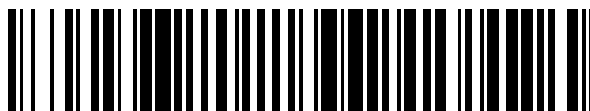


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 502 943**

51 Int. Cl.:

C12N 15/82 (2006.01)

C07K 14/415 (2006.01)

A01H 5/00 (2006.01)

A01H 5/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.09.2004 E 10175334 (1)**

97 Fecha y número de publicación de la concesión europea: **02.07.2014 EP 2298918**

54 Título: **Regulación de la biomasa y tolerancia al estrés de plantas**

30 Prioridad:

23.09.2003 US 669824

16.06.2004 US 870198

26.04.2004 US 565948 P

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

06.10.2014

73 Titular/es:

MENDEL BIOTECHNOLOGY, INC. (100.0%)

3935 Point Eden Way

Hayward, CA 94545, US

72 Inventor/es:

GUTTERSON, NEAL;

KUMIMOTO, RODERICK W.;

KEDDIE, JAMES S.;

SHERMAN, BRADLEY K.;

LIBBY, JEFFREY M.;

JIANG, CAI-ZHONG;

HEARD, JACQUELINE E.;

RATCLIFFE, OLIVER y

HEMPEL, FREDERICK D.

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 502 943 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

Regulación de la biomasa y tolerancia al estrés de plantas.

5 **Relación con solicitudes relacionadas**

La presente solicitud reivindica el beneficio de la Solicitud Provisional de Estados Unidos de cesión común N° 60/565.948, presentada el 26 de abril de 2004; y es una continuación en parte de la Solicitud de Patente no provisional de Estados Unidos N° 10/669.824, presentada el 23 de septiembre de 2003 y la Solicitud de Patente no provisional de Estados Unidos N° 10/870.198, presentada el 16 de junio de 2004.

Campo de la invención

La presente invención se refiere a aumentar la tolerancia de una planta a tensiones abióticas, incluyendo tensiones por frío y osmóticas.

Antecedentes de la invención

Los rasgos de una planta pueden controlarse mediante varios procesos celulares. Un modo importante de manipular ese control es mediante factores de transcripción, proteínas que influyen en la expresión de un gen o conjuntos de genes particulares. Debido a que los factores de transcripción son elementos de control clave de las rutas biológicas, la alteración de los niveles de expresión de uno o más factores de transcripción puede cambiar rutas biológicas completas en un organismo. Las estrategias para manipular las características bioquímicas, de desarrollo o fenotípicas de una planta alterando la expresión de un factor de transcripción pueden dar como resultado plantas y cultivos con propiedades comercialmente valiosas nuevas y/o mejoradas, incluyendo rasgos que mejoran el rendimiento en condiciones sin estrés, o supervivencia y producción durante periodos de estrés abiótico.

Conveniencia de aumentar la biomasa. La capacidad de aumentar la biomasa o tamaño de una planta tendría varias aplicaciones comerciales importantes. Pueden generarse especies de cultivo que producen cultivos mayores, generando mayor rendimiento, por ejemplo, en plantas en las que la parte vegetativa de la planta es comestible.

El aumento del tamaño de la hoja puede ser de interés particular. El aumento de la biomasa de la hoja puede usarse para aumentar la producción de productos farmacéuticos o industriales derivados de plantas. Un aumento de la fotosíntesis vegetal total se consigue normalmente aumentando el área de hoja de la planta. Puede usarse capacidad fotosintética adicional para aumentar el rendimiento derivado de tejido vegetal particular, incluyendo las hojas, raíces, frutos o semillas, o permitir el crecimiento de una planta bajo una menor intensidad de luz o bajo alta intensidad de luz.

La modificación de la biomasa del tejido de la raíz puede ser útil para mejorar la capacidad de una planta para crecer en condiciones ambientales rigurosas, incluyendo sequía o privación de nutrientes, debido a que las raíces mayores pueden alcanzar mejor el agua o los nutrientes o captar agua o nutrientes.

Para algunas plantas ornamentales, la capacidad para proporcionar variedades mayores sería altamente deseable. Para muchas plantas, incluyendo árboles frutales, árboles que se usan para producción de madera, o árboles y arbustos que actúan como pantallas contra el viento o para la vista, el aumento de la estatura proporciona beneficios mejorados en forma de mayor rendimiento o mejor protección.

Problemas asociados con la sequía. Una sequía es un periodo de clima anormalmente seco que persiste el suficiente tiempo para producir un desequilibrio hidrológico grave (por ejemplo daño de cultivos, falta de aporte de agua, etc.). La sequía es el problema relacionado con el clima principal en agricultura y también se clasifica como uno de los principales desastres naturales de todos los tiempos, provocando no solamente daños económicos, sino también pérdidas de vidas humanas. Por ejemplo, la sequía de 1984-1985 en el Cuerno de África condujo a una hambruna que mató a 750.000 personas.

Los problemas para las plantas provocados por baja disponibilidad de agua incluyen tensiones mecánicas provocadas por la retirada del agua celular. La sequía también provoca que las plantas se hagan más susceptibles a diversas enfermedades (Simpson (1981) en Water Stress on Plants, (Simpson, G. M., ed.), Praeger, NY, pp. 235-265). El factor más importante en la resistencia a la sequía es la capacidad de la planta para mantener un alto estado de agua y turgencia manteniendo a la vez la fijación del carbono. Diversos mecanismos adaptativos influyen en esta capacidad, incluyendo el aumento del área de superficie o profundidad de las raíces, ajuste osmótico, y la acumulación de proteínas hidrófilas. El ácido abscísico (ABA) también es un componente regulador esencial de muchas de estas características protectoras.

Además de las muchas regiones de tierra del mundo que son demasiado áridas para la mayoría de, si no todas, las plantas de cultivo, el abuso y sobreutilización del agua disponible dan como resultado una pérdida creciente de tierra utilizable para agricultura, un proceso que, en su extremo, da como resultado desertización. El problema se

compone además del aumento de la acumulación de sal en los suelos, que aumenta la pérdida de agua disponible en suelos.

Problemas asociados con altos niveles salinos. Una de cada cinco hectáreas de tierra irrigada está dañada por la sal, un factor histórico importante en el deterioro de sociedades agrarias antiguas. Se espera que esta condición empeore, reduciendo adicionalmente la disponibilidad de tierra apta para el cultivo y producción de cultivos, ya que ninguno de los cinco cultivos alimentarios más importantes, trigo, maíz, arroz, patatas y soja, pueden tolerar la sal excesiva.

Los efectos perjudiciales de la sal en las plantas son una consecuencia tanto del déficit de agua que da como resultado estrés osmótico (similar al estrés por sequía) como de los efectos de iones de sodio en exceso en procesos bioquímicos críticos. Como con la congelación y la sequía, la alta salinidad provoca déficit de agua. La presencia de alta salinidad hace difícil que las raíces vegetales extraigan agua de su ambiente (Buchanan *et al.* (2000) en *Biochemistry and Molecular Biology of Plants*, American Society of Plant Physiologists, Rockville, MD). La salinidad del suelo es por lo tanto una de las variables más importantes que determinan donde puede desarrollarse una planta. En muchas partes del mundo, grandes áreas de tierra son incultivables debido a salinidad del suelo naturalmente alta. Para aumentar el problema, la salinización de suelos que se usan para producción agrícola es un problema significativo y creciente en regiones que se basan en gran medida en la agricultura.

La tolerancia a la sal es particularmente importante al principio del ciclo de vida de una planta, ya que la evaporación de la superficie del suelo provoca movimiento de agua hacia arriba, y la sal se acumula en la capa de suelo superior donde se sitúan las semillas. Por lo tanto, la germinación normalmente tiene lugar a una concentración salina mucho mayor que el nivel salino medio en el perfil de suelo completo.

Problemas asociados con el calor excesivo. La germinación de muchos cultivos es muy sensible a la temperatura. Un factor de transcripción que potenciaría la germinación en condiciones cálidas sería útil para cultivos que se plantan tarde en la temporada o en climas cálidos. Las plántulas y plantas maduras que se exponen a exceso de calor pueden experimentar choque térmico, que puede surgir en diversos órganos incluyendo hojas y particularmente frutos, cuando la transpiración es insuficiente para superar el estrés por calor. El calor también daña las estructuras celulares, incluyendo orgánulos y el citoesqueleto, y altera la función de membrana (Buchanan *et al.* (2000) mencionado anteriormente).

El choque térmico puede producir una reducción de la síntesis proteica general, acompañada de expresión de proteínas de choque térmico. Las proteínas de choque térmico actúan como chaperonas y están implicadas en el plegamiento de proteínas desnaturalizadas por calor.

El estrés por calor con frecuencia acompaña a condiciones de baja disponibilidad de agua. El calor en sí mismo se ve como un estrés de interacción y aumenta los efectos perjudiciales provocados por condiciones de déficit de agua. La demanda evaporativa muestra aumentos casi exponenciales con aumentos de las temperaturas durante el día, y puede dar como resultado altas tasas de transpiración y bajos potenciales de agua de la planta (Hall *et al.* (2000) *Plant Physiol.* 123: 1449-1458). El daño de alta temperatura al polen se produce casi siempre junto con estrés por sequía, y se produce pocas veces en condiciones de buen riego.

Problemas asociados con condiciones de frío excesivo o enfriamiento. La expresión "sensibilidad al enfriamiento" se ha usado para describir muchos tipos de daño fisiológico producido a temperaturas bajas, pero por encima de la de congelación. La mayoría de los cultivos de orígenes tropicales tales como la soja, arroz, maíz y algodón se dañan fácilmente por enfriamiento. El daño de frío típico incluye marchitamiento, necrosis, clorosis o filtración de iones de las membranas celulares. Los mecanismos subyacentes de sensibilidad al frío no se entienden aún completamente, pero probablemente implican el nivel de saturación de membrana y otras deficiencias fisiológicas. Por ejemplo, se produce con frecuencia fotoinhibición de la fotosíntesis (alteración de la fotosíntesis debido a intensidades de luz altas) en condiciones atmosféricas despejadas después de noches de final del verano/otoño frías. El enfriamiento puede conducir a pérdidas de rendimiento y menor calidad del producto debido al retardo de la maduración del maíz. Otra consecuencia del mal crecimiento es la escasa cobertura del terreno de los campos de maíz en primavera, lo que con frecuencia da como resultado erosión del suelo, aumento de la aparición de malas hierbas, y reducción de la captación de nutrientes. Una captación retardada del nitrógeno mineral también podría conducir a pérdidas aumentadas de nitrato en el agua del suelo. Según algunas estimaciones, el enfriamiento supone pérdidas monetarias en los Estados Unidos (EE.UU.) solamente por detrás de la sequía y la inundación.

Conveniencia de detectar azúcares alterados. Los azúcares son moléculas reguladoras clave que afectan a diversos procesos en plantas superiores incluyendo germinación, crecimiento, floración, senescencia, metabolismo de los azúcares y fotosíntesis. La sacarosa, por ejemplo, es la forma de transporte principal de fotosintatos y se ha mostrado que su flujo a través de las células afecta a la expresión génica y altera la acumulación de compuestos de almacenamiento en semillas (relaciones de fuente-sumidero). También se ha descrito en plantas la detección de hexosa específica de glucosa y está implicada en la división celular y represión de genes de "hambruna" (ciclos fotosintéticos o de glioxilato).

El déficit de agua es un componente habitual de muchas tensiones vegetales. Se produce déficit de agua en células vegetales cuando la tasa de transpiración de la planta completa excede la captación de agua. Además de la sequía, otras tensiones, tales como salinidad y baja temperatura producen deshidratación celular (McCue y Hanson (1990) Trends Biotechnol. 8: 358-362).

La transducción de señal de estrés salino y de sequía consiste en rutas de señalización de homeostasis iónica y osmótica. El aspecto iónico del estrés salino se señaliza mediante la ruta de SOS en la que un complejo de proteína quinasa SOS3-SOS2 sensible a calcio controla la expresión y actividad de transportadores iónicos tales como SOS1. La ruta que regula la homeostasis iónica en respuesta al estrés salino se ha revisado recientemente en Xiong y Zhu (Xiong y Zhu (2002) Plant Cell Environ. 25: 131-139).

El componente osmótico del estrés salino implica reacciones vegetales complejas que solapan con respuestas al estrés por sequía y/o frío. Se han revisado recientemente aspectos comunes de la respuesta al estrés por sequía, frío y salina en Xiong y Zhu (2002) mencionado anteriormente. Estos incluyen:

- (a) cambios transitorios en los niveles de calcio citoplasmático muy pronto en el acontecimiento de señalización (Knight, (2000) Int. Rev. Cytol. 195: 269-324; Sanders *et al.* (1999) Plant Cell 11: 691-706);
- (b) transducción de la señal mediante proteína quinasas dependientes de calcio y/o activadas por mitógeno (CDPK; véase Xiong y Zhu (2002) mencionado anteriormente) y proteína fosfatasa (Merlot *et al.* (2001) Plant J. 25: 295-303; Tähtiharju y Palva (2001) Plant J. 26: 461-470);
- (c) aumentos de los niveles de ABA en respuesta al estrés que desencadena un subconjunto de respuestas (Xiong y Zhu (2002) mencionado anteriormente, y referencias en la misma);
- (d) inositol fosfatos como moléculas señal (al menos para un subconjunto de los cambios de transcripción sensibles a estrés (Xiong *et al.* (2001) Genes Dev. 15: 1971-1984));
- (e) activación de fosfolipasas que a su vez generan una serie diversa de moléculas mensajeras secundarias, algunas de las cuales podrían regular la actividad de las quinasas sensibles al estrés (la fosfolipasa D actúa en una ruta independiente de ABA, Frank *et al.* (2000) Plant Cell 12: 111-124);
- (f) inducción de genes de tipo abundante en embriogénesis tardía (LEA) incluyendo los genes COR/RD que contienen CRT/DRE (Xiong y Zhu (2002) mencionado anteriormente);
- (g) niveles aumentados de antioxidantes y osmolitos compatibles tales como prolina y azúcares solubles (Hasegawa *et al.* (2000) Annu. Rev. Plant Mol. Plant Physiol. 51: 463-499);
- (h) acumulación de especies de oxígeno reactivas tales como superóxido, peróxido de hidrógeno y radicales de hidroxilo (Hasegawa *et al.* (2000) mencionado anteriormente).

La biosíntesis de ABA está regulada por estrés osmótico en múltiples etapas. La señalización de estrés osmótico tanto dependiente de ABA como independiente de ABA modifica en primer lugar factores de transcripción expresados de forma constitutiva, lo que conduce a la expresión de activadores de transcripción de respuesta temprana, que después activan genes efectores de tolerancia al estrés cadena abajo.

Basándose en las características comunes de muchos aspectos de las respuestas al estrés por frío, sequía y salina, puede concluirse que los genes que aumentan la tolerancia al estrés por frío o salina también pueden mejorar la protección frente al estrés por sequía. De hecho, esto ya se ha demostrado para factores de transcripción (en el caso de AtCBF/DREB1) y para otros genes tales como OsCDPK7 (Saijo *et al.* (2000) Plant J. 23: 319-327), o AVP1 (una bomba de protones de pirofosfatasa vacuolar; Gaxiola *et al.* (2001) Proc. Natl. Acad. Sci. USA 98: 11444-11449).

La presente invención se refiere a métodos y composiciones para producir plantas transgénicas con rasgos modificados, particularmente rasgos que abordan la tolerancia al estrés y el aumento de la biomasa. Estos rasgos pueden proporcionar un valor significativo por que puede conseguirse un mayor rendimiento y/o la planta puede después desarrollarse en ambientes hostiles, en los que, por ejemplo, la alta o baja temperatura, baja disponibilidad de agua o alta salinidad pueden limitar o evitar el crecimiento de plantas transgénicas.

Los inventores han identificado polinucleótidos que codifican factores de transcripción, incluyendo G1073 (atHRC1), y equivalentes en el clado G1073 de polipéptidos de factores de transcripción, han desarrollado numerosas plantas transgénicas usando estos polinucleótidos y han analizado las plantas con respecto a su biomasa y tolerancia a tensiones abióticas. Al hacerlo, han identificado importantes secuencias polinucleotídicas y polipeptídicas para producir plantas y cultivos comercialmente valiosos así como los métodos para realizarlos y usarlos. Otros aspectos y realizaciones de la invención se describen posteriormente y pueden derivarse de las enseñanzas de la presente divulgación en su totalidad.

Sumario de la invención

La invención proporciona una planta transgénica que tiene mayor tolerancia a una estrés abiótico que una planta de control, comprendiendo dicha planta transgénica un polinucleótido recombinante que codifica un polipéptido; en la que el polipéptido comprende un primer dominio que comprende SEC ID N°: 72, y un segundo dominio que tiene al menos 62 % de identidad de secuencia de aminoácidos con el segundo dominio conservado de SEC ID N°: 2 como

se expone en la Tabla 1 y el segundo dominio comprende SEC ID N°: 63 o SEC ID N°: 64; en la que la expresión del polipéptido en la planta transgénica está regulada por un promotor específico de raíz, un promotor específico de tejido de hoja, un promotor específico epidérmico, o un promotor inducible por estrés, y la expresión del polipéptido en la planta transgénica confiere a la planta transgénica dicha mayor tolerancia al estrés abiótico; no comprendiendo dicha planta de control el polinucleótido recombinante; y dicha planta transgénica tiene menos anomalías del desarrollo indeseables o menos efectos morfológicos adversos que una segunda planta de control que expresa el polipéptido bajo el control regulador de un promotor constitutivo.

La invención se define adicionalmente en las reivindicaciones adjuntas.

Un vector, plásmido u otra construcción de ácido nucleico que contiene un polinucleótido de factor de transcripción y elementos reguladores para regulación de la transcripción del polinucleótido se describe en el presente documento. SEC ID N°: 2 es una secuencia denominada en el presente documento G1073. El clado G1073 de polipéptidos de factores de transcripción deriva de un ancestro polipeptídico común (Figura 4), y comprende un dominio de gancho AT y un segundo dominio conservado. Las secuencias de miembros del clado G1073 que se han usado con éxito para conferir tolerancia aumentada al estrés abiótico derivan de varias especies diversas, incluyendo dicotiledóneas tales como *Arabidopsis* y soja, y monocotiledóneas tales como arroz. Los polipéptidos de miembros del clado G1073 comprenden un dominio de gancho AT y un segundo dominio conservado, que a su vez comprende las secuencias SEC ID N°: 72 (en el dominio de gancho At) y SEC ID N°: 63 o SEC ID N°: 64 (en el segundo dominio conservado). El vector, plásmido o construcción de ácido nucleico también puede contener un elemento regulador. Este puede ser un promotor constitutivo, inducible o específico de tejido que controla la expresión de la secuencia polinucleotídica. El vector, plásmido o construcción de ácido nucleico se introduce después en una planta diana (una planta que no se ha transformado aún con el vector, plásmido o construcción de ácido nucleico), transformado de este modo la planta en una que tiene biomasa y/o tolerancia aumentada al estrés abiótico, en relación con plantas de control. Los promotores inducibles pueden incluir, por ejemplo, los promotores DREB2A y RD29A. El promotor RD29A se ha usado con éxito para regular la expresión del polinucleótido G1073 y conferir tolerancia a aumento del estrés abiótico. Los ejemplos de promotores específicos de tejido que se han usado de esta manera incluyen el promotor ARSK1 (específico de raíz), el promotor CUT1 (específico de epidermis) y el promotor RBSC3 (específico de hoja). El uso de promotores específicos de tejido o inducibles mitiga los efectos morfológicos indeseables que pueden asociarse con la sobreexpresión constitutiva de miembros del clado G1073 (por ejemplo, cuando no es deseable un tamaño aumentado).

El método también abarca aumentar la tolerancia de una planta al estrés abiótico con un enfoque de vector múltiple. En este caso, un primer vector que comprende un promotor clonado delante de un dominio de unión a ADN LexA fusionado con un dominio de activación de GAL4 se introduce en la planta. Se introduce después un segundo vector en la misma planta; este segundo vector comprende una secuencia polinucleotídica que codifica un miembro del clado del polipéptido G1073. Después se permite que la planta sobreexpresa el polipéptido miembro de G1073, lo que aumenta la tolerancia al estrés abiótico. El promotor clonado delante de un dominio de unión a ADN LexA puede ser, por ejemplo, el promotor RD29A, aunque se prevén fácilmente otros promotores que actúan de una manera similar y que pueden expresarse de una manera inducible o específica de tejido y también están abarcados por la presente invención.

Los métodos abarcados por la invención también pueden extenderse a técnicas de propagación usadas para generar plantas. Por ejemplo, una planta diana que se ha transformado con un polinucleótido que codifica un miembro del clado de polipéptidos G1073 y que tiene mayor biomasa y/o tolerancia al estrés abiótico que un control de tipo silvestre o no transformado puede "autofecundarse" (es decir, autopolinizarse) o cruzarse con otra planta para producir semillas que comprenden un polinucleótido recombinante como se define en la reivindicación 1. Pueden cultivarse plantas descendientes a partir de esta semilla, generando de este modo plantas descendientes transformadas con mayor tolerancia al estrés abiótico que las plantas de control.

La invención también abarca plantas transgénicas (y semillas que comprenden un polinucleótido recombinante como se define en la reivindicación 1 de estas plantas transgénicas) producidas por los presentes métodos.

Breve descripción del listado de secuencias y dibujos

El Listado de Secuencias proporciona secuencias polinucleotídicas y polipeptídicas ejemplares de la invención. Los rasgos asociados con el uso de las secuencias se incluyen en los ejemplos.

CD-ROM1 y CD-ROM2 son discos compactos leíbles por ordenador de memoria solamente de lectura. Cada uno contiene una copia del Listado de Secuencias en formato de texto ASCII. El Listado de Secuencias se llama "MBI0068PCT.ST25.txt" y tiene 145 kilobytes de tamaño. Las copias del Listado de Secuencias en los discos CD-ROM se incorporan por la presente por referencia en su totalidad.

La Figura 1 muestra una estimación conservadora de relaciones filogenéticas entre los órdenes de plantas con flores (modificada de Angiosperm Phylogeny Group (1998) Ann. Missouri Bot. Gard. 84: 1-49). Las plantas con un único cotiledón (monocotiledóneas) son un clado monofilético anidado dentro de al menos dos linajes principales de

dicotiledóneas; las eudicotiledóneas se dividen adicionalmente en rósidas y astéridas. La *Arabidopsis* es una eudicotiledónea rósida clasificada dentro del orden Brassicales; el arroz es un miembro del orden de monocotiledóneas Poales. La Figura 1 se adaptó de Daly *et al.* (2001) Plant Physiol. 127: 1328-1333.

- 5 La Figura 2 muestra un dendograma filogenético que representa relaciones filogenéticas de taxones de plantas superiores, incluyendo clados que contienen tomate y *Arabidopsis*; adaptado de Ku *et al.* (2000) Proc. Natl. Acad. Sci. USA 97: 9121-9126; y Chase *et al.* (1993) Ann. Missouri Bot. Gard. 80: 528-580.

- 10 La Figura 3 representa la estructura de dominio de proteínas de gancho AT, representadas por una representación esquemática de la proteína G1073. Las regiones que comprenden el dominio de unión a ADN, el segundo dominio conservado y la región ácida se representan por las letras "A", "B" y "C", respectivamente. Las flechas indican sitios de fosforilación CK2 y PKC potenciales. Se localiza un dominio de unión a ADN conservativo en las posiciones 34 a 42.

- 15 La Figura 4 muestra ortólogos de cultivo que se identificaron mediante análisis de BLAST de fuentes de datos patentadas y públicas. Se generó después un árbol filogenético usando ClustalX basado en secuencias proteicas completas. Las secuencias que se indican con un número de "GID" que comienza con la letra "G" seguida de "At" se refieren a secuencias de *Arabidopsis*; las secuencias con "Gm" son secuencias de soja, y las "Os" son secuencias de arroz. Pueden encontrarse veinticuatro miembros del clado G1073 derivados de *Arabidopsis*, maíz y soja dentro de la caja grande; estas secuencias representan ejemplos del clado y no se pretende que representen el clado completo, que tiene muchos más miembros de otras especies vegetales. Se ha mostrado que un número representativo de miembros del clado G1073 confieren propiedades ventajosas a plantas cuando se sobreexpresan; se ha mostrado que las secuencias que aparecen con un superíndice "a" confieren tolerancia aumentada a aumento del estrés abiótico, y se ha mostrado que las secuencias que aparecen con un superíndice "b" confieren biomasa aumentada. Muchas de las secuencias restantes no se han ensayado aún en plantas que las sobreexpresen. Varias secuencias de miembros del clado G1073 que también se ha mostrado que confieren estrés abiótico en plantas no se muestran en la Figura 4, pero se desvelan en el Ejemplo VII.

- 30 En las Figuras 5A-5H, se muestran los alineamientos de varias proteínas de gancho de AT identificadas en la Figura 4, e incluyen miembros del clado de *Arabidopsis* (G1067, G1069, G1073, G1667, G2153, G2156, G2789), soja (G3456, G3459, G3460), y arroz (G3399, G3400, G3401, G3407) que se ha mostrado que confieren rasgos similares en plantas cuando se sobreexpresan (el clado se indica por la caja grande). También se muestran los dominios conservados de gancho AT (Figura 5C) y los segundos dominios conservados que abarcan las Figuras 5D a 5F, como se indica por las barras en la parte superior e inferior de la página, respectivamente.

- 35 Las Figuras 6A y 6B muestran secciones transversales de tallo de *Arabidopsis* de tipo silvestre (izquierda) y que sobreexpresa G1073 (derecha). En el tallo de la planta que sobreexpresa G1073, los haces vasculares son mayores (contienen más células en las áreas del floema y xilema) y las células de la corteza están agrandadas.

- 40 Muchas plantas de *Arabidopsis* que sobreexpresan G1073 (Figura 7A, ejemplo a la derecha) son mayores que las plantas de control de tipo silvestre (Figura 7A, izquierda). Esta distinción también es cierta para los órganos florales, que, como se ve en la Figura 7B, son significativamente mayores en la planta que sobreexpresa G1073 a la derecha que en la de la planta de tipo silvestre a la izquierda.

- 45 La Figura 8 es una gráfica que compara el número de silicuas en plantas de control (tipo silvestre) y 35S::G1073 que indica como está asociado el número de semillas con el número aumentado de silicuas por planta visto en las líneas con sobreexpresión.

- 50 Como se ve en las Figuras 9A y 9B, G1073 actúa tanto en soja como en tomate para aumentar la biomasa. En la Figura 9A, la planta de soja mayor a la derecha sobreexpresa G1073. Las hojas de tomate de varias líneas que sobreexpresaban G1073 eran mucho mayores que las de plantas de tomate de tipo silvestre, como se ve en la Figura 9B comparando las hojas de la planta con sobreexpresión a la izquierda y la de una planta de tipo silvestre a la derecha.

- 55 La Figura 10A es una fotografía de una planta de *Arabidopsis* que sobreexpresa el gen de monocotiledónea G3399, un ortólogo de arroz de G1073. El fenotipo de mayor tamaño y masa es el mismo que el fenotipo conferido por G1073 de *Arabidopsis* y sus secuencias parálogas G1067, G2153 y G2157. La Figura 10B muestra de forma similar los efectos de otro ortólogo de arroz G3407, a los siete días. La planta con sobreexpresión a la izquierda es aproximadamente 50 % mayor que la planta de control a la derecha.

- 60 La Figura 11 muestra los efectos de la sobreexpresión de G3460, un ortólogo de soja de G1073, en la morfología de la planta. Treinta y ocho días después de plantar, la planta con sobreexpresión a la izquierda tiene hojas significativamente más anchas y más masivas que la planta de control a la derecha. El que sobreexpresa también demuestra desarrollo tardío, una característica vista también cuando se sobreexpresan G1073 o sus parálogos.

65

La Figura 12 muestra los efectos de la sobreexpresión de G3460, un ortólogo de soja de G1073, en plantas de *Arabidopsis* sometidas a un ensayo de desecación basado en placa. Las plántulas que sobreexpresan G3460 son más tolerantes al tratamiento por desecación, como se demuestra por el mayor tamaño, mayor masa de la raíz, y color más verde de las plantas a la izquierda que las plantas de control a la derecha.

Descripción detallada

La presente invención se refiere a polinucleótidos y polipéptidos para modificar fenotipos de plantas, particularmente los asociados con tolerancia a aumento del estrés abiótico. A lo largo de la presente divulgación, se hace referencia a diversas fuentes de información y/o se incorporan específicamente. Las fuentes de información incluyen artículos de revistas científicas, documentos de patente, libros de texto y direcciones de páginas inactivas en navegador Web.

Como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares “un” y “el” incluyen la referencia plural a no ser que el contexto claramente dicte otra cosa. Por lo tanto, por ejemplo, una referencia a “una célula hospedadora” incluye una pluralidad de dichas células hospedadoras, y una referencia a “un estrés” es una referencia a una o más tensiones y equivalentes de las mismas conocidas por los expertos en la materia, y así sucesivamente.

Definiciones

“Molécula de ácido nucleico” se refiere a un oligonucleótido, polinucleótido o cualquier fragmento de los mismos. Puede ser ADN o ARN de origen genómico o sintético, monocatenario o bicatenario, y combinado con carbohidratos, lípidos, proteínas u otros materiales para realizar una actividad particular tal como transformación o formar una composición útil tal como un ácido nucleico peptídico (PNA).

Un “polinucleótido” es una molécula de ácido nucleico que comprende una pluralidad de nucleótidos polimerizados, por ejemplo, al menos aproximadamente 15 nucleótidos polimerizados consecutivos. Un polinucleótido puede ser un ácido nucleico, oligonucleótido, nucleótido o cualquier fragmento de los mismos. En muchos casos, un polinucleótido comprende una secuencia de nucleótidos que codifica un polipéptido (o proteína) o un dominio o fragmento del mismo. Adicionalmente, el polinucleótido puede comprender un promotor, un intrón, una región potenciadora, un sitio de poliadenilación, un sitio de inicio de la traducción, regiones no traducidas 5' o 3', un gen indicador, un marcador seleccionable o similares. El polinucleótido puede ser ADN o ARN monocatenario o bicatenario. El polinucleótido comprende opcionalmente bases modificadas o una cadena principal modificada. El polinucleótido puede ser, por ejemplo, ADN o ARN genómico, un transcrito (tal como un ARNm), un ADNc, un producto de PCR, un ADN clonado, un ADN o ARN sintético, o similares. El polinucleótido puede estar combinado con carbohidratos, lípidos, proteínas u otros materiales para realizar una actividad particular tal como transformación o formar una composición útil tal como un ácido nucleico peptídico (PNA). El polinucleótido puede comprender una secuencia en orientaciones bien con sentido o bien antisentido. “Oligonucleótido” es sustancialmente equivalente a los términos amplímero, cebador, oligómero, elemento, diana y sonda y es preferentemente monocatenario.

“Gen” o “secuencia génica” se refiere a la secuencia codificante parcial o completa de un gen, su complemento, y sus regiones no traducidas 5' o 3'. Un gen también es una unidad de herencia funcional, y en término físicos es un segmento particular o secuencia de nucleótidos a lo largo de una molécula de ADN (o ARN, en el caso de virus de ARN) implicada en la producción de una cadena polipeptídica. Esta última puede someterse a procesamiento posterior tal como modificación química o plegamiento para obtener una proteína o un polipéptido funcional. Un gen puede aislarse, aislarse parcialmente, o encontrarse dentro del genoma de un organismo. Por ejemplo, un gen de factor de transcripción codifica un polipéptido de factor de transcripción, que puede ser funcional o requerir procesamiento para actuar como un iniciador de la transcripción.

Funcionalmente, los genes pueden definirse por el ensayo de cis-trans, un ensayo genético que determina si se producen dos mutaciones en el mismo gen y que puede usarse para determinar los límites de la unidad genéticamente activa (Rieger *et al.* (1976) Glossary of Genetics and Cytogenetics: Classical and Molecular, 4ª ed., Springer Verlag, Berlín). Un gen generalmente incluye las regiones precedentes (“líderes”; cadena arriba) y siguientes (“colas”; cadena abajo) a la región codificante. Un gen puede incluir también secuencias intermedias, no codificantes, denominadas “intrones”, localizadas entre segmentos codificantes individuales, denominados “exones”. La mayoría de los genes tienen una región promotora asociada, una secuencia reguladora 5' del codón de inicio de la transcripción (hay algunos genes que no tienen un promotor identificable). La función de un gen también puede calcularse por potenciadores, operadores y otros elementos reguladores.

Un “polinucleótido recombinante” es un polinucleótido que no está en su estado nativo, por ejemplo, el polinucleótido comprende una secuencia de nucleótidos no hallada en la naturaleza, o el polinucleótido está en un contexto distinto de en el que se encuentra en la naturaleza, por ejemplo, separado de secuencias de nucleótidos con las que normalmente está en proximidad en la naturaleza, o adyacente a (o contiguo con) secuencias de nucleótidos con las que normalmente no está en proximidad. Por ejemplo, la secuencia en cuestión puede clonarse en un vector, o recombinarse de otro modo con un o más ácidos nucleicos adicionales.

Un “polinucleótido aislado” es un polinucleótido, bien de origen natural o bien recombinante, que está presente fuera de la célula en la que normalmente se encuentra en la naturaleza, bien purificado o no. Opcionalmente, un polinucleótido aislado se somete a uno o más procedimientos de enriquecimiento o purificación, por ejemplo, lisis celular, extracción, centrifugación, precipitación o similares.

Un “polipéptido” es una secuencia de aminoácidos que comprende una pluralidad de restos de aminoácidos polimerizados consecutivos, por ejemplo, al menos aproximadamente 15 restos de aminoácidos polimerizados consecutivos. En muchos casos, un polipéptido comprende una secuencia de restos de aminoácidos polimerizados que es un factor de transcripción o un dominio o parte o fragmento del mismo. Adicionalmente, el polipéptido puede comprender: (i) un dominio de localización; (ii) un dominio de activación; (iii) un dominio de represión; (iv) un dominio de oligomerización; (v) un dominio de unión a ADN; o similares. El polipéptido comprende opcionalmente restos de aminoácidos modificados, restos de aminoácidos de origen natural no codificados por un codón, restos de aminoácidos de origen no natural.

“Proteína” se refiere a una secuencia de aminoácidos, oligopéptido, péptido, polipéptido o partes de los mismos bien de origen natural o bien sintético.

“Parte”, como se usa en el presente documento, se refiere a cualquier parte de una proteína usada para cualquier fin, pero especialmente para la exploración de una biblioteca de moléculas que se une específicamente con esa parte o para la producción de anticuerpos.

Un “polipéptido recombinante” es un polipéptido producido por la traducción de un polinucleótido recombinante. Un “polipéptido sintético” es un polipéptido creado por polimerización consecutiva de restos de aminoácidos aislados usando métodos bien conocidos en la técnica. Un “polipéptido aislado”, bien si es un polipéptido de origen natural o uno recombinante, está más enriquecido en (o fuera de) una célula que el polipéptido en su estado natural en una célula de tipo silvestre, por ejemplo, más de aproximadamente 5 % enriquecido, más de aproximadamente 10 % enriquecido, o más de aproximadamente 20 %, o más de aproximadamente 50 %, o más, enriquecido, es decir, indicado como alternativa: 105 %, 110 %, 120 %, 150 % o más, enriquecido en relación con el tipo silvestre normalizado al 100 %. Dicho enriquecimiento no es el resultado de una respuesta natural de una planta de tipo silvestre. Como alternativa, o adicionalmente, el polipéptido aislado se separa de otros componentes celulares con los que está normalmente asociado, por ejemplo, por cualquiera de los diversos métodos de purificación de proteínas del presente documento.

“Homología” se refiere a similitud de secuencia entre una secuencia de referencia y al menos un fragmento de un inserto de clon recién secuenciado o su secuencia de aminoácidos codificada.

“Identidad” o “similitud” se refiere a similitud de secuencia entre dos secuencias polinucleotídicas o entre dos secuencias polipeptídicas, siendo la identidad una comparación más estricta. Las frases “porcentaje de identidad” y “% de identidad” se refieren al porcentaje de similitud de secuencia hallado en una comparación de dos o más secuencias polinucleotídicas o dos o más secuencias polipeptídicas. La “similitud de secuencia” se refiere al porcentaje de similitud en secuencia de pares de bases (como se determina por cualquier método adecuado) entre dos o más secuencias polinucleotídicas. Dos o más secuencias pueden ser cualquiera de 0-100 % similares, o cualquier valor entero entre ellos. La identidad o similitud puede determinarse comparando una posición en cada secuencia que pueda alinearse para fines de comparación. Cuando una posición en la secuencia comparada está ocupada por la misma base nucleotídica o aminoácido, entonces las moléculas son idénticas en esa posición. Un grado de similitud o identidad entre secuencias polinucleotídicas está en función del número de nucleótidos idénticos, coincidentes o correspondientes en posiciones compartidas por las secuencias polinucleotídicas. Un grado de identidad de secuencias polinucleotídicas está en función del número de aminoácidos idénticos en posiciones correspondientes compartidas por las secuencias polipeptídicas. Un grado de homología o similitud de las secuencias polipeptídicas está en función del número de aminoácidos en posiciones correspondientes compartidas por las secuencias polipeptídicas.

“Alineamiento” se refiere a varias bases de nucleótidos o secuencias de restos de aminoácidos alineadas por comparación a lo largo de modo que puedan visualizarse e identificarse fácilmente los componentes en común (es decir, bases de nucleótidos o restos de aminoácidos en posiciones correspondientes). La fracción o el porcentaje de componentes en común están relacionados con la homología o identidad entre las secuencias. Pueden usarse alineamientos tales como los de las Figuras 5A-5H para identificar dominios conservados y relación dentro de estos dominios. Un alineamiento puede determinarse adecuadamente por medio de programas informáticos conocidos en la técnica, tales como el software MACVECTOR (1999) (Accehys, Inc., San Diego, CA).

Un “dominio conservado” o “región conservada” como se usa en el presente documento se refiere a una región en secuencias polinucleotídicas o polipeptídicas heterólogas en las que hay un grado relativamente alto de identidad de secuencia entre las distintas secuencias. Un “dominio de gancho AT”, tal como se encuentra en un miembro polipeptídico de la familia del factor de transcripción de gancho AT, es un ejemplo de un dominio conservado. Con respecto a polinucleótidos que codifican factores de transcripción desvelados en el presente documento, un dominio conservado es preferentemente de al menos nueve pares de bases (pb) de longitud. Un “dominio conservado”, con

respecto a polipéptidos de gancho AT desvelados en el presente documento se refiere a un dominio dentro de una familia de factores de transcripción que muestra un mayor grado de homología de secuencia, tal como al menos aproximadamente 62 % de identidad de secuencia incluyendo sustituciones conservativas, al menos aproximadamente 63 %, al menos aproximadamente 65 %, al menos aproximadamente 67 %, al menos aproximadamente 68 %, al menos aproximadamente 69 %, al menos aproximadamente 71 %, al menos aproximadamente 78 % o al menos aproximadamente 89 % de identidad de secuencia de restos de aminoácidos con el dominio conservado. Las secuencias que poseen o codifican dominios conservados que cumplen estos criterios de porcentaje de identidad, y que tienen actividad biológica comparable con las presentes secuencias de factores de transcripción, siendo por lo tanto miembros del clado G1073 de polipéptidos de factores de transcripción, están abarcadas por la invención. Puede indicarse que un fragmento o dominio está fuera de un dominio conservado, fuera de una secuencia consenso, o fuera de un sitio de unión a ADN consenso que se sabe que existe o que existe para una clase, familia o subfamilia de factor de transcripción particular. En este caso, el fragmento o dominio no incluirá los aminoácidos exactos de una secuencia consenso o sitio de unión a ADN consenso de una clase, familia o subfamilia de factor de transcripción, o los aminoácidos exactos de una secuencia consenso de factor de transcripción particular o sitio de unión a ADN consenso. Además, un fragmento, región o dominio particular de un polipéptido, o un polinucleótido que codifica un polipéptido, puede estar “fuera de un dominio conservado” si todos los aminoácidos del fragmento, región o dominio quedan fuera de un dominio o dominios conservados definidos para un polipéptido o proteína. Las secuencias que tienen grados menores de identidad pero actividad biológica comparable se consideran equivalentes.

Como reconoce un experto en la materia, los dominios conservados pueden identificarse como regiones o dominios de identidad con una secuencia consenso específica (véase, por ejemplo, Riechmann *et al.* (2000) *Science* 290: 2105-2110). Por lo tanto, usando métodos de alineamiento bien conocidos en la técnica, pueden determinarse los dominios conservados de los factores de transcripción de plantas para las proteínas de gancho AT (Reeves y Beckerbauer (2001) *Biochim. Biophys. Acta* 1519: 13-29; y Reeves (2001) *Gene* 277: 63-81).

Los dominios conservados para SEC ID N°: 2, 4, 6, 8, 10, 12, 14, 16, 18, 26, 30, 38, 40, 42, 67 y 69 se enumeran en la Tabla 1. Además, los polipéptidos de la Tabla 1 tienen dominios de gancho AT y segundo conservado específicamente indicados por sitios de inicio y parada. Una comparación de las regiones de los polipéptidos en SEC ID N°: 2, 4, 6, 8, 10, 12, 14, 16, 18, 26, 30, 38, 40, 42, 67 y 69 permite que un experto en la materia (véase, por ejemplo, Reeves y Nisson (1995) *Biol. Chem.* 265: 8573-8582) identifique dominios de gancho AT o dominios conservados para cualquiera de los polipéptidos enumerados o indicados en la presente divulgación.

“Complementario” se refiere al enlace de hidrógeno natural por formación de pares de bases entre purinas y pirimidinas. Por ejemplo, la secuencia A-C-G-T (5' → 3') forma enlaces de hidrógeno con sus complementos A-C-G-T (5' → 3') o A-C-G-U (5' → 3'). Dos moléculas monocatenarias pueden considerarse parcialmente complementarias, si solamente algunos de los nucleótidos se enlazan, o “completamente complementarias” si todos los nucleótidos se enlazan. El grado de complementariedad entre cadenas de ácido nucleico afecta a la eficacia y fuerza de hibridación y reacciones de amplificación. “Completamente complementario” se refiere al caso en el que se produce enlace entre cada par de bases y su complemento en un par de secuencias, y las dos secuencias tienen el mismo número de nucleótidos.

Las expresiones “altamente riguroso” o “condición altamente rigurosa” se refieren a las condiciones que permiten la hibridación de cadenas de ADN cuyas secuencias son altamente complementarias, en las que estas mismas condiciones excluyen la hibridación de ADN significativamente desapareados. Las secuencias polinucleotídicas capaces de hibridar en condiciones rigurosas con los polinucleótidos de la presente invención pueden ser, por ejemplo, variantes de las secuencias polinucleotídicas desveladas, incluyendo variantes alélicas o de corte y empalme, o secuencias que codifican ortólogos o parálogos de polipéptidos desvelados en el presente documento. Se desvelan métodos de hibridación de ácidos nucleicos en detalle en Kashima *et al.* (1985) *Nature* 313: 402-404, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2ª Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (“Sambrook”), y en Haymes *et al.* “Nucleic Acid Hybridization: A Practical Approach”, IRL Press, Washington, D. C. (1985).

En general, la rigurosidad se determina por la temperatura, fuerza iónica y concentración de agentes desnaturizantes (por ejemplo, formamida) usados en un procedimiento de hibridación y lavado (para una descripción más detallada del establecimiento y determinación de la rigurosidad, véase la sección “Identificación de Polinucleótidos o Ácidos Nucleicos por Hibridación” posterior). El grado en el que dos ácidos nucleicos hibridan en diversas condiciones de rigurosidad se correlaciona con el alcance de su similitud. Por lo tanto, secuencias de aminoácidos similares de diversas fuentes, tales como dentro del genoma de una planta (como en el caso de parálogos) o de otra planta (como en el caso de ortólogos) que pueden realizar funciones similares pueden aislarse basándose en su capacidad para hibridar con secuencias de factores de transcripción conocidas. Son posibles numerosas variaciones en las condiciones y medios por lo que puede realizarse hibridación de ácido nucleico para aislar secuencias de factores de transcripción que tienen similitud con secuencias de factores de transcripción conocidas en la técnica y no se limitan a las desveladas de forma explícita en el presente documento. Dicho enfoque puede usarse para aislar secuencias polinucleotídicas que tienen diversos grados de similitud con secuencias de factores de transcripción desveladas, tales como, por ejemplo, factores de transcripción codificados que tienen 62 %

o mayor identidad con el dominio de gancho AT de factores de transcripción desvelados.

Los términos “parálogo” y “ortólogo” se definen posteriormente en la sección titulada “Ortólogos y Parálogos”. Brevemente, los ortólogos y parálogos son genes relacionados evolutivamente que tienen secuencias y funciones similares. Los ortólogos son genes estructuralmente relacionados en diferentes especies que derivan por un acontecimiento de especiación. Los parálogos son genes estructuralmente relacionados dentro de una única especie que derivan por un acontecimiento de duplicación.

El término “equiválogo” describe miembros de un conjunto de proteínas homólogas que están conservadas con respecto a función desde su último ancestro común. Las proteínas relacionadas se agrupan en familias de equiválogos, y de otro modo en familias de proteínas con otros tipos de homología definidos jerárquicamente. Esta definición se proporciona en el sitio Web de la red global (www) del Instituto para la Investigación Genómica (TIGR) “tigr.org” bajo el encabezamiento “Terms associated with TIGRFAMs”.

El término “variante”, como se usa en el presente documento, puede referirse a polinucleótidos o polipéptidos, que difieren de los polinucleótidos o polipéptidos desvelados en el presente documento, respectivamente, en su secuencia entre sí, y como se expone posteriormente.

Con respecto a variantes polinucleotídicas, las diferencias entre polinucleótidos y variantes polinucleotídicas desvelados en el presente documento están limitadas de modo que las secuencias de nucleótidos de las primeras y las últimas son muy similares en general y, en muchas regiones, idénticas. Debido a la degeneración del código genético, las diferencias entre las primeras y últimas secuencias de nucleótidos pueden ser silenciosas (es decir, los aminoácidos codificados por el polinucleótido son iguales, y las secuencias polinucleotídicas variantes codifican la misma secuencia de aminoácidos que el polinucleótido desvelado en el presente documento). Las secuencias de nucleótidos variantes pueden codificar diferentes secuencias de aminoácidos, en cuyo caso dichas diferencias de nucleótidos darán como resultado sustituciones, adiciones, delecciones, inserciones, truncamientos o fusiones de aminoácidos con respecto a las secuencias polinucleotídicas desveladas similares. Estas variaciones pueden dar como resultado variantes polinucleotídicas que codifican polipéptidos que comparten al menos una característica funcional. La degeneración del código genético también dicta que muchos polinucleótidos variantes diferentes pueden codificar polipéptidos idénticos y/o sustancialmente similares además de las secuencias ilustradas en el Listado de Secuencias.

Una variante de un ácido nucleico de factor de transcripción enumerado en el Listado de Secuencias es una que tiene una secuencia que difiere de una de las secuencias polinucleotídicas en el Listado de Secuencias, o una secuencia complementaria, que codifica un polipéptido funcionalmente equivalente (es decir, un polipéptido que tiene algún grado de actividad biológica equivalente o similar) pero difiere en secuencia de la secuencia en el Listado de Secuencias, debido a la degeneración del código genético. Se incluyen dentro de esta definición polimorfismos que pueden ser o no fácilmente detectables usando una sonda oligonucleotídica particular del polinucleótido que codifica el polipéptido, e hibridación inapropiada o inesperada con variantes alélicas, con un locus distinto del locus cromosómico normal para la secuencia polinucleotídica que codifica el polipéptido.

“Variante alélica” o “variante alélica polinucleotídica” se refiere a cualquiera de dos o más formas alternativas de un gen que ocupan el mismo locus cromosómico. La variación alélica surge de forma natural mediante mutación, y puede dar como resultado polimorfismo fenotípico dentro de las poblaciones. Las mutaciones génicas pueden ser “silenciosas” o pueden codificar polipéptidos que tienen la secuencia de aminoácidos alterada. “Variante alélica” y “variante alélica polipeptídica” también pueden usarse con respecto a polipéptidos, y en este caso el término se refiere a un polipéptido codificado por una variante alélica de un gen.

“Variante de corte y empalme” o “variante de corte y empalme polinucleotídica” como se usa en el presente documento se refieren a formas alternativas de ARN transcrito de un gen. La variación de corte y empalme se produce de forma natural como resultado de sitios alternativos que se cortan y empalman dentro de una única molécula de ARN transcrita o entre moléculas de ARN transcritas por separado, y puede dar como resultado varias formas diferentes de ARNm transcritas del mismo gen. Por lo tanto, las variantes de corte y empalme pueden codificar polipéptidos que tienen diferentes secuencias de aminoácidos, que pueden tener o no funciones similares en el organismo. La “variante de corte y empalme” o “variante de corte y empalme polipeptídica” también puede referirse a un polipéptido codificado por una variante de corte y empalme de un ARNm transcrito.

Como se usa en el presente documento “variantes polinucleotídicas” también puede referirse a secuencias polinucleotídicas que codifican parálogos y ortólogos de las secuencias polipeptídicas desveladas en el presente documento. “Variantes polipeptídicas” puede referirse a secuencias polipeptídicas que son parálogos y ortólogos de las secuencias polipeptídicas desveladas en el presente documento.

Las diferencias entre polipéptidos desvelados en el presente documento y variantes polipeptídicas están limitadas de modo que las secuencias de las primeras y las últimas son muy similares en general y, en muchas regiones, idénticas. Las secuencias polipeptídicas desveladas en el presente documento y variantes polipeptídicas similares pueden diferir en su secuencia de aminoácidos por una o más sustituciones, adicionales, delecciones, fusiones y

truncamientos, que pueden estar presentes en cualquier combinación. Estas diferencias pueden producir cambios silenciosos y dar como resultado un factor de transcripción funcionalmente equivalente. Por lo tanto, se apreciará fácilmente por los expertos en la materia, que cualquiera de diversas secuencias polinucleotídicas es capaz de codificar los factores de transcripción y polipéptidos homólogos de factores de transcripción descritos en el presente documento. Una variante de secuencia polipeptídica puede tener cambios “conservativos”, en los que un aminoácido sustituido tiene propiedades estructurales o químicas similares. Pueden realizarse por lo tanto sustituciones de aminoácidos deliberadas basándose en la similitud en polaridad, carga, solubilidad, hidrofobicidad, hidrofilia y/o la naturaleza anfipática de los restos, siempre que se conserve una cantidad significativa de la actividad funcional o biológica del factor de transcripción. Por ejemplo, los aminoácidos con carga negativa pueden incluir ácido aspártico y ácido glutámico, los aminoácidos con carga positiva pueden incluir lisina y arginina, y los aminoácidos con grupos de cabeza polar no cargada que tienen valores de hidrofilia similares pueden incluir leucina, isoleucina y valina; glicina y alanina; asparagina y glutamina; serina y treonina; y fenilalanina y tirosina (para más detalles sobre sustituciones conservativas, véase Tabla 3). De forma menos común, una variante puede tener cambios “no conservativos”, por ejemplo, reemplazo de una glicina con un triptófano. Las variaciones menores similares también pueden incluir delecciones o inserciones de aminoácidos, o ambas. Los polipéptidos relacionados pueden comprender, por ejemplo, adiciones y/o delecciones de uno o más sitios de glucosilación ligados a N o ligados a O, o una adición y/o una delección de uno o más restos de cisteína. Pueden encontrarse instrucciones para determinar qué y cuantos restos de aminoácidos pueden sustituirse, insertarse o suprimirse sin anular la actividad funcional o biológica usando programas informáticos bien conocidos en la técnica, por ejemplo, software DNASTAR (véase documento USPN 5.840.544).

“Fragmento”, con respecto a un polinucleótido, se refiere a un clon o cualquier parte de una molécula polinucleotídica que conserva una característica utilizable, funcional. Los fragmentos útiles incluyen oligonucleótidos y polinucleótidos que pueden usarse en tecnologías de hibridación o amplificación o en la regulación de la replicación, transcripción o traducción. Un “fragmento polinucleotídico” se refiere a cualquier subsecuencia de un polinucleótido, normalmente de al menos aproximadamente 9 nucleótidos consecutivos, preferentemente al menos aproximadamente 30 nucleótidos, más preferentemente al menos aproximadamente 50 nucleótidos de cualquiera de las secuencias proporcionadas en el presente documento. Son fragmentos polinucleotídicos ejemplares los primeros sesenta nucleótidos consecutivos de los polinucleótidos del factor de transcripción enumerados en el Listado de Secuencias. Los fragmentos ejemplares también incluyen fragmentos que comprenden una región que codifica un dominio de gancho AT de un factor de transcripción. Los fragmentos ejemplares también incluyen fragmentos que comprenden un dominio conservado de un factor de transcripción. Los fragmentos ejemplares también incluyen fragmentos que comprenden un gancho AT o segundo dominio conservado de un factor de transcripción de gancho AT, por ejemplo, los restos de aminoácidos 34-42 y 78-175 de G1073 (SEC ID N°: 2), como se indica en la Tabla 1.

Los fragmentos también pueden incluir subsecuencias de polipéptidos y moléculas proteicas, o una subsecuencia del polipéptido. Los fragmentos pueden tener usos porque pueden tener potencial antigénico. En algunos casos, el fragmento o dominio es una subsecuencia del polipéptido que realiza al menos una función biológica del polipéptido intacto sustancialmente de la misma manera, o en un grado similar, que lo hace el polipéptido intacto. Por ejemplo, un fragmento polipeptídico puede comprender un motivo estructural reconocible o dominio funcional tal como un sitio de unión a ADN o dominio que se une con una región promotora de ADN, un dominio de activación, o un dominio para interacciones proteína-proteína, y puede iniciar la transcripción. Los fragmentos pueden variar de tamaño desde tan poco como 3 restos de aminoácidos hasta la longitud completa del polipéptido intacto, pero son preferentemente de al menos aproximadamente 30 restos de aminoácidos de longitud y más preferentemente de al menos aproximadamente 60 restos de aminoácidos de longitud.

Las secuencias de ADN que codifican factores de transcripción y derivados de factores de transcripción, o fragmentos de los mismos, pueden producirse completamente por química sintética. Después de la producción, la secuencia sintética puede insertarse en cualquiera de los muchos vectores de expresión disponibles y sistemas celulares usando reactivos bien conocidos en la técnica. Además, puede usarse química sintética para introducir mutaciones en una secuencia codificante de factores de transcripción o cualquier fragmento de la misma.

“Derivado” se refiere a la modificación química de una molécula de ácido nucleico o secuencia de aminoácidos. Las modificaciones químicas pueden incluir reemplazo de hidrógeno por un grupo alquilo, acilo o amino o glucosilación, pegilación o cualquier proceso similar que conserve o potencie la actividad biológica o vida útil de la molécula o secuencia.

El término “planta” incluye plantas completas, órganos/estructuras vegetativos de brotes (por ejemplo, hojas, tallos y tubérculos), raíces, flores y órganos/estructuras florales (por ejemplo, brácteas, sépalos, pétalos, estambres, carpelos, anteras y óvulos), semillas (incluyendo embrión, endospermo y tegumento) y fruto (el ovario maduro), tejido vegetal (por ejemplo, tejido vascular, tejido terrestre y similares) y células (por ejemplo, células de guarda, células del óvulo y similares) y descendencia de los mismos. La clase de plantas que pueden usarse en el método de la invención generalmente es tan amplio como la clase de plantas superiores e inferiores susceptibles a técnicas de transformación, incluyendo angiospermas (plantas monocotiledóneas y dicotiledóneas), gimnospermas, helechos, equisetos, psilofitas, licofitas, briofitas y algas multicelulares (véase por ejemplo, Figura 1, adaptada de Daly *et al.* (2001) *Plant Physiol.* 127: 1328-1333; Figura 2, adaptada de Ku *et al.* (2000) *Proc. Natl. Acad. Sci. USA* 97: 9121-

9126; y véase también Tudge en *The Variety of Life*, Oxford University Press, Nueva York, NY (2000) pp. 547-606).

Una “planta transgénica” se refiere a una planta que contiene material genético no hallado en una planta de tipo silvestre de la misma especie, variedad o cultivar. El material genético puede incluir un transgén, un acontecimiento de mutagénesis de inserción (tal como por transposón o mutagénesis de inserción de ADN-T), una secuencia de marcaje de activación, una secuencia mutada, un acontecimiento de recombinación homóloga o secuencia modificada por queraplastia. Normalmente, el material genético ajeno se ha introducido en la planta por manipulación humana, pero puede usarse cualquier método como reconocerá un experto en la materia.

Una planta transgénica puede contener un vector o casete de expresión. El casete de expresión normalmente comprende una secuencia codificante de polipéptido unida operativamente con (es decir, bajo el control regulador de) secuencias reguladoras inducibles o constitutivas apropiadas que permiten la expresión controlada del polipéptido. El casete de expresión puede introducirse en una planta por transformación o por cultivo después de transformación de una planta parental. Una planta se refiere a una planta completa así como a una parte de planta, tal como una semilla, fruto, hoja, o raíz, tejido vegetal, células vegetales o cualquier otro material vegetal, por ejemplo, un explante vegetal, así como a descendencia de los mismos, y a sistemas *in vitro* que imitan los componentes bioquímicos o celulares o procesos en una célula.

“Tipo silvestre”, como se usa en el presente documento, se refiere a una célula vegetal, semilla, componente vegetal, tejido vegetal, órgano vegetal o planta completa que no se ha modificado genéticamente o no se ha tratado en un sentido experimental. Pueden usarse células, semillas, componentes, tejido, órganos o plantas completas de tipo silvestre como controles para comparar los niveles de expresión y el alcance y naturaleza de la modificación de rasgos con células, tejido o plantas de la misma especie en los que se ha alterado la expresión de un factor de transcripción, por ejemplo, porque se ha anulado, sobreexpresado o expresado de forma ectópica.

Una “planta de control” como se usa en la presente invención se refiere a una célula vegetal, semilla, componente vegetal, tejido vegetal, órgano vegetal o planta completa usado para comparar frente a una planta transgénica o modificada genéticamente para el fin de identificar un fenotipo potenciado en la planta transgénica o modificada genéticamente. Una planta de control puede en algunos casos ser una línea vegetal transgénica que comprende un vector vacío o gen marcador, pero no contiene el polinucleótido recombinante que se expresa en la planta transgénica o modificada genéticamente que se evalúa. En general, una planta de control es una planta de la misma línea o variedad que la planta transgénica o modificada genéticamente que se ensaya. Una planta de control adecuada incluiría una planta genéticamente no alterada o no transgénica de la línea parental usada para generar una planta transgénica del presente documento.

Un “rasgo” se refiere a una característica fisiológica, morfológica, bioquímica o física de una planta o un material o célula vegetal particular. En algunos casos, esta característica es visible al ojo humano, tal como el tamaño de la semilla o la planta, o puede medirse por técnicas bioquímicas, tales como detección del contenido de proteínas, almidón o aceite de la semilla o las hojas, o mediante observación de un proceso metabólico o fisiológico, por ejemplo midiendo la tolerancia a privación de agua o concentraciones de sales o azúcares particulares, o mediante la observación del nivel de expresión de un gen o genes, por ejemplo, empleando análisis de Northern, RT-PCR, ensayos de expresión génica de micromatrices o sistemas de expresión de genes indicadores, o mediante observaciones agrícolas tales como tolerancia al estrés osmótico o producción. Sin embargo, puede usarse cualquier técnica para medir la cantidad de, nivel comparativo de, o diferencia en cualquier compuesto químico seleccionado o macromolécula en las plantas transgénicas.

La “modificación de rasgos” se refiere a una diferencia detectable en una característica en una planta que expresa de forma ectópica un polinucleótido o polipéptido en relación con una planta que no lo hace, tal como una planta de tipo silvestre. En algunos casos, la modificación de rasgos puede evaluarse de forma cuantitativa. Por ejemplo, la modificación de rasgos puede suponer al menos aproximadamente un 2 % de aumento o reducción, o una diferencia aún mayor, en un rasgo observado en comparación con una planta de control o de tipo silvestre. Se sabe que puede haber una variación natural en el rasgo modificado. Por lo tanto, la modificación del rasgo observado implica un cambio de la distribución y magnitud normal del rasgo en las plantas en comparación con plantas de control o de tipo silvestre.

Cuando dos o más plantas tienen “morfologías similares”, “morfologías sustancialmente similares”, “una morfología que es sustancialmente similar” o son “morfológicamente similares” las plantas tienen formas o apariencias comparables, incluyendo características análogas tales como dimensiones generales, altura, anchura, masa, masa de la raíz, forma, brillo, color, diámetro del tallo, tamaño de la hoja, dimensión de la hoja, densidad de la hoja, distancia internodal, ramificación, ramificación de la raíz, número y forma de inflorescencias, y otras características macroscópicas, y las plantas individuales no son fácilmente distinguibles basándose en sus características morfológicas solamente.

“Modula” se refiere a un cambio en la actividad (biológica, química o inmunológica) o vida útil resultante de la unión específica entre una molécula y una molécula de ácido nucleico o una proteína.

La expresión “perfil de transcrito” se refiere a los niveles de expresión de un conjunto de genes en una célula en un estado particular, particularmente por comparación con los niveles de expresión de ese mismo conjunto de genes en una célula del mismo tipo en un estado de referencia. Por ejemplo, el perfil de transcrito de un factor de transcripción particular en una célula en suspensión es los niveles de expresión de un conjunto de genes en una célula anulando o sobreexpresando ese factor de transcripción en comparación con los niveles de expresión de ese mismo conjunto de genes en una célula en suspensión que tiene niveles normales de ese factor de transcripción. El perfil de transcrito puede presentarse como una lista de los genes cuyo nivel de expresión es significativamente diferente entre los dos tratamientos, y las relaciones de diferencia. Las diferencias y similitudes entre los niveles de expresión también pueden evaluarse y calcularse usando métodos estadísticos y de agrupamiento.

“Expresión ectópica o expresión alterada” en referencia a un polinucleótido indica que el patrón de expresión en, por ejemplo, una planta transgénica o tejido vegetal, es diferente del patrón de expresión en una planta de tipo silvestre o una planta de referencia de la misma especie. El patrón de expresión también puede compararse con un patrón de expresión de referencia en una planta de tipo silvestre de la misma especie. Por ejemplo, el polinucleótido o polipéptido se expresa en un tipo celular o tisular distinto de un tipo celular o tisular en el que la secuencia se expresa en la planta de tipo silvestre, o por expresión en un momento distinto al momento en que la secuencia se expresa en la planta de tipo silvestre, o por una respuesta a diferentes agentes inducibles, tales como hormonas o señales ambientales, o a niveles de expresión diferentes (bien superiores o bien inferiores) en comparación con los hallados en una planta de tipo silvestre. La expresión también se refiere a patrones de expresión alterados que se producen reduciendo los niveles de expresión por debajo del nivel de detección o anulando completamente la expresión. El patrón de expresión resultante puede ser transitorio o estable, constitutivo o inducible. En referencia a un polipéptido, la expresión “expresión ectópica o expresión alterada” puede relacionarse además con niveles de actividad alterados resultantes de las interacciones de los polipéptidos con moduladores exógenos o endógenos o de interacciones con factores o como resultado de la modificación química de los polipéptidos.

El término “sobreexpresión” como se usa en el presente documento se refiere a un mayor nivel de expresión de un gen en una planta, célula vegetal o tejido vegetal, en comparación con la expresión en una planta, célula o tejido de tipo silvestre, en cualquier estadio del desarrollo o temporal para el gen. Puede producirse sobreexpresión cuando, por ejemplo, los genes que codifican uno o más factores de transcripción están bajo el control de un promotor fuerte (por ejemplo, la región de inicio de la transcripción de 35S del virus del mosaico de la coliflor). La sobreexpresión también puede estar bajo el control de un promotor inducible o específico de tejido. Por lo tanto, la sobreexpresión puede producirse en toda una planta, en tejidos específicos de la planta, o en presencia o ausencia de señales ambientales particulares, dependiendo del promotor usado.

La sobreexpresión puede tener lugar en células vegetales que normalmente carecen de expresión de polipéptidos funcionalmente equivalentes o idénticos a los presentes factores de transcripción. La sobreexpresión también puede producirse en células vegetales en las que la expresión endógena de los presentes factores de transcripción o moléculas funcionalmente equivalentes normalmente se produce, pero dicha expresión normal es a un nivel menor. La sobreexpresión da como resultado por lo tanto una producción mayor de lo normal, o “sobreproducción” del factor de transcripción en la planta, célula o tejido.

La expresión “región reguladora de la transcripción” se refiere a una secuencia reguladora de ADN que regula la expresión de uno o más genes en una planta cuando un factor de transcripción que tiene uno o más dominios de unión específicos se une con la secuencia reguladora de ADN. Los factores de transcripción relacionados con la presente invención poseen un dominio de gancho AT y un segundo dominio conservado. Pueden encontrarse ejemplos de secuencias de dominio de gancho AT y segundo dominio conservado similares en la Tabla 1. Los factores de transcripción también comprenden una subsecuencia de aminoácidos que forma un dominio de activación de la transcripción que regula la expresión de uno o más genes de tolerancia al estrés abiótico en una planta cuando el factor de transcripción se une con la región reguladora.

Descripción de las realizaciones específicas

Los factores de transcripción modifican la expresión de genes endógenos

Un factor de transcripción puede incluir, pero sin limitación, cualquier polipéptido que pueda activar o reprimir la transcripción de un único gen o varios genes. Como reconocerá un experto habitual en la materia, los factores de transcripción pueden identificarse por la presencia de una región o dominio de similitud e identidad estructural con una secuencia consenso específica o la presencia de un sitio de unión a ADN consenso específico o motivo de sitio de unión a ADN (véase, por ejemplo, Riechmann *et al.* (2000) mencionado anteriormente). Los factores de transcripción de plantas de la presente invención pertenecen a la familia del factor de transcripción de gancho AT (Reeves y Beckerbauer (2001) mencionado anteriormente; y Reeves (2001) mencionado anteriormente).

En general, los factores de transcripción codificados por las presentes secuencias están implicados en diferenciación y proliferación celular y la regulación del crecimiento. En consecuencia, un experto en la materia reconocería que expresando las presentes secuencias en una planta, se puede cambiar la expresión de genes autólogos o inducir la expresión de genes introducidos. Afectando a la expresión de secuencias autólogas similares en una planta que

tiene la actividad biológica de las presentes secuencias, o introduciendo las presentes secuencias en una planta, se puede alterar el fenotipo de una planta a uno con rasgos mejorados relacionados con tensiones osmóticas. Las secuencias descritas en el presente documento también pueden usarse para transformar una planta e introducir rasgos deseables no hallados en el cultivar o cepa de tipo silvestre. Las plantas pueden seleccionarse después con respecto a las que producen el grado más deseable de sobre o infraexpresión de genes diana de interés y mejora de rasgos coincidente.

Las secuencias descritas en el presente documento pueden ser de cualquier especie, particularmente especies vegetales, en una forma de origen natural o de cualquier fuente bien natural, sintética, semisintética o recombinante. Las secuencias también pueden incluir fragmentos de las presentes secuencias de aminoácidos. Cuando se indica "secuencia de aminoácidos" para referirse a una secuencia de aminoácidos de una molécula proteica de origen natural, no se pretende que "secuencia de aminoácidos" y expresiones similares limiten la secuencia de aminoácidos a la secuencia de aminoácidos nativa completa asociada con la molécula proteica indicada.

Además de métodos para modificar un fenotipo vegetal empleando uno o más polinucleótidos y polipéptidos descritos en el presente documento, los polinucleótidos y polipéptidos tienen diversos usos adicionales. Estos usos incluyen su uso en la producción recombinante (es decir, expresión) de proteínas; como reguladores de la expresión génica de plantas, como sondas de diagnóstico para la presencia de ácidos nucleicos complementarios o parcialmente complementarios (incluyendo para detección de ácidos nucleicos codificantes naturales); como sustratos para reacciones adicionales, por ejemplo, reacciones de mutación, reacciones de PCR o similares; como sustratos para clonación por ejemplo, incluyendo reacciones de digestión o ligamiento; y para identificar moduladores exógenos o endógenos de los factores de transcripción. El polinucleótido puede ser, por ejemplo, ADN o ARN genómico, un transcrito (tal como un ARNm), un ADNc, un producto de PCR, un ADN clonado, un ADN o ARN sintético, o similares. El polinucleótido puede comprender una secuencia de orientaciones con sentido o antisentido.

La expresión de genes que codifican factores de transcripción que modifican la expresión de genes endógenos, polinucleótidos y proteínas se conocen bien en la técnica. Además, las plantas transgénicas que comprenden polinucleótidos aislados que codifican factores de transcripción también pueden modificar la expresión de genes endógenos, polinucleótidos y proteínas. Los ejemplos incluyen Peng *et al.* (1997) *Genes Development* 11: 3194-3205 y Peng *et al.* (1999) *Nature*, 400: 256-261). Además, muchos otros han demostrado que un factor de transcripción de *Arabidopsis* expresado en una especie vegetal exógena induce la misma o muy similar respuesta fenotípica. Véase, por ejemplo, Fu *et al.* (2001) *Plant Cell* 13: 1791-1802; Nandi *et al.* (2000) *Curr. Biol.* 10: 215-218; Coupland (1995) *Nature* 377: 482-483; y Weigel y Nilsson (1995) *Nature* 377: 482-500).

En otro ejemplo, Mandel *et al.* (1992) *Cell* 71:133-143, y Suzuki *et al.* (2001) *Plant J.* 28: 409-418, enseñan que un factor de transcripción expresado en otra especie vegetal induce la misma o una respuesta fenotípica muy similar de la secuencia endógena, como se ha predicho con frecuencia en estudios anteriores de factores de transcripción de *Arabidopsis* en *Arabidopsis* (véase Mandel *et al.* (1992) mencionado anteriormente; Suzuki *et al.* (2001) mencionado anteriormente). Otros ejemplos incluyen Müller *et al.* (2001) *Plant J.* 28: 169-179; Kim *et al.* (2001) *Plant J.* 25: 247-259; Kozuka y Shimamoto (2002) *Plant Cell Physiol.* 43: 130-135; Boss y Thomas (2002) *Nature*, 416: 847-850; He *et al.* (2000) *Transgenic Res.* 9: 223-227; y Robson *et al.* (2001) *Plant J.* 28: 619-631.

En otro ejemplo más, Gilmour *et al.* (1998) *Plant J.* 16: 433-442) enseñan un factor de transcripción de AP2 de *Arabidopsis*, CBF1, que, cuando se sobreexpresa en plantas transgénicas, aumenta la tolerancia a congelación de la planta. Jaglo *et al.* (2001) *Plant Physiol.* 127: 910-917, identificaron adicionalmente secuencias en *Brassica napus* que codifican genes de tipo CBF y que transcritos de estos genes se acumulaban rápidamente en respuesta a temperatura baja. También se descubrió que los transcritos que codificaban proteínas de tipo CBF se acumulaban rápidamente en respuesta a temperatura baja en trigo, así como en tomate. Un alineamiento de las proteínas de CBF de *Arabidopsis*, *B. napus*, trigo, centeno y tomate reveló la presencia de restos de aminoácidos consecutivos conservados, PKK/RPAGRxKFxETRHP y DSAWR, que flanquean los dominios de unión a ADN de AP2/EREBP de las proteínas y los distinguen de otros miembros de la familia de proteínas AP2/EREBP. (Jaglo *et al.* (2001) mencionado anteriormente).

Los factores de transcripción median en respuestas celulares y rasgos de control mediante expresión alterada de genes que contienen secuencias de nucleótidos de acción en cis que son dianas del factor de transcripción introducido. Se aprecia bien en la técnica que el efecto de un factor de transcripción en respuestas celulares o un rasgo celular se determina por los genes particulares cuya expresión está alterada directa o indirectamente (por ejemplo, por una cascada de acontecimientos de unión de factores de transcripción y cambios transcripcionales) por la unión de factores de transcripción. En un análisis global de la transcripción que compara una condición convencional con una en la que está sobreexpresado un factor de transcripción, el perfil del transcrito resultante asociado con la sobreexpresión del factor de transcripción está relacionado con el rasgo o proceso celular controlado por ese factor de transcripción. Por ejemplo, se ha mostrado que el gen PAP2 (y otros genes en la familia de MYB) controla la biosíntesis de antocianina mediante la regulación de la expresión de genes que se sabe que están implicados en la ruta biosintética de la antocianina (Bruce *et al.* (2000) *Plant Cell* 12: 65-79; y Borevitz *et al.* (2000) *Plant Cell* 12: 2383-2393). Además, se han usado perfiles de transcrito globales con éxito como herramientas de

diagnóstico para estados celulares específicos (por ejemplo, cancerosos frente a no cancerosos; Bhattacharjee *et al.* (2001) Proc. Natl. Acad. Sci. USA 98: 13790-13795; y Xu *et al.* (2001) Proc. Natl. Acad. Sci. USA 98: 15089-15094). En consecuencia, es evidente para un experto en la materia que la similitud del perfil de transcrito tras la sobreexpresión de diferentes factores de transcripción indicaría similitud de la función del factor de transcripción.

Polipéptidos y polinucleótidos

La presente divulgación se refiere, entre otras cosas, a factores de transcripción (TF), y polipéptidos homólogos de factores de transcripción, y polinucleótidos aislados o recombinantes que codifican los polipéptidos, o polipéptidos variantes de nueva secuencia o polinucleótidos que codifican nuevas variantes de factores de transcripción derivados de las secuencias específicas proporcionadas en el Listado de Secuencias. La presente divulgación proporciona métodos para modificar la biomasa de una planta modificando el tamaño o el número de hojas o semillas de una planta controlando varios procesos celulares, y para aumentar la tolerancia de una planta a tensiones abióticas. Estos métodos se basan en la capacidad para alterar la expresión de moléculas reguladoras críticas que pueden estar conservadas entre diversas especies vegetales. Pueden descubrirse originalmente moléculas reguladoras conservadas relacionadas en un sistema modelo tal como *Arabidopsis* y descubrirse después moléculas homólogas, funcionales, en otra especie vegetal. Esta última puede usarse después para conferir aumento de la biomasa o tolerancia al estrés abiótico en diversas especies vegetales.

Se identificaron polinucleótidos ejemplares relacionados con la presente invención en la base de datos de GenBank de *Arabidopsis thaliana* usando programas y parámetros de análisis de secuencia disponibles públicamente. Después se caracterizaron adicionalmente secuencias inicialmente definidas para identificar secuencias que comprendían cadenas de secuencias específicas correspondientes a motivos de secuencia presentes en familias de factores de transcripción conocidos. Además, se identificaron polinucleótidos ejemplares adicionales en la base de datos de GenBank de plantas usando programas y parámetros de análisis de secuencia disponibles públicamente. Después se caracterizaron adicionalmente secuencias inicialmente definidas para identificar secuencias que comprendían cadenas de secuencias específicas correspondientes a motivos de secuencia presentes en familias de factores de transcripción conocidos. Las secuencias polinucleótídicas que cumplían dichos criterios se confirmaron como factores de transcripción.

Se identificaron polinucleótidos adicionales explorando *Arabidopsis thaliana* y/u otras bibliotecas de ADNc vegetal con sondas correspondientes a factores de transcripción conocidos en condiciones de hibridación de baja rigurosidad. Se recuperaron posteriormente secuencias adicionales, incluyendo secuencias codificantes de longitud completa, por el procedimiento de amplificación rápida de extremos de ADNc (RACE) usando un kit disponible en el mercado de acuerdo con las instrucciones del fabricante. Cuando sea necesario, se realizan múltiples ciclos de RACE para aislar extremos 5' y 3'. Después se recuperó el ADNc de longitud completa por una reacción en cadena de la polimerasa (PCR) de extremo a extremo rutinaria usando cebadores específicos para los extremos 5' y 3' aislados. Se proporcionan secuencias ejemplares en el Listado de Secuencias.

Las secuencias polipeptídicas y polinucleótídicas de G1067 se identificaron previamente en la Solicitud de Patente Provisional de Estados Unidos 60/135.134, presentada el 20 de mayo de 1999. Las secuencias polipeptídicas y polinucleótídicas de G1073 se identificaron previamente en la Solicitud de Patente Provisional de Estados Unidos 60/125.814, presentada el 23 de marzo de 1999. La función de G1073 en el aumento de la biomasa se desveló en la Solicitud Provisional de Estados Unidos N° 60/227.439, presentada el 22 de agosto de 2000 y la utilidad para aumento de la tolerancia a sequía observado en líneas transgénicas 35S::G1073 se desveló en la Solicitud no Provisional de Estados Unidos N° 10/374.780, presentada el 25 de febrero de 2003. Las secuencias polipeptídicas y polinucleótídicas de G2153 y G2156 se identificaron previamente en la Solicitud Provisional de Patente de Estados Unidos N° 60/338.692, presentada el 11 de diciembre de 2001 y en las Solicitudes de Patente no Provisionales de Estados Unidos 10/225.066 y 10/225.068, ambas de las cuales se presentaron el 9 de agosto de 2002. El fenotipo de detección de azúcar alterada y de tolerancia al estrés osmótico conferido por la sobreexpresión de G2153 se desveló en estas presentaciones. En el momento en que se presentaron cada una de las solicitudes anteriores, se identificó que estas secuencias eran codificantes o eran factores de transcripción, que se definieron como polipéptidos que tienen la capacidad para efectuar transcripción de un gen diana. Se ha determinado que las secuencias que tienen actividad reguladora génica tienen utilidad específica y sustancial por la Oficina de Patentes y Marcas de Estados Unidos (Registro Federal (2001) 66(4): 1095).

Estas secuencias y otras derivadas de especies diversas y halladas en el Listado de Secuencias se han expresado de forma ectópica en plantas con sobreexpresión. Los cambios en la característica o las características o el rasgo o los rasgos de las plantas se observaron después y se descubrió que conferían aumento de la biomasa o tolerancia al estrés abiótico. Por lo tanto, los polinucleótidos y polipéptidos pueden usarse para mejorar características deseables de las plantas.

Los polinucleótidos relacionados con la presente invención también se expresaron de forma ectópica en células vegetales con sobreexpresión y los cambios en los niveles de expresión de varios genes, polinucleótidos y/o proteínas de las células vegetales observadas. Por lo tanto, los polinucleótidos y polipéptidos pueden usarse para cambiar los niveles de expresión de genes, polinucleótidos y/o proteínas de plantas o células vegetales.

La familia del factor de transcripción de gancho AT

En organismos superiores, se ensambla ADN genómico en complejos de múltiples niveles con una serie de proteínas de unión a ADN, incluyendo las bien conocidas histonas y proteínas no histonas tales como las proteínas del grupo de alta movilidad (HMG). Las proteínas HMGA se clasifican en diferentes grupos basándose en sus motivos de unión a ADN, y uno de dichos grupos es el subgrupo HMG-I(Y) (recientemente renombrado como HMGA). Se ha mostrado que las proteínas en este grupo se unen con el surco menor del ADN mediante un péptido de nueve aminoácidos conservado (KRPRGRPCK) llamado el motivo de gancho AT (Reeves y Nisson (1995) mencionado anteriormente). En el centro de este motivo de gancho AT hay un tripéptido corto, altamente conservado de glicina-arginina-prolina (GRP). Este sencillo motivo de gancho AT puede estar presente en un número variable de copias (1-15) en una proteína de gancho AT dada. Por ejemplo, la proteína HMGA1 de mamífero tiene tres copias de este motivo. Las proteínas HMGA de mamífero participan en una amplia diversidad de procesos nucleares que varían de remodelación del cromosoma y la cromatina, a actuar como factores de transcripción arquitectónicos que regulan la expresión de numerosos genes *in vivo*. Como resultado, estas proteínas influyen en una serie diversa de procesos celulares incluyendo crecimiento, proliferación, diferenciación y muerte a través de las interacciones proteína-ADN y proteína-proteína (para revisiones véase Reeves y Beckerbauer (2001) mencionado anteriormente; y Reeves (2001) mencionado anteriormente). Se ha mostrado que las proteínas HMGA interaccionan específicamente con un gran número de otras proteínas, la mayoría de las cuales son factores de transcripción (Reeves (2001) mencionado anteriormente). También se someten a muchos tipos de modificación postraduccional. Un ejemplo es la fosforilación, que influye notablemente en su capacidad para interaccionar con sustratos de ADN, otras proteínas y cromatina (Onate *et al.* (1994) Mol. Cell Biol. 14: 3376-3391; Falvo *et al.* (1995) Cell 83: 1101-1111; Reeves y Nissen (1995) mencionado anteriormente; Huth *et al.* (1997) Nat. Struct. Biol. 4, 657-665; y Girard *et al.* (1998) EMBO J. 17: 2079-2085).

En plantas, se identificó una proteína con motivos de unión a ADN de gancho AT en avena (Nieto-Sotelo y Quail (1994) Biochem. Soc. Symp. 60, 265-275). Esta proteína se une con la región PE1 del promotor del gen del fitocromo A3 de avena y puede estar implicada en la regulación positiva de la expresión del gen *PHYA3* (Nieto-Sotelo y Quail (1994) mencionado anteriormente). También se han identificado proteínas de unión a ADN que contienen dominios de gancho AT en diversas especies vegetales, incluyendo arroz, guisantes y *Arabidopsis* (Meijer *et al.* (1996) Plant Mol. Biol. 31: 607-618; y Gupta *et al.* (1997a) Plant Mol. Biol. 35: 987-992). Los genes de gancho AT del arroz se expresan predominantemente en tejidos jóvenes y meristemáticos, lo que sugiere que las proteínas de gancho AT pueden afectar a la expresión de genes que determinan el estado de diferenciación de las células. El gen de gancho AT del guisante se expresa en todos los órganos incluyendo raíces, tallos, hojas, flores, zarcillos y semillas en desarrollo (Gupta *et al.* (1997a) mencionado anteriormente). El análisis de transferencia de Northern reveló que se expresaba un gen de gancho AT de *Arabidopsis* en todos los órganos con la mayor expresión en flores y silicuas en desarrollo (Gupta *et al.* (1997b) Plant Mol. Biol. 34: 529-536).

Recientemente, también se ha mostrado que la expresión de una proteína de gancho AT del maíz en células de levadura produce mejor crecimiento en un medio que contiene altas concentraciones de níquel.

Nuevos genes de factores de transcripción de gancho AT y motivos de unión en *Arabidopsis* y otras especies diversas

Los inventores han identificado al menos treinta y cuatro genes de *Arabidopsis* que codifican proteínas con motivos de unión a ADN de gancho AT. De estos, hay veintidós genes que codifican un único motivo de unión a ADN de gancho AT; ocho genes que codifican dos motivos de unión a ADN de gancho AT; tres genes (G280, G1367 y G2787, SEC ID N°: 55, 57 y 59, respectivamente) que codifican cuatro motivos de unión a ADN de gancho AT y un único gen (G3045, SEC ID N°: 61) que codifica tres motivos de unión a ADN de gancho AT.

G1073 (SEC ID N°: 2), por ejemplo, contiene un único motivo de unión a ADN de gancho AT típico (RRPRGRPAG) correspondiente a las posiciones 34 a 42 dentro de la proteína. Puede identificarse un dominio de 129 restos de aminoácidos altamente conservado con función desconocida (denominado en lo sucesivo el “segundo dominio conservado”) en el único subgrupo de dominio de gancho AT, el “clado G1073 de polipéptidos de factores de transcripción”, o más sencillamente el “clado G1073”. Después de esta región, un dominio ácido potencial abarca de la posición 172 a 190. Adicionalmente, el análisis de la proteína usando PROSITE revela tres sitios de fosforilación C de proteína quinasa potenciales en Ser32, Thr83 y Thr102, y tres sitios de fosforilación de caseína quinasa II potenciales en Ser6, Ser70 y Ser247 (Figura 3). En comparación con muchas otras proteínas de gancho AT, la proteína G1073 contiene un extremo N terminal más corto (Figuras 5A-5C).

Los miembros del clado G1073 son estructuralmente distintos de otras proteínas relacionadas con el gancho AT, como puede verse en las Figuras 5C-5F, en comparación con G1073 y secuencias anteriores que están comprendidas dentro del clado G1073 (dentro de la caja grande) y las secuencias que incluyen G1945, G2155 y G3408, que representan secuencias de gancho AT que quedan fuera del clado (debajo de la caja grande).

La Tabla 1 muestra los polipéptidos identificados por: polipéptido SEC ID N° (primera columna); Gen ID o “GID” N° (segunda columna); las coordenadas del resto de aminoácido para el dominio de gancho AT y segundo dominio

conservado (tercera columna); secuencias de gancho AT de los polipéptidos respectivos (cuarta columna); la identidad en términos de porcentaje con el dominio de gancho AT de G1073 (quinta columna); secuencias del segundo dominio conservado de los polipéptidos respectivos (sexta columna); y la identidad en términos de porcentaje con el segundo dominio conservado de G1073 (séptima columna). Se ha mostrado que muchas de estas secuencias confieren fenotipos tolerantes al estrés abiótico cuando se sobreexpresan en plantas, como se indica en la penúltima columna de la Tabla 1. La última columna indica las secuencias que se ha observado que aumentan la biomasa vegetal en líneas con sobreexpresión en relación con controles de tipo silvestre. Las secuencias polipeptídicas que muestran una capacidad significativa de conferir tolerancia al estrés abiótico y aumento de la biomasa incluyen dominios de gancho At y segundo conservado con 78 % y 62 % o mayor identidad con los dominios de gancho At y segundo conservado de G1073, respectivamente.

Tabla 1. Familias génicas y dominios de unión

SEC ID N°:	GID N°	Dominios de gancho AT y Segundo Conservado en Coordenadas de AA y Bases	Dominio de gancho AT	% de ID con el Dominio de gancho AT de G1073	Segundo Dominio Conservado	% de ID con el Segundo Dominio Conservado de G1073	Tolerante al estrés abiótico	Mayor Biomasa
2	G1073	Coordenadas del polipéptido: 34-42; 78-175	RRPRGRPPAG	100 %	VSTYATRRGC GVCISGTGAV TNVTIRQPAAP AGGGVITLHG RFDILSLTGTA LPPAPPAGAG LTVLAGGQG QVGGNVAGS LIASGPVVL MAAF	100 %	Sí	Sí
26	G3406	Coordenadas del polipéptido: 82-90, 126-222	RRPRGRPPG	89 %	VSTYARRRQR GVCVLSGSGV VTNVTLRQPS APAGAVVSLH GRFEILSLSGSF LPPAPPAGATS LTIFLAGGQGG VGGNVVVGAL YAAGPVVIAA SF	71 %	Sí	No
10	G3399	Coordenadas del polipéptido: 99-107, 143-240	RRPRGRPPG	89 %	VAEYARRRGR GVCVLSGSGA VNVVALRQPG ASPPGSMVAT LRGRFEILSLT GTVLPPAPPAG ASGLTVFLSGG QGQVIGGSVW GPLVAAGPVW LMAAS	71 %	Sí	Sí

4	G1067	Coordenadas del polipéptido: 86-94, 130-235	KRPRGRPPG	78 %	VSTYARRRGR GVSVLGGNGT VSNVTLRQPV TPNGGGVSG GGVWTLHGR FEILSLTGTVLP PPAPPGAGGLS IFLAGGGQQV VGGSVVAPLIA SAPVILMAASF	69 %	Sin datos	No
16	G3459	Coordenadas del polipéptido: 76-84, 121-216	RRPRGRPPG	89 %	VTAYARRRQR GICVLSGGTV TNVSLRQPAA AGAWTLHGR FEILSLSGSFLP PPAPPGATSLTI YLAGGGQQV GGNVIGELTA AGPVIVIAASF	68 %	Sí	Sí
30	G3400	Coordenadas del polipéptido: 83-91, 127-225	RRPRGRPLG	89 %	VCEFARRRGR GVSVLSGGGA VANVALRQPG ASPPGSLVAT MRGQFEILSLT GTVLPPPPAPS ASGLTVFLSGG QQQVGGGSA GQLIAAGPVFL MAASF	68 %	Sí	Sí
67	G2789	Coordenadas del polipéptido: 59-67; 103-196	RRPRGRPAG	100 %	LAVFARRRQR GVCVLTNGA VTNVTVRQPG GGVSLHGRF EILSLSGSFLPP PAPPAASGLK VYLAGGGQQV IGGSVWGPLTA SSPVVMAAS F	67 %	Sí	No

18	G3460	Coordenadas del polipéptido: 74-82, 118-213	RRPRGRPSG	89 %	VTAYARRRQR GICVLSGSGTV TNVSLRQPAA AGAVVRLHGR FEILSLSGSFLP PPAPPGATSLTI YLAGGGQGVV GGNVVGELTA AGPVIVIAASF	67 %	Sí	Sí
69	G1667	Coordenadas del polipéptido: 53-61; 97-192	KRPRGRPAG	89 %	LSDFARRKQR GLCILSANGCV TNVTLRQPASS GAIVTLHGRY EILSLGSLPP PAPLGITGLTIY LAGPQGQVVG GGVVGGLIAS GPVLMMAASF	66 %	Sin datos	Sí
8	G2156	Coordenadas del polipéptido: 72-80, 116-220	KRPRGRPPG	78 %	VTTYARRRGR GVSILSGNGTV ANVSLRQPAT TAAHGANGGT GGVVALHGRF EILSLTGTVLPP PAPPGSGGLSI FLSGVQGVIG GNVVAPLVAS GPVELMAASF	65 %	Sí	Sí
14	G3456	Coordenadas del polipéptido: 62-70, 106-201	RRPRGRPPG	89 %	VAQFARRRQR GVSILSGSGTV VNVNLRQPTA PGAVMALHGR FDILSLTGSFLP GPSPPGATGLT IYLAGGGQVIV GGEVVGPLVA AGPVLVMAAT F	65 %	Sí	Sí

12	G3407	Coordenadas del polipéptido: 63-71, 106-208	RRPRGRPPG	89 %	L TAYARRRQR GVCVLSAAGT VANVTLRQPQ SAQPGPASPA VATLHGRFEIL SLAGSFLPPA PPGATSLAFL AGGQGVVVG GSVAGALIAA GPVVVAAASF	63 %	Sin datos	Si
38	G3401	Coordenadas del polipéptido: 35-43, 79-174	RRPRGRPPG	89 %	IAHFARRRQR GVCVLSGAGT VTDVALRQPA APSAWVALRG RFEILSLTGIFL PGPAPPGSTGL TVYLAGGQGG WGGSVVGT TAAGPVMVIA STF	63 %	Si	Si
6	G2153	Coordenadas del polipéptido: 80-88, 124-227	RRPRGRPPG	100 %	LATFARRRQR GICILSGNGTV ANVTLRQPST AAVAAAPGGA AVLALQGRFEI LSLTGSFLPGP APPGSTGLTY LAGGQGVVVG GSWGPPLMAA GPVMLIAATF	62 %	Si	Si
42	G1069	Coordenadas del polipéptido: 67-75, 111-206	RRPRGRPPG	89 %	IAHFSRRRQRG VCVLSGTGSV ANVTLRQAAA PGGVVSLQGR FEILSLTGAFPL GPSPPGSTGLT VYLAGVQGGV VGGSVVGPLL AIGSMVIAAT F	62 %	Si*	Si

40	G3556	Coordenadas del polipeptido: 45-53; 89-185	RRPRGRPPG	89 %	IAGFSRRRQRG VSVLSGSGAV TNVTLRQPAG TGAAVALRG RFEILMSGAF LPAPAPPGATG LAVLAGGQG QVVGGSVMGE LIASGPVMVIA ATF	62 %	Sin datos	Si
71	G2157	88-96, 132-228	RRPRGRPPG	89 %	LNAFARRRGR GVSVLSGSGL VTNVTLRQPA ASGGVVSIRG QFEILSMCGAF LPTSGSPAAAA GLTIYLAGAQ GQVVGSGVA GPLIASGPVVI AATF	60 %	No	Si**
* resultados de estudios previos, no mostrados ** observado solamente en una línea								

Dentro del clado G1073 de polipéptidos de factores de transcripción, el dominio de gancho AT comprende la secuencia consenso:

RPRGRPXXG (SEC ID N°: 72)
Arg-Pro-Arg-Gly-Arg-Pro-Xaa-Gly

en la que Xaa puede ser cualquiera de varios restos de aminoácidos; en los ejemplos que se han mostrado hasta la fecha que confieren tolerancia al estrés abiótico, se ha mostrado que Xaa representa un resto de alanina, leucina, prolina o serina.

También dentro del clado G1073, el segundo dominio conservado generalmente comprende la secuencia consenso:

Gly-Xaa-Phe-Xaa-Ile-Leu-Ser-(Xaa)₂-Gly-(Xaa)₂-Leu-
Pro-(Xaa)₃₋₄-Pro-(Xaa)₅-Leu-(Xaa)₂-Tyr/Phe-(Xaa)₂-Gly-(Xaa)₂-Gly-Gln.

Una subsecuencia menor de interés en las secuencias del clado G1073 comprende:

Pro-(Xaa)₅-Leu-(Xaa)₂-Tyr (SEC ID N°: 63); o
Pro-(Xaa)₅-Leu-(Xaa)₂-Phe (SEC ID N°: 64).

La décima posición de SEC ID N°: 63 y 64 es un resto aromático, específicamente tirosina o fenilalanina, en las secuencias del clado G1073 que se han examinado hasta la fecha. Hasta la fecha, no se han encontrado restos aromáticos en la posición correspondiente en los factores de transcripción de gancho AT que están fuera del clado G1073.

Por lo tanto, los factores de transcripción relacionados con la presente invención poseen cada uno un dominio de gancho AT y un segundo dominio conservado, e incluyen parálogos y ortólogos de G1073 hallados por análisis de BLAST, como se describe posteriormente. Como se muestra en la Tabla 1, los dominios de gancho AT de G1073 y secuencias relacionadas son al menos 78 % idénticos a los dominios de gancho AT de G1073 y al menos 62 % idénticos al segundo dominio conservado hallado en G1073. Estos factores de transcripción se basan en la especificidad de unión de sus dominios de gancho AT; se ha mostrado que muchos tienen funciones similares o idénticas en plantas aumentando el tamaño y la biomasa de una planta (véase también Ejemplo VII, posterior).

Producción de polipéptidos

Los polinucleótidos relacionados con la presente invención incluyen secuencias que codifican factores de transcripción y polipéptidos homólogos de factores de transcripción y secuencias complementarias de los mismos, así como fragmentos únicos de secuencia codificante, o secuencia complementaria de la misma. Dichos polinucleótidos pueden ser, por ejemplo, ADN o ARN, por ejemplo, ARNm, ARNc, ARN sintético, ADN genómico, ADNc, ADN sintético, oligonucleótidos, etc. Los polinucleótidos son bicatenarios o monocatenarios e incluyen bien secuencias con sentido (es decir, codificantes) o bien secuencias antisentido (es decir, no codificantes, complementarias) o ambas. Los polinucleótidos incluyen la secuencia codificante de un factor de transcripción, o polipéptido homólogo del factor de transcripción, aislado, en combinación con secuencias codificantes adicionales (por ejemplo, un marcador de purificación, una señal de localización, como una proteína de fusión, como una preproteína, o similares), en combinación con secuencias no codificantes (por ejemplo, intrones o inteínas, elementos reguladores tales como promotores, potenciadores, terminadores y similares), y/o en un ambiente de vector y hospedador en el que el polinucleótido que codifica un factor de transcripción o polipéptido homólogo del factor de transcripción es un gen endógeno o exógeno,

Existe diversos métodos para producir polinucleótidos. Se conocen bien por los expertos en la materia procedimientos para identificar y aislar clones de ADN y se describen en, por ejemplo, Berger y Kimmel (1987), "Guide to Molecular Cloning Techniques", en Methods in Enzymology, vol. 152, Academic Press, Inc., San Diego, CA ("Berger"); Sambrook *et al.* (1989) Molecular Cloning - A Laboratory Manual (2ª Edición), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, and Current Protocols in Molecular Biology, Ausubel *et al.* editores, Current Protocols, Greene Publishing Associates, Inc. y John Wiley & Sons, Inc. (complementado mediante 2000; "Ausubel").

Como alternativa, pueden producirse polinucleótidos por diversos métodos de amplificación *in vitro* adaptados a la presente invención mediante selección apropiada de cebadores específicos o degradados. Se encuentran ejemplos de protocolos suficientes para dirigir a personas expertas a través de los métodos de amplificación *in vitro*, incluyendo la reacción en cadena de la polimerasa (PCR), la región en cadena de la ligasa (LCR), amplificación por Q β -replicasa y otras técnicas mediadas por ARN polimerasa (por ejemplo, NASBA), por ejemplo, para la producción de los ácidos nucleicos homólogos en Berger (1987) mencionado anteriormente, Sambrook (1989) mencionado anteriormente y Ausubel (hasta 2000) mencionado anteriormente, así como Mullis *et al.* (1990) PCR Protocols A Guide to Methods and Applications (Innis *et al.*, eds) Academic Press Inc. San Diego, CA. Se describen métodos mejorados para clonación de ácidos nucleicos amplificados *in vitro* en Wallace *et al.* Patente de Estados Unidos N°

5.426.039. Se resumen métodos mejorados para amplificar ácidos nucleicos grandes por PCR en Cheng *et al.* (1994) Nature 369: 684-685 y las referencias citadas en la misma, en la que se generan amplicones de PCR de hasta 40 kb. Un experto en la materia apreciará que esencialmente cualquier ARN puede convertirse en un ADN bicatenario adecuado para digestión de restricción, expansión por PCR y secuenciación usando transcriptasa inversa y una polimerasa (Berger (1987) mencionado anteriormente; Sambrook (1989) mencionado anteriormente; y Ausubel (hasta 2000) mencionado anteriormente).

Como alternativa, pueden ensamblarse polinucleótidos y oligonucleótidos a partir de fragmentos producidos por métodos de síntesis de fase sólida. Normalmente, se sintetizan individualmente fragmentos de hasta aproximadamente 100 bases y después se ligan de forma enzimática o química para producir una secuencia deseada, por ejemplo, un polinucleótido que codifica todo o una parte de un factor de transcripción. Por ejemplo, se describe síntesis química usando el método de fosforamidita, por ejemplo, en Beaucage *et al.* (1981) Tetrahedron Letters 22: 1859-1869; y Matthes *et al.* (1984) EMBO J. 3: 801-805. De acuerdo con dichos métodos, se sintetizan oligonucleótidos, se purifican, se hibridan con su cadena complementaria, se ligan y después se clonan opcionalmente en vectores adecuados. Si se desea, los polinucleótidos y polipéptidos pueden pedirse a medida de cualquiera de varios proveedores comerciales.

Secuencias homólogas

También se describen en el presente documento secuencias homólogas a las proporcionadas en el Listado de Secuencias derivadas de *Arabidopsis thaliana* o de otras plantas elegidas. Pueden derivarse secuencias homólogas de cualquier planta incluyendo monocotiledóneas y dicotiledóneas y en particular especies vegetales agrícolaemente importantes, incluyendo pero sin limitación cultivos tales como soja, trigo, maíz, patata, algodón, arroz, colza, colza oleaginosa (incluyendo canola), girasol, alfalfa, trébol, caña de azúcar y pasto; o frutas y verduras, tales como plátano, mora, arándano, fresa y frambuesa, cantalupo, zanahoria, coliflor, café, pepino, berenjena, uvas, melón dulce, lechuga, mango, melón, cebolla, papaya, guisantes, pimientos, piña, calabaza gigante, espinaca, calabaza, maíz dulce, tabaco, tomate, tomatillo, sandía, frutas rosáceas (tales como manzana, melocotón, pera, cereza y ciruela) y hortalizas del género *Brassica* (tales como brócoli, repollo, coliflor, coles de Bruselas y colirrábano). Otros cultivos, incluyendo frutas y verduras, cuyo fenotipo puede cambiarse y que comprenden secuencias homólogas incluyen cebada; centeno; mijo; sorgo; grosella; aguacate; frutas cítricas tales como naranjas, limones, pomelos y mandarinas, alcachofa, cerezas; nueces tales como la nuez y el cacahuete; endivia; puerro; raíces tales como arruruz, remolacha, yuca, nabo, rábano, ñame y boniato; y judías. Las secuencias homólogas también pueden derivar de especies leñosas, tales como pino, chopo y eucalipto o menta u otras lamiáceas. Además, las secuencias homólogas pueden derivar de plantas que están relacionadas evolutivamente con plantas de cultivo, pero que pueden no haberse usado aún como plantas de cultivo. Los ejemplos incluyen belladona (*Atropa belladonna*), relacionada con el tomate; estramonio (*Datura strommium*), relacionada con el peyote; y teosinte (especie de *Zea*), relacionada con el maíz.

Ortólogos y parálogos

Las secuencias homólogas como se ha descrito anteriormente pueden comprender secuencias ortólogas o parálogos. Se conocen varios métodos diferentes por los expertos en la materia para identificar y definir estas secuencias funcionalmente homólogas. Se describen tres métodos generales para definir ortólogos y parálogos; un ortólogo o parálogo, incluyendo equivalentes, puede identificarse por uno o más de los métodos descritos posteriormente.

Dentro de una única especie vegetal, la duplicación génica puede provocar dos copias de un gen particular, dando lugar a dos o más genes con secuencia similar y con frecuencia función similar conocidos como parálogos. Un parálogo es por lo tanto un gen similar formado por duplicación dentro de la misma especie. Los parálogos normalmente se agrupan entre sí o en el mismo clado (un grupo de genes similares) cuando se analiza una filogenia familiar de un gen usando programas tales como CLUSTAL (Thompson *et al.* (1994) Nucleic Acids Res. 22: 4673-4680; Higgins *et al.* (1996) Methods Enzymol. 266: 383-402). También pueden identificarse grupos de genes similares con análisis de BLAST por pares (Feng y Doolittle (1987) J. Mol. Evol. 25: 351-360). Por ejemplo, un clado de factores de transcripción del dominio MADS muy similares de *Arabidopsis* comparten todos una función común en el tiempo de floración (Ratcliffe *et al.* (2001) Plant Physiol. 126: 122-132), y un grupo de factores de transcripción de dominio AP2 muy similares de *Arabidopsis* están implicados en la tolerancia de las plantas a congelación (Gilmour *et al.* (1998) Plant J. 16: 433-442). El análisis de grupos de genes similares con función similar que quedan dentro de un clado puede producir subsecuencias que son particulares del clado. Estas subsecuencias, conocidas como secuencias consenso, pueden usarse no solamente para definir la secuencia dentro de cada clado, sino que definen las funciones de estos genes; los genes dentro de un clado pueden contener secuencias parálogos, o secuencias ortólogas que comparten la misma función (véase también, por ejemplo, Mount (2001), en Bioinformatics: Sequence and Genome Analysis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York, página 543).

La especiación, la producción de nuevas especies a partir de una especie parental, también puede dar lugar a dos o más genes con secuencia similar y función similar. Estos genes, denominados ortólogos, con frecuencia tienen una función idéntica dentro de sus plantas hospedadoras y son con frecuencia intercambiables entre especies sin

- pérdida de función. Debido a que las plantas tienen ancestros comunes, muchos genes en cualquier especie vegetal tendrán un gen ortólogo correspondiente en otra especie vegetal. Una vez que se ha construido un árbol filogenético para una familia génica de una especie usando un programa tal como CLUSTAL (Thompson *et al.* (1994) Nucleic Acids Res. 22: 4673-4680; Higgins *et al.* (1996) mencionado anteriormente) pueden situarse secuencias ortólogas potenciales en el árbol filogenético y puede determinarse su relación con genes de la especie de interés. También pueden identificarse secuencias ortólogas por una estrategia de BLAST recíproca. Una vez que se ha identificado una secuencia ortóloga, puede deducirse la función del ortólogo a partir de la función identificada de la secuencia de referencia.
- 10 Las secuencias génicas del factor de transcripción están conservadas entre diversas líneas de especies eucariotas (Goodrich *et al.* (1993) Cell 75: 519-530; Lin *et al.* (1991) Nature 353: 569-571; Sadowski *et al.* (1988) Nature 335: 563-564). Las plantas no son una excepción a esta observación; diversas especies vegetales poseen factores de transcripción que tienen secuencias y funciones similares.
- 15 Los genes ortólogos de diferentes organismos tienen funciones altamente conservadas, y muy frecuentemente funciones esencialmente idénticas (Lee *et al.* (2002) Genome Res. 12: 493-502; Remm *et al.* (2001) J. Mol. Biol. 314: 1041-1052). Los genes parálogos que han divergido mediante duplicación génica, pueden conservar funciones similares de las proteínas codificadas. En dichos casos, pueden usarse parálogos de forma intercambiable con respecto a ciertas realizaciones de la presente invención (por ejemplo, expresión transgénica de una consecuencia codificante). Un ejemplo de dichos parálogos altamente relacionados es la familia CBF, con tres miembros bien definidos en *Arabidopsis* y al menos un ortólogo en *Brassica napus*, todos los cuales controlan rutas implicadas en el estrés tanto de congelación como de sequía (Gilmour *et al.* (1998) Plant J. 16: 433-442; Jaglo *et al.* (1998) Plant Physiol. 127: 910-917).
- 20 Las siguientes referencias representan una pequeña muestra de los muchos estudios que demuestran que los genes de factores de transcripción conservados de diversas especies probablemente actúan de forma similar (es decir, regulan secuencias diana similares y controlan los mismos rasgos), y que los factores de transcripción pueden transformarse en diversas especies para conferir o mejorar rasgos.
- 25 (1) Se ha mostrado que factores de transcripción de *Arabidopsis* definidos, incluyendo G28 (hallado en la Patente de Estados Unidos 6.664.446), G482 (hallado en la Solicitud de Patente de Estados Unidos 20040045049), G867 (hallado en la Solicitud de Patente de Estados Unidos 20040098764) y G1073 (hallado en la Patente de Estados Unidos 6.717.034) confieren tolerancia al estrés o biomasa aumentada cuando se sobreexpresan las secuencias. Las secuencias polipeptídicas pertenecen a distintos clados de polipéptidos de factores de transcripción que incluyen miembros de diversas especies. En cada caso, se ha mostrado que un número significativo de secuencias de miembros del clado derivadas tanto de dicotiledóneas como de monocotiledóneas confieren biomasa aumentada o tolerancia al estrés cuando las secuencias se sobreexpresaron (datos no publicados).
- 30 (2) El gen NPR1 de *Arabidopsis* regula la resistencia adquirida sistémica (SAR); la sobreexpresión de NPR1 conduce a resistencia potenciada en *Arabidopsis*. Cuando NPR1 de *Arabidopsis* o el ortólogo de NPR1 del arroz se sobreexpresó en arroz (que, como una monocotiledónea, es distinto de *Arabidopsis*), y se expuso al patógeno de roya bacteriana del arroz *Xanthomonas oryzae* pv. *Oryzae*, las plantas transgénicas presentaron resistencia potenciada (Chern *et al.* (2001) Plant J. 27: 101-113). NPR1 actúa mediante la activación de la expresión de genes de factores de transcripción, tales como TGA2 (Fan y Dong (2002) Plant Cell 14: 1377-1389).
- 35 (3) Genes de E2F están implicados en la transcripción de genes vegetales para antígeno nuclear celular en proliferación (PCNA). Los E2F vegetales comparten un alto grado de similitud en la secuencia de aminoácidos entre monocotiledóneas y dicotiledóneas, y son incluso similares a los dominios conservados de los E2F animales. Dicha conservación indica una similitud funcional entre E2F vegetales y animales. Los factores de transcripción de E2F que regulan el desarrollo del meristemo actúan mediante elementos en cis comunes, y regulan genes relacionados (PCNA) (Kosugi y Ohashi, (2002) Plant J. 29: 45-59).
- 40 (4) El gen de ABI5 (insensible a ABA 5) codifica un factor de cremallera de leucina básico requerido para la respuesta de ABA en los tejidos de semillas y vegetativos. Los experimentos de cotransformación con construcciones de ADNc de ABI5 en protoplastos de arroz dieron como resultado la transactivación específica de promotores de trigo, *Arabidopsis*, judías y cebada inducibles por ABA. Estos resultados demuestran que factores de transcripción de ABI5 secuencialmente similares son dianas clave de una ruta de señalización de ABA conservada en diversas plantas (Gampala *et al.* (2001) J. Biol. Chem. 277: 1689-1694).
- 45 (5) Se obtuvieron secuencias de tres genes de tipo GAMYB de *Arabidopsis* basándose en la similitud de secuencia con los genes de GAMYB de cebada, arroz y *L. temulentum*. Se determinó que estos tres genes de *Arabidopsis* codificaban factores de transcripción (AtMYB33, AtMYB65 y AtMYB101) y podrían sustituir a un GAMYB de cebada y controlar la expresión de alfa amilasa (Gocal *et al.* (2001) Plant Physiol. 127: 1682-1693).
- 50 (6) El gen de control floral LEAFY de *Arabidopsis* puede acelerar drásticamente la floración en numerosas plantas dicotiledóneas. La expresión constitutiva de LEAFY de *Arabidopsis* también provocó la floración temprana en arroz transgénico (una monocotiledónea), con una fecha de espigación que fue de 26-34 días más temprano que la de las plantas de tipo silvestre. Estas observaciones indican que los genes reguladores florales de *Arabidopsis* son herramientas útiles para mejorar la fecha de espigación en cultivos de cereales (He *et al.* (2000) Transgenic Res. 9: 223-227).
- 55 (7) Las giberelinas bioactivas (GA) son reguladores endógenos esenciales del crecimiento vegetal. La

señalización de GA tiende a estar conservada en todo el reino vegetal. La señalización de GA está mediada mediante GAI, un miembro nuclear de la familia de GRAS de factores de transcripción vegetales. Se ha mostrado que GAI de *Arabidopsis* actúa en el arroz para inhibir las rutas de respuesta a giberelina. (Fu *et al.* (2001) Plant Cell 13: 1791-1802).

(8) El gen de *Arabidopsis* SUPERMAN (SUP), codifica un factor de transcripción potencial que mantiene la separación entre los estambres y los carpelos. Sobreexpresando SUP de *Arabidopsis* en el arroz, se mostró que el efecto de la presencia del gen en las separaciones de verticilos estaba conservado. Esto demostró que SUP es un regulador conservado de las separaciones de verticilos florales y afecta a la proliferación celular (Nandi *et al.* (2000) Curr. Biol. 10: 215-218.)

(9) Factores de transcripción de myb de maíz, petunia y *Arabidopsis* que regulan la biosíntesis de flavonoides son genéticamente muy similares y afectan al mismo rasgo en su especie nativa, por lo tanto la secuencia y función de estos factores de transcripción de myb se correlacionan entre sí en estas diversas especies (Borevitz *et al.* (2000) Plant Cell 12: 2383-2394).

(10) Los genes reduced height-1 del trigo (Rht-B1/Rht-D1) y dwarf-8 de maíz (d8) son ortólogos del gen insensible a giberelina (GAI) de *Arabidopsis*. Ambos de estos genes se han usado para producir variedades de grano enano que han mejorado la producción de grano. Estos genes codifican proteínas que se asemejan a factores de transcripción nucleares y contienen un dominio de tipo SH2, lo que indica que la fosfotirosina puede participar en la señalización de giberelina. Se ha mostrado que las plantas de arroz transgénico que contienen un alelo de GAI mutante de *Arabidopsis* producen respuestas reducidas a giberelina y se hacen enanas, lo que indica que los ortólogos de GAI mutantes podrían usarse para aumentar el rendimiento en una amplia serie de especies de cultivo (Peng *et al.* (1999) Nature 400: 256-261.)

Los factores de transcripción que son homólogos de los factores de transcripción de gancho AT enumerados normalmente compartirán al menos aproximadamente 78 % y 62 % de la identidad de secuencia de aminoácidos en sus dominios de gancho AT y segundo conservado, respectivamente. Los factores de transcripción más estrechamente relacionados pueden compartir al menos aproximadamente el 89 % o aproximadamente el 100 % de identidad en sus dominios de gancho AT, y al menos aproximadamente 62 %, 63 %, 65 %, 66 %, 67 %, 68 %, 69 %, 71 %, o mayor identidad con el segundo dominio conservado de G1073, como se ve por los ejemplos que se ha mostrado que confieren tolerancia al estrés abiótico en la Tabla 1. Los factores de transcripción que son homólogos de las secuencias enumeradas deberían compartir al menos aproximadamente el 50 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 90 %, o al menos aproximadamente el 95 % de identidad de secuencia de aminoácidos sobre la longitud completa del polipéptido o el homólogo.

Al nivel de nucleótidos, las secuencias descritas en el presente documento normalmente compartirán al menos aproximadamente el 40 % de identidad de secuencia de nucleótidos, preferentemente al menos aproximadamente el 50 %, aproximadamente el 60 %, aproximadamente el 70 % o aproximadamente el 80 % de identidad de secuencia, y más preferentemente aproximadamente el 85 %, aproximadamente el 90 %, aproximadamente el 95 % o aproximadamente el 97 % o más de identidad de secuencia con una o más de las secuencias de longitud completa enumeradas, o con una secuencia enumerada pero excluyendo o fuera de la región o las regiones que codifican una secuencia consenso conocida o sitio de unión a ADN consenso, o fuera de la región o regiones que codifican uno o todos los dominios conservados. La regeneración del código genético permite variaciones importantes en la secuencia de nucleótidos de un polinucleótido manteniendo a la vez la secuencia de aminoácidos de la proteína codificada.

El porcentaje de identidad puede determinarse de forma electrónica, por ejemplo, usando el programa MEGALIGN (DNASTAR, Inc. Madison, Wis.). El programa MEGALIGN puede crear alineamientos entre dos o más secuencias de acuerdo con diferentes métodos, por ejemplo, el método de clustal (véase, por ejemplo, Higgins y Sharp (1988) Gene 73: 237-244). El algoritmo de clustal agrupa las secuencias en grupos examinando las distancias entre todos los pares. Los grupos se alinean por pares y después en grupos. Pueden usarse otros algoritmos de alineamiento o programas, incluyendo FASTA, BLAST o ENTREZ, FASTA y BLAST, y que pueden usarse para calcular el porcentaje de similitud. Estos están disponibles como parte del paquete de análisis de secuencias GCG (Universidad de Wisconsin, Madison, WI) y pueden usarse con o sin ajustes por defecto. ENTREZ está disponible a través del Centro Nacional para la Información Biotecnológica. En una realización, el porcentaje de identidad de dos secuencias puede determinarse por el programa GCG con un peso de hueco de 1, por ejemplo, cada hueco de aminoácido se pondera como si fuera un desapareamiento de un único aminoácido o nucleótido entre las dos secuencias (véase documento USPN 6.262.333).

Otras técnicas para alineamiento se describen en Methods in Enzymology, vol. 266, Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., San Diego, Calif., Estados Unidos. Preferentemente, se utiliza un programa de alineamiento que permite huecos en la secuencia para alinear las secuencias. El Smith-Waterman es un tipo de algoritmo que permite huecos en los alineamientos de secuencias (véase Shpaer (1997) Methods Mol. Biol. 70: 173-187). Además, el programa GAP que usa el método de alineamiento de Needleman y Wunsch puede utilizarse para alinear secuencias. Una estrategia de búsqueda alternativa usa software MPSRCH, que se procesa en un ordenador MASPAR. MPSRCH usa un algoritmo de Smith-Waterman para puntuar secuencias en una ordenador paralelo masivo. Este enfoque mejora la capacidad para seleccionar coincidencias lejanamente relacionadas, y es especialmente tolerante de huecos pequeños y errores de

secuencia de nucleótidos. Pueden usarse secuencias de aminoácidos codificadas por ácido nucleico para buscar bases de datos tanto proteicas como de ADN.

El porcentaje de similitud entre dos secuencias polipeptídicas, por ejemplo, secuencia A y secuencia B, se calcula dividiendo la longitud de la secuencia A, menos el número de restos con huecos en la secuencia A, menos el número de restos con huecos en la secuencia B, entre la suma de las coincidencias de restos entre la secuencia A y secuencia B, por cien. Los huecos de baja o ninguna similitud entre las dos secuencias de aminoácidos no se incluyen en la determinación del porcentaje de similitud. El porcentaje de identidad entre las secuencias polinucleotídicas también puede contarse o calcularse por otros métodos conocidos en la técnica, por ejemplo, el método de Jotun Hein (véase, por ejemplo, Hein (1990) *Methods Enzymol.* 183: 626-645). La identidad entre las secuencias también puede determinarse por otros métodos conocidos en la técnica, por ejemplo, variando las condiciones de hibridación (véase Solicitud de Patente de Estados Unidos N° 20010010913).

Por lo tanto, los métodos para identificar una secuencia similar o paróloga, ortóloga u homóloga de uno o más polinucleótidos como se observa en el presente documento, o uno o más polipéptidos diana codificados por los polinucleótidos, u observados de otro modo en el presente documento y pueden incluir enlace o asociación de un fenotipo vegetal dado o función génica con una secuencia. En los métodos, se proporciona una base de datos de secuencias (localmente o a través de Internet o una intranet) y se realiza una consulta frente a la base de datos de secuencias usando las secuencias relevantes en el presente documento y fenotipos vegetales o funciones génicas asociados.

Además, una o más secuencias polinucleotídicas o uno o más polipéptidos codificados por las secuencias polinucleotídicas pueden usarse para buscar frente a una base de datos BLOCKS (Bairoch *et al.* (1997) *Nucleic Acids Res.* 25: 217-221), PFAM y otras que contienen motivos, secuencias y funciones génicas previamente identificados y anotados. Los métodos que buscan patrones de secuencia primarios con penalizaciones de hueco de estructura secundarios (Smith *et al.* (1992) *Protein Engineering* 5: 35-51) así como algoritmos tales como la Herramienta de Búsqueda de Alineamiento Local Básica (BLAST; Altschul (1993) *J. Mol. Evol.* 36: 290-300; Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410), BLOCKS (Henikoff y Henikoff (1991) *Nucleic Acids Res.* 19: 6565-6572), Modelos de Markov Ocultos (HMM; Eddy (1996) *Curr. Opin. Str. Biol.* 6: 361-365; Sonnhammer *et al.* (1997) *Proteins* 28: 405-420) y similares, pueden usarse para manipular y analizar secuencias polinucleotídicas y polipeptídicas codificadas por polinucleótidos. Estas bases de datos, algoritmos y otros métodos se conocen bien en la técnica y se describen en Ausubel *et al.* (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, Nueva York, NY, unidad 7.7, y en Meyers (1995) *Molecular Biology and Biotechnology*, Wiley VCH, Nueva York, NY, p 856-853.

Un método adicional para identificar o confirmar que las secuencias homólogas específicos controlan la misma función es comparando el perfil o los perfiles del transcrito obtenidos tras la sobreexpresión o anulación de dos o más factores de transcripción relacionados. Ya que los perfiles de transcritos son diagnósticos para estados celulares específicos, un experto en la materia apreciará que los genes que tienen un perfil de transcritos altamente similar (por ejemplo, con más del 50 % de transcritos regulados en común, con más del 70 % de transcritos regulados en común, o con más del 90 % de transcritos regulados en común) tendrán funciones altamente similares. Fowler *et al.* (2002) *Plant Cell* 14: 1675-1679, han mostrado que tres genes parálogos de la familia AP2 (CBF1, CBF2 y CBF3), cada uno de los cuales se induce tras tratamiento con frío, y cada uno de los cuales puede condicionar la tolerancia a congelación mejorada, tienen perfiles de transcritos altamente similares. Una vez que se ha mostrado que un factor de transcripción proporciona una función específica, su perfil de transcritos se convierte en una herramienta de diagnóstico para determinar si los parálogos u ortólogos potenciales tienen la misma función.

Además, pueden usarse métodos que usan alineamiento manual de secuencias similares u homólogas con una o más secuencias polinucleotídicas o uno o más polipéptidos codificados por las secuencias polinucleotídicas para identificar regiones de similitud y dominios de gancho AT. Dichos métodos manuales se conocen bien por los expertos en la materia y pueden incluir, por ejemplo, comparaciones de estructura terciaria entre una secuencia polipeptídica codificada por un polinucleótido que comprende una función conocida con una secuencia polipeptídica codificada por una secuencia polinucleotídica que tiene una función aún no determinada. Dichos ejemplos de estructura terciaria pueden comprender hélices alfa predichas, láminas beta, hélices anfipáticas, motivos de cremallera de leucina, motivos de dedos de cinc, regiones ricas en prolina, motivos de repetición de cisteína y similares.

Los ortólogos y parálogos de factores de transcripción desvelados en el presente documento pueden clonarse usando composiciones proporcionadas por la presente invención de acuerdo con métodos bien conocidos en la técnica. Pueden clonarse ADNc usando ARNm de una célula o tejido vegetal que expresa uno de los presentes factores de transcripción. Las fuentes de ARNm apropiadas pueden identificarse explorando transferencias de Northern con sondas diseñadas a partir de las presentes secuencias de factores de transcripción, después de lo cual se prepara una biblioteca a partir del ARNm obtenido de una célula o tejido positivo. Después se aísla ADNc codificante del factor de transcripción usando, por ejemplo, PCR, usando cebadores diseñados a partir de una secuencia génica del factor de transcripción desvelado en el presente documento, o explorando con un ADNc parcial o completo o con uno o más conjuntos de sondas degradadas basándose en las secuencias desveladas. La biblioteca de ADNc puede usarse para transformar células vegetales. Se detecta expresión de los ADNc de interés

usando, por ejemplo, métodos desvelados en el presente documento tales como micromatrices, transferencias de Northern, PCR cuantitativa o cualquier otra técnica para controlar los cambios en la expresión. Los clones genómicos pueden aislarse usando técnicas similares a estas.

Los ejemplos de ortólogos de las secuencias polipeptídicas de *Arabidopsis* SEC ID N°: 2, 4, 6, 8, 42 y 69, incluyen SEC ID N°: 10, 12, 14, 16, 18, 26, 30, 38, 40, y otros ortólogos funcionalmente similares enumerados en el Listado de Secuencias. Además de las secuencias en el Listado de Secuencias, se describen en el presente documento secuencias de nucleótidos aisladas que son secuencial y estructuralmente similares a G1073, G1067, G2153, G2156, G3399, G3407, G3456, G3459, G3460, G3406, G3400, G3401, G3556, G1069, G2789 y G1667 (SEC ID N°: 1, 3, 5, 7, 9, 11, 13, 15, 17, 25, 29, 37, 39, 41, 66 y 68, respectivamente) y pueden actuar en una planta aumentando la biomasa y la tolerancia al estrés abiótico, particularmente cuando se sobreexpresan. Estas secuencias polipeptídicas representan miembros del clado que actúan de forma similar a G1073 confiriendo tolerancia al estrés abiótico y muestran similitud de secuencia significativa con G1073, particularmente en sus dominios conservados respectivos, como se identifica en la Tabla 1.

Ya que un número representativo de estas secuencias polinucleotídicas en el clado G1073 de polipéptidos de factores de transcripción están filogenética (Figura 4) y secuencialmente (Figura 5A-5H) relacionadas y se ha mostrado que aumentan la biomasa de una planta y su tolerancia al estrés abiótico, un experto en la materia podría predecir que otras secuencias similares, filogenéticamente relacionadas, que quedan dentro del clado G1073 también aumentarían la biomasa de una planta y su tolerancia al estrés abiótico cuando se sobreexpresen.

Identificación de polinucleótidos o ácidos nucleicos por hibridación

Pueden identificarse polinucleótidos homólogos de las secuencias ilustradas en el Listado de Secuencias y las tablas, por ejemplo, por hibridación entre sí en condiciones rigurosas o en condiciones altamente rigurosas. Los polinucleótidos monocatenarios hibridan cuando se asocian basándose en diversas fuerzas físico-químicas bien caracterizadas, tales como enlaces de hidrógeno, exclusión de disolvente, apilamiento de bases y similares. La rigurosidad de una hibridación refleja el grado de identidad de secuencia de los ácidos nucleicos implicados, de modo que cuanto mayor sea la rigurosidad, más similares serán las dos cadenas polinucleotídicas. La rigurosidad se ve influida por diversos factores, incluyendo temperatura, concentración salina y composición, aditivos orgánicos y no orgánicos, disolventes, etc. presentes en las soluciones tanto de hibridación como de lavado e incubaciones (y número de las mismas), como se describe en más detalle en las referencias citadas posteriormente (por ejemplo, Sambrook *et al.* (1989); Berger y Kimmel (1987); y Anderson y Young - (1985)).

Las secuencias polinucleotídicas pueden ser capaces de hibridar con los polinucleótidos del factor de transcripción dentro del Listado de Secuencias, y fragmentos de los mismos en diversas condiciones de rigurosidad (véase, por ejemplo, Wahl y Berger (1987) *Methods Enzymol.* 152: 399-407; y Kimmel (1987) *Methods Enzymol.* 152: 507-511). Además de las secuencias de nucleótidos enumeradas en el Listado de Secuencias, puede identificarse ADNc de longitud completa, ortólogos y parálogos de las presentes secuencias de nucleótidos y aislarse usando métodos bien conocidos. Las bibliotecas de ADNc, ortólogos y parálogos de las presentes secuencias de nucleótidos pueden explorarse usando métodos de hibridación para determinar su utilidad como sondas de diana de hibridación o amplificación.

Con respecto a la hibridación, se conocen bien en la técnica condiciones que son altamente rigurosas, y medios para conseguirlas. Véase, por ejemplo, Sambrook *et al.* (1989) "Molecular Cloning: A Laboratory Manual" (2ª ed., Cold Spring Harbor Laboratory); Berger (1987) mencionada anteriormente, páginas 467-469; y Anderson y Young (1985) "Quantitative Filter Hybridisation", En: Hames y Higgins, ed., *Nucleic Acid Hybridisation, A Practical Approach*. Oxford, IRL Press, 73-111.

La estabilidad de dobles cadenas de ADN se ve afectada por factores tales como composición de bases, longitud y grado de desapareamiento de pares de bases. Las condiciones de hibridación pueden ajustarse para permitir que hibriden ADN de diferente relación de secuencias. La temperatura de fusión (T_m) se define como la temperatura a la que el 50 % de las moléculas bicatenarias se han disociado en sus cadenas individuales constituyentes. La temperatura de fusión de una doble cadena perfectamente coincidente, cuando el tampón de hibridación contiene formamida como un agente desnaturalizante, puede estimarse por las siguientes ecuaciones:

(I) ADN-ADN:

$$T_m(^{\circ}\text{C}) = 81,5 + 16,6(\log[\text{Na}^+] + 0,41(\% \text{ de G+C}) - 0,62(\% \text{ de formamida}) - 500/L$$

(II) ADN-ARN:

$$T_m(^{\circ}\text{C}) = 79,8 + 18,5(\log[\text{Na}^+] + 0,58(\% \text{ de G+C}) - 0,12(\% \text{ de G+C})^2 - 0,5(\% \text{ de formamida}) - 820/L$$

(III) ARN-ARN:

$$T_m(^{\circ}\text{C}) = 79,8 + 18,5(\log[\text{Na}^+] + 0,58(\% \text{ de G+C}) - 0,12(\% \text{ de G+C})^2 - 0,35(\% \text{ de formamida}) - 820/L$$

en las que L es la longitud de la doble cadena formada, $[\text{Na}^+]$ es la concentración molar del ión sodio en la solución de hibridación o lavado y el % de G+C es el porcentaje de bases de (guanina+citosina) en el híbrido. Para híbridos que no coinciden perfectamente, se requiere reducir la temperatura de fusión aproximadamente 1 °C por cada desapareamiento del 1 %.

Se realizan en general experimentos de hibridación en un tampón de pH entre 6,8 y 7,4, aunque la tasa de hibridación es casi independiente del pH a fuerzas iónicas que probablemente se usen en el tampón de hibridación (Anderson y Young (1985) mencionado anteriormente). Además, puede usarse uno o más de los siguientes para reducir la hibridación no específica: ADN de esperma de salmón sonificado u otro ADN no complementario, albúmina de suero bovino, pirofosfato sódico, dodecilsulfato sódico (SDS), polivinilpirrolidona, ficoll y solución de Denhardt. El dextran sulfato y polietilenglicol 6000 actúan para excluir el ADN de la solución, aumentando de este modo la concentración de ADN de sonda eficaz y la señal de hibridación dentro de una unidad de tiempo dada. En algunos casos, pueden ser deseables o requerirse condiciones de rigurosidad aún mayor para reducir la hibridación no específica y/o de fondo. Estas condiciones pueden crearse con el uso de mayor temperatura, menor fuerza iónica y mayor concentración de un agente desnaturizante tal como formamida.

Las condiciones de rigurosidad pueden ajustarse para explorar con respecto a fragmentos moderadamente similares tales como secuencias homólogas de organismos lejanamente relacionados, o para fragmentos altamente similares tales como genes que duplican las enzimas funcionales de organismos estrechamente relacionados. La rigurosidad puede ajustarse durante la etapa de hibridación o en los lavados posthibridación. La concentración salina, concentración de formamida, temperatura de hibridación y longitudes de sonda son variables que pueden usarse para alterar la rigurosidad (como se ha descrito por la fórmula anterior). Como una directriz general la alta rigurosidad se realiza normalmente de $T_m - 5^{\circ}\text{C}$ a $T_m - 20^{\circ}\text{C}$, rigurosidad moderada de $T_m - 20^{\circ}\text{C}$ a $T_m - 35^{\circ}\text{C}$ y baja rigurosidad de $T_m - 35^{\circ}\text{C}$ a $T_m - 50^{\circ}\text{C}$ para doble cadena de > 150 pares de bases. Puede realizarse hibridación a rigurosidad de baja a moderada ($25 - 50^{\circ}\text{C}$ por debajo de la T_m), seguido de lavados posthibridación a rigurosidades crecientes. Se determina de forma empírica que las tasas máximas de hibridación en solución se producen a $T_m - 25^{\circ}\text{C}$ para dobles cadenas ADN-ADN y $T_m - 15^{\circ}\text{C}$ para dobles cadenas ARN-ADN. Opcionalmente, el grado de disociación puede evaluarse después de cada etapa de lavado para determinar la necesidad de etapas de lavado de mayor rigurosidad, posteriores.

Pueden usarse condiciones de alta rigurosidad para seleccionar secuencias de ácido nucleico con altos grados de identidad con las secuencias desveladas. Un ejemplo de condiciones de hibridación rigurosas obtenidas en un método basado en filtro tal como una transferencia de Southern o Northern para hibridación de ácidos nucleicos complementarios que tienen más de 100 restos complementarios es aproximadamente de 5°C a 20°C menor que el punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y pH definidos. Las condiciones usadas para hibridación pueden incluir cloruro sódico de aproximadamente 0,02 M a aproximadamente 0,15 M, caseína de aproximadamente 0,5 % a aproximadamente 5 %, SDS aproximadamente 0,02 % o N-laurilsarcosina 0,1 %, citrato sódico de aproximadamente 0,001 M a aproximadamente 0,03 M, a temperaturas de hibridación entre aproximadamente 50°C y aproximadamente 70°C . Más preferentemente, las condiciones de alta rigurosidad son cloruro sódico aproximadamente 0,02 M, caseína aproximadamente 0,5 %, SDS aproximadamente 0,02 %, citrato sódico aproximadamente 0,001 M, a una temperatura de aproximadamente 50°C . Las moléculas de ácido nucleico que hibridan en condiciones rigurosas normalmente hibridarán con una sonda basándose en la molécula de ADN completa o partes seleccionadas, por ejemplo, con una subsecuencia única, del ADN.

La concentración salina rigurosa será habitualmente menor de aproximadamente 750 mM de NaCl y 75 mM de citrato trisódico. Pueden obtenerse condiciones crecientemente rigurosas con menos de aproximadamente 500 mM de NaCl y 50 mM de citrato trisódico, hasta una rigurosidad aún mayor con menos de aproximadamente 250 mM de NaCl y 25 mM de citrato trisódico. Puede obtenerse hibridación de baja rigurosidad en ausencia de disolvente orgánico, por ejemplo, formamida, mientras que puede obtenerse hibridación de alta rigurosidad en presencia de al menos aproximadamente 35 % de formamida y más preferentemente al menos aproximadamente 50 % de formamida. Las condiciones de temperatura rigurosas habitualmente incluirán temperaturas de al menos aproximadamente 30°C , más preferentemente de al menos aproximadamente 37°C , y más preferentemente al menos aproximadamente 42°C con formamida presente. Se conocen bien por los expertos en la materia diversos parámetros adicionales, tales como tiempo de hibridación, la concentración de detergente, por ejemplo, dodecil sulfato sódico (SDS) y fuerza iónica. Se consiguen diversos niveles de rigurosidad combinando estas diversas condiciones según sea necesario.

Las etapas de lavado que siguen a la hibridación también pueden variar en rigurosidad; las etapas de lavado posthibridación principalmente determinan la especificidad de hibridación, siendo los factores más críticos la temperatura y la fuerza iónica de la solución de lavado final. La rigurosidad de lavado puede aumentarse reduciendo la concentración salina o aumentando la temperatura. La concentración salina rigurosa para las etapas de lavado será preferentemente menor de aproximadamente 30 mM de NaCl y 3 mM de citrato trisódico, y más preferentemente menor de aproximadamente 15 mM de NaCl y 1,5 mM de citrato trisódico.

Por lo tanto, las condiciones de hibridación y lavado que pueden usarse para unir y retirar polinucleótidos con menos de la homología deseada para las secuencias de ácido nucleico o sus complementos que codifican los presentes factores de transcripción incluyen, por ejemplo:

SSC 6X a 65 °C;
Formamida al 50 %, SSC 4X a 42 °C; o
SSC 0,5X, SDS 0,1 % a 65 °C;

con, por ejemplo, dos etapas de lavado de 10 - 30 minutos cada una. Resultarán fácilmente evidentes para los expertos en la materia variaciones útiles de estas condiciones.

Un experto en la materia no esperaría una variación sustancial entre especies de polinucleótidos abarcadas dentro del alcance de la presente invención debido a que las condiciones altamente rigurosas expuestas en las fórmulas anteriores producen polinucleótidos estructuralmente similares.

Si se desea, se pueden emplear etapas de lavado de aún mayor rigurosidad, incluyendo SSC aproximadamente 0,2 X, SDS 0,1 % a 65 °C y lavado dos veces, siendo cada etapa de lavado de aproximadamente 30 minutos, o SSC aproximadamente 0,1 x, SDS 0,1 % a 65 °C y lavado dos veces durante 30 minutos. La temperatura para las soluciones de lavado será habitualmente de al menos aproximadamente 25 °C, y para mayor rigurosidad al menos aproximadamente 42 °C. Pueden aumentarse la rigurosidad de hibridación adicionalmente usando las mismas condiciones que las etapas de hibridación, con la temperatura de lavado elevada de aproximadamente 3 °C a aproximadamente 5 °C, y la rigurosidad puede aumentarse aún más usando las mismas condiciones excepto que la temperatura de lavado se eleva de aproximadamente 6 °C a aproximadamente 9 °C. Para la identificación de homólogos menores estrechamente relacionados, pueden realizarse etapas de lavado a una temperatura menor; por ejemplo, 50 °C.

Un ejemplo de una etapa de lavado de baja rigurosidad emplea una solución y condiciones de al menos 25 °C en NaCl 30 mM, citrato trisódico 3 mM, y SDS 0,1 % durante 30 minutos. Puede obtenerse mayor rigurosidad a 42 °C en NaCl 15 mM, con citrato trisódico 1,5 mM y SDS 0,1 % durante 30 minutos. Se obtienen condiciones de lavado de rigurosidad aún mayor a 65 °C-68 °C en una solución de NaCl 15 mM, citrato trisódico 1,5 mM y SDS 0,1 %. Los procedimientos de lavado emplearán en general al menos dos etapas de lavado finales. Resultarán evidentes para los expertos en la materia variaciones adicionales de estas condiciones (véase, por ejemplo, Solicitud de Patente de Estados Unidos N° 20010010913).

Pueden seleccionarse condiciones de rigurosidad tales que un oligonucleótido que es perfectamente complementario del oligonucleótido codificante hibride con el oligonucleótido codificante con una relación de señal y ruido de al menos aproximadamente 5-10x mayor que la relación para hibridación del oligonucleótido perfectamente complementario y un ácido nucleico que codifica un factor de transcripción conocido en la fecha de presentación de la solicitud. Puede ser deseable seleccionar condiciones para un ensayo particular de modo que se obtenga una relación de señal y ruido superior, es decir, aproximadamente 15x o más. En consecuencia, un ácido nucleico objeto hibridará con un único oligonucleótido codificante con una relación de señal y ruido al menos 2x o mayor en comparación con hibridación del oligonucleótido codificante con un ácido nucleico que codifica un polipéptido conocido. La señal particular dependerá del marcador usado en el ensayo relevante, por ejemplo, un marcador fluorescente, un marcador colorimétrico, un marcador radiactivo o similares. Pueden producirse sondas de PCR o hibridación marcadas para detectar secuencias polinucleotídicas relacionadas mediante oligomarcage, traslación de muesca, marcate de extremos o amplificación por PCR usando un nucleótido marcado.

Están abarcadas por la invención secuencias polinucleotídicas que son capaces de hibridar con las secuencias polinucleotídicas reivindicadas, incluyendo cualquiera de los polinucleótidos del factor de transcripción dentro del Listado de Secuencias, y fragmentos de los mismos en diversas condiciones de rigurosidad (véase, por ejemplo, Wahl y Berger (1987) mencionado anteriormente, páginas 399-407; y Kimmel (1987) Methods Enzymol. 152: 507-511). Además de las secuencias de nucleótidos en el Listado de Secuencias, pueden identificarse y aislarse usando métodos conocidos ADNc de longitud completa, ortólogos y parálogos de las presentes secuencias de nucleótidos. Las bibliotecas, ortólogos y parálogos de ADNc de las presentes secuencias de nucleótidos pueden explorarse usando métodos de hibridación para determinar su utilidad como sondas de amplificación o diana de hibridación.

Identificación de polinucleótidos o ácidos nucleicos con bibliotecas de expresión

Además de los métodos de hibridación, pueden obtenerse polipéptidos homólogos de factores de transcripción explorando una biblioteca de expresión usando anticuerpos específicos para uno o más factores de transcripción. Con la provisión en el presente documento del factor de transcripción desvelado y secuencias de ácido nucleico homólogas del factor de transcripción, el polipéptido o los polipéptidos codificados pueden expresarse y purificarse en un sistema de expresión heterólogo (por ejemplo, *E. coli*) y usarse para inducir anticuerpos (monoclonales o policlonales) específicos para el polipéptido o los polipéptidos en cuestión. También pueden inducirse anticuerpos contra péptidos sintéticos derivados de secuencias de aminoácidos del factor de transcripción, u homólogo del factor

de transcripción. Se conocen bien en la técnica métodos para inducir anticuerpos y se describen en Harlow y Lane (1988), *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Nueva York. Dichos anticuerpos pueden usarse después para explorar una biblioteca de expresión producida a partir de la planta de la que se desea clonar homólogos del factor de transcripción adicionales, usando los métodos descritos anteriormente. Los ADNc seleccionados pueden confirmarse por secuenciación y actividad enzimática.

Variaciones de secuencia

Se apreciará fácilmente por los expertos en la materia que cualquiera de diversas secuencias polinucleotídicas son capaces de codificar los factores de transcripción descritos en el presente documento. Debido a la degeneración del código genético, muchos polinucleótidos diferentes pueden codificar polipéptidos idénticos y/o sustancialmente similares además de las secuencias ilustradas en el Listado de Secuencias. Los ácidos nucleicos que tienen una secuencia pueden diferir de las secuencias mostradas en el Listado de Secuencias, o secuencias complementarias, que codifican péptidos funcionalmente equivalentes (es decir, péptidos que tienen algún grado de actividad biológica equivalente o similar) pero difieren en secuencia de la secuencia mostrada en el Listado de Secuencias debido a la degeneración del código genético.

Las secuencias polinucleotídicas alteradas que codifican polipéptidos incluyen las secuencias con deleciones, inserciones o sustituciones de diferentes nucleótidos, dando como resultado un polinucleótido que codifica un polipéptido con al menos una característica funcional de los polipéptidos presentes. Se incluyen dentro de esta definición polimorfismos que pueden ser o no fácilmente detectables usando una sonda oligonucleotídica particular del polinucleótido que codifica los presentes polipéptidos, e hibridación inapropiada o inesperada de variantes alélicas, con un locus distinto del locus cromosómico normal para la secuencia polinucleotídica que codifica los presentes polipéptidos.

La variante alélicas se refiere a cualquiera de dos o más formas alternativas de un gen que ocupa el mismo locus cromosómico. La variación alélica surge de forma natural mediante la mutación, y puede dar como resultado polimorfismo fenotípico dentro de las poblaciones. Las mutaciones génicas pueden ser silenciosas (es decir, sin cambios en el polipéptido codificado) o pueden codificar polipéptidos que tienen secuencia de aminoácidos alterada. La expresión variante alélica también se usa en el presente documento para indicar una proteína codificada por una variante alélica de un gen. La variante de corte y empalme se refiere a formas alternativas de ARN transcrito de un gen. La variación de corte y empalme surge de forma natural mediante el uso de sitios de corte y empalme alternativos dentro de una molécula de ARN transcrita, o menos habitualmente entre moléculas de ARN transcritas por separado, y puede dar como resultado varios ARNm transcritos del mismo gen. Las variantes de corte y empalme pueden codificar polipéptidos que tienen secuencia de aminoácidos alterada. La expresión variante de corte y empalme también se usa en el presente documento para indicar una proteína codificada por una variante de corte y empalme de un ARNm transcrito de un gen.

Los expertos en la materia reconocerán que, por ejemplo, G1073, SEC ID N°: 2, representa un único factor de transcripción; puede esperarse que se produzca variación alélica y corte y empalme alternativo. Pueden clonarse variantes alélicas de SEC ID N°: 1 explorando ADNc o bibliotecas genómicas de diferentes organismos individuales de acuerdo con procedimientos convencionales. Las variantes alélicas de la secuencia de ADN mostrada en SEC ID N°: 1, incluyendo las que contienen mutaciones silenciosas y en las que las mutaciones dan como resultado cambios de secuencia de aminoácidos, están dentro del alcance de la presente divulgación, como lo están proteínas que son variantes alélicas de SEC ID N°: 2. Los ADNc generados a partir de ARNm de corte y empalme alternativo que conservan las propiedades del factor de transcripción están incluidos dentro del alcance de la presente divulgación, como lo están los polipéptidos codificados por dichos ADNc y ARNm. Pueden clonarse variantes alélicas y variantes de corte y empalme de estas secuencias explorando ADNc o bibliotecas genómicas a partir de organismos individuales o tejidos de acuerdo con procedimientos convencionales conocidos en la técnica (véase documento USPN 6.388.064).

Las moléculas de ácido nucleico relacionadas también incluyen secuencias de nucleótidos que codifican un polipéptido que comprende una sustitución, modificación, adición y/o delección de uno o más restos de aminoácidos. Dichos polipéptidos relacionados pueden comprender, por ejemplo, adiciones y/o deleciones de uno o más sitios de glucosilación ligados a N o ligados a O, o una adición y/o delección de uno o más restos de cisteína.

Por ejemplo, la Tabla 2 ilustra, por ejemplo, que los codones AGC, AGT, TCA, TCC, TCG y TCT codifican todos el mismo aminoácido: serina. En consecuencia, en cada posición en la secuencia en la que hay un codón que codifica serina, puede usarse cualquiera de las secuencias nucleotídicas anteriores sin alterar el polipéptido codificado.

Tabla 2

Aminoácido			Posibles codones					
Alanina	Ala	A	GCA	GCC	GCG	GCU	CGC	
Cisteína	Cys	C	TGC	TGT				
Ácido aspártico	Asp	D	GAC	GAT				
Ácido glutámico	Glu	E	GAA	GAG				
Fenilalanina	Phe	F	TTC	TTT				
Glicina	Gly	G	GGA	GGC	GGG	GGT		
Histidina	His	H	CAC	CAT				
Isoleucina	Ile	I	ATA	ATC	ATT			
Lisina	Lys	K	AAA	AAG				
Leucina	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT
Metionina	Met	M	ATG					
Asparagina	Asn	N	AAC	AAT				
Prolina	Pro	P	CCA	CCC	CCG	CCT		
Glutamina	Gln	Q	CAA	CAG				
Arginina	Arg	R	AGA	AGG	CGA		CGG	CGT
Serina	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT
Treonina	Thr	T	ACA	ACC	ACG	ACT		
Valina	Val	V	GTA	GTC	GTG	GTT		
Triptófano	Trp	W	TGG					
Tirosina	Tyr	Y	TAC	TAT				

Las alteraciones de secuencia que no cambian la secuencia de aminoácidos codificada por el polinucleótido se denominan variaciones “silenciosas”. Con la excepción de los codones ATG y TGG, que codifican metionina y triptófano, respectivamente, cualquiera de los posibles codones para el mismo aminoácido puede sustituirse por diversas técnicas, por ejemplo, mutagénesis dirigida, disponible en este campo.

Además de variaciones silenciosas, pueden realizarse otras variaciones conservativas que alteran uno, o algunos aminoácidos en el polipéptido codificado, sin alterar la función del polipéptido.

Por ejemplo, también se prevén sustituciones, deleciones e inserciones introducidas en las secuencias proporcionadas en el Listado de Secuencias. Dichas modificaciones de secuencia pueden introducirse por ingeniería genética en una secuencia por mutagénesis dirigida (Wu, editor; Methods Enzymol. (1993) vol. 217, Academic Press) o los otros métodos indicados posteriormente. Las sustituciones de aminoácidos son normalmente de restos individuales; las inserciones habitualmente serán en el orden de aproximadamente de 1 a 10 restos de aminoácidos; y las deleciones variarán de aproximadamente 1 a 30 restos. Preferentemente, se realizan deleciones o inserciones en pares adyacentes, por ejemplo, una delección de dos restos o inserción de dos restos. Pueden combinarse sustituciones, deleciones, inserciones o cualquier combinación de los mismos para llegar a una secuencia. Las mutaciones que se realizan en el polinucleótido que codifica el factor de transcripción no deberían colocar la secuencia fuera de la fase de lectura y no deberían crear regiones complementarias que pudieran producir una estructura de ARNm secundaria. Preferentemente, el polipéptido codificado por el ADN realiza la función deseada.

Las sustituciones conservativas son en las que al menos un resto en la secuencia de aminoácidos se ha retirado y se ha insertado un resto diferente en su lugar. Dichas sustituciones generalmente se realizan de acuerdo con la Tabla 3 cuando se desea mantener la actividad de la proteína. La Tabla 3 muestra aminoácidos que pueden sustituir a un aminoácido en una proteína y que normalmente se consideran sustituciones conservativas. Los factores de transcripción enumerados en el Listado de Secuencias pueden tener hasta 10 sustituciones conservativas y mantener su función. Los factores de transcripción enumerados en el Listado de Secuencias pueden tener más de 10 sustituciones conservativas y aún mantener su función.

Tabla 3

Resto	Sustituciones Conservativas
Ala	Ser
Arg	Lys
Asn	Gln; His

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

- 5 Son sustituciones similares en la que al menos un resto en la secuencia de aminoácidos se ha retirado y se ha insertado un resto diferente en su lugar. Dichas sustituciones se realizan generalmente de acuerdo con la Tabla 4 cuando se desea mantener la actividad de la proteína. La Tabla 4 muestra aminoácidos que pueden sustituir a un aminoácido en una proteína y que normalmente se consideran sustituciones estructurales y funcionales. Por ejemplo, un resto en la columna 1 de la Tabla 4 puede sustituirse con un resto en la columna 2; además, un resto en la columna 2 de la Tabla 4 puede sustituirse con el resto de la columna 1.

Tabla 4

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

Trp	Tyr; Phe; His
Tyr	Trp; Phe; His
Val	Ala; Ile; Leu; Gly; Thr; Ser; Glu

Pueden seleccionarse sustituciones que sean menos conservativas que las de la Tabla 4 seleccionando restos que difieren más significativamente en su efecto en el mantenimiento de (a) la estructura de la cadena principal polipeptídica en el área de la sustitución, por ejemplo como una conformación en lámina o helicoidal, (b) la carga o hidrofobicidad de la molécula en el sitio diana, o (c) el volumen de la cadena lateral. Las sustituciones que en general se espera que produzcan los mayores cambios en las propiedades de la proteína serán en las que (a) un resto hidrófilo, por ejemplo, serilo o treonilo, sustituye a (o se sustituye por) un resto hidrófobo, por ejemplo, leucilo, isoleucilo, fenilalanilo, valilo o alanilo; (b) una cisteína o prolina sustituye a (o se sustituye por) cualquier otro resto; (c) un resto que tiene una cadena lateral electropositiva, por ejemplo, lisilo, arginilo o histidilo, sustituye a (o se sustituye por) un resto electronegativo, por ejemplo, glutamilo o aspartilo; o (d) un resto que tiene una cadena lateral voluminosa, por ejemplo, fenilalanina, sustituye a (o se sustituye por) uno que no tiene una cadena lateral, por ejemplo, glicina.

Modificación adicional de secuencias por mutación/evolución forzada

Además de generar sustituciones silenciosas o conservativas como se ha observado anteriormente, pueden usarse métodos de modificación de ácidos nucleicos o proteínas para alterar secuencias dadas para producir nuevas secuencias y/o para modificar química o enzimáticamente secuencias dadas para cambiar las propiedades de los ácidos nucleicos o proteínas.

Por lo tanto, pueden modificarse secuencias de ácido nucleico dadas, por ejemplo, de acuerdo con mutagénesis convencional o métodos de evolución artificiales para producir secuencias modificadas. Las secuencias modificadas pueden crearse usando polinucleótidos naturales purificados aislados de cualquier organismo o pueden sintetizarse a partir de composiciones purificadas y productos químicos usando medios químicos bien conocidos por los expertos en la materia. Por ejemplo, Ausubel (1997 y 2000; mencionado anteriormente), proporciona detalles adicionales sobre métodos de mutagénesis. Se describen métodos de evolución forzada artificiales, por ejemplo, en Stemmer (1994; Nature 370: 389-391), Stemmer (1994; Proc. Natl. Acad. Sci. USA 91: 10747-10751) y las Patentes de Estados Unidos 5.811.238, 5.837.500 y 6.242.568. Se describen métodos para modificar por ingeniería genética factores de transcripción sintéticos y otros polipéptidos, por ejemplo, en Zhang *et al.* (2000) J. Biol. Chem. 275: 33850-33860, Liu *et al.* (2001) J. Biol. Chem. 276: 11323-11334, y Isalan *et al.* (2001) Nature Biotechnol. 19: 656-660. También están disponibles muchos otros métodos de mutación y evolución y se espera que estén dentro de la experiencia del experto en la materia.

De forma similar, puede realizarse una alteración química o enzimática de ácidos nucleicos y polipéptidos expresados por métodos convencionales. Por ejemplo, puede modificarse la secuencia mediante adición de lípidos, azúcares, péptidos, compuestos orgánicos o inorgánicos, mediante la inclusión de nucleótidos o aminoácidos modificados, o similares. Por ejemplo, se ilustran técnicas de modificación de proteínas en Ausubel (1997 y 2000; mencionado anteriormente). Pueden encontrarse en el presente documento detalles adicionales sobre modificaciones químicas y enzimáticas. Estos métodos de modificación pueden usarse para modificar cualquier secuencia dada, o para modificar cualquier secuencia producida por los diversos métodos de modificación por mutación y evolución artificial indicados en el presente documento.

En consecuencia, la modificación de cualquier ácido nucleico dado puede ser por mutación, evolución, modificación química o enzimática, u otros métodos disponibles, por ejemplo, usando las secuencias del presente documento como un sustrato de partida para los diversos enfoques de modificación.

Por ejemplo, puede usarse una secuencia codificante optimizada que contiene codones preferidos por un hospedador procariota o eucariota particular, por ejemplo, para aumentar la tasa de traducción o para producir transcritos de ARN recombinantes que tienen propiedades deseables, tales como mayor semivida, en comparación con transcritos producidos usando una secuencia no optimizada. También pueden modificarse codones de parada de la traducción para reflejar la preferencia del hospedador. Por ejemplo, los codones de parada preferidos para *Saccharomyces cerevisiae* y mamíferos son TAA y TGA, respectivamente. El codón de parada preferido para plantas monocotiledóneas es TGA, mientras que los insectos y *E. coli* prefieren usar TAA como el codón de parada.

Las secuencias polinucleotídicas pueden también modificarse por ingeniería genética para alterar una secuencia codificante por diversas razones, incluyendo pero sin limitación, alteraciones que modifican la secuencia para facilitar la clonación, el procesamiento y/o la expresión del producto génico. Por ejemplo, se introducen opcionalmente alteraciones usando técnicas que se conocen bien en este campo, por ejemplo, mutagénesis dirigida, para insertar nuevos sitios de restricción, para alterar los patrones de glucosilación, para cambiar la preferencia codónica, para introducir sitios de corte y empalme, etc.

Además, puede combinarse un fragmento o dominio derivado de cualquiera de los polipéptidos con dominios derivados de otros factores de transcripción o dominios sintéticos para modificar la actividad biológica de un factor de transcripción. Por ejemplo, puede combinarse un dominio de unión a ADN derivado de un factor de transcripción descrito en el presente documento con el dominio de activación de otro factor de transcripción o con un dominio de activación sintético. Un dominio de activación de la transcripción ayuda en el inicio de la transcripción de un sitio de unión a ADN. Los ejemplos incluyen la región de activación de la transcripción de VP 16 o GAL4 (Moore *et al.* (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; Aoyama *et al.* (1995) Plant Cell 7: 1773-1785), péptidos derivados de secuencias bacterianas (Ma y Ptashne (1987) Cell 51: 113-119) y péptidos sintéticos (Giniger y Ptashne (1987) Nature 330: 670-672).

Expresión y modificación de polipéptidos

Normalmente, se incorporan secuencias polinucleotídicas en moléculas de ADN (o ARN) recombinantes que dirigen la expresión de polipéptidos en células hospedadoras apropiadas, plantas transgénicas, sistemas de traducción *in vitro* o similares. Debido a la degeneración inherente del código genético, las secuencias de ácido nucleico que codifican sustancialmente la misma secuencia de aminoácidos o una funcionalmente equivalente pueden sustituir a cualquier secuencia enumerada para posibilitar la clonación y expresión del homólogo relevante.

Las plantas transgénicas de la presente invención que comprenden secuencias polinucleotídicas recombinantes generalmente derivan de plantas parentales, que pueden en sí mismas ser plantas no transformadas (o no transgénicas). Estas plantas transgénicas pueden tener un gen de factor de transcripción "anulado" (por ejemplo, con una inserción genómica por recombinación homóloga, una construcción antisentido o de ribozima) o expresado en un grado normal o de tipo silvestre. Sin embargo, las plantas "descendientes" transgénicas con sobreexpresión mostrarán mayores niveles de ARNm, codificando el ARNm un factor de transcripción, que es una proteína de unión a ADN que es capaz de unirse con una secuencia reguladora de ADN e inducir la transcripción, y preferentemente, expresión de un gen de rasgo vegetal. Preferentemente, el nivel de expresión de ARNm será al menos tres veces mayor que el de la planta parental, o más preferentemente niveles de ARNm al menos diez veces mayores en comparación con dicha planta parental, y más preferentemente al menos cincuenta veces mayores en comparación con dicha planta parental.

Vectores, promotores y sistemas de expresión

La presente divulgación incluye construcciones recombinantes que comprenden una o más de las secuencias de ácido nucleico del presente documento. Las construcciones normalmente comprende un vector, tal como un plásmido, un cósmido, un fago, un virus (por ejemplo, un virus vegetal), un cromosoma artificial bacteriano (BAC), un cromosoma artificial de levadura (YAC), o similares, en el que se ha insertado una secuencia de ácido nucleico de la invención, en orientación directa o inversa. Preferentemente, la construcción comprende además secuencias reguladoras, incluyendo, por ejemplo, un promotor, unido operativamente con la secuencia. Se conocen por los expertos en la materia grandes números de vectores y promotores adecuados, y están disponibles en el mercado.

Los textos generales que describen técnicas de biología molecular útiles en el presente documento, incluyendo el uso y producción de vectores, promotores y muchos otros asuntos relevantes, incluyen Berger (1987) mencionado anteriormente, Sambrook (1989), mencionado anteriormente, y Ausubel (hasta 2000) mencionado anteriormente. Puede incorporarse cualquiera de las secuencias identificadas en un casete o vector, por ejemplo, para la expresión en plantas. Se han descrito varios vectores de expresión adecuados para transformación estable de células vegetales o para el establecimiento de plantas transgénicas incluyendo las descritas en Weissbach y Weissbach (1989) Methods for Plant Molecular Biology, Academic Press, y Gelvin *et al.* (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Los ejemplos específicos incluyen los derivados de un plásmido Ti de *Agrobacterium tumefaciens*, así como los desvelados por Herrera-Estrella *et al.* (1983) Nature 303: 209, Bevan (1984) Nucleic Acids Res. 12: 8711-8721, Klee (1985) Biotechnology 3: 637-642, para plantas dicotiledóneas.

Como alternativa, pueden usarse vectores no Ti para transferir el ADN a plantas monocotiledóneas y células usando técnicas de suministro de ADN libre. Dichos métodos pueden implicar, por ejemplo, el uso de liposomas, electroporación, bombardeo con microproyectiles, filamentos de carburo de silicio y virus. Usando estos métodos pueden producirse plantas transgénicas tales como trigo, arroz (Christou (1991) Bio/Technology 9: 957-962) y maíz (Gordon-Kamm (1990) Plant Cell 2: 603-618). Un embrión inmaduro también puede ser un buen tejido diana para monocotiledóneas para técnicas de suministro de ADN directo usando la pistola de partículas (Weeks *et al.* (1993) Plant Physiol. 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan y Lemeaux (1994) Plant Physiol. 104: 37-48, y para transferencia de ADN mediada por *Agrobacterium* (Ishida *et al.* (1996) Nature Biotechnol. 14: 745-750).

Normalmente, los vectores de transformación de plantas incluyen una o más secuencias codificantes vegetales clonadas (genómicas o de ADNc) en el control de transcripción de secuencias reguladoras 5' y 3' y un marcador seleccionado dominante. Dichos vectores de transformación de plantas normalmente contienen también un promotor (por ejemplo, una región reguladora que controla la expresión inducible o constitutiva, regulada por ambiente o desarrollo, o específica de célula o tejido), un sitio de partida de inicio de la transcripción, una señal de procesamiento de ARN (tal como sitios de corte y empalme de intrones), un sitio de terminación de la transcripción

y/o una señal de poliadenilación.

Una utilidad potencial para los polinucleótidos del factor de transcripción desvelados en el presente documento es el aislamiento de elementos promotores de estos genes que pueden usarse para programar la expresión en plantas de cualquier gen. Cada gen de factor de transcripción desvelado en el presente documento se expresa de una manera única, como se determina por elementos promotores localizados cadena arriba del inicio de la traducción, y adicionalmente dentro de un intrón del gen del factor de transcripción o cadena abajo del codón de terminación del gen. Como se conoce bien en la técnica, para una parte significativa de los genes, las secuencias promotores se localizan completamente en la región directamente cadena arriba del inicio de la traducción. En dichos casos, normalmente las secuencias promotoras se localizan a una distancia de 2,0 kb del inicio de la traducción, o a una distancia de 1,5 kb del inicio de la traducción, frecuentemente a una distancia de 1,0 kb del inicio de la traducción, y en ocasiones a una distancia de 0,5 kb del inicio de la traducción.

Las secuencias promotoras pueden aislarse de acuerdo con métodos conocidos por los expertos en la materia.

Los ejemplos de promotores vegetales constitutivos que pueden ser útiles para expresar la secuencia de factor de transcripción incluyen: el promotor 35S del virus del mosaico de la coliflor (CaMV), que confiere expresión de alto nivel, constitutiva, en la mayoría de los tejidos vegetales (véase, por ejemplo, Odell *et al.* (1985) *Nature* 313: 810-812); el promotor de nopalina sintasa (An *et al.* (1988) *Plant Physiol.* 88: 547-552); y el promotor de octopina sintasa (Fromm *et al.* (1989) *Plant Cell* 1: 977-984).

Los factores de transcripción relacionados con la presente invención pueden unirse operativamente con un promotor específico que provoca que el factor de transcripción se exprese en respuesta a señales ambientales, específicas de tejido o temporales. Se conoce diversos promotores de genes vegetales que regulan la expresión génica en respuesta a señales ambientales, hormonales, químicas, del desarrollo y de una manera activa en tejido; muchos de estos pueden usarse para expresión de una secuencia del factor de transcripción en plantas. La elección de un promotor se basa en gran medida en el fenotipo de interés y se determina por factores tales como el tejido (por ejemplo, semilla, fruto, raíz, polen, tejido vascular, flor, carpelo, etc.), capacidad de inducción (por ejemplo, en respuesta a heridas, calor, frío, sequía, luz, patógenos, etc.), temporización, estadio del desarrollo y similares. Se han caracterizado numerosos promotores conocidos y pueden emplearse de forma favorable para promover la expresión de un polinucleótido en una planta transgénica o célula de interés. Por ejemplo, los promotores específicos de tejido incluyen: promotores específicos de semilla (tales como la napina, faseolina o promotor DC3 descritos en la Patente de Estados Unidos N° 5.773.697), promotores específicos de fruto que están activos durante la maduración del fruto (tales como el promotor *dru 1* (Patente de Estados Unidos N° 5.783.393) o el promotor 2A11 (Patente de Estados Unidos N° 4.943.674) y el promotor de poligalacturonasa del tomate (Bird *et al.* (1988) *Plant Mol. Biol.* 11: 651-662), promotores específicos de raíz, tales como ARSK1, y los desvelados en las Patentes de Estados Unidos N° 5.618.988, 5.837.848 y 5.905.186, promotores específicos de epidermis, incluyendo CUT1 (Kunst *et al.* (1999) *Biochem. Soc. Trans.* 28: 651-654), promotores activos en polen tales como *PTA29*, *PTA26* y *PTA13* (Patente de Estados Unidos N° 5.792.929), promotores activos en tejido vascular (Ringli y Keller (1998) *Plant Mol. Biol.* 37: 977-988), específicos de flor (Kaiser *et al.* (1995) *Plant Mol. Biol.* 28: 231-243), polen (Baerson *et al.* (1994) *Plant Mol. Biol.* 26: 1947-1959), carpelos (Ohl *et al.* (1990) *Plant Cell* 2: 837-848), polen y óvulos (Baerson *et al.* (1993) *Plant Mol. Biol.* 22: 255-267), promotores inducibles por auxina (tales como los descritos en van der Kop *et al.* (1999) *Plant Mol. Biol.* 39: 979-990 o Baumann *et al.* (1999) *Plant Cell* 11: 323-334), promotor inducible por citoquinina (Guevara-García (1998) *Plant Mol. Biol.* 38: 743-753), promotores sensibles a giberelina (Shi *et al.* (1998) *Plant Mol. Biol.* 38: 1053-1060, Willmott *et al.* (1998) *Plant Mol. Biol.* 38: 817-825) y similares. Son promotores adicionales los que inducen expresión en respuesta a calor (Ainley *et al.* (1993) *Plant Mol. Biol.* 22: 13-23), luz (por ejemplo, el promotor *rbcS-3A* del guisante, Kuhlemeier *et al.* (1989) *Plant Cell* 1: 471-478, y el promotor *rbcS* del maíz, Schaffner y Sheen (1991) *Plant Cell* 3: 997-1012); heridas (por ejemplo, wunl, Siebertz *et al.* (1989) *Plant Cell* 1: 961-968); patógenos (tales como el promotor PR-1 descrito en Buchel *et al.* (1999) *Plant Mol. Biol.* 40: 387-396, y el promotor PDF1.2 descrito en Manners *et al.* (1998) *Plant Mol. Biol.* 38: 1071-1080), y productos químicos tales como metil jasmonato o ácido salicílico (Gatz (1997) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48: 89-108). Además, el momento de la expresión puede controlarse usando promotores tales como los que actúan en la senescencia (Gan y Amasino (1995) *Science* 270: 1986-1988); o desarrollo de semillas tardío (Odell *et al.* (1994) *Plant Physiol.* 106: 447-458).

Los vectores de expresión de plantas también pueden incluir señales de procesamiento de ARN que pueden situarse dentro, cadena arriba o cadena abajo de la secuencia codificante. Además, los vectores de expresión pueden incluir las secuencias reguladoras adicionales de la región 3' no traducida de genes vegetales, por ejemplo, una región terminadora 3' para aumentar la estabilidad de ARNm del ARNm, tal como la región terminadora PI-II de la patata o las regiones terminadoras 3' de octopina o nopalina sintasa.

Elementos de expresión adicionales

Las señales de inicio específicas pueden ayudar a la traducción eficaz de secuencias codificantes. Estas señales pueden incluir, por ejemplo, el codón de inicio ATG y secuencias adyacentes. En casos en los que una secuencia codificante, su codón de inicio y secuencias cadena arriba se insertan en el vector de expresión apropiado, pueden

no ser necesarias señales de control de la traducción adicionales. Sin embargo, en casos en los que solamente se inserta una secuencia codificante (por ejemplo, una secuencia codificante de proteína madura), o una parte de la misma, pueden proporcionarse por separado señales de control de la transcripción exógenas que incluyen el codón de inicio ATG. El codón de inicio se proporciona en la fase de lectura correcta para facilitar la transcripción. Los elementos de transcripción exógenos y codones de inicio pueden ser de diversos orígenes, tanto naturales como sintéticos. La eficacia de la expresión puede potenciarse por la inclusión de potenciadores apropiados para el sistema celular que se use.

Hospedadores de expresión

La presente divulgación también se refiere a células hospedadoras que se transducen con vectores, y la producción de polipéptidos (incluyendo fragmentos de los mismos) por técnicas recombinantes. Las células hospedadoras se modifican por ingeniería genética (es decir, se introducen ácidos nucleicos, por ejemplo, se transducen, se transforman o se transfectan) con los vectores, que pueden ser, por ejemplo, un vector de clonación o un vector de expresión que comprende los ácidos nucleicos relevantes en el presente documento. El vector es opcionalmente un plásmido, una partícula viral, un fago, un ácido nucleico desnudo, etc. Las células hospedadoras modificadas por ingeniería genética pueden cultivarse en medios nutrientes convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes o amplificar el gen relevante. Las condiciones de cultivo, tales como temperatura, pH y similares, son las previamente usadas con la célula hospedadora seleccionada para expresión, y resultarán evidentes para los expertos en la materia y las referencias citadas en el presente documento, incluyendo Sambrook (1989) mencionado anteriormente, y Ausubel (hasta 2000) mencionado anteriormente.

La célula hospedadora puede ser una célula eucariota, tal como una célula de levadura, o una célula vegetal, o la célula hospedadora puede ser una célula procariota, tal como una célula bacteriana. También son adecuados para algunas aplicaciones protoplastos vegetales. Por ejemplo, los fragmentos de ADN se introducen en tejidos vegetales, células vegetales cultivadas o protoplastos vegetales por métodos convencionales incluyendo electroporación (Fromm *et al.* (1985) Proc. Natl. Acad. Sci. USA 82: 5824-5828), infección por vectores virales tales como virus del mosaico de la coliflor (CaMV) (Hohn *et al.* (1982) Molecular Biology of Plant Tumors, Academic Press, Nueva York, NY, pp. 549-560; documento US 4.407.956), penetración balística de alta velocidad por partículas pequeñas con el ácido nucleico bien dentro de la matriz de perlas pequeñas o partículas, o bien en la superficie (Kleun *et al.* (1987) Nature 327: 70-73), uso de polen como un vector (documento WO 85/01856), o uso de *Agrobacterium tumefaciens* o *A. rhizogenes* que porta un plásmido de ADN T en el que se clonan fragmentos de ADN. El plásmido de ADN-T se transmite a células vegetales tras su infección por *Agrobacterium tumefaciens*, y una parte se integra de forma estable en el genoma vegetal (Horsch *et al.* (1984) Science 233: 496-498; Fraley *et al.* (1983) Proc. Natl. Acad. Sci. USA 80: 4803-4807).

La célula puede incluir un ácido nucleico que codifica un polipéptido, en el que la célula expresa un polipéptido. La célula también puede incluir secuencias de vectores, o similares. Además, las células y plantas transgénicas pueden incluir cualquier polipéptido o ácido nucleico anterior o de la presente memoria descriptiva, por ejemplo, producido por transducción de un vector descrito en el presente documento.

Para producción de alto rendimiento, a largo plazo, de proteínas recombinantes, puede usarse expresión estable. Las células hospedadoras transformadas con una secuencia de nucleótidos que codifica un polipéptido de la invención se cultivan opcionalmente en condiciones adecuadas para la expresión y recuperación de la proteína codificada del cultivo celular. La proteína o fragmento de la misma producido por una célula recombinante puede secretarse, unirse a membrana o estar contenida de forma intracelular, dependiendo de la secuencia y/o el vector usado. Como se entenderá por los expertos en la materia, los vectores de expresión que contienen polinucleótidos que codifican proteínas maduras pueden diseñarse con secuencias señal que dirigen la secreción de los polipéptidos maduros a través de una membrana celular procariota o eucariota.

Restos de aminoácidos modificados

Los polipéptidos pueden contener uno o más restos de aminoácidos modificados. La presencia de aminoácidos modificados puede ser ventajosa en, por ejemplo, el aumento de la semivida del polipéptido, la reducción de la antigenicidad o toxicidad del polipéptido, el aumento de la estabilidad en el almacenamiento del polipéptido, o similares. Se modifican un resto o restos de aminoácidos, por ejemplo, de forma cotraduccional o postraduccional durante la producción recombinante o se modifican por medios sintéticos o químicos.

Los ejemplos no limitantes de un resto de aminoácido modificado incluyen la incorporación u otro uso de aminoácidos acetilados, aminoácidos glucosilados, aminoácidos sulfatados, aminoácidos prenilados (por ejemplo, farnesilados, geranilgeranilados), aminoácidos modificados con PEG (por ejemplo, "PEGilados"), aminoácidos biotinilados, aminoácidos carboxilados, aminoácidos fosforilados, etc. Toda la bibliografía está repleta de referencias adecuadas para guiar a un experto en la modificación de los restos de aminoácidos.

Los restos de aminoácidos modificados pueden prevenir o aumentar la afinidad del polipéptido por otra molécula, incluyendo, pero sin limitación, polinucleótido, proteínas, carbohidratos, lípidos y derivados de lípidos, y otros

compuestos orgánicos o sintéticos.

Identificación de factores proteicos adicionales

- 5 También puede usarse un factor de transcripción relacionado con la presente invención para identificar moléculas endógenas o exógenas adicionales que pueden afectar a un fenotipo o rasgo de interés. Dichas moléculas incluyen moléculas endógenas sobre las que se actúa a un nivel transcripcional por un factor de transcripción para modificar un fenotipo según se desee. Por ejemplo, los factores de transcripción pueden emplearse para identificar uno o más genes cadena abajo que están sujetos a un efecto regulador del factor de transcripción. En un enfoque, un factor de transcripción u homólogo del factor de transcripción se expresa en una célula hospedadora, por ejemplo, una célula vegetal transgénica, tejido o explante, y se controlan los productos de expresión, bien ARN o bien proteína, de dianas probables o aleatorias, por ejemplo, mediante hibridación con una micromatriz de sondas de ácido nucleico correspondientes a genes expresados en un tejido o tipo celular de interés, por electroforesis en gel bidimensional de productos proteicos, o por cualquier otro método conocido en la técnica para evaluar la expresión de productos génicos al nivel de ARN o proteína. Como alternativa, puede usarse un factor de transcripción para identificar secuencias promotoras (tales como sitios de unión en secuencias de ADN) implicadas en la regulación de una diana cadena abajo. Después de identificar una secuencia promotora, pueden modificarse interacciones entre el factor de transcripción y la secuencia promotora cambiando nucleótidos específicos en la secuencia promotora o aminoácidos específicos en el factor de transcripción que interaccionan con la secuencia promotora para alterar un rasgo de la planta. Normalmente, se identifican sitios de unión a ADN del factor de transcripción por ensayos de desplazamiento en gel. Después de identificar las regiones promotoras, las secuencias de región promotora pueden emplearse en matrices de ADN bicatenario para identificar moléculas que afectan a las interacciones de los factores de transcripción con sus promotores (Bulyk *et al.* (1999) *Nature Biotechnol.* 17: 573-577).
- 25 Los factores de transcripción identificados son también útiles para identificar proteínas que modifican la actividad del factor de transcripción. Dicha modificación puede producirse por modificación covalente, tal como por fosforilación, o por interacciones proteína-proteína (homo o heteropolímero). Puede emplearse cualquier método adecuado para detectar interacciones proteína-proteína. Entre los métodos que pueden emplearse están la coimmunoprecipitación, reticulación y copurificación mediante gradientes o columnas cromatográficas, y el sistema de levadura de dos híbridos.

El sistema de dos híbridos detecta interacciones de proteínas *in vivo* y se ha descrito previamente (Chien *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88: 9578-9582), y está disponibles en el mercado de Clontech (Palo Alto, Calif.). En dicho sistema, se construyen plásmidos que codifican dos proteínas híbridas: una consiste en el dominio de unión a ADN de una proteína activador de transcripción fusionada con el polipéptido del factor de transcripción y la otra consiste en el dominio de activación de la proteína activadora de transcripción fusionado con una proteína desconocida que se codifica por un ADNc que se ha recombinado en el plásmido como parte de una biblioteca de ADNc. El plásmido de fusión del dominio de unión a ADN y la biblioteca de ADNc se transforman en una cepa de la levadura *Saccharomyces cerevisiae* que contiene un gen indicador (por ejemplo, lacZ) cuya región reguladora contiene el sitio de unión del activador de la transcripción. Una de las proteínas híbridas por sí sola no puede activar la transcripción del gen indicador. La interacción de las dos proteínas híbridas reconstituye la proteína activadora funcional y da como resultado la expresión del gen indicador, que se detecta por un ensayo para el producto génico indicador. Después, los plásmidos de la biblioteca responsables de la expresión del gen indicador se aíslan y secuencian para identificar las proteínas codificadas por los plásmidos de la biblioteca. Después de identificar proteínas que interaccionan con los factores de transcripción, pueden realizarse ensayos con respecto a compuestos que interfieren con las interacciones proteína-proteína del factor de transcripción.

Subsecuencias

- 50 También se contemplan usos de polinucleótidos, también denominados en el presente documento oligonucleótidos, que tienen normalmente al menos 12 bases, preferentemente al menos 50 bases, que hibridan en condiciones rigurosas con una secuencia polinucleotídica descrita anteriormente. Los polinucleótidos pueden usarse como sondas, cebadores, agentes con sentido y antisentido, y similares, de acuerdo con métodos como se han observado anteriormente.
- 55 Subsecuencias de polinucleótidos, incluyendo fragmentos polinucleotídicos y oligonucleótidos son útiles como sondas y cebadores de ácidos nucleicos. Un oligonucleótido adecuado para su uso como una sonda o cebador es de al menos aproximadamente 15 nucleótidos de longitud, con más frecuencia de al menos aproximadamente 18 nucleótidos, frecuentemente de al menos aproximadamente 21 nucleótidos, frecuentemente de al menos aproximadamente 30 nucleótidos, o de aproximadamente 40 nucleótidos, o más de longitud. Una sonda de ácido nucleico es útil en protocolos de hibridación, por ejemplo, para identificar homólogos polipeptídicos adicionales, incluyendo protocolos para experimentos de micromatrices. Los cebadores pueden hibridarse con una cadena de ADN diana complementaria por hibridación de ácido nucleico para formar un híbrido entre el cebador y la cadena de ADN diana, y después extenderse a lo largo de la cadena de ADN diana por una enzima ADN polimerasa. Pueden usarse pares de cebadores para amplificación de una secuencia de ácido nucleico, por ejemplo, mediante la reacción en cadena de la polimerasa (PCR) u otros métodos de amplificación de ácido nucleico (Sambrook (1989)

5 mencionado anteriormente y Ausubel (hasta 2000) mencionado anteriormente).

Además, puede usarse como inmunógenos un polipéptido aislado o recombinante que incluye una subsecuencia de al menos aproximadamente 15 aminoácidos contiguos codificados por los polinucleótidos recombinantes o aislados, o dominios o fragmentos de los mismos, por ejemplo, para producir anticuerpos específicos para la secuencia polipeptídica, o como sondas para detectar una secuencia de interés. Una subsecuencia puede variar de tamaño de aproximadamente 15 aminoácidos de longitud hasta e incluyendo la longitud completa del polipéptido.

Un polipéptido expresado que comprende dicha subsecuencia polipeptídica realiza al menos una función biológica del polipéptido intacto de una manera sustancialmente igual, o en un grado similar, que el polipéptido intacto. Por ejemplo, un fragmento polipeptídico puede comprender un motivo estructural reconocible o dominio funcional tal como un dominio de unión a ADN que activa la transcripción, por ejemplo, uniendo con una región promotora de ADN específica un dominio de activación, o un dominio para interacciones proteína-proteína.

15 Producción de plantas transgénicas

Modificación de rasgos

Los polinucleótidos descritos en el presente documento se emplean favorablemente para producir plantas transgénicas con diversos rasgos, o características, que se han modificado de una manera deseable, por ejemplo, para mejorar las características de semillas de una planta. Por ejemplo, puede usarse la alteración de los niveles o patrones de expresión (por ejemplo, patrones de expresión espacial o temporal) de uno o más de los factores de transcripción (u homólogos de los factores de transcripción), en comparación con los niveles de la misma proteína hallada en una planta de tipo silvestre, para modificar los rasgos de una planta. Un ejemplo ilustrativo de modificación de rasgos, características mejoradas, alterando los niveles de expresión de un factor de transcripción particular se describe adicionalmente en los ejemplos y el Listado de Secuencias.

Arabidopsis como un sistema modelo

La *Arabidopsis thaliana* es un objeto de atención rápidamente creciente como un modelo para genética y metabolismo en plantas. La *Arabidopsis* tiene un genoma pequeño, y están disponibles estudios bien documentados. Es fácil de cultivar en grandes números y están disponibles, o pueden obtenerse fácilmente, mutantes que definen mecanismos controlados genéticamente importantes. Están disponibles diversos métodos para introducir y expresar genes homólogos aislados (véase Koncz *et al.*, editores, *Methods in Arabidopsis Research* (1992) World Scientific, Nueva Jersey NJ, en "Prefacio"). Debido a su pequeño tamaño, ciclo de vida corto, autogamia obligada y alta fertilidad, la *Arabidopsis* también es un organismo elegido para el aislamiento de mutantes y estudios en las rutas morfogénicas y del desarrollo, y el control de estas rutas por factores de transcripción (Koncz (1992) mencionado anteriormente, p. 72). Varios estudios de introducción de factores de transcripción en *A. thaliana* han demostrado la utilidad de esta planta para entender los mecanismos de la regulación génica y la alteración de rasgos en plantas (véase, por ejemplo, factores (Koncz (1992) mencionado anteriormente, y Patente de Estados Unidos Número 6.417.428)).

Genes de Arabidopsis en plantas transgénicas

La expresión de genes que codifican factores de transcripción que modifican la expresión de genes endógenos, polinucleótidos y proteínas se conoce bien en la técnica. Además las plantas transgénicas que comprenden polinucleótidos aislados que codifican factores de transcripción también pueden modificar la expresión de genes endógenos, polinucleótidos y proteínas. Los ejemplos incluyen Peng *et al.* (1997) *et al.* *Genes and Development* 11: 3194-3205, y Peng *et al.* (1999) *Nature* 400: 256-261. Además, muchos otros han demostrado que un factor de transcripción de *Arabidopsis* expresado en una especie vegetal exógena induce la misma respuesta fenotípica o una muy similar. Véase, por ejemplo, Fu *et al.* (2001) *Plant Cell* 13: 1791-1802; Nandi *et al.* (2000) *Curr. Biol.* 10: 215-218; Coupland (1995) *Nature* 377: 482-483; y Weigel y Nilsson (1995) *Nature* 377: 482-500.

Genes homólogos introducidos en plantas transgénicas

Pueden introducirse genes homólogos que pueden derivar de cualquier planta, o de cualquier fuente, bien natural, sintética, semisintética o recombinante, y que comparten identidad o similitud de secuencia significativa con los descritos en el presente documento, en plantas, por ejemplo, plantas de cultivo, para conferir rasgos deseables o mejorados. En consecuencia, pueden producirse plantas transgénicas que comprenden un vector o casete de expresión recombinante con un promotor unido operativamente con una o más secuencias homólogas de las secuencias desveladas en el presente documento. El promotor puede ser, por ejemplo, un promotor vegetal o viral.

La invención proporciona por lo tanto métodos para preparar plantas transgénicas, y para modificar rasgos vegetales, como se define por las reivindicaciones. Las plantas y partes vegetales que resultan de la aplicación de estos métodos también están abarcadas por la presente invención, como se define por las reivindicaciones.

Factores de transcripción de interés para la modificación de rasgos vegetales

En consecuencia, la existencia de una serie de grupos de madurez para diferentes latitudes representa una barrera importante para la introducción de nuevos rasgos valiosos. Cualquier rasgo (por ejemplo tolerancia al estrés abiótico o biomasa aumentada) tiene que cultivarse en cada uno de los grupos de madurez diferentes por separado, un ejercicio laborioso y costoso. La disponibilidad de cepa única, que podría cultivarse en cualquier latitud, aumentaría por lo tanto en gran medida el potencial de introducir nuevos rasgos en especies de cultivo tales como la soja y el algodón.

Para los efectos específicos, rasgos y utilidades conferidos a plantas, pueden usarse uno o más genes de factores de transcripción de la presente invención para aumentar o reducir, o mejorar o ser deletéreos para un rasgo dado. Por ejemplo, la anulación de un gen de factor de transcripción que aparece de forma natural en una planta, o supresión del gen (con, por ejemplo, supresión antisentido), pueden provocar tolerancia reducida a un estrés osmótico en relación con plantas no transformadas o de tipo silvestre. Sobreexpresando este gen, la planta puede experimentar tolerancia aumentada al mismo estrés. Puede introducirse más de un gen de factor de transcripción en una planta, transformando la planta con uno o más vectores que comprenden dos o más factores de transcripción, o mediante cultivo selectivo de plantas para producir cruces híbridos que comprenden más de un factor de transcripción introducido.

Genes, rasgos y utilidades que afectan a las características de la planta

Los factores de transcripción de las plantas pueden modular la expresión génica y, a su vez, modularse por la experiencia ambiental de una planta. Las alteraciones significativas en el ambiente de una planta dan como resultado invariablemente un cambio en el patrón de expresión génica del factor de transcripción de la planta. Los patrones de expresión de factores de transcripción alterados generalmente dan como resultado cambios fenotípicos en la planta. El producto o los productos génicos de factores de transcripción en plantas transgénicas difieren entonces en las cantidades o proporciones de los hallados en plantas de tipo silvestre o no transformadas, y esos factores de transcripción probablemente representen polipéptidos que se usan para alterar la respuesta al cambio ambiental. Como ejemplo, está bien aceptado en la técnica que pueden usarse métodos analíticos basados en la alteración de patrones de expresión para explorar con respecto a cambios fenotípicos en una planta de forma mucho más eficaz que lo que puede conseguirse usando métodos tradicionales.

I. Biomasa aumentada.

Las plantas que sobreexpresan nueve factores de transcripción de gancho AT relacionados distintos descritos en el presente documento, incluyendo secuencias de diversas especies de monocotiledóneas y dicotiledóneas tales como los polipéptidos de *Arabidopsis thaliana* G1073, G1067, G1667, G2153 y G2156, G2157, polipéptidos de *Oryza sativa* G3399, G3400, G3401, G3407, G3556, y polipéptidos de *Glycine max* G3456, G3459 y G3460, se hacen mayores que las plantas de control o de tipo silvestre, y generalmente producen hojas más anchas que las plantas de control o de tipo silvestre. Para algunas plantas ornamentales, la capacidad de proporcionar variedades mayores con estos genes o sus equivalentes puede ser altamente deseable. De forma más significativa, las especies de cultivo que sobreexpresan estos genes de diversas especies también producirían cultivos mayores, y por lo tanto mayores producciones, particularmente en las plantas cuya parte vegetativa de la planta es comestible (por ejemplo, lechuga, acelga, etc.). Esto ya se ha observado en *Arabidopsis* y plantas de tomate. Se ha descubierto que las plantas de tomate que sobreexpresan los polipéptidos G2153 y G2157 de *A. thaliana* son significativamente mayores que las plantas de tomate de control de tipo silvestre. Numerosas líneas de *Arabidopsis* que sobreexpresan G3399, G3400, G3401, G3407, o G3556, que son genes del arroz, y G3456, G3459 o G3460, que son genes de la soja, desarrollan rosetas y hojas significativamente mayores que los controles de *Arabidopsis* de tipo silvestre.

II. Tolerancia al aumento del estrés abiótico.

La sobreexpresión de muchos de los factores de transcripción en el clado G1073 de los polipéptidos de factores de transcripción confieren tolerancia al aumento del estrés cuando las secuencias se sobreexpresan en plantas. La biomasa aumentada observada en muchas de estas plantas parece estar relacionada con el mecanismo particular de tolerancia al estrés mostrada por estos genes. La decisión de que un órgano lateral continúe su crecimiento y expansión frente a entrar en las fases de desarrollo tardío (cese del crecimiento y senescencia) se controla genéticamente y hormonalmente, incluyendo regulación en un punto de control del tamaño de los órganos (por ejemplo, Mizukami (2001) Curr Opin Plant Biol. 4: 533-39; Mizukami y Fisher (2000) Proc. Natl. Acad. Sci. USA 97: 942-947; Hu *et al.* 2003 Plant Cell 15: 1591). El tamaño de los órganos se controla por la competencia meristemática de células del órgano, conduciendo la competencia meristemática aumentada a mayor tamaño del órgano (tanto hojas como tallos). Las hormonas vegetales pueden influir en el tamaño de los órganos de la planta, por ejemplo, conduciendo la sobreexpresión de la ruta del etileno a tamaño del órgano reducido. También hay sugerencias de que la auxina desempeña un papel determinante en el tamaño de los órganos. Las respuestas al estrés pueden influir en los niveles de hormonas en tejidos vegetales, incluyendo los niveles de ABA y etileno, modificando de este modo la competencia meristemática y tamaño de los órganos final. Por lo tanto, la sobreexpresión de genes *HRC* relacionados con G1073 altera las aportaciones ambientales (por ejemplo, estrés) al

punto de control de tamaño de los órganos, potenciando de este modo el tamaño de los órganos en condiciones de crecimiento típicas.

Debido a la frecuente exposición a tensiones en condiciones de cultivo de plantas típicas, el tamaño de los órganos genéticamente programado máximo no se alcanza con frecuencia. Se aprecia bien que el tamaño de los órganos de la hoja aumentado puede dar como resultado aumento de la producción de semillas, mediante actividad de fuente y captura de energía potenciada. Por lo tanto, una estrategia importante para optimización de la producción es la alteración de características del sensor que integra las aportaciones de estrés ambiental externa en la competencia meristemática y control de tamaño de los órganos. Los genes *HRC* que son el objeto de la presente invención representan un componente de este mecanismo de control. El aumento de la expresión de los genes *HRC* conduce a sensibilidad reducida del sensor ambiental para control del tamaño de los órganos para estas aportaciones de estrés. Este aumento del umbral de estrés para competencia meristemática reducida da como resultado aumento de la producción vegetativa y de semillas en condiciones de cultivo de plantas típicas. Se sabe que las proteínas de gancho AT modulan la expresión génica mediante interacciones con otras proteínas. Por lo tanto, el mecanismo de integración ambiental para el control del tamaño de los órganos instanciado por proteínas *HRC* tendrá componentes adicionales cuya función se reconocerá por la capacidad de las proteínas codificadas para participar en la regulación de conjuntos génicos que están regulados por proteínas *HRC*. La identificación de componentes adicionales de la integración puede conseguirse identificando otros factores de transcripción que se unen a regiones reguladoras cadena arriba, detectando proteínas que interaccionan directamente con proteínas *HRC*.

A. Respuestas a altas concentraciones de azúcar: detección de azúcar.

Además de su importante papel como una fuente de energía y componente estructural de la célula vegetal, los azúcares son moléculas reguladoras centrales que controlan varios aspectos de la fisiología, el metabolismo y el desarrollo de la planta (Hsieh *et al.* (1998) Proc. Natl. Acad. Sci. USA 95: 13965-13970). Se cree que este control se consigue regulando la expresión génica y, en plantas superiores, se ha mostrado que los azúcares reprimen o activan genes vegetales implicados en muchos procesos esenciales tales como fotosíntesis, metabolismo del glicoxilato, respiración, síntesis y degradación de almidón y sacarosa, respuesta a patógenos, respuesta a heridas, regulación del ciclo celular, pigmentación, floración y senescencia. Los mecanismos por los que los azúcares controlan la expresión génica no se entienden.

Varios mutantes de detección de azúcares han resultado ser alélicos para mutantes de ABA y etileno. Se encuentra ABA en todos los organismos fotosintéticos y actúa como un regulador clave de la transpiración, respuestas al estrés, embriogénesis y germinación de semillas. La mayoría de los efectos de ABA están relacionados con la actuación del compuesto como una señal de disponibilidad de agua reducida, por lo que desencadena una reducción de la pérdida de agua, ralentiza el crecimiento y media en respuestas adaptativas. Sin embargo, ABA también influye en el crecimiento y desarrollo de la planta mediante interacciones con otras fitohormonas. Los estudios fisiológicos y moleculares indican que el maíz y *Arabidopsis* tienen rutas casi idénticas con respecto a la biosíntesis de ABA y transducción de señal. Para una revisión adicional, véase Finkelstein y Rock (2002) "Absciscic acid biosynthesis and response", en The *Arabidopsis* Book, Editores: Somerville y Meyerowitz (American Society of Plant Biologists, Rockville, MD).

Por lo tanto, G1067, G2153, G2156, G2789, G3401, G3406, G3456, y factores de transcripción relacionados están probablemente implicados en la señalización de hormonas basándose en el fenotipo de detección de azúcares de sacarosa de líneas transgénicas con sobreexpresión. Por otro lado, el tratamiento de sacarosa usado en estos experimentos (9,5 % p/v) podría ser también un estrés osmótico. Por lo tanto, se podrían interpretar estos datos como un indicio de que estas líneas transgénicas son más tolerantes al estrés osmótico. Sin embargo, se sabe bien que las respuestas de plantas a ABA, estrés osmótico y otros pueden estar ligadas, y estos diferentes tratamientos pueden incluso actuar de una manera sinérgica para aumentar el grado de una respuesta. Por ejemplo, Xiong, Ishitani, y Zhu ((1999) Plant Physiol. 119: 205-212) han mostrado que pueden usarse estudios genéticos y moleculares para mostrar interacción extensiva entre estrés osmótico, estrés por temperatura y respuestas de ABA en plantas. Esos investigadores analizaron la expresión de *RD29A-LUC* en respuesta a diversos regímenes de tratamiento en *Arabidopsis*. El promotor *RD29A* contiene el elemento tanto sensible a ABA como el sensible a deshidratación, también denominada la repetición C, y puede activarse por estrés osmótico, baja temperatura o tratamiento con ABA; la transcripción del gen de *RD29A* en respuesta a tensiones osmóticas y de frío está mediada por rutas tanto dependientes de ABA como independientes de ABA (Xiong, Ishitani, y Zhu (1999) mencionado anteriormente). *LUC* se refiere a la secuencia codificante de luciferasa de luciérnaga, que, en este caso, estaba conducida por el promotor *RD29A* sensible al estrés. Los resultados revelaron interacciones tanto positivas como negativas, dependiendo de la naturaleza y duración de los tratamientos. Se descubrió que el estrés por baja temperatura alteraba la señalización osmótica pero el estrés por calor moderado potenciaba fuertemente la inducción de estrés osmótico, actuando de este modo de forma sinérgica con las rutas de señalización osmóticas. En este estudio, los autores indicaron que el estrés osmótico y ABA pueden actuar de forma sinérgica mostrando que los tratamientos indujeron simultáneamente la expresión del transgén y el gen endógeno. Se indicaron resultados similares por Bostock y Quatrano ((1992) Plant Physiol. 98: 1356-1363), que descubrieron que el estrés osmótico y ABA actúan de forma sinérgica e inducen la expresión del gen *Em* del maíz. Ishitani *et al.* ((1997) Plant Cell 9: 1935-1949) aislaron un grupo de mutaciones de un único gen de *Arabidopsis* que confieren respuestas

potenciadas tanto a tensiones osmóticas como a ABA. La naturaleza de la recuperación de estos mutantes de estrés osmótico y tratamiento con ABA sugirió que aunque existen rutas de señalizaciones separadas para el estrés osmótico y ABA, las rutas comparten varios componentes; estos componentes comunes pueden mediar en interacciones sinérgicas entre el estrés osmótico y ABA. Por lo tanto, al contrario de la creencia anterior de que las rutas de señalización del estrés dependiente de ABA e independiente de ABA actúan de una manera paralela, los datos de los inventores revelan que estas rutas interaccionan y convergen para activar la expresión de genes de estrés.

Debido a que los azúcares son moléculas de señalización importantes, la capacidad para controlar la concentración de un azúcar de señalización o cómo una planta percibe o responde a un azúcar de señalización podría usarse para controlar el desarrollo, la fisiología o el metabolismo de la planta. Por ejemplo, se ha mostrado que el flujo de sacarosa (un azúcar disacárido usado para transportar de forma sistémica carbono y energía en la mayoría de las plantas) afecta a la expresión génica y altera la acumulación de compuestos de almacenamiento en semillas. La manipulación de la ruta de señalización de sacarosa en semillas puede por lo tanto provocar que las semillas tengan más proteína, aceite o carbohidratos, dependiente del tipo de manipulación. De forma similar, en tubérculos, la sacarosa se convierte en almidón que se usa como un almacén de energía. Se cree que las rutas de señalización de azúcares podrían determinar parcialmente los niveles del almidón sintetizado en los tubérculos. La manipulación de la señalización de azúcar en tubérculos podría conducir a tubérculos con un mayor contenido de almidón.

Por lo tanto, los genes de factores de transcripción desvelados en el presente documento, que manipulan la ruta de transducción de señal de azúcar, incluyendo, por ejemplo G1073 y G2156, junto con sus equivalentes, pueden conducir a expresión génica alterada para producir plantas con rasgos deseables. En particular, la manipulación de las rutas de transducción de señal de azúcar podría usarse para alterar relaciones de fuente-sumidero en semillas, tubérculos, raíces y otros órganos de almacenamiento conduciendo a un aumento de la producción.

B. Respuestas a tensiones osmóticas (alta salinidad, congelación, deshidratación y sequía)

Las plantas se someten a una serie de retos ambientales. Varios de estos, incluyendo estrés salino, estrés osmótico general, estrés por sequía y estrés por congelación, tienen la capacidad de influir en la disponibilidad de agua celular y de planta completa. No resulta sorprendente, por lo tanto, que las respuestas vegetales a esta colección de tensiones están relacionadas. En una revisión reciente, Zhu indica que "la mayoría de los estudios sobre la señalización de estrés hídrico se han centrado en el estrés salino principalmente por que las respuestas vegetales a salinidad y sequía están estrechamente relacionadas y los mecanismos se solapan" (Zhu (2002) Ann. Rev. Plant Biol. 53: 247-273). Se han documentado muchos ejemplos de respuestas similares (es decir, rutas genéticas para este conjunto de tensiones). Por ejemplo, se ha mostrado que los factores de transcripción de CBF condicionan la resistencia a salinidad, congelación y sequía (Kasuga *et al.* (1999) Nature Biotech. 17: 287-291). El gen *rd29B* de *Arabidopsis* se induce en respuesta al estrés tanto de salinidad como de deshidratación, un proceso que está mediado principalmente a través de un proceso de transducción de señal de ABA (Uno *et al.* (2000) Proc. Natl. Acad. Sci. USA 97: 11632-11637), dando como resultado actividad alterada de los factores de transcripción que se unen con un elemento cadena arriba dentro del promotor de *rd29B*. En *Meibryanthemum crystallinum* (hierba del hielo), Patharker y Cushman han mostrado que se induce una proteína quinasa dependiente de calcio (McCDPK1) por exposición a tensiones tanto por sequía como por salinidad (Patharker y Cushman (2000) Plant J. 24: 679-691). También se ha mostrado que la quinasa inducida por estrés fosforila un factor de transcripción, alterando supuestamente su actividad, aunque los niveles de transcrito del factor de transcripción diana no se alteran en respuesta al estrés por salinidad o sequía. De forma similar, Saijo *et al.* demostraron que una proteína quinasa dependiente de calmodulina inducida por salinidad/sequía del arroz (OsCDPK7) confería mayor tolerancia a salinidad y sequía para el arroz cuando se sobreexpresaba (Saijo *et al.* (2000) Plant J. 23: 319-327).

La deshidratación induce estrategias de supervivencia similares en plantas que el estrés por congelación (Yelenosky (1989) Plant Physiol 89: 444-451) y el estrés por sequía induce tolerancia a la congelación (Siminovitch *et al.* (1982) Plant Physiol 69: 250-255; y Guy *et al.* (1992) Planta 188: 265-270). Además de la inducción de proteínas de aclimatación al frío, las estrategias que permiten que las plantas sobrevivan en condiciones de agua baja pueden incluir reducción del área de superficie, o producción de aceite o cera en superficie.

En consecuencia, un experto en la materia esperaría que algunas rutas implicadas en la resistencia a una de estas tensiones, y por lo tanto reguladas por un factor de transcripción individual, también estuvieran implicadas en la resistencia a otra de estas tensiones, reguladas por los mismos factores de transcripción u homólogos. Por supuesto, las rutas de resistencia generales están relacionadas, no son idénticas, y por lo tanto no todos los factores de transcripción que controlan la resistencia a un estrés controlarán la resistencia a las otras tensiones. No obstante, si un factor de transcripción condiciona la resistencia a una de estas tensiones, resultaría evidente para un experto en la materia ensayar con respecto a resistencia a estas tensiones relacionadas.

Por lo tanto, la modificación de la expresión de miembros del clado G1073 puede usarse para aumentar la tolerancia de una planta a condiciones de agua baja y proporcionar los beneficios de supervivencia mejorada, producción aumentada y un alcance de plantación geográfico y temporal extendido. Se ha mostrado que las plantas que sobreexpresan G3399, G3400 y G3460 son más tolerantes a deshidratación en ensayos de desecación basados en

placa que las plantas de control de tipo silvestre (Ejemplo VII). Se ha mostrado que las líneas de plantas que sobreexpresan G1073, G1067, G2153 de *Arabidopsis*, G3399, G3400, G3401 de arroz y G3460 de soja son significativamente más tolerantes a la sequía que las plantas de control de tipo silvestre en ensayos basados en suelo, confirmando los resultados basados en placas.

Se ha mostrado que varias de las secuencias del clado G1073 (por ejemplo, G1073, G1069, G2153, G2156, G2789, G3401 y G3460) tienen un fenotipo de tolerancia al estrés osmótico alterado en virtud de la germinación mejorada de plantas que sobreexpresan estas secuencias en medios que contienen azúcares altos. También se ha mostrado que la mayoría de estos genes confieren mayor tolerancia al estrés por salinidad o desecación a plantas con sobreexpresión (se ha mostrado que todos aumentan la tolerancia al estrés osmótico en *Arabidopsis*, y se ha mostrado que G2153 hace lo mismo para plantas de tomate maduras). Por lo tanto, la modificación de la expresión de estos y otros genes de factores de transcripción desvelados estructuralmente relacionados pueden usarse para aumentar la tasa de germinación o crecimiento en condiciones osmóticas adversas, lo que podría influir en la supervivencia y producción de semillas y plantas. Las tensiones osmóticas pueden regularse por mecanismos de control molecular específicos que incluyen genes que controlan los movimientos de agua e iones, proteínas inducidas por estrés funcionales y estructurales, percepción y transducción de señales y eliminación de radicales libres, y muchos otros (Wang *et al.* (2001) Acta Hort. (ISHS) 560: 285-292). Los instigadores de estrés osmótico incluyen congelación, sequía y alta salinidad, cada uno de los cuales se analizan en más detalle posteriormente.

De muchas maneras, la congelación, alta salinidad y sequía tienen efectos similares en plantas, no siendo el menor de los cuales la inducción de polipéptidos comunes que responden a estas tensiones diferentes. Por ejemplo, la congelación es similar al déficit de agua porque la congelación reduce la cantidad de agua disponible para una planta. La exposición a temperaturas de congelación puede conducir a deshidratación celular a medida que el agua deja las células y forma cristales de hielo en espacios intercelulares (Buchanan (2000) mencionado anteriormente). Como con la alta concentración salina y congelación, los problemas para plantas provocados por la baja disponibilidad de agua incluyen tensiones mecánicas provocadas por la retirada de agua celular. Por lo tanto, la incorporación de factores de transcripción que modifican la respuesta de una planta al estrés osmótico en, por ejemplo, un cultivo o planta ornamental, puede ser útil para reducir el daño o la pérdida. Se abordan posteriormente efectos específicos provocados por la congelación, alta salinidad y sequía.

Los genes del Listado de Secuencias, incluyendo, por ejemplo, G1073, G2153, G2156, G3401, G3456, G3459 y G3460, que proporcionan tolerancia a salinidad pueden usarse para obtener por ingeniería genética cultivos y árboles tolerantes a salinidad que pueden prosperar en suelos con alto contenido salino o en condiciones de sequía. En particular, la tolerancia a la salinidad aumentada durante el estadio de germinación de una planta potencia su supervivencia y producción. Los genes de factores de transcripción desvelados en el presente documento que proporcionan aumento de la tolerancia a la salinidad durante la germinación, el estadio de plántula, y durante el ciclo de vida de una planta, encontrarían un valor particular para conferir supervivencia y producción en áreas en las que un cultivo particular normalmente no prosperaría.

C. Respuestas al estrés por frío

La tolerancia a enfriamiento potenciada puede extender el alcance del crecimiento eficaz de especies de cultivo sensibles a enfriamiento permitiendo una plantación más temprana o recolección más tardía. Puede conferirse tolerancia a enfriamiento mejorada por expresión aumentada de glicerol-3-fosfato acetiltransferasa en cloroplastos (véase, por ejemplo, Wolter *et al.* (1992) *et al.* EMBO J. 4685-4692, y Murata *et al.* (1992) Nature 356: 710-713).

La tolerancia al enfriamiento también podría actuar como un modelo para entender cómo se adaptan las plantas a déficit de agua. El estrés tanto de enfriamiento como de agua comparten rutas de transducción de señales similares y mecanismos de tolerancia/adaptación. Por ejemplo, la aclimatación a temperaturas de enfriamiento podría inducirse por estrés hídrico o tratamiento con ABA. Los genes inducidos por baja temperatura incluyen deshidrinas (o proteínas LEA). Las deshidrinas también se inducen por salinidad, ABA, estrés hídrico y durante los estadios tardíos de la embriogénesis.

Otra influencia importante del enfriamiento se produce durante el almacenamiento después de la recolección. Por ejemplo, algunas frutas y verduras no se almacenan bien a temperaturas bajas (por ejemplo, plátanos, aguacates, melones y tomates). El proceso de maduración normal del tomate se ve alterado si se expone a temperaturas frías. Los genes de factores de transcripción que confieren resistencia a temperaturas de enfriamiento potencian por lo tanto la tolerancia durante el almacenamiento después de la recolección. Se ha mostrado que varios de los genes de factores de transcripción desvelados en el presente documento confieren mejor germinación y crecimiento en condiciones frías. Por ejemplo, la germinación mejorada en condiciones frías vista con G1073, G2153 G2156, G3400, G3401, G3456, G3459 y G3460 indica un papel en la regulación de respuestas al frío por estos genes y otros miembros del clado G1073 de polipéptidos de factores de transcripción. Estos genes pueden por lo tanto sobreexpresarse o modificarse por ingeniería genética de otro modo para manipular la respuesta al estrés por baja temperatura. Los genes que permitieran germinación y vigor de la plántula en frío tendrían una utilidad altamente significativa para permitir que las semillas se planten más temprano en la estación con una alta tasa de supervivencia. Los genes de factores de transcripción que confieren mejor supervivencia en climas más fríos

5 permiten a un cultivador adelantar el momento de plantación en primavera y extender la estación de crecimiento hasta el otoño para mayores producciones de cultivo. La germinación de semillas y supervivencia a temperaturas significativamente por debajo de la temperatura media requerida para la germinación de semillas y supervivencia de plantas no transformadas aumentaría la distribución potencial de una planta de cultivo en regiones en las que de otro modo no conseguiría desarrollarse.

Biomasa aumentada

10 Se ha mostrado que la sobreexpresión de G1073 y varios otros miembros del clado G1073, incluyendo G1667, G2153, G2156, G3399, G3400, G3401, G3407, G3456, G3459, G3460 y G3556, produce plantas que son mayores que las de control, particularmente en estadios de crecimiento posteriores. Para algunas plantas ornamentales, la capacidad de proporcionar variedades mayores con estos genes o sus equivalentes puede ser altamente deseable. Para muchas plantas, incluyendo árboles frutales, árboles que se usan para la producción de madera o árboles y
15 arbustos que actúan como pantallas contra el viento o para la vista, la mayor estatura proporciona beneficios mejorados en forma de mayor producción o protección mejorada. Las especies de cultivo también pueden producir mayores rendimientos en cultivos mayores, particularmente en los que la parte vegetativa de la planta es comestible.

Floración retardada

20 En un número significativo de especies, por ejemplo, cultivos de raíces, en los que las partes vegetativas de las plantas constituyen el cultivo y los tejidos reproductivos se descartan, es ventajoso identificar e incorporar genes de factores de transcripción que retarden o eviten la floración para evitar que se desvíen recursos al desarrollo reproductivo. Por ejemplo, la sobreexpresión de G1073, G1067, G1667, G2153, G2156, G3399, G3401, G3406, G3459, G3460 o G3556 retarda el tiempo de floración en plantas transgénicas. La extensión del desarrollo
25 vegetativo con genes de factores de transcripción desvelados en el presente documento podría de este modo proporcionar grandes aumentos de la producción. La prevención de la floración puede ayudar a maximizar las producciones vegetativas y evitar el escape de polen de organismos modificados genéticamente (GMO).

Sumario de características de plantas alteradas.

30 Se ha mostrado en experimentos de laboratorio y de campo que miembros del clado G1073 de polipéptidos de factores de transcripción, que derivan de una amplia serie de plantas, confieren un mayor tamaño, tolerancia al estrés abiótico y fenotipos de floración retardada en plantas que sobreexpresan estas secuencias. La divulgación también proporciona polinucleótidos que codifican polipéptidos del clado G1073, fragmentos de los mismos, dominios conservados de los mismos, parálogos, ortólogos, equivalentes y fragmentos de los mismos. Estas secuencias se enumeran en el Listado de Secuencias, y debido al alto grado de similitud estructural de las presentes secuencias, se espera que muchas de las secuencias para las que no se han generado datos también actúen aumentando la biomasa vegetal y/o tolerancia al estrés abiótico. La divulgación también abarca los complementos de los polinucleótidos. Los polinucleótidos también son útiles para explorar bibliotecas de moléculas o compuestos con respecto a unión específica y para identificar otras secuencias de miembros del clado G1073 identificando ortólogos que tienen secuencias similares, particularmente en los dominios conservados.

Antisentido y cosupresión

45 Además de la expresión de los ácidos nucleicos descritos en el presente documento como reemplazo génico o ácidos nucleicos de modificación del fenotipo de la planta, los ácidos nucleicos también son útiles para la supresión con sentido y antisentido de la expresión, por ejemplo, para regular negativamente la expresión de un ácido nucleico, por ejemplo, como un mecanismo adicional para modular el fenotipo de la planta. Es decir, los ácidos nucleicos, o subsecuencias o secuencias antisentido de los mismos, pueden usarse para bloquear la expresión de ácidos nucleicos homólogos de origen natural. Se conoce en la técnica diversas tecnologías con sentido y antisentido, por ejemplo, como se expone en Lichtenstein y Nellen (1997) Antisense Technology: A Practical Approach IRL Press at Oxford University Press, Oxford, Reino Unido. La regulación antisentido también se describe en Crowley *et al.* (1985) Cell 43: 633-641; Rosenberg *et al.* (1985) Nature 313: 703-706; Preiss *et al.* (1985) Nature 313: 27-32; Melton (1985) Proc. Natl. Acad. Sci. USA 82: 144-148; Izant y Weintraub (1985) Science 229: 345-352; y Kim y Wold (1985) Cell 42: 129-138. Se conocen en la técnica métodos adicionales para la regulación antisentido. Se ha usado regulación antisentido para reducir o inhibir la expresión de genes de plantas en, por ejemplo, la Publicación de Patente Europea N° 271988. El ARN antisentido puede usarse para reducir la expresión génica para producir un cambio fenotípico visible o bioquímico en una planta (Smith *et al.* (1988) Nature 334: 724-726; Smith *et al.* (1990) Plant Mol. Biol. 14: 369-379). En general, se introducen secuencias con sentido o antisentido en una célula, donde
50 opcionalmente se amplifican, por ejemplo, por transcripción. Dichas secuencias incluyen tanto secuencias oligonucleotídicas sencillas como secuencias catalíticas tales como ribozimas.

65 Por ejemplo, una reducción o eliminación de la expresión (es decir, una "anulación") de un factor de transcripción o polipéptido homólogo de factor de transcripción en una planta transgénica, por ejemplo, para modificar un rasgo de la planta, puede obtenerse introduciendo una construcción antisentido correspondiente al polipéptido de interés

como un ADNc. Para supresión antisentido, el factor de transcripción o ADNc homólogo se dispone en orientación inversa (con respecto a la secuencia codificante) en relación con la secuencia promotora en el vector de expresión. No es necesario que la secuencia introducida sea el ADNc o gen de longitud completa, y no es necesario que sea idéntica al ADNc o gen hallado en el tipo de planta para transformar. Normalmente, es necesario que la secuencia antisentido sea solamente capaz de hibridar con el gen diana o ARN de interés. Por lo tanto, cuando la secuencia introducida es de longitud más corta, será necesario un mayor grado de homología con la secuencia del factor de transcripción endógena para supresión antisentido eficaz. Aunque pueden utilizarse secuencias antisentido de diversas longitudes, preferentemente, la secuencia antisentido introducida en el vector será de al menos 30 nucleótidos de longitud, y se observará normalmente supresión antisentido mejorada a medida que aumenta la longitud de la secuencia antisentido. Preferentemente, la longitud de la secuencia antisentido en el vector será de más de 100 nucleótidos. La transcripción de una construcción antisentido como se describe da como resultado la producción de moléculas de ARN que son el complemento inverso de moléculas de ARNm transcritas del gen de factor de transcripción endógeno en la célula vegetal.

También puede conseguirse supresión de la expresión de genes de factores de transcripción endógenos usando una ribozima. Las ribozimas son moléculas de ARN que poseen actividad endorribonucleasa altamente específica. La producción y uso de ribozimas se desvela en la Patente de Estados Unidos N° 4.987.071 y Patente de Estados Unidos N° 5.543.508. Pueden usarse secuencias de ribozimas sintéticas incluyendo ARN antisentido para conferir actividad de escisión de ARN en el ARN antisentido, de modo que las moléculas de ARNm endógenas que hibridan con el ARN antisentido se escinden, lo que a su vez conduce a una inhibición antisentido potenciada de la expresión de genes endógenos.

Los vectores en los que el ARN codificado por un factor de transcripción o ADNc homólogo del factor de transcripción está sobreexpresado también pueden usarse para obtener cosupresión de un gen endógeno correspondiente, por ejemplo, de la manera descrita en la Patente de Estados Unidos N° 5.231.020 de Jorgensen. Dicha cosupresión (también denominada supresión con sentido) no requiere que el ADNc del factor de transcripción completo se introduzca en las células vegetales, ni requiere que la secuencia introducida sea exactamente idéntica al gen de factor de transcripción endógeno de interés. Sin embargo, como con la supresión antisentido, la eficacia de supresión se potenciará a medida que aumenta la especificidad de la hibridación, por ejemplo, a medida que se alarga la secuencia introducida y/o a medida que aumenta la similitud de secuencia entre la secuencia introducida y el gen del factor de transcripción endógeno.

Los vectores que expresan una forma no traducible del ARNm del factor de transcripción, por ejemplo, secuencias que comprenden uno o más codones de parada, o mutación sin sentido, también pueden usarse para suprimir la expresión de un factor de transcripción endógeno, reduciendo o eliminando de este modo su actividad y modificando uno o más rasgos. Se describen métodos para producir dichas construcciones en la Patente de Estados Unidos N° 5.583.021. Preferentemente, dichas construcciones se realizan introduciendo un codón de parada prematuro en el gen del factor de transcripción. Como alternativa, puede modificarse un rasgo de la planta por silenciamiento génico usando ARN bicatenario (Sharp (1999) *Genes and Development* 13: 139-141). Otro método para anular la expresión de un gen es mediante mutagénesis de inserción usando el ADN-T de *Agrobacterium tumefaciens*. Después de generar los mutantes de inserción, los mutantes pueden explorarse para identificar los que contienen la inserción en un factor de transcripción o gen homólogo de factor de transcripción. Las plantas que contienen un único acontecimiento de inserción de transgén en el gen deseado pueden cruzarse para generar plantas homocigotas para la mutación. Dichos métodos se conocen bien por los expertos en la materia (véase por ejemplo Koncz *et al.* (1992) *Methods in Arabidopsis Research*, World Scientific Publishing Co. Pte. Ltd., River Edge NJ).

La supresión de la expresión del gen del factor de transcripción endógeno también puede conseguirse usando interferencia de ARN, o ARNi. La ARNi es una técnica de silenciamiento génico dirigida, postranscripcional, que usa ARN bicatenario (ARNbc) para inducir la degradación de ARN mensajero (ARNm) que contiene la misma secuencia que el ARNbc (Constans, (2002) *The Scientist* 16:36). Se producen ARN de interferencia pequeños, o ARNip, en al menos dos etapas: una ribonucleasa endógena escinde ARNbc más largo en ARN más cortos, de 21-23 nucleótidos de longitud. Los segmentos de ARNip median después en la degradación del ARNm diana (Zamore, (2001) *Nature Struct. Biol.*, 8: 746-50). Se ha usado ARNi para la determinación de la función génica de una manera similar a los oligonucleótidos antisentido (Constans, (2002) *The Scientist* 16: 36). Se han modificado por ingeniería genética vectores de expresión que expresan de forma continua ARNip en células transfectadas de forma transitoria y estable para expresar ARN en horquilla pequeños (ARNhp), que se procesan *in vivo* en moléculas de tipo ARNip capaces de llevar a cabo silenciamiento específico de genes (Brummelkamp *et al.*, (2002) *Science* 296: 550-553, y Paddison, *et al.* (2002) *Genes & Dev.* 16: 948-958). El silenciamiento génico postranscripcional por ARN bicatenario se analiza en más detalle en Hammond *et al.* (2001) *Nature Rev Gen* 2: 110-119, Fire *et al.* (1998) *Nature* 391: 806-811 y Timmons y Fire (1998) *Nature* 395: 854.

Como alternativa, puede alterarse un fenotipo vegetal eliminando un gen endógeno, tal como un factor de transcripción u homólogo de factor de transcripción, por ejemplo, por recombinación homóloga (Kempin *et al.* (1997) *Nature* 389: 802-803).

También puede modificarse un rasgo vegetal usando el sistema Cre-lox (por ejemplo, como se describe en la Patente de Estados Unidos N° 5.658.772). Puede modificarse un genoma vegetal para incluir primer y segundo sitios lox que después se ponen en contacto con una Cre recombinasa. Si los sitios lox están en la misma orientación, la secuencia de ADN intermedia entre los dos sitios se escinde. Si los sitios lox están en la orientación opuesta, se invierte la secuencia intermedia.

Los polinucleótidos y polipéptidos descritos en el presente documento también pueden expresarse en una planta en ausencia de un casete de expresión manipulando la actividad o el nivel de expresión del gen endógeno por otros medios, tales como, por ejemplo, expresando de forma ectópica un gen por marcaje de la activación de ADN-T (Ichikawa *et al.* (1997) Nature 390 698-701; Kakimoto *et al.* (1996) Science 274: 982-985). Este método implica la transformación de una planta con un marcador génico que contiene múltiples potenciadores de la transcripción y una vez que se ha insertado el marcador en el genoma, la expresión de una secuencia codificante del gen flanqueante se desregula. En otro ejemplo, la maquinaria transcripcional en una planta puede modificarse para aumentar los niveles de transcripción de un polinucleótido (véase, por ejemplo, Publicaciones de PCT WO 96/06166 y WO 98/53057 que describen la modificación de la especificidad de unión a ADN de proteínas de dedos de cinc cambiando aminoácidos particulares en el motivo de unión a ADN).

La planta transgénica también puede incluir la maquinaria necesaria para expresar o alterar la actividad de un polipéptido codificado por un gen endógeno, por ejemplo, alterando el estado de fosforilación del polipéptido para mantenerlo en un estado activado.

Pueden producirse plantas transgénicas (o células vegetales, o explantes vegetales o tejidos vegetales) incorporando los polinucleótidos y/o expresando los polipéptidos descritos en el presente documento por diversas técnicas bien establecidas como se ha descrito anteriormente. Después de la construcción de un vector, más normalmente un casete de expresión, que incluye un polinucleótido, por ejemplo, que codifica un factor de transcripción u homólogo de factor de transcripción, pueden usarse técnicas convencionales para introducir el polinucleótido en una planta, una célula vegetal, un explante vegetal o un tejido vegetal de interés. Opcionalmente, la célula, el explante o el tejido vegetal pueden regenerarse para producir una planta transgénica.

La planta puede ser cualquier planta superior, incluyendo gimnospermas, plantas monocotiledóneas y dicotiledóneas. Están disponibles protocolos adecuados para Leguminosas (alfalfa, soja, trébol, etc.), Umbelíferas (zanahoria, apio, chirivía), Crucíferas (repollo, rábano, colza, brócoli, etc.), Cucurbitáceas (melones y pepino), Gramíneas (trigo, maíz, arroz, cebada, mijo, etc.), Solanáceas (patata, tomate, tabaco, pimientos, etc.) y diversos otros cultivos. Véanse protocolos descritos en Ammirato *et al.*, Editores, (1984) Handbook of Plant Cell Culture, Vol. 2 y 3., Crop Species, Macmillan Publ. Co., Nueva York NY; Shimamoto *et al.* (1989) Nature 338: 274-276; Fromm *et al.* (1990) Bio/Technol. 8: 833-839; y Vasil *et al.* (1990) Bio/Technol. 8: 429-434.

La transformación y regeneración de células vegetales tanto monocotiledóneas como dicotiledóneas son ahora rutinarias, y la selección de la técnica de transformación más apropiada se determinará por el practicante. La elección del método variará con el tipo de planta para transformar; los expertos en la materia reconocerán la idoneidad de métodos particulares para tipos vegetales dados. Los métodos adecuados pueden incluir, pero sin limitación: electroporación de protoplastos vegetales; transformación mediada por liposomas; transformación mediada por polietilenglicol (PEG); transformación usando virus; microinyección de células vegetales; bombardeo de microproyectiles de células vegetales; infiltración en vacío; y transformación mediada por *Agrobacterium tumefaciens*. La transformación significa introducir una secuencia de nucleótidos en una planta de una manera que provoque expresión estable o transitoria de la secuencia.

Los ejemplos exitosos de la modificación de características vegetales por transformación con secuencias clonadas que actúan para ilustrar el conocimiento actual en este campo de la tecnología incluyen: Patentes de Estados Unidos N° 5.571.706; 5.677.175; 5.510.471; 5.750.386; 5.597.945; 5.589.615; 5.750.871; 5.268.526; 5.780.708; 5.538.880; 5.773.269; 5.736.369 y 5.610.042.

Después de la transformación, las plantas se seleccionan preferentemente usando un marcador seleccionable dominante incorporado en el vector de transformación. Normalmente, dicho marcador conferirá resistencia a antibióticos o herbicidas en las plantas transformadas, y puede conseguirse selección de transformantes exponiendo las plantas a concentraciones apropiadas del antibiótico o herbicida.

Después de seleccionarse y cultivarse hasta la madurez las plantas transformadas, se identifican las plantas que muestran un rasgo modificado.

El rasgo modificado puede ser cualquiera de los rasgos descritos anteriormente. Adicionalmente, para confirmar que el rasgo modificado se debe a cambios en los niveles de expresión o actividad del polipéptido o polinucleótido puede determinarse analizando la expresión de ARNm usando transferencias de Northern, RT-PCR o micromatrices, o expresión de proteínas usando inmunotransferencias o transferencias de Western o ensayos de desplazamiento en gel.

Sistemas integrados para determinar la identidad de secuencia

Además de proporcionar composiciones y métodos para mejorar rasgos vegetales, la presente invención puede ser un sistema integrado, ordenador o medio leíble por ordenador que comprende un conjunto de instrucciones para determinar la identidad de una o más secuencias en una base de datos. Además, el conjunto de instrucciones puede usarse para generar o identificar secuencias que cumplen cualquier criterio específico. Además, el conjunto de instrucciones puede usarse para asociar o ligar ciertos beneficios funcionales, tales como características mejoradas, con una o más secuencias identificadas.

Por ejemplo, el conjunto de instrucciones puede incluir, por ejemplo, una comparación de secuencias u otro programa de alineamiento, por ejemplo, un programa disponible tal como, por ejemplo, el Paquete de Wisconsin Versión 10.0, tal como BLAST, FASTA, PILEUP, FINDPATTERNS o similares (GCG, Madison, WI). Puede buscarse en bases de datos de secuencias públicas tales como GenBank, EMBL, Swiss-Prot y PIR o bases de datos de secuencias privadas tales como la base de datos de secuencias PHYTOSEQ (Incyte Genomics, Wilmington, DE).

Puede realizarse alineamiento de secuencias para comparación por el algoritmo de homología local de Smith y Waterman (1981) *Adv. Appl. Math.* 2: 482-489, por el algoritmo de alineamiento de homología de Needleman y Wunsch (1970) *J. Mol. Biol.* 48: 443-453, por el método de búsqueda de similitud de Pearson y Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444-2448, por implementaciones computarizadas de estos algoritmos. Después del alineamiento, se realizan normalmente comparaciones de secuencia entre dos (o más) polinucleótidos o polipéptidos comparando secuencias de las dos secuencias sobre una ventana de comparación para identificar y comparar regiones locales de similitud de secuencia. La ventana de comparación puede ser un segmento de al menos aproximadamente 20 posiciones contiguas, habitualmente de aproximadamente 50 a aproximadamente 200, más habitualmente de aproximadamente 100 a aproximadamente 150 posiciones contiguas. Se proporciona una descripción del método en Ausubel *et al.* (hasta 2000) mencionado anteriormente.

Puede usarse diversos métodos para determinar relaciones de secuencias, incluyendo alineamiento manual y alineamiento y análisis de secuencias asistidos por ordenador. Este último enfoque es un enfoque preferido en la presente invención, debido al rendimiento aumentado proporcionado por los métodos asistidos por ordenador. Como se ha indicado anteriormente, están disponibles diversos programas informáticos para realizar alineamiento de secuencias, o pueden producirse por un experto en la materia.

Un algoritmo ejemplar que es adecuado para determinar el porcentaje de identidad de secuencia y similitud de secuencia es el algoritmo BLAST, que se describe en Altschul *et al.* (1990) mencionado anteriormente. Está públicamente disponible software para realizar análisis de BLAST, por ejemplo, a través del Centro Nacional para la Información Biotecnológica de la Biblioteca Nacional de Medicina (nc-bi.nlm.nih.gov); véase en el sitio web de la red global (www) de los Institutos Nacionales de la Salud del gobierno de Estados Unidos (gov)). Este algoritmo implica identificar en primer lugar pares de secuencias de alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia de consulta, que coinciden con o satisfacen alguna puntuación T umbral de valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de bases de datos. T se denomina el umbral de puntuación de palabra vecina (Altschul *et al.* (1990, 1993) mencionado anteriormente). Estos aciertos de palabra vecina iniciales actúan como semillas para iniciar búsquedas para encontrar HSP más largos que los contengan. Los aciertos de palabra se extienden después en ambas direcciones a lo largo de cada secuencia en tanto pueda aumentarse la puntuación de alineamiento acumulada. Las puntuaciones acumuladas se calculan usando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de restos coincidentes; siempre > 0) y N (puntuación de penalización para restos desapareados; siempre < 0). Para secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulada. La extensión de los aciertos de palabra en cada dirección se detiene cuando: la puntuación de alineamiento acumulada cae en una cantidad X de su valor adquirido máximo; la puntuación acumulada llega a cero o por debajo, debido a la acumulación de uno o más alineamientos de restos de puntuación negativa; o se alcanza el final de una de las secuencias. Los parámetros del algoritmo BLAST W , T y X determinan la sensibilidad y velocidad del alineamiento. El programa BALSTN (para secuencias de nucleótidos) usa por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, un punto de corte de 100, $M=5$, $N=-4$, y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP usa por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10, y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919). A no ser que se indique de otro modo, "identidad de secuencia" se refiere en el presente documento al % de identidad de secuencia generado a partir de un *tblastx* usando la versión de NCBI del algoritmo a los ajustes por defecto usando alineamientos con huecos con el filtro "apagado" (véase, por ejemplo, el sitio web de NIH NLM NCBI en ncbi.nlm.nih.gov).

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787). Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de menor suma ($P(N)$), que proporciona un indicio de la probabilidad por la que se produciría al azar una coincidencia entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia (y, por lo tanto, en este contexto, homólogo) si la probabilidad de suma menor en una comparación del

ácido nucleico de ensayo con el ácido nucleico de referencia es menor de aproximadamente 0,1, o menor de aproximadamente 0,01, y/o incluso menor de aproximadamente 0,001. Un ejemplo adicional de un algoritmo de alineamiento de secuencias útil es PILEUP. PILEUP crea un alineamiento de secuencias múltiples de un grupo de secuencias relacionadas usando alineamientos por pares, progresivos. El programa puede alinear, por ejemplo, hasta 300 secuencias de una longitud máxima de 5.000 letras.

El sistema integrado u ordenador incluye normalmente una interfaz de entrada del usuario que permite a un usuario ver selectivamente uno o más registros de secuencias correspondientes a la o las cadenas de caracteres, así como un conjunto de instrucciones que alinean la o las cadenas de caracteres entre sí o con una cadena de caracteres adicional para identificar una o más regiones de similitud de secuencia. El sistema puede incluir un enlace de una o más cadenas de caracteres con un fenotipo o función génica particular. Normalmente, el sistema incluye un elemento de salida leíble por el usuario que presenta un alineamiento producido por el conjunto de instrucciones de alineamiento.

Los métodos de la presente invención pueden implementarse en un ambiente informático localizado o distribuido. En un ambiente distribuido, los métodos pueden implementarse en un único ordenador que comprende múltiples procesadores o en múltiples ordenadores. Los ordenadores pueden conectarse, por ejemplo mediante un bus común, pero más preferentemente el ordenador o los ordenadores son nodos en una red. La red puede ser una red de área local o extendida generalizada o dedicada y, en ciertas realizaciones preferidas, los ordenadores pueden ser componentes de una intranet o Internet.

Por lo tanto, la divulgación proporciona métodos para identificar una secuencia similar u homóloga de uno o más polinucleótidos como se indica en el presente documento, o uno o más polipéptidos diana codificados por los polinucleótidos, o de otro modo indicado en el presente documento y pueden incluir enlace o asociación del fenotipo o función génica de una planta dada con una secuencia. En los métodos, se proporciona una base de datos de secuencias (localmente o a través de una intra o Internet) y se realiza una consulta sobre la base de datos de secuencias usando las secuencias relevantes en el presente documento y fenotipos o funciones génicas de plantas asociados.

Puede introducirse cualquier secuencia en el presente documento en la base de datos, antes o después de consultar la base de datos. Esto proporciona tanto expansión de la base de datos como, si se realiza antes de la etapa de consulta, inserción de secuencias de control en la base de datos. Las secuencias de control pueden detectarse por la consulta para asegurar la integridad general tanto de la base de datos como de la consulta. Como se ha observado, la consulta puede realizarse usando una interfaz basada en un buscador web. Por ejemplo, la base de datos puede ser una base de datos pública centralizada tal como las indicadas en el presente documento, y la búsqueda puede realizarse desde una terminal remota u ordenador a través de una intranet o Internet.

Puede usarse cualquier secuencia en el presente documento para identificar una secuencia similar, homóloga, paróloga u ortóloga en otra planta. Esto proporciona medios para identificar secuencias endógenas en otras plantas que pueden ser útiles para alterar un rasgo de plantas descendientes, que resulta del cruce de dos plantas de diferente cepa. Por ejemplo, pueden identificarse secuencias que codifican un ortólogo de cualquiera de las secuencias del presente documento que aparecen de forma natural en una planta con un rasgo deseado usando las secuencias desveladas en el presente documento. La planta se cruza después con una segunda planta de la misma especie pero que no tiene el rasgo deseado para producir descendencia que después puede usarse en experimentos de cruce adicionales para producir el rasgo deseado en la segunda planta. Por lo tanto la planta descendiente resultante no contiene transgenes; la expresión de la secuencia endógena también puede regularse por tratamiento con un producto químico particular u otro medio, tal como EMR. Algunos ejemplos de dichos compuestos bien conocidos en la técnica incluyen: etileno; citoquininas; compuestos fenólicos, que estimulan la transcripción de los genes necesarios para la infección; monosacáridos específicos y ambientes ácidos que potencian la inducción del gen vir; polisacáridos ácidos que inducen uno o más genes cromosómicos; y opinas; otros mecanismos incluyen tratamiento con luz u oscuridad (para una revisión de los ejemplos de dichos tratamientos, véase, Winans (1992) Microbiol. Rev. 56: 12-31; Eyal *et al.* (1992) Plant Mol. Biol. 19: 589-599; Chrispeels *et al.* (2000) Plant Mol. Biol. 42: 279-290; Piazza *et al.* (2002) Plant Physiol. 128: 1077-1086).

La Tabla 5 enumera secuencias que se ha descubierto que son ortólogas de varios factores de transcripción representativos. Los encabezamientos de columnas incluyen los factores de transcripción enumerados por (a) la SEC ID N°: del ortólogo o nucleótido que codifica el ortólogo; (b) el Identificador de Secuencia o (c) Número de Referencia de GenBank; (d) la especie de la que derivan los ortólogos de los factores de transcripción; y (e) la probabilidad de menor suma durante el análisis de BLAST.

Tabla 5. Parálogos y ortólogos y otros genes relacionados de genes de factores de transcripción de *Arabidopsis* representativos identificados usando BLAST

SEC ID N°: del Ortólogo o Nucleótido que Codifica el Ortólogo	GID N°	Identificador de Secuencia o Número de Referencia	Especie de la que Deriva el Ortólogo	Probabilidad de Menor Suma para la Secuencia Polinucleotídica de <i>Arabidopsis</i>
1	G1073		<i>Arabidopsis thaliana</i>	
3	G1067		<i>Arabidopsis thaliana</i>	
5	G2153		<i>Arabidopsis thaliana</i>	
7	G2156		<i>Arabidopsis thaliana</i>	
41	G1069		<i>Arabidopsis thaliana</i>	5e-90**
43	G1945		<i>Arabidopsis thaliana</i>	5e-51**
45	G2155		<i>Arabidopsis thaliana</i>	6e-43**
47	G1070		<i>Arabidopsis thaliana</i>	5e-70**
49	G2657		<i>Arabidopsis thaliana</i>	3e-70†
51	G1075		<i>Arabidopsis thaliana</i>	8e-72**
53	G1076		<i>Arabidopsis thaliana</i>	9e-74**
9	G3399	AP004165	<i>Oryza sativa</i> (grupo de cultivar japónica)	1e-81†
11	G3407	AP004635	<i>Oryza sativa</i>	5e-90†
13	G3456	BM525692	<i>Glycine max</i>	2e-87**
39	G3556		<i>Oryza sativa</i>	7e-67††
15	G3459	C33095_1	<i>Glycine max</i>	6e-67††
17	G3460	C33095_2	<i>Glycine max</i>	1e-66*
		BH566718	<i>Brassica oleracea</i>	1e-129**
		BH685875	<i>Brassica oleracea</i>	1e-124†
		BZ432677	<i>Brassica oleracea</i>	1e-113**
		BZ433664	<i>Brassica oleracea</i>	1e-107†
		BH730050	<i>Brassica oleracea</i>	1e-104†
		AP004971	<i>Lotus corniculatus</i> var. <i>japonicus</i>	3e-91**
		CC729476	<i>Zea mays</i>	1e-83**
21	G3403	AP004020	<i>Oryza sativa</i> (grupo de cultivar japónica)	2e-81**
		AAAA01000486	<i>Oryza sativa</i> (grupo de cultivar índica)	7e-80*
		CB003423	<i>Vitis vinifera</i>	2e-76*
		CC645378	<i>Zea mays</i>	4e-75*
23	G3458	C32394_2	<i>Glycine max</i>	9e-73**
25	G3406	AL662981	<i>Oryza sativa</i>	7e-73*
		BQ785950	<i>Glycine max</i>	3e-73*
		BH975957	<i>Brassica oleracea</i>	9e-72*

		BQ865858	<i>Lactuca sativa</i>	7e-72*
		CB891166	<i>Medicago truncatula</i>	5e-72*
		CF229888	<i>Populus x canescens</i>	2e-71*
		BQ863249	<i>Lactuca sativa</i>	2e-71*
		BG134451	<i>Lycopersicon esculentum</i>	3e-70*
27	G3405	AP005653	<i>Oryza sativa</i> (grupo de cultivar japónica)	1e-69**
29	G3400	AP005477	<i>Oryza sativa</i> (grupo de cultivar japónica)	2e-67*
31	G3404	AP003526	<i>Oryza sativa</i> (grupo de cultivar japónica)	2e-67*
		AP004971	<i>Lotus corniculatus</i> var. <i>japonicus</i>	7e-66*
		BM110212	<i>Solanum tuberosum</i>	8e-65*
		AC124953	<i>Medicago truncatula</i>	2e-63*
35	G3462	BI321563	<i>Glycine max</i>	3e-61*
		BH660108	<i>Brassica oleracea</i>	2e-61†
		BQ838600	<i>Triticum aestivum</i>	2e-59*
		CD825510	<i>Brassica napus</i>	7e-58†
		BF254863	<i>Hordeum vulgare</i>	1e-56*
37	G3401	AAAA01017331 SC17331 AP004587	<i>Oryza sativa</i> (grupo de cultivar japónica)	9e-42*
* Comparación de probabilidad de menor suma con G1073 † Comparación de probabilidad de menor suma con G1067 ** Comparación de probabilidad de menor suma con G2153 †† Comparación de probabilidad de menor suma con G2156				

Modelación molecular

- Otro medio que puede usarse para confirmar la utilidad y función de secuencias de factores de transcripción que son ortólogos o parálogos de factores de transcripción desvelados en el presente documento es mediante el uso de software de modelación molecular. La modelación molecular se usa rutinariamente para predecir la estructura polipeptídica, y están disponibles para este fin diversos programas de modelación de estructura proteica, tales como "Insight H" (Accelrys, Inc.). La modelación puede usarse por lo tanto para predecir qué restos de un polipéptido pueden cambiarse sin alterar la función (Cramer *et al.* (2003) Patente de Estados Unidos N° 6.521.453). Por lo tanto, puede mostrarse que los polipéptidos que son secuencialmente similares tienen una mayor probabilidad de función similar por su similitud estructural, que puede, por ejemplo, establecerse por comparación de regiones de superestructura. Las tendencias relativas de los aminoácidos a formar regiones de superestructura (por ejemplo, hélices y láminas β) están bien establecidas. Por ejemplo, O'Neil *et al.* ((1990) Science 250: 646-651) han analizado en detalle las tendencias de formación de hélices de los aminoácidos. Pueden usarse tablas de actividad formadora de estructuras relativa para aminoácidos como tablas de sustitución para predecir qué restos pueden sustituirse funcionalmente en una región dada, por ejemplo, en dominios de unión a ADN de factores de transcripción conocidos y equivalentes. Pueden después identificarse homólogos que probablemente sean funcionalmente similares.
- Es de interés particular la estructura de un factor de transcripción en la región de sus dominios conservados, tales como los identificados en la Tabla 1. Pueden realizarse análisis estructurales comparando la estructura del factor de transcripción conocido alrededor de su dominio conservado con los de ortólogos y parálogos. El análisis de varios polipéptidos dentro de un grupo o clado de factores de transcripción, incluyendo los polipéptidos funcional o secuencialmente similares proporcionados en el Listado de Secuencias, también puede proporcionar un

entendimiento de elementos estructurales requeridos para regular la transcripción dentro de una familia dada.

Ejemplos

La invención, que se describe ahora de forma general, se entenderá más fácilmente por referencia a los siguientes ejemplos, que se incluyen únicamente para fines de ilustración de ciertos aspectos y realizaciones de la presente invención y no se pretende que limiten la invención. Se reconocerá por un experto en la materia que un factor de transcripción que está asociado con un primer rasgo particular puede también estar asociado con al menos otro segundo rasgo no relacionado e inherente que no se predijo por el primer rasgo.

Ejemplo I: identificación y clonación de genes de longitud completa

Se identificaron secuencias de factores de transcripción potenciales (genómicas o EST) relacionadas con factores de transcripción conocidos en la base de datos de GenBank de *Arabidopsis thaliana* usando el programa de análisis de secuencias tblastn usando parámetros por defecto y un umbral de punto de corte de P valor de -4 o -5 o menor, dependiendo de la longitud de la secuencia de consulta. Después se exploraron aciertos de secuencia de factor de transcripción potenciales para identificar los que contenían cadenas de secuencias particulares. Si los aciertos de secuencia contenían dichas cadenas de secuencias, se confirmaron las secuencias como factores de transcripción.

Como alternativa, se exploraron bibliotecas de ADNc de *Arabidopsis thaliana* derivadas de diferentes tejidos o tratamientos, o bibliotecas genómicas para identificar nuevos miembros de una familia de transcripción usando un enfoque de hibridación de baja rigurosidad. Las sondas se sintetizaron usando cebadores específicos de genes en una reacción de PCR convencional (temperatura de hibridación de 60 °C) y marcados con ³²P dCTP usando el Kit de Marcaje de ADN High Prime (Roche Diagnostics Corp., Indianápolis, IN). Se añadieron sondas radiomarcadas purificadas a filtros sumergidos en medio de hibridación Church (NaPO₄ 0,5 M pH 7,0, SDS 7 %, albúmina de suero bovino 1 % p/v) y se hibridaron durante una noche a 60 °C con agitación. Los filtros se lavaron dos veces durante 45 a 60 minutos con SSC 1x, SDS 1 % a 60 °C.

Para identificar la secuencia adicional 5' o 3' de una secuencia de ADNc parcial en una biblioteca de ADNc, se realizó amplificación rápida 5' y 3' de los extremos de ADNc (RACE) usando el kit de amplificación de ADNc MARATHON (Clontech, Palo Alto, CA). En general, el método implicó en primer lugar aislar ARNm poli(A), realizando síntesis de ADNc de primera y segunda cadena para generar ADNc bicatenario, haciendo los extremos del ADNc romos, seguido de ligamiento del Adaptador de MARATHON con el ADNc para formar una biblioteca de ADNc bc ligado a adaptador.

Se diseñaron cebadores específicos de genes para usar junto con cebadores específicos de adaptador para reacciones RACE tanto 5' como 3'. Se usaron cebadores anidados, en lugar de cebadores individuales, para aumentar la especificidad de PCR. Usando reacciones RACE 5' y 3', se obtuvieron fragmentos RACE 5' y 3', se secuenciaron y se clonaron. El proceso puede repetirse hasta que se identifiquen los extremos 5' y 3' del gen de longitud completa. Después se generó ADNc de longitud completa por PCR usando cebadores específicos para extremos 5' y 3' del gen por PCR de extremo a extremo.

Ejemplo II: construcción de vectores de expresión

La secuencia se amplificó a partir de una biblioteca genómica o de ADNc usando cebadores específicos para secuencias cadena arriba y cadena abajo de la región codificante. El vector de expresión fue pMEN20 o pMEN65, que derivan ambos de pMON316 (Sanders *et al.* (1987) Nucleic Acids Res. 15: 1543-1558) y contienen el promotor 35 S de CaMV para expresar transgenes. Para clonar la secuencia en el vector, se digirieron tanto pMEN20 como el fragmento de ADN amplificado por separado con enzimas de restricción Sall y NotI a 37 °C durante 2 horas. Los productos de digestión se sometieron a electroforesis en un gel de agarosa al 0,8 % y se visualizaron por tinción con bromuro de etidio. Los fragmentos de ADN que contenían la secuencia y el plásmido linealizado se escindieron y se purificaron usando un kit de extracción en gel QIAQUICK (Qiagen, Valencia, CA). Los fragmentos de interés se ligaron a una relación de 3:1 (vector e inserto). Se llevaron a cabo reacciones del ligamiento usando ADN ligasa T4 (New England Biolabs, Beverly MA) a 16 °C durante 16 horas. Los ADN ligados se transformaron en células competentes de la cepa de *E. coli* DH5α usando el método de choque térmico. Las transformaciones se sembraron en placas LB que contenían kanamicina 50 mg/l (Sigma Chemical Co. St. Louis MO). Las colonias individuales se cultivaron durante una noche en 5 ml de caldo de cultivo LB que contenía kanamicina 50 mg/l a 37 °C. Se purificó ADN plasmídico usando kits de Mini Prep Qiaquick (Qiagen, Valencia CA).

Para el sistema de dos componentes, se usan dos construcciones separadas: promotor p::LexA-GAL4TA y opLexA::factor de transcripción. El primero de estos (promotor::LexA-GAL4TA) comprendía un promotor deseado clonado delante de un dominio de unión a ADN LexA fusionado con un dominio de activación de GAL4. La cadena principal del vector de construcción (pMEN48, también conocido como P5375) también portaba un marcador de resistencia a kanamicina junto con un indicador opLexA::GFP. Se obtuvieron líneas transgénicas que contenían este primer componente y se seleccionó al menos una línea que mostraba expresión reproducible del gen indicador en el patrón deseado durante varias generaciones. Se estableció una población homocigota para esa línea y se

supertransformó la población (para producir una supertransformación de “supTfn”) portando la segunda construcción (opLexA:: factor de transcripción) el factor de transcripción de interés clonado detrás de un sitio operador LexA. Esta segunda cadena principal de vector de construcción (pMEN53, también conocido como P5381) también contenía un marcador de resistencia a sulfonamida.

5

Ejemplo III: transformación de *Agrobacterium* con el vector de expresión

Después de construirse el vector plasmídico que contenía el gen, el vector se usó para transformar células de *Agrobacterium tumefaciens* que expresaban los productos génicos. La reserva de células de *Agrobacterium tumefaciens* para transformación se realizó como se ha descrito en Nagel *et al.* (1990) FEMS Microbiol Letts. 67: 325-328. La cepa de *Agrobacterium* ABI se cultivó en 250 ml de medio LB (Sigma) durante una noche a 28 °C con agitación hasta que se alcanzó una absorbancia sobre 1 cm a 600 nm (A_{600}) de 0,5 - 1,0. Las células se recogieron por centrifugación a 4.000 x g durante 15 minutos a 4 °C. Después se resuspendieron las células en 250 µl de tampón helado (HEPES 1 mM, pH ajustado a 7,0 con KOH). Las células se centrifugaron de nuevo como se ha descrito anteriormente y se resuspendieron en 125 µl de tampón helado. Las células se centrifugaron después y se resuspendieron dos veces más en el mismo tampón de HEPES como se ha descrito anteriormente a un volumen de 100 µl y 750 µl, respectivamente. Después se distribuyeron células resuspendidas en alícuotas de 40 µl, se congelaron rápidamente en nitrógeno líquido y se almacenaron a -80 °C.

Las células de *Agrobacterium* se transformaron con plásmidos preparados como se ha descrito anteriormente siguiendo el protocolo descrito en Nagel *et al.* 1990) mencionado anteriormente. Para cada construcción de ADN para transformar, se mezclaron 50 - 100 ng de ADN (generalmente resuspendidos en Tris-HCl 10 mM, EDTA 1 mM, pH 8,0) con 40 µl de células de *Agrobacterium*. La mezcla de ADN/células se transfirió después a una cubeta helada con un hueco de electrodo de 2 mm y se sometió a una carga de 2,5 kV disipada a 25 µF y 200 µF usando un aparato Gene Pulser II (Bio-Rad, Hercules, CA). Después de electroporación, las células se resuspendieron inmediatamente en 1,0 ml de LB y se permitió que se recuperaran sin selección de antibióticos durante 2 - 4 horas a 28 °C en un incubador de agitación. Después de la recuperación, las células se sembraron en placas en medio selectivo de caldo LB que contenía espectinomicina 100 µg/ml (Sigma) y se incubaron durante 24-48 horas a 28 °C. Después se seleccionaron colonias individuales y se inocularon en medio nuevo. La presencia de la construcción plasmídica se verificó por amplificación por PCR y análisis de secuencia.

Ejemplo IV: transformación de plantas de *Arabidopsis* con *Agrobacterium tumefaciens* con vector de expresión

Después de la transformación de *Agrobacterium tumefaciens* con vectores plasmídicos que contenían el gen, se identificaron colonias individuales de *Agrobacterium*, se propagaron y se usaron para transformar plantas de *Arabidopsis*. Brevemente, se inocularon cultivos de 500 ml de medio LB que contenía kanamicina 50 mg/l con las colonias y se cultivaron a 28 °C con agitación durante 2 días hasta que se alcanzó una absorbancia óptica a longitud de onda de 600 nm sobre 1 cm (A_{600}) de > 2,0. Después se recogieron las células por centrifugación a 4.000 x g durante 10 minutos, y se resuspendieron en medio de infiltración (sales de Murashige y Skoog 1/2 X (Sigma), vitaminas B-5 de Gamborg 1 X (Sigma), sacarosa 5,0 % (p/v) (Sigma), bencilamino purina 0,044 µM (Sigma), L-77 Silwet 200 µl/l (Lehle Seeds)) hasta que se alcanzó una A_{600} de 0,8.

Antes de la transformación, se sembraron semillas de *Arabidopsis thaliana* (ecotipo Columbia) a una densidad de ~10 plantas por maceta de 10,16 cm en medio de siembra Pro-Mix BX (Hummert International) cubierto con malla de fibra de vidrio (18 mm X 16 mm). Las plantas se cultivaron con iluminación continua (50-75 µE/m²/segundo) a 22-23 °C con humedad relativa del 65-70 %. Después de aproximadamente 4 semanas, los tallos de inflorescencia primarios (espigamientos) se cortan para promover el crecimiento de múltiples espigamientos secundarios. Después del florecimiento de los espigamientos secundarios maduros, las plantas se prepararon para transformación por retirada de todas las silicuas y flores abiertas.

Las macetas se sumergieron después boca abajo en la mezcla de medio de filtración de *Agrobacterium* como se ha descrito anteriormente durante 30 segundos, y se colocaron en sus laterales para permitir el secado en una superficie plana de 30 cm x 60 cm cubierta con un envoltorio plástico. Después de 24 horas, el envoltorio de plástico se retiró y las macetas se pusieron en posición vertical. El procedimiento de inmersión se repitió una semana después, para un total de dos inmersiones por maceta. Las semillas se recogieron después de cada maceta de transformación y se analizaron siguiendo el protocolo descrito posteriormente.

Ejemplo V: identificación de transformantes primarios de *Arabidopsis*

Se esterilizaron semillas recogidas de las macetas de transformación esencialmente de la siguiente manera. Las semillas se dispersaron en una solución que contenía Tritón X-100 0,1 % (v/v) (Sigma) y agua estéril y se lavaron por agitación de la suspensión durante 20 minutos. La solución de lavado se secó después y se reemplazó con solución de lavado nueva para lavar las semillas durante 20 minutos con agitación. Después de la retirada de la solución de etanol/detergente, se añadió una solución que contenía Tritón X-100 0,1 % (v/v) y lejía 30 % (v/v)

(CLOROX; Clorox Corp. Oakland CA) a las semillas, y la suspensión se agitó durante 10 minutos. Después de la retirada de la solución de lejía/detergente, las semillas se lavaron después cinco veces en agua destilada estéril. Las semillas se almacenaron en el último agua de lavado a 4 °C durante 2 días en oscuridad antes de sembrarse en placas en medio de selección antibiótico (sales de Murashige y Skoog 1 X (pH ajustado a 5,7 con KOH 1 M), vitaminas B-5 de Gamborg 1 X, fitagar 0,9 % (Life Technologies) y kanamicina 50 mg/l). Las semillas se germinaron bajo iluminación continua (50-75 $\mu\text{E}/\text{m}^2/\text{segundo}$) a 22-23 °C. Después de 7-10 días de cultivo en estas condiciones, eran visibles y se obtuvieron transformantes primarios resistentes a kanamicina (generación de T₁). Estas plántulas se transfirieron en primer lugar a placas de selección nuevas en las que las plántulas continuaron creciendo durante 3-5 días más, y después a suelo (medio para maceta Pro-Mix BX).

Se cruzaron transformantes primarios y se recogieron semillas descendientes (T₂); se seleccionaron y analizaron plántulas resistentes a kanamicina. Los niveles de expresión de los polinucleótidos recombinantes en los transformantes variaron de un aumento del nivel de expresión de aproximadamente 5 % a un aumento del nivel de expresión de al menos 100 %. Se realizan observaciones similares con respecto a la expresión del nivel de los polipéptidos.

Ejemplo VI: identificación de fenotipos modificados en plantas con sobreexpresión

En algunos casos, los patrones de expresión de los genes inducidos por estrés pueden controlarse por experimentos de micromatrices. En estos experimentos, se generan ADNc por PCR y se resuspenden a una concentración final de ~100 ng/ μl en SSC 3X o Na-fosfato 150 mM (Eisen y Brown (1999) *Methods Enzymol.* 303: 179-205). Los ADNc se aplican puntualmente en portaobjetos de vidrio de microscopio recubiertos con polilisina. Los ADNc preparados se siembran en alícuotas en placas de 384 pocillos y se aplican puntualmente a los portaobjetos usando, por ejemplo, un caballete x-y-z (OmniGrid) que puede obtenerse de GeneMachines (Menlo Park, CA) equipado con tornillos de tipo pluma que pueden obtenerse de Telechem International (Sunnyvale, CA). Después de la aplicación puntual, las matrices se curan durante un mínimo de una semana a temperatura ambiente, se rehidratan y se bloquean siguiendo el protocolo recomendado por Eisen y Brown (1999; mencionado anteriormente).

Se marcan muestras de ARN total de muestra (10 μg) usando colorantes Cy3 y Cy5 fluorescentes. Las muestras marcadas se resuspenden en SSC 4X/SDS 0,03 %/4 μg de ADN de esperma de salmón/2 μg de ARNt/Na-pirofosfato 50 mM, se calientan a 95 °C durante 2,5 minutos, se sedimentan por centrifugación y se colocan en la matriz. La matriz se cubre después con un cubreobjetos de vidrio y se coloca en una cámara sellada. La cámara se mantiene después en un baño de agua a 62 °C durante una noche. Las matrices se lavan como se ha descrito en Eisen y Brown (1999) mencionado anteriormente y se exploran en un explorador láser General Scanning 3000. Los archivos resultantes se cuantifican posteriormente usando software IMAGEGENE (BioDiscovery, Los Angeles CA).

Los fenotipos modificados observados para plantas con sobreexpresión particulares pueden incluir biomasa aumentada y/o tolerancia o resistencia al estrés abiótico aumentada o reducida. Para una planta con sobreexpresión particular que muestra una característica menos beneficiosa, tal como tolerancia o resistencia al estrés abiótico reducida, puede ser más útil seleccionar una planta con una expresión reducida del factor de transcripción particular. Para una anulación particular que muestra una característica menos beneficiosa, tal como tolerancia al estrés abiótico reducida, puede ser más útil seleccionar una planta con una expresión aumentada del factor de transcripción particular.

Los ensayos de germinación en este ejemplo siguieron modificaciones del mismo protocolo básico. Se sembraron semillas estériles en el medio condicional enumerado posteriormente. Las placas se incubaron a 22 °C bajo luz durante 24 horas (120-130 $\mu\text{Ein}/\text{m}^2/\text{s}$) en una cámara de cultivo. La evaluación de la germinación y el vigor de la plántula se realizaron de 3 a 15 días después de plantar. Los medios basales fueron medio de Murashige-Skoog 80 % (MS) + vitaminas.

Para experimentos de estrés realizados con plantas más maduras, las semillas se germinaron y cultivaron durante siete días en MS + vitaminas + sacarosa 1 % a 22 °C y después se transfirieron a condiciones de estrés por frío y calor. Las plantas se expusieron al estrés por frío (6 horas de exposición a 4-8 °C), o estrés por calor (se aplicaron 32 °C durante cinco días, después de lo cual las plantas se transfirieron de nuevo a 22 °C para recuperación y se evaluaron después de 5 días en relación con controles no expuestos a la temperatura deprimida o elevada).

Se pretendió que los ensayos de estrés por salinidad encontraran genes que confirieran mejor germinación, vigor de la plántula o crecimiento en alta salinidad. La evaporación de la superficie del suelo provoca movimiento de agua hacia arriba y acumulación de sal en la capa superior del suelo donde se colocan las semillas. Por lo tanto, la germinación normalmente tiene lugar a una concentración salina mucho mayor que la concentración salina media en el perfil de suelo completo. Las plantas difieren en su tolerancia a NaCl dependiendo de su estadio del desarrollo, por lo tanto se evaluaron la germinación de las semillas, el vigor de las plántulas y las respuestas de crecimiento de las plantas.

Se realizaron ensayos de estrés osmótico (incluyendo ensayos de NaCl y manitol) para determinar si un fenotipo de estrés osmótico era específico de NaCl o si era un fenotipo relacionado con estrés osmótico general. Las plantas tolerantes al estrés osmótico también podrían tener más tolerancia a sequía y/o congelación.

- 5 Para experimentos de germinación en estrés salino y osmótico, el medio se complementó con NaCl 150 mM o manitol 300 mM. Se realizaron ensayos de sensibilidad de regulador del crecimiento en medio MS, vitaminas y ABA 0,3 μ M, sacarosa 9,4 % o glucosa 5 %.

10 Se realizaron ensayos de sequía para encontrar genes que median en mejor supervivencia de las plantas después de privación de agua grave, a corto plazo. Se mide la filtración de iones si es necesario. Los resultados de tolerancia al estrés osmótico positiva también apoyan un fenotipo tolerante a sequía. Se realizaron exploraciones de sequía basados en suelo con plantas de *Arabidopsis* que sobreexpresaban los factores de transcripción enumerados en el Listado de Secuencias, donde se indica posteriormente. Se estratificaron semillas de plantas de *Arabidopsis* de tipo silvestre, o plantas que sobreexpresaban un polipéptido, durante tres días a 4 °C en agarosa 0,1 %. Se sembraron después catorce semillas de cada planta con sobreexpresión o de tipo silvestre en macetas de cerámica de 7,62 cm que contenían una mezcla 50:50 de vermiculita:perlita con una capa pequeña superior de MetroMix 200 y se cultivaron durante quince días con luz las 24 horas. Las macetas que contenían plántulas de tipo silvestre y con sobreexpresión se colocaron en pisos en orden aleatorio. Se inició el estrés por sequía colocando las macetas en papel absorbente durante de siete a ocho días. Se consideró que las plántulas tenían suficiente estrés cuando la mayoría de las macetas que contenían plántulas de tipo silvestre dentro de un piso se habían marchitado gravemente. Se volvieron a regar las macetas y se puntuó la supervivencia de cuatro a siete días después. Las plantas se clasificaron frente a controles de tipo silvestre para cada uno de dos criterios: tolerancia a las condiciones de sequía y recuperación (supervivencia) después de volver a regar. Al final del periodo de sequía inicial, se asignó a cada maceta una puntuación de valor numérico dependiendo de los criterios anteriores. Se asignó un valor bajo a las plantas con una apariencia extremadamente mala (es decir, las plantas eran uniformemente marrones) y se dio un valor alto a plantas que se clasificaron con apariencia muy sana (es decir, las plantas estaban completamente verdes). Después de volver a regar las plantas e incubarlas durante de cuatro a siete días adicionales, las plantas se reevaluaron para indicar el grado de recuperación del tratamiento con privación de agua. Se realizó después un análisis para determinar qué plantas sobrevivían mejor a la privación de agua, identificando los transgenes que conferían uniformemente fenotipos tolerantes a la sequía y su capacidad para recuperarse de este tratamiento. Los análisis se realizaron comparando tabulaciones generales y dentro del piso con un conjunto de modelos estadísticos para explicar las variaciones entre lotes. Se tabularon varias medidas de la supervivencia, incluyendo: (a) la proporción media de plantas que sobrevivían en relación con la supervivencia de tipo silvestre dentro del mismo piso; (b) la mediana de la proporción de supervivencia en relación con supervivencia de tipo silvestre dentro del mismo piso; (c) la supervivencia media general (tomando todos los lotes, pisos y macetas); (d) la supervivencia media general en relación con las supervivencia de tipo silvestre general; y (e) la puntuación visual media de la salud de la planta antes de volver a regar.

40 Se realizaron experimentos para identificar los transformantes que mostraban detección de azúcar modificada. Para dichos estudios, se germinaron semillas de transformantes en medio que contenía alto azúcar (glucosa 5 %, sacarosa 9,4 %) que normalmente restringe parcialmente la elongación del hipocótilo. Las plantas con detección de azúcar alterada pueden tener hipocótilos más largos o más cortos que las plantas normales cuando se cultivan en este medio. Adicionalmente, pueden variarse otros rasgos vegetales tales como la masa de la raíz. Se pretendió que los ensayos de detección de azúcar encontraran genes implicados en la detección de azúcar germinando semillas en altas concentraciones de sacarosa y glucosa y buscando grados de elongación del hipocótilo. El ensayo de germinación en manitol controló con respecto a respuestas relacionadas con el estrés osmótico. Los azúcares son moléculas reguladoras clave que afectan a diversos procesos en plantas superiores incluyendo germinación, crecimiento, floración, senescencia, metabolismo de los azúcares y fotosíntesis. La sacarosa es la principal forma de transporte de fotosintatos y se ha mostrado que su flujo a través de las células afecta a la expresión génica y altera la acumulación de compuestos de almacenamiento en las semillas (relaciones de fuente-sumidero). También se ha descrito la detección de hexosa específica de glucosa en plantas y está implicada en la división celular y represión de genes de "hambruna" (ciclos fotosintéticos o de glioxilato).

55 Se llevaron a cabo ensayos de estrés por temperatura para encontrar genes que conferían mejor germinación, vigor de la plántula o crecimiento de la planta con estrés por temperatura (frío, congelación y calor). Se llevaron a cabo experimentos de germinación en frío con estrés por temperatura a 8 °C. Se realizaron experimentos de germinación con estrés por calor de 32 °C a 37 °C durante 6 horas de exposición.

60 Se midió el tiempo de floración por el número de hojas de roseta presentes cuando era evidente una inflorescencia visible de aproximadamente 3 cm. El número de hojas totales y de la roseta en el tallo descendiente están estrechamente correlacionados con el momento de la floración (Koornneef *et al.* (1991) Mol. Gen. Genet. 229: 57-66). También se midió la respuesta de vernalización. Para tratamientos de vernalización, se sembraron semillas en placas de agar MS, se sellaron con una cinta microporosa, y se colocaron en una habitación fría a 4 °C con niveles de luz bajos durante 6-8 semanas. Las placas se transfirieron después a las habitaciones de cultivo junto con placas que contenían controles no verbalizados recién sembrados. Se contaron las hojas de las rosetas cuando fue evidente una inflorescencia visible de aproximadamente 3 cm.

Las secuencias de factores de transcripción del Listado de Secuencias, o las de las presentes Tabla o Figuras, y sus equivalentes, pueden usarse para preparar plantas transgénicas y plantas con rasgos alterados. Las plantas transgénicas específicas enumeradas posteriormente se producen a partir de las secuencias del Listado de Secuencias, como se ha indicado.

Ejemplo VII: genes que confieren mejoras significativas a plantas

Este ejemplo proporciona pruebas experimentales de biomasa aumentada y tolerancia al estrés abiótico controlada por polipéptidos de factores de transcripción y polipéptidos de la presente divulgación.

Se realizaron experimentos para identificar los transformantes que mostraban una diferencia morfológica en relación con plantas de control de tipo silvestre, es decir, una estructura y/o características del desarrollo modificadas. Para dichos estudios, los transformantes se observaron a simple vista para identificar nuevas características estructurales o del desarrollo asociadas con la expresión ectópica de los polinucleótidos o polipéptidos de la invención. Se indican posteriormente ejemplos de genes y equivalentes que confieren mejoras significativas a plantas con sobreexpresión. También se presentan observaciones experimentales realizadas con respecto a genes específicos cuya expresión se ha modificado en plantas con sobreexpresión, y aplicaciones potenciales basadas en estas observaciones.

Las secuencias de factores de transcripción del Listado de Secuencias pueden usarse para preparar plantas transgénicas con rasgos alterados. A partir de los resultados experimentales de los ensayos de fisiología basados en placa presentados en las tablas del presente ejemplo, se puede inferir que un número representativo de secuencias de diversas especies vegetales proporcionaba tolerancia al aumento del estrés en una serie de ensayos de estrés abiótico. También se indican efectos observados de la sobreexpresión en el tiempo de floración en el texto posterior. Estos efectos comparables indican que las secuencias halladas dentro del clado G1073 de polipéptidos de factores de transcripción están relacionadas funcionalmente y pueden usarse para conferir diversos tipos de tolerancia al estrés abiótico en plantas. Varios de estos genes confieren simultáneamente biomasa aumentada y tolerancia aumentada a múltiples tensiones abióticas. Como se demuestra por las tablas de este ejemplo, no todos los transformantes son tolerantes al estrés abiótico, y no todas las líneas fueron tolerantes en el mismo grado. Es generalmente ventajoso usar métodos presentados en el Ejemplo VI para seleccionar las líneas con sobreexpresión más tolerantes y útiles.

Resultados:

Varios miembros representativos del clado G1073 de polipéptidos de factores de transcripción, incluyendo las secuencias de *Arabidopsis* G1073, G1067, G1069, G2153, G2156, G2789, secuencias de arroz G3399, G3400, G3401, G3406, G3407, y secuencias de soja G3456, G3459 y G3460, aumentaron la tolerancia al estrés abiótico cuando se sobreexpresaron estas secuencias. Varias secuencias de miembros del clado también aumentaron la biomasa, como en plantas que sobreexpresaban las secuencias de *Arabidopsis* G1073, G1067, G1069, G1667, G2153, G2156, G2157, G2789, secuencias de arroz G3399, G3400, G3401, G3406, G3407, G3556, y secuencias de soja G3456, G3459 y G3460.

G1073 (SEC ID N°: 1 y 2)

Los inventores han demostrado previamente que la sobreexpresión de G1073 proporciona tolerancia a la sequía y producción potenciada en líneas 35S::G1073. Los inventores han diseñado ahora este locus como *HERCULES 1* (*HRC1*).

El objetivo de este estudio era reevaluar las líneas 35S::G1073 y comparar sus efectos de sobreexpresión con los de sus parálogos y ortólogos potenciales. Los inventores también buscaron ensayar si el uso de un sistema de sobreexpresión de dos componentes produciría cualquier refuerzo del fenotipo en relación con el uso de una fusión de promotor directo 35S.

La sobreexpresión de G1073 mediante el sistema de dos componentes dio como resultado fenotipos similares a los observados previamente con experimentos de fusión de promotor directo realizados previamente. En ambos proyectos, las plantas con sobreexpresión de G1073 mostraron un aumento de la biomasa en relación con plantas de control de tipo silvestre junto con cambios en la morfología de las hojas y un retardo de leve a moderado en el tiempo de floración.

Las plantas transgénicas que sobreexpresaban G1073 eran sustancialmente mayores que los controles de tipo silvestre, con al menos un 60 % de aumento de la biomasa (Figuras 6A y 6B, 7A y 7B). La masa aumentada de plantas transgénicas 35S::G1073 se atribuyó al agrandamiento de múltiples tipos de órganos incluyendo tallos, raíces y órganos florales; aparte de las diferencias de tamaño, estos órganos no se vieron afectados en su morfología general. Las plantas 35S::G1073 mostraron un aumento de la anchura (pero no longitud) de órganos de hojas maduros, produjeron 2-3 hojas más de roseta, y tuvieron hojas caulinares agrandadas en comparación con las hojas de tipo silvestre correspondientes. La sobreexpresión de G1073 dio como resultado un aumento tanto de la masa de la hoja como del área de la hoja por planta, y la morfología de la hoja (las plantas con sobreexpresión de

- 5 G1073 tendían a producir hojas más serradas). Los inventores también descubrieron que la masa de la raíz aumentaba en las plantas transgénicas, y que los órganos florales también se agrandaban (Figura 7B). Se observó un aumento de aproximadamente el 40 % en el diámetro del tallo en las plantas transgénicas. Las imágenes de las secciones transversales del tallo de plantas 35S::G1073 revelaron que las células corticales son grandes y que los haces vasculares contenían más células en el floema y xilema en relación con los controles de tipo silvestre (Figuras 6A y 6B). El tamaño de los pétalos en las líneas 35S::G1073 aumentó en el 40-50 % en comparación con controles de tipo silvestre. Las células epidérmicas de los pétalos en las mismas líneas fueron aproximadamente 25-30 % mayores que las de las plantas de control. Además, se produjeron 15-20 % más células epidérmicas por pétalo en comparación con controles de tipo silvestre. Por lo tanto, en pétalos y tallos, el aumento del tamaño se asoció con un aumento del tamaño celular así como del número de células.
- 10 La producción de semillas también aumentó en comparación con las plantas de control. Las líneas 35S::G1073 mostraron un aumento de al menos el 70 % en la producción de semillas. Esta producción de semillas aumentada se asoció con un número aumentado de silicuas por planta, más que de semillas por silicua.
- 15 Las líneas de dos componentes 35S::G1073 mostraron un retardo de leve a moderado en la aparición de la floración y desarrollaron hojas mayores más anchas que las de los controles de tipo silvestre. Estos efectos fueron de penetración inmediata, observándose en diversos grados en ocho de veinte líneas T1.
- 20 G1073 actúa tanto en soja como en tomate para aumentar la biomasa. En la Figura 9A, la planta de soja mayor a la derecha sobreexpresa G1073. Las hojas de tomate de varias líneas con sobreexpresión de G1073 fueron mucho mayores que las de las plantas de tomate de tipo silvestre, como se ve en la Figura 9B comparando las hojas de la planta con sobreexpresión a la izquierda con las de una planta de tipo silvestre a la derecha.
- 25 Como se ve en la tabla posterior, las líneas de dos componentes 35S::G1073 también presentaron una tolerancia notablemente aumentada a altos niveles de sal y sacarosa durante la germinación. Los estudios con fusión de promotor directo 35S dieron como resultado plantas con tolerancia a la sequía significativamente mayor que el tipo silvestre en ensayos basados en suelo.
- 30 Los inventores han obtenido fenotipos morfológicos y/o fisiológicos similares a partir de la sobreexpresión de los genes de *Arabidopsis* relacionados con G1073 (G1067, G1069, G1667, G2153, G2156, G2157, G2789), arroz (G3399, G3400, G3401, G3406, G3407, G3556) y genes de soja (G3456, G3459, G3460), lo que indica que estos genes probablemente estén funcionalmente relacionados.

Tabla 6. G1073 de *Arabidopsis thaliana* Supertransformación (supTfn) de 2 componentes de 35S

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
supTfn de 2 componentes	301	++	wt	++	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	304	++	wt	++	wt	+	wt	wt	wt	wt
supTfn de 2 componentes	305-	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	306	++	wt	++	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	308	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	310	+	wt	++	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	311	++	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	314	+	wt	++	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	319	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	320	+	wt	wt	wt	wt	wt	wt	wt	wt
+ más tolerante que las plantas de control de tipo silvestre ++ mucho más tolerante que las plantas de control de tipo silvestre										

Ejemplo VIII. Mitigación de efectos morfológicos indeseables por la sobreexpresión de polipéptidos del clado GH1073.

Los resultados de estrés abiótico mostrados anteriormente proporcionan pruebas de que los miembros del clado G1073 de polipéptidos de factores de transcripción pueden usarse para crear plantas con las características de producción, rendimiento y/o distribución mejorados. Sin embargo, la sobreexpresión de estos miembros del clado puede proporcionar efectos morfológicos no deseados, incluyendo menor tamaño de las plantas. Esto se ha observado con muchas de, pero no todas, las líneas generadas en el presente estudio. Ya que es con frecuencia deseable generar plantas con estatura normal o casi normal, una reducción o eliminación de otras características morfológicas proporcionadas por la sobreexpresión de un miembro del clado G1073 bajo el control regulador de un promotor constitutivo puede no ser siempre el mejor enfoque para mejorar la tolerancia al estrés.

Este presente estudio también incluyó una investigación en el uso de sistemas de sobreexpresión de dos componentes o promotor alternativos para el fin de conferir tolerancia al aumento del estrés y eliminar anomalías del desarrollo tales como tamaño reducido que se han asociado con la sobreexpresión constitutiva de G1073. A este respecto, la presente invención también se refiere a métodos y composiciones para producir plantas transgénicas con rendimiento ante estrés mejorado conseguido alterando la expresión de G1073 y secuencias relacionadas con combinaciones de gen-promotor específicas u otros medios reguladores. Estas combinaciones pueden regular los patrones de expresión del factor de transcripción de una manera transitoria, inducible o específica de órgano o tejido. Como se muestra posteriormente, este enfoque puede usarse para generar plantas que son morfológicamente similares a las plantas de control de tipo silvestre que no se han transformado con un polinucleótido que codifica G1073 o un equivalente. Por lo tanto, el tipo de elemento regulador usado para controlar la regulación del gen del factor de transcripción puede usarse para aliviar anomalías del desarrollo indeseables o efectos morfológicos adversos que resultarían de otro modo por la sobreexpresión de los mismos genes de factores de transcripción con un promotor constitutivo tal como el promotor 35S.

ARSK1::G1073 de raíz (*Arabidopsis*)

Los inventores han obtenido líneas ARSK1::G1073 usando un enfoque de dos componentes; no fueron evidentes efectos uniformes en la morfología entre estos transformantes y no se observaron alteraciones en el tamaño de las hojas. Por lo tanto, o bien la expresión del promotor ARSK1 era demasiado débil o la expresión en la raíz no era suficiente para desencadenar las alteraciones en el tamaño de las hojas que eran evidentes en las líneas 35S::G1073.

Resulta interesante que, aunque las líneas ARSK1::G1073 no mostraron cambios morfológicos claros, cinco de diez de estas líneas sí mostraron tolerancia potenciada a cloruro sódico en un ensayo de germinación basado en placa. Otras dos líneas rindieron mejor que los controles de tipo silvestre en un ensayo de germinación en frío. Estos fenotipos de tolerancia al estrés osmótico son de interés particular, ya que muestran que G1073 puede proporcionar tolerancia al estrés independientemente de cambios en el tamaño de los órganos. Adicionalmente, ya que ARSK1 no se expresa de forma significativa en el tejido de los brotes, los resultados sugieren que la expresión de G1073 no se requiere en el brote para conseguir tolerancia al estrés.

Sumario de la morfología

Las líneas de *Arabidopsis* en las que se expresó G1073 a partir del promotor ARSK1 (mediante el sistema de dos componentes) no presentaron diferencias uniformes en la morfología en comparación con los controles.

Se examinaron veinte líneas T1 (341-360); se observó que tres líneas (Nº 342, 346, 357) eran ligeramente pequeñas y de desarrollo lento. Sin embargo el resto de las líneas mostraron morfología de tipo silvestre en todos los estadios.

De las líneas presentadas para ensayos fisiológicos, todas excepto la línea 556 mostraron segregación en placas de selección en la generación T2 que era compatible con el transgén que estaba presente en un único locus. La línea 556 mostró segregación que era compatible con inserciones en múltiples loci.

Sumario de la fisiología

Las plántulas de cinco líneas ARSK1::G1073 tenían más vigor de la plántula cuando germinaron en placas que contenían cloruro sódico. Las plántulas de otras dos líneas rindieron mejor que los controles de tipo silvestre en un ensayo de germinación en frío, y dos líneas rindieron mejor en un ensayo de sequía.

Tabla 19. G1073 (*Arabidopsis*) - resultados del ensayo de estrés abiótico de RSK1 de raíz

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frio	Crecimiento con Calor	Desecación	Crecimiento con Frio
sup Tfn de 2 componentes	342	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	344	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	345	+	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	346	wt	wt	wt	wt	wt	+	wt	wt	wt
sup Tfn de 2 componentes	347	wt	wt	wt	wt	wt	+	wt	wt	wt
sup Tfn de 2 componentes	351	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	354	+	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	355	+	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	356	+	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	359	+	wt	wt	wt	wt	wt	wt	wt	wt

CUT1::G1073 epidérmico (*Arabidopsis*)

Los inventores han obtenido líneas CUT1::G1073 usando un enfoque de dos componentes; no fueron evidentes efectos uniformes en la morfología entre estos transformantes y no se observaron alteraciones en el tamaño de las hojas. Por lo tanto, o bien la expresión del promotor CUT1 era demasiado débil o bien la expresión epidérmica no era suficiente para desencadenar las alteraciones en el tamaño de las hojas que son evidentes en las líneas 35S::G1073.

Aunque las líneas CUT1::G1073 no mostraron cambios morfológicos claros, tres de diez de estas líneas sí mostraron tolerancia potenciada a cloruro sódico en un ensayo de germinación basado en placa. Dos de estas líneas también rindieron mejor que los controles de tipo silvestre en un ensayo de germinación de sacarosa, mientras que la tercera línea germinó mejor que los controles de tipo silvestre en medio de manitol. Una cuarta línea CUT1::G1073 dio un resultado positivo en el ensayo de sacarosa solamente. Aunque estos fenotipos de tolerancia al estrés osmótico se vieron en un número de líneas relativamente pequeño, son de interés particular, ya que sugieren que G1073 puede proporcionar tolerancia al estrés independientemente de cambios en el tamaño de los órganos. Adicionalmente, la línea conductora CUT1 no proporciona expresión significativa en la raíz, lo que sugiere que la expresión de G1073 no es necesaria en la raíz para conseguir dicha tolerancia.

Sumario de la morfología

Se han generado ahora líneas de *Arabidopsis* que expresan G1073 a partir del promotor CUT1 (usando el sistema de dos componentes; CUT1::LexA; opLexA::G1073). Se obtuvieron un total de diecinueve líneas (381-399). Fue evidente alguna variación de tamaño en los estadios tempranos del crecimiento, pero en general, las plantas no mostraron diferencias uniformes en la morfología con respecto a los controles.

De las líneas presentadas para ensayos fisiológicos, las siguientes mostraron una segregación en placas de selección en la generación T2 que era compatible con que el transgén estuviera presente en un único locus: 384, 391, 392, 394, 396. Las líneas 381, 390, 393, 395, 397 mostraron segregación que era compatible con inserciones en múltiples loci.

Sumario de la fisiología (ensayos en placa)

Tres líneas CUT1::G1073 mostraron aumento del vigor de las plántulas cuando germinaron en placas que contenían cloruro sódico. De estas tres líneas, las plántulas de dos líneas también rindieron mejor que los controles de tipo silvestre cuando germinaron en sacarosa mientras que las plántulas de la tercera línea tuvieron mejor vigor cuando germinaron en placas que contenían manitol. Una cuarta línea mostró un mejor rendimiento en un ensayo de germinación de sacarosa.

Tabla 20. G1073 (*Arabidopsis*) - resultados del ensayo de estrés abiótico de CUT1 epidérmico

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
sup Tfn de 2 componentes	381	wt	wt	+	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	384	wt	wt	wt	wt	wt	wt	wt	wt	wt
2-componentes	390	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	391	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	392	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	393	+ + +	++	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	394	+	wt	+	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	395	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	396	+	wt	+	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	397	wt	wt	wt	wt	wt	wt	wt	wt	wt

RBCS3::G1073 de hoja (*Arabidopsis*)

Los inventores han obtenido líneas RBCS3::G1073 del promotor específico de tejido (hoja) usando un enfoque de dos componentes.

5

Sumario de la morfología

Las líneas 541 y 542 pueden haber sido marginalmente tardías pero no mostraron de otro modo ninguna diferencia morfológica obvia en relación con los controles de tipo silvestre.

10

Las líneas 961-973 fueron ligeramente más lentas en el crecimiento que los controles de tipo silvestre, pero fueron de otro modo morfológicamente similares a los controles.

Sumario de la fisiología

15

Más notablemente, las plántulas de estas plantas con sobreexpresión mostraron tolerancia aumentada a las tensiones osmóticas de sal y calor en ensayos de germinación. Dos líneas mostraron tolerancia aumentada al frío en ensayos de crecimiento.

Tabla 22. G1073 (*Arabidopsis*) - resultados del ensayo de estrés abiótico de RBCS3 de hoja

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
sup Tfn de 2 componentes	541	wt	wt	wt	+	wt	wt	+	wt	+
sup Tfn de 2 componentes	542	wt	wt	wt	wt	wt	wt	+	wt	wt
sup Tfn de 2 componentes	961	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	962	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	965	wt	wt	wt	wt	wt	wt	+	wt	+
sup Tfn de 2 componentes	966	wt	wt	wt	wt	wt	wt	+	wt	wt
sup Tfn de 2 componentes	967	+	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	968	+	wt	wt	wt	wt	wt	+	wt	wt
sup Tfn de 2 componentes	969	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	973	+	wt	wt	wt	wt	wt	wt	wt	wt

Superactivación (N-GAL4-TA) de G1073 (*Arabidopsis*)

Los inventores han aislado ahora líneas que sobreexpresan una versión de la proteína G1073 que tiene un dominio de activación de GAL4 fusionado con el extremo N terminal.

Sumario de la morfología

En su mayor parte, las líneas eran morfológicamente indistinguibles de los controles de tipo silvestre, teniendo varias líneas fisiologías normales o casi normales. Sin embargo, hubo un número pequeño de plantas que mostraron floración retardada y cambios en el tamaño de las hojas. Además, se observó que algunas eran de color oscuro.

Se generaron tres lotes de líneas que sobreexpresaban una forma superactiva de G1073 que comprendía un dominio de transactivación GAL4 fusionado con el extremo N terminal de la proteína: líneas 841-852, 981-991 y 1441-1460.

La mayoría de las plantas en cada una de las plantaciones anteriores parecían de tipo silvestre; sin embargo fueron evidentes floración retardada y cambios en el tamaño de las hojas en un número pequeño de las líneas en cada conjunto. Las plantas que mostraban este fenotipo florecieron hasta 3-4 semanas después de los controles de tipo silvestre (con luz las 24 horas), eran de color oscuro, y tenían hojas que se habían rizado y enrollado (particularmente en los estadios tardíos del ciclo de vida).

Este fenotipo anterior se observó con las siguientes frecuencias:

- 3/12 líneas (846, 851, 852) del conjunto 841-852
- 2/11 líneas (983, 989) del conjunto 981-991
- 7/20 líneas (1442, 1443, 1449, 1452, 1453, 1454, 1455) del conjunto 1441-1460.

De las plantas en este conjunto final, sin embargo, solamente la N° 1442 mostró un fuerte fenotipo mientras que las otras presentaron efectos relativamente leves.

Es quizás notable que además de los efectos en el tiempo de floración y desarrollo de las hojas, se observó que un número pequeño de las líneas T1 obtenidas en el segundo lote (981-991) estaban más avanzadas que los controles de tipo silvestre en el estado de día 7. Sin embargo, este efecto no se observó en la descendencia de T2 de ninguna de estas líneas o en ninguno de los otros dos conjuntos de plantas T1.

Sumario de la fisiología

Las plántulas de dos de estas líneas superactivadas mostraron tolerancia aumentada a ABA, y germinaron en placas en condiciones frías. Las plántulas de dos líneas también rindieron mejor que los controles de tipo silvestre en un ensayo de sequía basado en placa. Otras líneas mostraron un mejor rendimiento en un ensayo de estrés osmótico basado en sacarosa, o en ensayos de cultivo en condiciones de frío o calor.

Tabla 23. G1073 (*Arabidopsis*) - resultados del ensayo de estrés abiótico de superactivación (N-GAL4-TA)

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
N terminal de GAL4	842	wt	wt	wt	wt	wt	wt	+	wt	wt
N terminal de GAL4	843	wt	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	849	wt	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	850	wt	wt	wt	wt	wt	wt	wt	+	+
N terminal de GAL4	851	wt	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	981	wt	wt	+	+	wt	wt	wt	wt	wt
N terminal de GAL4	983	wt	wt	wt	+	wt	wt	wt	+	wt
N terminal de GAL4	984	wt	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	985	wt	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	986	wt	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	989	wt	wt	wt	+	wt	wt	wt	wt	wt

Superactivación (C-GAL4-TA) de G1073 (*Arabidopsis*)

Los inventores han aislado ahora líneas que sobreexpresan una versión de la proteína G1073 que tiene un dominio de activación de GAL4 fusionado con el extremo C terminal.

5

Sumario de la morfología

En diversos estadios del crecimiento, algunas de las plantas con un dominio de activación de GAL4 fusionado con el extremo C terminal eran algo pequeñas. Sin embargo, muchas de las plantas que eran más tolerantes a las tensiones abióticas, indicadas en la tabla posterior, eran solamente ligeramente menores que los controles de tipo silvestre en algunos estadios del crecimiento, y muchas de las líneas eran morfológicamente muy similares a plantas de control de tipo silvestre.

10

Sumario de la fisiología

La mayoría de las líneas superactivadas C-GAL4 ensayadas eran más tolerantes al estrés osmótico en un ensayo de desecación grave basado en placa que las plantas de control de tipo silvestre. Dos líneas eran más tolerantes a alto manitol.

15

Tabla 24. G1073 (*Arabidopsis*) - resultados del ensayo de estrés abiótico de superactivación (C-GAL4-TA)

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frio	Crecimiento con Calor	Desecación	Crecimiento con Frio
C terminal de GAL4	1542	wt	wt	wt	wt	wt	wt	wt	+	wt
C terminal de GAL4	1543	wt	wt	wt	wt	wt	wt	wt	wt	wt
C terminal de GAL4	1544	wt	wt	wt	wt	wt	wt	wt	+	wt
C terminal de GAL4	1545	wt	wt	wt	wt	wt	wt	wt	wt	wt
C terminal de GAL4	1546	wt	wt	wt	wt	wt	wt	wt	+	wt
C terminal de GAL4	1547	wt	wt	wt	wt	wt	wt	wt	+	wt
C terminal de GAL4	1551	wt	+	wt	wt	wt	wt	wt	+	wt
C terminal de GAL4	1552	wt	+	wt	wt	wt	wt	wt	+	wt

ARSK1::G1067 de raíz (*Arabidopsis*)

Los inventores han obtenido líneas ARSK1::G1067 específicas de tejido (raíz) usando un enfoque de dos componentes.

- 5 La mayoría (18 de 26) de estos transformantes parecieron de tipo silvestre, y no presentaban pruebas de hojas rizadas ni enanismo grave. Ocho de 26 líneas mostraron reducciones del tamaño y se desarrollaron más lentamente que los controles, en varios grados.
- 10 En ensayos de estrés basados en placas, cuatro de diez líneas ARSK1::G1067 mostraron tolerancia potenciada en un ensayo de deshidratación grave. Estas cuatro líneas habían mostrado un fenotipo de tipo silvestre en las exploraciones morfológicas, demostrando que G1067 podría potenciar la tolerancia a la sequía sin producir efectos negativos evidentes en el tamaño de la planta. Otras tres líneas ARSK1::G1067 rindieron mejor que las plantas de control de tipo silvestre en un ensayo de germinación con alto NaCl.

Tabla 25. Resultados del ensayo de estrés abiótico de ARSK1::G1067

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
supTfn de 2 componentes	344	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	345	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	346	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	347	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	401	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	402	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	403	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	406	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	407	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	408	wt	wt	wt	wt	wt	wt	wt		wt

RBCS3::G1067 de hoja (*Arabidopsis*)

Los inventores han obtenido líneas RBCS3::G1067 del promotor específico de tejido (hoja) usando un enfoque de dos componentes.

5 Varias líneas RBCS3::G1067 producidas con un enfoque de dos componentes fueron generalmente pequeñas en estadios tempranos, tenían hojas redondeadas, cortas y florecieron ligeramente tarde. En estadios posteriores del crecimiento, las hojas se retorcieron y rizaron, pero en líneas ocasionales las hojas parecieron más anchas que las de los controles. La apariencia de hojas anchas, aunque a una frecuencia baja, sugiere que G1073 y G1067 podrían, al menos en algún grado, estar funcionalmente relacionadas.

Las líneas 581-590 mostraron un ligero retardo en la aparición de la floración (aproximadamente 1-5 días con luz las 24 horas).

15 En estadios tempranos del crecimiento, las líneas 621-629 fueron ligeramente pequeñas en estadios tempranos y tuvieron hojas cortas, redondas, más bien anchas. No se observó floración retardada en este conjunto de líneas. En el estadio de floración tardío, las líneas 621-629 no tuvieron diferencias morfológicas uniformes en relación con plantas de control de tipo silvestre, excepto por la líneas 622, 624, que tenían hojas planas, ligeramente anchas. En estadios posteriores del crecimiento no hubo diferencias uniformes en la morfología entre las líneas con sobreexpresión 621-629 y plantas de control de tipo silvestre.

20 Varias de estas líneas tuvieron mayor tolerancia al estrés que las plantas de control de tipo silvestre, como se ve en la siguiente tabla.

Tabla 26. Resultados del ensayo de estrés abiótico de RBCS3::G1067

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frio	Crecimiento con Calor	Desecación	Crecimiento con Frio
sup Tfñ de 2 componentes	581	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfñ de 2 componentes	582	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfñ de 2 componentes	586	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfñ de 2 componentes	587	+	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfñ de 2 componentes	588	+	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfñ de 2 componentes	622	+	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	624	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfñ de 2 componentes	627	+	wt	wt	wt	wt	wt	wt	wt	+
sup Tfñ de 2 componentes	628	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfñ de 2 componentes	629	wt	wt	wt	wt	wt	wt	wt	wt	wt

RD29A::G1067 inducible por estrés (*Arabidopsis*)

Los inventores han obtenido líneas RD29A::G1067 de promotor inducible por estrés usando un enfoque de dos componentes.

5 La mayoría de estas líneas RD29A::LexA;opLexA::G1067 en el fondo de línea RD29A 5 no mostraron alteraciones uniformes en la morfología en relación con los controles. Un menor número de los transformantes sí mostró una reducción pequeña en el tamaño y hojas ligeramente más redondeadas que los controles. Por lo tanto, en estas líneas, la expresión constitutiva baja producida por la línea conductora podría haber desencadenado dichos efectos.

10 Sin embargo, ninguna de las líneas mostró el enanismo extremo y hojas rizadas vistos en las líneas 35S::G1067.

Varias de estas líneas tenían mayor tolerancia al estrés que las plantas de control de tipo silvestre, particularmente en los ensayos de desecación grave basados en placa, como se ve en la siguiente tabla.

Tabla 27. Resultados del ensayo de estrés abiótico de RD29A::G1067

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
supTfu de 2 componentes	661	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	663	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	664	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	668	+	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	704	+	+	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	707	wt	wt	wt	wt	wt	wt	wt	wt	wt
N terminal de GAL4	708	wt	+	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	710	wt	+	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	711	wt	wt	+	+	wt	+	wt	+	wt
supTfn de 2 componentes	717	+	wt	wt	+	wt	wt	wt	+	wt

ARSK1::G2156 de raíz (*Arabidopsis*)

Los inventores han obtenido líneas ARSK1::G2156 de promotor específico de tejido (raíz) usando un enfoque de dos componentes.

5 Aproximadamente la mitad de las líneas de uno de estos lotes presentó un retardo muy marginal en la aparición de la floración, pero la mayoría de las líneas no presentaron diferencias evidentes en el crecimiento y desarrollo con los controles de tipo silvestre. Por lo tanto, el uso de un promotor de raíz en combinación con G2156 eliminó en gran medida las morfologías indeseables producidas por la sobreexpresión de ese gen.

10 Varias de estas líneas también tuvieron mayor tolerancia al estrés que la plantas de control de tipo silvestre, particularmente en ensayos de desecación grave basados en placas como se ve en la siguiente tabla.

Tabla 28. Resultados del ensayo de estrés abiótico de ARSK1::G2156

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
sup Tfn de 2 componentes	363	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	364	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	365	wt	wt	wt	wt	wt	wt	wt	+	wt
sup Tfn de 2 componentes	368	wt	wt	wt	wt	wt	wt	wt	+	wt
sup Tfn de 2 componentes	370	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	486	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	488	wt	wt	wt	wt	wt	wt	wt	+	wt
sup Tfn de 2 componentes	490	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	492	wt	wt	wt	wt	wt	wt	wt	wt	wt
sup Tfn de 2 componentes	494	wt	wt	wt	wt	wt	wt	wt	+	wt

RBCS3::G2156 de hojas (*Arabidopsis*)

Los inventores han obtenido líneas RBCS3::G2156 del promotor específico de tejido (hoja) usando un enfoque de dos componentes.

5 En estadios tempranos, estas plantas eran ligeramente pequeñas y mostraron hojas más bien redondeadas. Sin embargo, en estadios posteriores, el 50 % de las líneas desarrollaron hojas agrandadas y mostraron biomasa de la roseta aumentada en comparación con los controles. La mayoría de las líneas que mostraban este fenotipo también presentaron un ligero retardo en la aparición de la floración. Los inventores han observado previamente hojas grandes en plantas con sobreexpresión constitutiva de 35S::G2156. Sin embargo, los agrandamientos de hojas se vieron a menor frecuencia en el estudio de 35S::G2156 que en el estudio de RBCS3::G2156. Adicionalmente muchas de las líneas del experimento de 35S::G2156 fueron muy pequeñas y tuvieron múltiples defectos; estos efectos parecen haberse evitado mediante el uso del promotor de RBCS3. El tamaño de la hoja aumentado visto en el presente estudio fue comparable con los efectos producidos por el aumento de la expresión de G1073 y sirvió para fortalecer la conclusión de que los dos genes tienen papeles relacionados.

RBCS3 produce expresión en tejido de hoja fotosintético, relativamente maduro. Por lo tanto, G2156 cuando se expresó en un estadio relativamente tardío del desarrollo de las hojas produjo señales del desarrollo que mantenían el crecimiento de la hoja. Sin embargo, sigue existiendo la posibilidad de que G2156 desencadenara la producción de señales del desarrollo en hojas maduras que se transmitieron después a primordios de hojas más jóvenes, y los hizo crecer en exceso en un estadio temprano.

Varias de estas líneas tuvieron mayor tolerancia al estrés abiótico que los controles de tipo silvestre (Tabla 29).

Tabla 29. Resultados del ensayo de estrés abiótico de RBCS3::G2156

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frio	Crecimiento con Calor	Desecación	Crecimiento con Frio
supTfn de 2 componentes	543	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	544	wt	wt	wt	wt	wt	wt	wt	+	wt
supTfn de 2 componentes	551	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	552	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	553	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	554	wt	wt	wt	+	wt	wt	wt	wt	wt
supTfn de 2 componentes	557	wt	wt	wt	+	wt	wt	wt	wt	wt
supTfn de 2 componentes	582	wt	wt	wt	+	wt	wt	wt	wt	wt
supTfn de 2 componentes	584	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	587	wt	wt	wt	wt	wt	wt	wt	wt	wt

123D29A::G2356 inducible por estrés (*Arabidopsis*)

Los inventores han obtenido líneas RD29A::G2156 del promotor inducible por estrés usando un enfoque de dos componentes.

5 La mayoría de los transformantes RD29A::LexA;opLexA::G2156 de dos componentes en el fondo de línea RD29A 5 no mostraron diferencias uniformes en la morfología con los controles. Un número menor de líneas sí floreció ligeramente tarde y desarrolló hojas agrandadas más tarde en el desarrollo. Por lo tanto, en estas líneas, la expresión constitutiva baja producida por la línea conductora podría haber desencadenado dichos efectos.

10 Varias de estas líneas tuvieron mayor tolerancia al estrés que las plantas de control de tipo silvestre, como se ve en la siguiente tabla. Fueron particularmente notables los resultados obtenidos que mostraban que la mayoría de las líneas eran menos sensibles en el ensayo de germinación de ABA, lo que indica un fenotipo tolerante al estrés osmótico.

Tabla 30. Resultados del ensayo de estrés abiótico de RD29A::G2156

Tipo de proyecto	Línea	Germ en Alto NaCl	Germ en Alto Manitol	Germ en Alta Sacarosa	ABA	Germ con Calor	Germ con Frío	Crecimiento con Calor	Desecación	Crecimiento con Frío
supTfn de 2 componentes	622	wt	wt	wt	wt	wt	wt	-	wt	wt
supTfn de 2 componentes	624	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	625	wt	wt	wt	+	wt	wt	wt	wt	wt
supTfn de 2 componentes	626	wt	wt	wt	+	wt	wt	wt	wt	wt
supTfn de 2 componentes	628	wt	wt	wt	+	wt	wt	wt	wt	wt
supTfn de 2 componentes	686	+	wt	+	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	688	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	689	wt	wt	wt	wt	wt	wt	wt	wt	wt
supTfn de 2 componentes	690	wt	wt	wt	+	wt	wt	wt	wt	wt
supTfn de 2 componentes	692	wt	wt	+	++	wt	wt	wt	wt	wt

Utilidades para miembros del clado G1073 bajo control regulador no constitutivo

Los resultados de estos estudios con el control regulador no constitutivo de numerosos miembros del clado G1073 indican que las secuencias polinucleotídicas y polipeptídicas pueden usarse para mejorar la tolerancia al estrés relacionada con sequía manteniendo a la vez morfología normal o casi normal en condiciones sin estrés o de estrés bajo, y tamaño y vigor mejorados en relación con plantas de control de tipo silvestre en condiciones de estrés abiótico. Los datos también confirman las conclusiones de los inventores de que G1073 y otros miembros del clado G1073 pueden ser herramientas valiosas para el fin de aumentar la producción, biomasa y modificar el tiempo de floración.

Se realizó análisis de combinaciones de combinaciones de miembros del clado G1073 con elementos reguladores para 1) proporcionar información mecánica sobre la función de miembros del clado G1073, y 2) identificar patrones optimizados de la expresión de miembros del clado G1073. La expresión diferencial de G1073 y secuencias relacionadas ha revelado que puede obtenerse cierto grado de tolerancia al estrés osmótico sin un impacto significativo en el tamaño de la planta o los órganos. Los ejemplos específicos incluyen expresión con promotores específicos de tejido, incluyendo los promotores de CUT1 (específico de epidermis), ARSK1 (específico de raíz), RBCS3 (específico de hoja) y promotores inducibles por estrés, incluyendo el promotor RD29A. Las líneas que sobreexpresaban una forma superactiva de G1073 que comprendía un dominio de transactivación de GAL4 fusionado con el extremo N o C terminal del polipéptido también eran más tolerantes a las tensiones abióticas, y generalmente eran morfológicamente similares a las plantas de control de tipo silvestre. Estas combinaciones de elementos reguladores de factores de transcripción demuestran que pueden usarse miembros del clado G1073 específicos de tejido, inducibles y transactivados para proporcionar tolerancia al estrés abiótico con poco o ningún impacto en el crecimiento general de la planta o la producción en condiciones de estrés abiótico bajo, y mejoran significativamente la producción y el vigor en condiciones de estrés abiótico.

Ejemplo IX: identificación de secuencias homólogas por búsqueda de homología por ordenador

Este ejemplo describe la identificación de genes que son ortólogos de factores de transcripción de *Arabidopsis thaliana* a partir de una búsqueda de homología por ordenador.

Se identificaron secuencias homólogas, incluyendo las de parálogos y ortólogos de *Arabidopsis* y otras especies vegetales, usando herramientas de búsqueda de secuencias en bases de datos, tales como la Herramienta de Búsqueda de Alineamiento Local Básica (BLAST) (Altschul *et al.* (1990) mencionado anteriormente; y Altschul *et al.* (1997) Nucleic Acid Res. 25: 3389-3402). Los programas de análisis de secuencia tblastx se emplearon usando la matriz de puntuación BLOSUM-62 (Henikoff y Henikoff (1992) Proc. Natl. Acad. Sci. USA 89: 10915-10919). La base de datos de GenBank del NCBI completa se filtró para buscar secuencias de todas las plantas excepto *Arabidopsis thaliana* seleccionando todas las entradas en la base de datos de GenBank del NCBI asociadas con el ID taxonómico del NCBI 33090 (Viridiplantae; todas las plantas) y excluyendo entradas asociadas con el ID taxonómico 3701 (*Arabidopsis thaliana*).

Estas secuencias se comparan con secuencias que representan genes de factores de transcripción presentados en el Listado de Secuencias, usando el algoritmo de TBLASTX de la Universidad de Washington (versión 2.0a19MP) con los ajustes por defecto usando alineamientos con huecos con el filtro "apagado". Para cada gen de factor de transcripción en el Listado de Secuencias, se pidieron comparaciones individuales por puntuación de probabilidad (P-valor), en las que la puntuación refleja la probabilidad de que se produzca un alineamiento particular al azar. Por ejemplo, una puntuación de $3,6e-59$ es $3,6 \times 10^{-59}$. Además de los P valores, las comparaciones también se puntuaron por porcentaje de identidad. El porcentaje de identidad refleja el grado en el que dos segmentos de ADN o proteína son idénticos sobre una longitud particular. Los ejemplos de secuencias identificadas de este modo se presentan en, por ejemplo, el Listado de Secuencias, y la Tabla 5. Se identificaron fácilmente secuencias parálogas u ortólogas y están disponibles en GenBank por el número de Referencia (Tabla 5; Identificador de Secuencia o Número de Referencia). El porcentaje de identidad de secuencia entre estas secuencias puede ser tan bajo como 49 %, o incluso menor identidad de secuencia.

Se identificaron secuencias parálogas candidatas entre factores de transcripción de *Arabidopsis* mediante alineamiento, identidad y relaciones filogenéticas. G1067, G2153 y G2156 (SEC ID N°: 4, 6 y 8, respectivamente), parálogos de G1073, pueden encontrarse en el Listado de Secuencias.

Se identificaron secuencias ortólogas candidatas a partir de conjuntos unigénicos patentados de secuencias génicas vegetales en *Zea mays*, *Glycine max* y *Oryza sativa* basándose en la homología significativa con factores de transcripción de *Arabidopsis*. Estos candidatos se compararon de forma recíproca con el conjunto de factores de transcripción de *Arabidopsis*. Si el candidato mostró similitud máxima en el dominio proteico con el factor de transcripción inductor o con un parólogo del factor de transcripción inductor, entonces se consideró un ortólogo. Se proporcionan secuencias no de *Arabidopsis* identificadas que se había mostrado de esa manera que eran ortólogas de las secuencias de *Arabidopsis*, por ejemplo, en la Tabla 5.

Ejemplo X: identificación de secuencias ortólogas y parálogas por PCR

Los ortólogos de genes de *Arabidopsis* pueden identificarse por varios métodos, incluyendo hibridación, amplificación o bioinformática. Este ejemplo describe cómo se pueden identificar equivalentes del factor de transcripción de la familia AP2 de *Arabidopsis* CBF1, que confiere tolerancia a tensiones abióticas (Thomashow *et al.* (2002) Patente de Estados Unidos N° 6.417.428) y un ejemplo para confirmar la función de secuencias homólogas. En este ejemplo, se encontraron ortólogos de CBF1 en colza (*Brassica napus*) usando reacción en cadena de la polimerasa (PCR).

Se diseñaron cebadores degradados para regiones del dominio de unión AP2 y fuera del AP2 (dominio carboxilo terminal):

Mol 368 (inverso) 5'- CAY CCN ATH TAY MGN GGN GT -3'
(Solicitud de Patente de Estados Unidos 20040098764)

Mol 378 (directo) 5'- GGN ARN ARC ATN CCY TCN GCC -3'
(Solicitud de Patente de Estados Unidos 20040098764)

(Y: C/T, N: A/C/G/T, H: A/C/T, M: A/C, R: A/G)

El cebador Mol 368 está en el dominio de unión AP2 de CBF1 (secuencia de aminoácidos: His-Pro-Ile-Tyr-Arg-Gly-Val) mientras que el cebador Mol 378 están fuera del dominio AP2 (dominio carboxilo terminal) (secuencia de aminoácidos: Met-Ala-Glu-Gly-Met-Leu-Leu-Pro).

El ADN genómico aislado de *B. napus* se amplificó por PCR usando estos cebadores siguiendo estas condiciones: una etapa de desnaturalización inicial de 2 minutos a 93 °C; 35 ciclos de 93 °C durante 1 minuto, 55 °C durante 1 minuto y 72 °C durante 1 minuto; y una incubación final de 7 minutos a 72 °C al final del ciclo.

Los productos de PCR se separaron por electroforesis en un gel de agarosa al 1,2 % y se transfirieron a una membrana de nylon y se hibridaron con la sonda de CBF1 de AT preparada a partir del ADN genómico de *Arabidopsis* por amplificación por PCR. Los productos hibridados se visualizaron por el sistema de detección colorimétrico (Boehringer Mannheim) y las bandas correspondientes de un gel de agarosa similar se aislaron usando el Kit de Extracción Qiagen (Qiagen, Valencia CA). Los fragmentos de ADN se ligaron en el vector del clon de TA del Kit de Clonación TOPO TA (Invitrogen Corporation, Carlsbad CA) y se transformaron en la cepa de *E. coli* TOP10 (Invitrogen).

Se seleccionaron siete colonias y los insertos se secuenciaron en una máquina ABI 377 a partir de ambas cadenas con sentido y antisentido después del aislamiento del ADN plasmídico. La secuencia de ADN se editó por un secuenciador y se alineó con el AtCBF1 por software GCG y búsqueda de blast del NCBI.

La secuencia de ácido nucleico y secuencia de aminoácidos de un ortólogo de colza hallado de esta manera (bnCBF1) identificado por este procedimiento se muestran en la Solicitud de Patente de Estados Unidos 20040098764.

Las secuencias de aminoácidos alineadas muestran que el gen bnCBF1 tiene 88 % de identidad con la secuencia de *Arabidopsis* en la región del dominio AP2 y 85 % de identidad con la secuencia de *Arabidopsis* fuera del dominio AP2 cuando se alinea para dos secuencias de inserción que están fuera del dominio AP2.

De forma similar, también pueden identificarse secuencias parálogas de los genes de *Arabidopsis*, tales como CBF1.

Dos parálogos de CBF1 de *Arabidopsis thaliana*: CBF2 y CBF3. CBF2 y CBF3 se han clonado y secuenciado como se describe posteriormente

Se exploró una biblioteca de ADNc lambda preparada a partir de ARN aislado de *Arabidopsis thaliana* ecotipo Columbia (Lin y Thomashow (1992) Plant Physiol. 99: 519-525) con respecto a clones recombinantes que portaban insertos relacionados con el gen CBF1 (Stockinger *et al.* (1997) Proc. Natl. Acad. Sci. USA 94: 1035-1040). CBF1 se radiomarcó con ³²P por cebadores aleatorios (Sambrook *et al.* mencionado anteriormente) y se usó para explorar la biblioteca por la técnica de transferencia de placas usando condiciones de hibridación y lavado rigurosas convencionales (Hajela *et al.* (1990) Plant Physiol. 93: 1246-1252; Sambrook *et al.* mencionado anteriormente) (tampón SSPE 6 X, 60 °C para hibridación y tampón SSPE 0,1 X y 60 °C para lavados). Se obtuvieron doce clones de hibridación positiva y se determinaron las secuencias de ADN de los insertos de ADNc. Los resultados indicaron que los clones quedaban en tres clases. Una clase portaba insertos correspondientes a CBF1. Las otras dos clases portaban secuencias correspondientes a dos homólogos diferentes de CBF1, designados CBF2 y CBF3. Las secuencias de ácido nucleico y secuencias codificantes de proteínas predichas para CBF1, CBF2 y CBF3 de *Arabidopsis* pueden encontrarse, por ejemplo, en la Solicitud de Patente de Estados Unidos 20040098764, al igual que las secuencias de ácido nucleico y secuencia codificante de proteína predicha para el ortólogo de CBF de

Brassica napus bnCBF1.

Una comparación de las secuencias de ácido nucleico de *CBF1*, *CBF2* y *CBF3* de *Arabidopsis* indica que son de 83 a 85 % idénticas como se muestra en la Tabla 31.

TABLA 31. Comparación de identidad de *CBF1*, *CBF2* y *CBF3* de *Arabidopsis*

	Porcentaje de identidad ^a	
	ADN ^b	Polipéptido
cbf1/cbf2	85	86
cbf1/cbf3	83	84
cbf2/cbf3	84	85
^a El porcentaje de identidad se determinó usando el algoritmo <i>Clustal</i> del programa Megalign (DNASTAR, Inc.).		
^b Se muestran comparaciones de las secuencias de ácido nucleico de las fases abiertas de lectura.		

De forma similar, las secuencias de aminoácidos de los tres polipéptidos de CBF varían del 84 al 86 % de identidad. Un alineamiento de las tres secuencias de aminoácidos revela que la mayoría de las diferencias en la secuencia de aminoácidos se producen en la mitad C terminal del polipéptido. Esta región de CBF1 actúa como un dominio de activación tanto en levadura como en *Arabidopsis* (no mostrado).

Los restos 47 a 106 de CBF1 corresponden al dominio AP2 de la proteína, un motivo de unión a ADN que hasta la fecha solamente se ha encontrado en proteínas vegetales. Una comparación de los dominios AP2 de CBF1, CBF2 y CBF3 indica que hay algunas diferencias en la secuencia de aminoácidos. Estas diferencias en la secuencia de aminoácidos podrían tener un efecto en la especificidad de unión al ADN.

Ejemplo XI: transformación de colza con un plásmido que contiene CBF1, CBF2 o CBF3

Después de identificar genes homólogos de CBF1, se transformó colza con un plásmido que contenía los genes de CBF1, CBF2 o CBF3 de *Arabidopsis* en el vector pGA643 (An (1987) Methods Enzymol. 253: 292). En estas construcciones los genes de CBF se expresaron de forma constitutiva bajo el promotor 35S de CaMV. Además, el gen CBF1 se clonó bajo el control del promotor COR15 de *Arabidopsis* en el mismo vector pGA643. Cada construcción se transformó en la cepa de *Agrobacterium* GV3101. Se cultivaron *Agrobacterias* transformadas durante 2 días en medio AB mínimo que contenía antibióticos apropiados.

Se transformó colza de primavera (*B. napus* cv. Westar) usando el protocolo de Moloney *et al.* (1989) Plant Cell Reports 8: 238, con algunas modificaciones como se describe. Brevemente, las semillas se esterilizaron y se sembraron en placas en medio MS de media concentración, que contenía sacarosa al 1 %. Las placas se incubaron a 24 °C con luz 60-80 $\mu\text{E}/\text{m}^2\text{s}$ usando un fotoperiodo de 16 horas de luz/8 horas de oscuridad. Se recogieron cotiledones de plántulas de 4-5 días de edad, los peciolo se cortaron y se sumergieron en la solución de *Agrobacterium*. Los cotiledones sumergidos se colocaron en medio de cocultivo a una densidad de 20 cotiledones/placa y se incubaron como se ha descrito anteriormente durante 3 días. Los explantes se transfirieron al mismo medio, pero conteniendo timentina 300 mg/l (SmithKline Beecham, PA) y se diluyeron a 10 cotiledones/placa. Después de 7 días los explantes se transfirieron a medio de selección/regeneración. Las transferencias se continuaron cada 2-3 semanas (2 o 3 veces) hasta que se habían desarrollado los brotes. Los brotes se transfirieron a medio de elongación de brotes cada 2-3 semanas. Se transfirieron los brotes de aspecto sano a medio de enraizamiento. Una vez que se habían desarrollado buenas raíces, las plantas se colocaron en suelo para maceta húmedo.

Las plantas transformadas se analizaron después con respecto a la presencia del gen NPTII/resistencia a kanamicina por ELISA, usando el kit de ELISA NPTII de 5Prime-3Prime Inc. (Boulder, CO). Aproximadamente el 70 % de las plantas exploradas fueron positivas para NPTII. Solamente esas plantas se analizaron adicionalmente.

A partir de la transferencia de Northern el análisis de las plantas que se transformaron con las construcciones de expresión constitutiva, mostró expresión de los genes CBF y todos los genes CBF fueron capaces de inducir el gen regulado por frío de *Brassica napus* BN115 (homólogo del gen COR15 de *Arabidopsis*). La mayoría de las plantas transgénicas parecen mostrar un fenotipo de crecimiento normal. Como se esperaba, las plantas transgénicas son más tolerantes a la congelación que las plantas de control de tipo silvestre. Usando el ensayo de filtración de electrolitos de las hojas, el control mostró una filtración del 50 % de -2 a -3 °C. La colza de primavera transformada con CBF1 o CBF2 mostró una filtración del 50 % de -6 a -7 °C. La colza de primavera transformada con CBF3 muestra una filtración del 50 % de aproximadamente -10 a -15 °C. La colza de invierno transformada con CBF3 puede mostrar una filtración del 50 % de aproximadamente -16 a -20 °C. Además, si la colza de primavera o invierno

está aclimatada al frío las plantas transformadas pueden mostrar un aumento adicional en la tolerancia a la congelación de al menos -2 °C.

Para ensayar la tolerancia a salinidad de las plantas transformadas, las plantas se regaron con NaCl 150 mM. Las plantas que sobreexpresaban CBF1, CBF2 o CBF3 crecieron mejor en comparación con plantas que no se habían transformado con CBF1, CBF2, o CBF3.

Estos resultados demuestran que pueden identificarse equívocos de factores de transcripción de *Arabidopsis* y mostrarse que confieren funciones similares en especies vegetales no *Arabidopsis*.

Ejemplo XII: ensayos de desplazamiento en gel.

La presencia de un factor de transcripción que comprende un dominio de unión a ADN que se une con un elemento de unión al factor de transcripción de ADN se evalúa usando el siguiente ensayo de desplazamiento en gel. El factor de transcripción se expresa de forma recombinante y se aísla a partir de *E. coli* o se aísla de material vegetal. Se incubaba proteína soluble total, incluyendo el factor de transcripción, (40 ng) a temperatura ambiente en 10 µl de tampón de unión 1X (HEPES 15 mM (pH 7,9), EDTA 1 mM, KCl 30 mM, glicerol 5 %, albúmina de suero bovino 5 %, DTT 1 mM) más 50 ng de poli(dI-dC):poli(dI-dC) (Pharmacia, Piscataway NJ) con o sin 100 ng de ADN competidor. Después de 10 minutos de incubación, se añade ADN de sonda que comprende un elemento de unión al factor de transcripción de ADN (1 ng) que se ha marcado con ³²P por relleno de extremos (Sambrook *et al.* (1989) mencionado anteriormente) y la mezcla se incubaba durante 10 minutos adicionales. Las muestras se cargan en geles de poliacrilamida (4 % p/v) y se fraccionan por electroforesis a 150V durante 2 horas (Sambrook *et al.* mencionado anteriormente). El grado de unión del ADN de sonda-factor de transcripción se visualiza usando autorradiografía. Se preparan ADN de sondas y competidores a partir de insertos de oligonucleótidos ligados en el sitio BamHI de pUC118 (Vieira *et al.* (1987) Methods Enzymol. 153: 3-11). La orientación y el número de concatenación de los insertos se determinan por análisis de secuencia de ADN dideoxi (Sambrook *et al.* mencionado anteriormente). Los insertos se recuperan después de digestión de restricción con EcoRI y HindIII y fraccionamiento en geles de poliacrilamida (12 % p/v) (Sambrook *et al.* mencionado anteriormente).

Ejemplo XIII. Transformación de dicotiledóneas

Se ha mostrado experimentalmente que especies de cultivo que sobreexpresan miembros del clado G1073 de polipéptidos de factores de transcripción (por ejemplo G2153) producen plantas con biomasa aumentada en ensayos de campo. Esta observación indica que estos genes, cuando se sobreexpresan, darán como resultado mayores producciones de diversas especies vegetales, que pueden ser más significativas en las plantas en las que la parte vegetativa de la planta es comestible. Se ha descubierto que las plantas de tomate que sobreexpresan el polipéptido G2153 de *A. thaliana* son mayores que las plantas de tomate de control de tipo silvestre.

Por lo tanto, las secuencias de factores de transcripción enumeradas en el Listado de Secuencias recombinadas en vectores de expresión pMEN20 o pMEN65 pueden transformarse en una planta para el fin de modificar rasgos de la planta. El vector de clonación puede introducirse en diversas plantas de cereales por medios bien conocidos en la técnica tales como, por ejemplo, transferencia de ADN directa o transformación mediada por *Agrobacterium tumefaciens*. Es ahora rutinario producir plantas transgénicas usando la mayoría de plantas dicotiledóneas (véase Weissbach y Weissbach, (1989) mencionado anteriormente; Gelvin *et al.* (1990) mencionado anteriormente; Herrera-Estrella *et al.* (1983) mencionado anteriormente; Bevan (1984) mencionado anteriormente; y Klee (1985) mencionado anteriormente). Los métodos para el análisis de rasgos son rutinarios en la técnica y se han desvelado anteriormente ejemplos.

Se han descrito previamente numerosos protocolos para la transformación de plantas de tomate y soja, y se conocen bien en la técnica. Gruber *et al.* ((1993) en Methods in Plant Molecular Biology and Biotechnology, p. 89-119, Glick y Thompson, eds., CRC Press, Inc., Boca Raton) describen varios vectores de expresión y métodos de cultivo que pueden usarse para transformación de células o tejidos y posterior regeneración. Para la transformación de soja, se describen métodos en Miki *et al.* (1993) en Methods in Plant Molecular Biology and Biotechnology, p. 67-88, Glick y Thompson, eds., CRC Press, Inc., Boca Raton; y Patente de Estados Unidos N° 5.563.055, (Townsend y Thomas), presentada el 8 de octubre de 1996.

Hay un número sustancial de alternativas a protocolos de transformación mediados por *Agrobacterium*, otros métodos para el fin de transferir genes exógenos en soja o tomate. Uno de dichos métodos es la transformación mediada por microproyectiles, en la que el ADN en la superficie de partículas de microproyectiles se introduce en tejidos vegetales con un dispositivo biolístico (véase, por ejemplo, Sanford *et al.*, (1987) Part. Sci. Technol. 5: 27-37; Christou *et al.* (1992) Plant. J. 2: 275-281; Sanford (1993) Methods Enzymol. 217: 483-509; Klein *et al.* (1987) Nature 327: 70-73; Patente de Estados Unidos N° 5.015.580 (Christou *et al.*), presentada el 14 de mayo de 1991; y Patente de Estados Unidos N° 5.322.783 (Tomes *et al.*), presentada el 21 de junio de 1994.

Como alternativa, se han usado métodos de sonicación (véase, por ejemplo, Zhang *et al.* (1991) Bio/Technology 9: 996-997); captación directa de ADN en protoplastos usando precipitación con CaCl₂, alcohol polivinílico o poli-L-

ornitina (véase, por ejemplo, Hain *et al.* (1985) *Mol. Gen. Genet.* 199: 161-168; Draper *et al.*, *Plant Cell Physiol.* 23: 451-458 (1982)); fusión de liposomas o esferoplastos (véase, por ejemplo, Deshayes *et al.* (1985) *EMBO J.*, 4: 2731-2737; Christou *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84: 3962-3966); y electroporación de protoplastos y células y tejidos completos (véase, por ejemplo, Donn *et al.* (1990) en *Abstracts of VIth International Congress on Plant Cell and Tissue Culture IAPTC*, A2-38: 53; D'Halluin *et al.* (1992) *Plant Cell* 4: 1495-1505; y Spencer *et al.* (1994) *Plant Mol. Biol.* 24: 51-61) para introducir ADN ajeno y vectores de expresión en plantas.

Después de transformarse una planta o célula vegetal (y la última regenerada en una planta), la planta transformada puede cruzarse consigo misma o una planta de la misma línea, una planta no transformada o de tipo silvestre, u otra planta transformada de una línea de plantas transgénica diferente. El cruzamiento proporciona las ventajas de producir variedades transgénicas nuevas y con frecuencia estables. Los genes y rasgos que confieren que se han introducido en una línea de tomate o soja pueden moverse a una línea distinta de plantas usando técnicas de retrocruzamiento tradicionales bien conocidas en este campo. La transformación de plantas de tomate puede realizarse usando los protocolos de Koornneef *et al.* (1986) en *Tomato Biotechnology*: Alan R. Liss, Inc., 169-178, y en la Patente de Estados Unidos 6.613.962, este último método descrito brevemente en el presente documento. Se precultivan explantes de cotiledones de ocho días de edad durante 24 horas en placas de Petri que contienen una capa de alimentación de células en suspensión de petunia híbrida sembradas en medio MS con sacarosa 2 % (p/v) y agar 0,8 % complementado con ácido α -naftaleno acético 10 μ M y 6-bencilaminopurina 4,4 μ M. Los explantes se infectan después con un cultivo diluido de una noche de *Agrobacterium tumefaciens* que contiene un vector de expresión que comprende un polinucleótido de la invención durante 5-10 minutos, se secan por transferencia en papel de filtro estéril y se cultivan conjuntamente durante 48 horas en las placas de capa de alimentación originales. Las condiciones de cultivo son como se ha descrito anteriormente. Se diluyen cultivos de una noche de *Agrobacterium tumefaciens* en medio MS líquido con sacarosa 2 % (p/v), pH 5,7 hasta una DO₆₀₀ de 0,8.

Después del cultivo conjunto, los explantes de cotiledones se transfieren a placas de Petri con medio selectivo que comprende medio MS con zeatina 4,56 μ M, vancomicina 67,3 μ M, cefotaxima 418,9 μ M y sulfato de kanamicina 171,6 μ M, y se cultivan en las condiciones de cultivo descritas anteriormente. Los explantes se subcultivan cada tres semanas en medio nuevo. Los brotes emergentes se disecan del callo subyacente y se transfieren a tarros de vidrio con medio selectivo sin zeatina para formar raíces. La formación de las raíces en un medio que contiene sulfato de kanamicina es un indicio positivo de una transformación exitosa.

Puede realizarse transformación de plantas de soja usando los métodos hallados en, por ejemplo, la Patente de Estados Unidos 5.563.055 (Townsend *et al.*, presentada el 8 de octubre de 1996), descrito brevemente en el presente documento. En este método la semilla de soja se esteriliza en superficie por exposición a gas de cloro introducido en un tarro de vidrio curvado. Las semillas se germinan sembrando en placas en medio solidificado de agar de concentración 1/10 sin reguladores del crecimiento de las plantas y cultivando a 28 °C con una duración del día de 16 horas. Después de tres o cuatro días, las semillas pueden prepararse para cocultivo. El tegumento se retira y la radícula que se elonga se retira 3-4 mm por debajo de los cotiledones.

Se dejan crecer cultivos de una noche de *Agrobacterium tumefaciens* que albergan el vector de expresión que comprende un polinucleótido de la invención hasta fase logarítmica, se agrupan y se concentran por centrifugación. Se realizan inoculaciones en lotes de modo que cada placa de semillas se trate con un nuevo sedimento resuspendido de *Agrobacterium*. Los sedimentos se resuspenden en 20 ml de medio de inoculación. El inóculo se vierte en una placa de Petri que contiene semillas preparadas y los nodos de cotiledones se maceran con una hoja quirúrgica. Después de 30 minutos, los explantes se transfieren a placas del mismo medio que se ha solidificado. Los explantes se incluyen con el lado adaxial hacia arriba y nivelados con la superficie del medio y se cultivan a 22 °C durante tres días con luz fluorescente blanca. Estas plantas pueden regenerarse después de acuerdo con métodos bien establecidos en la técnica, tal como moviendo las plantas después de tres días a un medio de contraselección líquido (véase Patente de Estados Unidos 5.563.055).

Los explantes pueden después seleccionarse, incluirse y cultivarse en medio de selección solidificado. Después de un mes en medio selectivo el tejido transformado se hace visible como sectores verdes de tejido en regeneración frente a un fondo de tejido blanqueado, menos sano. Los explantes con sectores verdes se transfieren a un medio de elongación. El cultivo se continúa en este medio con transferencias a placas nuevas cada dos semanas. Cuando los brotes son de 0,5 cm de longitud pueden escindirse en la base y colocarse en un medio de enraizamiento.

Ejemplo XIV: biomasa y tolerancia al aumento de estrés abiótico en monocotiledóneas

Pueden transformarse plantas de cereales tales como, pero sin limitación, maíz, trigo, arroz, sorgo o cebada con las presentes secuencias polinucleótídicas, incluyendo secuencias derivadas de monocotiledóneas o dicotiledóneas tales como las presentadas en las Tablas 2 o 5, clonadas en un vector tal como pGA643 y que contienen un marcador de resistencia a kanamicina, y expresadas de forma constitutiva bajo, por ejemplo, los promotores 35S de CaMV o COR15. También pueden usarse pMEN20 o pMEN65 y otros vectores de expresión para el fin de modificar rasgos de la planta. Por ejemplo, puede modificarse pMEN20 para reemplazar la región codificante de NptII con el gen BAR de *Streptomyces hygroscopicus* que confiere resistencia a fosfotricina. Los sitios KpnI y BglII del gen Bar se retiran por mutagénesis dirigida con cambios codónicos silenciosos.

El vector de clonación puede introducirse en diversas plantas de cereales por medios bien conocidos en la técnica incluyendo transferencia directa de ADN o transformación mediada por *Agrobacterium tumefaciens*. Este último enfoque puede conseguirse por diversos medios, incluyendo, por ejemplo, el de la Patente de Estados Unidos N° 5.591.616, en el que se transforma un callo de monocotiledónea poniendo en contacto el tejido desdiferenciador con la *Agrobacterium* que contiene el vector de clonación.

Los tejidos de muestra se sumergen en una suspensión de 3×10^{-9} células de *Agrobacterium* que contiene el vector de clonación durante 3-10 minutos. El material del callo se cultiva en medio sólido a 25 °C en oscuridad durante varios días. Los callos cultivados en este medio se transfieren a medio de Regeneración. Se continúan las transferencias cada 2-3 semanas (2 o 3 veces) hasta que se desarrollan brotes. Los brotes se transfieren después a medio de Elongación de Brotes cada 2-3 semanas. Los brotes de aspecto sano se transfieren a medio de enraizamiento y después de haberse desarrollado las raíces, las plantas se colocan en suelo para maceta húmedo.

Las plantas transformadas se analizan después con respecto a la presencia del gen de NPTII/resistencia a kanamicina por ELISA, usando el kit de ELISA NPTII de 5Prime-3Prime Inc. (Boulder, CO).

También es rutinario usar otros métodos para producir plantas transgénicas de la mayoría de los cultivos de cereales (Vasil (1994) Plant Mol. Biol. 25: 925-937) tales como maíz, trigo, arroz, sorgo (Cassas *et al.* (1993) Proc. Natl. Acad. Sci. USA 90: 11212-11216), y cebada (Wan y Lemeaux (1994) Plant Physiol. 104:37-48). Pueden usarse métodos de transferencia de ADN tales como el método de microproyectiles para el maíz (Fromm *et al.* (1990) Bio/Technol. 8: 833-839; Gordon-Kamm *et al.* (1990) Plant Cell 2: 603-618; Ishida (1990) Nature Biotechnol. 14:745-750), el trigo (Vasil *et al.* (1992) Bio/Technol. 10: 667-674; Vasil *et al.* (1993) Bio/Technol. 11: 1553-1558; Weeks *et al.* (1993) Plant Physiol. 102: 1077-1084), y el arroz (Christou (1991) Bio/Technol. 9: 957-962; Hiei *et al.* (1994) Plant J. 6: 271-282; Aldemita y Hodges (1996) Planta 199: 612-617; y Hiei *et al.* (1997) Plant Mol. Biol. 35:205-218). Para la mayoría de las plantas de cereales, las células embriogénicas derivadas de tejidos del escutelo inmaduros son las dianas celulares preferidas para transformación (Hiei *et al.* (1997) Plant Mol. Biol. 35: 205-218; Vasil (1994) Plant Mol. Biol. 25: 925-937). Para transformar células embriogénicas de maíz derivadas de tejido escutelar inmaduro usando bombardeo de microproyectiles, el genotipo A188XB73 es el genotipo preferido (Fromm *et al.* (1990) Bio/Technol. 8: 833-839; Gordon-Kamm *et al.* (1990) Plant Cell 2: 603-618). Después del bombardeo de microproyectiles los tejidos se seleccionan en fosfinotricina para identificar las células embriogénicas transgénicas (Gordon-Kamm *et al.* (1990) Plant Cell 2: 603-618). Las plantas transgénicas se regeneran por técnicas de regeneración de maíz convencionales (Fromm *et al.* (1990) Bio/Technol. 8: 833-839; Gordon-Kamm *et al.* (1990) Plant Cell 2: 603-618).

El análisis de transferencia de Northern, RT-PCR o análisis de micromatriz de las plantas transformadas, regeneradas, pueden usarse para mostrar la expresión de G1073 y genes relacionados que son capaces de inducir tolerancia al estrés abiótico y mayor tamaño.

Para verificar la capacidad para conferir tolerancia al estrés abiótico, pueden exponerse plantas maduras que sobreexpresan un G1073 o equivalente, o como alternativa, plántula descendiente de estas plantas, a un estrés osmótico, tal como sequía, calor, alta salinidad o congelación. Como alternativa, estas plantas pueden exponerse en una condición de estrés osmótico que también puede medir la detección de azúcar alterada, tal como una condición de alto azúcar. Comparando las plantas de tipo silvestre y transgénicas tratadas de forma similar, puede mostrarse que las plantas transgénicas tienen mayor tolerancia al estrés abiótico.

Después de haberse transformado una planta o célula vegetal monocotiledónea (y haberse regenerado esta última a una planta) y haberse mostrado que tiene mayor tamaño o tolerancia al estrés abiótico, o que produce mayor rendimiento en relación con una planta de control en las condiciones de estrés, la planta monocotiledónea transformada puede cruzarse consigo misma o una planta de la misma línea, una planta monocotiledónea no transformada o de tipo silvestre, u otra planta monocotiledónea transformada de una línea transgénica de plantas diferente.

Estos experimentos demostrarían que pueden identificarse miembros del clado G1073 de polipéptidos del factor de transcripción y mostrarse que confieren mayor tamaño, mayor producción y/o mayor tolerancia al estrés abiótico en monocotiledóneas, incluyendo tolerancia o resistencia a múltiples tensiones.

Ejemplo XV: Secuencias que confieren mejoras significativas a especies no *Arabidopsis*

La función de ortólogos específicos de G1073 se ha analizado y puede caracterizarse adicionalmente e incorporarse en plantas de cultivo. La sobreexpresión ectópica de estos ortólogos puede regularse usando elementos reguladores constitutivos, inducibles o específicos de tejido. Los genes que se han examinado y se ha mostrado que modifican rasgos de plantas (incluyendo aumento de la biomasa y tolerancia al estrés abiótico) codifican miembros del clado G1073 de polipéptidos de factores de transcripción, tales como los hallados en *Arabidopsis thaliana* (SEC ID N°: 2, 4, 6, 8, 42, 67 y 69), *Oryza sativa* (SEC ID N°: 10, 12, 26, 30 y 38) y *Glycine max* (SEC ID N°: 14, 16, 18 y 40). Además de estas secuencias, se espera que secuencias polinucleotídicas relacionadas que codifican polipéptidos hallados en el Listado de Secuencias puedan inducir también rasgos alterados, incluyendo biomasa aumentada y

tolerancia al estrés abiótico, cuando se transformen en diversas plantas considerable de diferentes especies, e incluyendo dicotiledóneas y monocotiledóneas. Las secuencias polinucleotídicas y polipeptídicas derivadas de monocotiledóneas (por ejemplo, las secuencias de arroz) pueden usarse para transformar plantas tanto monocotiledóneas como dicotiledóneas, y las derivadas de dicotiledóneas (por ejemplo, los genes de *Arabidopsis* y soja) pueden usarse para transformar uno de los grupos, aunque se espera que algunas de estas secuencias actúen mejor si el gen se transforma en una planta del mismo grupo del que deriva la secuencia.

Las semillas de estas plantas transgénicas se someten a ensayos de germinación para medir la detección de sacarosa. Las semillas monocotiledóneas estériles, incluyendo, pero sin limitación, maíz, arroz, trigo, centeno y sorgo, así como dicotiledóneas incluyendo, pero sin limitación soja y alfalfa, se siembran en medio MS al 80 % más vitaminas con sacarosa 9,4 %; el medio de control carece de sacarosa. Todas las placas de ensayo se incuban después a 22 °C con luz las 24 horas, 120-130 $\mu\text{Ein}/\text{m}^2/\text{s}$, en una cámara de cultivo. La evaluación de la germinación y el vigor de la plántula se realizan entonces tres días después de plantar. Puede descubrirse que los que sobreexpresan estos genes son más tolerantes a alta sacarosa por tener mejor germinación, radículas más largas y más expansión de cotiledones. Estos resultados han indicado que los que sobreexpresan G1073, G1067, G1069, G2153, G2156, G2789, G3401 y G3460 están implicados en detección de azúcar específica de sacarosa; se espera que ortólogos estructuralmente similares de estas secuencias, incluyendo los hallados en el Listado de Secuencias, también estén implicados en la detección de azúcar, un indicio de tolerancia al estrés osmótico alterada.

Las plantas que sobreexpresan las secuencias de factores de transcripción descritas en el presente documento también pueden someterse a ensayos de sequía basados en suelo para identificar las líneas que son más tolerantes a privación de agua que las plantas de control de tipo silvestre. Varias líneas de plantas que sobreexpresan un miembro del clado G1073 de polipéptidos de factores de transcripción serán significativamente mayores y más verdes, con menor marchitamiento o desecación, que las plantas de control de tipo silvestre, particularmente después de un periodo de privación de agua seguido de nuevo riego y un periodo de incubación posterior. La secuencia del miembro del clado G1073 puede sobreexpresarse bajo el control regulador de promotores constitutivos, específicos de tejido o inducibles, o puede comprender un dominio de transactivación de GAL4 fusionado con el extremo N o C terminal del polipéptido. Los resultados presentados en el Ejemplo VIII indican que miembros del clado G1073 pueden conferir tolerancia al estrés cuando se sobreexpresan bajo el control regulador de promotores no constitutivos o un dominio de transactivación fusionado con el miembro del clado sin un impacto significativo y la morfología de la planta. Las líneas que presentan rasgos útiles pueden seleccionarse para estudio adicional o desarrollo comercial.

Las plantas monocotiledóneas, incluyendo arroz, maíz, trigo, centeno, sorgo, cebada y otros, pueden transformarse con un plásmido que contiene un miembro del clado G1073 de polipéptidos de factores de transcripción. La secuencia del clado G1073 puede incluir secuencias derivadas de dicotiledóneas o monocotiledóneas tales como las presentadas en la Tabla 1 o Tabla 5. Estos genes de factores de transcripción de gancho AT pueden clonarse en un vector tal como pGA643 y que contiene un marcador de resistencia a kanamicina, y después expresarse de forma constitutiva bajo el promotor 35S de CaMV o promotor COR15.

El vector de clonación puede introducirse en monocotiledóneas, por ejemplo, por medios descritos en detalle en el Ejemplo XIV, incluyendo transferencia de ADN directa o transformación mediada por *Agrobacterium tumefaciens*. Este último enfoque puede conseguirse por diversos medios, incluyendo, por ejemplo, el de la Patente de Estados Unidos N° 5.591.616, en el que se transforma un callo de monocotiledónea poniendo en contacto tejido desdiferenciador con la *Agrobacterium* que contiene el vector de clonación.

Los tejidos de muestra se sumergen en una suspensión de 3×10^{-9} células de *Agrobacterium* que contienen el vector de clonación durante 3-10 minutos. El material del callo se cultiva en medio sólido a 25 °C en oscuridad durante varios días. Los callos cultivados en este medio se transfieren a medio de Regeneración. Las transferencias se continúan cada 2-3 semanas (2 o 3 veces) hasta que se desarrollan brotes. Los brotes se transfieren después a medio de Elongación de Brotes cada 2-3 semanas. Los brotes de aspecto sano se transfieren a medio de enraizamiento y después de haberse desarrollado las raíces, las plantas se colocan en suelo para maceta húmedo.

Las plantas transformadas se analizan después con respecto a la presencia del gen de NPTII/resistencia a kanamicina por ELISA, usando el kit de ELISA NPTII de 5Prime-3Prime Inc. (Boulder, CO).

Pueden usarse análisis de transferencia de Northern, RT-PCR o análisis de micromatrices de las plantas regeneradas, transformadas, para mostrar la expresión de un miembro del clado G1073 de polipéptidos de factores de transcripción que es capaz de inducir tolerancia al estrés abiótico.

Para verificar la capacidad de conferir tolerancia al estrés abiótico, las plantas maduras que expresan un gen de equivalente derivado de monocotiledónea o, como alternativa, plántula descendiente de estas plantas, pueden exponerse usando métodos descritos en el Ejemplo VI. Comparando plantas de tipo silvestre y las transgénicas, se ha mostrado que estas últimas son más tolerantes a estrés abiótico y/o tienen biomasa aumentada, en comparación con plantas de control de tipo silvestre tratadas de forma similar.

Estos experimentos demuestran que pueden identificarse varios miembros representativos del clado G1073 de polipéptidos de factores de transcripción, incluyendo G1067, G1069, G1073, G1667, G2153, G2156, G2157, G2789, G3399, G3400, G3401, G3406, G3407, G3456, G3459, G3460 y G3556, y mostrarse que aumentan la biomasa y/o mejoran la tolerancia de estrés abiótico, incluyendo tensiones osmóticas tales como estrés por sequía o salinidad. Se espera que los mismos métodos puedan aplicarse para identificar otros miembros útiles y valiosos del clado de una amplia serie de especies.

La divulgación también se refiere a los aspectos expuestos en los siguientes párrafos numerados:

1. Una planta transgénica que tiene tolerancia aumentada al estrés abiótico, comprendiendo dicha planta transgénica un polinucleótido recombinante que codifica un miembro del clado G1073 de polipéptidos de factores de transcripción, en la que el miembro del clado G1073 de polipéptidos de factores de transcripción comprende:

un dominio de gancho AT; y

un segundo dominio conservado que comprende SEC ID N°: 63 o SEC ID N°: 64.

2. La planta transgénica del párrafo 1, en la que el dominio de gancho AT y el segundo dominio conservado son al menos 78 % y 62 % idénticos al dominio de gancho AT y el segundo dominio conservado de SEC ID N°: 2, respectivamente.

3. La planta transgénica del párrafo 1, en la que el estrés abiótico es frío.

4. La planta transgénica del párrafo 1, en la que el estrés abiótico es un estrés osmótico.

5. La planta transgénica del párrafo 4, en la que el estrés osmótico se selecciona del grupo que consiste en calor, sequía, desecación, congelación y alta salinidad.

6. La planta transgénica del párrafo 1, en la que el miembro del clado G1073 de polipéptidos de factores de transcripción se selecciona del grupo que consiste en SEC ID N°: 2, 4, 6, 8, 10, 12, 14, 16, 18, 26, 30, 38, 42 y 67.

7. La planta transgénica del párrafo 1, en la que el polinucleótido recombinante está unido operativamente con un promotor inducible o específico de tejido que regula la expresión del polinucleótido recombinante cuando el polinucleótido recombinante se transforma en una planta.

8. La planta transgénica del párrafo 7, en la que el promotor específico de tejido es un promotor vascular, epidérmico, de hoja o de raíz.

9. La planta transgénica del párrafo 7, en la que el promotor inducible o específico de tejido se selecciona del grupo que consiste en un promotor SUC2, un promotor CUT1, un promotor RBCS3, un promotor ARSK1 y un promotor RD29A.

10. Semilla producida a partir de la planta transgénica de acuerdo con el párrafo 1.

11. Un método para producir una planta transformada con mayor tolerancia al estrés abiótico que una planta de control, comprendiendo el método:

(a) proporcionar un vector de expresión que comprende una secuencia polinucleotídica que codifica un miembro del clado G1073 de polipéptidos de factores de transcripción;

en el que el miembro del clado G1073 de polipéptidos de factores de transcripción comprende un dominio de gancho AT y un segundo dominio conservado, en orden de N terminal a C terminal, en el que el segundo dominio conservado comprende SEC ID N°: 63 o SEC ID N°: 64; y

en el que la secuencia polinucleotídica está unida operativamente con un elemento regulador que controla la expresión de la secuencia polinucleotídica;

(b) transformar una planta diana con el vector de expresión para producir la planta transformada; y

(c) cultivar la planta transformada que comprende el vector de expresión.

12. El método del párrafo 11, en el que el dominio de gancho AT y el segundo dominio conservado son al menos 78 % y 62 % idénticos al dominio de gancho AT y el segundo dominio conservado de SEC ID N°: 2, respectivamente.

13. El método del párrafo 11, en el que dicha tolerancia al estrés abiótico es a frío.

14. El método del párrafo 11, en el que dicha tolerancia al estrés abiótico es al estrés osmótico.

15. El método del párrafo 14, en el que dicho estrés osmótico se selecciona del grupo que consiste en calor,

deseccación, sequía, congelación y alta salinidad.

16. El método del párrafo 11, en el que el miembro del clado G1073 de polipéptidos de factores de transcripción comprende una secuencia polipeptídica seleccionada del grupo que consiste en SEC ID N°: 2, 4, 6, 8, 10, 12, 14, 16, 18, 26, 30, 38, 42 y 67.

17. El método del párrafo 11, en el que el elemento regulador es un promotor inducible o específico de tejido.

18. El método del párrafo 17, en el que la planta transformada tiene una morfología que es sustancialmente similar a la planta de control.

19. El método del párrafo 17, en el que el promotor específico de tejido es un promotor vascular, epidérmico, de hoja o de raíz.

20. El método del párrafo 17, en el que el promotor inducible o específico de tejido se selecciona del grupo que consiste en un promotor SUC2, un promotor CUT1, un promotor RBCS3, un promotor ARSK1 y un promotor RD29A.

21. El método del párrafo 11, comprendiendo adicionalmente las etapas del método:

(d) autopolinizar o cruzar la planta transformada consigo misma u otra planta, respectivamente, para producir semillas; y

(e) cultivar una planta descendiente de la semilla;

en el que la planta descendiente tiene mayor tolerancia al estrés abiótico que la planta de control.

22. Una planta transgénica que tiene biomasa aumentada, comprendiendo dicha planta transgénica un polinucleótido recombinante que comprende una secuencia de nucleótidos que codifica un miembro del clado G1073 de polipéptidos de factores de transcripción;

en la que el miembro del clado G1073 de polipéptidos de factores de transcripción comprende:

un dominio de gancho AT; y

un segundo dominio conservado que comprende SEC ID N°: 63 o SEC ID N°: 64.

23. La planta transgénica del párrafo 22, en la que el dominio de gancho AT y el segundo dominio conservado son al menos 78 % y 62 % idénticos al dominio de gancho AT y el segundo dominio conservado de SEC ID N°: 2, respectivamente.

24. La planta transgénica del párrafo 22, en la que el miembro del clado G1073 de polipéptidos de factores de transcripción se selecciona del grupo que consiste en SEC ID N°: 4, 6, 8, 10, 12, 14, 16, 18, 30, 38, 42, 67 y 69.

25. La planta transgénica del párrafo 22, en la que el polinucleótido recombinante está unido operativamente con al menos un elemento regulador capaz de regular la expresión del polinucleótido recombinante cuando el polinucleótido recombinante se transforma en una planta.

26. Semillas producidas a partir de la planta transgénica de acuerdo con el párrafo 22.

27. Un método para producir una planta transformada con mayor biomasa que una planta de control, comprendiendo el método:

(a) proporcionar un vector de expresión que comprende una secuencia polinucleotídica que codifica un miembro del clado G1073 de polipéptidos de factores de transcripción;

en el que el miembro del clado G1073 de polipéptidos de factores de transcripción comprende un dominio de gancho AT y un segundo dominio conservado, en orden de N terminal a C terminal, en el que el segundo dominio conservado comprende SEC ID N°: 63 o SEC ID N°: 64; y

en el que la secuencia polinucleotídica está unida operativamente con un elemento regulador que controla la expresión de la secuencia polinucleotídica;

(b) transformar una planta diana con el vector de expresión para producir la planta transformada; y

(c) cultivar la planta transformada que comprende el vector de expresión.

28. El método del párrafo 27, en el que el dominio de gancho AT y el segundo dominio conservado son al menos 78 % y 62 % idénticos al dominio de gancho AT y el segundo dominio conservado de SEC ID N°: 2, respectivamente.

29. El método del párrafo 27, en el que el miembro del clado G1073 de polipéptidos de factores de transcripción comprende una secuencia polipeptídica seleccionada del grupo que consiste en SEC ID N°: 4, 6, 8, 10, 12, 14,

16, 18, 26, 30, 38, 42, 67 y 69.

30. El método del párrafo 27, comprendiendo las etapas del método además:

- 5 (d) autopolinizar o cruzar la planta transformada consigo misma u otra planta, respectivamente, para producir semillas; y
(e) cultivar una planta descendiente de la semilla;

en el que la planta descendiente tiene mayor biomasa que la planta de control.

10

LISTADO DE SECUENCIAS

<110> Mendel Biotechnology, Inc.

15

<120> Regulación Transcripcional de Biomasa Vegetal y Tolerancia a Estrés abiótico

<130> AHB/FP6714109

20

<140>

<141> 17-09-2004

<150> 60/565.948

<151> 26-04-2004

25

<150> 10/669.824

<151> 23-09-2003

<150> 10/870.198

<151> 16-06-2004

30

<150> PCT/US2004/030739

<151> 17-09-2004

<150> 04784569.8

35

<151> 17-09-2004

<160> 72

<170> PatentIn versión 3.2

40

<210> 1

<211> 974

<212> ADN

<213> Arabidopsis thaliana

45

<220>

<223> G1073

<400> 1

ES 2 502 943 T3

```

ccccccgacc tgcctctaca gagacctgaa gattccagaa cccacactga tcaaaaataa 60
catggaactt aacagatctg aagcagacga agcaaaggcc gagaccactc ccaccggtgg 120
agccaccagc tcagccacag cctctggctc ttcctccgga cgtcgtccac gtggtcgtcc 180
tgcaggttcc aaaaacaaac ccaaacctcc gacgattata actagagata gtcctaactg 240
ccttagatca cacgttcttg aagtcacctc cgggtcggac atatccgagg cagtctccac 300
ctacgccact cgtcgcggct gcggcggttg cattataagc ggcacgggtg cggtcactaa 360
cgtcacgata cggcaacctg cggctccggc tggaggagggt gtgattaccg tgcattggtc 420
gtttgacatt ttgtctttga ccggtactgc gcttccaccg cctgcaccac cgggagcagg 480
aggtttgacg gtgtatctag ccggagggtc aggacaagtt gtaggaggga atgtggctgg 540
ttcgtaatt gcttcgggac cggtagtggt gatggctgct tcttttgcaa acgcagttta 600
tgataggtta ccgattgaag aggaagaaac cccaccgccg agaaccaccg ggggtgcagca 660
gcagcagccg gaggcgtctc agtcgtcggg gggtacgggg agtggggccc aggcgtgtga 720
gtcaaacctc caagggtgaa atggtggagg aggtgttgct ttctacaatc ttggaatgaa 780
tatgaacaat tttcaattct ccgggggaga tatttacggt atgagcggcg gtagcggagg 840

aggtggtggc ggtgcgacta gaccgcggtt ttagagtttt agcgttttgg tgacaccttt 900
tgttgcgttt gcgtgtttga cctcaaacta ctaggctact agctatagcg gttgcgaaat 960
gcgaatatta gggt 974

```

<210> 2
 <211> 270
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G1073

<400>2

ES 2 502 943 T3

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

Pro Thr Gly Gly Ala Thr Ser Ser Ala Thr Ala Ser Gly Ser Ser Ser
20 25 30

Gly Arg Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys Asn Lys Pro Lys
35 40 45

Pro Pro Thr Ile Ile Thr Arg Asp Ser Pro Asn Val Leu Arg Ser His
50 55 60

Val Leu Glu Val Thr Ser Gly Ser Asp Ile Ser Glu Ala Val Ser Thr
65 70 75 80

Tyr Ala Thr Arg Arg Gly Cys Gly Val Cys Ile Ile Ser Gly Thr Gly
85 90 95

Ala Val Thr Asn Val Thr Ile Arg Gln Pro Ala Ala Pro Ala Gly Gly
100 105 110

Gly Val Ile Thr Leu His Gly Arg Phe Asp Ile Leu Ser Leu Thr Gly
115 120 125

Thr Ala Leu Pro Pro Pro Ala Pro Pro Gly Ala Gly Gly Leu Thr Val
130 135 140

Tyr Leu Ala Gly Gly Gln Gly Gln Val Val Gly Gly Asn Val Ala Gly
145 150 155 160

Ser Leu Ile Ala Ser Gly Pro Val Val Leu Met Ala Ala Ser Phe Ala
165 170 175

Asn Ala Val Tyr Asp Arg Leu Pro Ile Glu Glu Glu Glu Thr Pro Pro
180 185 190

Pro Arg Thr Thr Gly Val Gln Gln Gln Gln Pro Glu Ala Ser Gln Ser
195 200 205

Ser Glu Val Thr Gly Ser Gly Ala Gln Ala Cys Glu Ser Asn Leu Gln
210 215 220

Gly Gly Asn Gly Gly Gly Gly Val Ala Phe Tyr Asn Leu Gly Met Asn
225 230 235 240

Met Asn Asn Phe Gln Phe Ser Gly Gly Asp Ile Tyr Gly Met Ser Gly
245 250 255

Gly Ser Gly Gly Gly Gly Gly Gly Ala Thr Arg Pro Ala Phe
260 265 270

<210> 3
<211> 1473
<212> ADN
<213> Arabidopsis thaliana

ES 2 502 943 T3

<220>
<223> G1067

<400> 3

5

```
tctcaagctt ctctctcctt tttttcccat agcacatcag aatcgctaaa tacgactcct 60
atgcaaagaa gaagctactt ctttctcttg ccctaattaa tctacctaac taggggtttcc 120

tcttaccttt catgagagag atcatttaac ataagtcacc ttttttatat cttttgcttc 180
gtctttaatt tagttctgtt cttggtctgt ttctatatatt tgtcggcttg cgtaaccgat 240
cacaccttaa tgcttttagct attgtttcct caaaatcatg agttttgact tctcgatctg 300
agttttcttt ttctctcttt acgctcttct tcacctagct accaatatat gaacgagcag 360
gatcaagaat cgagaaattg atttgagctg gcgaataagc agtggtggga tagggaatta 420
gtagatgcgg cggcgatgga aggcggttac gagcaaggcg gtggagcttc tagatacttc 480
cataacctct ttagaccgga gattcaccac caacagcttc aaccgcaggg cgggatcaat 540
cttatcgacc agcatcatca tcagcaccag caacatcaac aacaacaaca accgtcggat 600
gattcaagag aatctgacca ttcaacaaa gatcatcatc aacagggctg acccgattca 660
gacccgaata catcaagctc agcaccggga aaacgtccac gtggacgtcc accaggatct 720
aagaacaaag ccaagccacc gatcatagta actcgtgata gcccacacgc gcttagatct 780
cacgttcttg aagtatctcc tggagctgac atagttgaga gtgtttccac gtacgctagg 840
aggagagggg gaggcgtctc cgttttagga ggaaacggca ccgtatctaa cgtcactctc 900
cgtcagccag tcaactcctg aaatggcggg ggtgtgtccg gaggaggagg agttgtgact 960
ttacatggaa ggtttgagat tctttcgcta acggggactg ttttgccacc tcctgcaccg 1020
cctggtgccg gtggtttgtc tatattttta gccggagggc aaggtcaggt ggtcggagga 1080
agcgtttgtg cttccccttat tgcattcagct ccggttatac taatggcggc ttcgttctca 1140
aatgcggttt tcgagagact accgattgag gaggaggaag aagaagggtg tgggtggcga 1200
ggaggaggag gaggagggcc accgcagatg caacaagctc catcagcatc tccgccgtct 1260

ggagtgaccg gtcagggaca gttaggaggt aatgtgggtg gttatgggtt ttctggtgat 1320
cctcatttgc ttggatgggg agctggaaca cttcaagac caccttttta attgaatttt 1380
aatgtccgga aatttatgtg tttttatcat cttgaggagt cgtctttcct ttgggatatt 1440
tggtgtttta tgtttagttg atatgcatat ttt 1473
```

<210> 4
<211> 311
<212> PRT
<213> Arabidopsis thaliana

10

<220>
<223> Polipéptido G1067

15

<400> 4

ES 2 502 943 T3

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

ES 2 502 943 T3

Pro Ala Pro Pro Gly Ala Gly Gly Leu Ser Ile Phe Leu Ala Gly Gly
 195 200 205

Gln Gly Gln Val Val Gly Gly Ser Val Val Ala Pro Leu Ile Ala Ser
 210 215 220

Ala Pro Val Ile Leu Met Ala Ala Ser Phe Ser Asn Ala Val Phe Glu
 225 230 235 240

Arg Leu Pro Ile Glu Glu Glu Glu Glu Gly Gly Gly Gly Gly Gly
 245 250 255

Gly Gly Gly Gly Gly Pro Pro Gln Met Gln Gln Ala Pro Ser Ala Ser
 260 265 270

Pro Pro Ser Gly Val Thr Gly Gln Gly Gln Leu Gly Gly Asn Val Gly
 275 280 285

Gly Tyr Gly Phe Ser Gly Asp Pro His Leu Leu Gly Trp Gly Ala Gly
 290 295 300

Thr Pro Ser Arg Pro Pro Phe
 305 310

<210> 5
 <211> 1383
 <212> ADN
 <213> Arabidopsis thaliana

<220>
 <223> G2153

<400> 5

ES 2 502 943 T3

```

ttcttgctta gtatcattct ttgtcgtggt cttttaatta accttttgca atttgtcttg 60
tgttttctcac aacacaaaaa ctgttaaaag tggttaaaaaa tcaagatctg aaaaatctta 120
tcaccgcttc taggtttttc agtttttttt cttccttttc ctgatctaaa ttaacttata 180
tttcttaggg tttcacttct tgaaacattt aatcagaatt aattaacctc tctagggctt 240
tcatggcgaa tccatggtgg acaggacaag tgaacctatc cggcctcgaa acgacgccgc 300
ctggttcctc tcagttaaag aaaccagatc tccacatctc catgaacatg gccatggact 360
caggtcacaa taatcatcac catcaccaag aagtcgataa caacaacaac gacgacgata 420
gagacaactt gagtggagac gaccacgagc cacgtgaagg agccgtagaa gccccacgc 480
gccgtccacg tggacgtcct gctggttcca agaacaaacc aaagccaccg atcttcgtca 540
ctcgcgattc tccaaatgct ctcaagagcc atgtcatgga gatcgctagt gggactgacg 600
tcatcgaaac cctagctact ttgctaggc ggcgtcaacg tggcatctgc atcttgagcg 660
gaaatggcac agtggctaac gtcaccctcc gtcaaccctc gaccgctgcc gttgcgcgcg 720
ctcctggtgg tgcggctggt ttggctttac aagggaggtt tgagattctt tctttaaccg 780

gttctttctt gccaggaccg gctccacctg gttccaccgg tttaacgatt tacttagccg 840
gtggtcaagg tcaggttggt ggaggaagcg tgggtggccc attgatggca gcagggtccg 900
tgatgctgat cgccgccacg ttctctaacg cgacttacga gagattgcca ttggaggagg 960
aagaggcagc agagagaggc ggtggtggag gcagcggagg agtggttccg gggcagctcg 1020
gaggcggagg ttcgccacta agcagcggtg ctggtggagg cgacggtaac caaggacttc 1080
cgggtgtataa tatgccggga aatcttggtt ctaatggtgg cagtgggtgga ggaggacaga 1140
tgagcggcca agaagcttat ggttgggctc aagctagggtc aggattttta cgtgcgttaa 1200
aatgggtttt aatttacaga agttaacaat aagattataa tgatgtttat tatgatgatg 1260
aaaaccagtc agttgctact tgttactagt gagctatata gtttgtggac attatattat 1320
gttctctctt gactatgatt attatttgct aaatttcact tagctaaaaa aaaaaaaaaa 1380
aaa 1383

```

<210> 6
 <211> 315
 <212> PRT
 <213> Arabidopsis thaliana

 <220>
 <223> Polipéptido G2153

 <400> 6

5

10

ES 2 502 943 T3

Met Ala Asn Pro Trp Trp Thr Gly Gln Val Asn Leu Ser Gly Leu Glu
 1 5 10 15
 Thr Thr Pro Pro Gly Ser Ser Gln Leu Lys Lys Pro Asp Leu His Ile
 20 25 30
 Ser Met Asn Met Ala Met Asp Ser Gly His Asn Asn His His His His
 35 40 45
 Gln Glu Val Asp Asn Asn Asn Asn Asp Asp Asp Arg Asp Asn Leu Ser
 50 55 60
 Gly Asp Asp His Glu Pro Arg Glu Gly Ala Val Glu Ala Pro Thr Arg
 65 70 75 80
 Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys Asn Lys Pro Lys Pro Pro
 85 90 95
 Ile Phe Val Thr Arg Asp Ser Pro Asn Ala Leu Lys Ser His Val Met
 100 105 110
 Glu Ile Ala Ser Gly Thr Asp Val Ile Glu Thr Leu Ala Thr Phe Ala
 115 120 125
 Arg Arg Arg Gln Arg Gly Ile Cys Ile Leu Ser Gly Asn Gly Thr Val
 130 135 140

ES 2 502 943 T3

Ala Asn Val Thr Leu Arg Gln Pro Ser Thr Ala Ala Val Ala Ala Ala
145 150 155 160

Pro Gly Gly Ala Ala Val Leu Ala Leu Gln Gly Arg Phe Glu Ile Leu
165 170 175

Ser Leu Thr Gly Ser Phe Leu Pro Gly Pro Ala Pro Pro Gly Ser Thr
180 185 190

Gly Leu Thr Ile Tyr Leu Ala Gly Gly Gln Gly Gln Val Val Gly Gly
195 200 205

Ser Val Val Gly Pro Leu Met Ala Ala Gly Pro Val Met Leu Ile Ala
210 215 220

Ala Thr Phe Ser Asn Ala Thr Tyr Glu Arg Leu Pro Leu Glu Glu Glu
225 230 235 240

Glu Ala Ala Glu Arg Gly Gly Gly Gly Gly Ser Gly Gly Val Val Pro
245 250 255

Gly Gln Leu Gly Gly Gly Gly Ser Pro Leu Ser Ser Gly Ala Gly Gly
260 265 270

Gly Asp Gly Asn Gln Gly Leu Pro Val Tyr Asn Met Pro Gly Asn Leu
275 280 285

Val Ser Asn Gly Gly Ser Gly Gly Gly Gly Gln Met Ser Gly Gln Glu
290 295 300

Ala Tyr Gly Trp Ala Gln Ala Arg Ser Gly Phe
305 310 315

<210> 7
<211> 1361
<212> ADN
<213> Arabidopsis thaliana

<220>
<223> G2156

<400> 7

ES 2 502 943 T3

```

ttttttttcc ctttcctcgt tcaaaaaaag tacttgcaga gtcactcact ctcagtctca    60
gcacatgaat taatttgaag cttccctaga attctttcac atcaattaat acgacaccgt    120
ctcgggtgaa gaatctctcc tctcttgccc taaagcgagt tagggtttaa cacacaaagc    180
atacccttta gatttgtgtc tcttagctct gtttttgcg gcttgtgtaa ccgatcaact    240
caagctattg gtcctcacc tcttgaaatt tgacttctcc aatggatctc aaagtttctc    300
ttatatgaat tctatcttca ccttcacaat atctttatat atatgagcca caagaacaag    360
aagagtcagt agatgcggct gccatggacg gtggttacga tcaatccgga ggagcttcta    420

gatactttca caacctcttc aggcctgagc ttcacacca gcttcaacct cagcctcaac    480
ttaccctttt gcctcagcct cagcctcaac ctcagcctca gcagcagaat tcagatgatg    540
aatctgactc caacaaggat ccgggttccg acccagttac ctctggttca accgggaaac    600
gtccacgtgg acgtcctccg ggatccaaga acaagccgaa gccaccggtg atagtgacta    660
gagatagccc caacgtgctt agatctcatg ttcttgaagt ctcatctgga gccgacatag    720
tcgagagcgt taccacttac gctcgcagga gaggaagagg agtctccatt ctcagtggta    780
acggcacggg ggctaacgtc agtctccggc agccggcaac gacagcggct catggggcaa    840
atggtggaac cggagggtgt gtggctctac atggaaggtt tgagatactt tcctcacag    900
gtacggtgtt gccgccctt gcgccgccag gatccggtgg tctttctatc tttctttccg    960
gcgttcaagg tcaggtgatt ggaggaaacg tgggtggctcc gcttgtggct tcgggtccag   1020
tgatactaatt ggctgcatcg ttctctaatt caactttcga aaggcttccc cttgaagatg   1080
aaggaggaga aggtggagag ggaggagaag ttggagaggg aggaggagga gaaggtggct   1140
caccgccggc cacgtcatca tcaccaccat ctggagccgg tcaaggacag ttaagaggta   1200
acatgagtggt ttatgatcag ttgcccgtg atcctcattt gcttggttgg ggagccgcag   1260
ccgcagccgc accaccaaga ccagcctttt agaattgaaa attatgtccg taacatagct   1320
gtaaccaaatt ttcatttctc aaaattaaaa gaaaaaaaaa a                        1361

```

<210> 8
 <211> 302
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G2156

<400> 8

ES 2 502 943 T3

Met Asp Gly Gly Tyr Asp Gln Ser Gly Gly Ala Ser Arg Tyr Phe His
 1 5 10 15
 Asn Leu Phe Arg Pro Glu Leu His His Gln Leu Gln Pro Gln Pro Gln
 20 25 30
 Leu His Pro Leu Pro Gln Pro Gln Pro Gln Pro Gln Pro Gln Gln Gln
 35 40 45
 Asn Ser Asp Asp Glu Ser Asp Ser Asn Lys Asp Pro Gly Ser Asp Pro
 50 55 60
 Val Thr Ser Gly Ser Thr Gly Lys Arg Pro Arg Gly Arg Pro Pro Gly
 65 70 75 80
 Ser Lys Asn Lys Pro Lys Pro Pro Val Ile Val Thr Arg Asp Ser Pro
 85 90 95

ES 2 502 943 T3

Asn Val Leu Arg Ser His Val Leu Glu Val Ser Ser Gly Ala Asp Ile
 100 105 110
 Val Glu Ser Val Thr Thr Tyr Ala Arg Arg Arg Gly Arg Gly Val Ser
 115 120 125
 Ile Leu Ser Gly Asn Gly Thr Val Ala Asn Val Ser Leu Arg Gln Pro
 130 135 140
 Ala Thr Thr Ala Ala His Gly Ala Asn Gly Gly Thr Gly Gly Val Val
 145 150 155
 Ala Leu His Gly Arg Phe Glu Ile Leu Ser Leu Thr Gly Thr Val Leu
 165 170 175
 Pro Pro Pro Ala Pro Pro Gly Ser Gly Gly Leu Ser Ile Phe Leu Ser
 180 185 190
 Gly Val Gln Gly Gln Val Ile Gly Gly Asn Val Val Ala Pro Leu Val
 195 200 205
 Ala Ser Gly Pro Val Ile Leu Met Ala Ala Ser Phe Ser Asn Ala Thr
 210 215 220
 Phe Glu Arg Leu Pro Leu Glu Asp Glu Gly Gly Glu Gly Gly Glu Gly
 225 230 235 240
 Gly Glu Val Gly Glu Gly Gly Gly Gly Glu Gly Gly Pro Pro Pro Ala
 245 250 255
 Thr Ser Ser Ser Pro Pro Ser Gly Ala Gly Gln Gly Gln Leu Arg Gly
 260 265 270
 Asn Met Ser Gly Tyr Asp Gln Phe Ala Gly Asp Pro His Leu Leu Gly
 275 280 285
 Trp Gly Ala Ala Ala Ala Ala Ala Pro Pro Arg Pro Ala Phe
 290 295 300

<210> 9
 <211> 1160
 <212> ADN
 <213> Oryza sativa

<220>
 <223> G3399

<400> 9

ES 2 502 943 T3

```

cgggtgaggag cgagttgtag aggaggaggg gcaccgatca accggaagc tggcgccgg 60
gaagtgggtgg gagcgccag atgacacggc catggccggg atggaccctg gcggggcg 120
cgccggcgcc ggcagctcac ggtacttcca ccatctgctc cgaccgcagc agccgtcgcc 180

gctgtcaccg ctgtcgccga catcccatgt caagatggag cactccaaga tgtcaccga 240
caagagcccc gtggcgagg gagatcacgc gggagggagt ggaagcggcg gcgtcgccgg 300
tgaccaccag ccgtcgctgt cggccatggt gcccgctcag ggtggcagcg gcagcgccgg 360
cggtagtggtc tcgggtgggc cgacgcggcg cccgcgcggg cgccgccccg ggtccaagaa 420
caagccgaag ccgcccata tcgtgacgcg cgacagcccc aacgcgctgc actcgcacgt 480
gctcgaggtc gccggcgggc ccgacgtcgt cgactgcgtg gccgagtacg cccgcgcgg 540
agggcgcggc gtgtgcgtgc tgagcgcgcg cggcgccgtc gtcaacgtgg cgctcgcgca 600
gccggcgcg tcgccggcg gcagcatggt ggccacgctg cggggccggg tcgagatcct 660
atctctcacg ggcacgggtc tgccgcctcc cgcgccaccg ggcgcgagcg gcctcacgt 720
gttcctctcc ggcggccagg gccaggtgat cggcggcagc gtggtgggccc cgctggtcgc 780
cgcgggggccc gtcgtcctga tggcggcctc attcgcgaac gccgtgtacg agcggctgcc 840
gctggagggc gaggaagagg aggtcgccgc gcccgccgcc ggaggcgaag cacaagatca 900
agtggcaca tcagctggac cccagggca gcaaccggcg gcgtcacagt cctccggcgt 960
gacaggaggc gacggcaccg gcggcgccgg tggcatgtcg ctctacaacc tcgccgggaa 1020
tgtgggaggc tatcagctcc ccggagacaa cttcggaggt tggagcggcg ccggcgccgg 1080
cggagtcagg ccaccgttct gacccatgtc ttagcatcca gttcaaaaat tctccaaatt 1140
aagaattgcg cagtgcaggc 1160

```

<210> 10
 <211> 336
 <212> PRT
 <213> Oryza sativa

<220>
 <223> Polipéptido G3399

<400> 10

ES 2 502 943 T3

Met Ala Gly Met Asp Pro Gly Gly Gly Gly Ala Gly Ala Gly Ser Ser
 1 5 10 15
 Arg Tyr Phe His His Leu Leu Arg Pro Gln Gln Pro Ser Pro Leu Ser
 20 25 30
 Pro Leu Ser Pro Thr Ser His Val Lys Met Glu His Ser Lys Met Ser
 35 40 45
 Pro Asp Lys Ser Pro Val Gly Glu Gly Asp His Ala Gly Gly Ser Gly
 50 55 60
 Ser Gly Gly Val Gly Gly Asp His Gln Pro Ser Ser Ser Ala Met Val
 65 70 75 80
 Pro Val Glu Gly Gly Ser Gly Ser Ala Gly Gly Ser Gly Ser Gly Gly
 85 90 95

ES 2 502 943 T3

Pro Thr Arg Arg Pro Arg Gly Arg Pro Pro Gly Ser Lys Asn Lys Pro
100 105 110

Lys Pro Pro Ile Ile Val Thr Arg Asp Ser Pro Asn Ala Leu His Ser
115 120 125

His Val Leu Glu Val Ala Gly Gly Ala Asp Val Val Asp Cys Val Ala
130 135 140

Glu Tyr Ala Arg Arg Arg Gly Arg Gly Val Cys Val Leu Ser Gly Gly
145 150 155 160

Gly Ala Val Val Asn Val Ala Leu Arg Gln Pro Gly Ala Ser Pro Pro
165 170 175

Gly Ser Met Val Ala Thr Leu Arg Gly Arg Phe Glu Ile Leu Ser Leu
180 185 190

Thr Gly Thr Val Leu Pro Pro Pro Ala Pro Pro Gly Ala Ser Gly Leu
195 200 205

Thr Val Phe Leu Ser Gly Gly Gln Gly Gln Val Ile Gly Gly Ser Val
210 215 220

Val Gly Pro Leu Val Ala Ala Gly Pro Val Val Leu Met Ala Ala Ser
225 230 235 240

Phe Ala Asn Ala Val Tyr Glu Arg Leu Pro Leu Glu Gly Glu Glu Glu
245 250 255

Glu Val Ala Ala Pro Ala Ala Gly Gly Glu Ala Gln Asp Gln Val Ala
260 265 270

Gln Ser Ala Gly Pro Pro Gly Gln Gln Pro Ala Ala Ser Gln Ser Ser
275 280 285

Gly Val Thr Gly Gly Asp Gly Thr Gly Gly Ala Gly Gly Met Ser Leu
290 295 300

Tyr Asn Leu Ala Gly Asn Val Gly Gly Tyr Gln Leu Pro Gly Asp Asn
305 310 315 320

Phe Gly Gly Trp Ser Gly Ala Gly Ala Gly Gly Val Arg Pro Pro Phe
325 330 335

<210> 11
<211> 987
<212> ADN
<213> Oryza sativa

<220>
<223> G3407

<400> 11

ES 2 502 943 T3

```

tcatcatcat catcacttgc atatcgaatt attaataata tcgaagatgg caggccttga 60
tttgggcacc agctacctcc accaccacca atcactgcat ctccgccacg acgatggcgg 120
cgccgggtcc gacgacggcg gccacgacga cctctcgccg gggagcggcg gtggcggcgg 180
gcccagcagc acggccgggtg gcgccgggat cggcggcggc gaggtcgtcg ctcgccgccc 240
ccgcgggcgc ccgccgggt ccaagaacaa gcccaagccg ccggtgatca tcaccagga 300
gagcgccaac gcgctcaggg cgcatactct cgaggtagcc gccggttgcg atgtgttcga 360
ggcgtgacg gcgtacgcgc gccgccggca gcgcgggggtg tgcgtgctct cggcggcggg 420
gacagtggcg aacgtcacgc tccggcagcc gcagtcggcg cagcccgggc cggcctcgcc 480
ggcgggtggcg acgctgcacg gcaggttcga gatactctcc ctcgcgggct ctttcctgcc 540
cccggccgcg ccgccgggcg ccaccagcct cgccgcgttc ctcgccggcg ggcaggggca 600
ggtcgtcggg ggcagcgtcg ccggcgcgct catcgcgggc gggcccgtcg tcgtcgtcgc 660
cgcgctgttc agcaacgtgg cgtacgagag gctgccgctc gaggacggcg acgaggtggt 720
ccccccggcg ccggcagga gcgaccaggg cggcggcggc agcggcggca tgccaccatt 780
aggcgttgat ccgtcgggcg gcgccgccac cggtgggctc ccgttcttca acatgccgtt 840
cgggatgccg ccaatgccgg tggacggcca cgccggctgg cctggcgccg gcgtcgggag 900
gccaccgttc tcatgatgga tggatcccca tattccggcg agcggccggc ggcgaggtgg 960
tcggcaagat tgaagacatg gacatgg 987

```

<210> 12
 <211> 289
 <212> PRT
 <213> Oryza sativa

<220>
 <223> Polipéptido G3407

<400> 12

```

Met Ala Gly Leu Asp Leu Gly Thr Ser Tyr Leu His His His Gln Ser
1           5           10           15

Leu His Leu Arg His Asp Asp Gly Gly Ala Gly Ser Asp Asp Gly Gly
20           25           30

His Asp Asp Leu Ser Pro Gly Ser Gly Gly Gly Gly Gly Pro Ser Ser
35           40           45

Thr Ala Gly Gly Ala Gly Ile Gly Gly Gly Glu Val Val Ala Arg Arg
50           55           60

Pro Arg Gly Arg Pro Pro Gly Ser Lys Asn Lys Pro Lys Pro Pro Val
65           70           75           80

```

ES 2 502 943 T3

Ile Ile Thr Arg Glu Ser Ala Asn Ala Leu Arg Ala His Ile Leu Glu
85 90 95
Val Ala Ala Gly Cys Asp Val Phe Glu Ala Leu Thr Ala Tyr Ala Arg
100 105 110
Arg Arg Gln Arg Gly Val Cys Val Leu Ser Ala Ala Gly Thr Val Ala
115 120 125
Asn Val Thr Leu Arg Gln Pro Gln Ser Ala Gln Pro Gly Pro Ala Ser
130 135 140
Pro Ala Val Ala Thr Leu His Gly Arg Phe Glu Ile Leu Ser Leu Ala
145 150 155 160
Gly Ser Phe Leu Pro Pro Pro Ala Pro Pro Gly Ala Thr Ser Leu Ala
165 170 175
Ala Phe Leu Ala Gly Gly Gln Gly Gln Val Val Gly Gly Ser Val Ala
180 185 190
Gly Ala Leu Ile Ala Ala Gly Pro Val Val Val Val Ala Ala Ser Phe
195 200 205
Ser Asn Val Ala Tyr Glu Arg Leu Pro Leu Glu Asp Gly Asp Glu Val
210 215 220
Val Pro Pro Ala Pro Ala Gly Ser Asp Gln Gly Gly Gly Gly Ser Gly
225 230 235 240
Gly Met Pro Pro Leu Gly Val Asp Pro Ser Gly Gly Ala Ala Thr Gly
245 250 255
Gly Leu Pro Phe Phe Asn Met Pro Phe Gly Met Pro Pro Met Pro Val
260 265 270
Asp Gly His Ala Gly Trp Pro Gly Ala Gly Val Gly Arg Pro Pro Phe
275 280 285
Ser

<210> 13
<211> 869
<212> ADN
<213> Glycine max

<220>
<223> G3456

<400> 13

ES 2 502 943 T3

```

tcgacaactt caacttttgt tatggccaac cggtggtgga ccgggtcggt ggggtctagag    60

aactctggcc actcgatgaa aaaaccggat ctgggggtttt ccatgaacga gagtacggtg    120
acggggaacc atataggaga agaagatgag gacagagaaa acagcgacga gccaagagag    180
ggagctattg acgtcgccac cacgcgccgc cctaggggac gtccaccggg ctccagaaac    240
aagccgaaac cgccgatatt cgtcaccga gacagcccta acgcgctgcg gagccacgtc    300
atggagattg ccgtcggagc cgacatcgcc gactgcgtgg cgcagttcgc tcggaggcgc    360
cagcgcgggg tttccattct cagcggcagc gggaccgtcg tcaacgtcaa tctccggcaa    420
cccacggcac ccggcgccgt catggcgctc cacggccgct tcgacatcct ctccctcacc    480
ggctcctttc tccctgggcc gtccctccc ggcgccaccg ggctcacaat ctacctcgcc    540
ggaggccagg ggcagatcgt cggcggcgga gtggtgggcc cgctcgtggc ggcgggcccc    600
gtattggtaa tggcggctac tttttccaat gctacgtatg aaagattgcc tttagaggat    660
gatgatcagg aacaacacgg cggcggaggc ggaggagggt cgccgcagga aaaaaccggg    720
gggtccggcg aggcgtcgtc gtcgatttcg gtttataaca ataatgttcc tccgagttta    780
ggctctccga atgggcaaca tctgaaccat gaagcttatt cttctccttg gggtcattct    840
cctcatgcca gacctcttt ctaattagc                                     869

```

<210> 14
 <211> 262
 <212> PRT
 <213> Glycine max

<220>
 <223> Polipéptido G3456

<400> 14

ES 2 502 943 T3

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

<210> 15
<211> 1596
<212> ADN
<213> Glycine max
<220>

ES 2 502 943 T3

<223> G3459

<400> 15

```

ctgtcgcgtg ggaacaaaat ggctgcattg tgagttcttt gtccccctca acctcatttc 60
aattctctct cccccccatt cttacttcac ccgcgcccc tcccccgccc gctcccgtcc 120
cttttctttc tctgcactcc atctttcttt ccaaaaccca cccttttcta ttctctttcc 180
tcttcctcct tttcccttct ttttatttcc ttacactcac aacatttccc ttaaaataaa 240
cataaacaaa ccagcactgt tcttgacccc caaaaaaaaa aaatctctac tatttattaa 300
ctatattaat tcctccataa tataatcatt tgttttcctt gttttctggt ttctcttata 360
atatataacc ttcttttata ttttttttct gttttgcacc ttgtgattgt gagttatatc 420
tatttatatt tatatatcat tctctctctt ttttttggat gtgtctatgg ctggtttgga 480
tttaggaagc gcctcacgct ttgttcaaaa cttcacaga ccagacttgc acttgcaaca 540
aaatttccag cagcaccagg accagcagca ccagcgtgat ttggaggagc agaaaactcc 600

tccgaatcac agaatggggg cgccgttcga cgatgatagc gatgatagaa gcccgggcct 660
ggagctcact tcaggctcctg gcgacatcgt cggacggcgc ccgcgtggca ggccctcctgg 720
gtcgaagaac aagcctaagc cgcccgctcat aatcacccgg gagagcgcca acacgctgag 780
ggcgcacatc ctcgaggtcg gaagcggctc cgacgtcttc gactgtgtca ccgcgtatgc 840
ccggcggcgc cagcgtggga tctgcgtcct cagtggcagc ggcaccgtca ccaatgtcag 900
tctccggcag cctgcagctg ccgggtgccg cgtcacgctg cacggcagg tgcagattct 960
ctccctctct ggctcgttcc tcccgccgcc ggctccgccg ggagccacca gcctcacaat 1020
ctacctggcc ggcgggcagg ggcaggttgt cggaggaaac gtcactggag aattaaccgc 1080
agcagggcca gtaatcgtca tcgcagcgtc gttcaccaac gtggcttacg agaggttacc 1140
cttagaagaa gatgaacaac agcagcaaca acagcagctt cagattcagc cacctgcaac 1200
gacgtcgtct caaggaaaca acaacaaca taaccctttc cccgaccctt cttcaggact 1260
tcccttcttc aatttaccac tcaatatgca gaatgttcag ttaccagttg agggttgggc 1320
tgtaaaccct gcttcacgtc cacaaccttt ttgagagttc atgaagatgt tgacggagga 1380
tttatatcac aaaaggcttt atattatttt aaggtcagca aattaatatt catggactac 1440
aacatatata taaactatat gtttttctt cttcttcatt ttattttggt tttttcttat 1500
gttggttaat gatataatat gacatgataa ttattatgta gtctgatttt catctccttg 1560
gaattttata tacttatttc ccctgttaaa aaaaaa 1596

```

5

<210> 16

<211> 295

<212> PRT

10 <213> Glycine max

<220>

<223> Polipéptido G3459

15 <400> 16

ES 2 502 943 T3

Met	Ala	Gly	Leu	Asp	Leu	Gly	Ser	Ala	Ser	Arg	Phe	Val	Gln	Asn	Leu
1				5					10					15	
His	Arg	Pro	Asp	Leu	His	Leu	Gln	Gln	Asn	Phe	Gln	Gln	His	Gln	Asp
			20					25					30		
Gln	Gln	His	Gln	Arg	Asp	Leu	Glu	Glu	Gln	Lys	Thr	Pro	Pro	Asn	His
		35					40					45			
Arg	Met	Gly	Ala	Pro	Phe	Asp	Asp	Asp	Ser	Asp	Asp	Arg	Ser	Pro	Gly
	50					55					60				
Leu	Glu	Leu	Thr	Ser	Gly	Pro	Gly	Asp	Ile	Val	Gly	Arg	Arg	Pro	Arg
65					70					75					80
Gly	Arg	Pro	Pro	Gly	Ser	Lys	Asn	Lys	Pro	Lys	Pro	Pro	Val	Ile	Ile
				85					90					95	

ES 2 502 943 T3

Thr Arg Glu Ser Ala Asn Thr Leu Arg Ala His Ile Leu Glu Val Gly
 100 105 110
 Ser Gly Ser Asp Val Phe Asp Cys Val Thr Ala Tyr Ala Arg Arg Arg
 115 120 125
 Gln Arg Gly Ile Cys Val Leu Ser Gly Ser Gly Thr Val Thr Asn Val
 130 135 140
 Ser Leu Arg Gln Pro Ala Ala Ala Gly Ala Val Val Thr Leu His Gly
 145 150 155 160
 Arg Phe Glu Ile Leu Ser Leu Ser Gly Ser Phe Leu Pro Pro Pro Ala
 165 170 175
 Pro Pro Gly Ala Thr Ser Leu Thr Ile Tyr Leu Ala Gly Gly Gln Gly
 180 185 190
 Gln Val Val Gly Gly Asn Val Ile Gly Glu Leu Thr Ala Ala Gly Pro
 195 200 205
 Val Ile Val Ile Ala Ala Ser Phe Thr Asn Val Ala Tyr Glu Arg Leu
 210 215 220
 Pro Leu Glu Glu Asp Glu Gln Gln Gln Gln Gln Gln Leu Gln Ile
 225 230 235 240
 Gln Pro Pro Ala Thr Thr Ser Ser Gln Gly Asn Asn Asn Asn Asn Asn
 245 250 255
 Pro Phe Pro Asp Pro Ser Ser Gly Leu Pro Phe Phe Asn Leu Pro Leu
 260 265 270
 Asn Met Gln Asn Val Gln Leu Pro Val Glu Gly Trp Ala Val Asn Pro
 275 280 285
 Ala Ser Arg Pro Gln Pro Phe
 290 295

<210> 17
 <211> 862
 <212> ADN
 <213> Glycine max

<220>
 <223> G3460

<400> 17

ES 2 502 943 T3

tatggctggt ttggatttag gaagcgcgtc acgctttgtt cagaatcttc acttaccgga	60
cttgcacttg caacaaaatt accagcaacc ccggcacaag cgcgattcgg aggagcaaga	120
gactcctccg aaccgggaa cagcgtggc gccgttcgac aacgatgatg aaaaaagcca	180
gggcctggag ctggcttcag gccctgggga catcgttgga cggcgcccac gcggcagacc	240
ttccgggtcc aagaacaagc cgaagccacc ggtgataatc acccgggaga gcgccaacac	300
gctgagggcg cacattctcg aggtaggaag cggctccgac gtcttcgact gtgtcaccgc	360
ttatgcgcgg cggcgccagc gcgggatctg cgtcctcagc ggcatggta ccgtcaccaa	420
tgctagtctc cggcagcctg cggctgccgg agccgtcgtc aggctgcacg gaaggttcga	480
gattctctct ctctccggct cgttcctccc gccgccggct ccgccgggag ccaccagtct	540
cacaatctac ctgcgcggcg ggcagggcca ggtcgtcgga ggaaacgtcg tgggagaatt	600
aaccgcggca gggccagtaa tcgtcatcgc agcatcgttc accaacgtgg cttacgagag	660
gtccccctta gaagaagatg aacaacaaca acaacagctt cagattcagt caccgcgaac	720
gacgtcatct caaggaaaca acaacaataa ccctttccct gacccttctt caggacttcc	780
cttcttcaac ttaccactca atatgcagaa tgttcagtta ccacctttt gagggttcat	840
gaatctgata atatgagact ga	862

<210> 18
 <211> 276
 <212> PRT
 <213> Glycine max

<220>
 <223> Polipéptido G3460

<400> 18

ES 2 502 943 T3

Met Ala Gly Leu Asp Leu Gly Ser Ala Ser Arg Phe Val Gln Asn Leu
 1 5 10 15
 His Leu Pro Asp Leu His Leu Gln Gln Asn Tyr Gln Gln Pro Arg His
 20 25 30
 Lys Arg Asp Ser Glu Glu Gln Glu Thr Pro Pro Asn Pro Gly Thr Ala
 35 40 45
 Leu Ala Pro Phe Asp Asn Asp Asp Lys Ser Gln Gly Leu Glu Leu
 50 55 60
 Ala Ser Gly Pro Gly Asp Ile Val Gly Arg Arg Pro Arg Gly Arg Pro
 65 70 75 80
 Ser Gly Ser Lys Asn Lys Pro Lys Pro Pro Val Ile Ile Thr Arg Glu
 85 90 95
 Ser Ala Asn Thr Leu Arg Ala His Ile Leu Glu Val Gly Ser Gly Ser
 100 105 110
 Asp Val Phe Asp Cys Val Thr Ala Tyr Ala Arg Arg Arg Gln Arg Gly
 115 120 125
 Ile Cys Val Leu Ser Gly Ser Gly Thr Val Thr Asn Val Ser Leu Arg
 130 135 140
 Gln Pro Ala Ala Ala Gly Ala Val Val Arg Leu His Gly Arg Phe Glu
 145 150 155 160
 Ile Leu Ser Leu Ser Gly Ser Phe Leu Pro Pro Pro Ala Pro Pro Gly
 165 170 175
 Ala Thr Ser Leu Thr Ile Tyr Leu Ala Gly Gly Gln Gly Gln Val Val
 180 185 190
 Gly Gly Asn Val Val Gly Glu Leu Thr Ala Ala Gly Pro Val Ile Val
 195 200 205
 Ile Ala Ala Ser Phe Thr Asn Val Ala Tyr Glu Arg Leu Pro Leu Glu
 210 215 220
 Glu Asp Glu Gln Gln Gln Gln Gln Leu Gln Ile Gln Ser Pro Ala Thr
 225 230 235 240
 Thr Ser Ser Gln Gly Asn Asn Asn Asn Asn Pro Phe Pro Asp Pro Ser
 245 250 255
 Ser Gly Leu Pro Phe Phe Asn Leu Pro Leu Asn Met Gln Asn Val Gln
 260 265 270
 Leu Pro Pro Phe
 275

ES 2 502 943 T3

<210> 19
<211> 1064
<212> ADN
<213> Oryza sativa

5

<220>
<223> G3408

10

<400> 19

```
gtggttttgg ttgattgcta ctctgtgcca tgtcgttctg cgagagggac atgaacaagg 60
agagcatgta ccaagaacgg gacgacatgg cggggatacg gttcgcgacg ccgccgctgc 120
ctcagcagca gcagcagcag cagctggtgg agtgcttctc cgacgaggtg gacagccgcg 180
ggagtggcgg cgagatgaag gatgccgtgg ggagcgggag tgggcagctg gtcgttgttg 240
gtggcgggga tggggcgagc atcgaggtgg cgaagaagag gagggggagg ccgccggggg 300
ccaagaacaa gccgaagcca cccgtggtga tcacgcggga ggcgagagccg gcggcgggcga 360
tgcgcccgca cgtgatcgag atccccggcg ggcgggacgt cgcgagaggcg ctcgcgcggt 420
tctcgagccg tcggaacctc gggatctgcg tgctcgccgg caccggcgcg gtcgccaacg 480
tgtcgtccg ccacccgtca cccgggggtcc cgggctcagc tccggctgcg atcgtgttcc 540
acggccggta cgagatcctc tccctgtcgg ccacgttcct gcctccggcc atgtcctccg 600

tggcgcccc a ggccgcggtc gccgccgcgg gcctctccat ctgctcgc ccggccgcacg 660
gccagatcgt cggcgggggc gtggcaggcc cgctctacgc cgcgaccacc gtcgtggtcg 720
tcgcccgccg cttcaccaac cccaccttc accgcctccc cgccgacgac gacgcgtcgg 780
tgtccgtctc ggtgtcactc tccggcagcg gcgacgcgga cgaacaccgg ggccaccagc 840
acaaacctga gccgcaagaa ccgcgccaac ttcgacggcc gccaccgcac ctgtcagcag 900
ccgccgccgt ctcagcagca cagccggtgg agccatgcgg cgcgcccattg tacgcctgcc 960
accctcagcc acaggaggtg atgtggccgc cgccggctcg tacgccgcac ccgccgccgc 1020
cgccgccgta ctaatccgac cgaattggta cgccattgcc acat 1064
```

<210> 20
<211> 334
<212> PRT
<213> Oryza sativa

15

<220>
<223> Polipéptido G3408

20

<400> 20

ES 2 502 943 T3

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

ES 2 502 943 T3

	165		170		175										
Leu	Ser	Leu	Ser	Ala	Thr	Phe	Leu	Pro	Pro	Ala	Met	Ser	Ser	Val	Ala
			180					185					190		
Pro	Gln	Ala	Ala	Val	Ala	Ala	Ala	Gly	Leu	Ser	Ile	Ser	Leu	Ala	Gly
		195					200					205			
Pro	His	Gly	Gln	Ile	Val	Gly	Gly	Ala	Val	Ala	Gly	Pro	Leu	Tyr	Ala
	210					215					220				
Ala	Thr	Thr	Val	Val	Val	Val	Ala	Ala	Ala	Phe	Thr	Asn	Pro	Thr	Phe
	225					230				235					240
His	Arg	Leu	Pro	Ala	Asp	Asp	Asp	Ala	Ser	Val	Ser	Val	Ser	Val	Ser
				245					250					255	
Leu	Ser	Gly	Ser	Gly	Asp	Ala	Asp	Glu	His	Arg	Gly	His	Gln	His	Lys
			260					265					270		
Pro	Glu	Pro	Gln	Glu	Pro	Arg	Gln	Leu	Arg	Arg	Pro	Pro	Pro	His	Leu
		275					280					285			
Ser	Ala	Ala	Ala	Ala	Val	Ser	Ala	Ala	Gln	Pro	Val	Glu	Pro	Cys	Gly
	290					295					300				
Ala	Pro	Met	Tyr	Ala	Cys	His	Pro	Gln	Pro	Gln	Glu	Val	Met	Trp	Pro
	305				310					315					320
Pro	Pro	Ala	Arg	Thr	Pro	His	Pro	Pro	Pro	Pro	Pro	Pro	Tyr		
				325					330						

<210> 21
 <211> 902
 <212> ADN
 <213> Oryza sativa

<220>
 <223> G3403

<400> 21

ES 2 502 943 T3

```

tcgaccagtg cattaattgt gcttctcctg gagcagggtc aagctggcga acaggtcgtg    60
gtggtcgggc cccatgggct tgccggagca gccgtccggc tcgtcgggcc ccaaggcgga    120
gctcccgggtg gccaaggagc cggaggcgag cccgacgggg ggcgcgggcg cggaaccacgc    180
cgacgagaac aacgaatccg gcggcgggcga gccgcgggag ggcgccgtgg tggcgggcgc    240
caaccggcgc ccccgcggcc gcccgccggg ctccaagaac aagccgaagc cgcccatctt    300
cgtgacgcgc gacagcccca acgcgctgcg cagccacgtc atggaggtgg ccggcgggcg    360
cgacgtcgcc gacgccatcg cgcagttctc gcgcggccgc cagcgcgggc tctgctgtgt    420
cagcgggcgc gggacggtcg ccaacgtcgc gctgcccag ccgtcggcgc ccggcgccgt    480

cgtcgccctg cacggccgct tcgagatcct ctccctcacc ggcaccttcc tccccggccc    540
ggcgctccg ggctccacgg ggctcaccgt ctacctgcc ggcgggcagg gccaggttgt    600
cggcggcagc gtcgtggggt cgctcatcgc cgcggggccc gtcattgtga tcgctccac    660
gttcgccaac gccacctacg agcgccctgc attggaggaa gaagaggagg gtcaggcccc    720
gccccatgcc ggcggcgcgc agccccctcat ggccggcggc cagggcatcg ccgacccttc    780
ggcgctgcca atgttcaacc tgccgccgag caacgggctc ggcgggcggc gcgacggttt    840
cccatgggcg gcgcaccccc gcccacgta ctgatgatg gcaatagcat ggaagattca    900
tt                                                                    902

```

<210> 22
 <211> 266
 <212> PRT
 <213> Oryza sativa

<220>
 <223> Polipéptido G3403

<400> 22

ES 2 502 943 T3

Met Gly Leu Pro Glu Gln Pro Ser Gly Ser Ser Gly Pro Lys Ala Glu
 1 5 10 15
 Leu Pro Val Ala Lys Glu Pro Glu Ala Ser Pro Thr Gly Gly Ala Ala
 20 25 30
 Ala Asp His Ala Asp Glu Asn Asn Glu Ser Gly Gly Gly Glu Pro Arg
 35 40 45
 Glu Gly Ala Val Val Ala Ala Pro Asn Arg Arg Pro Arg Gly Arg Pro
 50 55 60
 Pro Gly Ser Lys Asn Lys Pro Lys Pro Pro Ile Phe Val Thr Arg Asp
 65 70 75 80
 Ser Pro Asn Ala Leu Arg Ser His Val Met Glu Val Ala Gly Gly Ala
 85 90 95
 Asp Val Ala Asp Ala Ile Ala Gln Phe Ser Arg Arg Arg Gln Arg Gly
 100 105 110
 Val Cys Val Leu Ser Gly Ala Gly Thr Val Ala Asn Val Ala Leu Arg
 115 120 125
 Gln Pro Ser Ala Pro Gly Ala Val Val Ala Leu His Gly Arg Phe Glu
 130 135 140
 Ile Leu Ser Leu Thr Gly Thr Phe Leu Pro Gly Pro Ala Pro Pro Gly
 145 150 155 160
 Ser Thr Gly Leu Thr Val Tyr Leu Ala Gly Gly Gln Gly Gln Val Val
 165 170 175
 Gly Gly Ser Val Val Gly Ser Leu Ile Ala Ala Gly Pro Val Met Val
 180 185 190
 Ile Ala Ser Thr Phe Ala Asn Ala Thr Tyr Glu Arg Leu Pro Leu Glu
 195 200 205
 Glu Glu Glu Glu Gly Ser Gly Pro Pro Met Pro Gly Gly Ala Glu Pro
 210 215 220
 Leu Met Ala Gly Gly His Gly Ile Ala Asp Pro Ser Ala Leu Pro Met
 225 230 235 240
 Phe Asn Leu Pro Pro Ser Asn Gly Leu Gly Gly Gly Gly Asp Gly Phe
 245 250 255
 Pro Trp Ala Ala His Pro Arg Pro Pro Tyr
 260 265

<210> 23
 <211> 1153
 <212> ADN
 <213> Glycine max

<220>
<223> G3458

5 <400> 23

```

tcgcccacgc gtccgtacgg ctgcgagaag acgacagaag gggccacttt atttgtctct 60
ctctttccct tccaacctca tcccattccg ttttctctgc agtactcaat tgatcccttt 120
gtttttctat tcgtttctgag agctttgtgt gtatggccgg catagacttg gggttcagcat 180
cacattttgt tcatcatcgc cttgaacgcc ctgaccttga agacgatgag aaccaacaag 240
accaagacaa caaccttaac aatcacgaag ggcttgacct agttacacca aattcagggtc 300
ctgggtgatgt tgttggtcgc aggccaagag gaagacctcc aggttcaaag aacaagccaa 360
aaccaccagt tatcatcaca agagagagtg caaacacctc tagggctcac atccttgaag 420
ttagtagtgg ttgtgacgtc tttgaatcgg tcgctaccta tgcaaggaag cgacaaagag 480
ggatctgtgt cctcagtggg agtggcaccg tgaccaacgt gacattgagg cagccggccg 540
cggctgggtgc cgtcgtcacg ctgcacggaa ggtttgagat cctctctttg tcaggatcat 600
tcctcccacc tccagctcca ccagggtgcta caagtttgac tgtgttcctt ggtggaggac 660
agggccaagt ggtgggagga aatgttggtt gtcctttggt ggcttctggg cctgttattg 720
ttattgcttc atcttttact aatgtagcat atgagagggt gcctttggat gaagatgaat 780
ctatgcagat gcaacaaggg caatcatcag ctgggtgatg tagcggtgac catggtggtg 840
gagttagtaa taactctttt ccggatccgt cttccgggct tccattcttc aatttgccac 900
taaacatgcc tcagttacct gttgatgggt gggctggcaa ctctggtgga aggcaatctt 960

actgatccag agtctttggg ggcacaaagg tgagaagttg aattgatctc atatatttg 1020
gtcttctcta atctttcctc tgaatattgc ttgtgaagaa gtactgattt ttctattgaa 1080
gaaatcgttt gtttggctag gtttgttgta aggacgatca gtttctagga acaactgtaa 1140
aacgttttct ctt 1153

```

10 <210> 24
<211> 270
<212> PRT
<213> Glycine max

15 <220>
<223> Polipéptido G3458
<400> 24

ES 2 502 943 T3

Met Ala Gly Ile Asp Leu Gly Ser Ala Ser His Phe Val His His Arg
 1 5 10 15
 Leu Glu Arg Pro Asp Leu Glu Asp Asp Glu Asn Gln Gln Asp Gln Asp
 20 25 30
 Asn Asn Leu Asn Asn His Glu Gly Leu Asp Leu Val Thr Pro Asn Ser
 35 40 45
 Gly Pro Gly Asp Val Val Gly Arg Arg Pro Arg Gly Arg Pro Pro Gly
 50 55 60
 Ser Lys Asn Lys Pro Lys Pro Pro Val Ile Ile Thr Arg Glu Ser Ala
 65 70 75 80
 Asn Thr Leu Arg Ala His Ile Leu Glu Val Ser Ser Gly Cys Asp Val
 85 90 95
 Phe Glu Ser Val Ala Thr Tyr Ala Arg Lys Arg Gln Arg Gly Ile Cys
 100 105 110
 Val Leu Ser Gly Ser Gly Thr Val Thr Asn Val Thr Leu Arg Gln Pro
 115 120 125
 Ala Ala Ala Gly Ala Val Val Thr Leu His Gly Arg Phe Glu Ile Leu
 130 135 140
 Ser Leu Ser Gly Ser Phe Leu Pro Pro Pro Ala Pro Pro Gly Ala Thr
 145 150 155 160
 Ser Leu Thr Val Phe Leu Gly Gly Gly Gln Gly Gln Val Val Gly Gly
 165 170 175
 Asn Val Val Gly Pro Leu Val Ala Ser Gly Pro Val Ile Val Ile Ala
 180 185 190
 Ser Ser Phe Thr Asn Val Ala Tyr Glu Arg Leu Pro Leu Asp Glu Asp
 195 200 205
 Glu Ser Met Gln Met Gln Gln Gly Gln Ser Ser Ala Gly Asp Gly Ser
 210 215 220
 Gly Asp His Gly Gly Gly Val Ser Asn Asn Ser Phe Pro Asp Pro Ser
 225 230 235 240
 Ser Gly Leu Pro Phe Phe Asn Leu Pro Leu Asn Met Pro Gln Leu Pro
 245 250 255
 Val Asp Gly Trp Ala Gly Asn Ser Gly Gly Arg Gln Ser Tyr
 260 265 270

<210> 25
 <211> 918

ES 2 502 943 T3

<212> ADN
<213> Oryza sativa

5 <220>
<223> G3406

<400> 25

```

atggcagggtc tgcacctcgg caccgccgcg acgcgctacg tccaccagct ccaccacctc   60
caccccgacc tccagctgca gcacagctac gccaagcagc acgagccgtc cgacgacgac   120
cccaacggca gcggcgggcg cggaacagc aacggcgggc cgtacgggga ccatgacggc   180
gggtcctcgt cgtcagggtc tgccaccgac ggcgcggtcg gcgggcccgg cgacgtggtg   240
gcgcgccggc gcggggggcg cccgcctggc tccaagaaca agccgaagcc gccggtgatc   300
atcacgcggg agagcgccaa cacgtgcgc gccacatcc tggaggctcg gagcggtgctg   360
gacgtgttcg agtgctctc cacgtacgcg gccggcggc agcgcggtgt gtgctgtctg   420
agcggcagcg gcgtgggtac caacgtgacg ctgcgtcagc cgtcggcgcc cgcgggcgcc   480
gtcgtgtcgc tgcacgggag gttcgagatc ctgtcgtctt cgggctcctt cctcccgcgg   540
ccggctcccc ccggcgccac cagcctcacc atcttcctcg ccggggggcca gggacagggtc   600
gtcggcgcca acgtcgtcgg cgcgtcttac gccgcgggcc cggtcatcgt catcgcggcg   660
tccttcgcca acgtcgccta cgagcgctc cacttgagg aggaggagg gccgcccgcg   720
caggccggcc tgcagatgca gcagcccggc ggcgcgccg atgctggtgg catgggtggc   780
gcgttccccg cggaaccgct tgccgcggc ctcccgttct tcaacctgcc gctcaacaac   840
atgcccgggtg gcggcggtc acagctccct cccggcgccg acggccatgg ctgggccggc   900
gcacggccac cgttctga                                     918

```

10 <210> 26
<211> 305
<212> PRT
<213> Oryza sativa

15 <220>
<223> Polipéptido G3406

20 <400> 26

ES 2 502 943 T3

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

ES 2 502 943 T3

Gly Met Gly Gly Ala Phe Pro Pro Asp Pro Ser Ala Ala Gly Leu Pro
 260 265 270
 Phe Phe Asn Leu Pro Leu Asn Asn Met Pro Gly Gly Gly Gly Ser Gln
 275 280 285
 Leu Pro Pro Gly Ala Asp Gly His Gly Trp Ala Gly Ala Arg Pro Pro
 290 295 300
 Phe
 305

<210> 27
 <211> 1013
 <212> ADN
 <213> Oryza sativa

<220>
 <223> G3405

<400> 27

ctagctaatt gattgctagc ttgcaaggat ggatccgggtc acggcatcaa tacacggtca 60
 ccatcttcct ccaccgttca acaccgcga cttccatcac catctccagc agcagcagca 120
 ccagctgcat ctcaagaccg aggatgacca aggcggcggc actccgggtg tcttcggcag 180
 ccgcggcacc aagcgcgacc acgacgacga cgagaacagt ggcaacggcc atggaagcgg 240
 tgggtgacggc ggtgacctcg cgctggtacc cccctcgggt ggcgggccgg acggcgccgg 300
 gagcgagagc gccacgcgcc gcccgagggg acgcccggcg gggccaaga acaagccgaa 360
 gccaccgatc atcatcacca gggacagcgc caacacgctc cggacgcacg tcatggaggt 420
 ggccggcggc tgcgacatct ccgagagcat caccacgttc gcgcgacgcc ggcagcgccg 480
 ggtttgctg ctcagcggcg ccggcaccgt cactaacgtc acgctgcggc agcccgcac 540
 gcagggagcg gtcgttgcg tccacggccg gtccgagata ctctccctct ccggctcctt 600
 cctcccgccg cccgccccgc cggaggccac ggggctcacc gtctacctgg ccggaggcca 660
 gggccaggtc gtgggcggca gcgtcgctcg cgcgctgacc gcggctgggc ctgtggtgat 720
 aatggcggcg tcttttgca acgcggtgta cgagcggctg ccgttgagg acgacgagct 780
 actggcggct caagggcaag ccgacagcgc tgggttgctc gccgcggggc agcaagcggc 840
 gcagctcgcc ggcggggccg tcgatccaag cctcttccaa ggactaccac caaacctact 900
 cggaaacgtg cagctgccgc cggaagccgc ctacggatgg aaccctggag ccggcggtgg 960
 ccgcccggcg ccgttctgag atggatcgat tccgcgacag caacgcagca tag 1013

<210> 28
 <211> 316
 <212> PRT
 <213> Oryza sativa

<220>
 <223> Polipéptido G3405

<400> 28

ES 2 502 943 T3

Met Asp Pro Val Thr Ala Ser Ile His Gly His His Leu Pro Pro Pro
 1 5 10 15
 Phe Asn Thr Arg Asp Phe His His His Leu Gln Gln Gln Gln His Gln
 20 25 30
 Leu His Leu Lys Thr Glu Asp Asp Gln Gly Gly Gly Thr Pro Gly Val
 35 40 45
 Phe Gly Ser Arg Gly Thr Lys Arg Asp His Asp Asp Asp Glu Asn Ser
 50 55 60
 Gly Asn Gly His Gly Ser Gly Gly Asp Gly Gly Asp Leu Ala Leu Val
 65 70 75 80
 Pro Pro Ser Gly Gly Gly Pro Asp Gly Ala Gly Ser Glu Ser Ala Thr
 85 90 95
 Arg Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys Asn Lys Pro Lys Pro
 100 105 110
 Pro Ile Ile Ile Thr Arg Asp Ser Ala Asn Thr Leu Arg Thr His Val
 115 120 125
 Met Glu Val Ala Gly Gly Cys Asp Ile Ser Glu Ser Ile Thr Thr Phe
 130 135 140
 Ala Arg Arg Arg Gln Arg Gly Val Cys Val Leu Ser Gly Ala Gly Thr
 145 150 155 160
 Val Thr Asn Val Thr Leu Arg Gln Pro Ala Ser Gln Gly Ala Val Val
 165 170 175
 Ala Leu His Gly Arg Phe Glu Ile Leu Ser Leu Ser Gly Ser Phe Leu
 180 185 190
 Pro Pro Pro Ala Pro Pro Glu Ala Thr Gly Leu Thr Val Tyr Leu Ala
 195 200 205
 Gly Gly Gln Gly Gln Val Val Gly Gly Ser Val Val Gly Ala Leu Thr
 210 215 220
 Ala Ala Gly Pro Val Val Ile Met Ala Ala Ser Phe Ala Asn Ala Val
 225 230 235 240
 Tyr Glu Arg Leu Pro Leu Glu Asp Asp Glu Leu Leu Ala Ala Gln Gly
 245 250 255
 Gln Ala Asp Ser Ala Gly Leu Leu Ala Ala Gly Gln Gln Ala Ala Gln
 260 265 270

ES 2 502 943 T3

Leu Ala Gly Gly Ala Val Asp Pro Ser Leu Phe Gln Gly Leu Pro Pro
275 280 285

Asn Leu Leu Gly Asn Val Gln Leu Pro Pro Glu Ala Ala Tyr Gly Trp
290 295 300

Asn Pro Gly Ala Gly Gly Gly Arg Pro Ala Pro Phe
305 310 315

<210> 29
<211> 1143
<212> ADN
<213> Oryza sativa

<220>
<223> G3400

<400> 29

```
ctctgcaggt gaggtgaggt gaggggtggg atcgagggat cgagaagctg gcggcgagca 60
agtgggtggga gcgcgccgat gagggcgcca tggccgggat ggatcccacc ggcggcggtg 120
gcggcgggcgg cgtggcgggc cactacctac acatgctccg gcgcgagcag caccagccac 180
tgtccccggc aggtgacgtc aaggcggagc ggtccatgct gtcgccggat gagagccccg 240
gcgcggacgc cgacctagga tcggaccacc cgacgtcgtc ggccatggtg gcggcgagg 300
acagcggcgg cggcagcggg tcgggtggcc cgatgcggcg cccccgcggg aggccgctgg 360
gctccaagaa caagcccaag ccgcccata tcgtgacgcg ggacagcccc aacgcgttcc 420
actccacgt cctcgaggtc gccgcgggaa ccgacatcgt cgagtgcgtc tgcgagttcg 480
cgcgccgccg cggccgcggc gtctccgtgc tcagcgggtg cggcgccgtc gccaacgtcg 540
cgctccgccg gccaggcgcg tcgcccccg gcagcctggt cgccaccatg cgcggccagt 600
tcgagatcct gtccctcacg ggcaccgtcc tccgcgcc ccgcgccccc agcgccagcg 660
gcctcaccgt cttcctctcc ggcgggcagg gccagggtgt cggcgggagc gtggccggcc 720
agctcatcgc cgcggggcca gtcttctca tggccgcctc gttcgccaat gccgtctacg 780
agcgtctgcc actcgatggg gaggatccgg aggcagaggc tgccgccgcc acccctcccg 840
gcgatgcggc gcagccaacc ggcaccac caccgcagca gcagcccaca gcctcgagc 900
cctctgaggt gaccgccggt gacggcgggc gcggcgggcg tctcggcatg tatcttggag 960
gccatgtggg atcctaccag cagcagcagc agcaacttcc cggaccagga gacaacttcg 1020
gtagctggag cggcagcatc aggcgcggc cattctgatc caaacacctc aaatcaagct 1080
ctccccaac aacgccatgc atgtctaaat cctcacaaga ttactccaa gaagacgaag 1140
ctg 1143
```

<210> 30
<211> 322
<212> PRT
<213> Oryza sativa

<220>
<223> Polipéptido G3400

<400> 30

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

ES 2 502 943 T3

245

250

255

Thr Gly Pro Pro Pro Pro Gln Gln Gln Pro Thr Ala Ser Gln Ser Ser
260 265 270

Glu Val Thr Ala Gly Asp Gly Gly Gly Gly Gly Gly Leu Gly Met Tyr
275 280 285

Leu Gly Gly His Val Gly Ser Tyr Gln Gln Gln Gln Gln Gln Leu Pro
290 295 300

Gly Pro Gly Asp Asn Phe Gly Ser Trp Ser Gly Ser Ile Arg Pro Pro
305 310 315 320

Pro Phe

5 <210> 31
<211> 1152
<212> ADN
<213> Oryza sativa

10 <220>
<223> G3404
<400> 31

ES 2 502 943 T3

```

accgtgagag agggagacag atggatccgg tgacggcggc ggcggcgcac gggggtgggc 60
accaccacca ccaccacttc ggagcgccac cggcggcggc gttccaccac caccggttcc 120
accacggcgg cggggcgcac taccggcg cgttcagca gtttcaggag gagcagcagc 180
agcttgtggc ggcggcggcg gcggctggtg ggatggcgaa gcaggagctg gtggatgaga 240
gcaacaacac catcaacagc ggcgggagca acgggagcgg cggggaggag cagaggcagc 300
agtccgggga ggagcagcac cagcaagggg cggcggcgcc ggtggtgatc cggcgtccca 360
ggggccgccc cgccggctcc aagaacaagc ccaagcctcc ggtcatcatc acgcgcgaca 420
gcgccagcgc gctgcgggcg cacgtcctcg aggtcgctc cgggtgcgac ctcgtcgaca 480
gcgtcgccac gttcgcgcg cgcgccagg tcggtgtctg cgtgctcagc gccaccggcg 540
ccgtcaccaa cgtctccgtc cggcagcccc gcgcgggccc cggcgccgtc gtcaacctca 600

ccggccgctt cgacatcctc tcgctgtccg gtccttcct cccgccgcgc gcgcctccct 660
ccgccaccgg cctcaccgtc tacgtctccg gcggccaggg gcaggctctg ggcggcacgg 720
tcgccggacc gtcctcgcc gtcggccccg tcgtcatcat ggccgcctcg ttcgggaacg 780
ccgcctacga gcgcctcccc ctcgaggacg acgagccgcc gcagcacatg gcgggaggcg 840
gccagtcttc gccgcggccg ccgccgctgc cattaccacc acaccagcag ccgattcttc 900
aagaccatct gccacacaac ctgatgaacg gaatccacct ccccggcgac gccgcctacg 960
gctggaccag cggcggcggc ggcggcggcc gcgcggcgcc gtactgatca acatcgatct 1020
cgccggagag aaaaaaatg gaggagaagg atcggagcga gccgtgcatg gtgtaggatg 1080

aattaagcta agagttaatt tcttcttccg cttttgctaa tcatgatgct ctcgtgttgt 1140
ttaatctgtg gc 1152

```

<210> 32
 <211> 328
 <212> PRT
 <213> Oryza sativa

<220>
 <223> Polipéptido G3404

<400> 32

ES 2 502 943 T3

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

ES 2 502 943 T3

Pro Pro Ala Pro Pro Ser Ala Thr Gly Leu Thr Val Tyr Val Ser Gly
 210 215 220
 Gly Gln Gly Gln Val Val Gly Gly Thr Val Ala Gly Pro Leu Ile Ala
 225 230 235 240
 Val Gly Pro Val Val Ile Met Ala Ala Ser Phe Gly Asn Ala Ala Tyr
 245 250 255
 Glu Arg Leu Pro Leu Glu Asp Asp Glu Pro Pro Gln His Met Ala Gly
 260 265 270
 Gly Gly Gln Ser Ser Pro Pro Pro Pro Pro Leu Pro Leu Pro Pro His
 275 280 285
 Gln Gln Pro Ile Leu Gln Asp His Leu Pro His Asn Leu Met Asn Gly
 290 295 300
 Ile His Leu Pro Gly Asp Ala Ala Tyr Gly Trp Thr Ser Gly Gly Gly
 305 310 315 320
 Gly Gly Gly Arg Ala Ala Pro Tyr
 325

<210> 33
 <211> 987
 <212> ADN
 <213> Oryza sativa

<220>
 <223> G3407

<400> 33

ES 2 502 943 T3

```

tcatcatcat catcacttgc atatcgaatt attaataata tcgaagatgg caggccttga      60
tttgggcacc agctacctcc accaccacca atcactgcat ctccgccacg acgatggcgg      120
cgccggctcc gacgacggcg gccacgacga cctctcgccg gggagcggcg gtggcggcgg      180
gcccagcagc acggccgggtg gcgccgggat cggcggcggc gaggtcgtcg ctccgccccc      240
ccgcggccgc ccgccgggct ccaagaacaa gccaagccg ccggtgatca tcaccaggga      300
gagcgccaac gcgctcaggg cgcataacct cgaggtagcc gccggttgcg atgtgttcga      360
ggcgctgacg gcgtacgcgc gccgccggca gcgcgggggtg tgcgtgctct cggcggcggg      420
gacagtggcg aacgtcacgc tccggcagcc gcagtcggcg cagcccgggc cggcctcgcc      480
ggcgggtggcg acgctgcacg gcaggttcga gatactctcc ctccggggct ccttcctgcc      540
cccgcccgcg ccgccgggcg ccaccagcct cgccgcgttc ctccgccggc ggcaggggca      600
ggtcgtcggt ggacgcgtcg ccggcgcgct catcgcgggc gggcccgtcg tcgtcgtcgc      660
cgcgtcgttc agcaacgtgg cgtacgagag gctgccgctc gaggacggcg acgaggtggt      720
ccccccggcg ccggcaggga gcgaccaggg cggcggcggc agcggcggca tgccaccatt      780

aggcgttgat ccgtcgggcg gcgccgccac cgggtgggctc ccgttcttca acatgccgtt      840
cgggatgccg ccaatgccgg tggacggcca cgccggctgg cctggcgccg gcgtcgggag      900
gccaccgttc tcatgatgga tggatcccca tattccggcg agcggccggc ggcgaggtgg      960
tcggcaagat tgaagacatg gacatgg                                     987

```

<210> 34
 <211> 289
 <212> PRT
 <213> Oryza sativa

<220>
 <223> Polipéptido G3407

<400> 34

ES 2 502 943 T3

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

ES 2 502 943 T3

	195		200		205
Ser	Asn	Val	Ala	Tyr	Glu
	210				215
		Leu	Pro	Leu	Glu
				Asp	Gly
				220	Asp
					Glu
					Val
Val	Pro	Pro	Ala	Pro	Ala
225				230	Gly
					Ser
					Asp
					Gln
					235
					Gly
					Gly
					Gly
					Ser
					240
Gly	Met	Pro	Pro	Leu	Gly
				245	Val
					Asp
					Pro
					Ser
					250
					Gly
					Gly
					Ala
					Ala
					Thr
					255
					Gly
Gly	Leu	Pro	Phe	Phe	Asn
			260		Met
					Pro
					Phe
					265
					Gly
					Met
					Pro
					Pro
					Met
					270
					Pro
					Val
Asp	Gly	His	Ala	Gly	Trp
		275			Pro
					Gly
					280
					Ala
					Gly
					Val
					Gly
					Arg
					285
					Pro
					Pro
					Phe

Ser

<210> 35
 <211> 1611
 <212> ADN
 <213> Glycine max

<220>
 <223> G3462

<400> 35

ES 2 502 943 T3

```

ccacgcgtcc gcactctaag gtacttcccc actgttccca ttttatgccc gtactagtct 60
gaccctttta gtggcataga ttgaggtaca attacaaaag catattaccc gtatcgatc 120
ctttctgcta caaccacccc gaaacacact gcgttttttt tgtgtcaaac ttcaaccggg 180
gctttcttac ttctcttctc ttcaactttt tagtgttttt ttttttttt tctcgtgtcg 240
ggtccttgag cttctatttg tttatttatt tccggcagcg agttgagtta attttcgtgt 300
catagtctca attaggtgga tgtgacttgt gtggttaggg ttacatgca aaaacgcatg 360
tgcttcgggt ctaattcatc ctgaaacaag gttttgggac agagatctga atatttcaag 420
ctatctattt tgtttctcta tctagaaggt tttgagaaat caaagatggc gaatagatgg 480
tgggctggga atgtgggaat gataagagag caagagttga tggaaaacag caacaacaac 540
aacaacaaca acaacgctac tactactaca ccgacgacga ggagcaacag caacactaac 600
gcgaacacca acaccaacac gaccgaggaa gaggtgagca gggataacgg agaggaccag 660
aaccaaaacc tcggcagcca cgaagggtcg gagcccgga gacgcggtcg gaggccacgt 720
ggcaggccag cggggtccaa gaacaagccc aagccgccc tagtcataat ttttttaagc 780
cccaacgcgc tcggaagcca cgtcctggaa atcgccctcg gccgcgatgt cgccgagagc 840
atcgccgcct tcgccaaccg ccgccaccgt ggcgtgtcgg tcctcagcgg gagtggcatt 900
gtagccaacg tcactctccg ccagccccgc gccccgcgc gcgtcataac cctccacggg 960

aggttcgaga tactctccct ctggggtgcc tttttgccgt cccctcgcc gtccggcgcc 1020
accggactga ccgtctacct agccggcggg caggggcagg ttgtcggcgg caacgtggcg 1080
ggctctctcg tcgcctccgg accggtgatg gtgatcgccg ccactttcgc taatgccact 1140
tatgagaggt tgcctctgga ggatgatcaa ggtgaggagg aaatgcaagt gcagcagcag 1200
cagcagcagc agcaacagca gcagcagcag cagcagcaac aacaatctca aggtttgggg 1260
gaacaggttt caatgcctat gtataatttg cctcctaatt tgctacacaa tggtcagaac 1320
atgcctcatg atgtgttctg gggagctcca cctcgccctc ctccttcctt ctgatcacc 1380
ttgccaatat gatcatgtct ttaatctctc actgacttgc gaattaagta ctatgttaat 1440
taattttctc cggtttttct tgcaagcata gctagctagc tagcaagggt agttattagg 1500
atggttttgt taatttgtgc ttcttagaga ctcgagtcaa gtagatgatg ttcttatctt 1560
taatatactt tgtagtacta ctggtttgtt tattgttttt tttaaaaaa a 1611

```

<210> 36
 <211> 302
 <212> PRT
 <213> Glycine max

<220>
 <223> Polipéptido G3462

<400> 36

ES 2 502 943 T3

Met Ala Asn Arg Trp Trp Ala Gly Asn Val Gly Met Ile Arg Glu Gln
 1 5 10 15
 Glu Leu Met Glu Asn Ser Asn Asn Asn Asn Asn Asn Ala Thr
 20 25 30
 Thr Thr Thr Pro Thr Thr Arg Ser Asn Ser Asn Thr Asn Ala Asn Thr
 35 40 45
 Asn Thr Asn Thr Thr Glu Glu Glu Val Ser Arg Asp Asn Gly Glu Asp
 50 55 60
 Gln Asn Gln Asn Leu Gly Ser His Glu Gly Ser Glu Pro Gly Ser Ser
 65 70 75 80
 Gly Arg Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys Asn Lys Pro Lys
 85 90 95
 Pro Pro Ile Val Ile Ile Phe Leu Ser Pro Asn Ala Leu Arg Ser His
 100 105 110
 Val Leu Glu Ile Ala Ser Gly Arg Asp Val Ala Glu Ser Ile Ala Ala
 115 120 125
 Phe Ala Asn Arg Arg His Arg Gly Val Ser Val Leu Ser Gly Ser Gly
 130 135 140

Ile Val Ala Asn Val Thr Leu Arg Gln Pro Ala Ala Pro Ala Gly Val
145 150 155 160

Ile Thr Leu His Gly Arg Phe Glu Ile Leu Ser Leu Ser Gly Ala Phe
165 170 175

Leu Pro Ser Pro Ser Pro Ser Gly Ala Thr Gly Leu Thr Val Tyr Leu
180 185 190

Ala Gly Gly Gln Gly Gln Val Val Gly Gly Asn Val Ala Gly Ser Leu
195 200 205

Val Ala Ser Gly Pro Val Met Val Ile Ala Ala Thr Phe Ala Asn Ala
210 215 220

Thr Tyr Glu Arg Leu Pro Leu Glu Asp Asp Gln Gly Glu Glu Glu Met
225 230 235 240

Gln Val Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
245 250 255

Gln Gln Gln Gln Ser Gln Gly Leu Gly Glu Gln Val Ser Met Pro Met
260 265 270

Tyr Asn Leu Pro Pro Asn Leu Leu His Asn Gly Gln Asn Met Pro His
275 280 285

Asp Val Phe Trp Gly Ala Pro Pro Arg Pro Pro Pro Ser Phe
290 295 300

<210> 37
<211> 1003
<212> ADN
<213> Oryza sativa

<220>
<223> G3401

<400> 37

ES 2 502 943 T3

```

ggatatatag gccgcaggta gaagctggcg agccggtggt gggacgaggg gcggctcagc    60
ctgccggtga cagccgccga cccatccctc ggccttggca acaacgacac caccggcgcc    120
ggtgacgacg acatggcgtc caaggagcca agcggcgacc acgaccacga gatgaacggg    180
accagcgccg gggggcgcgga gccaaggac ggcgcggtgg tgaccggccg caaccggcgc    240
ccccgcggac ggccgccggg ctccaagaac aagcccaagc cgcccatctt cgtgacgcgg    300
gacagcccga acgcgctgcg cagccacgtc atggaggtgg cggcgggcgc cgatgtcgcc    360
gagtccatcg cgcacttcgc gcggcgggcg cagcgcgggc tctgcgtgct cagcgggggc    420
ggcaccgtga ccgacgtggc cctgcgccag ccggccgcgc cgagcgccgt ggtggcgctc    480
cgtgggcggt tcgagatcct gtccctgacg gggacgttcc tgccggggcc ggcgcgcgcg    540

ggctccaccg ggctgaccgt gtacctcgcc ggcgggcagg ggcaggtggt gggcggcagc    600
gtggtgggga cgctcaccgc ggcgggggcg gtcattgtga tcgcctccac cttcgccaac    660
gccacctacg agaggctgcc gctggatcag gaggaggagg aagcagcggc aggcggcatg    720
atggcgccgc cgccactcat ggccggcgcc gccgatccac tacttttcgg cgggggaatg    780
cacgacgccg ggcttgctgc atggcaccat gcccgcctc cgccgccgcc gccctactag    840
ctctctagct agattaatta attaaagctt atgaattaaa tccttgctta attaatggc    900
atatatatat actcacatga tatatatgat caacaagcta tatatagcta gccagctcga    960
gctaattaat taacaaaagg tgagccaata taatgcaaga tcg                                1003

```

<210> 38
 <211> 235
 <212> PRT
 <213> Oryza sativa

<220>
 <223> Polipéptido G3401

<400> 38

Met Ala Ser Lys Glu Pro Ser Gly Asp His Asp His Glu Met Asn Gly
1 5 10 15

Thr Ser Ala Gly Gly Gly Glu Pro Lys Asp Gly Ala Val Val Thr Gly
20 25 30

Arg Asn Arg Arg Pro Arg Gly Arg Pro Pro Gly Ser Lys Asn Lys Pro
35 40 45

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

His Val Met Glu Val Ala Gly Gly Ala Asp Val Ala Glu Ser Ile Ala
65 70 75 80

His Phe Ala Arg Arg Arg Gln Arg Gly Val Cys Val Leu Ser Gly Ala
85 90 95

Gly Thr Val Thr Asp Val Ala Leu Arg Gln Pro Ala Ala Pro Ser Ala
100 105 110

Val Val Ala Leu Arg Gly Arg Phe Glu Ile Leu Ser Leu Thr Gly Thr
115 120 125

Phe Leu Pro Gly Pro Ala Pro Pro Gly Ser Thr Gly Leu Thr Val Tyr
130 135 140

Leu Ala Gly Gly Gln Gly Gln Val Val Gly Gly Ser Val Val Gly Thr
145 150 155 160

Leu Thr Ala Ala Gly Pro Val Met Val Ile Ala Ser Thr Phe Ala Asn
165 170 175

Ala Thr Tyr Glu Arg Leu Pro Leu Asp Gln Glu Glu Glu Glu Ala Ala
180 185 190

Ala Gly Gly Met Met Ala Pro Pro Pro Leu Met Ala Gly Ala Ala Asp
195 200 205

Pro Leu Leu Phe Gly Gly Gly Met His Asp Ala Gly Leu Ala Ala Trp
210 215 220

His His Ala Arg Pro Pro Pro Pro Pro Tyr
225 230 235

<210> 39
<211> 803
<212> ADN
<213> Oryza sativa

ES 2 502 943 T3

<220>
<223> G3556

<400> 39

5

```
tcgacgggat ggggagcatc gacggccact cgctgcagca gcatcagggg tactcccacg 60
gcggcggcgc gggagggagc aacgaggagg aggaggcgtc gccgccgcc gccggtggct 120
cggctacggg gtcggcgggc cgccggccga gggggaggcc gccgggctcc aagaacaagc 180
cgaagccgcc cgtcgtggtg acgcgggaga gcccacacgc gatgcgttcc cacgtgctgg 240
agatcgccag cggcgccgac atcgtcgagg ccatcgcggg cttctcccgc cgcaggcagc 300
gcggcgtctc cgtgctcagc gggagcggcg ccgtcaccaa cgtcacgctc cggcagcccg 360
cggggactgg ggccgccgcc gtcgcgctgc gggggagggt cgagatattg tccatgtctg 420
gcgcccttcct cccggcgccg gcgccgccag gggccacggg gctcgccgtg tacctcgccg 480
gcgggcaggg gcagggtggtg ggtgggagcg tcatggggga gctgatcgcg tcgggccccg 540
tcatggtgat cgcggccacg ttcggcaacg ccacgtacga gaggtgccg ctggaccagg 600
aaggcgagga gggcgccgtg ctgtccgggt cggagggcgc cgccgcgcag atggagcagc 660
agagcagcgg aggcgccgtc gtgccccgc cgatgtacgc cgccgtccag cagacgccgc 720
cgcacgacat gttcgggcag tgggggcatg cagcgggtggc tcggccgccg ccgacatcgt 780
tctagcactg gcaccgggta att 803
```

<210> 40
<211> 258
<212> PRT
<213> Oryza sativa

10

<220>
<223> Polipéptido G3556

15

<400> 40

Met Gly Ser Ile Asp Gly His Ser Leu Gln Gln His Gln Gly Tyr Ser

ES 2 502 943 T3

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

Ser Phe

<210> 41
<211> 1116

ES 2 502 943 T3

<212> ADN
<213> Arabidopsis thaliana

5 <220>
<223> G1069

<400> 41

```

ttggaaccct agaggccttt caagcaaadc atcagggtaa caatttcttg atctttcttt    60
ttagcgaatt tccagttttt ggtcaatcat ggcaaaccct tgggtggacga accagagtgg    120
tttagcgggc atggtggacc attcgggtctc ctcaggccat caccaaaacc atcaccacca    180
aagtcttctt accaaaggag atcttggaat agccatgaat cagagccaag acaacgacca    240
agacgaagaa gatgatccta gagaaggagc cgttgaggtg gtcaaccgta gaccaagagg    300
tagaccacca ggatccaaaa acaaacccaa agctccaatc tttgtgacaa gagacagccc    360
caacgcactc cgtagccatg tcttgagat ctcgcacggc agtgacgtcg ccgacacaat    420
cgctcacttc tcaagacgca ggcaacgcgg cgtttgcggt ctcagcggga caggctcagt    480
cgctaacgtc accctccgcc aagccgccgc accaggaggt gtggtctctc tccaaggcag    540
gtttgaaatc ttatctttaa ccggtgcttt cctccctgga ccttccccac ccgggtcaac    600
cggtttaacg gtttacttag ccgggggtcca gggtcaggtc gttggaggta gcgttgtagg    660
cccactctta gccatagggt cgggtcatggt gattgctgct actttctcta acgctactta    720
tgagagattg cccatggaag aagaggaaga cgggtggcggc tcaagacaga ttcacggagg    780
cggtgactca ccgccagaa tcggtagtaa cctgcctgat ctatcaggga tggccggggc    840
aggctacaat atgccgccgc atctgattcc aaatggggct ggtcagctag ggcacgaacc    900
atatacatgg gtccacgcaa gaccaccta ctgactcagt gagccatttc tatatataat    960
ggtctatata aataaatata tagatgaata taagcaagca atttgaggta gtctattaca   1020
aagcttttgc tctggttga aaaataaata agtatcaaag ctttgtttgt tcttaatgga   1080
aatatagagc ttgggaagggt agaaagagac gacatt                               1116

```

10 <210> 42
<211> 281
<212> PRT
<213> Arabidopsis thaliana

15 <220>
<223> Polipéptido G1069

20 <400> 42

```

Met Ala Asn Pro Trp Trp Thr Asn Gln Ser Gly Leu Ala Gly Met Val
 1              5              10              15

Asp His Ser Val Ser Ser Gly His His Gln Asn His His His Gln Ser
          20              25              30

Leu Leu Thr Lys Gly Asp Leu Gly Ile Ala Met Asn Gln Ser Gln Asp
 35              40              45

```

Asn Asp Gln Asp Glu Glu Asp Asp Pro Arg Glu Gly Ala Val Glu Val
 50 55 60
 Val Asn Arg Arg Pro Arg Gly Arg Pro Pro Gly Ser Lys Asn Lys Pro
 65 70 75 80
 Lys Ala Pro Ile Phe Val Thr Arg Asp Ser Pro Asn Ala Leu Arg Ser
 85 90 95
 His Val Leu Glu Ile Ser Asp Gly Ser Asp Val Ala Asp Thr Ile Ala
 100 105 110
 His Phe Ser Arg Arg Arg Gln Arg Gly Val Cys Val Leu Ser Gly Thr
 115 120 125
 Gly Ser Val Ala Asn Val Thr Leu Arg Gln Ala Ala Ala Pro Gly Gly
 130 135 140
 Val Val Ser Leu Gln Gly Arg Phe Glu Ile Leu Ser Leu Thr Gly Ala
 145 150 155 160
 Phe Leu Pro Gly Pro Ser Pro Pro Gly Ser Thr Gly Leu Thr Val Tyr
 165 170 175
 Leu Ala Gly Val Gln Gly Gln Val Val Gly Gly Ser Val Val Gly Pro
 180 185 190
 Leu Leu Ala Ile Gly Ser Val Met Val Ile Ala Ala Thr Phe Ser Asn
 195 200 205
 Ala Thr Tyr Glu Arg Leu Pro Met Glu Glu Glu Glu Asp Gly Gly Gly
 210 215 220
 Ser Arg Gln Ile His Gly Gly Gly Asp Ser Pro Pro Arg Ile Gly Ser
 225 230 235 240
 Asn Leu Pro Asp Leu Ser Gly Met Ala Gly Pro Gly Tyr Asn Met Pro
 245 250 255
 Pro His Leu Ile Pro Asn Gly Ala Gly Gln Leu Gly His Glu Pro Tyr
 260 265 270
 Thr Trp Val His Ala Arg Pro Pro Tyr
 275 280

<210> 43
 <211> 1130
 <212> ADN
 <213> Arabidopsis thaliana

<220>
 <223> G1945

<400> 43

```

atttcccaaa gggattttacg aaaagtcctt ctcctctatc atctctttat tcaccccata    60
ccaacaacct ctacatcttc ttcttcttct tcctcctctt ttattttctt tttaaatacat    120
ttacacaaaa atccaaagac aaatctgaaa tctctaataa acaaataccat aaaataagaa    180
aaacaaagat gaaaggtgaa tacagagagc aaaagagtaa cgaaatgttt tccaagcttc    240
ctcatcatca acaacaacag caacaacaac aacaacaaca ctctcttacc tctcacttcc    300
acctctcctc caccgtaacc cccaccgtcg atgactcctc catcgaagtg gtccgacgtc    360
cacgtggcag accaccaggt tccaaaaaca aacctaaacc acccgtcttc gtcacacgtg    420
acaccgaccc tcctatgagt ccttacatcc tcgaagttcc ttcaggaaac gacgtcgtcg    480
aagccatcaa ccgtttctgc cgccgtaaat ccatcgaggt ctgcgtcctt agtggctctg    540
gctctgtagc taacgtcact ttacgtcagc catcacggc agctcttggc tctaccataa    600
ctttccatgg aaagtttgat ctctctccg tctccgaac gtttctcct cctccgcctc    660
gtacttctct gtctcctccc gtttctaact tcttcaccgt ctctctcgct ggacctcaag    720
gacaaatcat cggaggggtc gtcgctggc cacttatttc ggcaggaaca gtttacgtca    780
tcgccgcaag tttcaacaac ccttcttacc accggttacc ggcggaagaa gagcaaaaac    840
actcggcggg gacaggggaa agagaggagc aatctccgcc ggtctctggt ggcggtgaag    900
agtcaggaca gatggcggga agtgaggag agtcgtgtgg ggtatcaatg tacagttgcc    960
acatgggtgg ctctgatgtt atttgggccc ctacagccag agctccaccg ccatactaac   1020
caatccttct ttcacaaatc tctttcttct tttttttgtt tttttttgtt ttgggttagg   1080
atgaatcaag aaactagggt tttttttttt tttttttaaa aaaaaaaaaa   1130

```

5 <210> 44
 <211> 276
 <212> PRT
 <213> Arabidopsis thaliana

10 <220>
 <223> Polipéptido G1945

<400> 44

```

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

```

15

ES 2 502 943 T3

65		70		75		80									
Pro	Pro	Met	Ser	Pro	Tyr	Ile	Leu	Glu	Val	Pro	Ser	Gly	Asn	Asp	Val
				85					90					95	
Val	Glu	Ala	Ile	Asn	Arg	Phe	Cys	Arg	Arg	Lys	Ser	Ile	Gly	Val	Cys
			100					105					110		
Val	Leu	Ser	Gly	Ser	Gly	Ser	Val	Ala	Asn	Val	Thr	Leu	Arg	Gln	Pro
		115					120					125			
Ser	Pro	Ala	Ala	Leu	Gly	Ser	Thr	Ile	Thr	Phe	His	Gly	Lys	Phe	Asp
	130					135					140				
Leu	Leu	Ser	Val	Ser	Ala	Thr	Phe	Leu	Pro	Pro	Pro	Pro	Arg	Thr	Ser
145					150					155					160
Leu	Ser	Pro	Pro	Val	Ser	Asn	Phe	Phe	Thr	Val	Ser	Leu	Ala	Gly	Pro
				165					170					175	
Gln	Gly	Gln	Ile	Ile	Gly	Gly	Phe	Val	Ala	Gly	Pro	Leu	Ile	Ser	Ala
			180					185					190		
Gly	Thr	Val	Tyr	Val	Ile	Ala	Ala	Ser	Phe	Asn	Asn	Pro	Ser	Tyr	His
		195					200					205			
Arg	Leu	Pro	Ala	Glu	Glu	Glu	Gln	Lys	His	Ser	Ala	Gly	Thr	Gly	Glu
	210					215					220				
Arg	Glu	Gly	Gln	Ser	Pro	Pro	Val	Ser	Gly	Gly	Gly	Glu	Glu	Ser	Gly
225					230					235					240
Gln	Met	Ala	Gly	Ser	Gly	Gly	Glu	Ser	Cys	Gly	Val	Ser	Met	Tyr	Ser
				245					250					255	
Cys	His	Met	Gly	Gly	Ser	Asp	Val	Ile	Trp	Ala	Pro	Thr	Ala	Arg	Ala
			260					265					270		
Pro	Pro	Pro	Tyr												
		275													

<210> 45
 <211> 1050
 <212> ADN
 <213> Arabidopsis thaliana

<220>
 <223> G2155

<400> 45

ES 2 502 943 T3

```

ctcatatata ccaaccaaac ctctctctgc atctttatta acacaaaatt ccaaaagatt 60
aaatgttgtc gaagctccct acacagcgac acttgcacct ctctccctcc tctccctcca 120

tggaaaccgt cgggcggtcca cgtggcagac ctcgagggtc caaaaacaaa cctaaagctc 180
caatctttgt caccattgac cctcctatga gtccttacat cctcgaagtg ccatccggaa 240
acgatgtcgt tgaagcccta aaccgtttct gccgcggtaa agccatcggc ttttgcgctc 300
tcagtggctc aggtctccgtt gctgatgtca ctttgcgtca gccttctccg gcagctcctg 360
gctcaaccat tactttccac ggaaagtctg atcttctctc tgtctccgcc actttcctcc 420
ctcctctacc tcctacctcc ttgtccctc ccgtctccaa tttcttcacc gtctctctcg 480
ccggacctca ggggaaagtc atcgggtgat tcgtcgtcgg tcctctcggt gccgccggaa 540
ctgtttactt cgtcgccact agtttcaaga acccttccta tcaccggtta cctgctacgg 600
aggaagagca aagaaactcg gcggaagggg aagaggaggg acaatcgccg ccggtctctg 660
gaggtgggtg agagtcgatg tacgtgggtg gctctgatgt catttgggat cccaacgcca 720
aagctccatc gccgtactga ccacaaatcc atctcgttca aactagggtt tcttcttctt 780
tagatcatca agaatcaaca aaaagattgc attttagat tctttgtaat atcataattg 840
actcactctt taatctctct atcacttctt ctttagcttt ttctgcagtg tcaaacttca 900
catatttgta gtttgatttg actatcccca agttttgtat ttatcatcac aaatttttgc 960
ctgtctctaa tgggtgtttt ttcgtttgta taatcttatg cattgtttat tggagctcca 1020
gagattgaat gtataatata atggtttaat 1050

```

<210> 46
 <211> 225
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G2155

<400> 46

ES 2 502 943 T3

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

<210> 47
 <211> 1295
 <212> ADN
 <213> Arabidopsis thaliana
 <220>

ES 2 502 943 T3

<223> G1070

<400> 47

```

tcgaccagct tggatttcgt tgttcatcat tactactctc tttcttcttc tagctagcta    60
gttttgacag caaaataaga agcaaaaaaa aggtcaacta aaaaagatct gttcttagat    120
cactctcttc ttcttttttt gatccaattc caccattgaa tcatagatca tggatccagt    180
acaatctcat ggatcacaaa gctctctacc tcctcctttc cacgcaagag actttcaatt    240
acatcttcaa caacagcaac aagagttctt cctccaccat caccagcaac aaagaaacca    300
aaccgatggt gaccaacaag gaggatcagg aggaaaccga caaatcaaga tggatcgtga    360
agagacaagc gacaacatag acaacatagc taacaacagc ggtagtgaag gtaaagacat    420
agatatacac ggtgggttcag gagaaggagg tgggtggctcc ggaggagatc atcagatgac    480
aagaagacca agaggaagac cagcgggatc caagaacaaa caaaaccac cgattatcat    540
cacacgggac agcgcaaagc cgcttagaac ccacgtgatg gagatcggag atggctgcga    600
cttagtcgaa agcgttgcca cttttgcacg aagacgcaa cgcggcgttt gcgttatgag    660
cgggtactgga aatgttacta acgtcactat acgtcagcct ggatctcatc cttctcctgg    720

ctcggtagtt agtcttcacg gaaggttcga gattctatct ctctcaggat cttttctccc    780
tcctccggct ctcctacag ccaccggatt gagtgtttac ctcgctggag gacaaggaca    840
ggtaggttga ggaagcgtag ttgggtccgtt gttatgtgct ggtcctgtcg ttgtcatggc    900
tgcgtctttt agcaatgcgg cgtacgaaag gttgccttta gaggaagatg agatgcagac    960
gccggttcat ggcggaggag gaggaggatc attggagtcg ccgccaatga tgggacaaca   1020
actgcaacat cagcaacaag ctatgtcagg tcatcaaggg ttaccaccta atcttcttgg   1080
ttcggttcag ttgcagcagc aacatgatca gtcttattgg tcaacgggac gaccaccgta   1140
ttgatcaaat atacacacac actcataatc gttgctagct agctaacgat gaatcatgag   1200
tttagtggat atatatatga ttaaaagagg ttagcttatg aacattaata agagtttgga   1260
ttctatcgag cttcattatg tttgggtcat cgttc                                1295

```

5

<210> 48

<211> 324

<212> PRT

10 <213> Arabidopsis thaliana

<220>

<223> Polipéptido G1070

15

<400> 48

ES 2 502 943 T3

Met Asp Pro Val Gln Ser His Gly Ser Gln Ser Ser Leu Pro Pro Pro
 1 5 10
 Phe His Ala Arg Asp Phe Gln Leu His Leu Gln Gln Gln Gln Gln Glu
 20 25 30
 Phe Phe Leu His His His Gln Gln Gln Arg Asn Gln Thr Asp Gly Asp
 35 40 45
 Gln Gln Gly Gly Ser Gly Gly Asn Arg Gln Ile Lys Met Asp Arg Glu
 50 55 60
 Glu Thr Ser Asp Asn Ile Asp Asn Ile Ala Asn Asn Ser Gly Ser Glu
 65 70 75 80
 Gly Lys Asp Ile Asp Ile His Gly Gly Ser Gly Glu Gly Gly Gly Gly
 85 90 95
 Ser Gly Gly Asp His Gln Met Thr Arg Arg Pro Arg Gly Arg Pro Ala
 100 105 110
 Gly Ser Lys Asn Lys Pro Lys Pro Pro Ile Ile Ile Thr Arg Asp Ser
 115 120 125
 Ala Asn Ala Leu Arg Thr His Val Met Glu Ile Gly Asp Gly Cys Asp
 130 135 140
 Leu Val Glu Ser Val Ala Thr Phe Ala Arg Arg Arg Gln Arg Gly Val

[illegible]

<210> 49
<211> 1020
<212> ADN
<213> *Arabidopsis thaliana*

<220>
<223> G2657

<400> 49

ES 2 502 943 T3

```

tcaatacggg gccgacccg tagaccaata ctgctgatca ttctgttggt gcggtggcaa 60
ctgaaccgaa ccaagaagat tcggtggttag tccttgagcc gccgccatag ctgccatagc 120
ttgttgctgt cccatcatcg ggggagatcc cattccacca ccacctctc ctctccacc 180
gcctccttga actggcgtct gcatctcatc ttcttccaaa ggcagccttt cgtacgccgc 240
attgctaaaa gaagccgcca taaccaccac aggacccgaa cacaacaaag gtcccaccac 300
actacctcca acgacctgcc cttgtcctcc ggctaggtaa acgcttagtc cgggtggctgc 360

aggcggcgca ggcggaggca agaaagatcc cgaaagagag aggatttcaa accggccgtg 420
aaggctaacc accgagccag gtggcgatcc aggctgacgt atagtacgt tagtaacgt 480
tcctgtaccg ctcataacgc aaacgcctct ttggcggcgt ctagcgaacg tagccataca 540
gtcaactatg tcacatccgt ctcttatctc catgacgtga gttcgaagcg cgtttgcgt 600
gtctcttggt atgattattg gagctttagg ttgttcttg gatcctgctg gtcttctct 660
tggccttctt gtcactgtgt ctccacttcc tccaccaccg ctctctctt ctctccgtg 720
taaactcatc tctttacctt cgctaccgct gttggtatta gcgatgttgt ccatgttatc 780
gcttgctctt tcgcatcca tcttgataga tctattcaat attgaccctc cttgctgctc 840
gtgatcttga tcaagggttc tttgtggttg ctgatgatgg tggagaaaga actgttggtg 900
ttgttggtgt tgatgttgtt gttgatgttg ttgttggtgt tgaagatgta attggaaatc 960
tctagcatgg aaaggaggag gaagagagct ttgtgatcca tgagattgaa ctggatccat 1020

```

<210> 50
 <211> 339
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G2657

<400> 50

ES 2 502 943 T3

Met Asp Pro Val Gln Ser His Gly Ser Gln Ser Ser Leu Pro Pro Pro
 1 5 10 15
 Phe His Ala Arg Asp Phe Gln Leu His Leu Gln Gln Gln Gln Gln His
 20 25 30
 Gln Gln Gln His Gln Gln Gln Gln Gln Gln Gln Phe Phe Leu His His
 35 40 45
 His Gln Gln Pro Gln Arg Asn Leu Asp Gln Asp His Glu Gln Gln Gly
 50 55 60
 Gly Ser Ile Leu Asn Arg Ser Ile Lys Met Asp Arg Glu Glu Thr Ser
 65 70 75 80
 Asp Asn Met Asp Asn Ile Ala Asn Thr Asn Ser Gly Ser Glu Gly Lys
 85 90 95
 Glu Met Ser Leu His Gly Gly Glu Gly Gly Ser Gly Gly Gly Gly Ser
 100 105 110
 Gly Glu Gln Met Thr Arg Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys
 115 120 125
 Asn Lys Pro Lys Ala Pro Ile Ile Ile Thr Arg Asp Ser Ala Asn Ala
 130 135 140

Leu Arg Thr His Val Met Glu Ile Gly Asp Gly Cys Asp Ile Val Asp
 145 150 155 160
 Cys Met Ala Thr Phe Ala Arg Arg Arg Gln Arg Gly Val Cys Val Met
 165 170 175
 Ser Gly Thr Gly Ser Val Thr Asn Val Thr Ile Arg Gln Pro Gly Ser
 180 185 190
 Pro Pro Gly Ser Val Val Ser Leu His Gly Arg Phe Glu Ile Leu Ser
 195 200 205
 Leu Ser Gly Ser Phe Leu Pro Pro Pro Ala Pro Pro Ala Ala Thr Gly
 210 215 220
 Leu Ser Val Tyr Leu Ala Gly Gly Gln Gly Gln Val Val Gly Gly Ser
 225 230 235 240
 Val Val Gly Pro Leu Leu Cys Ser Gly Pro Val Val Val Met Ala Ala
 245 250 255
 Ser Phe Ser Asn Ala Ala Tyr Glu Arg Leu Pro Leu Glu Glu Asp Glu
 260 265 270
 Met Gln Thr Pro Val Gln Gly Gly Gly Gly Gly Gly Gly Gly Gly
 275 280 285
 Gly Met Gly Ser Pro Pro Met Met Gly Gln Gln Gln Ala Met Ala Ala
 290 295 300
 Met Ala Ala Ala Gln Gly Leu Pro Pro Asn Leu Leu Gly Ser Val Gln
 305 310 315 320
 Leu Pro Pro Pro Gln Gln Asn Asp Gln Gln Tyr Trp Ser Thr Gly Arg
 325 330 335
 Pro Pro Tyr

<210> 51
 <211> 1084
 <212> ADN
 <213> Arabidopsis thaliana

<220>
 <223> G1075

<400> 51

ttgtgtttg gtgctggcat ggctggcttc gatctaggca caacttctcg ctacgtccac 60
 aacgtcgatg gtggcggcgg cggacagttc accaccgaca accaccacga agatgacggt 120
 ggcgctggag gaaaccacca tcataccat cataatcata atcaccatca aggttttagat 180

ES 2 502 943 T3

```

ttaatagctt ctaatgataa ctctggacta ggcggcgggtg gaggaggagg gagcggtgac 240
ctcgtcatgc gtcggccacg tggccgtcca gctggatcga agaacaaacc gaagccgccg 300
gtgattgtca cgcgcgagag cgcaaacact cttagggctc acattcttga agttggaagt 360
ggctgcgacg ttttcgaatg tatctccact tacgctcgtc ggagacagcg cgggatttgc 420
gttttatccg ggacgggaac cgtcactaac gtcagcatcc gtcagcctac ggcggccgga 480
gctgttgtga ctctgcgggg tacttttgag attctttccc tctccggatc ttttcttccg 540
ccacctgctc ctccaggggc gactagcttg acgatattcc tcgctggagc tcaaggacag 600
gtcgtcggag gtaacgtagt tggtgagtta atggcggcgg ggcgggtaat ggtcatggca 660
gcgtctttta caaacgtggc ttacgaaagg ttgcctttgg acgagcatga ggagcacttg 720
caaagtggcg gcggcggagg tggagggaat atgtactcgg aagccactgg cgggtggcga 780
gggttgcctt tctttaattt gccgatgagt atgcctcaga ttggagtga aagttggcag 840
gggaatcacg ccggcgccgg tagggctccg ttttagcaat ttaagaaact ttaattgttt 900
tttccacttt tttgtttttc tccgaatttt atgaaattat gatttaagaa aaaaaacgat 960
attgttcatt tattgacctt ctactgcat ggtttcttct attgggttaa ttggctagct 1020
cataagaatt gtttaatttg gttattgtca tcaaatttgc ccacatataa agcttctagc 1080
aat 1084

```

<210> 52
 <211> 285
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G1075

<400> 52

```

Met Ala Gly Leu Asp Leu Gly Thr Thr Ser Arg Tyr Val His Asn Val
1          5          10
Asp Gly Gly Gly Gly Gly Gln Phe Thr Thr Asp Asn His His Glu Asp
20        25        30
Asp Gly Gly Ala Gly Gly Asn His His His His His Asn His Asn
35        40        45
His His Gln Gly Leu Asp Leu Ile Ala Ser Asn Asp Asn Ser Gly Leu
50        55        60
Gly Gly Gly Gly Gly Gly Gly Ser Gly Asp Leu Val Met Arg Arg Pro
65        70        75        80
Arg Gly Arg Pro Ala Gly Ser Lys Asn Lys Pro Lys Pro Pro Val Ile
85        90        95

```

Val Thr Arg Glu Ser Ala Asn Thr Leu Arg Ala His Ile Leu Glu Val
100 105 110

Gly Ser Gly Cys Asp Val Phe Glu Cys Ile Ser Thr Tyr Ala Arg Arg
115 120 125

Arg Gln Arg Gly Ile Cys Val Leu Ser Gly Thr Gly Thr Val Thr Asn
130 135 140

Val Ser Ile Arg Gln Pro Thr Ala Ala Gly Ala Val Val Thr Leu Arg
145 150 155 160

Gly Thr Phe Glu Ile Leu Ser Leu Ser Gly Ser Phe Leu Pro Pro Pro
165 170 175

Ala Pro Pro Gly Ala Thr Ser Leu Thr Ile Phe Leu Ala Gly Ala Gln
180 185 190

Gly Gln Val Val Gly Gly Asn Val Val Gly Glu Leu Met Ala Ala Gly
195 200 205

Pro Val Met Val Met Ala Ala Ser Phe Thr Asn Val Ala Tyr Glu Arg
210 215 220

Leu Pro Leu Asp Glu His Glu Glu His Leu Gln Ser Gly Gly Gly Gly
225 230 235 240

Gly Gly Gly Asn Met Tyr Ser Glu Ala Thr Gly Gly Gly Gly Gly Leu
245 250 255

Pro Phe Phe Asn Leu Pro Met Ser Met Pro Gln Ile Gly Val Glu Ser
260 265 270

Trp Gln Gly Asn His Ala Gly Ala Gly Arg Ala Pro Phe
275 280 285

<210> 53
<211> 1342
<212> ADN
<213> Arabidopsis thaliana

<220>
<223> G1076

<400> 53

ES 2 502 943 T3

```

attttagtct tcctataact tcttctcaat cctctctcat atcttttttc ttagtttaaa 60
tttcaataaa atagaaaaaa acatatacaa atctacagag aagagaagct ttattttaat 120
cttgtgtgtg tgtgtgtgtt ttatataatt tttatttttt ttcaaattaa aatctcttct 180
ttgcttttga tgtgggcatg gctggctctg atctaggcac agcttttcgt tacgttaatc 240
accagctcca tcgtcccgat ctccaccttc accacaattc ctctccgat gacgtcactc 300
ccggagccgg gatgggtcat ttcaccgtcg acgacgaaga caacaacaac aaccatcaag 360

gtcttgactt agcctctggt ggaggatcag gaagctctgg aggaggagga ggtcacggcg 420
ggggaggaga cgtcgttggc cgctgtccac gtggcagacc accgggatcc aagaacaaac 480
cgaaacctcc ggtaattatc acgcgcgaga gcgcaaacac tctaagagct cacattcttg 540
aagtaacaaa cggtcgcgat gttttcgact gcgttgcgac ttatgctcgt cggagacagc 600
gagggatctg cgttctgagc ggtagcggaa cggtcacgaa cgtcagcata cgtcagccat 660
ctgcggcttg agcggttgtg acgctacaag gaacgttcga gattctttct ctctccgat 720
cgtttcttcc tcctccggca cctcccggag caacgagttt gacaattttc ttagccggag 780
gacaaggcca ggtggttga ggaagcgttg tgggtgagct tacggcggct ggaccggtga 840
ttgtgattgc agcttcgttt actaatgttg cttatgagag acttccttta gaagaagatg 900
agcagcagca acagcttgga ggaggatcta acggcggagg taatttggtt ccggaggtgg 960
cagctggagg aggaggagga cttccgttct ttaatttacc gatgaatatg caaccaaatg 1020
tgcaacttcc ggtggaaggt tggccgggga attccggtgg aagaggtcct ttctgatgtg 1080
tatatatga taatcattat atatataccg gcggagaagc ttttccggcg aagaatttgc 1140
gagagtgaag aaaggttaga aaagctttta atggactaat gaatttcaa ttatcatcgt 1200
gatttcggac attgtcttgt tcatcatgtt aagcttaggt ttattttttg tcgtttgtag 1260
aattttatgt ttgaatcctt ttttttttct gtgaaactct attgtgttcg tctgcgaagg 1320
aaaaaaaaat tctcaaaaaa aa 1342

```

<210> 54
 <211> 292
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G1076

<400> 54

ES 2 502 943 T3

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

ES 2 502 943 T3

[illegible]

<210> 55
<211> 983
<212> ADN
<213> Arabidopsis thaliana

<220>
<223> G280

<400> 55

ES 2 502 943 T3

aagttaatat gagaataatg agaaaaccac tttcccaa at tgctttttaa aatccctcct 60
cacacagatt ccttccttca tcacctcaca cactctctac gcttgacatg gccttcgac 120

tccaccatgg ctacgcttca gatacgcatt catcagaact tccgctggtt tctctccac 180
cttatcctca gatgataatg gaagcgattg agtccttgaa cgataagaac ggctgcaaca 240
aaacgacgat tgctaagcac atcgagtcga ctcaacaaac tctaccgccg tcacacatga 300
cgctgctcag ctaccatctc aaccagatga agaaaaccgg tcagctaatac atggtgaaga 360
acaattatat gaaaccagat ccagatgctc ctcttaagcg tggctggtggc cgtcctccga 420
agcagaagac tcaggccgaa tctgacgccg ctgctgctgc tgttggtgct gccaccgctg 480
tctctacaga tccgcctaga tctcgtggcc gtccaccgaa gccgaaagat ccacgagac 540
ctcccagga gaaggtcatt accggatctg gaaggccacg aggacgacca ccgaagagac 600
cgagaacaga ttcggagacg gttgctgcgc cggaaccggc agctcaggcg acaggtgagc 660
gtaggggacg tgggagacct ccgaaggtga agccgacggt ggttgctccg gttgggtgct 720
gaattaatcg gtacttatgc aatttcggaa tctttagtta ctgaaaaatg gaatctctta 780
gagagtaaga gagtgttcta atttagctta attagattta tttggatttc tttcagtatt 840
tggattgtaa actttagaat ttgtgtgtgt gttgttgctt agtcctgaga taagatataa 900
cattagcgac tgtgtattat tattattact gcattgtggt atgtgaaact ttgttctctt 960
gttgaaaaaa aaaaaaaaaa aaa 983

<210> 56
<211> 204
<212> PRT
<213> Arabidopsis thaliana

<220>
<223> Polipéptido G280

<400> 56

ES 2 502 943 T3

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

<210> 57
<211> 1964
<212> ADN
<213> Arabidopsis thaliana

<220>
<223> G1367

<400> 57

ES 2 502 943 T3

```

tccttccaca aaactttttt aatttttatct gaaaaattaa aacaaccgaa acaaaaaaaaa 60
aaaactaaaa atcaaaaatc tcatcacctt ctttgctctg tatttttttct ctctcactaa 120
atcctccatg gatccttctc tctctgcaac caatgatcct catcatcctc ctctcctca 180
gttcacatct ttccctcctt tcaccaacac caaccccttc gcctctccaa accacccctt 240
cttcaccgga cccaccgccg tcgcgccgcc aaacaacatc catctctatc aagcagctcc 300
tccgcagcag ccacaaacat ctccagttcc tcctcatcca tctattttccc accctcctta 360
ctctgacatg atttgacagg cgattgcagc gttaaacgaa ccagatgggt caagcaagca 420
agctattttc aggtacatag agagaattta cactgggatt cctactgctc atggagcttt 480
gttgacacac catctcaaga ctttgaagac cagtgggatt cttgtcatgg ttaagaaatc 540
ttacaagctt gcttctactc ctctcctcc tcctcctact agtgtagctc ctagtcttga 600
acctcccaga tctgatttca tagtcaacga gaaccaacct ttacctgac cggttttggc 660
ttcttctact cctcagacta ttaaacgtgg tcgtggctga cctccaaaag ctaaaccaga 720
tgttgttcaa cctcaacctc tgactaatgg aaaactcacc tgggaacaga gtgaattacc 780
tgtctctcga ccagaggaga tacagatata gccgccacag ttaccgttac agccacagca 840
gccggttaag agaccgccgg gtcgtcctag aaaagatgga acttcgccga cgggtaagcc 900
agctgcttct gtttccggtg gtgtggagac tgtgaaacga agaggtagac ctccgagtgg 960
aagagctgct gggagggaga gaaagcctat agtagtctca gctccagctt cagtgttccc 1020

gtatgttgct aatggtggtg ttagacgccg agggagacca aagagagttg acgctgggtg 1080
tgcttctctt gttgctccac caccaccacc accaactaac gtagagagtg gaggagagga 1140
ggttgcagtc aagaaacgag gaagaggacg gcctcctaag attggaggtg ttatcaggaa 1200
gcctatgaag ccgatgagaa gctttgctcg tactggaaaa cccgtaggaa gaccagaaa 1260
gaatgcggtg tcagtgggag cttctggacg acaagatggt gactatggag aactgaagaa 1320
gaagtttgag ttgtttcaag cgagagctaa ggatattgta attgtgttga aatccgagat 1380
aggaggaagt ggaaatcaag cagtggttca agccatacag gacctggaag ggatagcaga 1440
gacaacaaac gagccaaagc acatggaaga agtgcagctg ccagacgagg aacaccttga 1500
aaccgaacca gaagcagagg gtcaaggaca gacagaagca gaggcaatgc aagaagctct 1560
gttctaaaga taaagccttg acataaaaag ctagcaagtg gtgggtttac ttgttgtgtg 1620
ttacatgaaa tttttaatct tataagggtg tttgcaggag aaaaacaaaa agaacaatgt 1680
gatgaactga tgatgatgat tgtgtctcta accaaacaac aaggagaggt agggtaatgt 1740
ctgtaaagtg aattaggatg ttaccattgt tcatgcttcc catctctctc catcgccat 1800
atctgtgtag gcagctttgt tctttgttcc ctctgttttt ttttagactg ttgtgtctct 1860
tattctatct tgtctcctta ggcttttttag gagtgtgtgt tgatgtttat caaaaacgct 1920
tatgtaattt ttatgaccac ttctactttt tatgatgggt tctt 1964

```

<210> 58
 <211> 479
 <212> PRT

<213> Arabidopsis thaliana

<220>

<223> Polipéptido G1367

5

<400> 58

Met Asp Pro Ser Leu Ser Ala Thr Asn Asp Pro His His Pro Pro Pro
1 5 10 15

Pro Gln Phe Thr Ser Phe Pro Pro Phe Thr Asn Thr Asn Pro Phe Ala
20 25 30

Ser Pro Asn His Pro Phe Phe Thr Gly Pro Thr Ala Val Ala Pro Pro
35 40 45

Asn Asn Ile His Leu Tyr Gln Ala Ala Pro Pro Gln Gln Pro Gln Thr
50 55 60

Ser Pro Val Pro Pro His Pro Ser Ile Ser His Pro Pro Tyr Ser Asp
65 70 75 80

Met Ile Cys Thr Ala Ile Ala Ala Leu Asn Glu Pro Asp Gly Ser Ser
85 90 95

Lys Gln Ala Ile Ser Arg Tyr Ile Glu Arg Ile Tyr Thr Gly Ile Pro
 100 105 110
 Thr Ala His Gly Ala Leu Leu Thr His His Leu Lys Thr Leu Lys Thr
 115 120 125
 Ser Gly Ile Leu Val Met Val Lys Lys Ser Tyr Lys Leu Ala Ser Thr
 130 135 140
 Pro Pro Pro Pro Pro Pro Thr Ser Val Ala Pro Ser Leu Glu Pro Pro
 145 150 155 160
 Arg Ser Asp Phe Ile Val Asn Glu Asn Gln Pro Leu Pro Asp Pro Val
 165 170 175
 Leu Ala Ser Ser Thr Pro Gln Thr Ile Lys Arg Gly Arg Gly Arg Pro
 180 185 190
 Pro Lys Ala Lys Pro Asp Val Val Gln Pro Gln Pro Leu Thr Asn Gly
 195 200 205
 Lys Leu Thr Trp Glu Gln Ser Glu Leu Pro Val Ser Arg Pro Glu Glu
 210 215 220
 Ile Gln Ile Gln Pro Pro Gln Leu Pro Leu Gln Pro Gln Gln Pro Val
 225 230 235 240
 Lys Arg Pro Pro Gly Arg Pro Arg Lys Asp Gly Thr Ser Pro Thr Val
 245 250 255
 Lys Pro Ala Ala Ser Val Ser Gly Gly Val Glu Thr Val Lys Arg Arg
 260 265 270
 Gly Arg Pro Pro Ser Gly Arg Ala Ala Gly Arg Glu Arg Lys Pro Ile
 275 280 285
 Val Val Ser Ala Pro Ala Ser Val Phe Pro Tyr Val Ala Asn Gly Gly
 290 295 300
 Val Arg Arg Arg Gly Arg Pro Lys Arg Val Asp Ala Gly Gly Ala Ser
 305 310 315 320
 Ser Val Ala Pro Pro Pro Pro Pro Pro Thr Asn Val Glu Ser Gly Gly
 325 330 335
 Glu Glu Val Ala Val Lys Lys Arg Gly Arg Gly Arg Pro Pro Lys Ile
 340 345 350
 Gly Gly Val Ile Arg Lys Pro Met Lys Pro Met Arg Ser Phe Ala Arg
 355 360 365

Thr Gly Lys Pro Val Gly Arg Pro Arg Lys Asn Ala Val Ser Val Gly
 370 375 380

Ala Ser Gly Arg Gln Asp Gly Asp Tyr Gly Glu Leu Lys Lys Lys Phe
 385 390 395 400

Glu Leu Phe Gln Ala Arg Ala Lys Asp Ile Val Ile Val Leu Lys Ser
 405 410 415

Glu Ile Gly Gly Ser Gly Asn Gln Ala Val Val Gln Ala Ile Gln Asp
 420 425 430

Leu Glu Gly Ile Ala Glu Thr Thr Asn Glu Pro Lys His Met Glu Glu
 435 440 445

Val Gln Leu Pro Asp Glu Glu His Leu Glu Thr Glu Pro Glu Ala Glu
 450 455 460

Gly Gln Gly Gln Thr Glu Ala Glu Ala Met Gln Glu Ala Leu Phe
 465 470 475

<210> 59
 <211> 1878
 <212> ADN
 <213> Arabidopsis thaliana

<220>
 <223> G2787

<400> 59

ES 2 502 943 T3

```

tctcagagca aaaaacaaaa aaaaagaaaa aaaaacccta aatctaaatc tcaccttcca 60
cctctgtcctt tttttttttt gttctttttt ttttttttac tgtatcttct cttctctttg 120
ctctgcaaaa atctcacatc catggatcca tctcttggtg atcctcatca tcctcctcag 180
ttcacccctt ttctcattt tcccacctcc aatcatcatc ctttaggacc aaatccgtac 240
aataaccatg tcgtcttcca accgcagccg caaacgcaaa cgcaaattccc gcaaccgcag 300
atgtttcagt tatctccaca tgtttcaatg ccccacctc cttactccga aatgatttgc 360
gctgcgattg cggcgtaaaa cgaaccggat ggttcgagca agatggcaat ttcgagatac 420
atcgagagat gttacaccgg tttaacttct gctcatgctg ctttgttgac tcaccatctc 480
aagactttga agaccagtgg tgttctttct atgggttaaga aatcttaca aattgctggt 540
tcttctactc ctctgctag ttagctgtt gctgctgctg ccgccgctca aggtctcgat 600
gttcccagat ctgagattct ccattcaagt aacaacgatc ccatggcttc tggctctgct 660
tctcagcctc tgaaacgagg tcgtggtcgt cctcctaagc ctaaacctga atctcaacca 720
caaccactac agcaacttcc accgaccaat caagtccagg ctaacggaca gccaatctgg 780

gaacagcagc aagttcaatc acctgttccg gttccgactc cggttacaga gtcggcgaag 840
agaggacctg gtcgtccaag gaagaacggt tctgctgctc ctgctactgc accaatcggt 900

caagcttcgg ttatggctgg aattatgaaa cgtagaggta gaccaccggg tcgtcgagct 960
gctgggagac agaggaagcc caaatccgtt tcttctactg cctctgtgta tccttatgtt 1020
gctaattggtg ctagacgcag aggaaggcct aggagagttg ttgaccctag cagtattgtt 1080
agtgttgctc cagtaggtgg tgaaaatgtg gcagcggttg cgccagggat gaagcgtgga 1140
cgtggacgac cacctaagat tgggtggtgtt atcagtaggc ttattatgaa gcctaagaga 1200
ggacgaggac gtcctgtagg tagaccaga aagattggaa catcagtcac gactgggaca 1260
caagattctg gagaactcaa gaagaagttt gatatttttc aagagaaagt gaaagaaatt 1320
gtgaagggtg tgaaggatgg agttacaagt gagaatcaag cagtggtgca agccataaaa 1380
gatctggaag cactaacagt gacggagacc gttgagccac aagttatgga agaagtgacg 1440
ccagaggaga ctgcagcacc acagactgaa gctcaacaaa ctgaagctgc tgagacacaa 1500
ggaggacaag aagaaggaca agaaagagaa ggagaaacac agaccagac agaagcagag 1560
gcaatgcaag aagctctgtt ctgaagaata ataatgatct agaaaacaac ctagacataa 1620
tagccttggt gtttggcgtt aggagtgtt ttttttagtt gttttagggtg ttggaatcgc 1680
atcttaaatt atataaaaat ctataaggaa ttttaatttt tctaggtttt gttgtctgca 1740
gaagaagaaa tagtagactc gttaatggtg ttgtgtcgg tgtgtcttta accaaaccat 1800
aagacgtggc tgtaaattag cgatgtttct agtcttccat ctttaataat ctcttattgc 1860
gtctgtgcct ttgttttt 1878

```

<210> 60
 <211> 480
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G2787

<400> 60

ES 2 502 943 T3

Met Asp Pro Ser Leu Gly Asp Pro His His Pro Pro Gln Phe Thr Pro
 1 5 10 15
 Phe Pro His Phe Pro Thr Ser Asn His His Pro Leu Gly Pro Asn Pro
 20 25 30
 Tyr Asn Asn His Val Val Phe Gln Pro Gln Pro Gln Thr Gln Thr Gln
 35 40 45
 Ile Pro Gln Pro Gln Met Phe Gln Leu Ser Pro His Val Ser Met Pro
 50 55 60
 His Pro Pro Tyr Ser Glu Met Ile Cys Ala Ala Ile Ala Ala Leu Asn
 65 70 75 80
 Glu Pro Asp Gly Ser Ser Lys Met Ala Ile Ser Arg Tyr Ile Glu Arg
 85 90 95

ES 2 502 943 T3

Cys Tyr Thr Gly₁₀₀ Leu Thr Ser Ala His₁₀₅ Ala Ala Leu Leu Thr₁₁₀ His His
 Leu Lys Thr₁₁₅ Leu Lys Thr Ser Gly₁₂₀ Val Leu Ser Met Val₁₂₅ Lys Lys Ser
 Tyr Lys₁₃₀ Ile Ala Gly Ser Ser₁₃₅ Thr Pro Pro Ala Ser₁₄₀ Val Ala Val Ala
 Ala₁₄₅ Ala Ala Ala Ala Gln₁₅₀ Gly Leu Asp Val Pro₁₅₅ Arg Ser Glu Ile Leu₁₆₀
 His Ser Ser Asn Asn₁₆₅ Asp Pro Met Ala Ser₁₇₀ Gly Ser Ala Ser Gln₁₇₅ Pro
 Leu Lys Arg Gly₁₈₀ Arg Gly Arg Pro Pro₁₈₅ Lys Pro Lys Pro Glu₁₉₀ Ser Gln
 Pro Gln₁₉₅ Pro Leu Gln Gln Leu Pro Pro Thr Asn Gln Val₂₀₅ Gln Ala Asn
 Gly Gln₂₁₀ Pro Ile Trp Glu Gln₂₁₅ Gln Gln Val Gln Ser₂₂₀ Pro Val Pro Val
 Pro₂₂₅ Thr Pro Val Thr Glu₂₃₀ Ser Ala Lys Arg Gly₂₃₅ Pro Gly Arg Pro Arg₂₄₀
 Lys Asn Gly Ser Ala₂₄₅ Ala Pro Ala Thr Ala₂₅₀ Pro Ile Val Gln Ala Ser₂₅₅
 Val Met Ala Gly₂₆₀ Ile Met Lys Arg Arg₂₆₅ Gly Arg Pro Pro Gly₂₇₀ Arg Arg
 Ala Ala Gly₂₇₅ Arg Gln Arg Lys Pro₂₈₀ Lys Ser Val Ser Ser₂₈₅ Thr Ala Ser
 Val Tyr₂₉₀ Pro Tyr Val Ala Asn₂₉₅ Gly Ala Arg Arg Arg₃₀₀ Gly Arg Pro Arg
 Arg₃₀₅ Val Val Asp Pro Ser₃₁₀ Ser Ile Val Ser Val₃₁₅ Ala Pro Val Gly Gly₃₂₀
 Glu Asn Val Ala₃₂₅ Ala Val Ala Pro Gly Met₃₃₀ Lys Arg Gly Arg Gly₃₃₅ Arg
 Pro Pro Lys Ile₃₄₀ Gly Gly Val Ile Ser₃₄₅ Arg Leu Ile Met Lys₃₅₀ Pro Lys
 Arg Gly Arg₃₅₅ Gly Arg Pro Val Gly₃₆₀ Arg Pro Arg Lys Ile₃₆₅ Gly Thr Ser

ES 2 502 943 T3

Val Thr Thr Gly Thr Gln Asp Ser Gly Glu Leu Lys Lys Lys Phe Asp
370 375 380

Ile Phe Gln Glu Lys Val Lys Glu Ile Val Lys Val Leu Lys Asp Gly
385 390 395 400

Val Thr Ser Glu Asn Gln Ala Val Val Gln Ala Ile Lys Asp Leu Glu
405 410 415

Ala Leu Thr Val Thr Glu Thr Val Glu Pro Gln Val Met Glu Glu Val
420 425 430

Gln Pro Glu Glu Thr Ala Ala Pro Gln Thr Glu Ala Gln Gln Thr Glu
435 440 445

Ala Ala Glu Thr Gln Gly Gly Gln Glu Glu Gly Gln Glu Arg Glu Gly
450 455 460

Glu Thr Gln Thr Gln Thr Glu Ala Glu Ala Met Gln Glu Ala Leu Phe
465 470 475 480

<210 > 61
<211 > 1772
<212> ADN
<213> Arabidopsis thaliana

<220>
<223> G3045

<400> 61

ES 2 502 943 T3

ttacattcca tccgctaact, tctggacctc gtcaattgct gttactgctt gtttcaactt 60
agctgcagct tccttcactt tcttttgctg caacattttt catttcagta acttatcatc 120
agattctttc tttttagttg aaatgaatcc attttaatta actaatcaaa tgaccattca 180
attccatctt ctaggctata cgacataatc taacaattct gttgacttgc tagtatcctt 240
tgtgctccac acaacataat gtctaaatca aattgatgca gggatacagt aatggtttacg 300
aaaaaccatt atagaagcta agtggggata gttcacttac taagagtgcg gttcttttct 360
tgagatctcc gacgtttgca gctaccggtg ccactgaagt gctcttttgt tcatgcatag 420
aaaacccgac tttgacataa gtatacacac agttgtataa gcatggttat gtcttacatg 480
aactcttgta gatattgact caaatgaaat gataatgact aaccaaatag atttcaagaa 540
atacaccaaa tccagatact atacacatct tttcaaaata ttacgaatca tttcaaattc 600
tgcagaacct aaaattaacc agatttgaga ccaccagaga caaataacat acaactctaa 660
actttttcca ctatatatgc agaacaaaca gtcaagaaca accgtataat tgggtatatac 720
cttttgtaa aattatacat taagcattgt tatgtctaac atgaactaaa cacttgtgaa 780
atttatttgg actcaaatta catgataact tcttaccaaa tagaccaatc actttcactt 840
ccacattata caaaaaaaga tttaatgaaa tacacaaaaa tccagataag atgcacatct 900

tttcaaagaa attacgaata atatcagata cttcacactc acaatagacc acatttgaga 960
caataaaga cattactctg aactttatct actatatgca gaagaaacag tcaagaagaa 1020
caatattaaa taagacattt tcccaaaata caccaaaatc cagataagat acacattttt 1080
ctaaaaatac ggggaatttc agatactgca atcctaaaag tagaccacat ttgagaccag 1140
agtcaaataa gacattaccc tgaattatct ccacactata cagaacaaac agtcaagaac 1200
aatcatataa ttggtatcag accattttcta aattttctttt gacattttgt gaataaagat 1260
aatgaaatta aagagaaaca taccttccta gtcctgcgca caggctgtgc agctacttcg 1320
acagtagggt tccttccacg tctcttagct ggagccacca cagtttctgc tggaacagtt 1380
gcagccgcca cgtcatcttt ctttggcctc cctcgttttc tagaaccctc tccagtagca 1440
gtagtaacca cagccgccgt tacagtcgac ctctttgccc taccacgttt cttcacagcc 1500
gccgccacag acgtagaagg aacagcctga gctacgtttc ttttcggacg accacttgggt 1560
ttagtactcg ccttcgcaga aacggcaccg atttgggaag agtcagattt agcctttggc 1620
ggtcgaccac gaccgcgttt ctgagaagca gtatcagtag ctgagacgcc ggtggcatcc 1680
gtttgagggt tgttgccaga cgcactctca ggtgtaccgg atcttggaac ttcggaacca 1740
gacgcgggag gagtaagagc tgttttcgcc at 1772

<210> 62
<211> 189
<212> PRT
<213> Arabidopsis thaliana

<220>
<223> Polipéptido G3045

<400> 62

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

<210> 63
<211> 10
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> misc_feature
<222> (2) .. (6)
<223> xaa puede ser cualquier aminoácido de origen natural

<220>
<221> misc_feature
<222> (8) .. (9)
<223> xaa puede ser cualquier aminoácido de origen natural

<220>
<223> Secuencia consenso del segundo dominio conservado

<400> 63

Pro Xaa Xaa Xaa Xaa Xaa Xaa Leu Xaa Xaa Tyr
1 5 10

5 <210> 64
<211> 10
<212> PRT
<213> Arabidopsis thaliana

10 <220>
<221> misc_feature
<222> (2) .. (6)
<223> xaa puede ser cualquier aminoácido de origen natural

15 <220>
<221> misc_feature
<222> (8) .. (9)
<223> xaa puede ser cualquier aminoácido de origen natural

20 <220>
<223> Secuencia consenso del segundo dominio conservado

<400> 64

Pro Xaa Xaa Xaa Xaa Xaa Xaa Leu Xaa Xaa Phe
1 5 10

25
30 <210> 65
<211> 1157
<212> ADN
<213> Secuencia artificial

<220>
<223> Secuencia artificial

35 <220>
<223> P21269

<400> 65

ES 2 502 943 T3

ggtgaggagc gagttgtaga ggaggagggg caccgatcaa ccgggaagct ggcggccggg 60
 aagtgggtggg agcggccaga tgacacggcc atggccggga tggaccctgg cgggggaggc 120
 gccggcgccg gcagctcacg gtacttccac catctgctcc gaccgcagca gccgtcgccg 180
 ctgtcaccgc tgtcgccgac atcccatgtc aagatggagc actccaagat gtcacccgac 240
 aagagccccg tgggcgaggg agatcacgcg ggaggggagtg gaagcggcgg cgtcggcggg 300
 gaccaccagc cgtcgtcgtc ggccatgggtg cccgtcgagg gtggcagcgg cagcggcggc 360
 ggtagtggct cgggtggggc gacgcggcgc ccgcgcgggc gccgccccg gtccaagaac 420
 aagccgaagc cgcccatcat cgtgacgcgc gacagccga acgcgctgca ctcgcacgtg 480
 ctcgaggtcg ccggcgggcg cgacgtcgtc gactgcgtgg ccgagtacgc ccgccgccga 540
 gggcgcgggc tgtgcgtgct gagcggcggc ggcgccgctc tcaacgtggc gctgcggcag 600
 ccgggcgcgt cgccgccccg cagcatgggtg gccacgctgc ggggcccgtt cgagatccta 660
 tctctcacgg gcacggctct gcagcctccc gcgccacccg gcgcgagcgg cctcaccgtg 720
 ttctctctcg gcggccaggg ccaggatgat ggcggcagcg tgggtggccc gctggtcgcc 780
 gcggggcccc tcgtcctgat ggcggcctca ttgcggaacg ccgtgtacga gcggtcgccg 840
 ctggagggcg aggaagagga ggtcgccgcg cccgcccgcg gaggcgaagc acaagatcaa 900
 gtggcacaaat cagctggacc cccagggcag caaccggcgg cgtcacagtc ctccggcgtg 960
 acaggaggcg acggcaccgg cggcgccggt ggcattgtgc tctacaacct cgccgggaat 1020
 gtgggaggct atcagctccc cggagacaac ttcgagggtt ggagcggcgc cggcgccggc 1080
 ggagtcaggc caccgttctg acccatgtct tagcatccag ttcaaaaatt ctccaaatta 1140
 agaattgcgc agtgcag 1157

<210> 66
 <211> 1040
 <212> ADN
 <213> Arabidopsis thaliana

<220>
 <223> G2789

<400> 66

ctttagggac accaaatcta ttcaacctaa aagccttctt ttcccctata ttgaccaact 60

ES 2 502 943 T3

```

ttttagcgaa tcagaagagg aatggatgag gtatctcgtt ctcatacacc gcaatttcta 120
tcaagtgatc atcagcacta tcaccatcaa aacgctggac gacaaaaacg cggcagagaa 180
gaagaaggag ttgaaccaa caatataggg gaagacctag ccacctttcc ttccggagaa 240
gagaatatca agaagagaag gccacgtggc agacctgctg gttccaagaa caaacccaaa 300
gcaccaatca tagtcactcg cgactccgcg aacgccttca gatgtcacgt catggagata 360
accaacgcct gcgatgtaat ggaaagccta gccgtcttcg ctagacgccg tcagcgtggc 420
gtttgcgtct tgaccggaaa cggggccggtt acaaacgtca ccgttagaca acctggcgga 480
ggcgtcgtca gttttacacg acggtttgag attctttctc tctcgggttc gtttcttctt 540
ccaccggcac caccagctgc gtctgggtta aagggtttact tagccggtgg tcaagggtcaa 600
gtgatcggag gcagtgtggt gggaccgctt acggcatcaa gtccggtggt cgttatggca 660
gcttcatttg gaaacgcac ttacgagagg ctgccactag aggaggagga ggaaactgaa 720
agagaaatag atggaaacgc ggctagggcg attggaacgc aaacgcagaa acagttaatg 780
caagatgcga catcgtttat tgggtcgccg tcgaatttaa ttaactctgt ttcgttgcca 840
ggatgaagctt attggggaac gcaacgaccg tctttctaag ataatatcat tgataatata 900
agtttcgtct tcttattctt tttcactttt tacctttttc actttcttag gttttgtttt 960
aacgtttgat taatacctga aggtttttgg aaaattttcg atcggataaa aggatttatg 1020
ttgcgagccg aaacgcggcc 1040

```

<210> 67
 <211> 265
 <212> PRT
 <213> Arabidopsis thaliana

<220>
 <223> Polipéptido G2789

<400> 67

ES 2 502 943 T3

Met Asp Glu Val Ser Arg Ser His Thr Pro Gln Phe Leu Ser Ser Asp
 1 5 10 15
 His Gln His Tyr His His Gln Asn Ala Gly Arg Gln Lys Arg Gly Arg
 20 25 30
 Glu Glu Glu Gly Val Glu Pro Asn Asn Ile Gly Glu Asp Leu Ala Thr
 35 40 45
 Phe Pro Ser Gly Glu Glu Asn Ile Lys Lys Arg Arg Pro Arg Gly Arg
 50 55 60
 Pro Ala Gly Ser Lys Asn Lys Pro Lys Ala Pro Ile Ile Val Thr Arg
 65 70 75 80
 Asp Ser Ala Asn Ala Phe Arg Cys His Val Met Glu Ile Thr Asn Ala
 85 90 95
 Cys Asp Val Met Glu Ser Leu Ala Val Phe Ala Arg Arg Arg Gln Arg
 100 105 110
 Gly Val Cys Val Leu Thr Gly Asn Gly Ala Val Thr Asn Val Thr Val
 115 120 125
 Arg Gln Pro Gly Gly Gly Val Val Ser Leu His Gly Arg Phe Glu Ile
 130 135 140
 Leu Ser Leu Ser Gly Ser Phe Leu Pro Pro Pro Ala Pro Pro Ala Ala
 145 150 155 160
 Ser Gly Leu Lys Val Tyr Leu Ala Gly Gly Gln Gly Gln Val Ile Gly
 165 170 175
 Gly Ser Val Val Gly Pro Leu Thr Ala Ser Ser Pro Val Val Val Met
 180 185 190
 Ala Ala Ser Phe Gly Asn Ala Ser Tyr Glu Arg Leu Pro Leu Glu Glu
 195 200 205
 Glu Glu Glu Thr Glu Arg Glu Ile Asp Gly Asn Ala Ala Arg Ala Ile
 210 215 220
 Gly Thr Gln Thr Gln Lys Gln Leu Met Gln Asp Ala Thr Ser Phe Ile
 225 230 235 240
 Gly Ser Pro Ser Asn Leu Ile Asn Ser Val Ser Leu Pro Gly Glu Ala
 245 250 255
 Tyr Trp Gly Thr Gln Arg Pro Ser Phe
 260 265

<210> 68
 <211> 938
 <212> ADN
 <213> Arabidopsis thaliana

5

<220>
 <223> G1667

<400> 68

10

```

tcgacccatc cctactcctc acaatcattc ttttggtgtc tctaccattt aataattcac    60
caaaatctcc ttacttaaac tcacaaactc ctcacaaatt ttctgaatct ttcagttgaa    120
catataacaa cattcataac aatggctgga ggtacagctc taactccaac ctctgtagga    180
tccaagtctg ttccaatgag gaacctgaa gcaacagaga gaggcaacac caacaacaac    240
ctgagagcat tacccaaagc cgtccaaccg gtttcatcaa tcgaaggaga gatggctaag    300
aggccacgtg gcagacccgc tggctccaag aacaaaccca aaccaccaat cattgtgact    360
cacgacagtc caaatccct cagagctaac gccgttgaga tcagctcagg ttgtgacatc    420

tgtgagactt tatcggattt tgcaagaagg aaacagagag gtctctgcat tctcagtgcc    480
aatggttggtg tcaccaatgt gacattaagg caaccagctt catcaggagc aattgtcaca    540
ttacacggac gttacgagat cctctcattg cttggatcaa tcttgccctc accagacca    600
cttggaataa ctggtctgac catttactta gccggacctc aaggacaggt tgttggtgga    660
ggagtgggtg gtgggctaatt cgcctctggt cctgttggtc tcatggctgc atctttcatg    720
aatgctgttt ttgatcgtct tcctatggat gatgatgaag ctgcctctat gcagaaccag    780
cagtactacc agaatggaag atcccgtcct ttagatgaca ttcattggact gcctcaaaat    840
ctgctcacta atggaaactc ggcttctgat atctactctt gggggccttg gaatcaaaga    900
taaagtgtgc tgtaggttga gagagaaccg taagtctg                                938
    
```

<210> 69
 <211> 253
 <212> PRT
 <213> Arabidopsis thaliana

15

<220>
 <223> Polipéptido G1667

20

<400> 69

ES 2 502 943 T3

Met Ala Gly Gly Thr Ala Leu Thr Pro Thr Ser Val Gly Ser Lys Ser
1 5 10 15

Val Pro Met Arg Asn His Glu Ala Thr Glu Arg Gly Asn Thr Asn Asn
20 25 30

Asn Leu Arg Ala Leu Pro Lys Ala Val Gln Pro Val Ser Ser Ile Glu
35 40 45

Gly Glu Met Ala Lys Arg Pro Arg Gly Arg Pro Ala Gly Ser Lys Asn
50 55 60

Lys Pro Lys Pro Pro Ile Ile Val Thr His Asp Ser Pro Asn Ser Leu
65 70 75 80

Arg Ala Asn Ala Val Glu Ile Ser Ser Gly Cys Asp Ile Cys Glu Thr
85 90 95

Leu Ser Asp Phe Ala Arg Arg Lys Gln Arg Gly Leu Cys Ile Leu Ser
100 105 110

Ala Asn Gly Cys Val Thr Asn Val Thr Leu Arg Gln Pro Ala Ser Ser
115 120 125

Gly Ala Ile Val Thr Leu His Gly Arg Tyr Glu Ile Leu Ser Leu Leu
130 135 140

Gly Ser Ile Leu Pro Pro Pro Ala Pro Leu Gly Ile Thr Gly Leu Thr
145 150 155 160

Ile Tyr Leu Ala Gly Pro Gln Gly Gln Val Val Gly Gly Gly Val Val
165 170 175

Gly Gly Leu Ile Ala Ser Gly Pro Val Val Leu Met Ala Ala Ser Phe
180 185 190

Met Asn Ala Val Phe Asp Arg Leu Pro Met Asp Asp Asp Glu Ala Ala
195 200 205

Ser Met Gln Asn Gln Gln Tyr Tyr Gln Asn Gly Arg Ser Arg Pro Leu
210 215 220

Asp Asp Ile His Gly Leu Pro Gln Asn Leu Leu Thr Asn Gly Asn Ser
225 230 235 240

Ala Ser Asp Ile Tyr Ser Trp Gly Pro Trp Asn Gln Arg
245 250

<210> 70
<211> 1496
<212> ADN
<213> Arabidopsis thaliana

<220>
<223> G2157

<400> 70

```

tcttttgatt ttaacctttt ttcagtagca agccaaaaaa aaaaaacaga caaagaagtt    60
ccttttatga taaaggatg atgatagcaa acaaatgata ccccatgtc ttgtgtgtct    120
gcttcacgca acatgttggt ttggatttgg ttaatctaaa agtttaagat aaggttttcg    180
gattctcttc ctgtcttgta atagtttctt gtcggagagc catcaacacc aacttcaaca    240
aaaaaaacaa gaaaaagaaa aagattctct ttctcgtttt atttccatta gagaagaaaa    300
aaagaatggc gaatccttgg tgggtaggga atgttgcgat cgggtggagt gagagtccag    360
tgacgtcatc agctccttct ttgcaccaca gaaacagtaa caacaacaac ccaccgacta    420
tgactcgttc ggatccaaga ttggaccatg acttcaccac caacaacagt ggaagcccta    480
ataccagac tcagagccaa gaagaacaga acagcagaga cgagcaacca gctgttgaaac    540
ccggatccgg atccgggtct acgggtcgtc gtcctagagg tagacctctt ggttccaaga    600
acaaacaaa gagtccagtt gttgttacca aagaaagccc taactctctc cagagccatg    660
ttcttgagat tgctacggga gctgacgtgg cggaaagctt aaacgccttt gctcgtagac    720
gcggccgggg cgtttcgggt ctgagcggta gtggtttggt tactaatgtt actctgcgtc    780
agcctgctgc atccggtgga gttgttagtt tacgtggtca gtttgagatc ttgtctatgt    840

gtggggcttt tcttcctacg tctggctctc ctgctgcagc cgctggttta accatttact    900
tagctggagc tcaagggtcaa gttgtgggag gtggagttgc tggcccgctt attgcctctg    960
gacccgttat tgtgatagct gctacgtttt gcaatgccac ttatgagagg ttaccgattg   1020

aggaagaaca acagcaagag cagccgcttc aactagaaga tgggaagaag cagaaagaag   1080
agaatgatga taacgagagt ggggaataacg gaaacgaagg atcgatgcag ccgccgatgt   1140
ataatatgcc tcctaatttt atcccaaagt gtcacaaat ggctcaacac gacgtgtatt   1200
gggggtggtcc tccgcctcgt gctcctcctt cgtattgatt agttagatag gcggtggttg   1260
gtgcgttctt ttacttgaa tgattatatt ttccattagg atggttaggc ttttgtttat   1320
taaagctatc aagtttcttt tttttttacg gataattcgg atgacaatta gctagtgttt   1380
gtttgtttgt tttgtggcgg cttttctgac ttgactatct tgatcgcgga tagctttgta   1440
tgaaagtgaa ttgattgtag aatcgtcttt tgaattttga tgttggaata aaccaa      1496

```

5

<210>71
 <211> 310
 <212> PRT
 <213> Arabidopsis thaliana

10

<220>
 <223> Polipéptido G2157

15

<400> 71

ES 2 502 943 T3

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

Ala Ala Ser Gly Gly Val Val Ser Leu Arg Gly Gln Phe Glu Ile Le
165 170 175
Ser Met Cys Gly Ala Phe Leu Pro Thr Ser Gly Ser Pro Ala Ala Al
180 185 190
Ala Gly Leu Thr Ile Tyr Leu Ala Gly Ala Gln Gly Gln Val Val Gl
195 200 205
Gly Gly Val Ala Gly Pro Leu Ile Ala Ser Gly Pro Val Ile Val Il
210 215 220
Ala Ala Thr Phe Cys Asn Ala Thr Tyr Glu Arg Leu Pro Ile Glu Gl
225 230 235 24
Glu Gln Gln Gln Glu Gln Pro Leu Gln Leu Glu Asp Gly Lys Lys Gl
245 250 255
Lys Glu Glu Asn Asp Asp Asn Glu Ser Gly Asn Asn Gly Asn Glu Gl
260 265 270
Ser Met Gln Pro Pro Met Tyr Asn Met Pro Pro Asn Phe Ile Pro As
275 280 285
Gly His Gln Met Ala Gln His Asp Val Tyr Trp Gly Gly Pro Pro Pr
290 295 300
Arg Ala Pro Pro Ser Tyr
305 310

<210> 72
<211> 8
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> misc_feature
<222> (7)..(7)
<223> xaa puede ser cualquier aminoácido de origen natural

<220>
<223> Secuencia consenso del dominio de gancho AT

<400> 72

Arg Pro Arg Gly Arg Pro Xaa Gly
1 5

REIVINDICACIONES

1. Una planta transgénica que tiene mayor tolerancia a un estrés abiótico que una planta de control, comprendiendo dicha planta transgénica un polinucleótido recombinante que codifica un polipéptido;
 5 en donde el polipéptido comprende un primer dominio que comprende SEC ID N°: 72, y un segundo dominio que tiene al menos un 62 % de identidad de secuencia de aminoácidos con el segundo dominio conservado de SEC ID N°: 2 como se expone en la Tabla 1 y el segundo dominio comprende SEC ID N°: 63 o SEC ID N°: 64;
 en donde la expresión del polipéptido en la planta transgénica está regulada por un promotor específico de raíz, un promotor específico de tejido de hoja, un promotor específico epidérmico o un promotor inducible por estrés, y la
 10 expresión del polipéptido en la planta transgénica confiere a la planta transgénica dicha mayor tolerancia al estrés abiótico;
 en donde dicha planta de control no comprende el polinucleótido recombinante; y dicha planta transgénica tiene menos anomalías del desarrollo indeseables o menos efectos morfológicos adversos que una segunda planta de control que expresa el polipéptido bajo el control regulador de un promotor constitutivo.
 15
2. Una planta transgénica de acuerdo con la reivindicación 1, en la que el promotor específico de raíz es ARSK1, el promotor específico de hoja es el promotor RBCS3, el promotor específico epidérmico es CUT1 y el promotor inducible por estrés es RD29A.
- 20 3. Una planta transgénica de acuerdo con las reivindicaciones 1 o 2, en la que el segundo dominio es al menos un 71 % idéntico al segundo dominio conservado de SEC ID N°: 2 como se expone en la Tabla 1.
4. Una planta transgénica de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en la que el polipéptido tiene al menos un 80 % de identidad de secuencia con SEC ID N°: 2.
 25
5. Una planta transgénica de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en la que la resistencia al estrés abiótico es al frío o al estrés osmótico tal como estrés por calor, desecación, sequía, congelación y alta salinidad.
- 30 6. Una planta transgénica de acuerdo con una cualquiera de las reivindicaciones 1 a 5, que es una planta de cultivo tal como soja, tomate, alfalfa, trébol, zanahoria, apio, chirivía, repollo, rábano, colza, brócoli, melón, pepino, patata, tabaco, pimienta, mijo, arroz, maíz, trigo, sorgo, centeno o cebada.
7. Un método para producir una planta transformada con mayor tolerancia a un estrés abiótico que una planta de control, comprendiendo el método:
 35
 - (a) proporcionar un vector de expresión que comprende una secuencia polinucleotídica que codifica un polipéptido en donde el polipéptido comprende un primer dominio que comprende SEC ID N°: 72, y un segundo dominio que tiene al menos un 72 % de identidad de secuencia de aminoácidos con el segundo dominio conservado de SEC ID N°: 2 como se expone en la Tabla 1 y el segundo dominio conservado comprende SEC ID N°: 63 o SEC ID N°: 64;
 40 en donde la expresión del polipéptido en la planta transgénica está regulada por un promotor específico de raíz, un promotor específico de tejido de hoja, un promotor específico de tejido epidérmico y un promotor inducible por estrés, y la expresión del polipéptido en la planta transgénica confiere a la planta transgénica dicha mayor tolerancia al estrés abiótico;
 - (b) transformar una planta diana con el vector de expresión para producir la planta transformada; y
 - (c) cultivar la planta transformada que comprende el vector de expresión,
 45
 en donde dicha planta de control no comprende el polinucleótido recombinante; y dicha planta transgénica tiene menos anomalías del desarrollo indeseables o menos efectos morfológicos adversos que una segunda planta de control que expresa el polipéptido bajo el control regulador de un promotor constitutivo.
 50
8. Un método de acuerdo con la reivindicación 7, en la que el promotor específico de raíz es ARSK1, el promotor específico de tejido de hoja es el promotor RBCS3, el promotor específico epidérmico es CUT1 y el promotor inducible por estrés es RD29A.
 55
9. Un método de acuerdo con las reivindicaciones 7 u 8, en el que dicho polipéptido es como se ha definido en las reivindicaciones 3 o 4.
- 60 10. Un método de acuerdo con las reivindicaciones 7, 8 o 9 en el que dicha planta es una planta de soja, tomate, alfalfa, trébol, zanahoria, apio, chirivía, repollo, rábano, colza, brócoli, melón, pepino, patata, tabaco, pimienta, mijo, arroz, maíz, trigo, sorgo, centeno o cebada.
11. Una parte de planta de la planta transgénica de una cualquiera de las reivindicaciones 1 a 6, o producida por el método de una cualquiera de las reivindicaciones 7 a 10, comprendiendo dicha parte de planta fruto, hoja, raíz o tejido vegetal, incluyendo tejido de suelo.
 65

12. Una semilla de una planta transgénica de una cualquiera de las reivindicaciones 1 a 6, o producida por el método de una cualquiera de las reivindicaciones 7 a 10, comprendiendo dicha semilla el polinucleótido recombinante como se ha definido en la reivindicación 1.

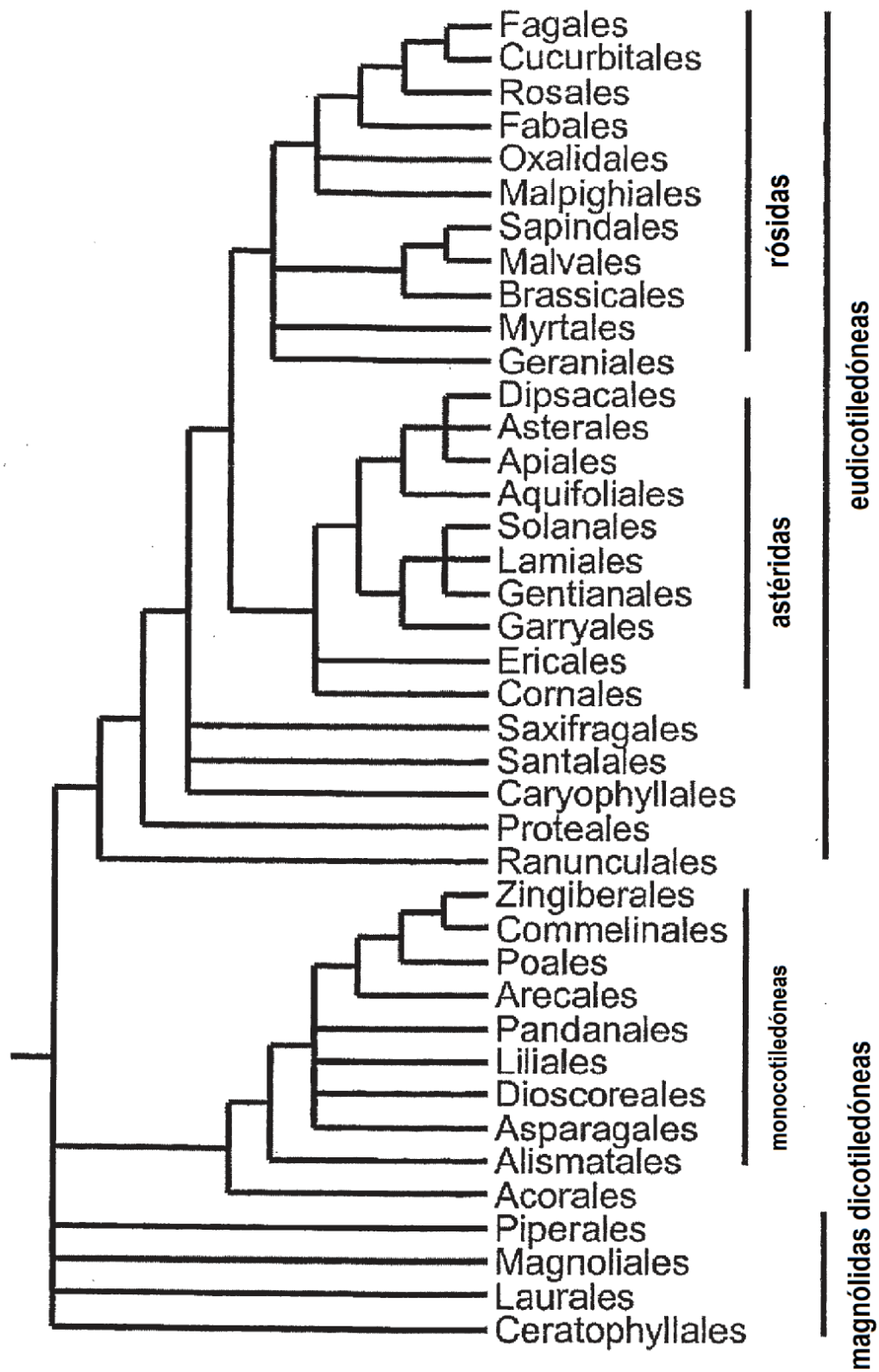


FIGURA 1

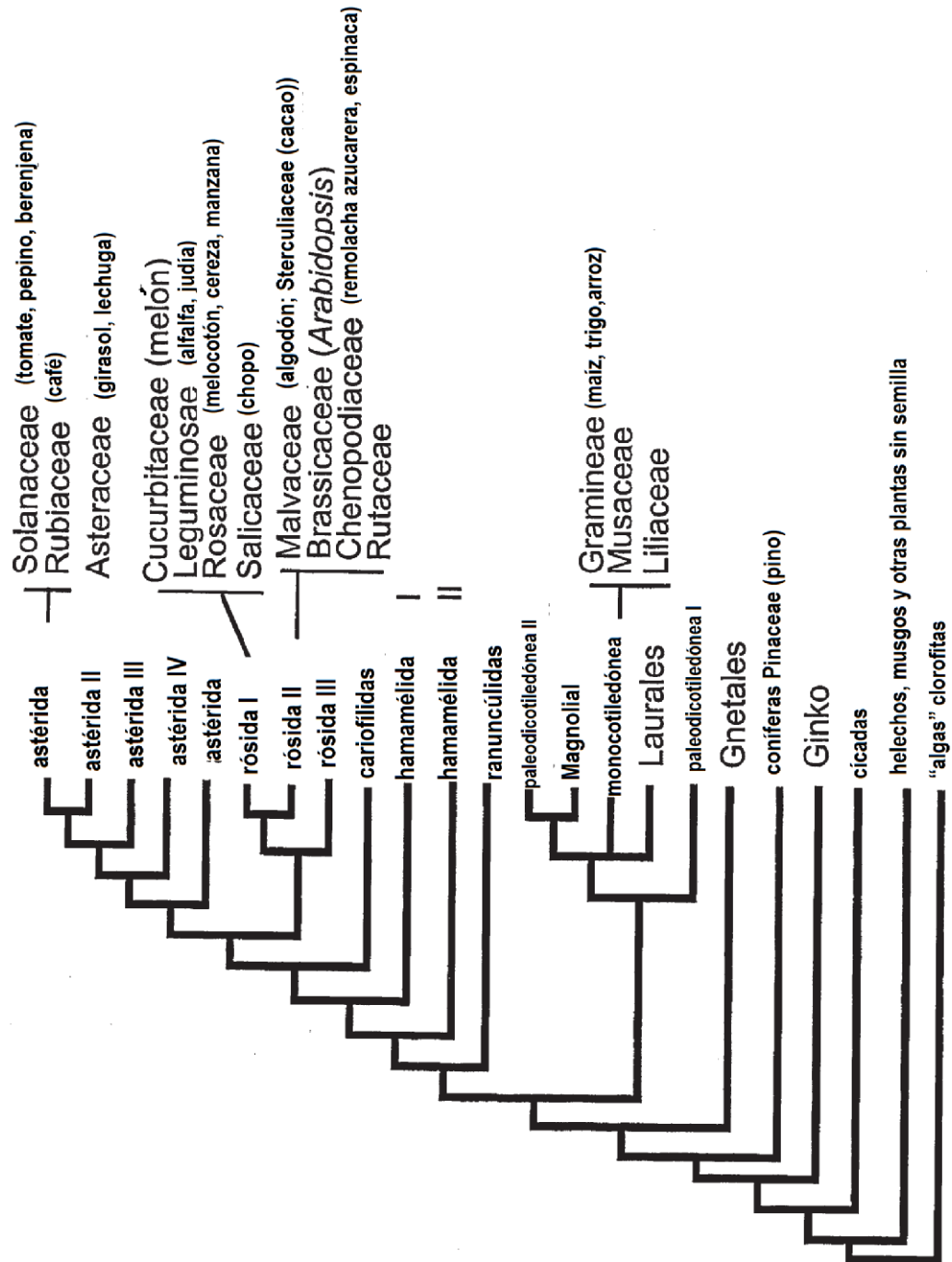


FIGURA 2

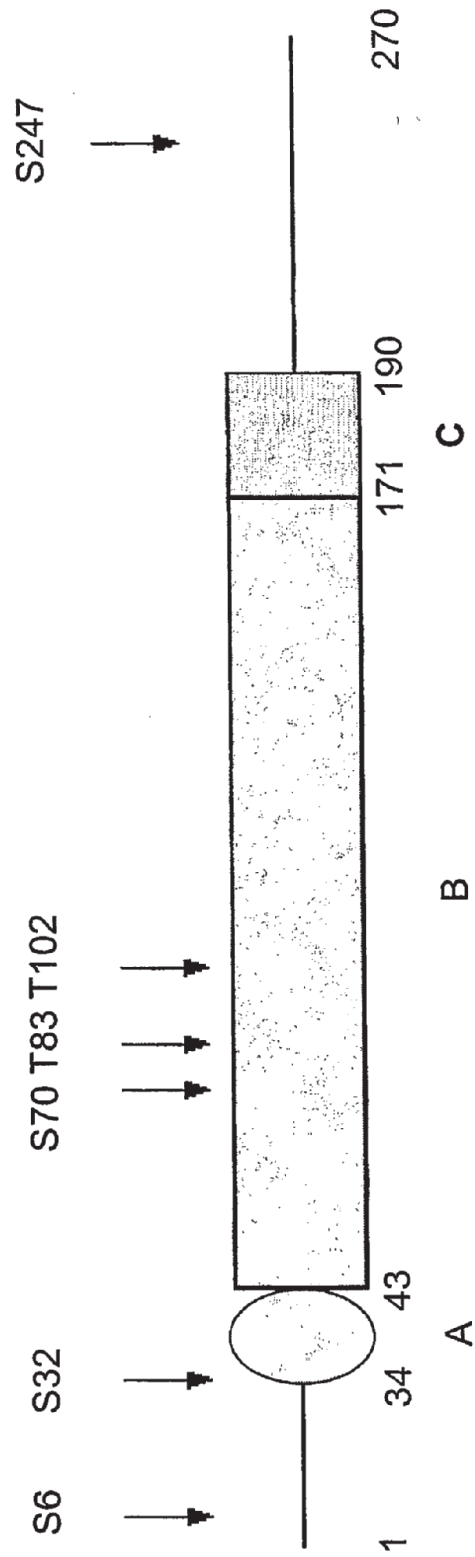


FIGURA 3

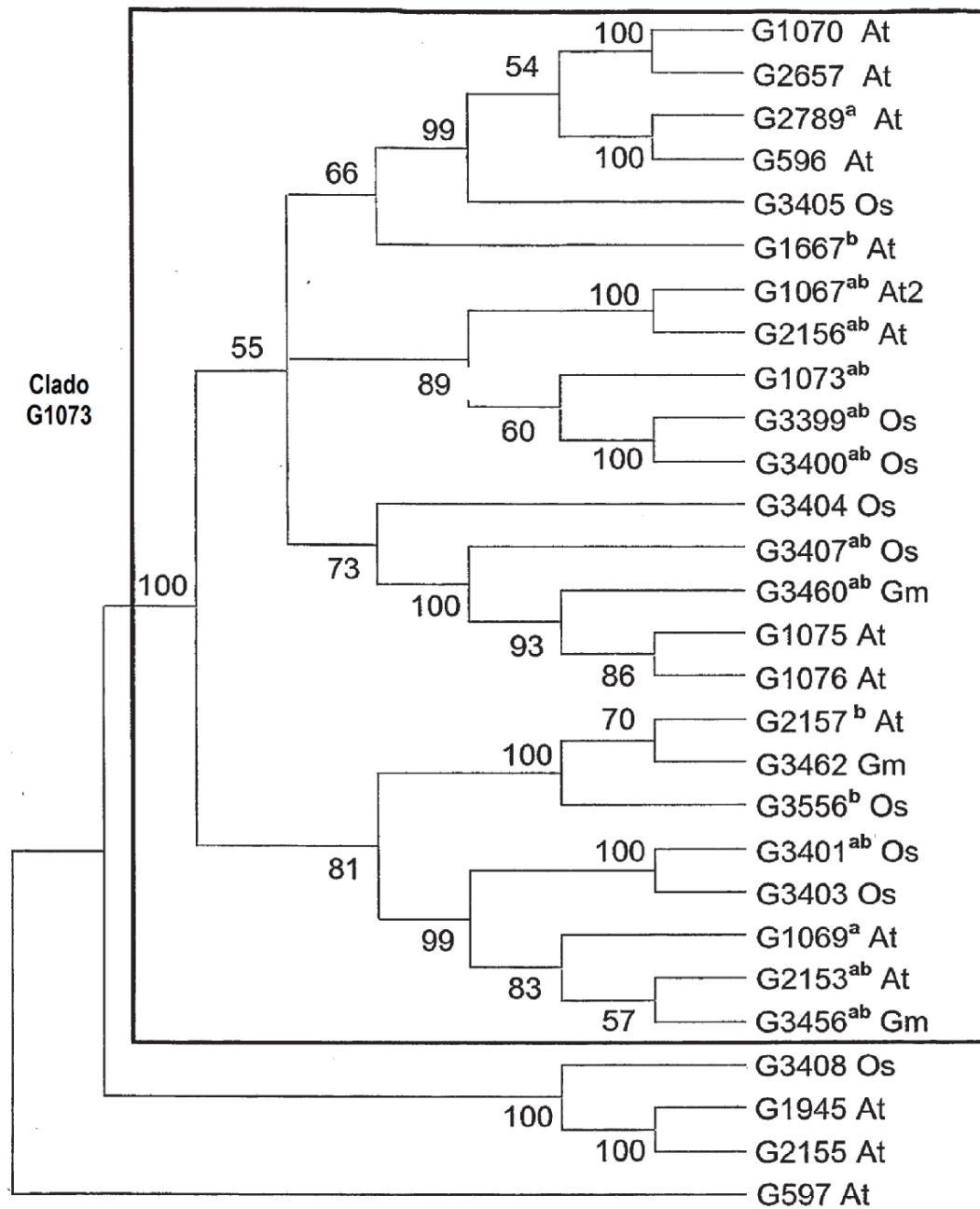


FIGURA 4

G2789	---MDEVs---	RSHTPQFLSSD-HQ---	HYHHQNA----	GR
G596	---MDQVS---	RSLPFPFLSRDLHL---	HPHHQFQ----	HQ
G1070	---MDPVQSHGSQ---	SSLPPFPFHARDFQL---	HLQQQQQEFFLHHHQ	
G3405	---MDPVTASIHG---	HHLPPFPFTRDFHH---	HLQQQHQ-----	QL
G3404	---MDPVTAAAHGGHHHHF-	GAPPVAAAFHHH---	PFHHGGGAHYPAAFQ	
G1075	---MAGLDLGTTS---	RYVHNV-DGGGGG---	QFTTDNHHEDDG-GA	
G1076	---MAGLDLGTAF---	RYVNHQ-LHRPDL---	HLHNSSSDDVTPGA	
G3460	---MAGLDLGSAS---	RFVQN--LHLPDL---	HLQONYQQPR---HK	
G3459	---MAGLDLGSAS---	RFVQN--LHRPDL---	HLQQNEFQQHQDQQHQ	
G3458	---MAGIDLGSAS---	HFVHHR-LERPDL---	EDDENQQDQDNNLN	
G3406	---MAGLDLGTAAAT---	RYVHQLHHLHPDL---	QLQHSYAKQHEPSDD	
G3407	---MAGLDLGTSYL---	HHHQSLHLRHDDG---	GAGSDDGGHDDLSPG	
G1069	---MANPWWTNQSLAGMVDH-	SVSSGH-----	HQN-HHHQSL---LT	
G2153	---MANPWWTGQVNLGLETTP-	PGSSQ-----	LKKPDLHISMNMAMD	
G3456	---	---	MKKPDLGFSMN---E	
G3401	---	MASKE-PSGD-----	---	
G3403	---	MGLPEQPSGSS-----	GPKAELPVAKE--PE	
G2157	---MANPWWVGNVAIGGVES-	PVTSSAPSLHHRNS---	NNNNPPTMTSRDPRL	
G3462	---MANRWWAGNVGMIREQEL-	MENS-----NN---	NNNNNATTTPTTRS	
G3556	---	MGSIDGHSLOQHQG---	---	
G1067	---MEGGEQGGGASRYFHNLF-	FRPEIHHQQLQP---	QGINLIDQHHHHQHQ	
G2156	---MDGGYDQSGGASRYFHNLF-	FRPELHH-QLQP---	QPQLHPLPQ---PQP	
G3400	MAGMDPTGGGGGGVAAHYLH-	MLRAQQHQPLS---	PAGDVKAERSMLSP	
G3399	MAGMDPGGGGAGAGSSRYFH-	HLRLRPQQPSPLSPL---	SPTSHVKMEHSMKSP	
G1073	---	MELNRSE-----	ADEAKAET-----	
G1945	---MKGEYREQKSNEMF---	SKLPHH-----	QQQQQQQQQSHSLTS	
G2155	---	ML-----	SKLPTQ-----R	
G3408	-MSFCERDMNKESMYQERDD-	MAGIRFATPPLPQQ---	QQQQQLVECFSD	DEVD

FIGURA 5A

G2789	QKRGR--EEEG--VE--PN--NIGEDLA--
G596	QQQQQ--QNHGHDIDQHRIGGLKRDADID--PNEHSSAGKDQS--
G1070	QQRNQTGDQQGGGGRQIKMDREETSDNIDN-IANNSGSECKDID--
G3405	HLKTED--DQGGGTPGVFGSRGTRKRDHDDDSNGHSGGDDLA--
G3404	QFQEEQ--QQLVAAAAAGMAKQELVDESNTINSGGSGSGGEEQ--
G1075	GGNHH--HHHHNNHHQGLDLIASN--
G1076	GMGHFT--VDEEDNNNHQGLDLASGG--
G3460	RDSEE--QETPPNPGTALAPFDND--
G3459	RDLEE--QKTPPNHRMG-APFDDD--
G3458	HEGLD--LVP--
G3406	DPNGS--GGGNSNGGPGYGDHDGGS--
G3407	SGGG--
G1069	KGDLG--IAMNQSDNDQDEED--
G2153	SGHNN--HHHQEVDDNNNDDDR--
G3456	STVTG--NHIGEE-DEDRENSD--
G3401	--HDEMNGTSAGG--
G3403	ASPTG--GAAADHADENNESGG--
G2157	DHDFT--TNNSGSPNTQTQSQEEQ--
G3462	NSNTN--ANTNTTTEEEVSRDNG--
G3556	-YSHG--GGAGGSNEEEEAASPPG--
G1067	QHQQQQQPSDDS--RESHNSNKD--
G2156	QPQPQQNSDD--ESD-SNKD--
G3400	DESPGADAD--LGSDH-PTSSAM--
G3399	DKSPVGEGDHAGG--SGSGGVGGDHPSSAM--
G1073	--TPTGG--ATSSA--
G1945	HFHLSS--TVTPTV--
G2155	HLHLSP--SSP--
G3408	SRGSGG--EMKDAVGS--

FIGURA 5B

Dominio de gancho AT

G2789	-----TFPSGEE-----NICKRRPRGRPAGSKNKPAPIIVTR--
G596	-----TPSGGEGSGGGGDNHITRRPRGRPAGSKNKPPIIITR--
G1070	-----IHGSGEGSGGGGDHQMTRRPRGRPAGSKNKPPIIITR--
G3405	-----LVPPSGGGPDGAG-SESATRRPRGRPAGSKNKPPIIITR--
G3404	-----RQSGEEHQHQAAPVIRRRPRGRPAGSKNKPPIIITR--
G1075	-----DNSGLGGGG-GGSGDLVMRRPRGRPAGSKNKPPIIITR--
G1076	-----GSGSSGGGGHGGGDVVGRRPRGRPPGSKNKPPIIITR--
G3460	-----DDKSQGLELASGPGDIVGRRPRGRPPGSKNKPPIIITR--
G3459	-----SDDRSPGLELTSGPGDIVGRRPRGRPPGSKNKPPIIITR--
G3458	-----NSGPGDVVGRRRPRGRPPGSKNKPPIIITR--
G3406	-----SSSGPATDGA VGPGDVVARRPRGRPPGSKNKPPIIITR--
G3407	-----GPSSTAGGAGIGGGEVARRPRGRPPGSKNKPPIIITR--
G1069	-----DPREGAVEVN-RRPRGRPPGSKNKPPIIITR--
G2153	-----NLSGDDHEPREGAVEAPT-RRPRGRPAGSKNKPPIIITR--
G3456	-----EPREGAIDVATRRPRGRPPGSKNKPPIIITR--
G3401	-----EPKDGAVVTGRNRRPRGRPPGSKNKPPIIITR--
G3403	-----EPREGAVVAAPNRRPRGRPPGSKNKPPIIITR--
G2157	-----NSRDEQPAVEPGSGSGSTGRRPRGRPPGSKNKPPIIITR--
G3462	-----EDQNQLGSEHGESEPGSGRRPRGRPAGSKNKPPIIITR--
G3556	-----G-----GSATGSAGRRPRGRPPGSKNKPPIIITR--
G1067	-----HHQGRPDSDPNTSSAPGKRPRGRPPGSKNKPPIIITR--
G2156	-----PGSDPVTSGT-GKRPRGRPPGSKNKPPIIITR--
G3400	-----VAAED-----SGGSGSGGPMRRPRGRPLGSKNKPPIIITR--
G3399	-----VPVEGSGSAGSGSGGPTRRPRGRPPGSKNKPPIIITR--
G1073	-----TASGSSG-----RRPRGRPAGSKNKPPIIITR--
G1945	-----DDSSIEVVRPRGRPPGSKNKPPIIITR--
G2155	-----SMEIVGRPRGRPPGSKNKPPIIITR--
G3408	-----GSGQLVVVGGGDGASIEVAKRRRGRPPGSKNKPPIIITR--

FIGURA 5C

Segundo dominio conservado

FIGURA 5D

G2789	----PGGG--VVS L HGRFEILSLSGSFLPPP-APPAASG-----LKVYLAGGQGVIG
G596	----PGGSSVNLHGRFEILSLSGSFLPPP-APPAASG-----LTIYLAGGQGVVG
G1070	----PSPGS-VVSLHGRFEILSLSGSFLPPP-APPTATG-----LSVYLAGGQGVVG
G3405	----PASQGAVALHGRFEILSLSGSFLPPP-APPEATG-----LTVYLAGGQGVVG
G3404	----GAGPGAVNLTGFRDILSLSGSFLPPP-APPSATG-----LTVVSGGQGVVG
G1075	-----AGAVVTLRGTFEILSLSGSFLPPP-APPGATS-----LTIFLAGAQQGVVG
G1076	-----AGAVVTLQGTFEILSLSGSFLPPP-APPGATS-----LTIFLAGGQGVVG
G3460	-----AGAVVRLHGRFEILSLSGSFLPPP-APPGATS-----LTIYLAGGQGVVG
G3459	-----AGAVVTLHGRFEILSLSGSFLPPP-APPGATS-----LTIYLAGGQGVVG
G3458	-----AGAVVTLHGRFEILSLSGSFLPPP-APPGATS-----LTVFLGGGQGVVG
G3406	-----AGAVVSLHGRFEILSLSGSFLPPP-APPGATS-----LTIFLAGGQGVVG
G3407	--P-GPASPAVATLHGRFEILSLAGSFLPPP-APPGATS-----LAAFLAGGQGVVG
G1069	-AA-APGG--VVS L QGRFEILSLTGAF L PGP-SPPGSTG-----LTVYLAGVQGVVG
G2153	VAA-APGGAVALQGRFEILSLTG SFL PGP-APPGSTG-----LTIYLAGGQGVVG
G3456	--T-APGA--VMALHGRFDILSLTG SFL PGP-SPPGATG-----LTIYLAGGQGVIG
G3401	--A-APSA--VVALRGRFEILSLTG TFL PGP-APPGSTG-----LTVYLAGGQGVVG
G3403	--S-APGA--VVALHGRFEILSLTG TFL PGP-APPGSTG-----LTVYLAGGQGVVG
G2157	----AAS-GGVVSLRGQFEILSMCGAF L P TSGSPAAAAG-----LTIYLAGAQQGVVG
G3462	----AAP-AGVITLHGRFEILSLSGAF L P SP-SPGATG-----LTVYLAGGQGVVG
G3556	----AGTGAAVALRGRFEILSM SGAFL P AP-APPGATG-----LAVYLAGGQGVVG
G1067	NGGVSGGGVVTLHGRFEILSLTG T V L P P P-APPGAGG-----LSIFLAGGQGVVG
G2156	AHGANGGTGGVVALHGRFEILSLTG T V L P P P-APPGSGG-----LSIFLSGVQGVIG
G3400	PGS-----LVATMRGQFEILSLTG T V L P P P-APPSASG-----LTVFLSGGQGVVG
G3399	PGS-----MVATLRGRFEILSLTG T V L P P P-APPGASG-----LTVFLSGGQGVIG
G1073	AGG-----GVITLHGRFDILSLTG T A L P P P-APPGAGG-----LTVYLAGGQGVVG
G1945	----LGST---ITFHGKFDLLSVSATFLPPPPRTSLSP-----VSNFFTVSLAGPQGVIG
G2155	----PGST---ITFHGKFDLLSVSATFLPPPPRTSLSP-----VSNFFTVSLAGPQGVIG
G3408	----PGSAPAAIVFHGRYEILSLSATFLPP-AMSSVAPQAAVAAAGLSISLAGPHQVIG

Segundo dominio conservado

FIGURA 5E

G2789	GSVVGPLTASSPVVMAASFNGNASYERLPLEEEETER-----EIDGNAARAIG
G596	GSVVGPLMASGPVIMAAASFNGAAAYERLPLEEDDQEEQTA---GAVANNIDGNATMGG
G1070	GSVVGPLLACAGPVVMAASFNSAAAYERLPLEEDEMQTP-----VHGGGGGSL
G3405	GSVVGALTAAGPVVIMAAASFANAVYERLPLEDELLAAQG-----QADSAGLLAAG
G3404	GTVAGPLIAGPVVIMAAASFNGAAAYERLPLEDDEFPQH-----MAGGQSSPP
G1075	GNVVGELMAAGPVVMAASFNTNVAYERLPLDEHEEHLQS-----GGG---GG
G1076	GSVVGELTAAGPVIVIAAASFTNVAYERLPLEEDEQQQL-----GGSN---GG
G3460	GNVVGELTAAGPVIVIAAASFTNVAYERLPLEEDEQQQQ-----LQIQSPATTSS---QG
G3459	GNVIGELTAAGPVIVIAAASFTNVAYERLPLEEDEQQQQ-----LQIQPPATTSS---QG
G3458	GNVVGPLVASGPVIVIASSFTNVAYERLPLDEDESMQM-----QGSSAGDGS GDHG
G3406	GNVVGALYAAAGPVIVIAASFANVAYERLPLEEEEAAPPQA---GLQMQQPGGAD---AG
G3407	GSVAGALIAAGPVVVAASFNSVAYERLPLEDGDDEVVPPA---PAGSDQGGGSG---GM
G1069	GSVVGPLLAIGSVMVIAATFSNATYERLPMEEEDGG-----SRQIHGGGDSF
G2153	GSVVGPLMAAGPVMLIAATFSNATYERLPLEEEEAER-----GGGSGSGVVP
G3456	GGVVGPLVAAAGPVLVMAATFSNATYERLPLEDDQEQH-----GG---GGGSPQ
G3401	GSVVGTLTAAGPVMVIASTFANATYERLPLDQEEEE-----AAAGMMAPPP
G3403	GSVVGSLIAAGPVMVIASTFANATYERLPLEEEEG-----SGPPMPGGAEP
G2157	GGVAGPLIASGPVIVIAATFCNATYERLPIEEEQ-QQE-----QPLQLEDGKKQ
G3462	GNVAGSLVASGPVMVIAATFANATYERLPLEDDQEGEE-----MQVQQQQQQQ
G3556	GSVMGELIASGPVMVIAATFGNATYERLPLDQEG-EEG-----AVLSGSEGAAA
G1067	GSVVAPLIASAPVILMAASFNSAVFERLPIEEEE-----GGG---GGGGGPPQ
G2156	GNVVAPLVASGPVILMAASFNSATFERLPLEDEGGE-----GEGGEGVGGGEGG
G3400	GSVAGQLIAAGPVFLMAASFANAVYERLPLDGEDPEAEAA---AATPPGDAAPTGPFP
G3399	GSVVGPLVAAAGPVVLMMAASFANAVYERLPLEGEEEEEVAAP---AAGGEAQDQVAQSAGP
G1073	GNVAGSLIASGPVLMMAASFANAVYDRPLPIEEETPPRT---TGVOQQQPEASQSSEV
G1945	GFVAGPLISAGTVYVIAASFNNPSYHRLPAEEEEQ-----KHSAGTGEREG
G2155	GFVAGPLVAAAGTVYFVATSFKNPSYHRLPATEEE-----QRNSAEGEEEG
G3408	GAVAGPLYAAATTVVVVAAAFTNPTFHRPLPADDDASVSVSVLSGSGDADEHRGHQHKPE

Segundo dominio conservado

FIGURA 5F

G2789	TQTQ	-----KQLMQDATSFIG-----	-----SPSNLINSVSLP-----
G596	TQTQTQTQ	-----QQLMQDPTSFIOG-----	-----LPPNLMNSVQLP-----
G1070	ESPPMMG	-----QQLQHQQQAMSGHQG-----	-----LPPNLLGVSQLOQQ-----H
G3405	QQA	-----QLAGGAVDPSLFIQG-----	-----LPPNLLIGNVQLPP-----
G3404	PPPLP	-----LPPHQQPILQDH-----	-----LPHNLMNGIHLPG-----
G1075	GNNYSE	-----AT-GGGG-----	-----LPPFNLPN-SMP-----Q
G1076	GNLFPE	-----VAAGGGG-----	-----LPPFNLPN-NMQPN-----VQ
G3460	NNNNN	-----PPDPSSG-----	-----LPPFNLPN-NMQN-----VQ
G3459	NNNNN	-----PPDPSSG-----	-----LPPFNLPN-NMQN-----VQ
G3458	GGVSNN	-----SFPDPSSG-----	-----LPPFNLPN-NMP-----Q
G3406	GMGGA	-----PPDPSAAG-----	-----LPPFNLPNMPGGG-----GSQ
G3407	PPLGVD	-----PSGGAATG-----	-----LPPFNMPF-GMP-----P
G1069	PRIGS	-----NLPDLSGMAGPG-----	-----YNMPPHL-IPNG-----AG-----QL-
G2153	GQLGG	-----GGSPSSGAGGDNQGLPVYNMPGNL-VSNGSGGG-----	-----QMS
G3456	EKTGG	-----PGEASSISVYN-----	-----NNVPPSLGLPNG-----QHL
G3401	LMAG	-----AADP--LLFGG-----	-----GMHDAGLAAW-----
G3403	LMAGGH	-----GIADPSALPMFN-----	-----LPPSNGLGCG-----
G2157	KEEND	-----NESGNNGNEGSMQ-----	-----PPMYNMPNFI PNHG-----QMA
G3462	QQQQQ	-----QQQSQGLGEQVS-----	-----MPMYNLPNLLHNGQ-----NMP
G3556	QMEQ	-----SSG-GAVVP-----	-----PPMY-----AAVQ-----QTP
G1067	MQQAPS	-----ASPPSGVTGQOQ-----	-----LGGNVGGYG-----FSGDP
G2156	PPATS	-----SSPPSG-AGQOQ-----	-----LRGNMSGYDQ-----FAGDP
G3400	PQQQPT	-----ASQSEVTAGDG-GGGGGLGMY-LGGHVGSYQQQOQLP-----	-----GPGDN
G3399	PGQQA	-----ASQSSGVTGGGTGGAGGMSLYNLAGNVGGYQL-----	-----PGDN
G1073	TGSGAQ	-----ACESN-LQGGNG-----	-----GGGVAFYNLGMNNNFQS-----GGDI
G1945	QSP	-----PVSGGEESSGMAGSGGESCGVSMYSCHMGGSD-----	-----V
G2155	QSP	-----PVSGGG-----	-----GESMY-----VGGSD-----V
G3408	QEPRQLR	-----RPPPHLSAAAAVSAAPVEPCGAPMYACHPQOE-----	-----V

FIGURA 5G

G2789	GEAYWGTQ-----RPS-F-----
G596	AEAYWGT-----RPS-F-----
G1070	DQSYWSTG-----RPP-Y-----
G3405	EAYGWNFGAGG-RPAF-----
G3404	DAAYGWTSGGGGGAAPY-----
G1075	IGVESWQGNH-AGAGRAF-----
G1076	LPVEGWPGNS-G--GRGPF-----
G3460	LP-----PF-----
G3459	LPVEGWAVNP-AS-RPQPF-----
G3458	LPVDGWAGNS-GG-R-QSY-----
G3406	LPPGADGHGW-AG-ARPPF-----
G3407	MPVDGHAGWPGAGVGRPPFS-----
G1069	GHEPYT--W---VHARPPY-----
G2153	GQEAYG--W---AQARSGF-----
G3456	NHEAYSSPWGHSFHARPPF-----
G3401	-HHARP-----PPPPPY-----
G3403	-GDGFPW-----AAHPRPPY-----
G2157	QHDVYWGG-----PPPRAPP-SY-----
G3462	-HDVFWG-----APPRPPP-SF-----
G3556	PHDMFGQWG-HAAVARPPPTSF-----
G1067	HLL-GWGAG---TPSRPPF-----
G2156	HLL-GWGAAAAAAPPRAAF-----
G3400	FG--SWSGS-----IRPPPF-----
G3399	FG--GWSGAGAGGVR-PPF-----
G1073	YGMSSGSGGGGGGATRAAF-----
G1945	IWA--PTARAP-----PPY-----
G2155	IWD--PNAKAP-----SPY-----
G3408	MWP--PPARTPHPPPPPPY-----

FIGURA 5H

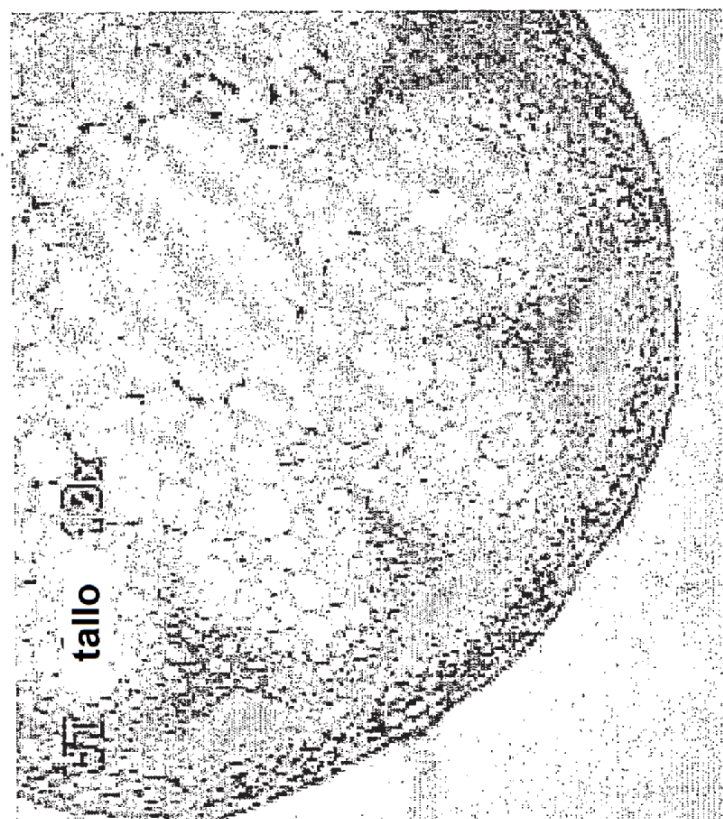


FIGURA 6A

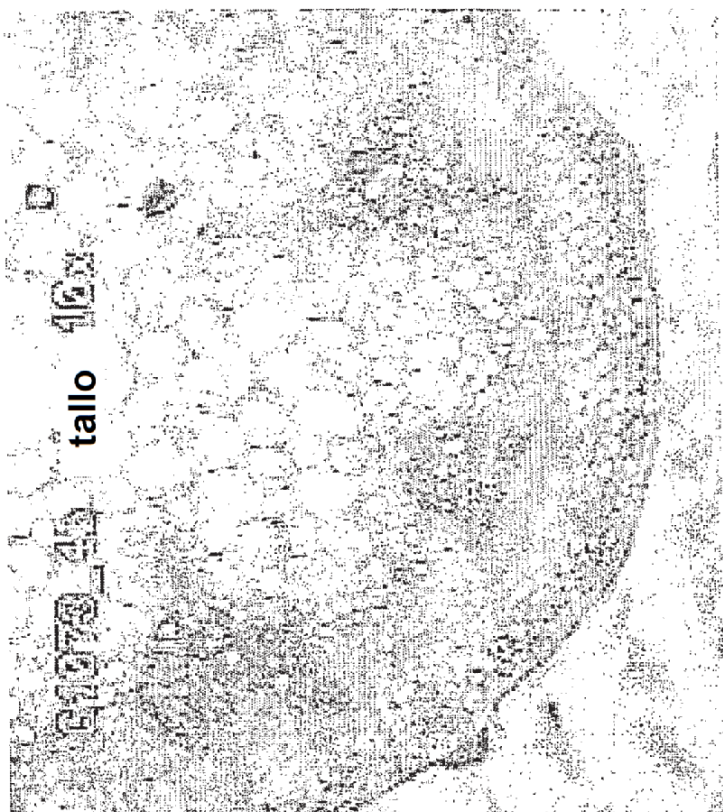


FIGURA 6B

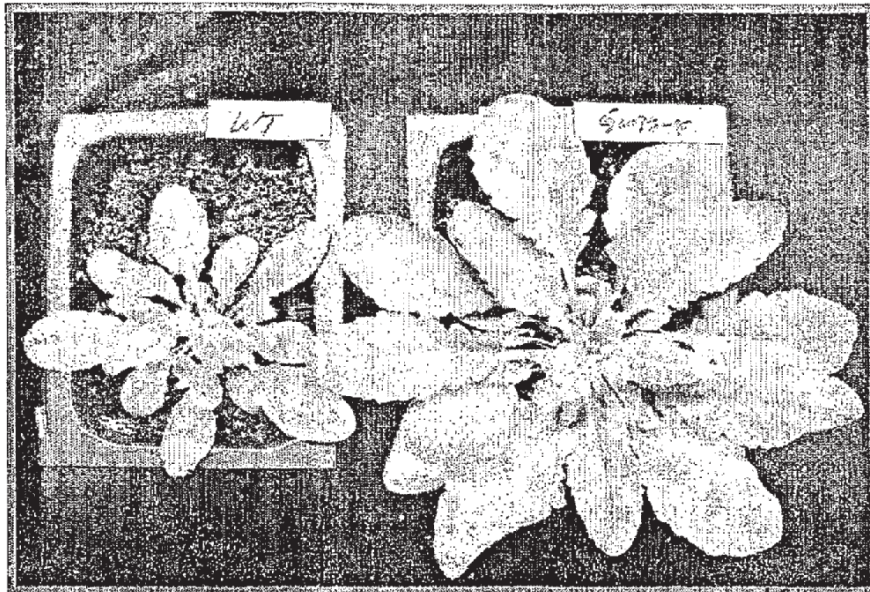


FIGURA 7A

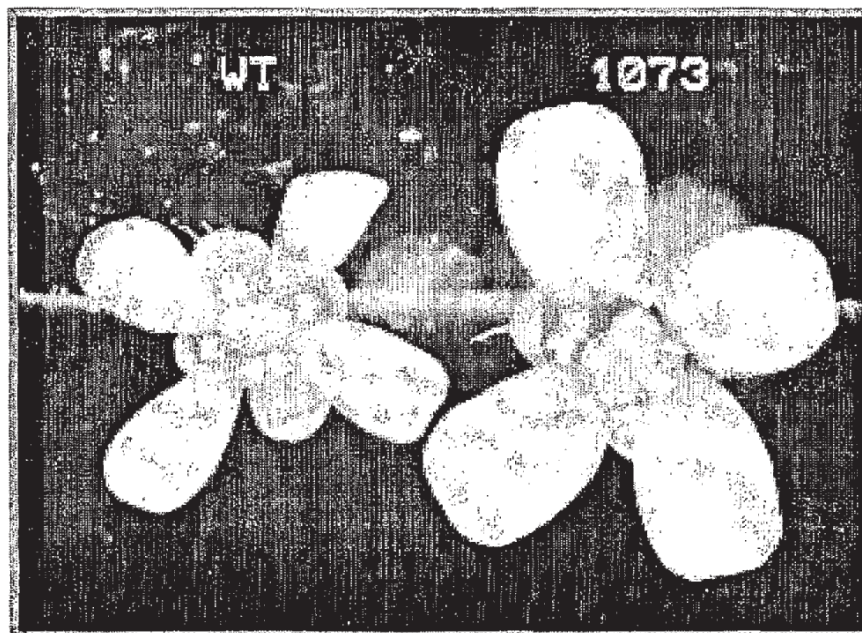


FIGURA 7B

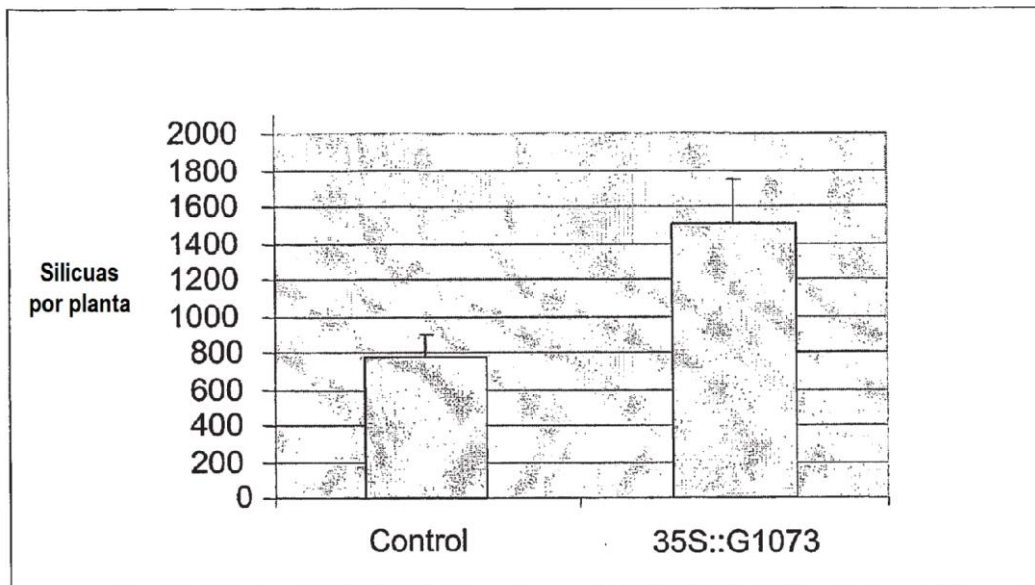


FIGURA 8



FIGURA 9A

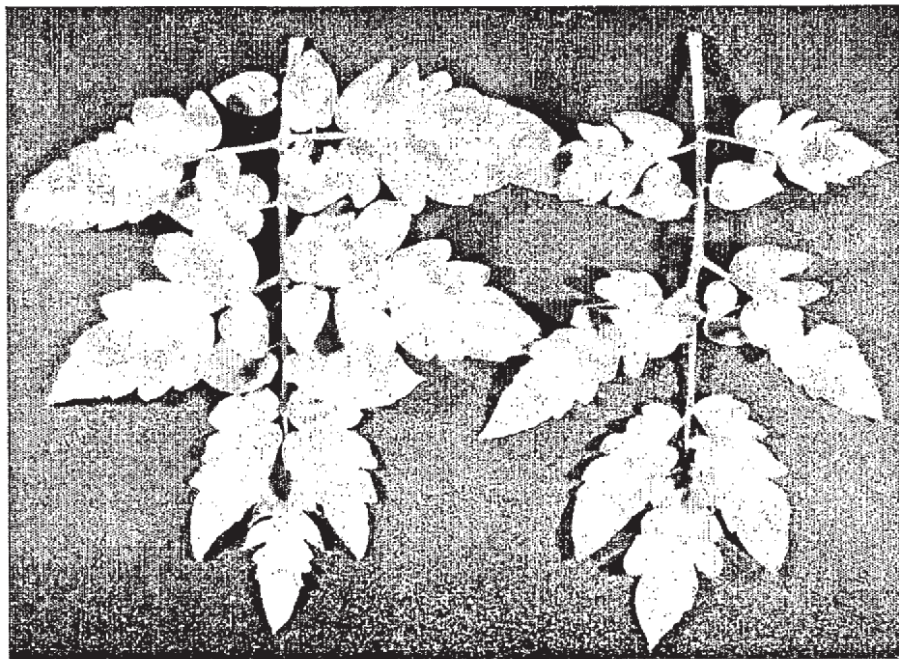


FIGURA 9B

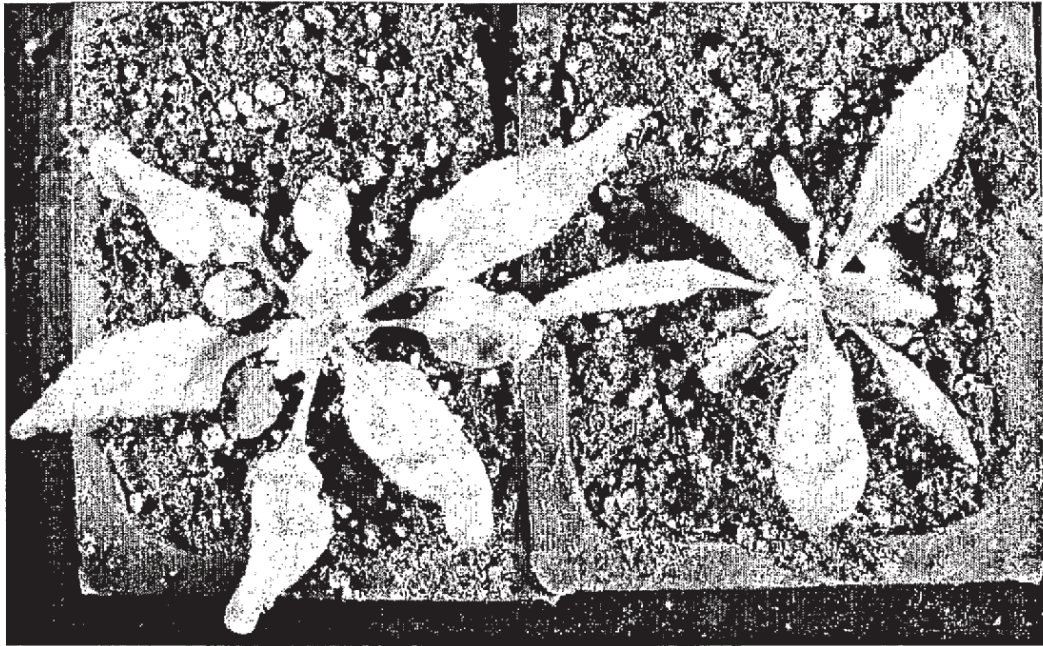


FIGURA 10A

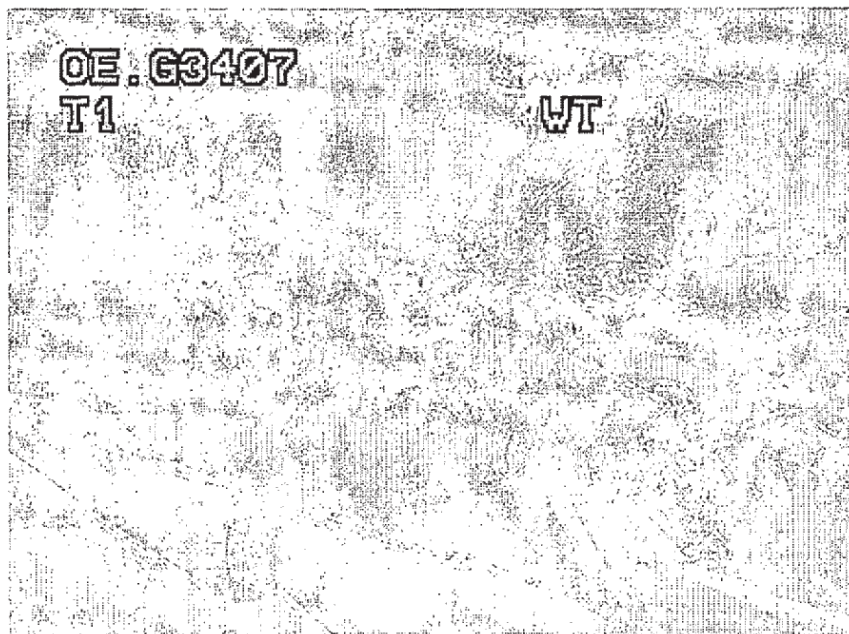


FIGURA 10B

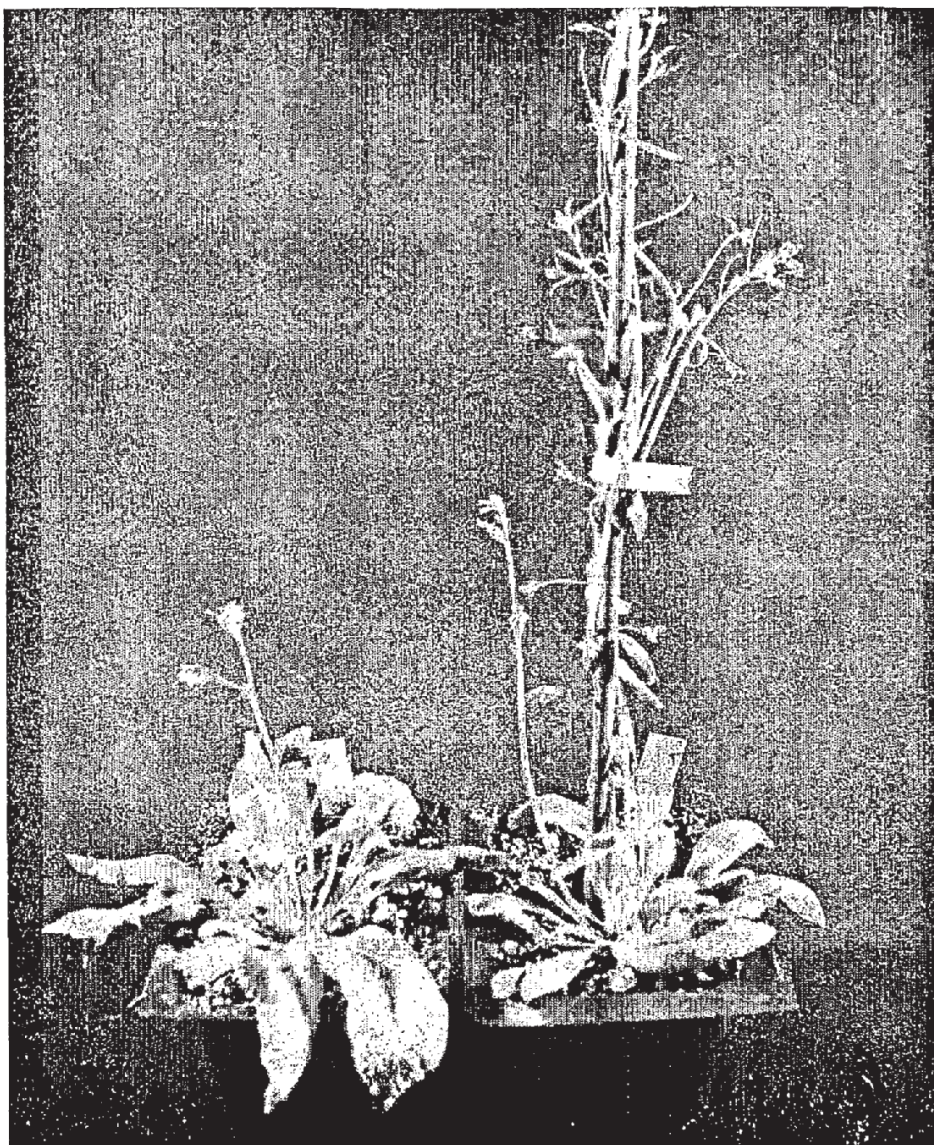


FIGURA 11



FIGURA 12