., Mullineaux, P. and Klee H., [(2000) "A guide to Agrobacterium binary vectors", Trends in Plant Science, Vol. 5 No.10, 446-451. Dependiendo del vector usado, la clonación puede ventajosamente ser llevada a cabo también a través de otras enzimas de restricción. Sitios de escisión ventajosos adecuados pueden ser agregados al ORF utilizando cebadores adecuados para la amplificación por PCR.

30 Se mezclaron aproximadamente 30 ng de vector preparado de una cantidad definida de amplificado preparado y se ligaron mediante la adición de ligasa.

35 Los vectores ligados fueron transformados en el mismo recipiente de reacción por adición de células de E. coli competentes (cepas DH5alfa) e incubación durante 20 minutos a 1°C seguido por un choque de calor durante 90 segundos a 42°C y enfriamiento a 4°C. Luego, se agregó medio (SOC) completo y la mezcla se incubó durante 45 minutos a 37°C. La mezcla completa fue sembrada subsecuentemente sobre una placa de agar con antibióticos (seleccionados como una función del vector binario usado) y se incubó durante la noche a 37°C.

40 El resultado de la etapa de clonación fue verificado por amplificación con la ayuda de cebadores que se enlazan corriente arriba y corriente abajo del sitio de integración, permitiendo así la amplificación de la inserción. Además se usaron combinaciones de los cebadores específicos para los genes antes mencionados y cebadores corriente arriba y corriente abajo en reacciones de PCR para identificar clones con la orientación de inserción correcta. Las amplificaciones fueron llevadas a cabo como se describe en el protocolo de la T aq DNA polimerasa (Gibco-BRL).

Los ciclos de amplificación fueron como sigue: 1 ciclo de 5 minutos a 94°C, seguido por 35 ciclos de en cada caso 15 segundos a 94°C, 15 segundos a 50-66°C y 5 minutos a 72°C seguido por 1 ciclo de 10 minutos a 72°C, luego 4°C.

45 Se verificaron varias colonias, pero solamente se detectó una colonia para la cual se detectó un producto de PCR del tamaño esperado en las siguientes etapas.

Una porción de esta colonia positiva fue transferida a un recipiente de reacción lleno con medio completo (LB) y se incubó durante la noche a 37°C. El medio LB contenía un antibiótico escogido para coincidir con el vector binario (véase más arriba) usado y el gen de resistencia presente aquí con el fin de seleccionar el clon.

La preparación de plásmido fue llevada a cabo como se especifica en el protocolo estándar Quiaprep (Quiagen).

5 Ejemplo 2: Generación de plantas transgénicas que expresan SEQ ID NO: 689

1 ng del ADN de plásmido aislado fue transformado por electroporación en células competentes de *Agrobacterium tumefaciens*, de cepa GV 3101, pMP90 (Koncz and Schell, Mol. Gen. Genet. 204, 383-396, 1986). La selección de la cepa agrobacteriana depende de la selección del vector binario. Una revisión de cepas posibles y sus propiedades se encuentra en Hellens, R., Mullineaux, P. and Klee H., (2000) " A guide to *Agrobacterium* binary vectors, Trends in Plant Science, Vol. 5 No.10, 446-451. Después de esto, se agregó medio completo (YEP) y la mezcla fue transferida en un recipiente de reacción nuevo durante 3 horas a 28°C. Después, toda la mezcla de reacción fue sembrada sobre placas de agar YEP suplementadas con los respectivos antibióticos, por ejemplo, rifampicina y gentamicina para GV3101 pMP90, y un antibiótico adicional para la selección sobre el vector binario, fue sembrado e incubado por 48 horas a 28°C.

15 Las agrobacterias generadas en el ejemplo 2, que contienen el constructo de plásmido fueron utilizadas entonces para la transformación de las plantas.

Se tomó una colonia de la placa de agar con la ayuda de una punta de pipeta y se colocó en 3 ml de medio TB líquido, que también contenía antibióticos adecuados, dependiendo de la cepa agrobacteriana y el plásmido binario. El precultivo fue cultivado a 48 horas a 28°C y 120 rpm.

20 Se usaron 400 ml de medio LB que contiene los mismos antibióticos indicados anteriormente para el cultivo principal. El precultivo fue transferido al cultivo principal. Se cultivó durante 18 horas a 28°C y 120 rpm. Después de centrifugar a 4000 rpm, la pella fue resuspendida en medios de infiltración (medio MS, sacarosa al 10%).

25 Con el fin de cultivar las plantas para la transformación, se llenaron placas (Piki Saat 80, verdes, provistos con un botón verde, 30 x 20 x 4.5 cm de Wiesauplast, Kunststofftechnik, Alemania) con un sustrato de GS 90 (solución estándar, Werkverband E.V., Alemania). Las placas fueron lavadas durante la noche con solución Proplant al 0.05% (Chimac-Apriphar, Bélgica). Se esparcieron semillas de *Arabidopsis thaliana* C24 (Nottingham Arabidopsis Stock Centre, Reino Unido; NASC Stock N906), aproximadamente 1000 semillas por placa. Las placas fueron cubiertas con una capucha y colocadas en la instalación de estratificación (8 h, 110 mmol/m²/s-1, 22°C; 16 h, oscuridad, 6°C). Después de 5 días, las placas fueron colocadas en la cámara ambiental controlada de día corto (8 h 130 mmol/m²/s-1, 22°C; 16 h, oscuridad 20°C), donde permanecieron durante aproximadamente 10 días hasta que se formaron las primeras hojas de planta.

30 Los plantones fueron transferidos a macetas que contenían el mismo sustrato (maceta Teku, 7 cm, serie LC, manufacturadas por Poppelmann GmbH & Co, Alemania). Se sembraron cinco plantas en cada maceta. Las macetas fueron regresadas a la cámara ambiental controlada de día corto para que las plantas continuaran con su crecimiento.

35 Después de 10 días, las plantas fueron transferidas a la cabina del invernadero (iluminación suplementaria 16 h, 340 µE, 22°C; 8 h, oscuridad, 20°C), donde se dejaron crecer durante 17 días adicionales.

40 Para la transformación, se sumergieron plantas de *Arabidopsis* de 6 semanas de edad que habían acabado de empezar su florescencia durante 10 segundos en la suspensión agrobacteriana antes descrita la cual había sido tratada previamente con 10 µl de Silwett L77 (Crompton S.A., osi Specialties, Suiza). El método en cuestión está descrito en Clough and Bent, 1998 (Clough, JC and Bent, AF. 1998 Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*, Plant J. 16:735-743.

Las plantas fueron colocadas subsecuentemente durante 18 horas en una cámara húmeda. Después de esto, las macetas fueron regresadas al invernadero para que las plantas continuaran con su crecimiento. Las plantas permanecieron en el invernadero durante otras 10 semanas hasta que las semillas estaban listas para recolección.

45 Dependiendo del marcador de resistencia utilizado para la selección de las plantas transformadas las semillas recolectadas fueron plantadas en el invernadero y sometidas a una selección por aspersión o incluso más fueron primero esterilizadas luego cultivadas sobre placas de agar suplementadas con el agente de selección respectivo. En el caso de resistencia a BASTA®, las plántulas fueron asperjadas cuatro veces a un intervalo de 2 a 3 días con BASTA® al 0.02% y se dejó que las plantas transformadas formaran semillas. Las semillas de las plantas de *A. thaliana* transgénica fueron almacenadas en el congelador (a -20°C).

Ejemplo 3: Análisis del contenido de nitrógeno

La determinación del nitrógeno en las muestras se lleva a cabo usando el método de Dumas el cual se basa en la combustión completa del material de prueba. La muestra se calienta en un horno a temperatura y rápidamente hace combustión en presencia de oxígeno puro. Los productos de combustión (principalmente CO₂, H₂O, NO_x y N₂) se recogen y se dejan equilibrar. Se pasa una alícuota de la mezcla de gas sobre cobre caliente para mover cualquier oxígeno y convertir el NO₂ en N₂. La muestra se pasa entonces a través de una trampa que retira el CO₂ y el H₂O. El nitrógeno restante es medido mediante un detector de conductividad térmica.

Para los análisis de materiales de hojas o de cáscaras de semillas, se utiliza material homogenizado secado por congelación. En el caso de semillas de Arabidopsis, las semillas se analizan directamente sin pretratamiento.

Se pesaron 4-7 mg de la muestra en un recipiente de hoja de estaño junto con 15 mg de óxido (WO₃) de tungsteno (VI). El análisis fue llevado a cabo utilizando un analizador elemental comercial (por ejemplo ELEMENTAR vario EL III, ELEMENTAR, Hanau, Alemania).

La Tabla VI muestra el contenido de nitrógeno total incrementado en semillas de plantas transgénicas transformadas con el ORF YPR138C de levadura, correspondiente a SEQ ID NO: 689 bajo control del promotor doble 35S. La columna 1 muestra los elementos medidos, la columna 2 muestra la variabilidad tipo silvestre como desviación estándar relativa, la columna 3 muestra los cambios principales en el contenido de elementos para las diferentes líneas transgénicas transformadas por SEQ ID NO: 689 con respecto al control tipo silvestre el cual es estandarizado como "1", la columna 4 muestra la desviación estándar para las diferentes líneas transgénicas y la columna 5 muestra el cambio máximo observado. Como se esperaba, el incremento relativo en nitrógeno corresponde a un descenso relativo en el contenido de carbono.

Parámetro	Variabilidad en WT (RSD; %)	Promedio	SD	Cambio máximo
%N	0.05	1.17	0.06	1.24
%C	0.01	0.90	0.03	0.87

Ejemplo 4: Eficiencia potenciada en el uso de nitrógeno

Selección de plantas (Arabidopsis) para crecimiento bajo suministro limitado de nitrógeno.

Para seleccionar plantas transgénicas se utilizó una instalación de cultivo específica. Para propósitos de alto rendimiento las plantas fueron seleccionadas en cuanto a la producción de biomasa sobre placas de agar con suministro limitado de nitrógeno (adaptado de Estelle and Somerville, 1987). El flujo de selección consiste de tres niveles. Las líneas transgénicas son sometidas a niveles subsecuentes si la producción de biomasa se mejoró significativamente en comparación con las plantas tipo silvestre. Con cada nivel se incrementó el número de replicados y la restricción estadística.

Por cada siembra, las semillas, que habían sido almacenadas en el refrigerador (a -20°C), fueron retiradas de los tubos Eppendorf con la ayuda de un palillo y transferidas sobre placas. En total, se distribuyeron aproximadamente 15-30 semillas horizontalmente sobre cada placa (12 x 12 cm).

Después de que las semillas habían sido sembradas, las placas se sometieron a estratificación durante 2-4 días en la oscuridad a 4 °C. Después de la estratificación, las plantas de prueba fueron cultivadas durante 22 a 25 días con un ritmo de 16 horas de luz, 8 horas de oscuridad a 20°C, una humedad atmosférica de 60% y una concentración de CO₂ de aproximadamente 400 ppm. Las fuentes de luz utilizadas generan una luz que recuerda el espectro de color solar con una intensidad de luz de aproximadamente 100 µE/m²/s-1.

El crecimiento mejorado bajo condiciones limitadas de nitrógeno fue establecido por la producción de biomasa de los brotes y raíces de plantas transgénicas en comparación con plantas de control tipo silvestre después de 20-25 días de crecimiento.

Las líneas transgénicas que expresan el ORF YPR138C de levadura correspondiente a SEQ ID NO: 689 bajo control de un fuerte promotor constitutivo como el promotor doble 35S, mostraron producción potenciada de biomasa de brotes y raíces bajo condiciones limitadas de nitrógeno en comparación con las plantas de control de tipo silvestre.

Ejemplo 5: Manipulación de plantas de ballico por sobreexpresión del polinucleótido caracterizado en la invención, por ejemplo, derivado de Saccharomyces, E. coli u otro organismo.

5 Pueden utilizarse semillas de diferentes variedades de ballico como fuentes explanta para la transformación, incluyendo la variedad comercial Gunne disponible en SvalofWeibull see company o la variedad Affinity. Las semillas son esterilizadas en superficies secuencialmente con Tween-20 al 1% durante 1 minuto, blanqueamiento al 100% durante 60 minutos, 3 enjuagues con 5 minutos cada uno con H2O desionizada y destilada, y luego germinada durante 3-4 días en papel de filtro estéril húmedo en la oscuridad. Los plantones son esterilizados adicionalmente durante 1 minuto con Tween-20 al 1%, 5 minutos con blanqueador al 75% y enjuagado 3 veces con ddH2O 5 minutos cada vez.

10 Las semillas esterilizadas en superficie se colocan sobre el medio de inducción de callo que contiene sales basales de Murashige and Skoog y vitaminas, 20 g/l de sacarosa, 150 mg/l de aspargina, 500 mg/l de hidrolizado de caseína, 3 g/l de Phytigel, 10 mg/l de BAP y 5 mg/l de dicamba. Las placas se incuban en la oscuridad a 25 °C durante 4 semanas para la germinación de las semillas y la inducción de callos embriogénicos.

15 Después de 4 semanas sobre el medio de inducción de callos, los brotes y raíces de los plantones son podados, el callo es transferido a medio fresco, se mantiene el cultivo durante otras 4 semanas y luego es transferido a un medio MSO en luz durante 2 semanas. Se drena entonces varias piezas de callos (11-17 semanas de edad) a través de un tamiz de malla 20 y se colocan sobre un medio de inducción de callos, o se cultivan en 100 ml de medio de inducción de callos líquido para ballico (el mismo medio que para la inducción de callos con agar) en un matraz de 250 ml; el matraz es envuelto en papel de aluminio y se agita a 175 rpm en la oscuridad a 23°C de la 1 semana. Se recolectan las células tamizando el cultivo líquido con un tamiz de malla 40. La fracción recolectada sobre el tamiz es sembrada y cultivada sobre medio de inducción de callos sólido para ballico durante 1 semana en la oscuridad a 25°C. los callo son transferidos entonces a y cultivados en medio MS que contiene 1% de sacarosa durante 2 semanas.

20 La transformación puede ser lograda bien sea con Agrobacterium o con métodos con bombardeo con partículas. Se crea un vector de expresión que contiene un promotor constitutivo de la planta y el ADNc del gen en un vector pUC. El ADN plásmido es preparado a partir de células de E. Coli utilizando el kit Quiagen de acuerdo con la instrucción del fabricante. Aproximadamente, se esparcen 2 g de callo embriogénico en el centro de un papel de filtro estéril en una placa de Petri. Se agrega al papel de filtro una alícuota de MSO líquido con 10 g/l de sacarosa. Se recubren placas de oro (1.0 µm de tamaño) con el plásmido de ADN de acuerdo con el método de Sanford et al., 1993 y se administran al callo embriogénico con los siguientes parámetros: 500 µg de partículas y 2 µg de ADN por disparo, 1300 psi y una distancia al objetivo de 8.5 cm desde la placa de detención a la placa del callo y un disparo por placa de callo.

25 Después del bombardeo los callos son transferidos de nuevo al medio de desarrollo de callos fresco y se mantiene en la oscuridad a temperatura ambiente durante un periodo de 1 semana. El callo es transferido entonces a condiciones de crecimiento a la luz a 25°C para iniciar la diferenciación de los embriones con el agente de selección apropiado, por ejemplo, Arsenal 250 nM, 5 mg/l de PPT o 50 mg/L de kanamicina. Los brotes resistentes al agente de selección aparecen y una vez enraizados son transferidos al suelo.

30 Las muestras de las plantas transgénicas primarias (T0) se analizan por PCR para confirmar la presencia de T-ADN. Estos resultados son confirmados por hibridación Southern en la cual el ADN se somete a electroforesis en un gel de agarosa al 1% y se transfiere a una membrana de nylon cargada positivamente (Roche Diagnostics). El PCR DIG Probe Synthesis Kit (Roche Diagnostics) se utiliza para preparar una sonda marcada con digoxigenina por PCR, y se utiliza como lo recomienda el fabricante.

35 Las plantas de ballico transgénicas T0 se propagan vegetativamente por cañones de escisión. Los cañones trasplantados son mantenidos en el invernadero durante 2 mese hasta que estén bien establecidos. Los lotes son desfoliados y se dejan crecer durante 2 semanas.

Las semillas del ballico transgénico pueden ser analizadas en cuanto al contenido incrementado de nitrógeno con el analizador elemental como se describe bajo el ejemplo 3.

Ejemplo 6: Manipulación de plantas de soja por sobreexpresión del polinucleótido caracterizado en la invención, por ejemplo, derivado de Saccharomyces, E. coli u otro organismo

45 La soja puede ser transformada de acuerdo con las siguientes modificación del método descrito en la Patente US 5,164,310 DE Texas A&M. Varias variedades de soja comercial son muy susceptibles de transformación por este método. El cultivar Jack (disponible de la Illinois Seed Foundation) se utiliza comúnmente para la transformación. Las semillas son esterilizadas por inmersión en etanol al 70% (v/v) durante 6 minutos y en legía comercial al 25% NaOCl) complementada con Tween al 0.1% (v/v) durante 20 minutos, seguido por enjuague 4 veces con agua doblemente destilada estéril. La eliminación del radículo, el hipocótilo y un cotiledón de cada plantón propaga plantones de siete días. Luego, el hepocotilo con los cotiledones es transferido a un medio de germinación fresco en cajas de petri e incubado a 25°C bajo un fotoperiodo de 16 horas (aproximadamente 100 µE-m-2s-1) durante tres semanas. Se cortan los nódulos axilares (de aproximadamente 4 mm de longitud) de plantas de 3-4 semanas de edad. Los nódulos axilares son escindidos e incubados en medio de cultivo Agrobacterium LBA4404.

Se han descrito muchos sistemas de vectores binarios para la transformación de plantas (por ejemplo, An, G, in *Agrobacterium Protocols. Methods in Molecular Biology* vol 44, pp 47-62, Gartland KMA and MR Davey eds. Humana Press, Totowa, New Jersey). Muchos están basados en el vector pBIN19 descrito por Bevan (Nucleic Acid Research. 1984, 12: 8711-8721) que incluye un casete de expresión genética de plantas flanqueado por las secuencias fronteras izquierda y derecha del plásmido Ti de *Agrobacterium tumefaciens*. Un casete de expresión de genes de plantas consiste de al menos dos genes – un gen marcador de selección y un promotor vegetal que regula la transcripción del ADNc o ADN genómico del gen característico. Pueden utilizarse diversos genes marcadores de selección, como se describe anteriormente, utilizando el gen de *Arabidopsis* que codifica una enzima acetohidroxi ácido sintasa (AHAS) sometida a mutación (Patentes de los Estados Unidos 57673666 y 6225105). De la misma forma, pueden utilizarse diversos promotores para regular el gen característico para proveer una regulación constitutiva, de desarrollo, de tejido, o ambiental de la transcripción genética como se describió anteriormente.

En este ejemplo, el promotor 34s (números GenBank Accession M59930 y X16673) se utilizan para proveer expresión constitutiva del gen característico.

Después del tratamiento en cocultivo, los explantes fueron lavados y transferidos a medios de selección suplementados con timentina 500 mg/L. los brotes fueron escindidos y colocados sobre un medio de elongación de brotes. Los brotes de más de 1 cm de longitud fueron colocados sobre medios de enraizamiento por dos a cuatro semanas antes de trasplantarlos al suelo.

Las plantas transgénicas primarias (T0) se analizan por PCR para confirmar la presencia de T-ADN. Estos resultados son confirmados por hibridación Southern en la cual el ADN es sometido a electroforesis sobre un gel de agarosa al 1% y transferido a una membrana de nylon cargada positivamente (Roche Diagnostics). El PCR DIG Probe Synthesis Kit (Roche Diagnostics) se utiliza para preparar una sonda marcada con dioxigenina por PCR, y se utiliza como lo recomienda el fabricante.

Las semillas de soja transgénica pueden ser analizadas en cuanto al contenido incrementado de nitrógeno con el analizador elemental tal como se describió abajo en el ejemplo 3.

25 **Ejemplo 7. Manipulación de plantas de maíz sobreexpresando el polinucleótido caracterizado en la invención, por ejemplo, derivado de *Saccharomyces*, *E. coli* u otro organismo.**

La transformación del maíz (*Zea Mays* L.) puede llevarse a cabo con una modificación del método descrito por Ishida et al. (1996. *Nature Biotech* 14745-50). La transformación es dependiente del genotipo del maíz y solamente los genotipos específicos son susceptibles de transformación y regeneración. La línea de cruzamiento interno A188 (University of Minnesota) o híbridos con A188 aparentemente son buenas fuentes de material donante para la transformación (Fromm et al., 1990. *Biotech* 8:833-839), pero pueden utilizarse exitosamente también otros genotipos. Se recolectan las ears de plantas de maíz a aproximadamente 11 días después de la polinización (DAP) cuando la longitud de los embriones inmaduros es aproximadamente 1 a 1.2 mm. Los embriones inmaduros son cocultivados con *Agrobacterium tumefaciens* que corta vectores “superbinarios” y se recuperan las plantas transgénicas mediante organogénesis. El sistema de vector superbinario de Japan Tobacco está escrito en las Patentes WO 94/00977 y WO 95/06722. Los vectores pueden ser contruidos como se describe. Pueden utilizarse diversos genes marcadores de selección incluyendo el gen de maíz que codifica una enzima acetohidroxi ácido sintasa (AHAS) que ha sufrido mutación (Patente de los Estados Unidos 6025541). De igual manera pueden utilizarse diversos promotores para regular el gen característico para proveer una regulación constitutiva, de desarrollo, de tejido o ambiental de la transcripción del gen. En este ejemplo, el promotor 34S (números de GenBank Accession M59930 y X16673) puede utilizarse para proveer expresión constitutiva del gen característico.

Los embriones escindidos pueden cultivarse sobre medio de inducción de callos, luego sobre medio de regeneración de maíz, que contiene imidazolinona como agente de selección. Las placas de Petri pueden ser incubadas en luz a 25°C durante 2-3 semanas, o hasta que se desarrollen los brotes. Los brotes verdes pueden ser transferidos desde cada embrión a medio de enraizamiento de maíz e incubados a 25°C durante 2-3 semanas, hasta que se desarrollen las raíces. Los brotes enraizados pueden ser trasplantados al suelo en el invernadero. Pueden producirse semillas T1 a partir de plantas que exhiban tolerancia a los herbicidas de imidazolinona y que pueden ser PCR positivos para los transgenes.

La generación T1 de inserciones en locus individual del T-ADN pueden segregar el transgen en una relación 3:1. Esta progenie que contiene una o dos copias del transgen puede ser tolerante para el herbicida imidazolinona. Las semillas de maíz transgénico pueden ser analizadas en cuanto al contenido incrementado de nitrógeno con el analizador elemental tal como se describe bajo el ejemplo 3.

Ejemplo 8: Manipulación de plantas de trigo por sobreexpresión del polinucleótido caracterizado en la invención, por ejemplo, derivado de *Saccharomyces*, *E. coli* u otro organismo.

La transformación del trigo puede ejecutarse con el método descrito por Ishida et al., (1996 Nature Biotech, 14745-50). El cultivar BobWhite (disponible de CYMMIT, México) puede ser utilizado comúnmente en la transformación. Pueden cultivarse embriones inmaduros con *Agrobacterium tumefaciens* que porta vectores "superbinarios", y se recuperan las plantas transgénicas a través de organogénesis. El sistema de vector superbinario de Japan Tobacco está descrito en las Patentes WO 94/00977 y WO 95/06722. Los vectores pueden ser construidos tal como se describió. Pueden utilizarse diversos genes marcadores de selección incluyendo el gen de maíz que codifica una enzima acetohidroxi ácido sintasa (AHAS) que ha sufrido mutación (Patente de los Estados Unidos 6025541) de la misma forma, pueden utilizarse diversos promotores para regular el gen característico para proveer regulación constitutiva, de desarrollo, de tejido o ambiental de la transcripción genética. El promotor 34S (número de GenBank Accession M59930 y X16673) puede utilizarse para proveer expresión constitutiva del gen característico.

Después de la incubación con *Agrobacterium*, los embriones pueden ser cultivados sobre medio de inducción de callos, luego en medio de regeneración, que contiene imidazolinona como agente de selección. Las placas de Petri pueden ser incubadas a la luz a 25°C durante 2-3 semanas, o hasta que se desarrollen los brotes. Los brotes enraizados pueden ser trasplantados al suelo en el invernadero. Pueden producirse semillas T1 a partir de plantas que exhiben tolerancia a los herbicidas de imidazolinona y los cuales son PCR positivos para los transgenes.

La generación T1 de inserciones en locus individual del T-ADN pueden segregar el transgen en una relación 3:1. Esta progenie que contiene una o dos copias del transgen puede ser tolerante al herbicida imidazolinona. Las plantas homocigóticas T2 exhiben fenotipos similares. Pueden analizarse semillas de trigo transgénico en cuanto al contenido incrementado de nitrógeno con el analizador elemental tal como se describe bajo el ejemplo 3.

Ejemplo 9: Manipulación de plantas de colza/canola por sobreexpresión del polinucleótido caracterizado en la invención, por ejemplo, derivado de *Saccharomyces cerevisiae*, *E. coli* u otro organismo.

Peciolos e hipocotilos de cotiledones de plántulas de 5-6 días de edad pueden ser utilizados como explantes para cultivo de tejidos y transformados de acuerdo con Babic et al., (1998, Plant Cell Rep 17: 183-188). El cultivar comercial Westar (Agriculture Canada) puede ser la variedad estándar utilizada para la transformación, pero pueden utilizarse otras variedades.

Puede utilizarse *Agrobacterium tumefaciens* LBA4404 que contiene un vector binario para la transformación de la canola. Muchos sistemas binarios diferentes han sido descritos para la transformación de plantas (por ejemplo An, G. in *Agrobacterium Protocols. Methods in Molecular Biology* vol 44, pp 47-62, Gartland KMA and MR Davey eds. Humana Press, Totowa, New Jersey). Muchos están basados en el vector pBIN19 descrito por Bevan (Nucleic Acid Research. 1984. 12:8711-8721) que incluye un casete de expresión de genes de plantas flanqueado por las secuencias de frontera izquierda y derecha del plásmido Ti de *Agrobacterium tumefaciens*. Un casete de expresión de gen de planta puede consistir de al menos dos genes – un gen marcador de selección y un promotor vegetal que regula la transcripción de ADNc o del ADN genómico del gen característico. Pueden utilizarse diversos genes marcadores de selección incluyendo el gen de *Arabidopsis* que codifica una enzima acetohidroxi ácido sintasa (AHAS) que ha sufrido mutación ((Patente de los Estados Unidos 57673666 y 6225105). De la misma forma, pueden utilizarse diversos promotores que regulan el gen característico para proveer regulación constitutiva, de desarrollo, de tejido o ambiental de la transcripción genética. El promotor 34S (número de GenBank Accession M59930 y X16673) puede utilizarse para proveer expresión constitutiva del gen característico.

Las semillas de canola pueden ser esterilizadas en superficie en etanol al 70% durante 2 minutos, y luego en clorox al 30% con una gota de Tween-20 durante 10 minutos, seguida por tres enjuagues con agua esterilizada destilada. Las semillas pueden ser germinadas *in vitro* durante 5 días en medio MS de fuerza media sin hormonas, sacarosa al 1%, Phytagar al 0.7% a 23°C, con 16 horas de luz. Los explantes del peciolo del cotiledón con el cotiledón unido pueden ser escindidos de los plántulas *in vitro*, y pueden ser inoculados con *Agrobacterium* sumergiendo el extremo cortado del explante del peciolo en la suspensión bacteriana. Los explantes pueden ser cultivados entonces durante 2 días sobre medio MSBAP-3 que contiene 3 mg/l de BAP, sacarosa al 3%, Phytagar al 0.7% a 23°C, con 16 horas de luz. Después de 2 días de cocultivo con *Agrobacterium*, los explantes del peciolo pueden ser transferidos al medio MSBAP-3 que contiene 3 mg/l de BAP, cefotaxima, carbenicilina, o timentina (300 mg/l) durante 7 días, y luego pueden ser cultivados sobre medio MSBAP-3 con cefotaxima, carbenicilina o timentina y el agente de selección hasta la regeneración de los brotes. Cuando los brotes tienen de 5-10 mm de longitud, pueden ser cortados y transferidos a un medio de elongación de brotes (MSBAP-0.5, que contiene 0.5 mg/l de BAP). Los brotes de aproximadamente 2 cm de longitud pueden ser transferidos al medio de enraizamiento (MS0) para inducción de la raíz.

Las muestras de las plantas transgénicas primarias (T0) pueden ser analizadas por PCR para confirmar la presencia de T-ADN. Estos resultados pueden ser confirmados por hibridación Southern en la cual el ADN se somete a electroforesis sobre un gel de agarosa al 1% y se transfieren a una membrana de nylon cargada positivamente (Roche Diagnostics). Puede utilizarse el PCR DIG Probe Synthesis Kit (Roche Diagnostics) para preparar una sonda marcada con digoxigenina por PCR y utilizada como lo recomienda el fabricante.

Las semillas de canola transgénicas pueden ser analizadas en cuanto al contenido incrementado de nitrógeno con el analizador elemental tal como se describe bajo el ejemplo 3.

Ejemplo 10: Plantas de alfalfa manipuladas por sobreexpresión del polinucleótido caracterizado en la invención, por ejemplo, derivado de Saccharomyces, E. coli u otro organismo.

5 Un clon regenerante de alfalfa (*Medicago sativa*) puede ser transformada utilizando el método de (McKersie et al., 1999 *Plant Physiol* 119: 839-847). La regeneración y la transformación de la alfalfa puede ser dependiente del genotipo y por lo tanto se requiere una planta regenerante. Se han descrito métodos para obtener plantas regenerantes. Por ejemplo, pueden seleccionarse del cultivar Rangelander (Agriculture Canada) o cualquier otra variedad de alfalfa comercial tal como lo describe Brown DCW and A Atanassov (1985. *Plant Cell Tissue Organ Culture* 4: 111-112). Alternativamente, la
10 variedad RA3 (University of Wisconsin) puede ser seleccionada para uso en el cultivo de tejidos (Walker et al., 1978 *Am J Bot* 65:654-659).

Los explantes de peciolo pueden ser cocultivados con un cultivo durante la noche de *Agrobacterium tumefaciens* C58C1 pMP90 (McKersie et al., 1999 *Plant Physiol* 119: 839-847) o LBA4404 que contiene un vector binario. Muchos sistemas diferentes de vectores binarios han sido descritos para la transformación de plantas (por ejemplo An, G. in *Agrobacterium Protocols. Methods in Molecular Biology* vol 44, pp 47-62, Gartland KMA and MR Davey eds. Humana Press, Totowa, New Jersey). Muchos están basados en el vector pBIN19 descrito por Bevan (*Nucleic Acid Research*. 1984. 12:8711-8721) que incluye un casete de expresión de genes vegetales flanqueado por las secuencias frontera izquierda y derecha del plásmido Ti de *Agrobacterium tumefaciens*. Un casete de expresión genética vegetal puede consistir de al menos dos genes – un gen marcador de selección y un promotor vegetal que regula la transcripción del
15 ADNC o el ADN genómico del gen característico. Pueden utilizarse diversos genes marcadores de selección que incluyen el gen *Arabidopsis* que codifica una enzima acetohidroxi ácido sintasa (AHAS) que ha sufrido mutación (Patentes de los Estados Unidos 57673666 y 6225105). De la misma forma, pueden utilizarse diversos promotores para regular el gen característico que provee regulación constitutiva. De desarrollo, de tejido o ambiental de la transcripción genética. El promotor 34S (números GenBank Accession M59930 y X16673) puede ser utilizados para
20 promover la expresión constitutiva del gen característico.

Los explantes pueden ser cocultivados durante 3 días en la oscuridad o en un medio de inducción SH que contiene 288 mg/L de Pro, 53 mg/L de tioprolina, 4.35 g/L de K₂SO₄ y 100 µm de acetosiringinona, los explantes pueden ser lavados en medio Murashige- Skoog de fuerza media (Murashige y Skoog, 1962) y sembrados sobre el mismo medio de inducción SH sin acetosiringinona pero con un agente de selección adecuado y un antibiótico adecuado para inhibir el
25 crecimiento del *Agrobacterium*. Después de varias semanas, los embriones somáticos pueden ser transferidos a medio de desarrollo BOi2Y que no contiene reguladores de crecimiento, ni antibióticos, y 50 g/L de sacarosa. Los embriones somáticos son germinados subsecuentemente sobre medio Murashige-Skoog de fuerza media. Los plantones enraizados pueden ser trasplantados en macetas y cultivados en un invernadero.

Las plantas transgénicas T0 son propagadas por cortes en los nódulos y enraizadas en medio de crecimiento Turface. Las plantas son desfoliadas y cultivadas hasta una altura de aproximadamente 10 cm (aproximadamente 2 semanas después de la desfoliación).

Las semillas de alfalfa transgénica pueden ser analizadas en cuanto al contenido incrementado de nitrógeno con el analizador elemental tal como se describe bajo el ejemplo 3.

Ejemplo 11: Preparación de secuencias homólogas de plantas. Pueden cultivarse plantas diferentes bajo condiciones estándar o variables en el invernadero.

Puede extraerse el ARN después del protocolo de Jones, Dunsmuir y Bedbrook (1985) *EMBO J.* 4: 2411-2418. Aproximadamente, 1 gramo de material de tejido de diversos órganos se tritura en nitrógeno líquido. El polvo es transferido a un tubo Falcon de 13 ml que contiene 4.5 ml de regulador NTES (NaCl 100 mM, Tris 10 mM/HCl pH 7.5, EDTA 1 mM, SDS al 1%; en agua libre de RNasa) y 3 ml de fenol/cloroformo/alcohol amílico (25/24/1), se mezclan
45 inmediatamente y se almacena sobre hielo. La mezcla se hace rotar durante 10 minutos a 7000 rpm utilizando una centrifuga (Sorval; rotor SM24 o SS34). El sobrenadante es transferido a un nuevo tubo, se agrega 1/10 del volumen de acetato de sodio 3 M (pH 5.2; en agua libre de RNasa) y un volumen de isopropanol, se mezcla y se almacena durante 1 hora u durante la noche a 20°C. la mezcla se hace rotar durante 10 minutos a 7000 rpm. El sobrenadante es descartado y la pella se lava con 70% de etanol (v/v). la mezcla se hace rotar durante 5 minutos a 7000 rpm, se
50 descarta el sobrenadante y la pella se seca al aire. Se agrega 1 ml de agua libre de RNasa y se deja que la pella de ADN/ARN se disuelva sobre hielo a 4 °C. La solución de ácidos nucleicos se transfiere a un tubo Eppendorf de 2 ml y se agrega 1 ml de acetato de litio 4 M. Después de mezclar la solución se mantiene al menos 3 horas, o durante la noche, a 4°C. La mezcla se hace rotar durante 10 minutos a 14000 rpm, se descarta el sobrenadante, la pella se lava con etanol al 70%, se seca al aire y se disuelve en 200 µl de agua libre de RNasa.

El ARN total puede ser utilizado para construir una biblioteca de ADNc de acuerdo con el protocolo del fabricante (por ejemplo usando el kit de síntesis y clonación ZAP-cDNA de Stratagene, La Jolla, Estados Unidos). Básicamente el ARN mensajero (ARNm) es cebado en la primera síntesis de cadena con un cebador enlazante oligo (dT) y se somete a transcripción reversa utilizando transcriptasa reversa. Después de la síntesis de la segunda cadena de ADNc, el ADNc de cadena doble es ligado al vector Uni-ZAP XR. El vector Uni-ZAP XR permite la escisión in vivo del fagémido pBluescript. El polienlazante del fagémido pBluescript tiene 25 sitios únicos de clonación flanqueados por promotores T3 y T7 y una selección de 6 sitios de cebador diferentes para el secuenciamiento del ADN. Un secuenciamiento en una ronda sencilla sistemática del extremo del cebador 5 esperado de los clones puede permitir la anotación preliminar de las secuencias por ejemplo con la ayuda del paquete pro Software (Biomax, Munich). Los clones de los ácidos nucleicos de la invención o usados en el proceso de acuerdo con la invención pueden ser identificados con base en la búsqueda de homología con algoritmos estándar de la lista blastp o gap. Los clones de longitud completa putativa identificada o de alta homología pueden ser sometidos a secuenciamiento adicional con el fin de obtener la secuencia completa.

Pueden identificarse secuencias completas homólogas nuevas adicionales de manera similar preparando bibliotecas de ADNc a partir de diversas fuentes vegetales tal como se describió anteriormente. Las bibliotecas pueden ser seleccionadas con secuencias disponibles de la invención bajo condiciones de baja restricción por ejemplo como las descritas en Sambrook et al., Molecular Cloning: A laboratory manual, Cold Spring Harbor 1989, Cold Spring Harbor Laboratory Press. Pueden someterse clones positivos purificados a la escisión in vivo y al secuenciamiento completo. Una alineamiento de secuencia tipo par de la secuencia original y de la nueva utilizando el programas blastp o gap permite la identificación de ortólogos, que significan secuencias homólogas de diferentes organismos las cuales tienen una identidad de secuencia de al menos 30%. Adicionalmente la conservación de los residuos de aminoácidos o dominios funcionalmente importantes, los cuales pueden ser identificados por el alineamiento de varios parálogos ya disponibles, pueden identificar una nueva secuencia como nuevos ortólogos.

Alternativamente las bibliotecas pueden ser sometidas a secuenciamiento de masas y las secuencias obtenidas pueden ser almacenadas en una base de datos de secuencias, las cuales pueden ser seleccionadas para ortólogos putativos mediante diferentes algoritmos de búsqueda, por ejemplo, el algoritmo Bastan para buscar las secuencias de ácidos nucleicos obtenidas con una secuencia de aminoácidos de la invención. Los clones con la identidad de secuencia más alta se utilizan para una determinación de la secuencia completa y los ortólogos pueden ser identificados como se describió anteriormente.

REIVINDICACIONES

1. Un proceso para la asimilación, acumulación y/o utilización potenciada de nitrógeno y el contenido incrementado de nitrógeno total y la eficiencia incrementadas en el uso de nitrógeno (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo nitrógeno de fertilizantes en organismos fotosintéticos activos, el cual comprende, incrementar o generar en un organismo o en una parte o en un compartimiento del mismo la expresión de al menos una molécula de ácido nucleico que comprende una molécula de ácido nucleico seleccionado del grupo consistente de:
- 5 a) una molécula de ácido nucleico que codifica un polipéptido como se muestra en SEQ ID NO: 690 o un fragmento del mismo con una actividad de proteína de transporte de amonio, la cual confiere una asimilación, acumulación y/o utilización potenciada de nitrógeno y contenido de nitrógeno total incrementado y eficiencia incrementada en el uso de nitrógeno,
- 10 b) una molécula de ácido nucleico que comprende una molécula de ácido nucleico como la mostrada en SEQ ID NO: 689 la cual confiere asimilación, acumulación y/o utilización potenciadas de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno;
- 15 c) molécula de ácido nucleico cuya secuencia puede ser deducida a partir de una secuencia de polipéptidos codificada por una molécula de ácido nucleico (a) o (b) como resultado de la degeneración del código genético y que confiere asimilación, acumulación y/o utilización potenciada de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno;
- 20 d) molécula de ácido nucleico que codifica un polipéptido con una actividad de proteína de transporte de amonio que tiene al menos 70% de identidad con la secuencia de aminoácidos del polipéptido codificado por la molécula de ácido nucleico de (a) a (c) y la cual confiere la asimilación, acumulación y/o utilización incrementadas de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno.
2. Un proceso de acuerdo con la reivindicación 1 en donde la molécula de ácido nucleico se origina a partir de un microorganismo.
- 25 3. Método para la identificación de un producto genético que confiere asimilación, acumulación y/o utilización incrementadas de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno (NUE) definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, en un organismos fotosintético o parte del mismo que comprende las siguientes etapas:
- 30 a) poner en contacto las moléculas de ácido nucleico de una muestra, la cual contiene un gen candidato que codifica un producto genético que confiere asimilación, acumulación y/o utilización potenciadas de nitrógeno y contenido incrementado de nitrógeno total en un organismo fotosintético o parte del mismo después de la expresión con la molécula de ácido nucleico tal como se define en la reivindicación 1;
- b) identificar las moléculas de ácido nucleico, que hibridan bajo condiciones restrictivas relajadas con la molécula de ácido nucleico como se define en la reivindicación 1 a) – c);
- 35 c) introducir las moléculas de ácido nucleico candidatas en células anfitrionas apropiadas para asimilación, acumulación y/o utilización potenciadas de nitrógeno y para contenido incrementado de nitrógeno y eficiencia incrementadas en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo nitrógeno de fertilizantes,
- 40 d) expresar las moléculas de ácido nucleico identificadas en las células anfitrionas;
- e) probar la asimilación, acumulación y/o utilización de nitrógeno y el contenido incrementado de nitrógeno total y la eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno; y
- 45 f) identificar la molécula de ácido nucleico y su producto genético cuya expresión confiere una asimilación, acumulación y/o utilización incrementadas de nitrógeno y un contenido incrementado de nitrógeno total y una eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, en la célula anfitriona y en la célula anfitriona después de la expresión en comparación con el tipo silvestre.

4. Un método para la identificación de un producto genético que confiere asimilación, acumulación y/o utilización potenciadas de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno (NUE), definido como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, en un organismo fotosintético o en una parte del mismo que comprende las siguientes etapas:
- 5 a) identificar en un banco de datos moléculas de ácido nucleico de un organismo; las cuales pueden tener un gen candidato que codifique un producto genético que confiere una asimilación, acumulación y/o utilización potenciadas de nitrógeno y un contenido incrementado de nitrógeno total en un organismo fotosintético o una parte del mismo después de la expresión, y el cual puede ser 20% homólogo con la molécula de ácido nucleico tal como se define en la reivindicación 1a) – c);
- 10 b) introducir las moléculas de ácido nucleico candidatas en células anfitrionas apropiadas para asimilación, acumulación y/o utilización potenciadas de nitrógeno y contenido incrementado de nitrógeno total y eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno,
- c) expresar las moléculas de ácido nucleico identificadas en las células anfitrionas;
- 15 d) probar la asimilación, acumulación y/o utilización de nitrógeno y el contenido incrementado de nitrógeno total y la eficiencia incrementada del uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno; y
- e) identificar la molécula de ácido nucleico y su producto genético cuya expresión confiere una asimilación, acumulación y/o utilización incrementadas de nitrógeno y un contenido incrementado de nitrógeno total y una eficiencia incrementada en el uso de nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, en la célula anfitriona y en la célula anfitriona después de la expresión en comparación con el tipo silvestre.
- 20
5. Uso de la molécula de ácido nucleico como se define en la reivindicación 1 para la identificación de una molécula de ácido nucleico que confiere una asimilación, acumulación y/o utilización incrementadas de nitrógeno y un contenido incrementado de nitrógeno total y una eficiencia incrementada en el uso del nitrógeno (NUE), definida como el rendimiento en biomasa por unidad de nitrógeno disponible en el suelo, incluyendo fertilizantes de nitrógeno, en la célula anfitriona después de la expresión en comparación con el tipo silvestre.
- 25
6. Un proceso de acuerdo con las reivindicaciones 1 a 2 caracterizado por un rendimiento total potenciado como se define en la reivindicación 1.
- 30
7. Un proceso de acuerdo con las reivindicaciones 1 a 2 caracterizado por el rendimiento potenciado de plantas tal como se define en la reivindicación 1 bajo condiciones limitantes de nitrógeno.
8. Uso de la molécula de ácido nucleico tal como se define en la reivindicación 1 para obtener un rendimiento total potenciado en biomasa, preferiblemente en semillas o rendimiento en masa proteínica.
- 35
9. Uso de acuerdo con la reivindicación 8 para obtener un rendimiento potenciado tal como se define en la reivindicación 1 bajo condiciones limitadas de nitrógeno.

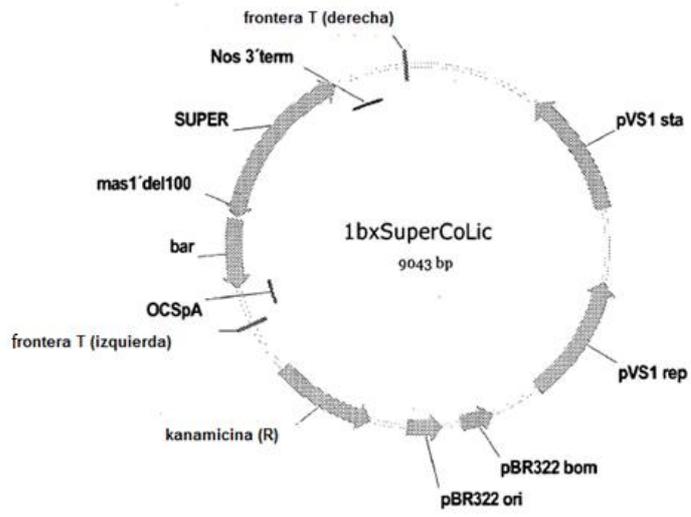


Fig. 1

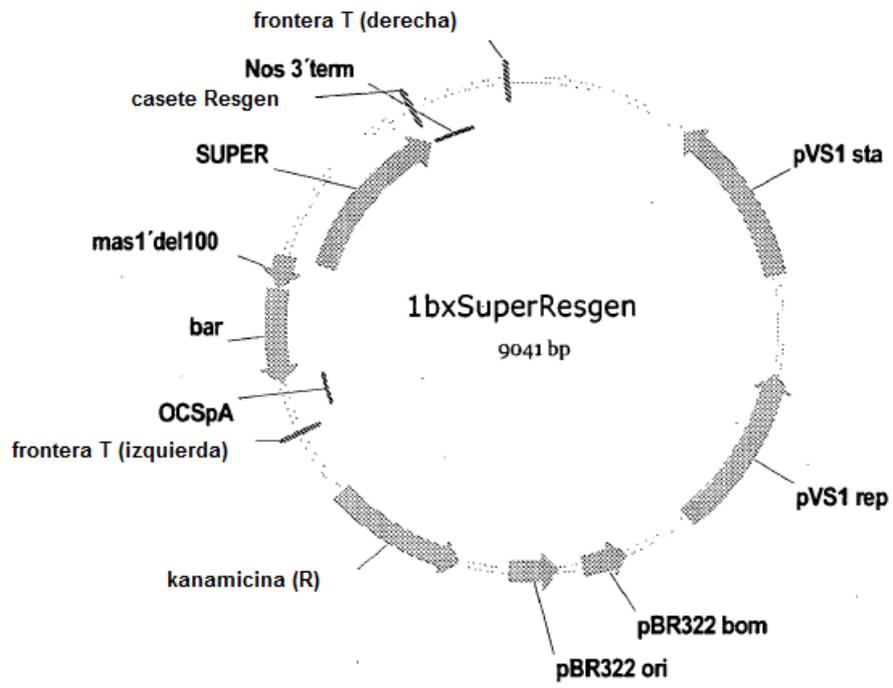


Fig. 2

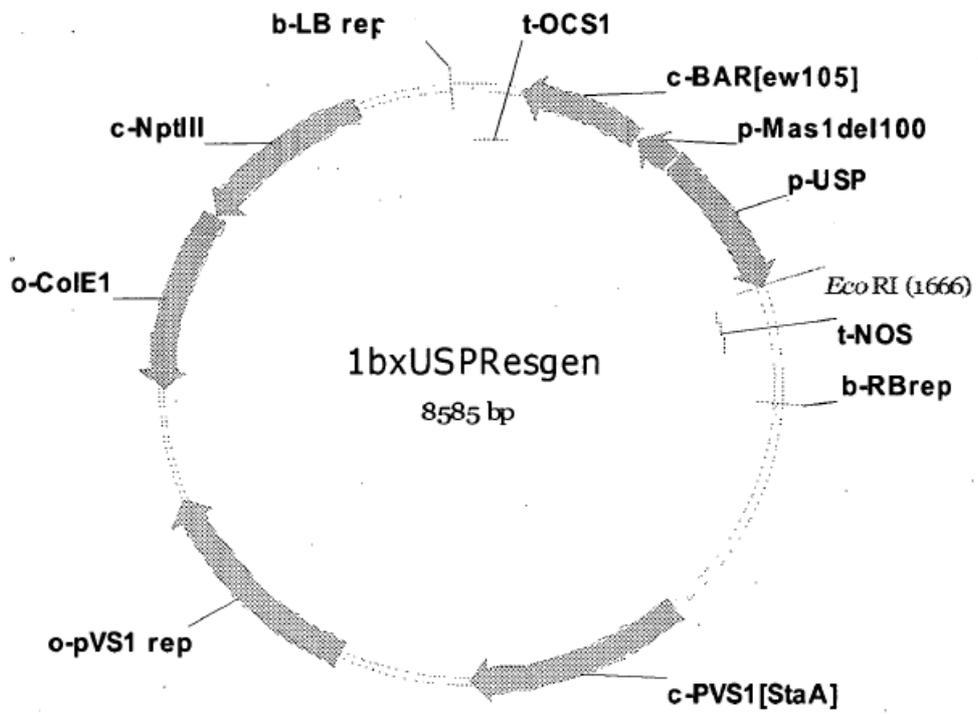


Fig. 3

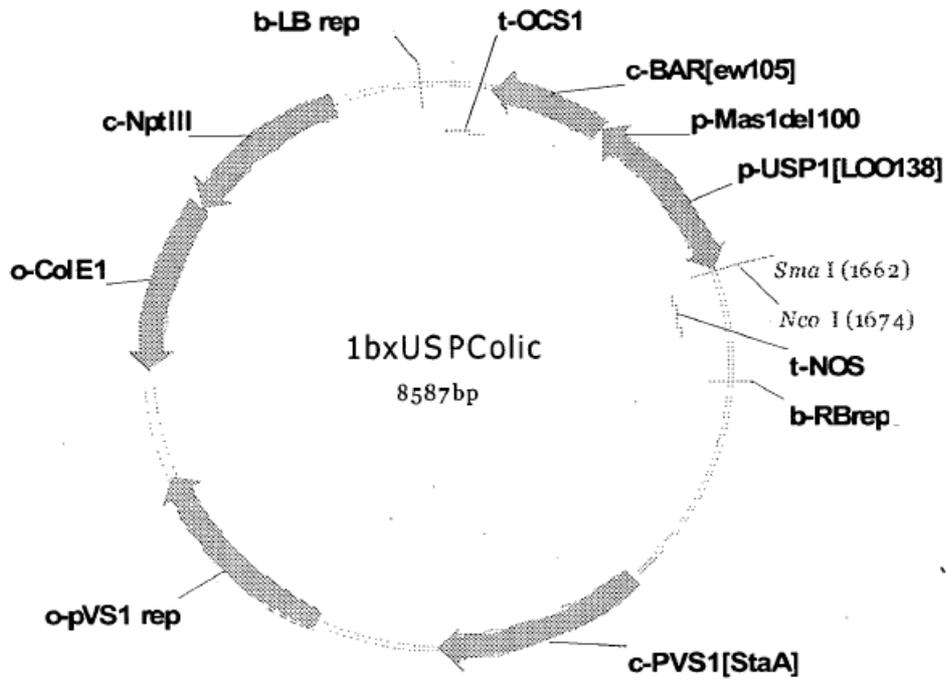


Fig. 4