



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

①① Número de publicación: **2 361 749**

⑤① Int. Cl.:
C12N 15/82 (2006.01)

⑫ **TRADUCCIÓN DE PATENTE EUROPEA**

T3

⑨⑥ Número de solicitud europea: **04816948 .6**

⑨⑥ Fecha de presentación : **12.11.2004**

⑨⑦ Número de publicación de la solicitud: **1682668**

⑨⑦ Fecha de publicación de la solicitud: **26.07.2006**

⑤④ Título: **Reguladores transcripcionales de plantas.**

③⑩ Prioridad: **13.11.2003 US 714887**
05.12.2003 US 527658 P
05.02.2004 US 542928 P

④⑤ Fecha de publicación de la mención BOPI:
21.06.2011

④⑤ Fecha de la publicación del folleto de la patente:
21.06.2011

⑦③ Titular/es: **Mendel Biotechnology, Inc.**
3935 Point Eden Way
Hayward, California 94545, US

⑦② Inventor/es: **Heard, Jacqueline, E.;**
Riechmann, José Luis;
Creelman, Robert, A.;
Ratcliffe, Oliver, J.;
Canales, Roger, D.;
Repetti, Peter;
Kumimoto, Roderick, W.;
Gutterson, Neal, I.;
Reuber, T., Lynne;
Pineda, Omaira;
Sherman, Bradley, K.;
Morrison, Tracy, A.;
Keddie, James, S.;
Jiang, Cai-Zhong;
Century, Karen, S.;
Adam, Luc;
Zhang, James, Z.;
Hempel, Frederick, D. y
Libby, Jeffrey, M.

⑦④ Agente: **Ponti Sales, Adelaida**

ES 2 361 749 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

Reguladores transcripcionales de plantas

Campo de la invención

5

[0001] La presente invención se refiere a composiciones y procedimientos para modificar el fenotipo de una planta, para proporcionar una mayor tolerancia a la sequía.

Antecedentes de la invención

10

[0002] Control de procesos celulares por factores de transcripción. Los estudios de una diversidad de organismos procariontas y eucariotas sugieren una evolución gradual de los mecanismos bioquímicos y fisiológicos y de las rutas metabólicas. A pesar de diferentes presiones evolutivas, las proteínas que regulan el ciclo celular en levaduras, plantas, nematodos, moscas, ratas y el hombre, tienen características químicas o estructurales comunes y modulan la misma actividad celular general. Una comparación de secuencias de genes con estructura y/o función conocidas de una especie de planta, por ejemplo, *Arabidopsis thaliana*, con las de otras plantas, permite a los investigadores desarrollar modelos para manipular los rasgos de una planta y desarrollar variedades con propiedades valiosas.

15

[0003] Los rasgos de una planta se pueden controlar mediante una serie de procesos celulares. Una forma importante de manipular este control es mediante factores de transcripción, proteínas que influyen en la expresión de un gen o conjunto 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 las rutas biológicas enteras 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 puede producir plantas y cultivos con propiedades nuevas y/o mejoradas comercialmente valiosas, incluyendo rasgos que mejoran el rendimiento o la supervivencia y el rendimiento durante periodos de estrés abiótico, mejoran la tolerancia a la sombra, o alteran la detección de la planta de su equilibrio de carbono/nitrógeno.

25

[0004] Problemas asociados con la falta de agua. En el entorno natural, las plantas a menudo crecen en condiciones que no son favorables, tales como sequía (poca disponibilidad de agua), alta salinidad, frío, heladas, o alta temperatura. Cualquiera de estos estreses abióticos puede retrasar el crecimiento y el desarrollo, reducir la productividad y en casos extremos, producir la muerte de la planta. De estos estreses, la poca disponibilidad de agua, que en una forma drástica se denomina una sequía, es un factor principal en la reducción del rendimiento de los cultivos en el mundo. Los problemas para las plantas producidos por la baja disponibilidad de agua incluyen estreses mecánicos causados por la extracción de agua celular. La sequía también provoca que las plantas se vuelvan más susceptibles a diferentes enfermedades (Simpson, ed. (1981) "The Value of Physiological Knowledge of Water Stress in Plants", en Water Stress on Plants, Praeger, NY, pág. 235-265).

35

[0005] Además de las muchas regiones en el mundo que son demasiado áridas para la mayoría si no para todas las plantas de cultivo, el uso excesivo del agua disponible está produciendo una mayor pérdida del suelo que se puede usar en agricultura, un proceso que, en el extremo, da como resultado la desertificación. El problema se combina además con la acumulación creciente de sal en los suelos, lo cual se añade a la pérdida de agua disponible en los suelos.

45

[0006] La falta de agua es un componente común de los estreses de muchas plantas. La falta de agua se produce en las células de las plantas cuando la velocidad de transpiración de la planta entera supera la absorción de agua. Además de la sequía, otros estreses tales como la salinidad y la baja temperatura, producen la deshidratación celular (McCue y Hanson (1990) *Trends Biotechnol.* 8: 358-362).

50

[0007] La transducción de señales del estrés salino y por sequía incluye las rutas de señalización de la homeostasis iónica y osmótica. El aspecto iónico del estrés salino es señalado por la ruta de SOS en la que un complejo de proteína quinasas SOS3-SOS2 sensible a calcio controla la expresión y la actividad de los transportadores de iones tales como SOS1. La ruta que regula la homeostasis iónica en respuesta al estrés salino se ha descrito recientemente por Xiong y Zhu (2002) *Plant Cell Environ.* 25: 131-139 y Ohta y col. (2003) *Proc. Natl. Acad. Sci. USA* 100:11771-11776.

55

[0008] El componente osmótico del estrés salino implica reacciones complejas de la planta que se solapan con respuestas al estrés por sequía y/o baja temperatura.

[0009] Recientemente se han revisado aspectos comunes de las respuestas al estrés por sequía, frío y salino por Xiong y Zhu (2002), véase antes. Estos incluyen

- 5 (a) cambios transitorios en los niveles citoplasmáticos de calcio muy pronto en el suceso de señalización (Knight, (2000) *Int. Rev. Cytol.* 195: 269-324; Sanders y col. (1999) *Plant. Cell* 11: 691-706);
- (b) transducción de señales a través de proteína quinasas (CDPKs; Xiong y Zhu (2002) véase antes) y proteína fosfatasa (Merlot y col. (2001) *Plant J.* 25: 295-303; Tahtiharju y Palva (2001) *Plant J.* 26: 461-470) activadas por mitógenos y/o dependientes de calcio;
- 10 (c) aumento en los niveles de ácido abscísico (ABA) en respuesta al estrés desencadenando un subconjunto de respuestas (Xiong y Zhu (2002); véase antes, y referencias citadas en el mismo);
- (d) fosfatos de inositol como moléculas señal (al menos para un subconjunto de cambios transcripcionales de respuesta al estrés (Xiong y col. (2001) *Genes Dev.* 15: 1971-1984);
- 15 (e) activación de fosfolipasas que a su vez genera un conjunto diverso de moléculas de segundo mensajero, algunas de las cuales pueden regular la actividad de las quinasas sensibles al estrés (la fosfolipasa D funciona en una ruta independiente del ABA Frank y col. (2000) *Plant Cell* 12:111-124);
- (f) inducción de genes de tipo de embriogénesis tardía abundantes (LEA) incluyendo genes COR/RD sensibles a CRT/DRE (Xiong y Zhu (2002) véase antes);
- 20 (g) mayores niveles de antioxidantes y osmolitos compatibles tales como prolina y azúcares solubles (Hasegawa y col. (2000) *Annu. Rev. Plant Mol. Plant Physiol.* 51: 463-499); y
- (h) acumulación de especies de oxígeno reactivas tales como superóxido, peróxido de hidrógeno y radicales hidróxido (Hasegawa y col. (2000) véase antes).

[0010] La biosíntesis del ácido abscísico se regula mediante el estrés osmótico en múltiples puntos. La señalización del estrés osmótico tanto dependiente como independiente de ABA modifica primero los factores de transcripción expresados de forma constitutiva, conduciendo a la expresión de activadores transcripcionales de respuesta temprana, que después activan activadores transcripcionales en dirección 3' y genes efectores de la tolerancia al estrés.

30 **[0011]** Basándose en la concordancia de muchos aspectos de las respuestas al estrés por temperatura baja, sequía y salino, se puede concluir que los genes que aumentan la tolerancia al estrés por temperatura baja o salinidad también pueden mejorar la protección frente al estrés por sequía. De hecho, esto ya se ha demostrado para factores de transcripción, como es el caso de AtCBF/DREB 1, y para otros genes tales como OsCDPK7 (Saijo y col. (2000) *Plant J.* 23: 319-327) o AVP1 (una bomba vacuolar de pirofosfatasa-protones, Gaxiola y col. (2001) *Proc. Natl. Acad. Sci. USA* 98: 11444-11449).

Resumen de la invención

[0012] El presente procedimiento se dirige a polinucleótidos recombinantes que confieren tolerancia al estrés abiótico en plantas cuando se altera la expresión de cualquiera de estos polinucleótidos recombinantes (p. ej., por exceso de expresión). Las secuencias relacionadas que se describen en el presente documento incluyen secuencias de nucleótidos que hibridan con el complemento de las secuencias de la invención en condiciones restrictivas. Un ejemplo de condiciones rigurosas que definen la invención incluye un procedimiento de hibridación que incorpora dos etapas de lavado de 6x SSC y 65°C, siendo cada etapa de 10-30 minutos de duración. Por ejemplo, G2133 (polinucleótido de la SEQ ID NO: 11 y polipéptido de la SEQ ID NO: 12) confiere tolerancia a la sequía, cuando éste polipéptido es expresado en exceso en plantas. Por lo tanto, la invención incluye el uso del polinucleótido y polipéptido G2133, así como de las secuencias de nucleótidos que son estructuralmente similares en cuanto que estas o su complemento hibridan con la SEQ ID NO: 11 en condiciones de hibridación restrictivas

50 **[0013]** En particular, la invención proporciona un procedimiento para producir una planta transgénica que tiene una mayor tolerancia a la falta de agua que una planta de control, comprendiendo el procedimiento las etapas de (a) producir un vector de expresión que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de al menos 70% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12; en el que el polipéptido tiene la propiedad de la SEQ ID NO: 12 de aumentar la tolerancia a la falta de agua en una planta con respecto a la planta de control; y (b) introducir el vector de expresión en una planta objetivo para producir una planta transgénica; en el que el polipéptido es expresado en exceso en la planta transgénica y dicho exceso de expresión da como resultado que la planta transgénica tiene una mayor tolerancia a la falta de agua que la planta de control.

[0014] La invención proporciona además una semilla transgénica que comprende un vector de expresión que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de al menos 70% con las coordenadas de los aminoácidos 10-77 de la SEQ ID NO: 12 producido por una planta transgénica producida por cualquiera de los procedimientos de las reivindicaciones 1 a 3, en el que una planta descendiente que ha crecido a partir de la semilla transgénica tiene una tolerancia mayor a la sequía con respecto a una planta de control; y en el que dicha secuencia está operativamente unida a un promotor que es inducible en respuesta al estrés por calor, sequía u osmótico, así como una semilla, fruta, hoja, raíz o descendiente de una planta que comprende un vector de expresión que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de al menos 70% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12; en el que el polipéptido tiene la propiedad de la SEQ ID NO:12 de aumentar la tolerancia a la falta de agua en una planta con respecto a la planta de control, que se puede obtener por el procedimiento de una cualquiera de las reivindicaciones 1 a 3, en el que dicha planta, semilla, fruta, hoja, raíz, célula vegetal o descendiente tiene una mayor tolerancia a la falta de agua que la planta de control, y en el que dicha secuencia está operativamente unida a un promotor que es inducible en respuesta al estrés por calor o sequía.

[0015] La invención proporciona además, el uso de un vector de expresión para producir una planta transgénica que tiene una mayor tolerancia a la falta de agua con respecto a una planta de control, en el que el vector de expresión codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de 70% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12.

[0016] La invención también se refiere a un procedimiento para producir una planta transgénica que tiene una mayor tolerancia a la sequía. Las etapas de este procedimiento incluyen primero proporcionar un vector de expresión que contiene una secuencia de nucleótidos que hibrida con el complemento de un polinucleótido de la invención en condiciones de hibridación restrictivas. El vector de expresión después se introduce en una célula de planta, la célula de planta se cultiva y a partir de esta se genera una planta. Debido a la presencia del vector de expresión en la planta, el polipéptido codificado por la secuencia de nucleótidos es expresado en exceso. Este polipéptido tiene la propiedad de regular la sequía en una planta, comparado con una planta de control que no expresa en exceso el polipéptido. Después de producir la planta transgénica tolerante a la sequía, se puede identificar comparándola con una o más plantas no transformadas que no expresan en exceso el polipéptido. Estas etapas del procedimiento pueden incluir además la autofecundación o cruce de la planta tolerante al estrés abiótico consigo misma o con otra planta, respectivamente, para producir la semilla. La "autofecundación" se refiere a la autopolinización, o el uso de polen de una planta para fertilizar la misma planta u otra planta en la misma línea, mientras que " cruzamiento" en general se refiere a la polinización cruzada con una planta de una línea diferente, tal como una planta silvestre o no transformada, u otra planta transformada de una línea transgénica de plantas diferentes. El cruzamiento proporciona la ventaja de poder producir nuevas variedades. La semilla resultante después se puede usar para cultivar una planta descendiente que es transgénica y tiene una mayor tolerancia al estrés abiótico.

[0017] Por lo tanto, en el presente documento se describe un procedimiento para aumentar la tolerancia de una planta a la sequía. Este procedimiento incluye proporcionar primero un vector que comprende (i) elementos reguladores eficaces en el control de la expresión de una secuencia de polinucleótido en una planta diana, en el que los elementos reguladores flanquean la secuencia de polinucleótido; y (ii) las propias secuencias de polinucleótidos que codifican un polipéptido que tiene la capacidad de regular la tolerancia a la sequía en una planta, cuando se compara con una planta de control de la misma especie que no expresa en exceso el polipéptido. La planta se transforma con el vector con el fin de generar una planta transformada con una mayor tolerancia a la sequía.

Breve descripción de la lista de secuencias y las figuras

[0018] La lista de secuencias proporciona secuencias de polinucleótidos y polipéptidos de ejemplo de la invención. Las características asociadas con el uso de las secuencias están incluidas en los ejemplos.

[0019] CD-ROM1 y CD-ROM2 son discos compactos de memoria de solo lectura legibles en el ordenador. El contenido de cada CD es idéntico, y cada uno contiene una copia de la Lista de Secuencias en formato de texto ASCII. La lista de secuencias se denomina "MBI0058PCT.ST25.txt" y tiene un tamaño de 2004 kilobytes. Las copias de la Lista de Secuencias en los discos de CD-ROM se incorporan en el presente documento por referencia en su totalidad.

La figura 1 muestra una evaluación conservativa de relaciones filogenéticas entre los órdenes de plantas con flores (modificado de Angiosperm Phylogeny Group (1998) *Ann. Missouri Bot. Gard.* 84: 1-49). Las plantas con un solo cotiledón (monocotiledóneas) son un clado monofilético anidado dentro de al menos dos linajes principales de dicotiledóneas; las eudicotiledóneas se dividen además en rósidas y astérides. *Arabidopsis* es una eucotiledónea rósida dentro del orden de las brasicáceas; el arroz es un miembro de las monocotiledóneas del orden poales. La figura 1 se ha adaptado de Daly y col. (2001) *Plant Physiol.* 127: 1328-1333.

La figura 2 muestra un dendograma filogenético que representa las relaciones de la taxonomía de las plantas superiores, incluyendo clados que contienen el tomate y *Arabidopsis*, adaptado de Ku y col. (2000) *Proc. Natl. Acad. Sci.* 97: 9121-9126; y Chase y col. (1993) *Ann. Missouri Bot. Gard.* 80: 528-580.

La figura 3 es un alineamiento múltiple de secuencias de aminoácidos dentro del dominio AP2 de G47, G2133 y sus ortólogos. Los ortólogos y parálogos del clado se indican mediante la barra negra en el lado izquierdo de la figura. De las secuencias examinadas hasta la fecha, se encontró que había dos restos de valina presentes en los miembros del clado G47 pero no fuera del clado (flechas). Los restos que se pueden usar para identificar un miembro del clado G47 se indican mediante los restos mostrados en las cajas en la figura 3.

La figura 4 ilustra la relación de G47 y secuencias relacionadas en este árbol filogenético del clado G47 y secuencias similares. El procedimiento de construcción del árbol usado fue la "Unión de vecinos" con "muerte rápida sistemática" y Bootstrapping con 1000 réplicas (no corregido ("p"), con huecos distribuidos de forma proporcional). Se usaron polipéptidos de longitud entera para construir la filogenia como se define en la figura 4. Se predice que los miembros del clado mostrados dentro de la caja contienen homólogos funcionales de G47. Abreviaturas: At *Arabidopsis thaliana*; Os *Oryza sativa*; Zm *Zea mays*; Gm *Glycine max*; Mt *Medicago truncatula*; Br *Brassica rapa*; Bo *Brassica oleracea*; Ze; *Zinnia elegans*.

Las figuras 5A y 5B comparan la recuperación de un tratamiento de sequía de controles silvestres y dos líneas de plantas *Arabidopsis* que expresan en exceso G2133, un parálogo de G47. Las figuras 5A y 5B muestran dos líneas de plantas 35S::G2133 (una línea en cada figura) en la maceta a la izquierda de cada figura y plantas de control a la derecha en cada figura. Cada maceta contenía varias plantas cultivadas con 24 h de luz. Se privó a todas de agua durante 8 días, y se muestran después de volver a regarlas. Todas las plantas de las líneas que expresaban en exceso G2133 se recuperaron, y todas las plantas de control murieron o estaban adversa y gravemente afectadas por el tratamiento de la sequía.

Descripción detallada de las realizaciones específicas

[0020] Los datos presentados en el presente documento representan los resultados de un cribado de una colección de factores de transcripción para identificar genes que se pueden aplicar para reducir las pérdidas de rendimiento que se producen por respuestas a la falta de nutrientes, estrés relacionado con la sequía y/o evitación de sombra.

[0021] Los autores de la invención han identificado numerosos genes de factores de transcripción que confieren una mayor tolerancia a la sequía con respecto a las plantas silvestres cuando se altera su expresión, tal como por un exceso de expresión o inactivación del gen en plantas transgénicas. Por lo tanto, la presente invención se dirige en parte a polinucleótidos recombinantes que confieren tolerancia al estrés relacionado con la sequía, a las plantas cuando se altera la expresión de los polinucleótidos recombinantes de la invención (p. ej., por exceso de expresión). En los presentes estudios, se llevaron a cabo ensayos basados en el suelo en los que primero se priva de agua a las plantas transgénicas, se evalúan por comparación con plantas de control, se vuelven a regar, y se evalúa también su recuperación en comparación con las plantas de control tratadas de forma similar.

[0022] Se llevó a cabo un segundo cribado en el que 2 o 3 líneas individuales de expresión en exceso (o un lote de semillas homocigotas diferentes, en el caso de líneas con genes inactivados) se volvieron a probar en el ensayo. A las líneas transgénicas individuales que mostraron fenotipos destacados en este segundo ciclo de ensayos se les dio una clasificación prioritaria "A". El conjunto de secuencias a las que se asignó una clasificación de prioridad "B" en la tabla de resultados debían confirmarse todavía en un segundo ensayo o no tenían un fenotipo destacado.

[0023] La presente invención se refiere en parte a polinucleótidos y polipéptidos, por ejemplo, para modificar fenotipos de plantas, en particular los asociados con la mejora del estrés por sequía. A lo largo de toda esta descripción, se hace referencia a diferentes fuentes de información. Las fuentes de información incluyen artículos de

revistas científicas, documentos de patentes, libros de texto, y direcciones de páginas inactivas del explorador de la Red. La referencia a estas fuentes de información indican claramente que las puede usar el experto en la materia. Se puede contar con y usar los contenidos y las enseñanzas de todas y cada una de las fuentes de información para hacer y usar las realizaciones de la invención.

5

[0024] Como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares “un”, “una” y “el”, “la” incluyen las referencias plurales salvo que se indique de otra forma claramente en el contexto. Por lo tanto, por ejemplo, una referencia a “una planta” incluye una pluralidad de dichas plantas, y una referencia a “un estrés” es una referencia a uno o más estreses y equivalentes de los mismos conocidos por los expertos en la materia, etc.

10

Definiciones

[0025] “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, de doble cadena o una cadena, y combinado con hidratos de carbono, lípidos, proteínas u otros materiales para llevar a cabo una actividad particular tal como transformación o formación de una composición útil tal como un ácido nucleico peptídico (PNA).

[0026] “Polinucleótido” es una molécula de ácido nucleico que comprende una pluralidad de nucleótidos polimerizados, por ejemplo, al menos aproximadamente 15 o más 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. Además, el polinucleótido puede comprender un promotor, un intrón, una región potenciadora, un sitio de poliadenilación, un sitio de inicio de la traducción, regiones 5' o 3' no traducidas, un gen indicador, un marcador seleccionable, o similares. El polinucleótido puede ser ADN o ARN de una cadena o de doble cadena. 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 la PCR, un ADN clonado, un ADN o ARN sintético, o similares. El polinucleótido puede estar combinado con hidratos de carbono, lípidos, proteínas u otros materiales para realizar una actividad particular tal como una transformación o formar una composición útil tal como un ácido nucleico-peptídico (PNA). El polinucleótido puede comprender una secuencia en orientación de sentido directo o sentido contrario. “Oligonucleótido” es sustancialmente equivalente a los términos amplímero, cebador, oligómero, elemento, diana y sonda y es preferiblemente monocatenario.

[0027] “Gen” o “secuencia génica” se refiere a la secuencia codificante parcial o completa de un gen, su complemento y sus regiones 5' o 3' no traducidas. Un gen también es una unidad funcional de herencia, y en términos físicos es un segmento o secuencia particular 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 de polipéptido. Esta última se puede someter a un procesamiento posterior tal como modificación química, corte y empalme, y plegado para obtener una proteína o polipéptido funcional. Un gen se puede aislar parcialmente o puede encontrarse con un genoma del organismo. A modo de ejemplo, un gen de un factor de transcripción codifica un polipéptido de factor de transcripción, que puede ser funcional o requiere procesamiento para funcionar como un iniciador de la transcripción.

[0028] Desde un punto de vista operativo, los genes se pueden definir por el ensayo cis-trans, un ensayo genético que determina si hay dos mutaciones en el mismo gen y esto se puede usar para determinar los límites de la unidad genéticamente activa (Rieger y col. (1976) “Glossary of Genetics and Cytogenetics: Classical and Molecular”, 4ª ed., Springer Verlag, Berlín). Un gen en general incluye regiones que preceden (“líderes”; en la dirección 5') y que siguen (“remolques”; dirección 3') a la región codificante. Un gen puede incluir también secuencias no codificantes, intercaladas, denominadas “intrones” localizadas entre segmentos codificantes individuales denominados “exones”. La mayoría de los genes tienen una región promotora asociada, una secuencia 5' reguladora del codón de inicio de la transcripción (hay algunos genes que no tienen un promotor identificable). La función de un gen también puede ser regulada por potenciadores, operadores y otros elementos reguladores.

[0029] “Un polinucleótido recombinante” es un polinucleótido que no está en su estado natural, por ejemplo, el polinucleótido comprende una secuencia de nucleótidos que no se encuentra en la naturaleza, o el polinucleótido está en un contexto distinto del que se encuentra de forma natural, por ejemplo, separado de secuencias de nucleótidos de las que normalmente está próximo en la naturaleza, o secuencias de nucleótidos adyacentes (o contiguas) con las que normalmente no está próximo. Por ejemplo, la secuencia en cuestión se puede clonar en un vector, o recombinar de otra forma con uno o más ácidos nucleicos adicionales.

[0030] Un “polinucleótido aislado” es un polinucleótido, sea natural o recombinante, que está presente fuera de la célula en la que normalmente se encuentra en la naturaleza, sea 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.

[0031] 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 un fragmento del mismo. Además, el polipéptido puede comprender: (i) un dominio de localización; (ii) un dominio de activación; (iii) un dominio de represión; (iv) un dominio de oligomerización; o (v) un dominio de unión a ADN, o similares. El polipéptido comprende opcionalmente restos de aminoácidos modificados, restos de aminoácidos naturales que no codifican un codón, o restos de aminoácidos no naturales.

[0032] “Proteína” se refiere a una secuencia de aminoácidos, oligopéptido, péptido, polipéptido o partes de los mismos sean naturales o sintéticos.

[0033] “Parte”, como se usa en el presente documento, se refiere a cualquier parte de una proteína usada para cualquier propósito, pero en especial para el cribado de una biblioteca de moléculas que se unen específicamente a esa parte o para la producción de anticuerpos.

[0034] 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 procedimientos conocidos en la técnica. Un “polipéptido aislado”, sea un polipéptido natural o recombinante, está más enriquecido en (o fuera de) una célula que el polipéptido en su estado natural en una célula salvaje, por ejemplo enriquecido en más de aproximadamente 5%, o al menos 105% con respecto al tipo salvaje normalizado a 100%. Dicho enriquecimiento no es el resultado de una respuesta natural de una planta silvestre. Alternativa o adicionalmente, el polipéptido aislado se separa de otros componentes celulares con los que está normalmente asociado, por ejemplo, por cualquiera de los diferentes procedimientos de purificación de proteínas del presente documento.

[0035] “Homología” se refiere a la similitud de secuencia entre una secuencia de referencia y al menos un fragmento de un inserto de clon secuenciado nuevo o su secuencia de aminoácidos codificada. Además, las expresiones “homología” y “secuencia(s) homóloga(s)” se pueden referir a una o más secuencias de polipéptidos que se modifican por medios químicos o enzimáticos. La secuencia homóloga puede ser una secuencia modificada por lípidos, azúcares, péptidos, compuestos orgánicos o inorgánicos, usando aminoácidos modificados o similares. Las técnicas de modificación de proteínas se ilustran en Ausubel y col. (eds) “Current Protocols in Molecular Biology”, John Wiley & Sons (1998).

[0036] “Identidad” o “similitud” se refieren a la similitud de secuencias entre dos secuencias de polinucleótidos o entre dos secuencias de polipéptidos, siendo la identidad la comparación más rigurosa. Las frases “porcentaje de identidad” y “% de identidad” se refieren al porcentaje de similitud de secuencia encontrado en una comparación de dos o más secuencias de polinucleótidos o dos o más secuencias de polipéptidos. “Similitud de secuencia” se refiere al porcentaje de similitud en la secuencia de pares de bases (determinado por cualquier método adecuado) entre dos o más secuencias de polinucleótidos. Dos o más secuencias pueden tener una similitud en cualquier punto de 0-100%, o cualquier valor entero entre estos. La identidad o similitud se puede determinar comparando una posición en cada secuencia que se pueden alinear para los propósitos de comparación. Cuando una posición en la secuencia comparada está ocupada por la misma base de nucleótido o aminoácido, entonces las moléculas son idénticas en esa posición. El grado de similitud o identidad entre secuencias de polinucleótidos es una función del número de emparejamientos idénticos de nucleótidos correspondientes en posiciones compartidas por las secuencias de polinucleótidos. El grado de identidad de una secuencia de polipéptido es una función del número de aminoácidos idénticos en las posiciones correspondientes compartidas por las secuencias de polipéptidos. El grado de homología o similitud de secuencias de polipéptidos es una función del número de aminoácidos en las posiciones correspondientes compartidas por las secuencias de polipéptidos.

[0037] Con respecto a los polipéptidos, las expresiones “identidad sustancial” o “sustancialmente idénticos” puede referirse a secuencias de suficiente similitud y estructura con los factores de transcripción de la Lista de Secuencias para producir una función similar cuando son expresadas o expresadas en exceso en una planta; en la

presente invención esta función es la detección del C/N alterado o una mayor tolerancia a la sequía o la sombra. Las secuencias que son al menos aproximadamente 50% idénticas, y preferiblemente al menos 82% idénticas con las presentes secuencias de polipéptidos se considera que tienen una “identidad sustancial” con estas últimas. Las secuencias que tienen grados de identidad menores pero actividad biológica comparable se considera que son equivalentes. La estructura necesaria para mantener la funcionalidad adecuada está relacionada con la estructura terciaria del polipéptido. Hay dominios y patrones discretos dentro de un factor de transcripción que deben estar presentes en el polipéptido para conferir la función y especificidad. Estas estructuras específicas son necesarias de modo que las secuencias interactivas se orientarán de forma adecuada para retener la actividad deseada. Por lo tanto, la “identidad sustancial” también se puede usar con respecto a subsecuencias, por ejemplo, patrones, que tienen una estructura y similitud suficientes, que son al menos aproximadamente 50% idénticas, y preferiblemente al menos 82% idénticas, con patrones similares en otras secuencias relacionadas, de modo que cada una confiere o es necesaria para la detección del C/N alterado o la mayor tolerancia a la sequía o la sombra.

[0038] La expresión “patrón consenso de aminoácidos” se refiere a la parte o a la subsecuencia de una secuencia de polipéptido que está sustancialmente conservada entre los factores de transcripción polipeptídicos listados en la Lista de Secuencias.

[0039] “Alineamiento” se refiere a una serie de secuencias de nucleótidos o restos de aminoácidos alineadas para la comparación longitudinalmente de modo que los componentes comunes (es decir bases de nucleótidos o restos de aminoácidos) se pueden identificar de forma visual y fácil. La fracción o porcentaje de componentes en común está relacionada con la homología o identidad entre las secuencias. Los alineamientos como los encontrados en las figuras se pueden usar para identificar dominios conservados y relacionados dentro de estos dominios. Un alineamiento se puede determinar de forma adecuada mediante programas de ordenador conocidos en la técnica, tales como MacVector (1999) (Accelrys, Inc., San Diego, CA).

[0040] Un “dominio conservado” o “región conservada” como se usa en el presente documento, se refiere a una región en secuencias de polinucleótidos o polipéptidos heterólogas en la que hay un grado de identidad de secuencia relativamente alto entre las distintas secuencias. Los dominios AP2 son ejemplos de dominios conservados.

[0041] Con respecto a los polinucleótidos que codifican los presentes factores de transcripción descritos, un dominio conservado preferiblemente tiene al menos 10 pares de bases (pb) de longitud.

[0042] Un “dominio conservado”, con respecto a los presentes polipéptidos descritos se refiere a un dominio dentro de un familia de los factores de transcripción que presenta un grado de homología de secuencia superior, tal como al menos 70% de similitud de secuencia, incluyendo sustituciones conservativas, y más preferiblemente al menos 79% de identidad de secuencia, e incluso más preferiblemente al menos 81%, o al menos aproximadamente 86%, o al menos aproximadamente 89%, o al menos aproximadamente 91%, o al menos aproximadamente 95%, o al menos aproximadamente 98% de identidad de secuencia de los restos de aminoácidos respecto al dominio conservado. La invención también abarca las secuencias que tienen o codifican dominios conservados que están de forma reconocible dentro de un clado dado de polipéptidos de factores de transcripción y que tienen una actividad biológica comparable con las secuencias de esta invención. Un fragmento o dominio puede denominarse fuera de un dominio conservado, fuera de una secuencia consenso o fuera de un sitio de unión al ADN consenso, cuando se sabe que existen o que existe para una clase, familia o su familia de factores de transcripción particulares. En este caso, el fragmento o dominio no incluirá los aminoácidos exactos de una secuencia consenso o sitio de unión al ADN consenso de una clase, familia o subfamilia de factores de transcripción, o los aminoácidos exactos de una secuencia consenso o sitio de unión al ADN consenso de factor de transcripción particular. 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 están fuera de un dominio o dominios conservados definidos para un polipéptido o proteína. Las secuencias que tienen un grado de identidad menor pero actividad biológica comparable se considera que son equivalentes.

[0043] Como reconoce un experto en la materia, los dominios conservados se pueden identificar como regiones o dominios de identidad respecto a una secuencia consenso específica (por ejemplo, Riechmann y col. (2000) véase antes). Por lo tanto, usando los procedimientos de alineamiento conocidos en la técnica, se pueden determinar los dominios conservados de los factores de transcripción de plantas AP2.

[0044] Los dominios conservados para una serie de las secuencias que confieren tolerancia a la sequía y detección del C/N alterado, se encuentran en las tablas 1 y 3, respectivamente. Una comparación de las regiones de

los polipéptidos en las tablas 1 o 3 permite al experto en la materia identificar dominios conservados para cualquiera de los polipéptidos listados o a los que se hace referencia en esta descripción.

[0045] “Complementaria” se refiere a los enlaces de hidrógeno naturales de emparejamiento de bases entre purinas y pirimidinas. Por ejemplo, la secuencia A-C-G-T (5'→3') forma enlaces de hidrógeno con sus complementarias A-C-G-T (5' → 3') o A-C-G-U (5'→3'). Dos moléculas de monocatenarias se pueden considerar parcialmente complementarias si forman enlace sólo algunos de los nucleótidos, o “completamente complementarias” si forman enlace todos los nucleótidos. El grado de complementariedad entre cadenas de ácido nucleico afecta a la eficacia y fuerza de las reacciones de hibridación y amplificación. “Totalmente complementarias” se refiere al caso en el que la formación de enlaces se produce entre cada uno de los pares de bases y sus complementarios en una pareja de secuencias, y las dos secuencias tienen el mismo número de nucleótidos.

[0046] Las expresiones “muy riguroso” o “condiciones muy rigurosas” se refiere a condiciones que permiten la hibridación de cadenas de ADN cuyas secuencias son muy complementarias, en las que las mismas condiciones excluyen la hibridación de ADN con emparejamientos erróneos significativos. Las secuencias de polinucleótidos capaces de hibridar en condiciones rigurosas con los polinucleótidos la presente invención pueden ser, por ejemplo, variantes de las secuencias de polinucleótidos descritas, incluyendo variantes alélicas o de corte y empalme, o secuencias que codifican ortólogos o parálogos de los presentes polipéptidos descritos. Los procedimientos de hibridación de ácidos nucleicos se describen en detalle en Kashima y col. (1985) *Nature* 313:402-404, Sambrook y col. (1989) “Molecular Cloning: A Laboratory Manual”. 2ª Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (“Sambrook”), y en Hames y Higgins, “Nucleic Acid Hybridisation: A Practical Approach”, IRL Press, Washington, D.C. (1985), cuyas referencias se incorporan en el presente documento por referencia.

[0047] En general, las condiciones rigurosas se determinan por la temperatura, fuerza iónica y concentración de los agentes desnaturizantes (por ejemplo, formamida) usados en una hibridación y el procedimiento de lavado (para una descripción más detallada para establecer y determinar las condiciones rigurosas, véase a continuación). El grado con que hibridan dos ácidos nucleicos en diferentes condiciones rigurosas esta correlacionado con la extensión de su similitud. Por lo tanto, las secuencias de ácidos nucleicos similares de una variedad de fuentes, tal como en el genoma de una planta (como es el caso de parálogos) o de otra planta (como es el caso de ortólogos) que pueden realizar funciones similares, se pueden aislar basándose en su capacidad de hibridar con secuencias de factores de transcripción conocidas. Son posibles numerosas variaciones en las condiciones y medios con los que se puede realizar la hibridación de ácidos nucleicos para aislar secuencias de factores de transcripción que tienen similitud con las secuencias de factores de transcripción conocidas en la técnica, y no están limitados a los que se describen explícitamente en el presente documento. Dicho procedimiento se puede usar para aislar secuencias de polinucleótidos que tienen diferentes grados de similitud con las secuencias de factores de transcripción descritos, tal como por ejemplo factores de transcripción que tienen 60% de identidad, o más preferiblemente más de aproximadamente 70% de identidad, lo más preferiblemente 72% de identidad o identidad mayor con los factores de transcripción descritos.

[0048] En relación con los términos “parálogo” y “ortólogo”, las secuencias de polinucleótidos homólogas y las secuencias de polipéptidos homólogas pueden ser parálogas u ortólogas de la secuencia de polipéptido o polinucleótido reivindicada. Los ortólogos y parálogos son genes evolutivamente relacionados que tienen secuencia similar y funciones similares. Los ortólogos son genes estructuralmente relacionados en especies diferentes que han derivado de un suceso de especiación. Los parálogos son genes estructuralmente relacionados dentro de una sola especie que han derivado de un suceso de duplicación. Los expertos en la materia apreciarán las secuencias que son suficientemente similares entre sí y pueden basarse en el porcentaje de identidad de las secuencias completas, porcentaje de identidad de un dominio o secuencia conservados dentro de una secuencia completa, porcentaje de similitud con la secuencia completa, porcentaje de similitud con un dominio o secuencia conservados dentro de la secuencia completa, y/o una disposición de nucleótidos contiguos o péptidos particulares respecto a un dominio conservado o secuencia completa. Las secuencias que son suficientemente similares entre sí también se unirán de una forma similar a los mismos sitios de unión al ADN de los elementos reguladores de la transcripción usando procedimientos conocidos por el experto en la materia.

[0049] El término “equivalogo” describe miembros de un conjunto de proteínas homólogas que son conservadas con respecto a su función desde su último ancestro común. Las proteínas relacionadas se agrupan en familias de equivalogos, y por otra parte en familias de proteínas con otros tipos de homología jerárquicamente definidas. Esta definición se proporciona en el sitio web (www) del Institute for Genomic Research (TIGR), “tigr.org” bajo el título “Terms associated with TIGRFAMs”. Institute for Genomic Research (TIGR) página web (www), “tigr.org” bajo el título “Terms associated with TIGRFAMs”.

[0050] El término “variante”, como se usa en el presente documento, se puede referir a polinucleótidos o polipéptidos que difieren entre sí en la secuencia de los polinucleótidos o polipéptidos descritos en el presente documento, respectivamente, y se exponen más adelante.

5

[0051] Con respecto a las variantes de polinucleótidos, las diferencias entre los polinucleótidos descritos en el presente documento y las variantes de polinucleótidos están limitadas de modo que las secuencias de nucleótidos de los primeros y de estos últimos son en general muy similares y, en muchas regiones, idénticos. Debido a la degeneración del código genético, la diferencia entre las primeras y estas últimas secuencias de nucleótidos puede ser silenciosa (es decir, los aminoácidos codificados por el polinucleótido son los mismos, y la secuencia de polinucleótidos variante codifica la misma secuencia de aminoácidos que el presente polinucleótido descrito. Las secuencias de nucleótidos variantes pueden codificar secuencias de aminoácidos diferentes, en cuyo caso dichas diferencias de nucleótidos darán como resultado sustituciones, adiciones, eliminaciones, inserciones, truncados o fusiones de aminoácidos, con respecto a las secuencias de polinucleótidos descritas. Estas variaciones pueden dar como resultado variantes de polinucleótido que codifica 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 la Lista de Secuencias.

10

15

20

[0052] Las secuencias de polipéptidos descritas en el presente documento y variantes de polipéptidos similares pueden diferir en la secuencia de aminoácidos en una o más sustituciones, adiciones, eliminaciones, fusiones y truncados, 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, el experto en la materia apreciará fácilmente que cualquiera de una variedad de secuencias de polinucleótidos es capaz de codificar factores de transcripción y polipéptidos homólogos de los factores de transcripción de la invención. Una variante de la secuencia de polipéptido puede tener cambios “conservativos”, en los que un aminoácido sustituido tiene propiedades estructurales o químicas similares. Se pueden hacer sustituciones de aminoácidos deliberadas basándose en la similitud de polaridad, carga, solubilidad, hidrofobia, hidrofilia, y/o naturaleza anfipática de los restos, con la condición de que se retenga una cantidad sustancial 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 polares no cargados 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 de las sustituciones conservativas, véase la tabla 6). De forma más rara, una variante puede tener cambios “no conservativos”, por ejemplo, sustitución de una glicina por un triptófano. Las variaciones minoritarias similares también pueden incluir eliminaciones o inserciones de aminoácidos, o ambas. Los polipéptidos relacionados pueden comprender, por ejemplo, adiciones y/o eliminaciones de uno o más sitios de glicosilación unidos a N o unidos a O, o una adición y/o eliminación de uno o más restos de cisteína. Se puede encontrar orientación para determinar cuáles y cómo muchos restos de aminoácidos se pueden sustituir, insertar o eliminar sin abolir la actividad funcional o biológica, usando programas de ordenador conocidos por el experto en la materia, por ejemplo, el software DNASTAR (USPN 5.840.544).

25

30

35

40

[0053] También está dentro del alcance de la invención una variante de un ácido nucleico de factor de transcripción listado en la Lista de Secuencias, es decir, uno que tiene una secuencia que difiere de una de las secuencias de polinucleótidos de la Lista 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 la secuencia respecto a la secuencia en la Lista de Secuencias, debido a la degeneración en el código genético. Están incluidos en esta definición los polimorfismos que se pueden detectar fácilmente o no usando una sonda de oligonucleótido particular del polinucleótido que codifica el polipéptido, y la hibridación inadecuada o inesperada con variantes alélicas, con un locus distinto del locus cromosómico normal para la secuencia de polinucleótido que codifica el polipéptido.

45

50

[0054] “Variante alélica” o “variante alélica de polinucleótido” 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 por mutación, y puede dar como resultado el polimorfismo fenotípico en las poblaciones. Las mutaciones de genes pueden ser “silenciosas” o pueden codificar polipéptidos que tienen la secuencia de aminoácidos alterada. “Variante alélica” y “variante alélica de polipéptido” también se pueden usar con respecto a los polipéptidos, y en este caso la expresión se refiere a un polipéptido codificado por una variante alélica de un gen.

55

- [0055]** “Variante de corte y empalme” o “variante de corte y empalme de polinucleótido”, como se usa en el presente documento, se refiere a formas alternativas de ARN transcrito a partir de un gen. La variación por corte y empalme se produce de forma natural como resultado de sitios alternativos que son cortados y empalmados dentro de una sola molécula de ARN transcrita o entre moléculas de ARN transcritas por separado, y puede dar como resultado diferentes formas 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 de polipéptido” también se puede referir a un polipéptido codificado por una variante de corte y empalme de un ARNm transcrito.
- 5
- [0056]** También como se usa en el presente documento, las “variantes de polinucleótido” también se pueden referir a secuencias de polinucleótidos que codifican parálogos y ortólogos de las presentes secuencias de polipéptidos descritas. Las “variantes de polipéptido” también se pueden referir a secuencias de polipéptidos que son parálogos y ortólogas de las presentes secuencias de polipéptidos descritas.
- 10
- [0057]** “Ligando” se refiere a cualquier molécula, agente o compuesto que se unirá específicamente a un sitio complementario en una molécula de ácido nucleico o proteína. Dichos ligandos estabilizan o modulan la actividad de las moléculas de ácido nucleico o proteínas de la invención y pueden estar compuestas por al menos uno de los siguientes: sustancias inorgánicas y orgánicas incluyendo ácidos nucleicos, proteínas, hidratos de carbono, grasas y lípidos.
- 15
- [0058]** “Modula” se refiere a un cambio en actividad (biológica, química o inmunológica) o vida útil que resulta de la unión específica entre una molécula y una molécula de ácido nucleico o una proteína.
- 20
- [0059]** El término “planta” incluye las plantas enteras, órganos/estructuras vegetativas que brotan (por ejemplo, hojas, tallos y tubérculos), raíces, flores y órganos/estructuras florales (por ejemplo, brácteas, sépalos, pétalos, estambres, carpelos, anteras y óvulos), semillas (incluyendo embriones, endospermo y recubrimiento de semilla) y fruto (el ovario maduro), tejido de la planta (por ejemplo, tejido vascular, tejido del suelo y similares) y células (por ejemplo, células guarda, células huevo y similares), y descendientes de los mismos. La clase de plantas que se puede usar en el procedimiento de la invención en general es tan amplio como la clase de plantas superiores e inferiores susceptibles de técnicas de transformación, incluyendo angiospermas (plantas monocotiledóneas y dicotiledóneas); gimnospermas, helechos, belchos, psilofitos, licofitos, briofitos y algas multicelulares (como se muestra, por ejemplo, en la figura 1, adaptada de Daly y col. (2001) *Plant Physiol.* 127: 1328-1333, y en la Figura 2, adaptada de Ku y col. (2000) *Proc. Natl. Acad. Sci.* 97: 9121-9126; y en Tudge (2000) en “The Variety of Life”, Oxford University Press, Nueva York, NY, pág. 547-606).
- 25
- 30
- [0060]** Una “planta transgénica” se refiere a una planta que contiene material genético que no se encuentra en una planta silvestre de la misma especie, variedad o variedad cultivada. El material genético puede incluir un transgén, un suceso de mutagénesis por inserción (tal como mutagénesis por inserción de transposón o ADN-T), una secuencia marcadora de activación, una secuencia mutada, un suceso de recombinación homóloga o una secuencia modificada por quimeroplastia. Típicamente, el material genético extraño se ha introducido en la planta mediante manipulación humana, pero se puede usar cualquier procedimiento como reconocerá el experto en la materia.
- 35
- 40
- [0061]** Una planta transgénica puede contener un vector o casete de expresión. El casete de expresión típicamente comprende una secuencia que codifica un polipéptido operativamente unida (es decir, bajo el control regulador de) a secuencias reguladoras inducibles o constitutivas adecuadas que permiten la expresión del polipéptido. El casete de expresión se puede introducir en una planta mediante transformación o mediante cultivo después de transformación de una planta parental. Una planta se refiere a una planta entera así como a una parte de la planta, tal como la semilla, fruto, hoja o raíz, tejido de la planta, células de la planta o cualquier otro material de la planta, por ejemplo, un explante de planta, así como descendientes de la misma, y a sistemas in vitro que imitan los componentes o procedimientos bioquímicos o celulares en una célula.
- 45
- 50
- [0062]** “Silvestre”, tal como se usa en el presente documento, se refiere a una célula de planta, semilla, componente de planta, tejido de planta, órgano de planta o la planta entera que no se ha modificado genéticamente ni se ha tratado en un sentido experimental. Las células, semillas, componentes, tejidos, órganos o plantas enteras silvestres se pueden usar como controles para comparar los niveles de expresión y la extensión y naturaleza de la modificación del rasgo con células, tejidos o plantas de la misma especie en las que se ha alterado la expresión de un factor de transcripción, por ejemplo, se ha inactivado, expresado en exceso, o expresado ectópicamente.
- 55

[0063] Una “planta de control”, tal como se usa la presente invención, se refiere a una célula de planta, semilla, componente de planta, tejido de planta, órgano de planta o la planta entera usada para comparar frente a la planta transgénica o genéticamente modificada, con el propósito de identificar un fenotipo potenciado en la planta transgénica o modificada genéticamente. Una planta de control en algunos casos puede ser una línea de planta transgénica que comprende un vector vacío o un gen marcador, pero no contiene el polinucleótido recombinante de la presente invención que se expresa en la planta transgénica o genéticamente modificada que se está evaluando. En general, una planta de control es una planta de la misma línea o variedad que la planta transgénica o genéticamente modificada que se está ensayando. Una planta de control adecuada incluiría una planta no transgénica o no alterada genéticamente de la línea parental usada para generar una planta transgénica en el presente documento.

[0064] “Fragmento”, con respecto a un polinucleótido, se refiere a un clon o cualquier parte de una molécula de polinucleótido que retiene una característica funcional usable. Los fragmentos útiles incluyen oligonucleótidos y polinucleótidos que se pueden usar en las tecnologías de hibridación o amplificación o en la regulación de la replicación, transcripción o traducción. Un “fragmento de polinucleótido” se refiere a cualquier subsecuencia de un polinucleótido, típicamente, de al menos aproximadamente 9 nucleótidos consecutivos, preferiblemente al menos aproximadamente 30 nucleótidos, más preferiblemente al menos aproximadamente 50 nucleótidos, de cualquiera de las secuencias proporcionadas en el presente documento. Los fragmentos de polinucleótido de ejemplo son los primeros 60 nucleótidos consecutivos de los polinucleótidos de los factores de transcripción listados en la Lista de Secuencias. Los fragmentos de ejemplo incluyen fragmentos que comprenden una región que codifica un dominio conservado (por ejemplo, un dominio AP2) de un factor de transcripción.

[0065] Los fragmentos también pueden incluir subsecuencias de polipéptidos y moléculas de proteínas, o una subsecuencia de un polipéptido. Los fragmentos pueden tener uso en cuanto que 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 una extensión similar, a la que lo hace en el polipéptido intacto. Por ejemplo, un fragmento de polipéptido puede comprender un patrón estructural reconocible o dominio funcional tal como un sitio de unión al ADN o dominio que se une a 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 pocos como 3 restos de aminoácidos hasta la longitud entera del polipéptido intacto, pero preferiblemente tienen al menos aproximadamente 30 restos de aminoácidos de longitud y más preferiblemente al menos aproximadamente 60 restos de aminoácidos de longitud. Los fragmentos de polipéptidos de ejemplo son los primeros 20 aminoácidos consecutivos de los polipéptidos de los factores de transcripción listados en la Lista de Secuencias. Los fragmentos de ejemplo también incluyen fragmentos que comprenden un dominio AP2 de un factor de transcripción, por ejemplo, los restos de aminoácidos 10-77 de G2133 (SEQ ID NO: 12), como se indica en la tabla 1.

[0066] La invención también abarca la producción de secuencias de ADN que codifican factores de transcripción y derivados de factores de transcripción, o fragmentos de los mismos, totalmente por química sintética. Después de la producción, la secuencia sintética se puede insertar en cualquiera de los muchos vectores de expresión y sistemas celulares disponibles usando reactivos conocidos en la técnica. Además, se puede usar la química sintética para introducir mutaciones en una secuencia que codifica factores de transcripción o cualquier fragmento de la misma.

[0067] “Derivado” se refiere a la modificación química de una molécula de ácido nucleico o secuencia de aminoácidos. Las modificaciones químicas pueden incluir la sustitución de hidrógeno o un grupo alquilo, acilo o amino o la glicosilación, pegilación, o cualquier procedimiento similar que retenga o potencie la actividad biológica o la vida útil de la molécula o secuencia.

[0068] Un “rasgo” se refiere a una característica fisiológica, morfológica, bioquímica o física de una planta o un material o célula de planta particular. En algunos casos, esta característica es visible a simple vista, tal como el tamaño de la semilla o planta, o se puede medir por técnicas bioquímicas, tales como detectando el contenido de proteína, almidón o aceite de semillas u hojas, o por observación de un proceso metabólico o fisiológico, por ejemplo, midiendo la tolerancia a la falta de agua o a concentraciones de sales o azúcar particulares, o por la observación del nivel de expresión de un gen o genes, por ejemplo usando análisis Northern, RT-PCR, ensayos de expresión de genes de micromatrices, o sistemas de expresión de genes indicadores, o por observaciones agrícolas tales como la tolerancia al estrés por sequía o el rendimiento. Sin embargo, se puede usar cualquier técnica para medir la cantidad, el nivel comparativo o la diferencia en cualquier compuesto o macromolécula química seleccionada en las plantas transgénicas.

[0069] La “modificación de rasgo” se refiere a una diferencia detectable en una característica en una planta que expresa ectópicamente un polinucleótido o polipéptido de la presente invención con respecto a una planta que no lo hace, tal como una planta silvestre. En algunos casos, la modificación del rasgo se puede evaluar de forma cuantitativa. Por ejemplo, la modificación del rasgo puede conllevar un aumento o disminución de al menos 5 aproximadamente 2% en un rasgo observado, o una diferencia incluso mayor, comparado con una planta silvestre o de control. Se sabe que puede haber una variación natural en el rasgo modificado. Por lo tanto, la modificación del rasgo observado conlleva un cambio de la distribución y magnitud normales del rasgo de las plantas comparado con la distribución y magnitud observadas en las plantas silvestres.

[0070] La expresión “perfil del transcrito” se refiere a los niveles de expresión de un conjunto de genes en una célula en un estado particular, en particular por comparación con los niveles de expresión del mismo conjunto de genes en una célula del mismo tipo en un estado de referencia. Por ejemplo, el perfil de transcritos de un factor de transcripción particular en una célula en suspensión son los niveles de expresión de un conjunto de genes en una 15 célula que reprimen o expresan en exceso este factor de transcripción comparado con los niveles de expresión del mismo conjunto de genes en una célula en suspensión que tiene los niveles normales de este factor de transcripción. El perfil de transcritos se puede presentar como una lista de aquellos genes cuyo nivel de expresión es significativamente diferente entre los dos tratamientos, y las proporciones de la diferencia. Las diferencias y similitudes entre los niveles de expresión también se pueden evaluar y calcular usando procedimientos estadísticos 20 y de agrupamiento.

[0071] La “expresión ectópica o expresión alterada” en referencia a un polinucleótido indica que el patrón de expresión, por ejemplo, en una planta transgénica o tejido de planta, es diferente del patrón de expresión en una planta silvestre o una planta de referencia de la misma especie. El patrón de expresión también se puede comparar 25 con un patrón de expresión de referencia en una planta silvestre de la misma especie. Por ejemplo, el polinucleótido o polipéptido se expresa en una célula o tipo de tejido distintos de una célula o tipo de tejido en el que la secuencia se expresa en la planta silvestre, o por la expresión en un momento distinto al momento en el que la secuencia se expresa en la planta silvestre, o por una respuesta a diferentes agentes inducibles, tales como hormonas o señales del entorno, o con niveles de expresión diferentes (sean mayores o menores) comparados con los encontrados en 30 una planta silvestre. El término también se refiere a patrones de expresión alterados que se producen disminuyendo los niveles de expresión por debajo del nivel de detección o por eliminación completa de la expresión. El patrón de expresión resultante puede ser transitorio o estable, constitutivo o inducible. Con referencia a un polipéptido, la expresión “expresión ectópica o expresión alterada” se puede referir además a los niveles de actividad alterados que resultan de las interacciones de los polipéptidos con moduladores exógenos o endógenos o de interacciones con 35 factores o como resultado de la modificación química de los polipéptidos.

[0072] La expresión “exceso de expresión”, tal como se usa en el presente documento, se refiere a un nivel de expresión mayor de un gen en una planta, célula de planta o tejido de planta, comparado con la expresión en una planta, célula o tejido silvestre, en cualquier fase de desarrollo o temporal para el gen. El exceso de expresión se 40 produce cuando, por ejemplo, los genes que codifican uno o más factores de transcripción están bajo el control de un promotor fuerte descrito en el presente documento (por ejemplo, la región de inicio de la transcripción 35S del virus del mosaico de la coliflor), o el exceso de expresión puede ser inducido cuando está presente una señal adecuada en el entorno. El exceso de expresión se puede producir en toda la planta o en tejidos específicos de la planta, dependiendo del promotor usado, como se describe más adelante.

[0073] El exceso de expresión puede tener lugar en células de plantas que normalmente carecen de la expresión de polipéptidos opcionalmente equivalentes o idénticos a los presentes factores de transcripción. El exceso de expresión también se puede producir en células de plantas en las que normalmente se produce la expresión endógena de los presentes factores de transcripción o moléculas funcionalmente equivalentes, pero dicha 50 expresión normal es en un nivel inferior. Por lo tanto, el exceso de expresión da como resultado una producción mayor que la normal o “exceso de producción” del factor de transcripción en la planta, célula o tejido.

[0074] 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 55 dominios de unión específicos se une a la secuencia reguladora de ADN. Los factores de transcripción de la presente invención pueden tener, por ejemplo, un dominio AP2, en cuyo caso el dominio AP2 del factor de transcripción se une a una región reguladora de la transcripción, tal como AtERF1, que se une al patrón AGCCGCC (la “caja GCC”) que está presente en promotores de genes tales como PDF1.2. Los factores de transcripción de la invención 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 a la región reguladora.

[0075] Una “muestra” con respecto a un material que contiene moléculas de ácido nucleico, puede comprender un fluido corporal; un extracto de una célula, cromosoma, orgánulo o membrana aislados de una célula; ADN, ARN o ADNc genómicos en disolución o unidos a un sustrato; una célula; un tejido; una huella tisular; una muestra forense; y similares. En este contexto “sustrato” se refiere a cualquier soporte rígido o semirígido al que se unen moléculas de ácido nucleico o proteínas, que incluyen membranas, filtros, chips, portaobjetos, obleas, fibras, perlas magnéticas o nanomagnéticas, geles, capilares u otros tubos, placas, polímeros y micropartículas, con una variedad de formas de la superficie incluyendo pocillos, zanjias, agujas, canales y poros. Un sustrato también se puede referir a un reaccionante en una reacción química o biológica, o una sustancia que actúa sobre ella (por ejemplo, una enzima).

Los factores de transcripción modifican la expresión de genes endógenos

[0076] Un factor de transcripción puede incluir, pero sin limitar, cualquier polipéptido que pueda activar o reprimir la transcripción de un solo gen o de una serie de genes. Como reconoce el experto en la materia, los factores de transcripción se pueden identificar por la presencia de una región o dominio de similitud o identidad estructural con una secuencia consenso específica o la presencia de un sitio de unión al ADN consenso específico o patrón de sitio de unión al ADN (por ejemplo, en Riechmann y col. (2000) *Science* 290: 2105-2110).

[0077] En general, los factores de transcripción codificados por las presentes secuencias están implicados en la diferenciación y proliferación celular y la regulación del crecimiento. Por consiguiente, un experto en la materia reconocerá que mediante la expresión de 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 tienen 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 en relación con el estrés por sequía, la tolerancia a la sombra, o la detección del C/N. Las secuencias de la invención también se pueden usar para transformar una planta e introducir rasgos deseables que nos encuentran en la variedad cultivada o cepa silvestre. Después, se pueden seleccionar aquellas plantas que producen el grado más deseado de exceso o defecto de expresión de los genes diana de interés y la mejora simultánea del rasgo.

[0078] Las secuencias de la presente invención pueden ser de cualquier especie, en particular especies de plantas, en forma natural o de cualquier fuente sea natural, sintética, semisintética o recombinante. Las secuencias de la invención también pueden incluir fragmentos de las presentes secuencias de aminoácidos. Cuando se cita la “secuencia de aminoácidos” para referirse a una secuencia de aminoácidos de una molécula de proteína natural, no se pretende que la “secuencia de aminoácidos” y expresiones similares limiten la secuencia de aminoácidos a la secuencia de aminoácidos nativa completa asociada con la molécula de proteína citada.

[0079] Además de los procedimientos para modificar el fenotipo de una planta usando uno o más polinucleótidos y polipéptidos de la invención descritos en el presente documento, los polinucleótidos y polipéptidos de la invención tienen una variedad de usos adicionales. Estos usos incluyen su uso en la producción recombinante (es decir, expresión) de proteínas; como reguladores de la expresión de genes de la planta, como sondas de diagnóstico para la presencia de ácidos nucleicos complementarios o parcialmente complementarios (incluyendo para la detección de ácidos nucleicos codificantes naturales); como sustratos para otras reacciones, por ejemplo, reacciones de mutación, reacciones de PCR, o similares; como sustratos para la clonación por ejemplo, incluyendo reacciones de digestión o ligado; y para la identificación de moduladores exógenos o endógenos de los factores de transcripción. En muchos casos, un polinucleótido comprende una secuencia de nucleótidos que codifica un polipéptido (o proteína) o un dominio o fragmento del mismo. Además, el polinucleótido puede comprender un promotor, un intrón, una región potenciadora, un sitio de poliadenilación, un sitio de inicio de la traducción, regiones 5' o 3' no traducidas, un gen indicador, un marcador seleccionable, o similares. El polinucleótido puede ser ADN o ARN de una cadena o dos cadenas. El polinucleótido comprende opcionalmente bases modificadas o una cadena principal modificada. El polinucleótido puede ser, por ejemplo, ADN o ARN genómicos, un transcrito (tal como un ARNm), un ADNc, un producto de la PCR, un ADN clonado, un ADN o ARN sintéticos, o similares. El polinucleótido puede comprender una secuencia en las orientaciones de sentido directo o sentido contrario.

[0080] La expresión de genes que codifican factores de transcripción que modifican la expresión de genes, polinucleótidos y proteínas endógenos, es bien conocida en la materia. Además, las plantas transgénicas que comprenden polinucleótidos aislados que codifican factores de transcripción también pueden modificar la expresión

de genes, polinucleótidos y proteínas endógenos. Los ejemplos incluyen Peng y col. (1997) *Genes Development* 11: 3194-3205, y Peng y col. (1999) *Nature*, 400: 256-261). Además, muchos otros han demostrado que un factor de transcripción de *Arabidopsis* expresado en una especie de planta exógena produce la misma respuesta fenotípica o muy similar (por ejemplo, en Fu y col. (2001) *Plant Cell* 13: 1791-1802; Nandi y col. (2000) *Curr. Biol* 10: 215-218; 5 Coupland (1995) *Nature* 377:482483; y Weigel y Nilsson (1995) *Nature* 377: 482-500).

[0081] En otro ejemplo, Mandel y col. (1992) *Cell* 71-133-143, y Suzuki y col. (2001) *Plant J.* 28: 409-418 enseñan que un factor de transcripción expresado en otra especie de planta produce la misma respuesta fenotípica o muy similar a la de la secuencia endógena, como se ha predicho con frecuencia en estudios previos de factores de transcripción de *Arabidopsis*, en *Arabidopsis* (Mandel y col. (1992) véase antes; y Suzuki y col. (2001) véase antes). 10 Otros ejemplos incluyen Muller y col. (2001) *Plant J.* 28: 169-179; Kim y col. (2001) *Plant J.* 25: 247-259; Kyoizuka y Shimamoto (2002) *Plant Cell Physiol.* 43: 130-135; Boss y Thomas (2002) *Nature*, 416: 847-850; He y col. (2000) *Transgenic Res.* 9: 223-227; y Robson y col. (2001) *Plant J.* 28: 619-631.

[0082] En otro ejemplo más, Gilmour y col. (1998) *Plant J.* 16: 433-442, enseñan un factor de transcripción AP2 de *Arabidopsis*, CBF1, que cuando se expresa en exceso en plantas transgénicas aumenta la tolerancia a las heladas. Jaglo y col. (2001) *Plant Physiol.* 127: 910-917, identificaron además secuencias en *Brassica napus* que codifican genes de tipo CBF y que los transcritos para estos genes se acumulaban rápidamente en respuesta a la temperatura baja. También se encontró que los transcritos que codifican las proteínas CBF se acumulaban 20 rápidamente en respuesta a la temperatura baja en el trigo, así como en el tomate. El alineamiento de las proteínas CBF de *Arabidopsis*, *B. napus*, trigo, centeno y tomate pusieron de manifiesto la presencia de restos de aminoácidos consecutivos conservados, PKK/RPAGRxKFXETRHP y DSAWR, que encierran los dominios de unión al ADN AP2/ERE BP de las proteínas y las distingue de otros miembros de la familia de proteínas AP2/ERE BP (Jaglo y col. (2001) véase antes).

[0083] Los factores de transcripción median respuestas celulares y controlan los rasgos a través de la expresión alterada de genes que contienen secuencias de nucleótidos que actúan en cis, que son dianas del factor de transcripción introducido. Se apreciará en la técnica que el efecto de un factor de transcripción en las respuestas celulares o un rasgo celular está determinado por los genes particulares cuya expresión se altera de forma directa o 30 indirecta (por ejemplo, por una cascada de sucesos de unión de factores de transcripción y cambios transcripcionales) por la unión del factor de transcripción. En un análisis global de la transcripción comparando un estado estándar con uno en el que se expresa en exceso un factor de transcripción, el perfil de transcritos resultante asociado con el exceso de expresión 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 35 MYB) controla la biosíntesis de la antocianina a través de la regulación de la expresión de genes que se sabe que están implicados en la ruta biosintética de la antocianina (Bruce y col. (2000) *Plant Cell*, 12: 65-79; Borevitz y col. (2000) *Plant Cell* 12: 2383-93). Además, los perfiles de transcritos globales se han usado con éxito como herramientas de diagnóstico para estados celulares específicos (por ejemplo, cancerosos frente a no cancerosos; Bhattacharjee y col. (2001) *Proc. Natl. Acad. Sci., USA*, 98: 13790-13795; Xu y col. (2001) *Proc. Natl. Acad. Sci., 40 USA*, 98: 15089.15094). Por consiguiente, es evidente para el experto en la materia que la similitud del perfil de transcritos tras el exceso de expresión de diferentes factores de transcripción indicaría similitud de la función del factor de transcripción.

Polipéptidos y polinucleótidos de la solicitud

[0084] La presente invención proporciona, entre otras cosas, factores de transcripción (TF), y polipéptidos homólogos de factores de transcripción, y polinucleótidos aislados o recombinantes que codifican los polipéptidos o nuevos polipéptidos variantes de secuencia o polinucleótidos que codifican nuevas variantes de factores de transcripción derivados de las secuencias específicas proporcionadas en el presente documento. 45

[0085] Los polinucleótidos de la invención pueden ser expresados ectópicamente en células de plantas que producen exceso de expresión y observarse los cambios en los niveles de expresión de una serie de genes, polinucleótidos y/o proteínas de las células de la planta. Por lo tanto, los polinucleótidos y polipéptidos se pueden usar para cambiar los niveles de expresión de genes, polinucleótidos y/o proteínas de plantas o células de plantas. 55 Estos polipéptidos y polinucleótidos se pueden usar para modificar las características de una planta, en particular la tolerancia a la sequía. Los polinucleótidos de la invención se pueden expresar ectópicamente en plantas que producen un exceso de expresión o se encuentran inactivadas, y observar los cambios en la o las características, o el o los rasgos de las plantas. Por lo tanto, los polinucleótidos y polipéptidos se pueden usar para mejorar las características de las plantas. Se ha mostrado que las secuencias de polipéptidos de la Lista de Secuencia confieren

una mayor tolerancia a la sequía o a la sombra o la detección del C/N alterado, cuando estos polipéptidos son expresados en exceso en plantas de *Arabidopsis*. Se ha mostrado que estos polinucleótidos tienen una asociación fuerte con estos rasgos, en cuanto que las plantas que expresan en exceso estas secuencias son más tolerantes a la sequía, la sombra o tienen detección del C/N alterado, respectivamente. La invención también abarca un complemento de los polinucleótidos. Los polinucleótidos también son útiles para el cribado de bibliotecas de moléculas o compuestos para la unión específica y para crear plantas transgénicas que tienen rasgos mejorados. Se ha mostrado y se espera que la alteración de los niveles de expresión de los equivalentes de estas secuencias, incluyendo parálogos y ortólogos en la Lista de Secuencias, y otros ortólogos que son estructural y secuencialmente similares a los primeros ortólogos, confiera fenotipos similares, incluyendo la detección del C/N alterado, tolerancia a la sequía y/o a la sombra en las plantas.

[0086] En algunos casos, se identificaron polinucleótidos de ejemplo que codifican los polipéptidos de la lista de secuencias en la base de datos de GenBank de *Arabidopsis thaliana* usando programas y parámetros de análisis de secuencias disponibles al público. Las secuencias identificadas inicialmente después se caracterizaron adicionalmente para identificar las secuencias que comprenden cadenas de secuencia específicas correspondientes a patrones de secuencia presentes en familias de factores de transcripción conocidos. Además, se identifican otros polinucleótidos de ejemplo que codifican los polipéptidos de la invención en la base de datos de GenBank de plantas usando programas y parámetros de análisis de secuencias disponibles al público. Las secuencias identificadas inicialmente después se caracterizaron más para identificar las secuencias que comprenden cadenas de secuencia específicas correspondientes a patrones de secuencia presentes en familias de factores de transcripción conocidos. Las secuencias de polinucleótidos que cumplen dichos criterios se confirmaron como factores de transcripción.

[0087] Se identificaron polinucleótidos adicionales de la invención mediante el cribado de bibliotecas de ADNc de *Arabidopsis thaliana* y/u otras plantas con sondas correspondientes a factores de transcripción conocidos en condiciones de hibridación rigurosas bajas. Posteriormente se recuperaron secuencias adicionales, incluyendo secuencias codificantes de longitud entera, mediante el procedimiento de amplificación rápida de los extremos del ADNc (RACE), usando un kit disponible en el comercio de acuerdo con las instrucciones del fabricante. Cuando era necesario, se llevaron a cabo varios ciclos de RACE para aislar los extremos 5' y 3'. Después, se recuperó el ADNc de longitud entera mediante una reacción en cadena de la polimerasa de extremo a extremo rutinaria (PCR) usando celadores específicos para aislar los extremos 5' y 3'. Se proporcionan secuencias de ejemplo en la Lista de Secuencias.

[0088] Los polinucleótidos son particularmente útiles cuando son elementos de la matriz que se pueden hibridar en una micromatriz. Dicha micromatriz se puede usar para el seguimiento de la expresión de genes que se expresan de forma diferente en respuesta a la luz limitada, sequía otros estreses osmóticos, o baja disponibilidad de nitrógeno. La micromatriz se puede usar en análisis de expresión de genes o genético a gran escala de un gran número de polinucleótidos; o en el diagnóstico de, por ejemplo, el estrés por sequía antes de que los síntomas fenotípicos sean evidentes. Además, se puede usar la micromatriz para investigar la respuestas celulares, tales como la proliferación celular, transformación y similares.

[0089] Cuando los polinucleótidos de la invención también se pueden usar como elementos de matriz que se pueden hibridar en una micromatriz, los elementos de la matriz se organizan de una forma ordenada de modo que cada elemento esté presente en una posición especificada en el sustrato. Debido a que los elementos de la matriz están en posiciones especificadas sobre el sustrato, los patrones y las intensidades de hibridación (que juntos crean un perfil de expresión único) se pueden interpretar en términos de niveles de expresión de genes particulares y se pueden correlacionar con un estrés, patología o tratamiento particular.

[0090] La invención también implica una composición agrónoma que comprende un polinucleótido de la invención junto con un vehículo adecuado, y un procedimiento para alterar una unidad de planta usando la composición.

[0091] A continuación se proporcionan ejemplos de polinucleótidos y polipéptidos específicos de la invención, y secuencias de equivalentes, junto con las descripciones de las familias de genes que comprenden estos polinucleótidos y polipéptidos.

[0092] A continuación se proporcionan ejemplos de polinucleótidos y polipéptidos específicos de la invención, y secuencias de equivalentes.

[0093] Se ha demostrado que las secuencias de polipéptidos de la lista de secuencias, incluyendo, por

ejemplo, las secuencias de *Arabidopsis* G2133; G1274, G922, G2999, G3086, G354, G1792, G2053, G975, G1069, G916, G1820, G2701, G47, G2854, G2789, G634, G175, G2839, G1452, G3091, G489, G303, G2992, y G682 (SEQ ID NO: 12, 6, 4, 14, 16, 228, 8, 10, 238, 240, 236, 244, 246, 2, 252, 248, 232, 224, 250, 242, 254, 230, 226, 50 y 234, respectivamente) confieren una mayor tolerancia a la sequía cuando se altera la expresión de estos polipéptidos en plantas de *Arabidopsis*. Estos polinucleótidos han mostrado que tienen una asociación fuerte con la tolerancia al estrés por sequía, en cuanto que las plantas que expresan en exceso estas secuencias son más tolerantes a la sequía. Las secuencias de ejemplo de la invención incluyen la G2133.

[0094] La presente descripción también abarca los complementos de estos polinucleótidos. Los polinucleótidos también son útiles para cribar bibliotecas de moléculas o compuestos para la unión específica y para crear plantas transgénicas que tienen tolerancia a la sequía alterada. Se espera que los equivalentes de estas secuencias, incluyendo parálogos y ortólogos en la lista de secuencias, y otros ortólogos que son estructural y secuencialmente similares a los ortólogos anteriores, confieran tolerancia a la sequía cuando se altera su expresión.

[0095] La familia AP2 incluyendo los clados G47/G2133 y G1792. AP2 (APETALA2) y EREBP (proteínas de unión a elementos de respuesta a etileno) son los miembros prototípicos de una familia de factores de transcripción única de las plantas, cuya característica que los distingue es que contienen el dominio de unión al ADN llamado AP2 (Riechmann y Meyerowitz (1998) *Biol. Chem.* 379: 633-646). El dominio AP2 fue reconocido por primera vez como un patrón repetido en la proteína AP2 de *Arabidopsis thaliana* (Jofuku y col. (1994) *Plant Cell* 6: 1211-122). Poco después, se identificaron 4 proteínas de unión al ADN del tabaco, que interactúan con una secuencia que es esencial para la respuesta de algunos promotores a la hormona etileno de la planta, y se designaron como *proteínas de unión a elementos de respuesta al etileno* (EREbps; Ohme-Takagi y col. (1995) *Plant Cell* 7: 173-182). El dominio de unión al ADN de EREBP-2 se cartografió en una región que era común a las cuatro proteínas (Ohme-Takagi y col. (1995) véase antes), y que se encontró que estaba estrechamente relacionado con el dominio de AP2 (Weigel (1995) *Plant Cell* 7: 388-389) pero que no contenía similitud de secuencia con los patrones de unión al ADN previamente conocidos.

[0096] Los genes AP2/BREBP forman una gran familia, con muchos miembros conocidos en varias especies de plantas (Okamoto y col. (1997) *Proc Natl. Acad. Sci. USA* 94: 7076-7081; Riechmann y Meyerowitz (1998) véase antes). El número de genes AP2/EREBP en el genoma de *Arabidopsis thaliana* es aproximadamente 145 (Riechmann y col. (2000) *Science* 290: 2105-2110). La clase APETALA2 se caracteriza por la presencia de 2 dominios de unión al ADN AP2, y contiene 14 genes. La AP2/ERF es la superfamilia más grande, e incluye 125 genes que están implicados en las respuestas al estrés abiótico (subgrupo DREB) y biótico (subgrupo ERF) y el subgrupo RAV incluye 6 genes que tienen todos un dominio de unión al ADN B3 además del dominio de unión al ADN AP2 (Kagaya y col. (1999) *Nucleic Acids Res.* 27: 470-478).

[0097] AP2 de *Arabidopsis* está implicado en la especificación de la identidad de sépalos y pétalos a través de su actividad como un gen homeótico que forma parte del mecanismo genético combinatorio de la determinación de la identidad de los órganos florales y también es necesario para el desarrollo normal de óvulos y semillas (Bowman y col. (1991) *Development* 112: 1-20; Jofuku y col. (1994) véase antes). ANT de *Arabidopsis* es necesario para el desarrollo del óvulo y también tiene una función en el crecimiento de los órganos florales (Elliott y col. (1996) *Plant Cell* 8: 155-168; Klucher y col. (1996) *Plant Cell* 8: 137-153). Finalmente, Gll5 del maíz regula la identidad celular epidérmica de las hojas (Moose y col. (1996) *Genes Dev.* 10: 3018-3027).

[0098] El ataque de un patógeno a una planta puede inducir respuestas de defensa que conducen a la resistencia a la invasión, y estas respuestas están asociadas con la activación de la transcripción de genes relacionados con la defensa, entre ellos los que codifican las proteínas relacionadas con la patogénesis (PR). La implicación de los genes de tipo EREBP en el control de la respuesta de defensa de la planta se basa en la observación de que muchos promotores de genes de PR contiene un elemento de actuación en cis corto que media sus respuestas al etileno (parece que en etileno es una de varias moléculas señal que controlan la activación de las respuestas de defensa). Se ha mostrado que las proteínas EREBP-1, -2; -3, y -4 del tabaco, y las proteínas Pti4, Pti5 y Pti6 del tomate reconocen dichos elementos de actuación en cis (Ohme-Takagi (1995) véase antes; Zhou y col. (1997) *EMBO J.* 16: 3207-3218). Además, se ha mostrado que las proteínas Pti4, Pti5, y Pti6 interactúan directamente con Pto, una proteína quinasa que confiere resistencia frente a *Pseudomonas syringae* en el tomate (Zhou y col. (1997) véase antes). Las plantas también son estimuladas por diferentes condiciones ambientales adversas como el frío o la sequía, y parece que las proteínas de tipo EREBP también están implicadas en las respuestas a estos estrés abióticos. La expresión del gen de COR (regulado por el frío) es inducida durante la aclimatación al frío, procedimiento por el cual las plantas aumentan su resistencia a la helada en respuesta a temperaturas bajas descongelantes. El gen CBF1 de tipo EREBP de *Arabidopsis* (Stockinger y col. (1997) *Proc. Natl.*

Acad. Sci. USA 94: 1035-1040) es un regulador de la respuesta de aclimatación al frío, debido a que la expresión ectópica de CBF1 en plantas transgénicas de *Arabidopsis* inducía la expresión del gen COR en ausencia del estímulo del frío, y aumentaba la tolerancia de la planta a la helada (Jaglo-Ottosen y col. (1998) *Science* 280: 104-106). Finalmente, otro gen de tipo EREBP de *Arabidopsis*, ABI4, está implicado en la transducción de la señal de ABA, porque los mutantes *abi4* son insensibles al ABA (ABA es una hormona de planta que regula muchos aspectos importantes desde el punto de vista agronómico del desarrollo de la planta; Finkelstein y col. (1998) *Plant Cell* 10: 1043-1054).

[0099] De las secuencias examinadas hasta la fecha, se encontraron dos restos de valina que están presentes en miembros del clado G47 pero no fuera del clado (indicado mediante flechas en la figura 3). Todos los miembros del clado examinado hasta ahora tienen la subsecuencia:

V-(X)₁₇-A-A-V-A-H-D-X-A, en la que X es un aminoácido y los restos identificados se indican mediante los restos mostrados en las cajas en la figura 3.

[0100] La tabla 1 muestra secuencias de la memoria descriptiva que, en muchos casos, confieren tolerancia a la sequía cuando se expresan en exceso. Los polipéptidos se identifican por la SEQ ID NO del polipéptido e identificador (por ejemplo, ID de gen (GID) N°, número de acceso u otro nombre), se presentan en orden de similitud con la primera secuencia de *Arabidopsis* listada para cada grupo, e incluye los dominios conservados del polipéptido en coordenadas de aminoácidos, las respectivas secuencias de los dominios, y la extensión de la identidad en términos de porcentaje respecto a la primera secuencia de *Arabidopsis* listada para cada grupo.

Tabla 1. Familia de genes y dominios de unión para secuencias de ejemplo que confieren tolerancia a la sequía, incluyendo parálogos y ortólogos

SEQ ID NO	GID	Especie	Dominios conservados en coordenadas de aminoácidos del polipéptido	Dominios conservados en coordenadas de bases del polinucleótido	Secuencia del dominio conservado	% de ID en el dominio conservado
						% de ID respecto a G2133
12	G2133	<i>Arabidopsis thaliana</i>	AP2:10-77	AP2:53-256	DQSKYKGIRRRKWGK WVSEIRVPGTRQLWL GSFSTAEGAAVAHDVA FYCLHRPSSLDDESNF PHLL	100%
94	G3646	<i>Brassica oleracea</i>	AP2:10-77	AP2:203-406	HQAKYKGIRRRKWGK WVSEIRVPATRERLWL GSFSTAEGAAVAHDVA FYCLHRPSSLDNEAFN PHLL	91%
92	G3645	<i>Brassica rapa subsp. Pekinensis</i>	AP2: 10-75	AP2: 40-237	TQSKYKGIRRRKWGK WVSEIRVPGTRDRLWL GSFSTAEGAAVAHDVA FYCLHQPNLSLESLNFP LL	89%
2	G47	<i>Arabidopsis thaliana</i>	AP2:10-75	AP2:65-262	SQSKYKGIRRRKWGK WVSEIRVPGTRDRLWL GSFSTAEGAAVAHDVA FFCLHQPDSLESLNFP LL	88%
88	G3643	<i>Glycine max</i>	AP2: 13-78	AP2:101-298	TNNKLGVRRRKWGK WVSEIRVPGTQERLWL GTATPEAAAVAHDV AVYCLSRPSSLDKLNFP ETL	69%
96	G3647	<i>Zinnia elegans</i>	AP2: 13-78	AP2: 53-250	SQKTYKGVRCRRWGK WVSEIRVPGSRERLWL GTYSTPEGAAVAHDVA SYCLKGNTSFHKLNIPS ML	63%
90	G3644	<i>Oryza sativa (japonica cultivar-group)</i>	AP2: 52-122	AP2: 154-366	ERCRYRGVRRRRWGK WVSEIRVPGTRERLWL GSYATPEAAAVAHDTA VYFLRGGAGDGGGGG ATLNFPERA	54%
98	G3649	<i>Oryza sativa (japonica cultivar-group)</i>	AP2: 15-87	AP2: 43-261	EMMRYRGVRRRRWGK WVSEIRVPGTRERLWL GSYATAEAAAVAHDA AVCLLRGGGRRAAA GGGGOLNFPARA	53%

SEQ ID NO	GID	Especie	Dominios conservados en coordenadas de aminoácidos del polipéptido	Dominios conservados en coordenadas de bases del polinucleótido	Secuencia del dominio conservado	% de ID en el dominio conservado
						% de ID respecto a G2133
100	G3651	<i>Oryza sativa</i> (<i>japonica</i> cultivar- group)	AP2: 60-130	AP2: 178-390	ERCRYRGVRRRRWGK WVSEIRVPGTRERLWL GSYATPEAAVAHDTA VYFLRGGAGDGGGGG ATAQLPGAR	52%

[0101] La tabla 2 muestra en la primera columna la SEQ ID NO del polipéptido; en la segunda columna el identificador (por ejemplo, ID de gen (GID) N°); en la tercera columna la familia del factor de transcripción a la que pertenece el polinucleótido; y en la cuarta columna las posiciones de los restos de aminoácidos del dominio conservado en coordenadas de aminoácidos (AA):

Tabla 2. Familias de genes y dominios conservados

SEQ ID NO del polipéptido	identificador	Familia	Dominios conservados en coordenadas de aminoácidos
12	G2133	AP2	10-77

10 Producción de polipéptidos

[0102] Los polinucleótidos de la 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, incluyendo el último ARNm, ARNc, ARN sintético, ADN genómico, ADNc, ADN sintético, oligonucleótidos, etc. Los polinucleótidos son de doble cadena o de cadena sencilla, e incluyen tanto secuencias de sentido directo (es decir, codificantes) como secuencias de sentido contrario (es decir, no codificantes, complementarias). Los polinucleótidos incluyen la secuencia codificante de un factor de transcripción, o un 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 un vector o entorno de un 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.

[0103] Existe una variedad de procedimientos para producir los polinucleótidos de la invención. Los procedimientos para identificar y aislar clones de ADN son bien conocidos para el experto en la materia, y se describen, por ejemplo, en Berger y Kimmel, "Guide to Molecular Cloning Techniques, Methods in Enzymology", vol. 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook y col. "Molecular Cloning - A Laboratory Manual" (2ª ed.), Vol 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989 ("Sambrook") y "Current Protocols in Molecular Biology", Ausubel y col. eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

[0104] Alternativamente, los polinucleótidos de la invención se pueden producir por una variedad de procedimientos de amplificación in vitro adaptados a la presente invención mediante la selección adecuada de cebadores específicos o degenerados. Los ejemplos de protocolos suficiente para dirigir a los expertos en la materia a través de los procedimientos de amplificación in vitro, que incluyen la reacción en cadena de la polimerasa (PCR), reacción en cadena de la ligasa (LCR), amplificación por Q β -replicasa y otras técnicas mediadas por la ARN polimerasa (por ejemplo, NASBA), por ejemplo, para la producción de ácidos nucleicos homólogos de la invención, se encuentran en Berger (véase antes), Sambrook (véase antes), y Ausubel (véase antes), así como en Mullis y col. (1987) "PCR Protocols A Guide to Methods and Applications" (Innis y col. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Se describen procedimientos mejorados para la clonación de ácidos nucleicos amplificados in vitro en Wallace y col. patente de EE.UU. n° 5.426.039. Se resumen procedimientos mejorados para la amplificación de ácidos nucleicos largos mediante PCR en Cheng y col. (1994) *Nature* 369: 684-685, y las referencias citadas en los mismos, en los que se generan amplicones por PCR de hasta 40 kb. Un experto en la materia apreciará que

esencialmente se puede convertir cualquier ARN en un ADN de doble cadena adecuado para la digestión por restricción, expansión por PCR y secuenciación, usando transcriptasa inversa y una polimerasa (por ejemplo, en Ausubel, Sambrook and Berger, véanse todos antes).

- 5 **[0105]** Alternativamente, los polinucleótidos y oligonucleótidos de la invención se pueden formar a partir de fragmentos producidos por procedimientos de síntesis en fase sólida. Típicamente, se sintetizan individualmente fragmentos de hasta aproximadamente 100 bases y después se ligan enzimática o químicamente para producir una secuencia deseada, por ejemplo, un polinucleótido que codifica todo o parte de un factor de transcripción. Por ejemplo, se describe la síntesis química usando el procedimiento de la fosforamidita, en Beaucage y col. (1981)
- 10 *Tetrahedron Letters* 22: 1859-1869; y Matthes y col. (1984) *EMBO J.* 3: 801-805. De acuerdo con dichos procedimientos, se sintetizan oligonucleótidos, se purifican, se reasocian con su cadena complementaria, se ligan y después opcionalmente se clonan en vectores adecuados. Y si se desea, los polinucleótidos y polipéptidos de la invención se pueden encargar a cualquiera de una serie de proveedores comerciales.

15 Secuencias homólogas

- [0106]** Las secuencias homólogas, es decir, las que comparten una identidad o similitud de secuencia significativa con las proporcionadas en la Lista de Secuencias, derivadas de *Arabidopsis thaliana* o de otras plantas elegidas, también son un aspecto de la invención. Las secuencias homólogas se pueden obtener de cualquier planta
- 20 incluyendo monocotiledóneas y dicotiledóneas y en particular especies de plantas importantes en agricultura, incluyendo pero sin limitar, cultivos tales como soja, trigo, maíz, patata, algodón, arroz, colza, colza de semillas oleaginosas (incluida la canola), girasol, alfalfa, trébol, caña de azúcar, y césped, o frutas y verduras, como banano, mora, arándano, fresa, y frambuesa, melón, zanahoria, coliflor, café, pepino, berenjena, uva, melón, lechuga, mango, melón dulce, cebolla, papaya, guisantes, pimientos, piña, calabaza, espinaca, chayote, maíz dulce, tabaco,
- 25 tomate, tomatillo, sandía, frutas rosáceas (tales como manzana, melocotón, pera, cereza y ciruela) y vegetales Brassica (tales como el brócoli, repollo, coliflor, coles de Bruselas y colirrábano). Otros cultivos, que incluyen frutas y verduras, cuyo fenotipo se puede cambiar y que comprenden secuencias homólogas incluyen la cebada; centeno; mijo; sorgo; grosella; aguacate; frutas cítricas tales como naranjas, limones, pomelos y mandarinas, alcachofas, cerezas; frutos secos, tales como la nuez y el cacahuete, endibia, puerro, raíces como de arruruz, remolacha, yuca,
- 30 nabo, el rábano, batata, y boniato; y judías. Las secuencias homólogas también se pueden obtener de las especies leñosas, tales como el pino, álamo y eucalipto o menta u otras labiadas. Además, se pueden obtener secuencias homólogas de plantas que están evolutivamente relacionadas con las plantas de cultivo, pero que pueden no haberse usado todavía como plantas de cultivo. Los ejemplos incluyen belladona (*Atropa belladonna*), relacionada con el tomate; estramonio (*Datura stramonium*), relacionada con el peyote, y el teosinta (*Zea*), relacionado con el
- 35 maíz.

Ortólogos y parálogos

- [0107]** Las secuencias homólogas como las descritas antes pueden comprender secuencias ortólogas y
- 40 parálogos. Son conocidos varios procedimientos diferentes por los expertos en la materia para identificar y definir estas secuencias funcionalmente homólogas. Se han descrito 3 procedimientos generales para definir ortólogos y parálogos; un ortólogo, parólogo u homólogo se puede identificar por uno o más de los procedimientos descritos a continuación.
- 45 **[0108]** Dentro de una sola especie de planta, la duplicación de genes pueden producir dos copias de un gen particular, dando lugar a dos o más genes con secuencia similar y a menudo función similar, conocidos como parálogos. Por lo tanto, un parólogo es un gen similar formado por duplicación dentro de la misma especie. Los parálogos típicamente se agrupan entre sí o en el mismo clado (un grupo de genes similares) cuando se analiza la
- 50 filogenia de una familia de genes usando programas tales como CLUSTAL (Thompson y col. (1994) *Nucleic Acids Res.* 22: 4673-4680; Higgins y col. (1996) *Methods Enzymol.* 266: 383-402). Los grupos de genes similares también se pueden identificar con el análisis de BLAST por parejas (Feng y Doolittle (1987) *J. Mol. Evol.* 25: 351-360). Por ejemplo, un clado de factores de transcripción con dominios MADS muy similares de *Arabidopsis* comparten todos una función común en el momento de la floración (Ratcliffe y col. (2001) *Plant Physiol.* 126: 122-132), y un grupo de factores de transcripción con dominios AP2 muy similares de *Arabidopsis* está implicado en la tolerancia de las
- 55 plantas a la helada (Gilmour y col. (1998) *Plant J.* 16: 433-442). El análisis de grupos de genes similares con función similar que están dentro de un clado pueden dar subsecuencias que son particulares para el clado. Estas subsecuencias, conocidas como secuencias consenso, se pueden usar no sólo para definir las secuencias dentro de cada clado, sino para definir las funciones de esos genes; los genes dentro de un clado pueden contener secuencias parálogos o secuencias ortólogas que comparten la misma función (por ejemplo, en Mount (2001), en

"Bioinformatics: Sequence and Genome Analysis" Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York, página 543).

[0109] 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, a menudo tienen una función idéntica en sus plantas hospedadoras y a menudo son intercambiables entre especies sin perder la función. Debido a que las plantas tienen ancestros comunes, muchos genes en cualquier especie de planta tendrán un gen ortólogo correspondiente en otra especie de planta. Una vez que se ha construido un árbol filogenético para una familia de genes de una especie usando un programa tal como CLUSTAL (Thompson y col. (1994) *Nucleic Acids Res.* 22: 46734680; Higgins y col. (1996) véase antes), se pueden poner las potenciales secuencias ortólogas en el árbol filogenético y determinar su relación con los genes de las especies de interés. Las secuencias ortólogas también se pueden identificar por una estrategia de BLAST recíproca. Una vez que se ha identificado una secuencia ortóloga, la función del ortólogo se puede deducir de la función identificada de la secuencia de referencia.

[0110] Las secuencias de los genes de factores de transcripción están conservadas a través de diferentes líneas de especies eucariotas (Goodrich y col. (1993) *Cell* 75: 519-530; Lin y col. (1991) *Nature* 353: 569-571; Sadowski y col. (1988) *Nature* 335: 563-564). Las plantas no son una excepción a esta observación; diferentes especies de plantas tienen factores de transcripción que tienen secuencias y funciones similares.

[0111] Los genes ortólogos de organismos diferentes tienen funciones altamente conservadas, y muy a menudo funciones esencialmente idénticas (Lee y col. (2002) *Genome Res.* 12: 493-502; Remmy col. (2001) *J. Mol. Biol.* 314: 1041-1052). Los genes parálogos, que han divergido por la duplicación de genes, pueden retener funciones similares de las proteínas codificadas. En dichos casos, los parálogos se pueden usar de forma intercambiable con respecto a determinadas realizaciones de la presente invención (por ejemplo, expresión transgénica de una secuencia codificante). Un ejemplo de dichos parálogos muy relacionados es la familia CBF, con 4 miembros bien definidos en *Arabidopsis* (CBF 1, GBF2, CBF3 y CBF4) y al menos un ortólogo en *Brassica napus*, los cuales controlan todas las rutas implicadas en el estrés tanto por helada como por sequía (Gilmour y col. (1998) *Plant J.* 16: 433-442; Jaglo y col. (1998) *Plant Physiol.* 127: 910-917).

[0112] Las siguientes referencias representan una pequeña muestra de los muchos estudios que demuestran que los genes de factores de transcripción conservados de diferentes especies es probable que funcionen de forma similar (es decir, regulan secuencias diana similares y controlan los mismos rasgos), y que se pueden transformar los factores de transcripción en diversas especies para conferir o mejorar rasgos.

(1) Diferentes factores de transcripción de *Arabidopsis*, incluyendo G28 (patente de EE.UU. 6.664.446), G432 (solicitud de patente de EE.UU. 20040045049; SEQ ID NO: 290 en la presente Lista de Secuencias), G867 (solicitud de patente de EE.UU. 20040098764; SEQ ID NO: 630 en la presente Lista de Secuencias), y G1073 (solicitud de patente de EE.UU. 20040128712; SEQ ID NO: 300 en la presente Lista de Secuencias), han mostrado que confieren tolerancia al estrés abiótico cuando las secuencias se expresan en exceso. Las secuencias de polipéptidos pertenecen a distintos clados de polipéptidos de factores de transcripción que incluyen miembros de diferentes especies. En cada caso, se ha mostrado que un número significativo de secuencias derivadas tanto de dicotiledóneas como de monocotiledóneas confieren tolerancia a diferentes estreses abióticos cuando las secuencias se expresan en exceso.

(2) El gen NPR1 de *Arabidopsis* regula la resistencia sistémica adquirida (SAR) (Cao y col. (1997) *Cell* 88: 57-63); la expresión en exceso de NPR1 conduce a una resistencia potenciada en *Arabidopsis*. Cuando NPR1 de *Arabidopsis* o el ortólogo de NPR1 del arroz se expresaban en exceso en el arroz (el cual, como una monocotiledónea, es diferente de *Arabidopsis*), las plantas transgénicas presentaban una resistencia potenciada a la estimulación con el patógeno bacteriano del añublo del arroz *Xanthomonas oryzae* pv. *Oryzae* (Chern y col. (2001) *Plant J.* 27: 101-113). NPR1 actúa a través de la activación de la expresión de los genes de factores de transcripción, tales como TGA2 (Fan y Dong (2002) *Plant Cell* 14: 1377-1389).

(3) Los genes E2F están implicados en la transcripción de genes de plantas para el antígeno nuclear de proliferación celular (PCNA). Los E2F de plantas comparten un grado alto de similitud en la secuencia de aminoácidos entre monocotiledóneas y dicotiledóneas, y son incluso similares a los dominios conservados de los E2F de animales. Dicha conservación indica una similitud funcional entre E2F de plantas y animales. Los factores de transcripción E2F que regulan el desarrollo del meristemo actúan a través de elementos cis comunes, y regulan genes relacionados (PCNA) (Kosugi y Ohashi, (2002) *Plant J.* 29: 45-59).

- (4) El gen ABI5 (insensible a ABA 5) codifica un factor de la cremallera de leucina básico necesario para la respuesta de ABA en la semilla y los tejidos vegetativos. Los experimentos de transformación simultánea con construcciones de ADNc de ABI5 en protoplastos de arroz dieron como resultado la trasactivación específica de promotores del trigo, *Arabidopsis*, judía y cebada inducibles por ABA. Estos resultados demuestran que los factores de transcripción similares en secuencia a ABI5 son dianas clave de una ruta de señalización de ABA conservada en diversas plantas (Gampala y col. (2001) *J. Biol. Chem.* 277: 1689-1694).
- (5) Las secuencias de 3 genes de tipo GAMYB de *Arabidopsis* se obtuvieron basándose en la similitud de secuencia con los genes GAMYB de la cebada, arroz y *L. temulentum*. Se determinó que estos 3 genes de *Arabidopsis* codificaban factores de transcripción (AtMYB33, AtMYB65 y AtMYB101) y podían sustituir la expresión de GAMYB de la cebada y de la α -amilasa de control (Gocal y col. (2001) *Plant Physiol.* 127: 1682-1693).
- (6) El gen del control floral LEAFY de *Arabidopsis* puede acelerar espectacularmente la floración en numerosas plantas dicotiledóneas. La expresión constitutiva de LEAFY de *Arabidopsis* también producía la floración temprana en el arroz transgénico (una monocotiledóneas), con una fecha de espigamiento que era 26-34 días antes que la de las plantas silvestres. Estas observaciones indican que los genes reguladores de la floración de *Arabidopsis* son herramientas útiles para mejorar la fecha de espigamiento en cultivos de cereales (He y col. (2000) *Transgenic Res.* 9: 223-227).
- (7) Las giberelinas (GA) bioactivas son reguladores endógenos esenciales del crecimiento de la planta. La señalización de GA tiende a ser conservada a través del reino vegetal. La señalización de GA es mediada por GAI, un miembro nuclear de la familia GRAS de factores de transcripción de plantas. Se ha mostrado que GAI de *Arabidopsis* funciona en el arroz para inhibir las rutas de respuestas a las giberelinas (Fu y col. (2001) *Plant Cell* 13: 1791-1802).
- (8) El gen SUPERMAN (SUP) de *Arabidopsis* codifica un factor de transcripción presunto que mantiene el límite entre los estambres y los carpelos. Mediante la expresión en exceso de SUP de *Arabidopsis* en el arroz, se mostró que el efecto de la presencia de los genes en los bordes de los verticilos era conservado. Esto demostraba que SUP es un regulador conservado de los bordes de los verticilos florales y afecta a la proliferación celular (Nandi y col. (2000) *Curr. Biol.* 10: 215-218).
- (9) Los factores de transcripción myb de maíz, petunia y *Arabidopsis* que regulan la biosíntesis de flavonoides son genéticamente similares y afectan al mismo rasgo en su especie nativa; por lo tanto, la secuencia y la función de estos factores de transcripción myb se correlacionan entre sí en diversas especies (Borevitz y col. (2000) *Plant Cell* 12: 2383-2394).
- (10) Los genes de la altura reducida del trigo 1 (Rht-B1/Rht-D1) y maíz enano 8 (d8) son ortólogos del gen insensible a la giberelina (GAI) de *Arabidopsis*. Ambos genes se han usado para producir variedades de grano enano que tiene un mejor rendimiento de granos. Estos genes codifican proteínas que se parecen a los factores de transcripción nucleares y contienen un dominio de tipo SH-2, indicando que la fosfotirosina puede participar en la señalización de la giberelina. Las plantas de arroz transgénicas que contienen un alelo de GAI mutante de *Arabidopsis* han mostrado producir menores respuestas a la giberelina y las hacen enanas, indicando que los ortólogos de GAI mutantes se podrían usar para aumentar el rendimiento en una amplia variedad especies de cultivo (Peng y col. (1999) *Nature* 400: 256-261).

[0113] Los factores de transcripción que son homólogos a las secuencias listadas típicamente compartirán al menos aproximadamente una identidad de secuencia de aminoácidos de 70% en su dominio conservado. Los factores de transcripción más estrechamente relacionados pueden compartir al menos aproximadamente 79% o aproximadamente 90% o aproximadamente 95% o aproximadamente 98% o mayor identidad de secuencia con las secuencias listadas, o con las secuencias listadas pero excluyendo o fuera de una secuencia consenso conocida o sitio de unión al ADN consenso, o con las secuencias listadas excluyendo uno o todos los dominios conservados. Los factores que están más estrechamente relacionados con las secuencias listadas comparten, por ejemplo, al menos aproximadamente 85%, aproximadamente 90% o aproximadamente 95% o mayor % de identidad de secuencia con las secuencias listadas, o con las secuencias listadas pero excluyendo o fuera de una secuencia consenso conocida o sitio de unión al ADN consenso, o fuera de uno o todos los dominios conservados. A nivel de nucleótidos, las secuencias típicamente compartirán una identidad de secuencia de nucleótidos de al menos aproximadamente 40%, preferiblemente al menos aproximadamente 50%, aproximadamente 60%, aproximadamente 70% o aproximadamente 80% de identidad de secuencia y más preferiblemente

aproximadamente 85%, aproximadamente 90%, aproximadamente 95% o aproximadamente 97% o mayor identidad de secuencia con una o más de las secuencias listadas, o con una secuencia listada pero excluyendo o fuera de una secuencia consenso conocida o sitio de unión al ADN consenso, o fuera de uno o todos los dominios conservados. La degeneración del código genético permite variaciones importantes de la secuencia de nucleótidos de un polinucleótido mientras que se mantiene la secuencia de aminoácidos de la proteína codificada. Los dominios AP2 dentro de la familia de factores de transcripción AP2 puede presentar un grado mayor de homología de secuencia, tal como una identidad de secuencia de aminoácidos de al menos 70% incluyendo sustituciones conservativas, y preferiblemente una identidad de secuencia de al menos 80%, y más preferiblemente al menos 85%, o al menos aproximadamente 86%, o al menos aproximadamente 87%, o al menos aproximadamente 88%, o al menos aproximadamente 90%, o al menos aproximadamente 95%, o al menos aproximadamente 98% de identidad de secuencia. Los factores de transcripción que son homólogos con las secuencias listadas deben compartir una identidad de secuencia de aminoácidos de al menos 30%, o al menos aproximadamente 60%, o al menos aproximadamente 75%, o al menos aproximadamente 80%, o al menos aproximadamente 90%, o al menos aproximadamente 95% a lo largo de la longitud entera del polipéptido o el homólogo.

15

[0114] El porcentaje de identidad se puede determinar electrónicamente, 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 procedimientos, por ejemplo, el procedimiento clustal (por ejemplo, en Higgins y Sharp (1988) *Gene* 73: 237-244). El algoritmo clustal agrupa secuencias en agrupaciones examinando las distancias entre todos los pares. Las agrupaciones se alinean en parejas y después en grupos. Se pueden usar otros programas o algoritmos de alineamiento, incluyendo FASTA, BLAST, o ENTREZ, FASTA y BLAST, y se pueden usar para calcular el porcentaje de similitud. Éstos están disponibles como parte del paquete de análisis de secuencia GCG (Universidad de Wisconsin, Madison, Wis.), y se pueden usar con o sin los parámetros por defecto. ENTREZ está disponible en el National Center for Biotechnology Information. En una realización, el porcentaje de identidad de dos secuencias se puede determinar mediante el programa GCG con un peso de hueco de 1, por ejemplo, cada hueco de aminoácido se pesa como si fuera un solo aminoácido o emparejamiento erróneo de nucleótido entre las dos secuencias (USPN 6.262.333).

20

25

[0115] Se describen otras técnicas de alineamiento en *Enzymology*, vol. 266, "Computer Methods for Macromolecular Sequence Analysis" (1996), ed. Doolittle, Academic Press, Inc., San Diego, Calif., EE.UU. Preferiblemente, se usa 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 (Shpaer (1997) *Methods Mol. Biol.* 70: 173-187). También se puede usar el programa GAP, usando el procedimiento de alineamiento de Needleman y Wunsch para alinear las secuencias. Una estrategia de búsqueda alternativa usa el software MPSRCH, el cual funciona en un ordenador MASPARE. MPSRCH usa un algoritmo de Smith-Waterman para puntuar las secuencias en un ordenador de procesamiento paralelo masivo. Este procedimiento implica la capacidad de coger emparejamientos relacionados a distancia, y es especialmente tolerante para huecos pequeños y errores de secuencias de nucleótidos. Las secuencias de aminoácidos codificadas por ácidos nucleicos se pueden usar para buscar en bases de datos tanto de proteínas como de ADN.

40

[0116] El porcentaje de similitud entre dos secuencias de polipéptido, por ejemplo la secuencia A y la secuencia B, se calcula dividiendo la longitud de la secuencia A, menos el número de huecos de restos de la secuencia A, menos el número de restos de huecos en la secuencia B, entre la suma de los emparejamientos de restos entre la secuencia A y la secuencia B, por cien. Los huecos sin o con poca similitud entre las dos secuencias de aminoácidos no están incluidos en la determinación del porcentaje de similitud. El porcentaje de identidad entre secuencias de polinucleótidos también se puede contar o calcular por otros procedimientos conocidos en la técnica, por ejemplo, el procedimiento de Jotun Hein (por ejemplo, en Hein (1990) *Methods Enzymol.* 183: 626-645). La identidad entre secuencias también se puede determinar por otros procedimientos conocidos en la técnica, por ejemplo, variando las condiciones de hibridación (solicitud de patente de EE.UU. N° 20010010913).

50

[0117] Por lo tanto, la invención proporciona procedimientos para identificar una secuencia similar o paróloga u ortóloga 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 lo que se indique de otra manera en el presente documento, y puede incluir la conexión o asociación del fenotipo de una planta dada o función génica con una secuencia. En los procedimientos, se proporciona una base de datos de secuencias (local o a través de internet o intranet) y se plantea una cuestión frente a la base de datos de secuencias usando las secuencias relevantes del presente documento y los fenotipos de plantas o funciones de genes asociados.

55

[0118] Además, se pueden usar una o más secuencias de polinucleótidos o uno o más polipéptidos

codificados por las secuencias de polinucleótidos para buscar frente a BLOCKS (Bairoch y col. (1997) *Nucleic Acids Res.* 25: 217-221), PFAM, y otras bases de datos que contienen patrones, secuencias y funciones de genes previamente identificados y anotados. Se pueden usar los procedimientos que buscan patrones de secuencia primaria con penalizaciones por huecos en la estructura secundaria (Smith y col. (1992) *Protein Engineering* 5: 35-51) así como algoritmos tales como Basic Local Alignment Search Tool (BLAST; Altschul (1993) *J. Mol. Evol.* 36: 290-300; Altschul y col. (1990) *J. Mol. Biol.* 215: 403-410), BLOCKS (Henikoff y Henikoff (1991) *Nucleic Acids Res.* 19: 6565-6572), Hidden Markov Models (HMM; Eddy (1996) *Curr. Opin. Str. Biol.* 6: 361-365; Sonnhammer y col. (1997) *Proteins* 28: 405-420), y similares, para manipular y analizar secuencias de polinucleótidos y polipéptidos codificados por los polinucleótidos. Estas bases de datos, algoritmos y otros procedimientos son bien conocidos en la técnica y se describen en Ausubel y col. (1997) "Short Protocols in Molecular Biology", John Wiley & Sons, Nueva York, NY, unit 7.7) y en Meyers (1995) "Molecular Biology and Biotechnology", Wiley VCH, Nueva York, NY, p 856-853).

[0119] Un procedimiento adicional para identificar o confirmar que las secuencias homólogas específicas controlan la misma función es por comparación del o de los perfiles de los transcritos obtenidos tras la expresión en exceso o inactivación de dos o más factores de transcripción relacionados. Puesto que los perfiles de transcritos son diagnóstico para estados celulares específicos, un experto la materia apreciará que los genes que tienen un perfil de transcritos muy similar (por ejemplo, con más de 50% de los transcritos regulados en común, más preferiblemente con más de 70% de los transcritos regulados en común, lo más preferiblemente con más de 90% de los transcritos regulados en común) tendrán funciones muy similares. Fowler y col. (2002) *Plant Cell* 14: 1675-1679) han mostrado que 3 genes parálogos de la familia AP2 (CBF1, CBF2 y CBF3), cada uno de los cuales es inducido tras tratamiento con frío, y cada uno de los cuales pueden condicionar una mejor tolerancia a la helada, tienen perfiles de transcritos muy 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 parálogos u ortólogos presuntos tienen la misma función.

[0120] Además, se pueden usar procedimientos que usan el alineamiento manual de secuencias similares u homólogas con una o más secuencias de polinucleótidos o uno o más polipéptidos codificados por las secuencias de polinucleótidos, para identificar regiones de similitud y dominios de unión de AP2. Dichos procedimientos manuales son bien conocidos para el experto en la materia y pueden incluir, por ejemplo, comparaciones de la estructura terciaria entre una secuencia de polipéptido codificada por un polinucleótido que comprende una función conocida con una secuencia de polipéptido codificado por una secuencia de polinucleótido que tiene una función todavía no determinada. Dichos ejemplos de estructura terciaria pueden comprender hélices α , láminas β , hélices anfipáticas, patrones de cremallera de leucina, patrones de dedos de cinc, regiones ricas en prolina, patrones de repetición de cisteína, y similares.

[0121] Los ortólogos y parálogos de los presentes factores de transcripción descritos, se pueden clonar usando composiciones proporcionadas por la presente invención de acuerdo con procedimientos conocidos en la técnica. Los ADNc se pueden clonar usando ARNm de una célula o tejido de planta que expresa uno de los presentes factores de transcripción. Las fuentes de ARNm adecuadas se pueden identificar examinando transferencias Northern con sondas diseñadas a partir de secuencias de los presentes factores de transcripción, después de lo cual se prepara una biblioteca a partir del ARNm obtenido de una célula o tejido positivo. Después, el ADNc que codifica el factor de transcripción se aísla usando, por ejemplo, la PCR, usando cebadores diseñados a partir de la secuencia de genes del presente factor de transcripción descrito, o mediante hibridación con un ADNc parcial o completa o con uno o más conjuntos de sondas degeneradas basadas en las secuencias descritas. La biblioteca de ADNc se puede usar para transformar células de plantas. La expresión de los ADNc de interés se detecta usando, por ejemplo, los procedimientos descritos en el presente documento tales como micromatrices, transferencias Northern, PCR cuantitativa, o cualquier otra técnica para el seguimiento de los cambios en la expresión. Los clones genómicos se pueden aislar usando técnicas similares a estas.

Identificación de polinucleótidos o ácidos nucleicos por hibridación

[0122] Los polinucleótidos homólogos a las secuencias ilustradas en la Lista de Secuencias y en las tablas, se pueden identificar, por ejemplo, por hibridación entre sí en condiciones rigurosas o muy rigurosas. Los polinucleótidos monocatenarios hibridan cuando se asocian basándose en una variedad de fuerzas fisicoquímicas bien caracterizada, tales como enlaces de hidrógeno, exclusión de disolvente, apilamiento de bases y similares. La restricción de una hibridación refleja el grado de identidad de secuencia de los ácidos nucleicos implicados, de modo que cuanto más altas son las condiciones rigurosas, más similares son las dos cadenas de polinucleótido. En las condiciones rigurosas influyen una variedad de factores, incluyendo la temperatura, concentración de sales y

composición, aditivos orgánicos y no orgánicos, disolventes, etc. presentes tanto en las disoluciones de hibridación como en las de lavado, e incubaciones (y número de las mismas), como se describe con más detalle en las referencias citadas antes.

- 5 **[0123]** A la estabilidad de los dúplex de ADN le afecta factores tales como la composición de bases, longitud y grado de emparejamiento erróneo de los pares de bases. Las condiciones de hibridación se pueden ajustar para permitir que los ADN de diferentes secuencias relevantes hibriden. La temperatura de fusión (T_f) se define como la temperatura cuando 50% de las moléculas de dúplex se han disociado en sus constituyentes de cadenas sencillas. Las temperaturas de fusión de un dúplex perfectamente emparejado, en el que el tampón de hibridación contiene formamida como agente desnaturizante, se puede calcular por las siguientes ecuaciones:

(I) ADN-ADN:

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

15

(II) ADN-ARN:

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

20 (III) ARN-ARN:

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

en las que L es la longitud del dúplex formado, $[\text{Na}^+]$ es la concentración molar del ion sodio en la disolución de hibridación o de lavado, y % de G+C es el porcentaje de las bases (guanina + citosina) en el híbrido. Para híbridos emparejados de forma imperfecta, es necesario aproximadamente 1°C para reducir la temperatura de fusión para cada 1% de emparejamientos erróneos.

- [0124]** Los experimentos de hibridación en general se llevan a cabo en un tampón con pH entre 6,8 y 7,4, aunque la velocidad de hibridación es casi independiente del pH a las fuerzas iónicas que es probable que se usen en el tampón de hibridación (Anderson y Young (1985) "Quantitative Filter Hybridisation." En: Hames y Higgins, ed., Nucleic Acid Hybridisation, A Practical Approach. Oxford, IRL Press, 73-111). Además, se puede usar uno o más de los siguientes para reducir la hibridación no específica: ADN de esperma de salmón tratado con ultrasonidos u otro ADN no complementario, albúmina de suero bovino, pirofosfato de sodio, dodecilsulfato de sodio (SDS), polivinilpirrolidona, Ficoll y disolución de Denhardt. El sulfato de dextrano y el polietilenglicol 6000 actúan para excluir el ADN de la disolución, aumentando así la concentración de ADN sonda eficaz y la señal de hibridación en una unidad de tiempo dada. En algunos casos, pueden ser deseables o necesarias condiciones incluso más rigurosas para reducir la hibridación de fondo y/o no específica. Estas condiciones se pueden crear con el uso de temperatura mayor, fuerza iónica menor y mayor concentración de un agente desnaturizante tal como la formamida.

40

- [0125]** Las condiciones rigurosas se pueden ajustar para cribar fragmentos moderadamente similares tales como secuencias homólogas de organismos relacionados de forma distante, o fragmentos muy similares tales como genes que duplican enzimas funcionales de organismos relacionados de cerca. Las condiciones rigurosas se pueden ajustar durante la etapa de hibridación, o en los lavados posteriores a la hibridación. La concentración de sal, concentración de formamida, temperatura de hibridación y longitud de la sonda, son variables que se pueden usar para alterar las condiciones rigurosas (como se describe mediante la fórmula anterior). Como norma general las condiciones muy rigurosas se llevan a cabo típicamente de $T_f - 5^{\circ}\text{C}$ a $T_f - 20^{\circ}\text{C}$, condiciones rigurosas moderadas de $T_f - 20^{\circ}\text{C}$ a $T_f - 35^{\circ}\text{C}$ y condiciones poco rigurosas de $T_f - 35^{\circ}\text{C}$ a $T_f - 50^{\circ}\text{C}$, para dúplex de >150 pares de bases. La hibridación se puede llevar a cabo en condiciones de poco rigurosas a moderadas ($25-50^{\circ}\text{C}$ por debajo de la T_f), seguido de lavados posteriores a la hibridación con condiciones rigurosas crecientes. Las velocidades de hibridación máximas en disolución se determina empíricamente que se producen a $T_f - 25^{\circ}\text{C}$ para dúplex de ADN-ADN y $T_f - 15^{\circ}\text{C}$ para dúplex de ARN-ADN. Opcionalmente, el grado de disociación se puede evaluar después de cada etapa de lavado para determinar la necesidad de posteriores etapas de lavado, de condiciones más rigurosas.

- 55 **[0126]** Las condiciones rigurosas altas se pueden usar para seleccionar secuencias de ácidos nucleicos con

grados altos de identidad con las secuencias descritas. Un ejemplo de condiciones de hibridación rigurosas obtenidas en un procedimiento basado en filtro, tal como una transferencia Southern o Northern para la hibridación de ácidos nucleicos complementarios que tienen más de 100 restos complementarios, es aproximadamente de 5°C a 20°C inferior que el punto de fusión térmico (T_f) para la secuencia específica con una fuerza iónica y pH definidos.

5 Las condiciones usadas para la hibridación pueden incluir cloruro de sodio de aproximadamente 0,02 M a aproximadamente 0,15 M, caseína de aproximadamente 0,5% a aproximadamente 5%, SDS aproximadamente al 0,02% o N-laurilsarcosina aproximadamente al 0,1%, citrato de sodio de aproximadamente 0,001 M a aproximadamente 0,03 M, a temperaturas de hibridación entre aproximadamente 50°C y aproximadamente 70°C. Más preferiblemente, las condiciones muy rigurosas son cloruro de sodio aproximadamente 0,02 M, caseína

10 aproximadamente al 0,5%, SDS aproximadamente al 0,02%, citrato de sodio aproximadamente 0,001 M, a una temperatura de aproximadamente 50°C. Las moléculas de ácido nucleico que hibridan en condiciones rigurosas típicamente hibridarán con una sonda basada en la molécula de ADN entera o partes seleccionadas, por ejemplo, con una subsecuencia única del ADN.

15 **[0127]** La concentración de sal rigurosa normalmente será menor de aproximadamente NaCl 750 mM y citrato trisódico 75 mM. Se pueden obtener condiciones rigurosas crecientes con menos de aproximadamente NaCl 500 mM y citrato trisódico 50 mM, hasta condiciones incluso más rigurosas con menos de aproximadamente NaCl 250 mM y citrato trisódico 25 mM. La hibridación en condiciones poco rigurosas se puede obtener en ausencia de disolvente orgánico, por ejemplo, formamida, mientras que la hibridación en condiciones muy rigurosas se puede

20 obtener en presencia de formamida al menos aproximadamente al 35% y más preferiblemente formamida al menos aproximadamente al 50%. Las condiciones de temperatura rigurosas normalmente incluirán temperaturas de al menos aproximadamente 30°C, más preferiblemente de al menos aproximadamente 37°C y lo más preferiblemente de al menos aproximadamente 42°C, con formamida presente. Los diversos parámetros adicionales, tales como el tiempo de hibridación, la concentración de detergentes, por ejemplo, de dodecilsulfato de sodio (SDS) y la fuerza

25 iónica, son bien conocidas por el experto en la materia. Los diferentes niveles de condiciones rigurosas se logran combinando estas diferentes condiciones según sea necesario.

[0128] Las etapas de lavado que siguen a la hibridación también pueden tener distintas condiciones rigurosas; las etapas de lavado posteriores a la hibridación determinan principalmente la especificidad de la

30 hibridación, siendo los factores más críticos la temperatura y la fuerza iónica de la disolución de lavado final. Las condiciones de lavado rigurosas se pueden aumentar disminuyendo la concentración de sales o aumentando la temperatura. Las condiciones rigurosas de la concentración de sales para las etapas de lavado serán preferiblemente menores de aproximadamente NaCl 3 mM y citrato sódico 3 mM, y lo más preferiblemente menores de aproximadamente NaCl 1,5 mM y citrato sódico 1,5 mM.

35 **[0129]** Por lo tanto, las condiciones de hibridación y de lavado que se pueden usar para unir y eliminar polinucleótidos con homología menor de la deseada con las secuencias de ácidos nucleicos o sus complementarias que codifican los factores de transcripción presentes incluyen, por ejemplo:

40 6X SSC a 65°C;
formamida al 50%, 4X SSC a 42°C; o
0,5X SSC, SDS al 0,1% a 65°C;
con, por ejemplo, 2 etapas de lavado de 10-30 minutos cada una. Las variaciones útiles de estas condiciones serán fácilmente evidentes para el experto en la materia.

45 **[0130]** Un experto en la materia no esperará una variación sustancial entre las especies de polinucleótidos abarcadas en el alcance de la presente invención, porque las condiciones muy rigurosas expuestas en las fórmulas anteriores dan polinucleótidos estructuralmente similares.

50 **[0131]** Si se desea, se pueden usar etapas de lavado incluso más rigurosas, incluyendo aproximadamente 0,2X SSC, SDS al 0,1% a 65°C y lavado dos veces, siendo cada etapa de lavado de aproximadamente 30 min, o aproximadamente 0,1 X SSC, SDS al 0,1% a 65°C y lavado dos veces durante 30 min. La temperatura de las disoluciones de lavado normalmente será al menos aproximadamente 25°C, y para condiciones más rigurosas al menos aproximadamente 42°C. Las condiciones rigurosas de la hibridación se pueden aumentar incluso más

55 usando las mismas condiciones que en las etapas de hibridación, aumentando la temperatura de lavado de aproximadamente 3°C a aproximadamente 5°C, y las condiciones rigurosas se pueden aumentar incluso más usando las mismas condiciones excepto que la temperatura de lavado se aumenta de aproximadamente 6°C a aproximadamente 9°C. Para la identificación de homólogos menos relacionados, las etapas de lavado se pueden llevar a cabo a una temperatura inferior, por ejemplo, 50°C.

- [0132]** Un ejemplo de una etapa de lavado de condiciones rigurosas bajas usa una disolución y condiciones de al menos 25°C en NaCl 30 mM, citrato trisódico 3 mM y SDS al 0,1% durante 30 min. Se pueden obtener condiciones más rigurosas a 42°C en NaCl 15 mM, citrato trisódico 1,5 mM y SDS al 0,1% durante 30 min. Se obtienen condiciones incluso más rigurosas a 65-68°C en una disolución de NaCl 15 mM, citrato trisódico 1,5 mM y SDS al 0,1%. Los procedimientos de lavado usarán en general al menos 2 etapas finales de lavado. Las variaciones adicionales de estas condiciones serán fácilmente evidentes para el experto en la materia (por ejemplo, en la solicitud de patente de EE.UU. N° 20010010913).
- 10 **[0133]** Las condiciones rigurosas se pueden seleccionar de modo que un oligonucleótido que sea perfectamente complementario con el oligonucleótido codificante hibride con el oligonucleótido codificante con una relación de señal a ruido de al menos aproximadamente 5-10x mayor que la relación para la hibridación del oligonucleótido perfectamente complementario con 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 a ruido mayor, es decir, aproximadamente 15x o más. Por consiguiente, un ácido nucleico objeto hibridará con un oligonucleótido codificante único con una relación de señal a ruido de al menos 2x o mayor, comparado con la hibridación del oligonucleótido codificante con un ácido nucleico que codifica un polipéptido conocido. La señal particular dependerá del marcador usado en el ensayo pertinente, por ejemplo, un marcador fluorescente, un marcador colorimétrico, un marcador radiactivo, o similares. La hibridación marcada o sondas de PCR para detectar secuencias de polinucleótidos relacionadas se pueden producir por oligomarcaje, traslado de mellas, marcaje del extremo o amplificación por PCR usando un nucleótido marcado.

- [0134]** La invención abarca todas las secuencias de polinucleótidos que son capaces de hibridar con las secuencias de polinucleótidos de la Lista de Secuencias, y fragmentos de las mismas en diferentes condiciones rigurosas (por ejemplo, en Wahl y Berger (1987) *Methods Enzymol.* 152: 399-407, y Kimmel (1987) *Methods Enzymol.* 152: 507-511). Las valoraciones de homología se proporcionan por hibridación de ADN-ADN o ADN-ARN en condiciones rigurosas como entienden bien los expertos en la materia (Hames y Higgins, Eds. (1985) "Nucleic Acid Hybridisation: A Practical Approach", IRL Press, Oxford, Reino Unido). Las condiciones rigurosas se pueden ajustar para cribar fragmentos moderadamente similares, tales como secuencias homólogas de organismos relacionados lejanamente, hasta fragmentos muy similares, tales como genes que duplican enzimas funcionales de organismos estrechamente relacionados. Los lavados posteriores a la hibridación determinan las condiciones rigurosas.

Identificación de polinucleótidos o ácidos nucleicos con bibliotecas de expresión

- 35 **[0135]** Además de los procedimientos de hibridación, se pueden obtener polipéptidos homólogos de factores de transcripción mediante el cribado de una biblioteca de expresión usando anticuerpos específicos para uno o más factores de transcripción. Proporcionando en el presente documento el factor de transcripción descrito y secuencias de ácidos nucleicos homólogas de los factores de transcripción, el o los polipéptidos codificados se pueden expresar y purificar en un sistema de expresión heterólogo (por ejemplo, *E. coli*) y usar para producir anticuerpos (monoclonales o policlonales) específicos para el o los polipéptidos en cuestión. También se pueden producir anticuerpos contra péptidos sintéticos derivados de secuencias o subsecuencias de aminoácidos de un factor de transcripción u homólogo del factor de transcripción. Los procedimientos para producir anticuerpos son bien conocidos en la técnica y se describen en Harlow y Lane (1988), "Antibodies: A Laboratory Manual", Cold Spring Harbor Laboratory, Nueva York. Dichos anticuerpos se pueden usar después para cribar una biblioteca de expresión producida a partir de la planta de la cual se desean clonar homólogos de factores de transcripción adicionales, usando los procedimientos descritos antes. Los ADNc seleccionados se pueden confirmar mediante secuenciación y actividad enzimática.

50 Variaciones de secuencia

- [0136]** Los expertos en la materia apreciarán fácilmente que cualquiera de una variedad de secuencias de polinucleótidos son capaces de codificar los factores de transcripción y polipéptidos homólogos de factores de transcripción de la invención. Debido a la degeneración del código genético, muchos polinucleótidos diferentes pueden codificar polipéptidos idénticos y/o sustancialmente similares además de las secuencias ilustradas en la Lista de Secuencias. Los ácidos nucleicos que tienen una secuencia que difiere de las secuencias mostradas en la Lista de Secuencias, o secuencias complementarias, que codifican péptidos funcionalmente equivalentes (es decir, péptidos que tienen el mismo grado de actividad biológica equivalente o similar) pero difieren en la secuencia respecto a la secuencia mostrada en la Lista de Secuencias debido a la degeneración del código genético, también

están dentro del alcance de la invención.

[0137] Las secuencias de polinucleótidos alteradas que codifican polipéptidos incluyen las secuencias con eliminaciones, inserciones o sustituciones de diferentes nucleótidos, que dan como resultado un polinucleótido que codifica un polipéptido con al menos una característica funcional de los presentes polipéptidos. Están incluidos en esta definición los polimorfismos que pueden ser o no fácilmente detectables usando una sonda de oligonucleótido particular del polinucleótido que codifica los presentes polipéptidos, y la hibridación inadecuada o inesperada con variantes alélicas, con un locus distinto del locus cromosómico normal para la secuencia de polinucleótido que codifica los presentes polipéptidos.

[0138] Las variantes alélicas se refieren a cualquiera de dos o más formas alternativas de un gen que ocupan el mismo locus cromosómico. La variación alélica surge naturalmente por mutación, y puede dar como resultado polimorfismo fenotípico dentro de las poblaciones. Las mutaciones de genes pueden ser silenciosas (es decir no hay cambio en el polipéptido codificado) o pueden codificar polipéptidos que tienen la 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 por corte y empalme se refiere a formas alternativas de ARN transcritas a partir de un gen. La variación por corte y empalme surge de forma natural por el uso de sitios de corte y empalme alternativos dentro de una molécula de ARN transcrita, o de forma menos común entre moléculas de ARN transcritas por separado, y puede producir varios ARNm transcritos a partir del mismo gen. Las variantes por corte y empalme pueden codificar polipéptidos que tienen la secuencia de aminoácidos alterada. La expresión variante por corte y empalme también se usa en el presente documento para indicar una proteína codificada por una variante por corte y empalme de un ARNm transcrito a partir de un gen.

[0139] Los expertos en la materia reconocerán que, por ejemplo, G2133, SEQ ID NO: 12, representa un solo factor de transcripción, y se puede esperar que se produzca variación alélica y corte y empalme alternativo. Las variantes alélicas de la SEQ ID NO: 11 se pueden clonar mediante hibridación de bibliotecas de ADNc o genómicas de diferentes organismos individuales de acuerdo con procedimientos estándar. Las variantes alélicas de la secuencia de ADN mostrada en la SEQ ID NO: 11, incluyendo las que contienen mutaciones silenciosas y aquellas en las que las mutaciones producen cambios en la secuencia de aminoácidos, están dentro del alcance de la presente invención, como lo están las proteínas que son variantes alélicas de la SEQ ID NO: 12. Los ADNc generados a partir de ARNm de corte y empalme alternativo, que retienen las propiedades del factor de transcripción están incluidos dentro del alcance de la presente invención, como lo están los polipéptidos codificados por dichos ADNc y ARNm. Las variantes alélicas y variantes por corte y empalme de estas secuencias se pueden clonar por hibridación de bibliotecas de ADNc o genómicas de diferentes organismos o tejidos individuales de acuerdo con procedimientos estándar conocidos en la técnica (documento USPN 6.388.064).

[0140] Por lo tanto, además de las secuencias expuestas en la Lista de Secuencias, la invención también abarca moléculas de ácidos nucleicos relacionadas que incluyen variantes alélicas o por corte y empalme de las secuencias de la Lista de Secuencias, e incluyen secuencias que son complementarias de cualquiera de las secuencias de nucleótidos anteriores. Las moléculas de ácido nucleico relacionadas también incluyen secuencias de nucleótidos que codifican un polipéptido que comprende una sustitución, modificación, adición y/o eliminación de uno o más restos de aminoácidos comparado con las secuencias de polipéptidos de la Lista de Secuencias y equivalentes. Dichos polipéptidos relacionados pueden comprender, por ejemplo, adiciones y/o eliminaciones de uno o más sitios de glicosilación unidos a N o unidos a O, o una adición y/o una eliminación de uno o más restos de cisteína.

[0141] Por ejemplo, la tabla 3 ilustra, por ejemplo, que los codones AGC, AGT, TCA, TCC, TCG, y TCT codifican todos el mismo aminoácido: la serina. Por consiguiente, en cada posición en la secuencia en la que hay un codón que codifica una serina, se puede usar cualquiera de las secuencias trinucleótidas anteriores sin alterar el polipéptido codificado.

Tabla 3

Aminoácido			Posibles Codones			
Alanina	Ala	A	GCA	GCC	GCG	GCT
Cisteína	Cys	C	TGC	TGT		

Aminoácido			Posibles Codones						
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	CGC	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					

- [0142]** Las alteraciones de secuencia que no cambien la secuencia de aminoácidos codificada por el polinucleótido se denominan variaciones “silenciosas”. Con la excepción de los codones ATG y TGG, que codifican la metionina y el triptófano, respectivamente, cualquiera de los posibles codones para el mismo aminoácido se puede sustituir por una variedad de técnicas, por ejemplo, mutagénesis dirigida, disponible en la técnica. Por consiguiente, todas y cualquiera de dichas variaciones de una secuencia seleccionada de la tabla anterior son una característica de la invención.
- 10 **[0143]** Además de las variaciones silenciosas, se pueden hacer otras variaciones conservativas que alteran uno o algunos restos de aminoácidos en el polipéptido codificado, sin alterar la función del polipéptido, y estas variantes conservativas son también una característica de la invención.
- 15 **[0144]** Por ejemplo, la invención también prevé sustituciones, eliminaciones e inserciones introducidas en las secuencias proporcionadas en la Lista de Secuencias. Dichas modificaciones de secuencia se pueden diseñar en una secuencia mediante mutagénesis dirigida (Wu (ed.) *Methods Enzymol.* (1993) vol. 217, Academic Press) o por otros procedimientos indicados más adelante. Las sustituciones de aminoácidos normalmente son de restos individuales; las inserciones normalmente serán del orden de aproximadamente de 1 a 10 restos de aminoácidos; y las eliminaciones estarán en el intervalo de aproximadamente 1 a 30 restos. En una realización, las eliminaciones o inserciones se hacen en pares adyacentes, por ejemplo, una eliminación de dos restos o inserción de dos restos. Se pueden combinar sustituciones, eliminaciones, inserciones o cualquier combinación de los mismos para llegar a una secuencia. Las mutaciones que se hacen en el polinucleótido que codifica el factor de transcripción no deben poner a la secuencia fuera del marco de lectura y no deben crear regiones complementarias que puedan producir una estructura de ARNm secundaria. Preferiblemente, el polipéptido codificado por el ADN realiza la función deseada.
- 20
- 25 **[0145]** Las sustituciones conservativas son aquellas en las que se ha eliminado al menos un resto en la secuencia de aminoácidos y se ha insertado un resto diferente en su lugar. Dichas sustituciones en general se hacen de acuerdo con la tabla 4 cuando se desea mantener la actividad de la proteína. La tabla 4 muestra los aminoácidos que se pueden sustituir por un aminoácido en una proteína y que se considera normalmente como
- 30 sustituciones conservativas.

Tabla 4

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

[0146] Sustituciones similares son aquellas en las que al menos un resto de la secuencia de aminoácidos se ha eliminado y se ha insertado en su lugar un resto diferente. Dichas sustituciones en general se hacen de acuerdo con la tabla 5 cuando se desea mantener la actividad de la proteína. La tabla 5 muestra los aminoácidos que se pueden sustituir por un aminoácido en una proteína y que se consideran normalmente como sustituciones estructurales y funcionales. Por ejemplo, un resto en la columna 1 de la tabla 5 se puede sustituir por un resto en la columna 2; además, un resto en la columna 2 de la tabla 5 se puede sustituir por el resto de la columna 1.

10

Tabla 5

Resto	Sustituciones conservativas
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

5

- [0147]** Se pueden seleccionar sustituciones que son menos conservativas que las de la tabla 5 escogiendo restos que difieren de forma más significativa en su efecto en el mantenimiento de (a) la estructura de la cadena principal del polipéptido en la zona de la sustitución, por ejemplo, como una conformación de lámina o de hélice, (b) la carga o hidrofobia de la molécula en el sitio diana, o (c) el volumen de la cadena lateral. Las sustituciones que se espera que en general produzcan cambios más grandes en las propiedades de la proteína serán aquellas en las que (a) un resto hidrófobo, por ejemplo, leucilo, isoleucilo, fenilalanilo, valilo o alanilo, se sustituye por un resto hidrófilo, por ejemplo, serilo o treonilo, (o a la inversa); (b) cualquier otro resto se sustituye por una cisteína o prolina (o a la inversa); (c) un resto electronegativo, por ejemplo, glutamilo o aspartilo, se sustituye por un resto que tiene una cadena lateral electropositiva, por ejemplo, lisilo, arginilo o histidilo (o a la inversa); o (d) un resto que no tiene cadena lateral, por ejemplo, glicina, se sustituye por uno que tiene una cadena lateral voluminosa, por ejemplo, fenilalanina (o a la inversa).

Modificación adicional de las secuencias de la invención - mutación/evolución forzada

- 20 **[0148]** Además de generar sustituciones silenciosas o conservativas como se ha indicado antes, la presente invención incluye opcionalmente procedimientos para modificar las secuencias de la lista de secuencias. En los procedimientos, se usan procedimientos de modificación de ácidos nucleicos o proteínas para alterar las secuencias dadas para producir más secuencias y/o para modificar química o enzimáticamente secuencias dadas para cambiar las propiedades de los ácidos nucleicos o proteínas.

25

- [0149]** Por lo tanto, en una realización, las secuencias de ácidos nucleicos dadas se modifican, por ejemplo, de acuerdo con la mutagénesis estándar o procedimientos de evolución artificial para producir secuencias

modificadas. Las secuencias modificadas se pueden crear usando polinucleótidos naturales purificados aislados de cualquier organismo o se pueden sintetizar a partir de composiciones y productos químicos purificados usando medios químicos bien conocidos para el experto en la materia. Por ejemplo, Ausubel, véase antes, proporciona detalles adicionales de los procedimientos de mutagénesis. Los procedimientos de evolución forzada artificial se describen, por ejemplo, en Stemmer (1994) *Nature* 370: 389-391, Stemmer (1994) *Proc. Natl. Acad. Sci.* 91: 10747-10751, y patentes de EE.UU. 5.811.238, 5.837.500, y 6.242.568. Se describen procedimientos para diseñar factores de transcripción sintéticos y otros polipéptidos, por ejemplo, en Zhang y col. (2000) *J. Biol. Chem.* 275: 33850-33860, Liu y col. (2001) *J. Biol. Chem.* 276: 11323-11334, e Isalan y col. (2001) *Nature Biotechnol.* 19: 656-660. También están disponibles muchos otros procedimientos de mutación y de evolución y se basará en el criterio del experto.

[0150] Igualmente, se puede llevar a cabo la alteración química o enzimática de ácidos nucleicos expresados y polipéptidos mediante procedimientos estándar. Por ejemplo, una secuencia se puede modificar por adición de lípidos, azúcares, péptidos, compuestos orgánicos o inorgánicos, por la inclusión de nucleótidos o aminoácidos modificados, o similares. Por ejemplo, se ilustran técnicas de modificación de proteínas en Ausubel, véase antes. En el presente documento se pueden encontrar más detalles sobre las modificaciones químicas y enzimáticas. Estos procedimientos de modificación se pueden usar para modificar cualquier secuencia dada, o para modificar cualquier secuencia producida por los diferentes procedimientos de modificación por mutación y evolución artificial indicados en el presente documento.

[0151] Por consiguiente, la invención proporciona la modificación de cualquier ácido nucleico dado por mutación, evolución, modificación química o enzimática, u otros procedimientos disponibles, así como los productos producidos por la práctica de dichos procedimientos, por ejemplo, usando las secuencias del presente documento como un sustrato de partida para los diferentes procedimientos de modificación.

[0152] Por ejemplo, se puede usar una secuencia codificante optimizada que contiene codones preferidos por un hospedador procarionta o eucariota particular, por ejemplo, para aumentar la velocidad de traducción o para producir transcritos de ARN recombinante que tienen las propiedades deseadas, tales como una semivida más prolongada, comparado con los transcritos producidos usando una secuencia no optimizada. Los codones de parada de la traducción también se pueden modificar 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 las plantas monocotiledóneas es TGA, mientras que los insectos y *E. coli* prefieren usar TAA como codón de parada.

[0153] Las secuencias de polinucleótidos de la presente invención también se pueden diseñar con el fin de alterar una secuencia codificante por una variedad de razones, incluyendo, pero sin limitar, alteraciones que modifican la secuencia para facilitar la clonación, procesamiento y/o expresión del producto génico. Por ejemplo, se introducen opcionalmente alteraciones usando técnicas que son conocidas en la materia, por ejemplo mutagénesis dirigida, para insertar nuevos sitios de restricción, para alterar patrones de glicosilación, para cambiar preferencias de codones, para introducir sitios de corte y empalme, etc.

[0154] Además, un fragmento o dominio derivado de cualquiera de los polipéptidos de la invención se puede combinar con dominios derivados de otros factores de transcripción o dominios sintéticos para modificar la actividad biológica de un factor de transcripción. Por ejemplo, un dominio de unión al ADN derivado de un factor de transcripción de la invención se puede combinar 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 a partir de un sitio de unión al ADN. Los ejemplos incluyen la región de activación de la transcripción de VP16 o GAL4 (Moore y col. (1998) *Proc. Natl. Acad. Sci.* 95: 376-381; Aoyama y col. (1995) *Plant Cell* 7: 1773-1785), péptidos derivados de secuencias bacterianas (Ma y Ptashne (1987) *Cell* 51: 113-119) y péptido sintéticos (Giniger y Ptashne (1987) *Nature* 330: 670-672).

Expresión y modificación de polipéptidos

[0155] Típicamente, las secuencias de polinucleótidos de la invención se incorporan en moléculas de ADN (o ARN) recombinantes que dirigen la expresión de polipéptidos de la invención en células hospedadoras adecuadas, plantas transgénicas, sistemas de traducción in vitro, o similares. Debido a la degeneración inherente del código genético, cualquier secuencia listada se puede sustituir por secuencias de ácidos nucleicos que codifican sustancialmente la misma secuencia de aminoácidos o una funcionalmente equivalente, para proporcionar la clonación y expresión del homólogo pertinente.

[0156] Las plantas transgénicas de la presente invención que comprenden secuencias de polinucleótidos recombinantes en general derivan de plantas parentales, que pueden ser ellas mismas plantas no transformadas (o no transgénicas). Estas plantas transgénicas pueden tener un gen del factor de transcripción "inactivado" (por ejemplo, con una inserción genómica por recombinación homóloga, una construcción no codificante o de ribozima) o expresado en una extensión normal o silvestre. Sin embargo, las plantas "descendientes" transgénicas que expresan en exceso, presentarán mayores niveles de ARNm, en las que el ARNm codifica un factor de transcripción, es decir, una proteína de unión al ADN que es capaz de unirse a una secuencia reguladora de ADN y que incluye la transcripción, y preferiblemente, la expresión de un gen del rasgo de la planta. Preferiblemente, el nivel de expresión del ARNm será al menos 3 veces mayor que el de la planta parental, o más preferiblemente al menos 10 veces mayor que los niveles de ARNm comparados con dicha planta parental, y lo más preferiblemente al menos 50 veces mayor comparado con dicha planta parental.

Vectores, promotores y sistemas de expresión

[0157] La presente invención incluye construcciones recombinantes que comprenden una o más secuencias de ácidos nucleicos del presente documento. Las construcciones típicamente comprenden un vector, tal como un plásmido, un cósmido, un fago, un virus (por ejemplo un virus de planta), un cromosoma artificial bacteriano (BAC), un cromosoma artificial de levadura (YAC), o similares, en el que se habrá insertado una secuencia de ácido nucleico de la invención, en una orientación directa o inversa. En un aspecto preferido de esta realización, la construcción comprende además secuencias reguladoras, incluyendo, por ejemplo, un promotor operativamente unido a la secuencia. Los expertos en la materia conocen un gran número de vectores y promotores adecuados, y están disponibles en el comercio.

[0158] Los textos generales que describen técnicas de biología molecular útiles en el presente documento, incluyendo el uso y la producción de vectores, promotores y muchos otros temas relevantes, incluyen Berger, Sambrook, véase antes y Ausubel, véase antes. Cualquiera de las secuencias identificadas se puede incorporar en un casete o vector, por ejemplo, para la expresión en plantas. Se ha descrito una serie de vectores de expresión adecuados para la transformación estable de células de plantas o para el establecimiento de plantas transgénicas, incluyendo los descritos en Weissbach y Weissbach (1989) "Methods for Plant Molecular Biology" Academic Press, y Gelvin y col. (1990) "Plant Molecular Biology Manual", Kluwer Academic Publishers. Los ejemplos específicos incluyen los derivados de un plásmido Ti de *Agrobacterium tumefaciens*, así como los descritos en Herrera- Estrella y col. (1983) *Nature* 303: 209, Bevan (1984) *Nucleic Acids Res.* 12: 8711-8721, Klee (1985) *Bio/Technology* 3: 637-642, para plantas dicotiledóneas.

[0159] Alternativamente, se pueden usar vectores no-Ti para transferir el ADN a plantas monocotiledóneas y células usando técnicas de suministro de ADN libre. Dichos procedimientos pueden implicar, por ejemplo, el uso de liposomas, electroporación, bombardeo de microproyectiles, filamentos de carburo de silicio y virus. Usando estos procedimientos se pueden producir plantas transgénicas tales como trigo, arroz (Gordon-Kamm (1990) *Plant Cell* 2: 603-618). Un embrión inmaduro también puede ser un buen tejido diana para monocotiledóneas para las técnicas de suministro directo de ADN usando la pistola de partículas (Weeks y col. (1993) *Plant Physiol.* 102: 1077-1084; Vasil (1993) *Bio/Technology* 10: 667-674; Wan y Lemeaux (1994) *Plant Physiol.* 104: 37-48, y para la transferencia de ADN mediada por *Agrobacterium* (Ishida y col. (1996) *Nature Biotechnol.* 14: 745-750).

[0160] Típicamente, los vectores de transformación de plantas incluyen una o más secuencias codificantes clonadas de la planta (genómicas o ADNc) bajo el control transcripcional de secuencias reguladoras 5' y 3' y un marcador seleccionable dominante. Dichos vectores de transformación de plantas típicamente contienen también un promotor (por ejemplo, una región reguladora que controla la expresión inducible o constitutiva, regulada por el medio ambiente o el desarrollo, específica de célula o tejido), un sitio de inicio de la transcripción, una señal de procesamiento del ARN (tal como sitios de corte y empalme, intrones), un sitio de terminación de la transcripción y/o una señal de poliadenilación.

[0161] Una utilidad potencial de los polinucleótidos de los factores de transcripción descritos en el presente documento es el aislamiento de elementos promotores de estos genes, que se pueden usar para programar la expresión en plantas de cualesquiera genes. Cada gen de factor de transcripción descrito en el presente documento se expresa de una forma única, determinada por los elementos promotores localizados en la dirección 5' del inicio de la traducción, y además dentro de un intrón del gen del factor de transcripción o en la dirección 3' del codón determinación del gen. Como se sabe en la técnica, para una parte significativa de los genes, las secuencias promotoras están situadas totalmente en la región directamente en la dirección 5' del inicio de la traducción. En dichos casos, típicamente las secuencias promotoras están situadas dentro de 2,0 kb del inicio de la traducción, o

dentro de 1,5 kb del inicio de la traducción, con frecuencia dentro de 1,0 kb del inicio de la traducción, y a veces dentro de 0,5 kb del inicio de la traducción.

[0162] Las secuencias promotoras se pueden aislar de acuerdo con procedimientos conocidos para el experto en la materia.

[0163] Los ejemplos de promotores de plantas constitutivos que pueden ser útiles para expresar la secuencia de TF 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 de plantas (por ejemplo, en Odell y col. (1985) *Nature* 313: 810-812); el promotor de la nopalina sintasa (An y col. (1988) *Plant Physiol.* 88: 547-552); y el promotor de la octopina sintasa (Fromm y col. (1989) *Plant Cell* 1: 977-984).

[0164] Los factores de transcripción de la invención pueden estar operativamente unidos a un promotor específico que hace que el factor de transcripción sea expresado en respuesta a señales medioambientales, específicas de tejido o temporales. Se puede usar una variedad de promotores de genes de plantas que regulan la expresión de genes en respuesta a señales medioambientales, hormonales, químicas, de desarrollo, y de una forma activada por tejido, para la expresión de una secuencia de TF 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, fruta, raíz, polen, tejido vascular, flor, carpelo, etc.), capacidad de inducción (por ejemplo, en respuesta a heridas, calor, frío, sequía, luz, patógenos, etc.), el tiempo, la etapa de desarrollo, y similares. Se han caracterizado numerosos promotores conocidos y se pueden usar favorablemente para promover la expresión de un polinucleótido de la invención en una planta o célula transgénica de interés. Por ejemplo, los promotores específicos de tejido incluyen: promotores específicos de semilla (tal como el promotor de napina, faseolina o DC3 descritos en la patente de EE.UU. n° 5.773.697), promotores específicos de fruto que son activos durante la maduración del fruto (tal como el promotor dru 1 (patente de EE.UU. n° 5.783.393), o el promotor 2A11 (patente de EE.UU. n° 4.943.674) y el promotor de la poligalacturonasa del tomate (Bird y col. (1988) *Plant Mol. Biol.* 11: 651-662), promotores específicos de raíz, tales como los descritos en las patentes de EE.UU. N° 5.618.988, 5.837.848 y 5.905.186, promotores activos en el polen tales como PTA29, PTA26 y PTA13 (patente de EE.UU. n° 5.792.929), promotores activos en el tejido vascular (Ringli y Keller (1998) *Plant Mol. Biol.* 37: 977-988), específicos de flores (Kaiser y col. (1995) *Plant Mol. Biol.* 28: 231-243), polen (Baerson y col. (1994) *Plant Mol. Biol.* 26: 1947-1959), carpelos (Ohl y col. (1990) *Plant Cell* 2: 837-848), polen y óvulos (Baerson y col. (1993) *Plant Mol. Biol.* 22: 255-267), promotores inducibles por auxina (tales como los descritos en van der Kop y col. (1999) *Plant Mol. Biol.* 39: 979-990 o Baumann y col., (1999) *Plant Cell* 11: 323-334), promotor inducible por citoquinas (Guevara-García (1998) *Plant Mol. Biol.* 38: 743-753), promotores que responden a la giberelina (Shi y col. (1998) *Plant Mol. Biol.* 38: 1053-1060, Willmott y col. (1998) *Plant Molec. Biol.* 38: 817-825) y similares. Son promotores adicionales los que producen expresión en respuesta al calor (Ainley y col. (1993) *Plant Mol. Biol.* 22: 13-23), la luz (por ejemplo el promotor del guisante rbcS-3A, Kuhlemeier y col. (1989) *Plant Cell* 1: 471-478), y el promotor del maíz rbcS, Schaffner y Sheen (1991) *Plant Cell* 3: 997-1012); heridas (por ejemplo, *wun1*, Siebertz y col. (1989) *Plant Cell* 1: 961-968); patógenos (tales como el promotor PR-1 descrito en Buchel y col. (1999) *Plant Mol. Biol.* 40: 387-396, y el promotor PDFI.2 descrito en Manners y col. (1998) *Plant Mol. Biol.* 38: 1071-1080), y productos químicos tales como el jasmonato de metilo o ácido salicílico (Gatz (1997) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48: 89-108). Además, el momento de la expresión se puede controlar usando promotores tales como los que actúan en la senescencia (Gan y Amasino (1995) *Science* 270: 1986-1988); o tarde en el desarrollo de la semilla (Odell y col. (1994) *Plant Physiol.* 106: 447-458).

[0165] Los vectores de expresión de plantas también pueden incluir señales de procesamiento del ARN que pueden estar situadas dentro, en dirección 5' o en dirección 3' de la secuencia codificante. Además, los vectores de expresión pueden incluir secuencias reguladoras adicionales a partir de la región 3' no traducida de genes de plantas, por ejemplo, una región 3' terminadora para aumentar la estabilidad del ARNm, tal como la región terminadora PI-II de la patata o las regiones 3' terminadoras de la octopina o nopalina sintasa.

Elementos de expresión adicionales

[0166] Las señales de inicio específicas pueden ayudar a la traducción eficaz de las secuencias codificantes. Estas señales pueden incluir, por ejemplo, el codón de inicio ATG y secuencias adyacentes. Pueden no ser necesarias señales de control de la traducción adicionales cuando se insertan una secuencia codificante, su codón de inicio y secuencias en la dirección 5' en el vector de expresión adecuado. Sin embargo, en los casos en los que sólo se inserta la secuencia codificante (por ejemplo, una secuencia codificante de proteína madura) o una parte de ella, se pueden proporcionar por separado señales de control de la transcripción exógenas incluyendo el codón de

inicio ATG. El codón de inicio se proporciona en el marco de lectura correcto para facilitar la transcripción. Los elementos transcripcionales y codones de inicio exógenos pueden ser de diferentes orígenes, tanto natural como sintético. La eficacia de la expresión se puede potenciar por la inclusión de potenciadores adecuados para el sistema celular que se usa.

5

Hospedadores de expresión

[0167] La presente invención también se refiere a células hospedadoras que se transducen con vectores de la invención, y a la producción de polipéptidos de la invención (incluyendo fragmentos de los mismos) por técnicas recombinantes. Las células hospedadoras se transforman por ingeniería genética (es decir, se introducen ácidos nucleicos, por ejemplo, se transducen, transforman o transfectan), con los vectores de esta invención, que pueden ser, por ejemplo, un vector de clonación o un vector de expresión que comprenden los ácidos nucleicos relevantes del presente documento. El vector es opcionalmente un plásmido, una partícula vírica, un fago, un ácido nucleico desnudo, etc. Las células hospedadoras transformadas por ingeniería genética se pueden cultivar en medios nutrientes convencionales modificados según sea adecuado para la activación de promotores, selección de transformantes o amplificación el gen relevante. Las condiciones de cultivo, tales como la temperatura, pH y similares, son las usadas previamente con la célula hospedadora seleccionada para la expresión, y serán evidentes para el experto en la materia y se encuentran en la referencias citadas en el presente documento, incluyendo, Sambrook, véase antes, y Ausubel, véase antes.

20

[0168] La célula hospedadora puede ser una célula eucariota, tal como una célula de levadura o una célula de planta, o la célula hospedadora puede ser una célula procariota, tal como una célula bacteriana. Los protoplastos de planta también son adecuados para algunas aplicaciones. Por ejemplo, se introducen fragmentos de ADN en tejidos de planta, células de plantas cultivadas o protoplastos de planta por procedimientos estándar, que incluyen la electroporación (Fromm y col. (1985) *Proc. Natl. Acad. Sci.* 82: 5824-5828), infección por vectores víricos tales como el virus del mosaico de la coliflor (CaMV) (Hohn y col., (1982) "Molecular Biology of Plant Tumors" Academic Press, Nueva York, NY, pág. 549-560; documento US 4.407.956), penetración balística de alta velocidad mediante partículas pequeñas con el ácido nucleico dentro de la matriz de perlas o partículas pequeñas o en la superficie (Klein y col. (1987) *Nature* 327: 70-73), uso de polen como vector (documento WO 85/01856), o uso de *Agrobacterium tumefaciens* o *A. rhizogenes* que lleva un plásmido de ADN-T en el que se clona fragmentos de ADN. El plásmido ADN-T se transmite a las células de la planta por infección mediante *Agrobacterium tumefaciens*, y una parte se integra establemente en el genoma de la planta (Horsch y col. (1984) *Science* 233: 496-498; Fraley y col. (1983) *Proc. Natl. Acad. Sci.* 80: 4803-4807).

[0169] La célula puede incluir un ácido nucleico de la invención que codifica un polipéptido, en el que la célula expresa un polipéptido de la invención. La célula también puede incluir secuencias de vectores, o similares. Además, las células y plantas transgénicas que incluyen cualquier polipéptido o ácido nucleico anterior o a lo largo de esta memoria descriptiva, por ejemplo, producidos por transducción de un vector de la invención, son una característica adicional de la invención.

40

[0170] Para la producción de alto rendimiento y a largo plazo de proteínas recombinantes, se puede usar la 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 producida por una célula recombinante puede ser secretada, estar unida a membrana, o contenida intracelularmente, dependiendo de la secuencia y/o el vector usados. Como entenderá el experto en la materia, los vectores de expresión que contienen polinucleótidos que codifican proteínas maduras de la invención, se pueden diseñar con secuencias señal que dirigen la secreción de los polipéptidos maduros a través de la membrana celular procariota o eucariota.

50 Restos de aminoácidos modificados

[0171] Los polipéptidos de la invención pueden contener uno o más restos de aminoácidos modificados. La presencia de aminoácidos modificados puede ser ventajosa, por ejemplo, para aumentar la semivida del polipéptido, reducir la antigenicidad o toxicidad del polipéptido, aumentar la estabilidad en el almacenamiento del polipéptido, o similares. El o los restos de aminoácidos se modifican, por ejemplo, simultáneamente con la traducción o después de la traducción durante la producción recombinante o se modifican por medios sintéticos o químicos.

[0172] Los ejemplos no limitantes de un resto de aminoácido modificado incluyen la incorporación u otro uso de aminoácidos acetilados, aminoácidos glicosilados, aminoácidos sulfatados, aminoácidos prenilados (por ejemplo,

farnesilados, geranilgeranilados), aminoácidos modificados con PEG (por ejemplo, "PEGilados"), aminoácidos biotinilados, aminoácidos carboxilados, aminoácidos fosforilados, etc. La bibliografía está llena de referencias adecuadas para guiar al experto en la modificación de los restos de aminoácidos.

- 5 **[0173]** Los restos de aminoácidos modificados pueden prevenir o aumentar la afinidad del polipéptido por otra molécula, incluyendo, pero sin limitar, polinucleótido, proteínas, hidratos de carbono, lípidos y derivados de lípidos, y otros compuestos orgánicos o sintéticos.

Identificación de factores de proteínas adicionales

10

- [0174]** Un factor de transcripción proporcionado por la presente invención también se puede usar para identificar moléculas endógenas o exógenas adicionales que pueden afectar al fenotipo o a un rasgo de interés. Dichas moléculas incluyen moléculas endógenas que actúan a un nivel transcripcional mediante un factor de transcripción de la invención para modificar un fenotipo como se desea. Por ejemplo, los factores de transcripción se pueden usar para identificar uno o más genes en la dirección 3' que se someten a un efecto regulador del factor de transcripción. En un procedimiento, un factor de transcripción u homólogo del factor de transcripción de la invención se expresa en una célula hospedadora, por ejemplo, una célula de planta transgénica, tejido o explante, y se hace el seguimiento de los productos de expresión, ARN o proteína, de dianas probables o aleatorias, por ejemplo, por hibridación con una micromatriz de sondas de ácidos nucleicos correspondientes a los genes expresados en un tejido o tipo de célula de interés, mediante electroforesis en gel bidimensional de los productos proteínicos, o por cualquier otro procedimiento conocido en la técnica para evaluar la expresión de productos génicos a nivel del ARN o proteína. Alternativamente, se puede usar un factor de transcripción de la invención para identificar secuencias promotoras (tales como sitios de unión en secuencias de ADN) implicadas en la regulación de una diana en la dirección 3'. Después de identificar una secuencia promotora, se pueden modificar las interacciones entre el factor de transcripción y la secuencia promotora mediante cambios en nucleótidos específicos en la secuencia promotora o aminoácidos específicos en el factor de transcripción que interacciona con la secuencia promotora, para alterar un rasgo de la planta. Típicamente, los sitios de unión al ADN de factores de transcripción se identifican mediante ensayos de desplazamiento en gel. Después de identificar las regiones promotoras, las secuencias de la región promotora se pueden usar en matrices de ADN de doble cadena para identificar moléculas que afectan a las interacciones de los factores de transcripción con sus promotores (Bulyk y col. (1999) *Nature Biotechnol.* 17: 573-577).

- [0175]** Los factores de transcripción identificados también son útiles para identificar proteínas que modifican la actividad del factor de transcripción. Dicha modificación se puede producir por modificación covalente, tal como por fosforilación o por interacciones de proteína-proteína (homo o heteropolímero). Se puede usar cualquier procedimiento adecuado para detectar interacciones de proteína-proteína. Entre los procedimientos que se pueden usar están la inmunoprecipitación simultánea, entrecruzamiento y purificación simultánea a través de gradientes o columnas cromatográficas, y el sistema de doble híbrido en levaduras.

- 40 **[0176]** El sistema de doble híbrido detecta interacciones de proteínas in vivo y se describen en Chien y col. ((1991) *Proc. Natl. Acad. Sci.* 88: 9578-9582) y está disponible en el comercio en Clontech (Palo Alto, Calif.). En dicho sistema, se construyen plásmidos que codifican dos proteínas híbridas: una consiste en el dominio de unión al ADN de una proteína activadora de la transcripción fusionada con el polipéptido del TF y la otra consiste en el dominio de activación de la proteína activadora de la transcripción fusionado con una proteína desconocida que es codificada 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 al ADN y la biblioteca de ADNc se transforman en una cepa de levaduras de *Saccharomyces cerevisiae* que contiene un gen indicador (por ejemplo, lacZ) cuya región reguladora contiene el sitio de unión al activador de la transcripción. Cualquiera de las proteínas híbridas 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 de expresión del gen indicador, que se detecta mediante un ensayo para el producto del gen indicador. Después, los plásmidos de la biblioteca responsables de la expresión del gen indicador se aíslan y se secuencian para identificar las proteínas codificadas por los plásmidos de la biblioteca. Después de identificar las proteínas que interactúan con los factores de transcripción, se pueden llevar a cabo ensayos para los compuestos que interfieren con las interacciones de proteína-proteína del TF.

55

Subsecuencias

- [0177]** También se contemplan usos de los polinucleótidos, denominados también en el presente documento como oligonucleótidos, que típicamente tienen al menos 12 o más bases que hibridan en condiciones rigurosas o

muy rigurosas con una secuencia de polinucleótido descrita antes. Los polinucleótidos se pueden usar como sondas, cebadores, agentes de sentido directo y sentido contrario, y similares, de acuerdo con los procedimientos indicados antes.

- 5 **[0178]** Las subsecuencias de los polinucleótidos de la invención, incluyendo fragmentos de polinucleótido y oligonucleótidos, son útiles como sondas de ácidos nucleicos y cebadores. Un oligonucleótido adecuado para usar como sonda o cebador tiene al menos aproximadamente 15 nucleótidos de longitud, más a menudo al menos aproximadamente 18 nucleótidos, a menudo al menos aproximadamente 21 nucleótidos, con frecuencia al menos aproximadamente 30 nucleótidos, o aproximadamente 40 nucleótidos, o más de longitud. Una sonda de ácido
10 nucleico es útil en protocolos de hibridación, por ejemplo, para identificar homólogos de polipéptidos adicionales de la invención, incluyendo protocolos para experimentos de micromatrices. Los cebadores se pueden volver a asociar con una cadena de ADN diana complementaria mediante hibridación de ácido nucleico para formar un híbrido entre el cebador y la cadena de ADN diana, y después extender a lo largo de la cadena de ADN diana mediante una enzima ADN polimerasa. Se pueden usar parejas de cebadores para la amplificación de una secuencia de ácido
15 nucleico, por ejemplo, mediante la reacción en cadena de la polimerasa (PCR) u otros procedimientos de amplificación de ácidos nucleicos (Sambrook, véase antes, Ausubel, véase antes).

- [0179]** Además, la invención incluye un polipéptido aislado o recombinante que incluye una subsecuencia de al menos aproximadamente 15 aminoácidos contiguos codificados por los polinucleótidos recombinantes o aislados
20 de la invención. Por ejemplo, dichos polipéptidos, o dominios o fragmentos de los mismos, se pueden usar como inmunógenos, por ejemplo, para producir anticuerpos específicos para la secuencia de polipéptido, o como sondas para detectar una secuencia de interés. El tamaño de una subsecuencia puede estar en el intervalo de aproximadamente 15 aminoácidos de longitud hasta e incluyendo la longitud completa del polipéptido.

- 25 **[0180]** Para que lo abarque la presente invención, un polipéptido expresado que comprende dicha subsecuencia de polipéptido realiza al menos una función biológica del polipéptido intacto sustancialmente de la misma forma, o en una extensión similar, a como lo hace el polipéptido intacto. Por ejemplo, un fragmento de polipéptido puede comprender un patrón estructural reconocible o dominio funcional tal como un dominio de unión al
30 ADN que activa la transcripción, por ejemplo, mediante unión a una región promotora de ADN específica de un dominio de activación, o un dominio para interacciones de proteína-proteína.

Producción de plantas transgénicas

Modificación de rasgos

- 35 **[0181]** Los polinucleótidos de la invención se usan de forma favorable para producir plantas transgénicas con diferentes rasgos o características que se han modificado de una forma conveniente, por ejemplo, para mejorar las características de las semillas de una planta. Por ejemplo, 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
40 factores de transcripción) de la invención, comparado con los niveles de la misma proteína encontrados en una planta silvestre, se puede usar para modificar rasgos de una planta. Un ejemplo ilustrativo de la modificación de rasgos, características mejoradas, mediante la alteración de los niveles de expresión de un factor de transcripción particular se describe a continuación en los ejemplos y la lista de secuencias.

45 *Arabidopsis* como un sistema modelo

- [0182]** *Arabidopsis thaliana* es objeto de una atención que crece rápidamente como un modelo para la genética y el metabolismo en plantas. *Arabidopsis* tiene un genoma pequeño y están disponibles estudios bien documentados. Es fácil cultivarla en grandes cantidades y están disponibles mutantes que definen mecanismos
50 genéticamente controlados importantes, o se pueden obtener fácilmente. Están disponibles diferentes procedimientos para introducir y expresar genes homólogos aislados (Koncz y col., eds., "Methods in Arabidopsis Research" (1992) World Scientific, Nueva Jersey, NJ, en "Prólogo"). Debido a su pequeño tamaño, ciclo de vida corto, autogamia obligada y alta fertilidad, *Arabidopsis* también es un organismo de elección para el aislamiento de mutantes y estudios en rutas morfogénicas y de desarrollo, y el control de estas rutas por factores de transcripción
55 (Koncz (1992), véase antes, pág. 72). Una serie de estudios que introducen factores de transcripción en *A. thaliana* han demostrado la utilidad de esta planta para la comprensión de los mecanismos de regulación de genes y alteración de rasgos en plantas (por ejemplo, en Koncz (1992) véase antes, y en la patente de EE.UU. nº 6.417.428).

Genes de *Arabidopsis* en plantas transgénicas

[0183] La expresión de genes que codifican factores de transcripción que modifican la expresión de genes endógenos, polinucleótidos y proteínas son bien conocidos en la materia. 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 y col. (1997) *Genes and Development* 11: 3194-3205 y Peng y col. (1999) *Nature* 400: 256-261. Además, muchos otros han demostrado que un factor de transcripción de *Arabidopsis* expresado en una especie de planta exógena produce la misma respuesta fenotípica o muy similar (Fu y col. (2001) *Plant Cell* 13: 1791-1802; Nandi y col. (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

[0184] Se pueden introducir en plantas genes homólogos que se pueden obtener de cualquier planta, o de cualquier fuente sea natural, sintética o semisintética o recombinante, y que comparte una identidad o similitud de secuencia significativa con las proporcionadas por la presente invención, por ejemplo, en plantas cultivadas, para conferir los rasgos mejorados o deseados. Por consiguiente, se pueden producir plantas transgénicas que comprenden un vector o casete de expresión recombinante con un promotor operativamente unido a una o más secuencias homólogas con las presentes secuencias descritas. El promotor puede ser, por ejemplo, un promotor de planta o vírico.

[0185] La invención proporciona, por lo tanto, procedimientos para preparar plantas transgénicas y para modificar rasgos de las plantas. Estos procedimientos incluyen introducir en una planta un vector o casete de expresión recombinante que comprende un promotor funcional operativamente unido a una o más secuencias homólogas con las presentes secuencias descritas. Las plantas y los kits para producir estas plantas que resultan de la aplicación de estos procedimientos también están incluidos en la presente invención.

Factores de transcripción de interés para la modificación de rasgos de plantas

[0186] Actualmente, la existencia de una serie de grupos de maduración para diferentes latitudes representa una barrera importante para la producción de nuevos rasgos valiosos. Cualquier rasgo (por ejemplo, resistencia a una enfermedad) debe cultivarse en cada uno de los diferentes grupos de maduración por separado, un ejercicio laborioso y costoso. Por lo tanto, la disponibilidad de una sola cepa, que podría cultivarse a cualquier latitud, aumentaría mucho el potencial para introducir nuevos rasgos en especies de cultivo tales como la soja y el algodón.

[0187] Para los efectos, rasgos y utilidades específicos conferidos a las plantas, se puede usar uno o más de los genes de los factores de transcripción de la presente invención, para aumentar o disminuir, o mejorar o demostrar pernicioso para un rasgo dado. Por ejemplo, la inactivación de un gen de un factor de transcripción que se produce de forma natural en una planta, o la eliminación del gen (con, por ejemplo, eliminación de sentido contrario), puede producir una disminución de la tolerancia a la sombra o al estrés por sequía con respecto a las plantas no transformadas o silvestres. Mediante el exceso de expresión de este gen, la planta puede experimentar un aumento de la tolerancia a estos estreses. Se puede introducir 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 dar cruces de híbridos que comprenden más de un factor de transcripción introducido.

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

[0188] Los factores de transcripción de la planta puede modular la expresión de genes, y a su vez, ser modulados por la experiencia medioambiental de la planta. Las alteraciones significativas del entorno de una planta producen de forma invariable un cambio en el patrón de expresión de los genes de los factores de transcripción de la planta. Los patrones de expresión de los factores de transcripción alterados en general dan como resultado cambios fenotípicos en la planta. El o los productos génicos de factores de transcripción en las plantas transgénicas difieren entonces en las cantidades o proporciones de las encontradas en las plantas silvestres o no transformadas, y es probable que estos factores de transcripción representen polipéptidos que son usados para alterar la respuesta al cambio medioambiental. A modo de ejemplo, está aceptado en la técnica que los procedimientos analíticos basados en patrones de expresión alterados se pueden usar para cribar cambios fenotípicos en una planta de forma mucho más eficaz que la lograda usando procedimientos tradicionales.

[0189] Potenciales aplicaciones de las presentes secuencias descritas que regulan la tolerancia al estrés

abiótico

[0190] Sensores de azúcares. Además de su importante función como una fuente de energía y componente estructural de la célula de la planta, los azúcares son moléculas reguladoras centrales que controlan varios aspectos de la fisiología, metabolismo y desarrollo de la planta (Hsieh y col. (1998) *Proc. Natl. Acad. Sci.* 95: 13965-13970). Se cree que este control se logra mediante la regulación de la expresión de genes y, en plantas superiores, se ha mostrado que los azúcares reprimen o activan los genes de la planta implicados en muchos procesos esenciales tales como la fotosíntesis, metabolismo del glioxilato, respiración, síntesis y degradación de almidón y sacarosa respuesta a patógenos, respuesta a heridas, regulación del ciclo celular, pigmentación, florecimiento y senescencia. Los mecanismos por los cuales los azúcares controlan la expresión de genes no se comprenden.

[0191] Varios mutantes de la detección de azúcares han resultado ser alélicos de ABA y mutantes de etileno. ABA se encuentra en todos los organismos fotosintéticos y actúa como un regulador clave en la transcripción, respuestas al estrés, embriogénesis, y germinación de semillas. La mayoría de los efectos de ABA están relacionados con el compuesto que actúa como una señal de la menor disponibilidad de agua, por lo que produce una reducción en la pérdida de agua, ralentiza el crecimiento y media las respuestas adaptativas. Sin embargo, ABA también influye en el crecimiento y desarrollo de la planta mediante interacciones con otras fitohormonas. Estudios fisiológicos y moleculares indican que el maíz y *Arabidopsis* tienen rutas casi idénticas con respecto a la biosíntesis de ABA y la transducción de señales (por ejemplo, en Finkelstein y Rock (2002) "Abscisic acid biosynthesis and response", en *The Arabidopsis Book*, Somerville y Meyerowitz, editores (American Society of Plant Biologists, Rockville, MD).

[0192] Esto implica potencialmente que cuando las secuencias de la invención son expresadas en exceso, confieren un fenotipo de sensores de azúcares o señalización de hormonas en las plantas. Por otra parte, el tratamiento de sacarosa usado en estos experimentos (al 9,4% en p/v) también podría ser un estrés osmótico. Por lo tanto, se podrían interpretar estos datos como una indicación de que en estas líneas transgénicas son más tolerantes al estrés osmótico. Sin embargo, es bien conocido que las respuestas de la planta a ABA, estrés osmótico y otros pueden estar conectadas, y éstos diferentes tratamientos pueden incluso actuar de una forma sinérgica para aumentar el grado de una respuesta. Por ejemplo, Xiong, Ishitani, y Zhu ((1999) *Plant Physiol.* 119: 205-212) han mostrado que se pueden usar estudios moleculares y genéticos para mostrar la interacción extensiva entre el estrés osmótico, el estrés por temperatura y las respuestas al ABA en plantas. Estos investigadores analizaron la expresión de *RD29A-LUC* en respuesta a diferentes regímenes de tratamiento en *Arabidopsis*. El promotor *RD29A* contiene tanto elemento de respuesta a ABA como de respuesta a la deshidratación, denominado también la repetición C, y puede ser activado por estrés osmótico, temperatura baja o tratamiento con ABA; la transcripción del gen *RD29A* en respuesta a estreses osmóticos y por frío es mediada tanto por rutas dependientes del ABA como independientes del ABA (Xiong, Ishitani, y Zhu (1999) véase antes). *LUC* se refiere a la secuencia que codifica la luciferasa de luciérnaga, que en este caso era dirigida por el promotor *RD29A* de respuesta al estrés. Los resultados pusieron de manifiesto tanto interacciones positivas como negativas dependiendo de la naturaleza y la duración de los tratamientos. Se encontró que el estrés por temperatura baja deteriora la señalización osmótica pero el estrés por calor moderado potenciaba mucho la inducción de estrés osmótico, actuando así de forma sinérgica con las rutas de señalización osmótica. En este estudio, los autores describieron que el estrés osmótico y ABA pueden actuar de forma sinérgica mostrando que los tratamientos inducen simultáneamente expresión de transgenes y de genes endógenos. Describían resultados similares Bostock y Quatrano ((1992) *Plant Physiol.* 98: 1356-1363), quienes encontraron que el estrés osmótico y ABA actúan de forma sinérgica e inducen la expresión del gen *Em* del maíz. Ishitani y col. (1997) *Plant Cell* 9: 1935-1949) aislaron un grupo de mutaciones de un solo gen de *Arabidopsis* que conferían una mayor respuesta tanto al estrés osmótico como a ABA. La naturaleza de la recuperación de estos mutantes del estrés osmótico y del tratamiento con ABA sugería que aunque existen rutas de señalización separadas para el estrés osmótico y ABA, las rutas comparten una serie de componentes; estos componentes comunes pueden mediar interacciones sinérgicas entre el estrés osmótico y ABA. Por lo tanto, contrariamente a la creencia que se mantenía previamente de que las rutas de señalización del estrés dependientes de ABA e independientes de ABA actúan de una forma paralela, los datos de los autores de la invención ponen de manifiesto que estas rutas se superponen y convergen para activar la expresión del gen del estrés.

[0193] Como los azúcares son importantes moléculas de señalización, la capacidad para controlar la concentración de un azúcar de señalización o cómo la planta percibe o responde a un azúcar de señalización, se podría usar para controlar el desarrollo, fisiología o metabolismo de la planta. Por ejemplo, se ha mostrado que el flujo de sacarosa (un azúcar disacárido usado para el transporte sistémico de carbono y energía en la mayoría de las plantas) afecta a la expresión de genes y altera la acumulación de compuestos de almacenamiento en las semillas. Por lo tanto, la manipulación de la ruta de señalización de la sacarosa en semillas hace que las semillas

tengan más proteínas, aceite o hidratos de carbono, dependiendo del tipo de manipulación. Igualmente, en tubérculos, la sacarosa se convierte en almidón que es usado como un almacén de energía. Se cree que las rutas de señalización de azúcares pueden determinar parcialmente los niveles de almidón sintetizado en los tubérculos. La manipulación de la señalización de azúcares en tubérculos podría conducir a tubérculos con un mayor contenido de almidón.

[0194] Por lo tanto, la alteración de la expresión de los presentes genes de factores de transcripción descritos que manipulan la ruta de transducción de señales de azúcares, incluyendo, por ejemplo, G175, G303, G354, G481, G916, G922, G1069, G1073, G1820, G2053, G2701, G2789, G2839, G2854, junto con sus equivalentes, o que presentan un fenotipo de estrés osmótico, incluyendo, por ejemplo, G47, G482, G489 o G1069, G1073, como se pone de manifiesto por su tolerancia, por ejemplo, al manitol superior, sal o PEG, se puede usar para producir plantas con rasgos deseables, incluyendo una mayor tolerancia a la sequía. En particular, la manipulación de las rutas de transducción de señales de azúcares se podría usar para alterar las relaciones fuente-sumidero en semillas, tubérculos, raíces y otros órganos de almacenamiento conduciendo a un aumento del rendimiento.

[0195] Estrés abiótico: tolerancia a la sequía y la humedad baja. La exposición a la deshidratación invoca estrategias de supervivencia similares en plantas que el estrés por helada (por ejemplo, en Yelenosky (1989) *Plant Physiol* 89: 444-451) y el estrés por sequía induce tolerancia a la congelación (por ejemplo, en Siminovitch y col. (1982) *Plant Physiol* 69: 250-255; y Guy y col. (1992) *Planta* 188: 265-270). Además de la inducción de proteínas de aclimatación al frío, las estrategias que permiten a la planta sobrevivir en condiciones de poca agua pueden incluir, por ejemplo, una menor superficie específica, o producción de aceite o cera superficiales. La modificación de la expresión de los presentes genes de factores de transcripción descritos, incluyendo G2133, G1274, G922, G2999, G3086, G354, G1792, G2053, G975, G1069, G916, G1820, G2701, G47, G2854, G2789, G634, G175, G2839, G1452, G3083, G489, G303, G2992, y G682, y sus equivalentes, se puede usar para aumentar la tolerancia de una planta a condiciones de poca agua y proporcionar los beneficios de una mayor supervivencia, mayor rendimiento y zona de cultivo temporal y geográfica extendida.

[0196] Estrés osmótico. La modificación de la expresión de una serie de los presentes genes de factores de transcripción descritos, por ejemplo, G47, G482, G489 o G1069, G2053 y sus equivalentes, se puede usar para aumentar la velocidad de germinación o el crecimiento en condiciones osmóticas adversas, que podrían tener impacto en la supervivencia y rendimiento de las semillas y las plantas. Los estreses osmóticos se pueden regular mediante mecanismos de control molecular específicos que incluyen genes que controlan los movimientos de agua e iónicos, proteínas funcionales y estructurales inducidas por estrés, percepción y transducción de señales, y depuración de radicales libres, y muchos otros (Wang y col. (2001) *Acta Hort.* (ISHS) 560: 285-292). Los instigadores del estrés osmótico incluyen heladas, sequía y alta salinidad, cada uno de los cuales se discuten a continuación con más detalle.

[0197] De muchas maneras, la helada, salinidad alta y sequía tienen efectos similares en las plantas, no siendo el menor de ellos la inducción de polipéptidos comunes que responden a estos estreses diferentes. Por ejemplo, la helada es similar a la deficiencia de agua en cuanto que la helada reduce la cantidad de agua disponible para una planta. La exposición a temperaturas de helada puede conducir a la deshidratación celular ya que el agua sale de las células y forma cristales de hielo en espacios intracelulares (Buchanan y col. (2000) en "Biochemistry and Molecular Biology of Plants". American Society of Plant Physiologists, Rockville, MD). Como con la alta concentración de sales y la helada, los problemas de las plantas producidos por la poca disponibilidad de agua incluyen el estrés mecánico producido por la extracción del agua celular. Por lo tanto, la incorporación de factores de transcripción que modifican la respuesta de una planta al estrés osmótico, por ejemplo, en un cultivo o planta ornamental, puede ser útil para reducir los daños o pérdidas. Los defectos específicos producidos por helada, salinidad alta y sequía de abordan a continuación.

[0198] Relación entre la tolerancia a las sales, sequía y helada. Las plantas están sometidas a una variedad de estímulos medioambientales. Algunos de éstos, incluyendo el estrés por sequía, tienen la capacidad de impactar a la planta entera y a la disponibilidad de agua celular. Por lo tanto, no es sorprendente que las respuestas de las plantas a esta colección de estreses estén relacionadas. En una revisión reciente, Zhu indica que "la mayoría de los estudios sobre la señalización del estrés por agua se han centrado en el estrés salino principalmente porque las respuestas de la planta a las sales y a la sequía están estrechamente relacionadas y los mecanismos se superponen." (Zhu (2002) *Ann. Rev. Plant Biol.* 53: 247-273). Se han documentado muchos ejemplos de respuestas y rutas similares a este conjunto de estreses. Por ejemplo, se ha mostrado que los factores de transcripción CBF condicionan la resistencia a las sales, helada y sequía (Kasuga y col. (1999) *Nature Biotech.* 17: 287-291). El gen *rd29B* de *Arabidopsis* es inducido en respuesta tanto al estrés salino como por deshidratación, un procedimiento que

es mediado en gran medida a través del proceso de transducción de señales de ABA (Uno y col. (2000) *Proc. Natl. Acad. Sci. USA* 97:11632-11637), dando como resultado la actividad alterada de los factores de transcripción que se unen a un elemento en la dirección 5' en el promotor de *rd29B*. En *Mesembryanthemum crystallinum* (planta de hiel), Patharker y Cushman han mostrado que se induce una proteína quinasa dependiente de calcio (McCDPKI) por exposición tanto a estreses por sequía como salinos (Patharker y Cushman (2000) *Plant J.* 24: 679-691). También se mostró que la quinasa inducida por estrés fosforilaba un factor de transcripción, alterando presuntamente su actividad, aunque los niveles transcritos del factor de transcripción diana no se alteran en respuesta al estrés salino o por sequía. Igualmente, Saijo y col. Demostraron que una proteína quinasa dependiente de calmodulina inducida por sal/sequía en el arroz (*OsCDPR7*) confería una mayor tolerancia a la sal y a la sequía al arroz cuando se expresaba en exceso (Saijo y col. (2000) *Plant J.* 23: 319-327).

[0199] La exposición a la deshidratación invoca las estrategias de supervivencia similares en plantas a las del estrés por helada (por ejemplo, en Yelenosky (1989) *Plant Physiol.* 89: 444-451) y el estrés por sequía induce tolerancia a la helada (por ejemplo, en Siminovitch y col. (1982) *Plant Physiol.* 69: 250-255; y Guy y col. (1992) *Planta* 188: 265-270). Además de la inducción de proteínas de aclimatación al frío, las estrategias que permiten a las plantas sobrevivir en condiciones de poca agua pueden incluir, por ejemplo, una menor superficie específica o producción de aceite o cera superficiales.

[0200] Por consiguiente, un experto en la materia esperaría que algunas rutas implicadas en la resistencia a uno de estos estreses, y por lo tanto reguladas por un factor de transcripción individual, estuvieran también implicadas en la resistencia a otro de estos estreses, regulados por los mismos factores de transcripción u homólogos. Por supuesto, las rutas de resistencia generales están relacionadas, pero no son idénticas, y por lo tanto no todos los factores de transcripción que controlan la resistencia a uno de estos estreses controlarán la resistencia a los otros estreses. No obstante, si un factor de transcripción condiciona la resistencia a uno de estos estreses, será evidente para el experto en la materia probar la resistencia a los estreses relacionados.

[0201] Por lo tanto, los genes de la lista de secuencias, incluyendo por ejemplo, G175, G922, G1452, G1820, G2701, G2999, G3086 y sus equivalentes, que proporcionan tolerancia a las sales, se pueden usar para diseñar cultivos tolerantes a las sales y árboles que puedan florecer en suelos con un alto contenido salino o en condiciones de sequía. En particular, el aumento de la tolerancia salina durante la etapa de germinación de una planta potencia la supervivencia y el rendimiento. Los presentes genes de factores de transcripción descritos que proporcionar una mayor tolerancia salina durante la germinación, la etapa de plántulas y a lo largo de todo el ciclo de vida de la planta, tendrán un valor particular para impartir supervivencia y rendimiento en zonas en las que un cultivo particular normalmente no prosperaría.

[0202] Resumen de las características alteradas de la planta relacionadas con la sequía. Se proporcionan clados de secuencias estructural y funcionalmente relacionadas que derivan de una amplia variedad de plantas, incluyendo polinucleótidos de la lista de secuencias y sus polipéptidos codificados, fragmentos de los mismos, parálogos, ortólogos, equivalentes y fragmentos de los mismos. Se ha mostrado en el laboratorio y en experimentos de campo que estas secuencias confieren fenotipos alterados de tolerancia al estrés abiótico y de tamaño en las plantas. La invención también proporciona los polipéptidos de la lista de secuencias, fragmentos de los mismos, dominios conservados de los mismos, parálogos, ortólogos, equivalentes y fragmentos de los mismos. Se ha observado que las plantas que expresan en exceso estas secuencias presentan un fenotipo de sensores de azúcares y/o son más tolerantes a una amplia variedad de estreses abióticos, incluyendo estrés por sequía y salinidad alta. Muchos de los ortólogos de estas secuencias están listados en la lista de secuencias, y debido al alto grado de similitud estructural con las secuencias de la invención, se espera que estas secuencias también funcionen para aumentar la tolerancia al estrés por sequía. La invención también abarca los complementos de los polinucleótidos. Los polinucleótidos son útiles para cribar bibliotecas de moléculas o compuestos que tienen unión específica y para crear plantas transgénicas que tienen una mayor tolerancia al estrés por sequía.

Sentido contrario y supresión simultánea

[0203] Además de la expresión de los ácidos nucleicos de la invención como sustitución de genes o ácidos nucleicos modificadores del fenotipo de la planta, los ácidos nucleicos también son útiles para la supresión de la expresión en sentido directo y sentido contrario, por ejemplo, para reducir la expresión de un ácido nucleico de la invención. Es decir, los ácidos nucleicos de la invención, o subsecuencias o secuencias de sentido contrario de las mismas, se pueden usar para bloquear la expresión de ácidos nucleicos homólogos naturales. Se conocen en la técnica una variedad de tecnologías de sentido directo o sentido contrario, 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 de sentido contrario también se describe en Crowley y col. (1985) *Cell* 43: 633-641; Rosenberg y col. (1985) *Nature* 313: 703-706; Preiss y col. (1985) *Nature* 313: 27-32; Melton (1985) *Proc. Natl. Acad. Sci.* 82:144-148; Izant y Weintraub (1985) *Science* 229: 345-352; y Kim y Wold (1985) *Cell* 42:129-138. Se conocen procedimientos adicionales de regulación de sentido contrario en la técnica. La regulación de sentido contrario se ha usado para reducir o inhibir la expresión de genes de plantas, por ejemplo, en la publicación de patente europea nº 271988. El ARN de sentido contrario se puede usar para reducir la expresión de genes para producir un cambio fenotípico visible o bioquímico en la planta (Smith y col. (1988) *Nature*, 334: 724-726; Smith y col. (1990) *Plant Mol. Biol.* 14: 369-379). En general, las secuencias de sentido directo o sentido contrario se introducen en la célula donde son opcionalmente amplificadas, por ejemplo, por transcripción. Dichas secuencias incluyen tanto secuencias de oligonucleótidos sencillas como secuencias catalíticas tales como ribozimas.

[0204] Por ejemplo, una reducción o eliminación de la expresión (es decir, una "inactivación") de un factor de transcripción o polipéptido homólogo del factor de transcripción en una planta transgénica, por ejemplo, para modificar un rasgo de la planta, se puede obtener introduciendo una construcción de sentido contrario correspondiente al polipéptido de interés como un ADNc. Para la supresión de sentido contrario, el ADNc del factor de transcripción u homólogo se dispone en orientación contraria (con respecto a la secuencia codificante) con respecto a la secuencia del promotor 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 que se encuentra en el tipo de planta que se va a transformar. Típicamente, solo es necesario que la secuencia de sentido contrario sea capaz de hibridar con el gen o ARN diana de interés. Por lo tanto, cuando la secuencia introducida es de menor longitud, será necesario un mayor grado de homología con la secuencia del factor de transcripción endógeno para la supresión de sentido contrario eficaz. Aunque se pueden usar secuencias de sentido contrario de diferentes longitudes, preferiblemente, la secuencia de sentido contrario introducida en el vector será de al menos 30 nucleótidos de longitud, y típicamente se observará una mejor supresión de sentido contrario al aumentar la longitud de la secuencia de sentido contrario. Preferiblemente, la longitud de la secuencia de sentido contrario en el vector tendrá más de 100 nucleótidos. La transcripción de una construcción de sentido contrario como se describe, da como resultado la producción de moléculas de ARN que son el complemento inverso de las moléculas de ARNm transcritas a partir del gen del factor de transcripción endógeno en la célula de la planta.

[0205] La supresión de la expresión del gen del factor de transcripción endógeno también se puede lograr usando la interferencia de ARN, o ARNi. La ARNi es una técnica de silenciamiento de genes dirigida, postranscripcional, que usa ARN bicatenario (ARNdc) para incitar la degradación del ARN mensajero (ARNm) que contiene la misma secuencia que el ARNdc (Constans, (2002) *The Scientist* 16:36). Los ARN interferentes pequeños, o ARNip, se producen en al menos 2 etapas: una ribonucleasa endógena escinde el ARNdc más largo en ARN más cortos de 21-23 nucleótidos de longitud. Los segmentos de ARNip median entonces la degradación del ARNm diana (Zamore, (2001) *Nature Struct. Biol.*, 8:746-50). La ARNi se ha usado para la determinación de la función de genes de una forma similar a los oligonucleótidos de sentido contrario (Constans, (2002) *The Scientist* 16:36). Los vectores de expresión que expresan continuamente ARNip en células transfectadas transitoria y establemente se han diseñado para expresar ARN en horquilla pequeños (ARNhp), que son procesados in vivo en moléculas de tipo ARNip capaces de llevar a cabo el silenciamiento específico de genes (Brummelkamp y col., (2002) *Science* 296:550-553, y Paddison, y col. (2002) *Genes & Dev.* 16:948-958). El silenciamiento de genes postranscripcionales por el ARN bicatenario se discute con más detalle en Hammond y col. (2001) *Nature Rev Gen* 2: 110-119, Fire et, al. (1998) *Nature* 391: 806-811 y Timmons y Fire (1998) *Nature* 395: 854. Los vectores en las que el ARN codificado por un ADNc de factor de transcripción u homólogo del factor de transcripción es expresado en exceso, también se pueden usar para obtener la supresión simultánea de un gen endógeno correspondiente, por ejemplo de la forma descrita en la patente de EE.UU. nº 5.231.020 de Jorgensen. Dicha supresión simultánea (denominada también supresión de sentido contrario), no requiere que se introduzca el ADNc del factor de transcripción entero en las células de la planta, ni requiere que la secuencia introducida sea exactamente igual al gen del factor de transcripción endógeno de interés. Sin embargo, como con la supresión de sentido contrario, la eficacia de la supresión aumentará al aumentar la especificidad de la hibridación, por ejemplo, cuando la secuencia introducida se alarga y/o se aumenta la similitud de secuencia entre la secuencia introducida y el gen del factor de transcripción endógeno.

[0206] 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 mutaciones sin sentido) también se pueden usar para suprimir la expresión de un factor de transcripción endógeno, reduciendo o eliminando así su actividad y modificando uno o más rasgos. Se describen procedimientos para producir dichas construcciones en la patente de EE.UU. nº 5.583.021. Preferiblemente, dichas construcciones se hacen introduciendo un codón de parada prematuro en el gen del factor de transcripción. Alternativamente, se puede modificar un rasgo de la planta mediante el silenciamiento de

- genes usando ARN de doble cadena (Sharp (1999) *Genes and Development* 13: 139-141). Otro procedimiento para abolir la expresión de un gen es mediante inserción de mutagénesis usando el ADN-T de *Agrobacterium tumefaciens*. Después de generar los mutantes de inserción, los mutantes se pueden cribar para identificar aquellos que contienen la inserción en un gen del factor de transcripción u homólogo del factor de transcripción. Las plantas
- 5 que contienen un solo suceso de inserción de transgén en el gen deseado se pueden cruzar para generar plantas homocigotas para la mutación. Dichos procedimientos son bien conocidos para el experto en la materia (por ejemplo, en Koncz y col. (1992) "Methods in Arabidopsis Research". World Scientific Publishing Co. Pte. Ltd., River Edge, NJ).
- 10 **[0207]** Alternativamente, el fenotipo de una planta se puede alterar eliminando un gen endógeno, tal como un factor de transcripción u homólogo del factor de transcripción, por ejemplo, mediante recombinación homóloga (Kempin y col. (1997) *Nature* 389: 802-803).
- [0208]** Un rasgo de una planta también se puede modificar usando el sistema Cre-lox (por ejemplo, como se describe en la patente de EE.UU. n° 5.658.772). Se puede modificar el genoma de una planta para incluir un primer y un segundo sitio lox que después se ponen en contacto con una Cre recombinasa. Si los sitios lox están en la misma orientación, entonces la secuencia de ADN intermedia entre los dos sitios se escinde. Si los sitios lox están en orientación opuesta, la secuencia intermedia se invierte.
- 15 **[0209]** Los polinucleótidos y polipéptidos de esta invención también se pueden expresar en una planta en ausencia de un casete de expresión, manipulando la actividad o nivel de expresión del gen endógeno por otros medios, tales como por ejemplo, mediante la expresión ectópica de un gen mediante marcaje de activación de ADN-T (Ichikawa y col. (1997) *Nature* 390 698-701; Kakimoto y col. (1996) *Science* 274: 982-985). Este procedimiento conlleva transformar una planta con un marcador de gen que contiene múltiples potenciadores transcripcionales y una vez que se ha insertado el marcador en el genoma, queda desregulada la expresión de un gen flanqueador que codifica la secuencia. En otro ejemplo, la maquinaria transcripcional de una planta se puede modificar para que aumente los niveles de transcripción de un polinucleótido de la invención (por ejemplo, en las publicaciones PCT WO 96/06166 y WO98/53057, que describen la modificación de la especificidad de unión al ADN de proteínas de dedos de cinc cambiando aminoácidos particulares en el patrón de unión al ADN).
- 20 **[0210]** 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.
- 25 **[0211]** Las plantas transgénicas (o células de plantas o explantes de plantas o tejidos de plantas) que incorporan los polinucleótidos de la invención y/o expresan los polipéptidos de la invención se pueden producir por una variedad de técnicas bien establecidas como se ha descrito antes. Después de la construcción de un vector, lo más típico un casete de expresión, que incluye un polinucleótido que codifica, por ejemplo, un factor de transcripción u homólogo del factor de transcripción de la invención, se pueden usar técnicas estándar para introducir el polinucleótido en una planta, una célula de planta, un explante de planta o un tejido de planta de interés. Opcionalmente, la célula, explante o tejido de planta se pueden regenerar para producir una planta transgénica.
- 30 **[0212]** La planta puede ser cualquier planta superior, incluyendo plantas gimnospermas, monocotiledóneas y dicotiledóneas. Están disponibles protocolos adecuados para *Leguminosae* (alfalfa, soja, trébol, etc.), *Umbeliferae* (zanahoria, apio, chirivía), *Cruciferae* (col, rábano, colza, brócoli, etc.), *Cucurbitaceae* (melón y pepino), *Gramineae* (trigo, maíz, arroz, cebada, mijo, etc.), *Solanaceae* (patata, tomate, tabaco, pimientos, etc.), y otros cultivos diferentes (por ejemplo, en los protocolos descritos en Ammirato y col., eds., (1984) "Handbook of Plant Cell Culture-Crop Species", Macmillan Publ. Co., Nueva York, NY; Shimamoto y col. (1989) *Nature* 338: 274-276; Fromm y col. (1990) *Bio/Technol.* 8: 833-839; y Vasil y col. (1990) *Bio/Technol.* 8: 429-434).
- 35 **[0213]** La transformación y regeneración de células de plantas tanto monocotiledóneas como dicotiledóneas ahora es rutinario, y la selección de la técnica de transformación más adecuada la determinará el experto. La elección del procedimiento variará con el tipo de planta que se va a transformar; los expertos en la materia reconocerán la idoneidad de procedimientos particulares para tipos de plantas dados. Los procedimientos adecuados pueden incluir, pero sin limitar: electroporación de protoplastos de plantas; transformación mediada por liposomas; transformación mediada por polietilenglicol (PEG); transformación usando virus; microinyección de células de plantas; bombardeo con microproyectiles de células de plantas; infiltración al vacío; y transformaciones mediada por *Agrobacterium tumefaciens*. La transformación significa introducir una secuencia de nucleótidos en una planta de una forma para producir la expresión estable o transitoria de la secuencia.
- 40 45 50

[0214] Los ejemplos satisfactorios de modificación de características de las plantas mediante transformación con secuencias clonadas que sirven para ilustrar el conocimiento actual en este campo de tecnología, y que se incorporan en el presente documento por referencia, incluyen: patente de EE.UU. nº 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.

[0215] Después de la transformación, las plantas se seleccionan preferiblemente usando un marcador seleccionable dominante incorporado en el vector de transformación. Típicamente, dicho marcador conferirá resistencia a un antibiótico o herbicida a las plantas transformadoras, y la selección de los transformantes se puede llevar a cabo exponiendo las plantas a concentraciones adecuadas del antibiótico o herbicida.

[0216] Después de seleccionar las plantas transformadas y cultivarlas hasta la madurez, se identifican aquellas plantas que presentan un rasgo modificado. El rasgo modificado puede ser cualquiera de los rasgos descritos antes. Además, para confirmar que el rasgo modificado se debe a los cambios en los niveles de expresión o actividad del polipéptido o polinucleótido de la invención, se puede determinar analizando la expresión del ARNm usando transferencias Northern, RT-PCR o micromatrices, o la expresión de proteínas usando inmunotransferencias o transferencias Western o ensayos de desplazamiento en gel.

20 Sistemas integrados - identidad de secuencias

[0217] Además, la presente invención puede ser un sistema integrado, ordenador o medio de lectura 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 se puede usar para generar o identificar secuencias que cumplen cualquiera de los criterios especificados. Además, el conjunto de instrucciones se puede usar para asociar o conectar determinados beneficios opcionales, tales como características mejoradas, con una o más secuencias identificadas.

[0218] 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, Wisconsin Package Versión 10.0, tales como BLAST, FASTA, PILEUP, FINDPATTERNS o similares (GCG, Madison, WI). Se pueden buscar 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, Palo Alto, CA).

[0219] El alineamiento de secuencias para la comparación se puede llevar a cabo mediante el algoritmo de homología local de Smith y Waterman (1981) *Adv. Appl. Math.* 2: 482-489, mediante el algoritmo de alineamiento de homología de Needleman y Wunsch (1970) *J. Mol. Biol.* 48: 443-453, mediante la búsqueda por el procedimiento de similitud de Pearson y Lipman (1988) *Proc. Natl. Acad. Sci.* 85: 2444-2448, mediante implementaciones computerizadas de estos algoritmos. Después del alineamiento, típicamente se llevan a cabo las comparaciones de secuencias entre dos (o más) polinucleótidos o polipéptidos comparando las secuencias de las dos secuencias a lo largo de 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, normalmente de aproximadamente 50 a aproximadamente 200, más habitualmente de aproximadamente 100 a aproximadamente 150 posiciones contiguas. Se proporciona una descripción del procedimiento en Ausubel y col., véase antes.

[0220] Se puede usar una variedad de procedimientos para determinar las relaciones de las secuencias, incluyendo el alineamiento manual y el alineamiento y análisis de secuencias asistido por ordenador. Este último procedimiento es un procedimiento preferido en la presente invención, debido al mayor rendimiento que se logra mediante los procedimientos asistidos por ordenador. Como se ha indicado antes, están disponibles una variedad de programas de ordenador para llevar a cabo el alineamiento de secuencias, o los puede producir el experto.

[0221] Un algoritmo de ejemplo que es adecuado para determinar el porcentaje de identidad de secuencias y la similitud de secuencias es el algoritmo BLAST, que se describe en Altschul y col. (1990) *J. Mol. Biol.* 215: 403-410. El software para llevar a cabo los análisis con BLAST está disponible al público, por ejemplo, a través del National Library of Medicine's National Center for Biotechnology Information (ncbi.nlm.nih; página web (www) del National Institutes of Health US government (gov)). Este algoritmo implica identificar primero pares de secuencias con una puntuación alta (HSP) identificando palabras cortas de longitud W en la secuencia de consulta, que se corresponden o satisfacen alguna puntuación T umbral de valor positivo cuando se alinean con una palabra de la

misma longitud en una secuencia de la base de datos. T se denomina umbral de puntuación de palabras vecinas (Altschul y col., véase antes). Estos aciertos de palabras vecinas iniciales actúan como semillas para iniciar las búsquedas para encontrar HSP más largas que los contienen: los aciertos de palabras se extienden entonces en ambas direcciones a lo largo de cada secuencia hasta tan lejos como se pueda aumentar la puntuación acumulativa de alineamiento. Las puntuaciones acumulativas se calculan usando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de restos de emparejamiento; siempre >0) y N (puntuación de penalización para los restos de emparejamiento erróneo; siempre <0). Para secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabras en cada dirección se detiene cuando: la puntuación acumulativa del alineamiento disminuye en la cantidad X desde su máximo valor alcanzado; la puntuación acumulativa va a cero o por debajo, debido a la acumulación de uno o más alineamientos de restos de puntuación negativa; o se llega al extremo de cualquiera de las secuencias. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) usa por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, un punto de corte de 100, M=S, N=-4, y una comparación de ambas cadenas. Para las 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 (Henikoff y Henikoff (1992) *Proc. Natl. Acad. Sci.* 89: 10915-10919). Salvo que se indique lo contrario, la "identidad de secuencia" en el presente documento se refiere al % de identidad de secuencia generado a partir de tblasx usando la versión de NCBI del algoritmo con los parámetros por defecto usando alineamientos incompletos con el filtro apagado (por ejemplo, en la página web de NIH NLM NCBI ncbi.nlm.nih, véase antes).

[0222] Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también lleva a cabo un análisis estadístico de la similitud entre las dos secuencias (por ejemplo, en Karlin y Altschul (1993) *Proc. Natl. Acad. Sci.* 90: 5873-5877). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de la suma más pequeña (P(N)), que proporciona una indicación de la probabilidad con la que ocurriría por azar un emparejamiento entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, se considera que un ácido nucleico es similar a una secuencia de referencia (y, por lo tanto, en este contexto, homóloga) si la probabilidad de la suma más pequeña en una comparación del ácido nucleico de ensayo respecto al ácido nucleico de referencia es menor de aproximadamente 0,1, o menor de aproximadamente 0,01, o incluso menor de aproximadamente 0,001. Un ejemplo adicional de un algoritmo de alineamiento de secuencias útiles PILEUP. PILEUP crea un alineamiento de secuencias múltiples a partir de un grupo de secuencias relacionadas usando alineamientos por pares, progresivos. El programa puede lineal, por ejemplo, hasta 300 secuencias de una longitud máxima de 5.000 letras.

[0223] El sistema integrado, u ordenador, típicamente incluye un interfaz de entrada de usuario que permite al usuario ver selectivamente uno o más registros de secuencias correspondientes a una o más cadenas de caracteres, así como un conjunto de instrucciones que alinea la una o más 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 una conexión de una o más cadenas de caracteres con un fenotipo particular o función génica. Típicamente, el sistema incluye un elemento de salida de lectura por el usuario que presenta un alineamiento producido por el conjunto de instrucciones de alineamiento.

[0224] Los procedimientos de esta invención se pueden implementar en un entorno de computación localizado o distribuido. En un entorno distribuido, los procedimientos se pueden implementar en un solo ordenador que comprende múltiples procesadores o en una multiplicidad de ordenadores. Los ordenadores pueden estar conectados, por ejemplo, a través de un bus común, más preferiblemente el o los ordenadores son nodos en una red. La red ha de ser una red generalizada o local y dedicada o de área amplia, y en determinadas realizaciones preferidas, los ordenadores pueden ser componentes de una intranet o un internet.

[0225] Por lo tanto, la invención proporciona procedimientos para identificar una secuencia similar u homóloga a uno o más polinucleótidos como se indica en el presente documento, o uno o más polipéptidos diana codificados por los polinucleótidos, o indicados de otra forma en el presente documento, y pueden incluir la conexión o asociación de un fenotipo o función génica de una planta dada con una secuencia. En los procedimientos, se proporciona una base de datos de secuencias (localmente o a través de una intranet o internet) y se hace una consulta a la base de datos de secuencias usando las secuencias relevantes del presente documento y fenotipos o funciones génicas asociados de la planta.

[0226] Se puede introducir cualquier secuencia del presente documento en la base de datos, antes o después de la consulta en la base de datos. Esto proporciona tanto la expansión de la base de datos como, si se hace antes de la etapa de consulta, la inserción de secuencias de control en la base de datos. Las secuencias de

control se pueden detectar mediante la consulta para asegurar la integridad general tanto de la base de datos como de la consulta. Como se ha indicado, la consulta se puede llevar a cabo usando un interfaz basado en un navegador de internet. Por ejemplo, la base de datos puede ser una base de datos pública centralizada tal como las indicadas en el presente documento, y la consulta se puede hacer desde un terminal u ordenador remotos a través de internet o intranet.

[0227] Cualquier secuencia del presente documento se puede usar 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 las plantas descendientes, que resulta del cruce de dos plantas de cepas diferentes. Por ejemplo, se pueden identificar secuencias que codifican un ortólogo de cualquiera de las secuencias del presente documento que están de forma natural en una planta con un rasgo deseado, usando las secuencias descritas en el presente documento. Después, la planta se cruza con una segunda planta de la misma especie pero que no tiene el rasgo deseado, para producir descendencia que después se puede usar en experimentos de cruce posteriores 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 se puede regular por el tratamiento con un producto químico particular u otros medios, tales como EMR. Algunos ejemplos de dichos compuestos bien conocidos en la técnica incluyen: etileno; citoquinas; compuestos fenólicos; los cuales estimulan la transcripción de los genes necesarios para la infección; monosacáridos específicos y entornos ácidos que potencian la inducción de genes; polisacáridos ácidos que inducen uno o más genes cromosómicos; y opines; otros mecanismos incluyen tratamiento de luz u oscuridad (por ejemplo, en Winans (1992) *Microbiol. Rev.* 56: 12-31; Eyal y col. (1992) *Plant. Mol. Biol.* 19: 589-599; Chrispeels y col. (2000) *Plant Mol. Biol.* 42: 279-290; Piazza y col. (2002) *Plant Physiol.* 128:1077-1086).

[0228] La tabla 6 lista el número de identificación de gen (GID) y las relaciones homólogas encontradas usando análisis de acuerdo con los ejemplos para las secuencias de la lista de secuencias. Las secuencias listadas como "secuencias de referencia" se determinaron originalmente por experimentación para conferir tolerancia a la sequía cuando se alteraba su expresión. En general, cada secuencia de referencia se usó para identificar el clado en el que se pueden encontrar las secuencias homólogas funcionalmente relacionadas.

Tabla 6. Genes homólogos y otros genes relacionados de genes de factores de transcripción de *Arabidopsis* representativos identificados usando BLAST

SEQ ID NO	GID N°	Polinucleótido (ADN) o polipéptido (PRT)	Especies de las que se obtienen secuencias homólogas	Redacción de la SEQ ID NO con otros genes
1	G47	ADN	<i>Arabidopsis thaliana</i>	Secuencia de referencia; la secuencia del polipéptido predicha es paróloga con G2133
2	G47	PRT	<i>Arabidopsis thaliana</i>	Secuencia de referencia; paróloga con G2133
11	G2133	ADN	<i>Arabidopsis thaliana</i>	Secuencia de referencia; la secuencia del polipéptido predicha es paróloga con G47
12	G214	PRT	<i>Arabidopsis thaliana</i>	Secuencia de referencia; paróloga con G47

Modelización molecular

[0229] Otro medio que se puede usar para confirmar la utilidad y función de secuencias de factores de transcripción que son ortólogas o parálogas con los presentes factores de transcripción descritos, es mediante el uso de software de modelización molecular. La modelización molecular se usa rutinariamente para predecir la estructura de polipéptidos, y están disponibles en el comercio para este propósito una variedad de programas de modelización de estructuras de proteínas tales como "Insight II" (Accelrys, Inc.). Por lo tanto, la modelización se puede usar para predecir qué restos de un polipéptido se pueden cambiar sin alterar la función (Cramer y col. (2003) patente de EE.UU. n° 6.521.453). Por lo tanto, se puede mostrar que los polipéptidos que son de secuencia similar tienen una alta probabilidad de tener función similar por su similitud estructural, lo cual se puede establecer, por ejemplo, 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 y col. ((1990) *Science* 250: 646-651) han discutido con detalle las tendencias de los aminoácidos a formar hélices. Se pueden usar tablas

de actividad de formación de estructura relativa como tablas de sustitución para predecir qué restos se pueden sustituir funcionalmente en una región dada, por ejemplo, en dominios de unión al ADN de factores de transcripción conocidos y equivalentes. Se pueden identificar los homólogos que es probable que sean funcionalmente similares.

- 5 **[0230]** Es de interés particular la estructura de un factor de transcripción en la región de sus dominios conservados, tal como los identificados en la tabla 1 y tabla 3. Se pueden llevar a cabo análisis estructurales comparando la estructura del factor de transcripción conocido alrededor de su dominio conservado con las de ortólogos y parálogos. El análisis de una serie de polipéptidos dentro de un grupo o clado de factores de transcripción, incluyendo los polipéptidos funcional o secuencialmente similares proporcionados en la lista de
10 secuencias, también puede proporcionar comprensión de los elementos estructurales necesarios para regular la transcripción dentro de una familia dada.

EJEMPLOS

- 15 **[0231]** La invención, descrita ahora de forma general, se entenderá mejor por referencia a los siguientes ejemplos, que se incluyen simplemente con el propósito de ilustrar determinados aspectos y realizaciones de la presente invención y no se pretende que limiten la invención. El experto en la materia reconocerá que un factor de transcripción que está asociado con un primer rasgo particular también puede estar asociado con al menos otro
20 segundo rasgo no relacionado e inherente que no se había predicho por el primer rasgo.

[0232] Las descripciones completas de los rasgos asociados con cada polinucleótido de la invención se describen con detalle en los ejemplos VIII, IX y X.

Ejemplo I: identificación y clonación del gen de longitud completa

- 25 **[0233]** Se identificaron secuencias de factores de transcripción presuntos (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 los parámetros por defecto y un umbral del valor de punto de corte P de -4 o -5 o inferior, dependiendo de la longitud de la secuencia de consulta. Después se cribaron los
30 aciertos de la secuencia del factor de transcripción presunto para identificar aquellos que contienen cadenas de secuencia particulares. Si los aciertos de secuencia contenían dichas cadenas de secuencias, las secuencias se confirmaban como factores de transcripción.

- [0234]** Alternativamente, se cribaron bibliotecas de ADNc de *Arabidopsis thaliana* obtenidas de diferentes
35 tejidos o tratamientos, o bibliotecas genómicas, para identificar nuevos miembros de una familia de transcripción usando un procedimiento de hibridación en condiciones rigurosas bajas. Se sintetizaron sondas usando cebadores específicos de genes en una reacción de PCR estándar (temperatura de reasociación 60°C) y se marcaron con ³²P dCTP usando el kit de marcaje High Prime DNA Labeling (Roche Diagnostics Corp., Indianapolis, IN). Las sondas radiomarcadas purificadas se añadieron a filtros sumergidos en medio de hibridación Church (NaPO₄ 0,5 M pH 7,0,
40 SDS al 7%, albúmina de suero bovino al 1% en p/v) y se hibridaron durante la noche a 60°C con agitación. Los filtros se lavaron dos veces durante 45 a 60 minutos con 1 x SCC, SDS al 1% a 60°C.

- [0235]** Para identificar secuencias adicionales 5' o 3' de una secuencia de ADNc parcial en una biblioteca de ADNc, se llevó a cabo la amplificación rápida 5' y 3' de los extremos del ADNc (RACE) usando el kit de amplificación
45 de ADNc MARATHON (Clontech, Palo Alto, CA). En general, el procedimiento implica aislar primero el poli(A)-ARNm, llevando a cabo la síntesis de la primera y segunda cadena de ADNc para generar ADNc de doble cadena, hacer los extremos romos del ADNc, seguido de la unión del adaptador MARATHON al ADNc para formar una biblioteca de ADNc de doble cadena ligado al adaptador.

- 50 **[0236]** Se diseñaron cebadores específicos de gen para usar junto con cebadores específicos del adaptador para la reacciones RACE tanto 5' como 3'. Se usaron cebadores anidados en lugar de cebadores sencillos, para aumentar la especificidad de la PCR. Usando las reacciones RACE de 5' y 3', se obtuvieron fragmentos de RACE 5'
y 3', se secuenciaron y se clonaron. El procedimiento se puede repetir hasta identificar los extremos 5' y 3' del gen de longitud completa. Después se generó el ADNc de longitud completa por PCR usando cebadores específicos
55 para los extremos 5' y 3' del gen mediante PCR de extremo a extremo.

Ejemplo II: Construcción de vectores de expresión

- [0237]** La secuencia se amplificó a partir de una biblioteca genómica o de ADNc usando cebadores

específicos para las secuencias en dirección 5' y dirección 3' de la región codificante. El vector de expresión era pMEN20 o pMEN65, que derivan ambos de pMON316 (Sanders y col. (1987) *Nucleic Acids Res.* 15:1543-1558) y contienen el promotor 35S de CaMV. Para clonar la secuencia en el vector, se digirieron tanto pMEN20 como el fragmento de ADN amplificado por separado con las 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 mediante 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 de QIAQUICK (Qiagen, Valencia CA). Los fragmentos de interés se ligaron en una relación de 3:1 (vector a inserto). Las reacciones de ligación usando la ADN ligasa T4 (New England Biolabs, Beverly MA) se llevaron a cabo a 16°C durante 16 h. Los ADN ligados se transformaron en células competentes de la cepa de *E. coli* DH5 α usando el procedimiento de choque térmico. Las transformaciones se cultivaron en placas LB que contenían kanamicina 50 mg/l (Sigma Chemical Co. St. Louis MO). Las colonias individuales se cultivaron durante la noche en 5 ml de caldo LB que contenía kanamicina 50 mg/l a 37°C. El ADN plasmídico se purificó usando los kits Qiaquick Mini Prep kits (Qiagen).

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

[0238] Después de construir el vector plasmídico que contenía el gen, el vector se usó para transformar células de *Agrobacterium tumefaciens* que expresaban los productos génicos. La cepa de células de *Agrobacterium tumefaciens* para la transformación se hizo como describen Nagel y col. (1990) *FEMS Microbiol Letts.* 67: 325-328. Se cultivó la cepa ABI de *Agrobacterium* en 250 ml de medio LB (Sigma) durante la noche a 28°C con agitación hasta que se alcanzó una absorbancia a lo largo de 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 min a 4°C. Después las células se volvieron a suspender en 250 ml de tampón enfriado (HEPES 1 mM, pH ajustado a 7,0 con KOH). Las células se centrifugaron otra vez como se ha descrito antes y se volvieron a suspender en 125 μ l de tampón enfriado. Después, las células se centrifugación y se volvieron a suspender dos veces más en el mismo tampón HEPES como se ha descrito antes con un volumen de 100 μ l y 750 μ l, respectivamente. Después, las células resuspendidas se distribuyeron en partes alícuotas de 40 μ l, se congelaron rápidamente en nitrógeno líquido y se almacenaron a -80°C.

[0239] Las células de *Agrobacterium* se transformaron con plásmidos preparados como se ha descrito antes siguiendo el protocolo descrito por Nagel y col. (1990), véase antes. Para cada construcción de ADN que se iba a transformar, se mezclaron 50-100 ng de ADN (en general resuspendido en Tris-HCl 10 mM, EDTA 1 mM, pH 8,0) con 40 μ l de células de *Agrobacterium*. La mezcla de ADN/células después se transfirió a una cubeta enfriada con un hueco de electrodo de 2 mm y se sometió a una carga de 2,5 kV disipada a 25 μ F y 200 μ F usando un aparato Gene Pulser II (Bio-Rad, Hercules, CA). Después de electroporación, las células se volvieron a suspender inmediatamente en 1,0 ml de LB y se dejaron recuperar sin selección con antibiótico durante 2-4 h a 28°C en un incubador con agitación. Después de recuperación, las células se cultivaron en medio de caldo LB que contenía estreptomycinina 100 μ g/ml (Sigma) y se incubaron durante 24-48 horas a 28°C. Después se recogieron colonias individuales y se inocularon en medio reciente. La presencia de la construcción del plásmido se verificó por amplificación por PCR y análisis de secuencia.

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

[0240] 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/ml con las colonias y se cultivaron a 28°C con agitación durante 2 días hasta alcanzar una absorbancia óptica a 600 nm de longitud de onda a lo largo de 1 cm (A_{600}) de >2,0. Después, se recogieron las células por centrifugación a 4.000 x g durante 10 min y se volvieron a suspender en medio de infiltración (1/2 X sales de Murashige y Skoog (Sigma), 1 X vitaminas B-5 de Gamborg (Sigma), sacarosa al 5,0% (p/v) (Sigma), bencilamino-purina 0,044 μ M (Sigma), Silwet L-77 200 μ l/l (Lehle Seeds)) hasta alcanzar una A_{600} de 0,8.

[0241] Antes de la transformación, las semillas de *Arabidopsis thaliana* (ecotipo Columbia) se sembraron con una densidad de ~10 plantas por maceta de 5 cm en medio de siembra en maceta Pro-Mix BX (Hummert International) cubiertas con malla de fibra de vidrio (18 mm X 16 mm). Las plantas se cultivaron con iluminación continua (50-75 μ E/m²/s) a 22-23°C con una humedad relativa de 65-70%. Después de aproximadamente 4 semanas, se cortan los tallos de inflorescencias primarias (espigas) para potenciar el crecimiento de múltiples espigas secundarias. Después de florecer las espigas secundarias maduras, las plantas se prepararon para la

transformación retirando todas las silicuas y flores abiertas.

[0242] Después, las macetas se sumergieron boca abajo en la mezcla de medio de infiltración de *Agrobacterium* como se ha descrito antes durante 30 s, y se pusieron de lado para permitir el drenaje en una superficie plana de 2,54 x 5 cm cubiertas con una envoltura plástica. Después de 24 h, se retiró la envoltura plástica y las macetas se pusieron derechas. El procedimiento de inmersión se repitió una semana después, durante un total de 2 inmersiones por maceta. Después se recogieron las semillas de cada maceta de transformación y se analizaron siguiente el siguiente protocolo descrito.

10 Ejemplo V: Identificación de transformantes primarios de *Arabidopsis*

[0243] Las semillas recogidas de las macetas de transformación se esterilizaron esencialmente como sigue. Las semillas se dispersaron en una disolución que contenía Triton X-100 (Sigma) al 0,1% (v/v) y agua estéril y se lavaron agitando la suspensión durante 20 min. La disolución de lavado después se drenó y se substituyó por una disolución de lavado reciente para lavar las semillas durante 20 min con agitación. Después de eliminar la disolución de etanol/detergente, se añadió a las semillas una suspensión que contenía Triton X-100 al 0,1 % (v/v) y lejía al 30% (v/v) (CLOROX; Clorox Corp. Oakland CA), y la suspensión se agitó durante 10 min. Después de separar la disolución de lejía/detergente, las semillas se lavaron 5 veces en agua destilada estéril. Las semillas se almacenaron en el último agua de lavado a 4°C durante 2 días en la oscuridad antes de cultivarlas en placa en medio de selección con antibiótico (1 X sales de Murashige y Skoog (pH ajustado a 5,7 con KOH 1 M), 1 X vitaminas B-5 de Gamborg, Phytagar al 0,9% (Life Technologies) y kanamicina 50 mg/l). Las semillas se germinaron con iluminación continua (50-75 $\mu\text{E}/\text{m}^2/\text{s}$) a 22-23°C. Después de 7-10 días de crecimiento en estas condiciones, se obtuvieron los transformantes primarios resistentes a la kanamicina (generación T1) que eran visibles. Estas plántulas se transfirieron primero a placas de selección recientes, en las que las plántulas continuaron creciendo durante 3-5 días más, y después al suelo (medio de siembra en macetas Pro-Mix BX).

[0244] Los transformantes primarios se cruzaron con semillas de la descendencia (T₂) recogidas; se seleccionaron las plántulas resistentes a la kanamicina y se analizaron. Los niveles de expresión de los polinucleótidos recombinantes en los transformantes varía de aproximadamente un aumento del nivel de expresión de 5% hasta al menos un aumento del nivel de expresión de 100%. Se hicieron observaciones similares con respecto al nivel de expresión del polipéptido.

Ejemplo VI: Identificación de plantas de *Arabidopsis* con inactivaciones de genes de factores de transcripción

[0245] El cribado de las colecciones de *Arabidopsis* mutegenizadas por inserción para mutantes nulos en un gen diana conocido, se hizo esencialmente como describen Krysan y col. (1999) *Plant Cell* 11: 2283-2290. Brevemente, se diseñaron cebadores específicos de gen, anidados por 5-250 pares de bases entre sí, a partir de las regiones 5' y 3' de un gen diana conocido. Igualmente, también se crearon grupos de cebadores anidados específicos para cada uno de los ADN-T o extremos de transposones (los bordes "derecho" e "izquierdo"). Se usaron todas las posibles combinaciones de cebadores específicos de gen y ADN-T/transposón para detectar por PCR un suceso de inserción dentro o cerca del gen diana. Después, los fragmentos de ADN amplificados se secuenciaron, lo que permite la determinación precisa del punto de inserción del ADN-T/transposón con respecto al gen diana. Los sucesos de inserción dentro de la secuencia codificante o intermedia de los genes se desconvolvieron de un grupo que comprendía una pluralidad de sucesos de inserción en una sola planta mutante única para la caracterización funcional. El procedimiento se describe con más detalle en Yu y Adam, solicitud de EE.UU. n° de serie 09/177.733 presentada el 23 de octubre, 1998.

50 Ejemplo VII: Identificación de fenotipos modificados en plantas con expresión en exceso o inactivación de genes

[0246] Se usó *Arabidopsis thaliana* ecotipo Columbia (Col-0) para crear todas las líneas que expresaban en exceso. Las plantas de control para el ensayo eran plantas Col-0 transformadas con un vector de transformación vacío (pMEN65).

55 Experimentos en micromatrices

[0247] En algunos casos, los patrones de expresión de los genes inducidos por estrés se pueden seguir por experimentos de micromatrices. En estos experimentos, los ADNc se generan por la PCR y se vuelven a suspender

- a una concentración final de ~100 ng/μl en 3x SSC o fosfato-Na 150 mM (Eisen y Brown (1999) *Methods Enzymol.* 303: 179-205). Los ADNc se aplican como manchas puntuales sobre portaobjetos de vidrio de microscopio recubiertos con polilisina. Los ADNc preparados se dividen en partes alícuotas en placas de 384 pocillos y se aplican como manchas puntuales sobre los portaobjetos usando, por ejemplo, una estructura de soporte x-y-z (OmniGrid)
- 5 que se puede adquirir en (Menlo Park, CA) equipado con clavijas de tipo vaina que se pueden adquirir en Telechem International (Sunnyvale, CA). Después de la aplicación en manchas puntuales, las matrices se curan durante un mínimo de una semana a temperatura ambiente y, se rehidratan y se bloquean siguiendo el protocolo recomendado por Eisen y Brown (1999) véase antes.
- 10 **[0248]** Las muestras de ARN total (10 μg) se marcan usando colorantes de Cy3 y Cy5 fluorescentes. Las muestras marcadas se vuelven a suspender en 4X SSC/SDS al 0,03%/4 μg de ADN de esperma de salmón/2 μg de ARNt/pirofosfato de sodio 50 mM, se calientan a 95°C durante 2,5 min, se centrifugan y se ponen sobre la matriz. Después la matriz se cubre con un cubreobjetos de vidrio y se pone en una cámara herméticamente cerrada. Después la cámara se mantiene en un baño de agua a 62°C durante una noche. Las matrices se lavan como
- 15 describen Eisen y Brown (1999), véase antes, y se escanean en un escáner láser General Scanning 3000. Los archivos resultantes posteriormente se cuantifican usando el software IMAGE, (BioDiscovery, Los Angeles CA).
- [0249]** Se pueden llevar a cabo experimentos de RT-PCR para identificar aquellos genes inducidos después de exposición a estreses abióticos. En general, se examinan los patrones de expresión de genes de tejidos foliares
- 20 de plantas de suelo.
- [0250]** Se llevó a cabo la PCR con transcriptasa inversa usando cebadores específicos de gen con la región codificante de cada secuencia identificada. Los cebadores se diseñaron cerca de la región 3' de cada secuencia de unión al ADN identificada inicialmente.
- 25 **[0251]** Los ARN totales de estos tejidos foliares del suelo se aislaron usando los protocolos de extracción de CTAB. Una vez extraído, el ARN total se normalizó en concentración con todos los tipos de tejidos para asegurar que la reacción de la PCR para cada tejido recibía la misma cantidad de molde de ADNc usando la banda 28S como referencia. Se purificó el poli(A+) ARN usando un protocolo modificado del protocolo del lote del kit de purificación
- 30 Qiagen OLIGOTEX. El ADNc se sintetizó usando protocolos estándar. Después de la síntesis de la primera cadena de ADNc, se usaron cebadores para actina 2 para normalizar la concentración de ADNc en todos los tipos de tejidos. Se encontró que la actina 2 es expresada de forma constitutiva en niveles bastante iguales en todos los tipos de tejidos que se investigan.
- 35 **[0252]** Para la RT-PCR, los moldes de ADNc se mezclaron con los cebadores correspondientes y la ADN polimerasa Taq. Cada reacción consistía en 0,2 μl de molde de ADNc, 2 μl de 10X tampón de Tricina, 2 μl de 10X tampón de Tricina y 16,8 μl de agua, 0,05 μl de cebador 1, 0,05 μl, de cebador 2, 0,3 μl de ADN polimerasa Taq y 8,6 μl de agua.
- 40 **[0253]** La placa de 96 pocillos se cubre con micropelícula y se pone en el termociclador para iniciar el ciclo de reacción. A modo de ilustración, el ciclo de reacción puede comprender las siguientes etapas:
- ETAPA 1: 93°C durante 3 minutos;
 ETAPA 2: 93°C durante 30 segundos;
 45 ETAPA 3: 65°C durante 1 minuto;
 ETAPA 4: 72°C durante 2 minutos;
 Las ETAPAS 2, 3 y 4 se repiten durante 28 ciclos;
 ETAPA 5: 72°C durante 5 minutos; y
 ETAPA 6: 4°C.
- 50 **[0254]** Para amplificar más productos, por ejemplo, para identificar genes que tienen expresión muy baja, se pueden llevar a cabo etapas adicionales: el siguiente procedimiento ilustra un procedimiento que se puede usar en relación con esto; la placa de PCR se vuelve a poner en el termociclador durante 8 ciclos más de etapas 2-4.
- 55 ETAPA 2: 93°C durante 30 segundos;
 ETAPA 3: 65°C durante 1 minuto;
 ETAPA 4: 72°C durante 2 minutos, repetida 8 ciclos; y
 ETAPA 5: 4°C.

[0255] Se cargan 8 µl de producto de la PCR y 1,5 µl de colorante de carga en un gel de agarosa al 1,2% para el análisis después de 28 ciclos y 36 ciclos. Los niveles de expresión de transcritos específicos se consideran bajos si son detectables sólo después de 36 ciclos de PCR. Los niveles de expresión se consideran medios o altos dependiendo de los niveles de transcritos comparados con los niveles de transcritos observados para un control interno tal como la actina 2. Los niveles de transcritos se determinan en experimentos repetidos y se comparan con los niveles de transcritos en plantas de control (p. ej., no transformadas).

Ensayos de estrés abiótico

10 **[0256]** Los fenotipos modificados observados para plantas particulares que expresan en exceso pueden incluir una mayor biomasa y/o mayor o menor tolerancia o resistencia al estrés abiótico. Para una expresión en exceso en particular que muestra una característica menos beneficiosa, tal como una menor tolerancia o resistencia al estrés abiótico, puede ser más útil seleccionar una planta con una menor expresión de un factor de transcripción particular. Para una inactivación particular que muestra una característica menos beneficiosa, tal como una menor tolerancia al estrés abiótico, puede ser más útil seleccionar una planta con una mayor expresión de un factor de transcripción particular.

20 **[0257]** Los ensayos de germinación en este ejemplo siguieron las modificaciones del mismo protocolo básico. Se sembraron semillas estériles en los medios condicionados listados a continuación. Las placas se incubaron a 22°C con 24 h de luz (120-130 µEin/m²/s) en una cámara de crecimiento. La evaluación de la germinación y el vigor de las plántulas se llevó a cabo de 3 a 15 días después de plantarlas. El medio basal era medio Murashige-Skoog al 80% (MS) + vitaminas.

25 **[0258]** Para los experimentos de estrés llevados a cabo con plantas más maduras, las semillas se germinaron y cultivaron durante varios días en MS + vitaminas + sacarosa al 1% a 22°C y después se transfirieron a condiciones de estrés por frío y calor. Las plantas se expusieron o bien a estrés por frío (exposición de 6 h a 4-8°C) o a estrés por calor (se aplicaron 32°C durante 5 días), después de lo cual las plantas se volvieron a transferir a 22°C para la recuperación y se evaluaron 5 días después con respecto a los controles que no se habían expuesto a temperatura reducida o elevada.

30 **[0259]** Los ensayos de estrés salino pretendían encontrar genes que confirieran mejores germinación, vigor de las plántulas o crecimiento en concentración salina alta. La evaporación desde la superficie del suelo produce el movimiento del agua hacia arriba y la acumulación de sal en la capa superior del suelo donde se ponen las semillas. Por lo tanto, normalmente la germinación se produce con una concentración salina mucho más alta que la concentración salina media del perfil de todo el suelo. Las plantas difieren en su tolerancia al NaCl dependiendo de la etapa de desarrollo, por lo tanto se evaluaron las respuestas en la germinación de la semilla, vigor de la plántula, y crecimiento de la planta.

40 **[0260]** Los ensayos de estrés osmótico (incluyendo ensayos de NaCl y manitol) se llevaron a cabo para determinar si un fenotipo de estrés osmótico era específico de NaCl o era un fenotipo de estrés osmótico general. Las plantas tolerantes al estrés osmótico podrían haber tenido también mayor tolerancia a la sequía y/o helada.

45 **[0261]** Para los experimentos de germinación para el estrés salino y osmótico, el medio se complementó con NaCl 150 mM o manitol 300 mM. Se llevaron a cabo ensayos de sensibilidad de reguladores del crecimiento en medio MS, vitaminas y, o bien ABA 0,3 µM, sacarosa al 9,4%, o bien glucosa al 5%.

50 **[0262]** Los experimentos se llevaron a cabo para identificar los transformantes que presentaban sensores de azúcar modificados. Para dichos estudios, se germinaron semillas de transformantes en medio que contenía una alta concentración de azúcar (glucosa al 5%, sacarosa al 9,4%) que normalmente restringen parcialmente el alargamiento del hipocotilo. Las plantas con los sensores de azúcar alterados pueden tener hipocotilos más largos o más cortos que las plantas normales cuando crecen en este medio. Además, se pueden variar otros rasgos de la planta tales como la masa de la raíz. Los ensayos de sensores de azúcar pretendían encontrar genes implicados en los sensores de azúcar, por germinación de semillas en concentraciones altas de sacarosa y glucosa y buscar los grados de alargamiento de los hipocotilos. El ensayo de germinación en manitol controlaba las respuestas relacionadas con el estrés osmótico. Los azúcares son moléculas reguladoras clave que afectan a diversos procesos en las plantas superiores incluyendo la germinación, crecimiento, floración, senescencia, metabolismo de azúcares y fotosíntesis. La sacarosa es la forma de transporte principal de la fotosíntesis y su flujo a través de las células se ha mostrado que afecta a la expresión de genes y altera la acumulación de compuestos de almacenamiento en semillas (relaciones fuente–sumidero). También se han descrito los sensores de hexosa específicos de glucosa en plantas y

se han implicado en la división celular y en la represión de los genes “famina” (ciclos fotosintéticos o de glioxilato).

[0263] Los ensayos de estrés por temperatura se llevaron a cabo para descubrir genes que conferirían una mejor germinación, vigor a las plántulas o crecimiento de las plantas en condiciones de estrés por temperatura (frío, helada y calor). Los experimentos de germinación en frío para el estrés por temperatura se llevaron a cabo a 8°C. Los experimentos de germinación para el estrés por calor se llevaron a cabo de 32°C a 37°C durante 6 h de exposición.

[0264] Los cribados para la sequía basados en suelo se llevaron a cabo en plantas de *Arabidopsis* que expresaban en exceso los factores de transcripción listados en la Lista de Secuencias. Semillas de plantas de *Arabidopsis* silvestres o plantas que expresaban en exceso un polipéptido de la invención, se pusieron en capas durante 3 días a 4°C en agarosa al 0,1%. Después se sembraron 14 semillas de cada uno del tipo de expresión en exceso o silvestre en macetas de arcilla de 7,6 cm que contenían una mezcla de vermiculita:perlita 50:50 con una pequeña capa en la parte superior de MetroMix 200 y se cultivaron durante 15 días con luz las 24 h. Las macetas que contenían las plántulas silvestres y que expresaban en exceso se pusieron en pletinas en orden aleatorio. El estrés por sequía se inició poniendo las macetas sobre papel absorbente durante 7 a 8 días. Se consideró que las plántulas estaban suficientemente estresadas cuando la mayoría de las macetas que contenían las plántulas silvestres en una pletina se habían marchitado gravemente. Las macetas entonces se volvieron a regar y se puntuó la supervivencia de 4 a 7 días después. Las plantas se clasificaron frente a los controles silvestres para cada uno de 2 criterios: tolerancia a las condiciones de sequía y recuperación (supervivencia) después de volver a regar.

[0265] Al final del periodo de sequía inicial, se asignó a cada maceta una puntuación con valor numérico dependiendo de los criterios anteriores. Se asignaron puntuaciones de 0-6 (tabla 11), con el valor inferior de “0” asignado a plantas con un aspecto extremadamente pobre (es decir, plantas que estaban uniformemente marrones) y se dio un valor de “6” a plantas que se valoraron como de aspecto muy saludable (es decir, las plantas estaban todas verdes). Después de volver a regar las plantas e incubarlas durante 4 a 7 días adicionales, se volvieron a evaluar las plantas para indicar el grado de recuperación desde el tratamiento de privación de agua.

[0266] Después, se llevó a cabo un análisis para determinar que plantas sobrevivían mejor a la falta de agua, identificando los transgenes que conferirían de forma consistente los fenotipos de tolerancia a la sequía y la capacidad para recuperarse de este tratamiento. El análisis se llevó a cabo comparando tabulaciones generales y dentro de una pletina con un conjunto de modelos estadísticos para dar cuenta de las variaciones entre lotes. Se tabularon varias medidas de supervivencia, incluyendo (a) la proporción media de plantas que sobrevivían con respecto a la supervivencia de las silvestres dentro de la misma pletina; (b) la proporción mediana de supervivencia con respecto a la supervivencia de las silvestres dentro de la misma pletina; (c) la supervivencia media general (tomando todos los lotes, pletinas y macetas); (d) la supervivencia media general con respecto a la supervivencia de silvestres general; y (e) la puntuación visual media de la salud de las plantas antes de volver a regarlas.

[0267] A continuación se indican ejemplos de genes y homólogos que confieren mejoras significativas a plantas inactivadas o que expresan en exceso. También se presentan las observaciones experimentales hechas por los autores de la invención con respecto a genes específicos cuya expresión se ha modificado en plantas inactivadas o que expresan en exceso, y potenciales aplicaciones basadas en estas observaciones.

Ejemplo VIII: resultados del análisis de estrés por sequía

[0268] Este ejemplo proporciona pruebas experimentales para la mayor tolerancia al estrés abiótico controlada por polipéptidos de factores de transcripción y polipéptidos de la invención.

Resultados:

[0269] Como se indica a continuación, se mostró que el exceso de expresión de G2133 aumentaba la tolerancia al estrés por sequía en plantas. Una serie de ortólogos de algunas de estas secuencias también eran capaces de aumentar la tolerancia al estrés abiótico, como se indica a continuación.

55 El clado G47 de polipéptidos de factores de transcripción

G47 (SEQ ID NO: 1 y 2)

[0270] Información publicada. G47 corresponde al gen T22J18.2 (AAC25505). No hay información disponible

sobre la o las funciones de G47.

[0271] Observaciones experimentales. La función de G47 se estudió usando plantas de *Arabidopsis* transgénicas en las que el gen se expresaban bajo el control del promotor 35S. El exceso de expresión de G47 daba como resultado una variedad de alteraciones fenotípicas morfológicas y fisiológicas.

[0272] Las plantas 35S::G47 presentaban una mayor tolerancia al estrés osmótico; los ensayos de estrés osmótico se llevaron a cabo usando medio de crecimiento que contenían polietilenglicol (PEG). Después de germinación, las plántulas de las líneas 35S::G47 que expresaban en exceso, en general aparecieron más grandes y tenían más crecimiento de las raíces que las plántulas de control silvestre.

[0273] Como se había predicho por estos ensayos de estrés osmótico, las plantas G47 también presentaban mayor supervivencia y tolerancia a la sequía en un ensayo de sequía basado en el suelo.

[0274] La expresión en exceso de G47 también producía un retraso sustancial en el tiempo de floración y producía un cambio notable en la arquitectura de los brotes. Los transformantes 35S::G47 eran pequeños en las primeras etapas y cambiaban para florecer más de una semana más tarde que los controles silvestres (condiciones de luz continua). Es interesante que, las inflorescencias de estas plantas aparecían densas y carnosas, tenían menor dominancia apical y presentaban un menor alargamiento entre nudos conduciendo a una estatura compacta baja. El patrón de ramificación de los tallos también aparecía anormal, volviéndose los brotes primarios "retorcidos" en cada nudo de coflorescencia. Además, las plantas mostraron una fertilidad ligeramente reducida y formaban silicuas bastante pequeñas que nacían en pedicelos cortos y se mantenían en vertical, cercanos contra el tallo.

[0275] Se detectaron alteraciones adicionales en los tallos de inflorescencias de plantas 35S::G47. Las secciones de los tallos de plantas T2-21 y T2-24 eran de diámetro más ancho, y tenían haces vasculares irregulares grandes que contenían un número mucho mayor de vasos de xilema que el tipo silvestre. Además algunos de los vasos de xilema dentro de los haces aparecían estrechos y posiblemente estaban más lignificados que los de los controles.

[0276] G47 se expresaba con niveles mayores en hojas en roseta, y los transcritos se podían detectar en otros tejidos (flor, embrión, silicua y plántulas en germinación), pero aparentemente no en las raíces.

[0277] Utilidades. G47 o sus equivalentes se pueden usar para aumentar la tolerancia de las plantas a la sequía y a otros estreses osmóticos. G47 o sus equivalentes también se podían usar para manipular el tiempo de floración, para modificar la arquitectura de la planta y la estructura del tallo, incluyendo el desarrollo del tejido vascular y el contenido de lignina. El uso de G47 o sus equivalentes de especies de árboles podría ofrecer el potencial de modular el contenido de lignina. Esto podría permitir mejorar la calidad de la madera usada para el mobiliario o la construcción. G47 y sus equivalentes incluyen, por ejemplo *Arabidopsis thaliana* SEQ ID NO: 12 (G2133); *Oryza sativa* (grupo de variedad de cultivo *japonica*) SEQ ID NO: 98 (G3649), SEQ ID NO: 100 (G3651), y SEQ ID NO: 90 (G3644); *Glycine max* SEQ ID NO: 88 (G3643); *Zinnia elegans* SEQ ID NO: 96 (G3647); *Brassica rapa* subsp. *Pekinensis* SEQ ID NO: 92 (G3645); y *Brassica oleracea* SEQ ID NO: 94 (G3646).

G2133 (SEQ ID NO: 11 y 12)

[0278] Información publicada. G2133 es un parálogo de G47. G2133 corresponde al gen F26A9.11 (AAF23336). No hay información disponible sobre la o las funciones de G2133.

[0279] Observaciones experimentales. La función de G2133 se estudió usando plantas de *Arabidopsis* transgénicas en las que el gen se expresaban bajo el control del promotor 35S.

[0280] La expresión de G2133 se detectó en una variedad de tejidos: muestras de flor, hoja, embrión y silicua. Su expresión se puede alterar mediante varias condiciones, incluyendo el tratamiento con auxina, estrés osmótico, infección por *Fusarium*. El exceso de expresión de G2133 causaba una variedad de alteraciones en el crecimiento y desarrollo de la planta: floración retrasada, arquitectura alterada de la inflorescencia y una disminución en el tamaño general y fertilidad.

[0281] En las primeras etapas, los transformantes 35S::G47 eran notablemente más pequeños que los controles y presentaban hojas retorcidas y verde oscuro. La mayoría de estas plantas permanecía en una fase vegetativa del desarrollo sustancialmente más larga que la de los controles, y producía un mayor número de hojas

antes de la floración temprana. En las plantas más gravemente afectadas, la floración temprana se produjo más de un mes más tarde que en las silvestres (24 h de luz). Además, las plantas presentaron una reducción de la dominancia apical y formaron números mayores de brotes simultáneamente, de las axilas de las hojas en roseta. Estos tallos con inflorescencias tenían entrenudos más cortos y llevaban un número mayor de nudos de hojas caulinares, dándoles un aspecto muy carnoso. La fertilidad de las plantas 35S::G47 en general era muy baja. Además, se encontró que las líneas G2133 que expresaban en exceso, eran más resistentes al herbicida glifosato en los experimentos iniciales y repetidos.

10 **[0282]** No se detectaron alteraciones en las plantas 35S::G47 en los análisis bioquímicos que se llevaron a cabo.

15 **[0283]** G2133 es un parálogo de G47, sabiéndose de este último por estudios previos, que confiere un fenotipo de tolerancia a la sequía cuando se expresa en exceso. Por lo tanto, no fue sorprendente cuando se mostró que G213 también inducía tolerancia a la sequía en una serie de líneas de 35S::G2133 estimuladas en ensayos de sequía basados en suelo (tablas 11 y 12). Los experimentos que comparaban la recuperación de controles silvestres y dos líneas de plantas de *Arabidopsis* que expresaban en exceso G2133 (un parálogo de G47), de un tratamiento de sequía, se llevaron a cabo con luz constante. Las líneas de 35S::G2133 y de control se cultivaron en macetas, conteniendo cada maceta varias plantas. Se privó a todas de agua durante 8 días y después se volvieron a regar. Después de volver a regarlas, todas las plantas de ambas líneas de G2133 que expresaban en exceso se revigorizaron, y todas las plantas de control murieron o estaban gravemente afectadas por el tratamiento de sequía (tabla 12).

25 **[0284]** Utilidades. G2133 y sus equivalentes se pueden usar para aumentar la tolerancia de las plantas a la sequía y a otros estresos osmóticos. G2133 también se podría usar para la generación de plantas resistentes al glifosato y para aumentar la resistencia de la planta al estrés oxidativo. G2133 y sus equivalentes incluyen, por ejemplo, *Arabidopsis thaliana* SEQ ID NO: 2 (G47); *Oryza sativa* (grupo de variedad cultivada *japonica*) SEQ ID NO: 98 (G3649), SEQ ID NO: 100 (G3651), y SEQ ID NO: 90 (G3644); *Glycine max* SEQ ID NO: 88 (G3643); *Zinnia elegans* SEQ ID NO: 96 (G3647); *Brassica rapa* subsp. *Pekinensis* SEQ ID NO: 92 (G3645); y *Brassica oleracea* SEQ ID NO: 94 (G3646).

30 **G3643 (SEQ ID NO: 87 and 88)**

[0285] G3643 es un ortólogo de la soja de G47 y G2133.

35 **[0286]** Observaciones experimentales. La función de G3643 se estudió usando plantas de *Arabidopsis* transgénicas en las que el gen se expresaba bajo control del promotor 35S.

40 **[0287]** Las plantas de *Arabidopsis* que expresaban en exceso G3643 eran más tolerantes al frío que las plantas de control silvestres cultivadas en condiciones similares en ensayos de germinación basados en placa. Una de estas líneas también era más tolerante a la desecación y al crecimiento en condiciones de frío en ensayos basados en la placa.

45 **[0288]** Utilidades. G3643 y sus equivalentes se pueden usar para aumentar la tolerancia de las plantas a condiciones de frío y condiciones de poca agua, incluyendo sequía.

G3644 (SEQ ID NO: 89 y 90)

[0289] G3644 es un ortólogo del arroz de G47 y G2133.

50 **[0290]** Observaciones experimentales. La función de G3644 se estudió usando plantas de *Arabidopsis* transgénicas en las que el gen se expresaba bajo control del promotor 35S.

55 **[0291]** Se encontró que varias plantas de *Arabidopsis* que expresaban en exceso G3644 eran más tolerantes a la desecación que las plantas de control silvestres cultivadas en condiciones similares en ensayos basados en placa. Se mostró que dos líneas también eran más tolerantes a la sal que la silvestre.

[0292] Utilidades. G3644 o sus equivalentes se pueden usar para aumentar la tolerancia de las plantas a concentraciones altas de sal y condiciones de poca agua, incluyendo sequía.

G3649 (SEQ ID NO: 97 y 98)

[0293] G3649 es un ortólogo del arroz de G47 y G2133.

5 [0294] Observaciones experimentales. La función de G3649 se estudió usando plantas de *Arabidopsis* transgénicas en las que el gen se expresaba bajo control del promotor 35S.

[0295] Varias plantas de *Arabidopsis* que expresaban en exceso G3649 eran más tolerantes al frío que las plantas de control silvestres cultivadas en condiciones similares en ensayos basados en placa. Dos líneas que
10 expresaban en exceso eran más tolerantes al calor que las plantas silvestres, y una línea 35S::G47 se encontró que era más tolerante a la desecación que la silvestre.

[0296] Utilidades. G3649 o sus equivalentes se pueden usar para aumentar la tolerancia de las plantas a condiciones de frío y condiciones de poca agua, incluyendo sequía.

15

Resumen de los resultados de los ensayos de sequía

[0297] La tabla 7 presenta los resultados obtenidos en un ensayo en el que plantas de *Arabidopsis* se sometieron a la privación de agua durante 7 a 8 días. Al final de este periodo de sequía, se asignó a cada maceta una puntuación numérica dependiendo de la salud de sus plantas. Se asignó una puntuación de 0 a 6 basándose en
20 el color de la planta y el aspecto general, recibiendo "0" las plantas que eran todo marrones, y en el otro extremo del espectro, las plantas que tenían un aspecto excelente (todo verdes) recibieron un "6". La media de la puntuación numérica registrada de todas las macetas de un genotipo dado por línea de todas las pletinas ensayadas se presenta en orden de salud decreciente.

25

Tabla 7. Comparación de la puntuación numérica registrada para las plantas sometidas a tratamiento de sequía.

GID	Puntuación media
G2133	5.875
G634	4.778
G922	4.667
G916	4.6
G1274	4.273
G864	3.733
G2999	3.7
G2992	3.7
G353	3.6
G47	3.459
G2053	3.404
G975	3.393
G489	3.364
G1792	3.281
G1820	3.2
G24S3	3.2
G2140	3.139
G2701	3.108
G3086	3.056
G611	3.048
G1452	3.042
G481	3.041
G624	3.000
G2854	2.829

GID	Puntuación media
G303	2.812
G2839	2.783
G2789	2708
G188	2692
G325	2.556
G2776	2.513
G175	2.467
G2110	2.432
G1206	2.412
G682	2.381
G1730	2.341
G2969	2.333
G2998	2333
G1069	2.316
Wild-type	2.284

[0298] La tabla 8 compara las clasificaciones de supervivencia de plantas de *Arabidopsis* que expresan en exceso varios polipéptidos, evaluadas después de 7 a 8 días de tratamiento de sequía, volver a regarlas y un periodo de recuperación de 2 a 3 días. Los valores indican la probabilidad mediana de supervivencia dentro de una pletina dada (los percentiles 50 de supervivencia dentro de cada maceta de un genotipo dado por línea dividido entre la supervivencia media de las silvestres en la pletina).

10 Tabla 8. Valoraciones de supervivencias para plantas de *Arabidopsis* después de tratamiento de sequía y volver a regarlas

GID	Mediana por pletina
G2133	3.365
G1274	2.059
G922	1.406
G2999	1.255
G3086	1.179
G354	1.167
G1792	1.161
G2053	1.091
G975	1.090
G1069	1.037
G916	1.023
G2T01	1.000
G1820	1.000
G47	0.921
G2854	0.889
G2789	0.845
G481	0.843
G634	0.834
G175	0.814
G2839	0.805
G1452	0.803
Wild-type	0.800

Ejemplo IX: Identificación de secuencias homólogas

5

[0299] Este ejemplo describe la identificación de genes que son ortólogos a los factores de transcripción de *Arabidopsis thaliana* a partir de una búsqueda de homología por ordenador.

[0300] Las secuencias homólogas, incluyendo las de parálogos y ortólogos de *Arabidopsis* y otras especies de plantas, se identificaron usando herramientas de búsqueda de secuencias en bases de datos, tales como la herramienta Basic Local Alignment Search Tool (BLAST) (Altschul y col. (1990) véase antes; y Altschul y col. (1997) *Nucleic Acid Res.* 25: 3389-3402). Los programas de análisis de secuencias tblastx se usaron con la matriz de puntuación BLOSUM-62 (Henikoff y Henikoff (1992) *Proc Natl. Acad. Sci.* 89: 10915-10919). Se filtró la base de datos de NCBI GenBank entera para buscar secuencias de todas las plantas excepto de *Arabidopsis thaliana* seleccionando todas las entradas en la base de datos NCBI GenBank asociadas con el ID taxonómico de NCBI 33090 (Viridiplantae; todas las plantas) y excluyendo las entradas asociadas con el ID taxonómico 3701 (*Arabidopsis thaliana*).

[0301] Estas secuencias se comparan con secuencias que representan genes de la invención, por ejemplo, polinucleótidos encontrados en la lista de secuencias, usando el algoritmo TBLASTX de la universidad de Washington (versión 2.0a19MP) con los parámetros por defecto usando los alineamientos incompletos con el filtro "apagado". Para cada secuencia de polinucleótido encontrada en la Lista de Secuencias, se ordenaron comparaciones individuales por puntuaciones de probabilidad (valor P), donde las puntuaciones reflejan la

probabilidad de que se produzca un alineamiento particular al azar. Por ejemplo, una puntuación de $3,6E-40$ es $3,6 \times 10^{-40}$. Además de los valores P, también se puntuaron las comparaciones por el porcentaje de identidad. El porcentaje de identidad refleja el grado con el que dos segmentos de ADN o proteínas son idénticos a lo largo de una longitud particular. Los ejemplos de secuencias identificadas de esta forma se presentan en la tabla 6. El porcentaje de identidad de secuencias entre estas secuencias puede ser tan bajo como 47%, o incluso una identidad de secuencia menor.

[0302] Las secuencias parálogas candidatas se identificaron entre factores de transcripción de *Arabidopsis* por el alineamiento, identidad y relaciones filogenéticas. Se identificaron secuencias ortólogas candidatas del conjunto de la base de datos de Unigene de marca registrada de secuencias de genes de plantas en *Zea mays*, *Glycine max* y *Oryza sativa*, basándose en la homología significativa con los factores de transcripción de *Arabidopsis*. Estos candidatos se compararon recíprocamente con el conjunto de factores de transcripción de *Arabidopsis*. Si el candidato mostraba una similitud máxima en el dominio de proteína con el factor de transcripción que se provocaba o con un parólogo del factor de transcripción que se provocaba, entonces se consideraba que era un ortólogo. Las secuencias que no eran de *Arabidopsis* que se mostró de esta forma que eran ortólogas a las secuencias de *Arabidopsis* se proporcionan en la tabla 6.

Ejemplo 7. Cribado de biblioteca de ADNc de planta para la secuencia que codifica el dominio de unión al ADN del factor de transcripción que se une a un elemento promotor de unión al factor de transcripción y demostración de la actividad reguladora de transcripción de la proteína

[0303] Se usa la estrategia de “un híbrido” (Li y Herskowitz (1993) *Science* 262:1870-1874) para cribar clones de ADNc de planta que codifican un polipéptido que comprende un dominio de unión al ADN del factor de transcripción, un dominio conservado. Brevemente, se construyen cepas de levadura que contienen un gen indicador lacZ con las secuencias del elemento promotor de unión del factor de transcripción silvestre o mutante en lugar de la UAS normal (secuencia activadora en dirección 5') del promotor GALA. Se construyen cepas indicadoras de levadura que llevan las secuencias del elemento promotor de unión del factor de transcripción como elementos UAS que están operativamente unidos en dirección 5' de un gen indicador lacZ con un promotor GAL4 mínimo. Las cepas se transforman con una biblioteca de expresión de planta que contiene insertos de ADNc aleatorios fusionados con el dominio de activación de GALA (GAL4-ACT) y se criban las formaciones de colonias azules en filtros tratados con X-gal (X-gal: 5-bromo-4-cloro-3-indolil- β -D-galactosida; Invitrogen Corporation. Carlsbad CA). Alternativamente, las cepas se transformaron con un polinucleótido de ADNc que codifica una secuencia de polipéptido de dominio de unión al ADN de factor de transcripción conocida.

[0304] Las cepas de levadura que llevan estas construcciones indicadoras producen niveles bajos de β -galactosidasa y forman colonias blancas en los filtros que contienen X-gal. Las cepas indicadoras que llevan las secuencias del elemento promotor de unión del factor de transcripción silvestres se transforman con un polinucleótido que codifica un polipéptido que comprende un dominio de unión al ADN de factores de transcripción de planta operativamente unidos a un dominio activador ácido del factor de transcripción GALA de levadura, “GAL4-ACT”. Los clones que contienen un polinucleótido que codifica un dominio de unión al ADN de factor de transcripción operativamente unido a GALA-ACT pueden unirse en dirección 5' de los genes indicadores lacZ que llevan la secuencia del elemento promotor de unión del factor de transcripción silvestre, activar la transcripción del gen lacZ y dar como resultado levaduras que forman colonias azules en los filtros tratados con X-gal.

[0305] Tras el cribado de aproximadamente 2×10^6 transformantes de levadura, se aíslan clones de ADNc positivos; es decir, clones que hacen que cepas de levaduras que llevan los indicadores lacZ operativamente unidos a elementos promotores de unión de factores de transcripción silvestres formen colonias azules en filtros tratados con X-gal. Los clones de ADNc no hacen que una cepa de levaduras que lleva elementos promotores de unión de factor de transcripción de tipo mutante fusionados con LacZ se vuelva azul. Por lo tanto, se muestra que un polinucleótido que codifica el dominio de unión al ADN de factor de transcripción, un dominio conservado, activa la transcripción de un gen.

Ejemplo XI: Ensayos de desplazamiento en gel.

[0306] La presencia de un factor de transcripción que comprende un dominio de unión al ADN que se une a un elemento de unión al ADN del factor de transcripción se evalúa usando el siguiente ensayo de desplazamiento en gel. El factor de transcripción se expresa de forma recombinante y se aísla de *E. coli* o se aísla de un material de planta. La proteína soluble total, que incluye el factor de transcripción, (40 ng) se incubaba a temperatura ambiente en $10 \mu\text{l}$ de 1x tampón de unión (HEPES 15 mM (pH 7,9), EDTA 1 mM, KCl 30 mM, glicerol al 5%, albúmina de suero

bovino al 5%, DTT 1 mM) más 50 ng poli(dl-dC):poli(dl-dC) (Pharmacia, Piscataway NJ) con o sin 100 ng de ADN competidor. Después de 10 minutos de incubación, se añade un ADN sonda que comprende un elemento de unión al ADN del factor de transcripción (1 ng) que se ha marcado con ³²P mediante carga en el extremo (Sambrook y col. véase antes) y la mezcla se incuba durante 10 minutos adicionales. Las muestras se cargan en geles de 5 poliacrilamida (al 4% en p/v) y se fraccionan por electroforesis a 150 V durante 2 h (Sambrook y col. véase antes). El grado de unión del factor de transcripción-ADN sonda se visualiza mediante autorradiografía. Las sondas y ADN competidores se preparan a partir de insertos oligonucleótidos ligados en un sitio BamHI de pUC118 (Vieira y col. (1987) *Methods Enzymol.* 153: 3-11). La orientación y el número de concatenaciones de insertos se determinan mediante análisis de secuencia de ADN por el procedimiento dideoxi (Sambrook y col. véase antes). Los insertos 10 se recuperan después de digestión con enzimas de restricción con EcoRI y HindIII y se fraccionan en geles de poliacrilamida (al 12% en p/v) (Sambrook y col. véase antes).

Ejemplo XII. Transformación de dicotiledóneas

15 **[0307]** Se ha mostrado experimentalmente que las especies de cultivos que expresan en exceso miembros del clado G1792 de polipéptidos de factores de transcripción, producen plantas con una mayor tolerancia a la enfermedad. Esta observación indica que estos genes, cuando son expresados en exceso, darán como resultado rendimientos mayores de diferentes especies de plantas, en particular durante condiciones de estrés abiótico.

20 **[0308]** Por lo tanto, las secuencias de factores de transcripción listadas en la Lista de Secuencias recombinadas en vectores de expresión pMEN20 o pMEN65 se pueden transformar en una planta con el propósito de modificar rasgos de la planta. El vector de clonación se puede introducir en una variedad de plantas de cereales mediante medios bien conocidos en la técnica tales como, por ejemplo, la transferencia de ADN directa o transformación mediada por *Agrobacterium tumefaciens*. Ahora es rutinario producir plantas transgénicas usando la 25 mayoría de las plantas dicotiledóneas (véase Weissbach y Weissbach, (1989) véase antes; Gelvin y col. (1990) véase antes; Herrera-Estrella y col. (1983) véase antes; Bevan (1984) véase antes; y Klee (1985) véase antes). Los procedimientos de análisis de rasgos son rutinarios en la técnica y se han descritos ejemplos antes.

[0309] Se pueden encontrar procedimientos para transformar el algodón en las patentes de EE.UU. n° 30 5.004.863, 5.159.135 y 5.518.908; para transformar especies de brassica se pueden encontrar en la patente de EE.UU. n° 5.463.174; para transformar plantas del cacahuate se pueden encontrar en Cheng y col. (1996) *Plant Cell Rep.* 15: 653-657, y McKently y col. (1995) *Plant Cell Rep.* 14: 699-703; y para transformar plantas del guisante se pueden encontrar en Grant y col. (1995) *Plant Cell Rep.* 15: 254-258.

35 **[0310]** Se han descrito previamente numerosos protocolos para la transformación de plantas del tomate y la soja, y son bien conocidos en la técnica. Gruber y col. ((1993) en "Methods in Plant Molecular Biology and Biotechnology", pág. 89-119, Glick y Thompson, eds., CRC Press, Inc., Boca Raton) describen varios vectores de expresión y procedimientos de cultivo que se pueden usar para la transformación de células o tejidos y la posterior regeneración. Para la transformación de la soja, describen procedimientos Miki y col. (1993) en "Methods in Plant 40 Molecular Biology and Biotechnology", pág. 67-88, Glick and Thompson, eds., CRC Press, Inc., Boca Raton; y en la patente de EE.UU. n° 5.563.055, (Townsend and Thomas), presentada el 8 de octubre, 1996.

[0311] Hay un número sustancial de alternativas a los protocolos de transformación mediados por *Agrobacterium*, y otros procedimientos para el propósito de transferir genes endógenos a la soja o a tomates. Uno 45 de dichos procedimientos es la transformación mediada por microproyectiles, en la que el ADN sobre la superficie de partículas de microproyectiles se dirige a los tejidos de las plantas con un dispositivo biobalístico (véase, por ejemplo, Sanford y col., (1987) *Part. Sci. Technol.* 5:27-37; Christou y col. (1992) *Plant. J.* 2: 275-281; Sanford (1993) *Methods Enzymol.* 217: 483-509; Klein y col. (1987) *Nature* 327: 70-73; patente de EE.UU. n° 5.015.580 (Christou y col), presentada el 14 de mayo, 1991; y patente de EE.UU. n° 5.322.783 (Tomes y col.), presentada el 21 de junio, 50 1994.

[0312] Alternativamente, se han usado procedimientos de tratamiento por ultrasonidos (véase, por ejemplo, Zhang y col. (1991) *Bio/Technology* 9: 996-997); absorción directa de ADN en protoplastos usando precipitación con CaCl₂, poli(alcohol vinílico) o poli-L-ornitina (véase, por ejemplo, Hain y col. (1985) *Mol. Gen. Genet.* 199: 161-168; 55 Draper y col., *Plant Cell Physiol.* 23: 451-458 (1982)); fusión de liposomas o esferoplastos (véase, por ejemplo, Deshayes y col. (1985) *EMBO J.*, 4: 2731-2737; Christou y col. (1987) *Proc. Natl. Acad. Sci. USA* 84: 3962-3966); y electroporación de protoplastos y células enteras y tejidos (véase, por ejemplo, Donn y col.(1990) en Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC, A2-38: 53; D'Halluin y col. (1992) *Plant Cell* 4: 1495-1505; y Spencer y col. (1994) *Plant Mol. Biol.* 24: 51-61), para introducir ADN extraño y vectores de expresión

en plantas.

[0313] Después de transformar una planta o célula de planta (y esta última regenerada en una planta), la planta transformada se puede cruzar consigo misma o con una planta de la misma línea, una planta no transformada o silvestre, u otra planta transformada de una línea de plantas transgénicas diferentes. El cruzamiento proporciona la ventaja de producir variedades transgénicas nuevas y a menudo estables. Los genes y los rasgos que confieren, que se han introducido en una línea del tomate o la soja se puede mover a líneas de plantas diferentes usando técnicas de retrocruzamiento tradicionales conocidas en la técnica. La transformación de plantas de tomate se puede llevar a cabo usando los protocolos de Koornneef y col. (1986) en "Tomato Biotechnology": Alan R Liss, Inc., 169-178, y en la patente de EE.UU. 6.613.962, describiéndose este último procedimiento brevemente en el presente documento. Explantes de cotiledones de 8 días de edad se precultivan durante 24 h en placas Petri que contienen una capa alimentadora de células en suspensión de *Petunia* híbrida cultivadas en medio MS con sacarosa al 2% (p/v) y agar al 0,8% complementado con ácido α -naftalenoacético 10 μ M y 6-bencilaminopurina 4,4 μ M. Después los explantes se infectan 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 absorción sobre papel de filtro estéril y se cocultivan durante 48 horas en las placas de capa alimentadora originales. Las condiciones de cultivo son como se describen antes. Los cultivos durante la noche de *Agrobacterium tumefaciens* se diluyen en medio MS líquido con sacarosa al 2% (p/v), pH 5,7, hasta una DO₆₀₀ de 0,8.

[0314] Después de cultivar simultáneamente, los explantes de cotiledones se transfieren a placas Petri con medio selectivo que comprende medio MS con zeatina 4,56 μ M, vancomicina 67,3 mM, cefotaxima 418,9 mM y sulfato de kanamicina 171,6 mM, y se cultivan en las condiciones descritas antes. Los explantes de subcultivan cada 3 semanas en medio reciente. Los brotes que emergen se diseccionan del callo subyacente y se transfieren a jarras de vidrio con medio selectivo sin zeatina para formar raíces. La formación de raíces en un medio que contiene sulfato de kanamicina es una indicación positiva de una transformación exitosa.

[0315] La transformación de plantas de soja se puede llevar a cabo usando los procedimientos que se encuentran, por ejemplo, en la patente de EE.UU. 5.563.055. En este procedimiento, se esteriliza la superficie de la semilla de soja por exposición a cloro gaseoso que se desarrolla en una campana de vidrio. Las semillas se germinan mediante cultivo en placa en medio solidificado de agar con una concentración 1/10 sin reguladores del crecimiento de plantas y se cultivan a 28°C con una longitud del día de 16 h. Después de 3 o 4 días, la semilla se puede preparar para el cultivo simultáneo. Se elimina el recubrimiento de la semilla y el alargamiento radicular se elimina 3-4 mm por debajo de los cotiledones.

[0316] Los cultivos durante la noche de *Agrobacterium tumefaciens* que albergan el vector de expresión que comprende un polinucleótido de la invención se cultivan hasta la fase logarítmica, se agrupan y se concentran por centrifugación. Se llevan a cabo inoculaciones en lotes de modo que cada placa de semillas se trató con un sedimento recién suspendido de *Agrobacterium*. Los sedimentos se vuelven a suspender en 20 ml de medio de inoculación. El inóculo se vierte en una placa Petri que contiene semillas preparadas y los nodos cotiledonares se maceran con una hoja de bisturí. Después de 30 minutos, los explantes se transfieren a placas del mismo medio que se ha solidificado. Los explantes se insertan con la cara adaxial hacia arriba y se nivelan con la superficie del medio y se cultivan a 22°C durante 3 días con luz fluorescente blanca. Después, estas plantas se pueden regenerar de acuerdo con procedimientos bien establecidos en la técnica, tal como moviendo los explantes después de 3 días a un medio de contraselección líquido (véase la patente de EE.UU. 5.563.055).

[0317] Después se pueden coger los explantes, insertarlos y cultivarlos en medio de selección solidificado. Después de 1 mes en medio selectivo, el tejido transformado se hace visible como sectores verdes de tejido regenerado contra un fondo de tejido menos sano, decolorado. Los explantes con sectores verde se transfieren a un medio de alargamiento. El cultivo se continúa en este medio con transferencias a placas recientes cada 2 semanas. Cuando los brotes tienen 0,5 cm de longitud se pueden cortar por la base y poner en un medio de enraizamiento.

Ejemplo XIII: Tolerancia alterada al estrés abiótico en monocotiledóneas

[0318] Las plantas de cereales tales como, pero sin limitar, maíz, trigo, arroz, sorgo o cebada, se pueden transformar con las presentes secuencias de polinucleótidos, incluyendo secuencias derivadas de mono o dicotiledóneas tales como las presentadas en las tablas 1, 3 ó 6, clonadas en un vector tal como pGA643 y que contienen un marcador de resistencia a la kanamicina, y expresar de forma constitutiva, por ejemplo, bajo los promotores CaMV 35S o COR15. También se pueden usar pMEN20 o pMEN65 y otros vectores de expresión con el propósito de modificar rasgos de la planta. Por ejemplo, pMEN020 se puede modificar para sustituir la región

codificante NptII con el gen BAR de *Streptomyces hygrosopicus* que confiere resistencia a la fosfotricina. Los sitios KpnI y BglII del gen Bar se eliminan mediante mutagénesis dirigida con cambios de codones silenciosos.

5 **[0319]** El vector de clonación se puede introducir en una variedad de plantas de cereales por medios bien conocidos en la técnica, incluyendo la transferencia directa de ADN o transformación mediada por *Agrobacterium tumefaciens*. Este último procedimiento se puede llevar a cabo por una variedad de medios, incluyendo, por ejemplo, el de la patente de EE.UU. No. 5.591.616, en el que se transforma un callo monocotiledóneo poniendo en contacto tejido desdiferenciado de *Agrobacterium* que contiene el vector de clonación.

10 **[0320]** 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 calloso se cultiva en un medio sólido a 23°C en la oscuridad durante algunos días. Los callos desarrollados 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. Después, los brotes se transfieren a medio de brote-alargamiento cada 2-3 semanas. Los brotes que parecen sanos se transfieren a medio
15 de enraizamiento y después de desarrollarse las raíces, las plantas se ponen en suelo de cultivo en maceta húmedo.

[0321] Después, se analiza en las plantas transformadas la presencia del gen NPTII/resistencia a la kanamicina por ELISA, usando el kit ELISA NPTII kit de 5Prime-3Prime Inc. (Boulder, CO).

20 **[0322]** También es rutinario el uso de otros procedimientos para producir plantas transgénicas de la mayoría de cultivos de cereales (Vasil (1994) *Plant Mol. Biol.* 25: 925-937) tales como maíz, trigo, arroz, sorgo (Cassas y col. (1993) *Proc. Natl. Acad. Sci. USA* 90: 11212-11216, y cebada (Wan y Lemeaux (1994) *Plant Physiol.* 104:37-48). Los procedimientos de transferencia de ADN tales como el procedimiento de microproyectiles se pueden usar para el maíz (Fromm y col. (1990) *Bio/Technol.* 8: 833-839; Gordon-Kamm y col. (1990) *Plant Cell* 2: 603-618; Ishida
25 (1990) *Nature Biotechnol.* 14:745-750), trigo (Vasil y col. (1992) *Bio/Technol.* 10:667-674; Vasil y col. (1993) *Bio/Technol.* 11:1553-1558; Weeks y col. (1993) *Plant Physiol.* 102:1077-1084), y arroz (Christou (1991) *Bio/Technol.* 9:957-962; Hiei y col. (1994) *Plant J.* 6:271-282; Aldemita y Hodges (1996) *Planta* 199:612-617; y Hiei y col. (1997) *Plant Mol. Biol.* 35:205-218). Para la mayoría de las plantas de cereales, las células embrionarias despojadas de los tejidos escutelares inmaduros, son las dianas celulares preferidas para la transformación (Hiei y col. (1997) *Plant Mol. Biol.* 35:205-218; Vasil (1994) *Plant Mol. Biol.* 25: 925-937). Para la transformación de células embrionarias de maíz despojadas de los tejidos escutelares inmaduros, usando el bombardeo de microproyectiles, se prefiere el genotipo A188XB73 (Fromm y col. (1990) *Bio/ Technol.* 8: 833-839; Gordon-Kamm y col. (1990) *Plant Cell* 2: 603-618). Después del bombardeo con microproyectiles, los tejidos se seleccionan en fosfotricina para
30 identificar las células embrionarias transgénicas (Gordon-Kamm y col. (1990) *Plant Cell* 2: 603-618). Las plantas transgénicas se regeneran mediante técnicas estándar de regeneración del maíz (Fromm y col. (1990) *Biol/ Technol.* 8: 833-839; Gordon-Kamm y col. (1990) *Plant Cell* 2: 603-618).

[0323] Se pueden usar los análisis de transferencia Northern, RT-PCR o micromatrices de las plantas transformadas y regeneradas para mostrar la expresión de G1792 y genes relacionados que son capaces de conferir
40 tolerancia al estrés biótico o abiótico.

[0324] Para verificar la capacidad de conferir tolerancia al estrés abiótico, las plantas maduras que expresan en exceso un factor de transcripción de la invención, o alternativamente, plántulas descendientes de estas plantas, se pueden estimular mediante un ensayo de estrés abiótico, tal como un ensayo de sequía, calor, alta salinidad o
45 helada, en condiciones de estrés osmótico que también pueden medir los sensores de azúcar alterados, tales como condiciones de alta concentración de azúcar, en un ensayo de tolerancia a la sombra, o en un ensayo de detección de C/N para identificar plantas con estrés alterado o tolerancia a la sombra o detección de C/N alterada. Comparando las plantas silvestres y transgénicas tratadas de forma similar, se puede mostrar que las plantas transgénicas tienen una mayor tolerancia al estrés abiótico.

50 **[0325]** Después de que una planta monocotiledónea o célula de planta se ha transformado (y ésta última se ha regenerado en una planta) y se ha mostrado que tiene una mayor tolerancia al estrés biótico o abiótico, o produce un mayor rendimiento con respecto a una planta de control en las condiciones de estrés, la planta monocotiledónea transformada se puede cruzar consigo misma o con una planta de la misma línea, una planta monocotiledónea no
55 transformada o silvestre, u otra planta monocotiledónea transformada de una línea de plantas transgénicas diferente.

Ejemplo XIV: Genes que confieren mejoras significativas a especies que no son *Arabidopsis*

[0326] Se ha analizado función de ortólogos específicos de los factores de transcripción de la invención y se

pueden caracterizar mejor mediante la incorporación en plantas de cultivo. La función de ortólogos específicos de las secuencias en la Lista de Secuencias se puede analizar mediante su expresión alterada (p. ej., expresión en exceso ectópica o inactivación de la expresión) en plantas, usando elementos reguladores constitutivos, inducibles o específicos de tejido, como se ha descrito antes. Estas secuencias incluyen secuencias de polinucleótidos encontradas en la Lista de Secuencias tales como por ejemplo: SE ID NO: 12 (G2133).

[0327] Las secuencias de polinucleótidos y polipéptidos derivadas de monocotiledóneas se pueden usar para plantas tanto monocotiledóneas como dicotiledóneas aleatoriamente, y las derivadas de dicotiledóneas se pueden usar para transformar cualquiera de los grupos, aunque algunas de estas secuencias funcionarán mejor si el gen se transforma en una planta del mismo grupo del que deriva la secuencia.

[0328] En estos ejemplos se proporcionan procedimientos de transformación, y se pueden usar en un vector de expresión. Después de introducir el vector en una célula de planta, la planta se puede regenerar a partir de la célula, después de lo cual se puede dejar que la planta exprese en exceso uno de los polipéptidos de la invención que tienen la propiedad de aumentar la tolerancia al estrés abiótico, tolerancia a la sombra o detección de C/N alterada en la planta transgénica. Las plantas con estos rasgos alterados se pueden identificar mediante comparación con plantas silvestres o no transformadas que no expresan en exceso el polipéptido, después de lo cual se pueden seleccionar una o más plantas con el grado deseado de uno o más de los rasgos mejorados. De esta forma, se pueden seleccionar las plantas con una mayor tolerancia a la sombra, mayor tolerancia al estrés abiótico, detección alterada de C/N, o más de uno de estos rasgos alterados.

[0329] Para los análisis relacionados con la tolerancia a la sequía, las semillas de estas plantas transgénicas se someten a ensayos de germinación para medir los sensores de sacarosa. Las semillas monocotiledóneas estériles, incluyendo, pero sin limitar, maíz, arroz, trigo, centeno y sorgo, así como las dicotiledóneas incluyendo, pero sin limitar, soja y alfalfa, se siembran en medio MS al 80% más vitamina con sacarosa al 9,4%; el medio de control carece de sacarosa. Después, todas las placas de ensayo se incuban a 22°C con 24 horas de luz, 120-130 $\mu\text{Ein}/\text{m}^2/\text{s}$, en una cámara de crecimiento. La evaluación de la germinación y el vigor de las plántulas se lleva a cabo entonces 3 días después de la plantación. Se pueden encontrar que las que tienen exceso de expresión de estas secuencias son más tolerantes a un contenido alto de sacarosa al tener mejor germinación, radículas más largas y mayor expansión de cotiledones. Estos resultados indicarían que el exceso de expresión de los ortólogos de la Lista de Secuencias está implicada en los sensores de azúcar específicos de sacarosa.

[0330] Las plantas que expresan en exceso estos ortólogos se pueden someter a ensayos de sequía basados en suelo para identificar aquellas líneas que son más tolerantes a la falta de agua que las plantas de control silvestres. En general, las plantas que expresan en exceso ortólogos aparecerán significativamente más grandes y más verdes, con menos marchitamiento o desecado, que las plantas de control silvestres, en particular después de un periodo de carencia de agua al que le sigue el regado de nuevo y un periodo de incubación posterior.

[0331] Las plantas que expresan en exceso estos ortólogos se pueden someter a ensayos de estrés abiótico para identificar aquellas líneas que son más tolerantes al estrés abiótico que las plantas de control silvestres en estas condiciones. En general, las plantas que expresan en exceso los ortólogos presentarán características morfológicas que están asociadas con un fenotipo de evitación de sombra (p. ej., cotiledones alterados, hipocotilos alterados, orientación de las hojas alterada, peciolo alterado, y/o fotomorfogénesis constitutiva) y pueden aparecer también más grandes, más verdes y más sanas que las plantas de control silvestres.

Ejemplo XV: identificación de secuencias ortólogas y parálogas

[0332] Los ortólogos de genes de *Arabidopsis* se pueden identificar por varios procedimientos, incluyendo la amplificación por hibridación o por procedimientos bioinformáticos. Este ejemplo describe cómo se pueden identificar homólogos del factor de transcripción CBF1 de la familia AP2 de *Arabidopsis*, que confieren tolerancia a estreses abióticos (Thomashow y col. (2002) patente de EE.UU. n° 6.417.428), y un ejemplo para confirmar la función de secuencias homólogas. En este ejemplo se encontraron ortólogos de CBF1 en la canola (*Brassica napus*) usando la reacción en cadena de la polimerasa (PCR)

[0333] Se designaron cebadores degenerados para las regiones del dominio de unión AP2 y fuera del AP2 (dominio carboxi terminal patente de EE.UU. n° 6.417.428):

Mol 368 (inverso) 5'- CAY CCN ATH TAY MGN GGN GT -3'

Mol 378 (directo) 5'- GGN ARN ARC ATN CCY TCN GCC -3'

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

5 **[0334]** 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 está fuera del dominio AP2. (dominio carboxilo terminal) (secuencia de aminoácidos: Met-Ala-Glu-Gly- Met-Leu-Pro).

10 **[0335]** El ADN genómico aislado de *B. napus* se amplificó usando estos cebadores siguiendo estas condiciones: una etapa de desnaturalización de 2 min a 93°C; 35 ciclos de 93°C durante 1 min, 55°C durante 1 min, y 72°C durante 1 min; y una incubación final de 7 min a 72°C al final de los ciclos.

15 **[0336]** Los productos de la PCR se separaron por electroforesis en un gel de agarosa al 1,2% y se transfirieron a membranas de nailon y se hibridaron con la sonda AT CBF1 preparada a partir de ADN genómico de *Arabidopsis* por amplificación por la PCR. Los productos hibridados se visualizaron mediante un sistema de detección colorimétrico (Boehringer Mannheim) y las correspondientes bandas de un gel de agarosa similar se aislaron usando el kit de extracción de Qiagen (Qiagen). Los fragmentos de ADN se ligaron en el vector clon TA del kit de clonación TOPO TA (Invitrogen) y se transformaron en la cepa de *E. coli* TOP10 (Invitrogen).

20 **[0337]** Se recogieron 7 colonias y los insertos se secuenciaron en un aparato ABI 377 a partir de ambas cadenas de sentido directo y sentido contrario después de aislar los plásmidos de ADN. La secuencia de ADN se editó mediante el secuenciador y se alineó con AtCBF1 con el software GCG y la búsqueda con Blast de NCBI.

25 **[0338]** La secuencia de ácido nucleico y las secuencias de aminoácidos de un ortólogo de canola encontrado de esta forma (bnCBF1; patente de EE.UU. n°: 6.417.428) identificado por este procedimiento, se muestra en la Lista de Secuencias.

30 **[0339]** Las secuencias de aminoácidos alineadas muestran que el gen bnCBF1 tiene una identidad de 88% con la secuencia de *Arabidopsis* en la región del dominio AP2 y una identidad de 85% con la secuencia de *Arabidopsis* fuera del dominio AP2 cuando se alinean para dos secuencias de inserción que están fuera del dominio AP2.

35 **[0340]** De forma similar, también se pueden identificar las secuencias parálogas de los genes de *Arabidopsis*, tales como *CBF1*.

[0341] Se han clonado dos parálogos de CBF1 de *Arabidopsis thaliana*: *CBF2* y *CBF3*. *CBF2* y *CBF3* se han clonado y secuenciado como se describe a continuación. Las secuencias de ADN y las proteínas codificadas se presentan en la patente de EE.UU. n° 6.417.428.

40 **[0342]** Se cribaron en 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) los clones recombinantes que llevaban insertos relacionados con el gen *CBF1* (Stockinger y col. (1997) *Proc. Natl. Acad. Sci.* 94: 1035-1040). *CBF1* se radiomarcó con ³²P por cebado aleatorio (Sambrook y col. véase antes) y se usó para cribar una biblioteca mediante la técnica de placa elevada usando condiciones de hibridación y lavado rigurosas (Hajela y col. (1990) *Plant Physiol.* 93:1246-1252; Sambrook y col. véase antes, 6 X tampón SSPE, 60°C para la hibridación y 0,1 X tampón SSPE y 60°C para los lavados). Se obtuvieron 12 clones de hibridación positivos y se determinaron las secuencias de ADN de los insertos de ADNc. Los resultados indican que los clones entran en 3 clases. Una clase lleva insertos que corresponden a *CBF1*. Las otras dos clases llevan secuencias que corresponden a dos homólogos diferentes de *CBF1*, denominados *CBF2* y *CBF3*. Las secuencias de ácidos nucleicos y secuencias que codifican la proteína predicha para *CBF1*, *CBF2*, *CBF3* de *Arabidopsis* y el ortólogo CBF de *Brassica napus* se exponen en la patente de EE.UU. n° 6.417.428.

50 **[0343]** Una comparación de las secuencias de ácidos nucleicos de *CBF1*, *CBF2* y *CBF3* de *Arabidopsis* indica que son de 83 a 85% idénticas como se muestran en la tabla 9.

55

TABLA 9

	Porcentaje de identidad ^a	
	ADN ^b	Polipéptido
cbf1/cbf2	85	86
cbf1/cbf2	83	84
cbf2/cbf3	84	85

^aEl porcentaje de identidad se determinó usando el algoritmo Clustal del programa MEGALIGN (DNASTAR, Inc.).
^bSe muestran las comparaciones de las secuencias de ácidos nucleicos de los marcos de lectura abiertos

[0344] Igualmente, las secuencias de aminoácidos de los 3 polipéptidos CBP están en el intervalo de identidad de 84 a 86%. Un alineamiento de las 3 secuencias de aminoácidos pone de manifiesto que la mayoría de las diferencias en la secuencia de aminoácidos se producen en la mitad C-terminal ácida del polipéptido. Esta región de CBF1 sirve como dominio de activación tanto en levaduras como en *Arabidopsis* (no se muestra).

[0345] Los restos 47 a 106 de CBF1 corresponden al dominio AP2 de la proteína, un patrón de unión al ADN que, hasta la fecha, sólo se había encontrado en proteínas de plantas. 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 pueden tener un efecto en la especificidad de unión al ADN.

Ejemplo XVI: Transformación de canola con un plásmido que contiene CBF1, CBF2 o CBF3

[0346] Después de identificar los genes homólogos de CHF1, la canola se transformó con un plásmido que contenía los genes CBF1, CBF2 o CBF3 de *Arabidopsis* clonados en el vector pGA643 (An (1987) *Methods Enzymol.* 253: 292). En estas construcciones, los genes CBF se expresaban de forma constitutiva bajo el promotor 35S de CaMV, además, el gen CBF1 se clonó bajo los controles del promotor COR15 de *Arabidopsis* en el mismo vector pGA643. Cada construcción se transformó en la cepa de *Agrobacterium* GV3101. Las agrobacterias transformadas se cultivaron durante 2 días en medio AB mínimo que contenía los antibióticos adecuados.

[0347] La canola de primavera (*B. napus* variedad cultivada Westar) se transformó usando el protocolo de Moloney y col. ((1989) *Plant Cell Reports* 8: 238) con algunas modificaciones como se describe. Brevemente, las semillas se esterilizaron y se cultivaron en placa en medio MS de media concentración, que contenía sacarosa al 1%. Las placas se cultivaron a 24°C con luz de 60-80 $\mu\text{E}/\text{m}^2/\text{s}$ usando un periodo de iluminación de 16 horas de luz/8 horas de oscuridad. Los cotiledones de las plántulas de 4-5 días de edad se recogieron, se cortaron los peciolos y se sumergieron en disolución de *Agrobacterium*. Los cotiledones sumergidos se pusieron en medio de cultivo simultáneo con una densidad de 20 cotiledones/placa y se incubaron como se ha descrito antes durante 30 días. Los explantes se transfirieron al mismo medio, pero que contenía 300 mg/l de timentina (SmithKline Beecham, PA) y se redujeron 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 desarrollaron los brotes. Los brotes se transfirieron a medio de brote-alargamiento cada 2-3 semanas. Los lotes que tenían un aspecto sano se transfirieron a medio de enraizamiento. Una vez que se desarrollaron raíces buenas, las plantas se plantaron en suelo para macetas húmedo.

[0348] Después se analizó en las plantas transformadas la presencia del gen NPTII/resistencia a la kanamicina mediante ELISA, usando el kit de ELISA NPTII de 5Prime-3Prime Inc. (Boulder, CO). Aproximadamente 70% de las plantas cribadas eran positivas para NPTII. Solamente estas plantas se siguieron analizando.

[0349] Los análisis de transferencia Northern de las plantas que se transformaron con las construcciones de expresión constitutiva mostraron expresión de los genes CBF y todos los genes CBF fueron capaces de inducir el gen BN115 regulado por el frío de *Brassica napus* (homólogo del gen COR15 de *Arabidopsis*). La mayor parte de las plantas transgénicas presentaron un fenotipo de crecimiento normal. Como se esperaba, las plantas transgénicas son más tolerantes a la helada que las plantas silvestres. Usando el ensayo de pérdida de electrolitos de las hojas, el control mostró una pérdida de 50% de -2 a -3°C. La canola de primavera transformada como CBF1 o CBF2 mostró una pérdida de 30% de -6 a -7°C. La canola de primavera transformadas con CBF3 mostró una pérdida de 50% de aproximadamente -10 a -15°C. La canola de invierno transformada con CBF3 puede presentar una pérdida de 50% de aproximadamente -16°C a -20°C. Además, si la canola de primavera o invierno se aclimatan al frío, las plantas transformadas pueden presentar una mayor aumento de la tolerancia a la helada de al menos -2°C.

[0350] Para ensayar la tolerancia a la salinidad de las plantas transformadas, las plantas se regaron con NaCl 150 mM. Las plantas que expresaban en exceso CBF1, CBF2 o CBF3 crecieron mejor comparadas con las plantas que no se habían transformado con CBF1, CBF2 o CBF3.

5

[0351] Estos resultados demuestran que los homólogos de factores de transcripción de *Arabidopsis* se pueden identificar y se muestra que confieren funciones similares en especies de plantas que no son *Arabidopsis*.

Ejemplo XVII: clonación de promotores de factores de transcripción

10

[0352] Los promotores se aislaron de genes de factores de transcripción que tienen patrones de expresión de genes útiles para una serie de aplicaciones, determinadas por los procedimientos bien conocidos en la técnica (incluyendo el análisis del perfil de transcritos con micromatrices de ADNc u oligonucleótidos, análisis de transferencia Northern, RT-PCR semicuantitativa o cuantitativa). Se ponen de manifiesto perfiles de expresión de genes interesantes determinando la abundancia de transcritos para un gen de factor de transcripción seleccionado, después de la exposición de las plantas a una serie de condiciones experimentales diferentes, y en una serie de tejidos o tipos de órganos diferentes, o etapas del desarrollo. Las condiciones experimentales a las que se exponen las plantas para este propósito incluyen frío, calor, sequía, estimulación osmótica, concentraciones de hormonas diferentes (ABA, GA, auxina, citoquinina, ácido salicílico, brasinoesteroide), estimulación con patógenos y plagas. Los tipos de tejidos y que etapas del desarrollo incluyen tallo, raíz, flor, hojas en roseta, hojas caulinares, silicuas, semillas en germinación y meristema. El conjunto de niveles de expresión proporciona un patrón que se determina mediante el elemento regulador del promotor génico.

15

20

[0353] Los promotores de factores de transcripción para los genes descritos en el presente documento se obtienen mediante clonación de 1,5 kb a 2,0 kb de secuencia genómica inmediatamente en la dirección 5' del codón de inicio de la traducción para la secuencia codificante de la proteína del factor de transcripción codificado. Esta región incluye la 5'-UTR del gen del factor de transcripción, que puede comprender elementos reguladores. La región de 1,5 kb a 2,0 kb se clona por procedimientos de PCR, usando cebadores que incluyen uno en la dirección 3' situado en el codón de inicio de la traducción (incluyendo una secuencia adaptadora adecuada), y otro en la dirección 5' situado a partir de 1,5 kb a 2,0 kb en la dirección 5' del condón de inicio de la traducción (incluyendo una secuencia adaptadora adecuada). Los fragmentos deseados se amplificaron por la PCR a partir del ADN genómico de *Arabidopsis* Col-0 usando ADN polimerasa Taq de alta fidelidad para minimizar la incorporación de mutación o mutaciones puntuales. Los cebadores de clonación incorporan dos sitios de restricción raros, tales como NotI y Sfi1, encontrados con frecuencia baja en el genoma de *Arabidopsis*. Se usan sitios de restricción adicionales en los casos en los que está presente un sitio de restricción NotI o Sfi1 dentro del promotor.

25

30

35

[0354] El fragmento de 1,5-2,0 kb en la dirección 5' desde el codón de inicio de la traducción, incluyendo la región 5' no traducida del factor de transcripción, se clona en un vector de transformación binario inmediatamente en la dirección 5' de un gen indicador adecuado, o un gen transactivador que es capaz de programar la expresión de un gen indicador en una segunda construcción génica. Los genes indicadores usados incluyen la proteína fluorescente verde (y variantes de color de la proteína fluorescente verde), β -glucuronidasa y luciferasa. Los genes transactivadores adecuados incluyen LexA-GAL4, junto con un indicador transactivable en un segundo plásmido binario (como se describe en la solicitud de patente de EE.UU. 09/958.131, incorporada en el presente documento por referencia). El o los plásmidos binarios se transfieren a *Agrobacterium* y la estructura del plásmido se confirma por PCR. Estas cepas se introducen en plantas de *Arabidopsis* como se describe en otros ejemplos, y los patrones de expresión de genes se determinan de acuerdo con procedimientos estándar conocidos para el experto en la materia, para el seguimiento de la fluorescencia de la GFP, actividad de β -glucuronidasa, o luminiscencia.

40

45

[0355] La presente invención no está limitada por las realizaciones específicas descritas en el presente documento. Habiéndose descrito ahora la invención completamente, será evidente para el experto en la materia que se pueden hacer muchos cambios y modificaciones de la misma sin salirse del alcance de las reivindicaciones adjuntas. Las modificaciones que son evidentes a partir de la descripción anterior y las figuras que acompañan entran dentro del alcance de las reivindicaciones.

50

55 LISTA DE SECUENCIAS

[0356]

<110> Mendel Biotechnology, Inc.

<120> REGULADORES TRANSCRIPCIONALES DE PLANTAS

<130> AHB/FP6379341

<160> 64
 <170> PatentIn versión 3.2
 <210> 1
 <211> 785
 5 <212> ADN
 <213> Arabidopsis thaliana
 <220>
 <223> secuencia de referencia G47; identificador de clado
 <400> 1

```

cttcttcttc acatcgatca tcatacaaca acaaaaaatg gattacagag aatccaccgg      60
tgaaagtcag tcaaagtaca aaggaatccg tcgctcgaaa tggggcaaat gggtagatcaga    120
gattagagtt ccgggaactc gtgaccgtct ctggttaggt tcattctcaa cagcagaagg      180
tgccgccgta gcacacgacg ttgctttctt ctgtttacac caacctgatt ctttagaatc     240
tctcaatttc cctcatttgc ttaatccttc actcgtttcc agaacttctc cgagatctat     300
ccagcaagct gcttctaacg ccggcatggc cattgacgcc ggaatcgcc acagtaccag       360
cgtgaactct ggatgcccag atacgacgac gtattacgag aatggagctg atcaagtgga      420
gccgttgaat atttcagtgt atgattatct gggcggccac gatcacgttt gatttatctc     480
gacggtcatg atcacgtttg atcttctttt gagtaagatt ttgtaccata atcaaaacag     540
gtgtggtgct aaaatcttac tcaaaacaag attaggtacc acagagaaac aatcaaatgg      600
ttgtgaatat acattataag gttttgatta atgtttgttt cactgattta gtgaagtttg     660
gtccattgta taaaaatcta ttcaagaaac ctagcgcgag atcatgtttc gtgattgaag     720
attgagattt ttaagtattc gtaatatatt tgtaaaatac aaataaaaaa aaaaaaaaaa     780
aaaaa                                                                    785
    
```

10 <210> 2
 <211> 144
 <212> PRT
 <213> Arabidopsis thaliana
 15 <220>
 <223> secuencia de referencia del polipéptido G47; identificador de clado
 <400> 2

```

Met Asp Tyr Arg Glu Ser Thr Gly Glu Ser Gln Ser Lys Tyr Lys Gly
 1           5           10           15
    
```

Ile Arg Arg Arg Lys Trp Gly Lys Trp Val Ser Glu Ile Arg Val Pro
 20 25 30

Gly Thr Arg Asp Arg Leu Trp Leu Gly Ser Phe Ser Thr Ala Glu Gly
 35 40 45

Ala Ala Val Ala His Asp Val Ala Phe Phe Cys Leu His Gln Pro Asp
 50 55 60

Ser Leu Glu Ser Leu Asn Phe Pro His Leu Leu Asn Pro Ser Leu Val
 65 70 75 80

Ser Arg Thr Ser Pro Arg Ser Ile Gln Gln Ala Ala Ser Asn Ala Gly
 85 90 95

Met Ala Ile Asp Ala Gly Ile Val His Ser Thr Ser Val Asn Ser Gly
 100 105 110

Cys Gly Asp Thr Thr Thr Tyr Tyr Glu Asn Gly Ala Asp Gln Val Glu
 115 120 125

Pro Leu Asn Ile Ser Val Tyr Asp Tyr Leu Gly Gly His Asp His Val
 130 135 140

<210> 3
 <211> 1449
 <212> ADN

5 <213> Arabidopsis thaliana

<220>
 <223> secuencia de referencia G922; identificador de clado
 <400> 3

```

atggtggcta tgtttcaaga agataatgga acatcttctg tagcttcac accacttcaa      60
gtcttctcaa ctatgtcact caacagaccg actctctctg cttcttcac tccgtttcat      120
tgtctcaaag atctcaaacc agaggagcgt ggtctctact taatccacct cttgctaact      180
tgtgccaacc acgtggcttc aggtagcctc caaaacgcta acgcagcgtc cgagcagctc      240
tctcaccctg cttctcctga cggcgacacg atgcagcgaa tcgctgctta cttcaccgaa      300
gcgcttgcta acagaatcct taagtctctg cctggtcttt acaagctct taacgcaact      360
cagacaagaa ctaacaatgt ctctgaggag attcatgta gaagactctt cttgagatg      420
ttcccgatac tcaaagtctc ttacttctct actaatcgag ctatactcga ggctatggaa      480
ggagagaaga tggttcatgt gattgatctc gatgcttctg agccagctca atggcttgct      540
ttgcttcaag cttttaactc taggcctgaa ggtccacctc atttgagaat cactgggtgt      600
catcaccaga aggaagtgct tgaacaaatg gctcatagac tcattgagga agcagagaaa      660
ctcgatatcc cgtttcagtt taatcccgtt gtgagtaggt tagactggtt aatgtagaa      720
    
```

```

cagttgctggg ttaaacacagg agaggcctta gccgttagct cggttcttca attgcatacc 780
ttcttggcct ctgatgatga tctcatgaga aagaactgcg ctttacgggt tcagaacaac 840
cctagtggag ttgacttgca gagagttcta atgatgagcc atggetctgc agctgaggca 900
cgtgagaatg atatgagtaa caacaatggg tatagcccta gcggtgactc ggccctcatc 960
ttgcctttac caagttcagg aaggactgat agcttctca atgctatttg gggtttctc 1020
ccaaaggcca tgggtgtcac tgagcaagac tcagaccaca acggctccac actaatggag 1080
aggctattag aatcacttta cacctacgca gcattgtttg attgcttggg aacaaaagt 1140
ccaagaactg ctcaagatag gatcaaagtg gagaagatgc tcttcgggga ggagatcaag 1200
aacatcatat cctcggaggg atttgagaga agagaaagac acgagaagct tgagaaatgg 1260
agccagagga tcgatttggc tggttttggg aatgttctc ttagctatta tgcgatgttg 1320
caggetagga gattgcttca agggtgcggg tttgatgggt atagaatcaa ggaagagagc 1380
gggtgcgcag taatttctg gcaagatcga cctctatact cggtatcagc ttggagatgc 1440
aggaagtga 1449
    
```

<210> 4

<211> 482

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G922; identificador de clado

<400> 4

```

Met Val Ala Met Phe Gln Glu Asp Asn Gly Thr Ser Ser Val Ala Ser
 1           5           10           15

Ser Pro Leu Gln Val Phe Ser Thr Met Ser Leu Asn Arg Pro Thr Leu
          20           25           30

Leu Ala Ser Ser Ser Pro Phe His Cys Leu Lys Asp Leu Lys Pro Glu
          35           40           45

Glu Arg Gly Leu Tyr Leu Ile His Leu Leu Leu Thr Cys Ala Asn His
 50           55           60

Val Ala Ser Gly Ser Leu Gln Asn Ala Asn Ala Ala Leu Glu Gln Leu
 65           70           75           80

Ser His Leu Ala Ser Pro Asp Gly Asp Thr Met Gln Arg Ile Ala Ala
          85           90           95

Tyr Phe Thr Glu Ala Leu Ala Asn Arg Ile Leu Lys Ser Trp Pro Gly
          100          105          110
    
```

Leu Tyr Lys Ala Leu Asn Ala Thr Gln Thr Arg Thr Asn Asn Val Ser
 115 120 125
 Glu Glu Ile His Val Arg Arg Leu Phe Phe Glu Met Phe Pro Ile Leu
 130 135 140
 Lys Val Ser Tyr Leu Leu Thr Asn Arg Ala Ile Leu Glu Ala Met Glu
 145 150 155 160
 Gly Glu Lys Met Val His Val Ile Asp Leu Asp Ala Ser Glu Pro Ala
 165 170 175
 Gln Trp Leu Ala Leu Leu Gln Ala Phe Asn Ser Arg Pro Glu Gly Pro
 180 185 190
 Pro His Leu Arg Ile Thr Gly Val His His Gln Lys Glu Val Leu Glu
 195 200 205
 Gln Met Ala His Arg Leu Ile Glu Glu Ala Glu Lys Leu Asp Ile Pro
 210 215 220
 Phe Gln Phe Asn Pro Val Val Ser Arg Leu Asp Cys Leu Asn Val Glu
 225 230 235 240
 Gln Leu Arg Val Lys Thr Gly Glu Ala Leu Ala Val Ser Ser Val Leu
 245 250 255
 Gln Leu His Thr Phe Leu Ala Ser Asp Asp Asp Leu Met Arg Lys Asn
 260 265 270
 Cys Ala Leu Arg Phe Gln Asn Asn Pro Ser Gly Val Asp Leu Gln Arg
 275 280 285
 Val Leu Met Met Ser His Gly Ser Ala Ala Glu Ala Arg Glu Asn Asp
 290 295 300
 Met Ser Asn Asn Asn Gly Tyr Ser Pro Ser Gly Asp Ser Ala Ser Ser
 305 310 315 320
 Leu Pro Leu Pro Ser Ser Gly Arg Thr Asp Ser Phe Leu Asn Ala Ile
 325 330 335
 Trp Gly Leu Ser Pro Lys Val Met Val Val Thr Glu Gln Asp Ser Asp
 340 345 350
 His Asn Gly Ser Thr Leu Met Glu Arg Leu Leu Glu Ser Leu Tyr Thr

```

                355                360                365
Tyr Ala Ala Leu Phe Asp Cys Leu Glu Thr Lys Val Pro Arg Thr Ser
 370                375                380

Gln Asp Arg Ile Lys Val Glu Lys Met Leu Phe Gly Glu Glu Ile Lys
385                390                395                400

Asn Ile Ile Ser Cys Glu Gly Phe Glu Arg Arg Glu Arg His Glu Lys
                405                410                415

Leu Glu Lys Trp Ser Gln Arg Ile Asp Leu Ala Gly Phe Gly Asn Val
                420                425                430

Pro Leu Ser Tyr Tyr Ala Met Leu Gln Ala Arg Arg Leu Leu Gln Gly
                435                440                445

Cys Gly Phe Asp Gly Tyr Arg Ile Lys Glu Glu Ser Gly Cys Ala Val
                450                455                460

Ile Cys Trp Gln Asp Arg Pro Leu Tyr Ser Val Ser Ala Trp Arg Cys
465                470                475                480

Arg Lys

```

<210> 5
 <211> 585
 <212> ADN

5 <213> Arabidopsis thaliana

<220>
 <223> secuencia de referencia G1274; identificador de clado
 <400> 5

```

atgaatatct ctcaaaaccc tagccctaatt ttacgtact tctccgatga aaactttatt      60
aatccgttta tggataacaa cgatttctca aatttgatgt tctttgacat agatgaagga      120
ggtaacaatg gattaatcga ggaagagatc tcattctccga caagcatcgt ttcgctggag      180
acatttaccg gggaaagcgg cggatccggc agcgcacaaa cgttgagtaa aaaggaatca      240
actaatagag gaagtaaaga gagtgatcag acgaaggaga cgggtcatcg agttgcattt      300
agaacgagat cgaagattga tgtgatggat gatggtttta aatggaggaa gtatggcaag      360
aaatctgtca aaaacaacat taacaagagg aattactaca aatgctcaag tgaaggttgc      420
tcggtgaaga agagggtaga gagagatggt gacgatgcag cttatgtaat tacaacatat      480
gaaggagtcc ataaccatga gagtctctct aatgtctatt acaatgaaat ggttttatct      540
tatgatcatg ataactggaa ccaacactct cttcttcgat cttaa                          585

```

10 <210> 6
 <211> 194
 <212> PRT
 <213> Arabidopsis thaliana
 <220>

<223> secuencia de referencia del polipéptido G1274; identificador de clado
 <400> 6

```

Met Asn Ile Ser Gln Asn Pro Ser Pro Asn Phe Thr Tyr Phe Ser Asp
1          5          10          15

Glu Asn Phe Ile Asn Pro Phe Met Asp Asn Asn Asp Phe Ser Asn Leu
20          25          30

Met Phe Phe Asp Ile Asp Glu Gly Gly Asn Asn Gly Leu Ile Glu Glu
35          40          45

Glu Ile Ser Ser Pro Thr Ser Ile Val Ser Ser Glu Thr Phe Thr Gly
50          55          60

Glu Ser Gly Gly Ser Gly Ser Ala Thr Thr Leu Ser Lys Lys Glu Ser
65          70          75          80

Thr Asn Arg Gly Ser Lys Glu Ser Asp Gln Thr Lys Glu Thr Gly His
85          90          95

Arg Val Ala Phe Arg Thr Arg Ser Lys Ile Asp Val Met Asp Asp Gly
100         105         110

Phe Lys Trp Arg Lys Tyr Gly Lys Lys Ser Val Lys Asn Asn Ile Asn
115         120         125

Lys Arg Asn Tyr Tyr Lys Cys Ser Ser Glu Gly Cys Ser Val Lys Lys
130         135         140

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

Glu Gly Val His Asn His Glu Ser Leu Ser Asn Val Tyr Tyr Asn Glu
165         170         175

Met Val Leu Ser Tyr Asp His Asp Asn Trp Asn Gln His Ser Leu Leu
180         185         190

Arg Ser
    
```

<210> 7

5 <211> 696

<212> ADN

<213> Arabidopsis thaliana

<220>

<223> secuencia de referencia G1792; identificador de clado

10 <400> 7

```

aatccataga tctcttatta aataacagtg ctgaccaagc tcttacaag caaaccaatc    60
tagaacacca aagttaatgg agagctcaaa caggagcagc aacaaccaat cacaagatga    120
caagcaagct cgtttccggg gagttcgaag aaggccttgg ggaaagtttg cagcagagat    180
tcgagacccg tcgagaaacg gtgcccgctct ttggctcggg acatttgaga ccgctgagga    240
ggcagcaagg gcttatgacc gagcagcctt taaccttagg ggtcatctcg ctatactcaa    300
cttccctaat gagtattatc cacgtatgga cgactactcg cttcgccctc cttatgcttc    360
ttcttcttcg tcgctgctcat cgggttcaac ttctactaat gtgagtcgac aaaaccaaaag    420
agaagtttcc gagtttgagt atttggacga taaggttctt gaagaacttc ttgattcaga    480
agaaaggaag agataatcac gattagtttt gttttgatat tttatgtggc actgttgttg    540
ctacctacgt gcattatgtg catgtatagg tcgcttgatt agtactttat aacatgcatg    600
ccacgaccat aaattgtaag agaagacgta ctttgcgttt tcatgaaata tgaatgttag    660
atggtttgag tacaaaaaaaa aaaaaaaaaa aaaaaa                                696
    
```

<210> 8

<211> 139

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G1792; identificador de clado

<400> 8

```

Met Glu Ser Ser Asn Arg Ser Ser Asn Asn Gln Ser Gln Asp Asp Lys
 1                               5                               10                               15

Gln Ala Arg Phe Arg Gly Val Arg Arg Arg Pro Trp Gly Lys Phe Ala
 20                               25                               30

Ala Glu Ile Arg Asp Pro Ser Arg Asn Gly Ala Arg Leu Trp Leu Gly
 35                               40                               45

Thr Phe Glu Thr Ala Glu Glu Ala Ala Arg Ala Tyr Asp Arg Ala Ala
 50                               55                               60

Phe Asn Leu Arg Gly His Leu Ala Ile Leu Asn Phe Pro Asn Glu Tyr
 65                               70                               75                               80

Tyr Pro Arg Met Asp Asp Tyr Ser Leu Arg Pro Pro Tyr Ala Ser Ser
 85                               90                               95

Ser Ser Ser Ser Ser Ser Gly Ser Thr Ser Thr Asn Val Ser Arg Gln
 100                              105                              110

Asn Gln Arg Glu Val Phe Glu Phe Glu Tyr Leu Asp Asp Lys Val Leu
 115                              120                              125

Glu Glu Leu Leu Asp Ser Glu Glu Arg Lys Arg
 130                              135
    
```

10

<210> 9

<211> 1029

<212> ADN
 <213> Arabidopsis thaliana
 <220>
 <223> secuencia de referencia G2053; identificador de clado

5 <400> 9

```

atggagaatc cggtagggttt aagattccgt ccgaccgaca aggagatcgt cgtcgtattac    60
ctccgaccaa aaaactccga tagggacacg agtcatgttg atcgagtcac tagcacagtc    120
actatccgta gcttcgaccc ttgggagtta ccttgccagt ctaggatcaa actgaaagat    180
gagtccttggg gtttcttcag ccctaaggag aacaaatctg gcagaggtga tcaacaaatt    240
agaaaaacga aatctgggta ctggaagatt actggcaaac caaagcctat cttgcgtaac    300
cgccaagaga tcggtgagaa aaagggtttg atgttttaca tgagtaagga acttgggtgga    360
tccaaatccg actgggttat gcacgagtac catgctttct ctcctactca gatgatgatg    420
acatatacaa tatgtaaagt tatgtttaag ggtgacgtga gagagatttc ttcttcttct    480
gcttcttatg gtagtgaat tgacgagagc cgtgactcct taatccctct tcttgtgaac    540
gattctgagg aagaagctca aatcgaggat gctataccaa tagaggaatg ggaaacatgg    600
ttgactgatg atggtgttga tgacgaggtg aatcatatta tgaatatgaa agatgatcgc    660
aacaaccaca ggcctcaaaa gccattgact ggtgtcttga ttgacgatag tagtgatgat    720
gatgatgatt ctgatttgct atctccaaca acaaattcta ttgaaaattc gagcacttgt    780
gatagttttg gtagctcaga ccaaataaac ttagtgtcac taactcaaga ggtgagccaa    840
gctctgataa ccagtattga tacacccgag aagattaaga gtccttatga tgatgcacaa    900
gggactgggg ctggagggca aaaattgggt caagagactc gagagaagaa acgagctggt    960
ttctttcaca ggatgataca aatattcgtc aagaaaaatt accaatgttc ttctatctca   1020
agaacataa                                     1029
    
```

<210> 10
 <211> 342
 <212> PRT
 10 <213> Arabidopsis thaliana
 <220>
 <223> secuencia de referencia de polipéptido G2053; identificador de clado
 <400> 10

Met Glu Asn Pro Val Gly Leu Arg Phe Arg Pro Thr Asp Lys Glu Ile
1 5 10 15

Val Val Asp Tyr Leu Arg Pro Lys Asn Ser Asp Arg Asp Thr Ser His
20 25 30

Val Asp Arg Val Ile Ser Thr Val Thr Ile Arg Ser Phe Asp Pro Trp
35 40 45

Glu Leu Pro Cys Gln Ser Arg Ile Lys Leu Lys Asp Glu Ser Trp Cys
50 55 60

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

Arg Lys Thr Lys Ser Gly Tyr Trp Lys Ile Thr Gly Lys Pro Lys Pro
85 90 95

Ile Leu Arg Asn Arg Gln Glu Ile Gly Glu Lys Lys Val Leu Met Phe
100 105 110

Tyr Met Ser Lys Glu Leu Gly Gly Ser Lys Ser Asp Trp Val Met His
115 120 125

Glu Tyr His Ala Phe Ser Pro Thr Gln Met Met Met Thr Tyr Thr Ile
130 135 140

Cys Lys Val Met Phe Lys Gly Asp Val Arg Glu Ile Ser Ser Ser Ser
145 150 155 160

Ala Ser Tyr Gly Ser Glu Ile Glu Gln Ser Arg Asp Ser Leu Ile Pro
165 170 175

Leu Leu Val Asn Asp Ser Glu Glu Glu Ala Gln Ile Glu Asp Ala Ile
180 185 190

Pro Ile Glu Glu Trp Glu Thr Trp Leu Thr Asp Asp Gly Val Asp Glu
195 200 205

Gln Val Asn His Ile Met Asn Met Lys Asp Asp Arg Asn Asn His Arg
210 215 220

Pro Gln Lys Pro Leu Thr Gly Val Leu Ile Asp Asp Ser Ser Asp Asp
 225 230 235 240

Asp Asp Asp Ser Asp Leu Leu Ser Pro Thr Thr Asn Ser Ile Glu Asn
 245 250 255

Ser Ser Thr Cys Asp Ser Phe Gly Ser Ser Asp Gln Ile Asn Leu Val
 260 265 270

Ser Leu Thr Gln Glu Val Ser Gln Ala Leu Ile Thr Ser Ile Asp Thr
 275 280 285

Pro Glu Lys Ile Lys Ser Pro Tyr Asp Asp Ala Gln Gly Thr Gly Ala
 290 295 300

Gly Gly Gln Lys Leu Gly Gln Glu Thr Arg Glu Lys Lys Arg Ala Gly
 305 310 315 320

Phe Phe His Arg Met Ile Gln Ile Phe Val Lys Lys Ile His Gln Cys
 325 330 335

Ser Ser Ile Ser Arg Thr
 340

<210> 11
 <211> 485
 <212> ADN

5 <213> Arabidopsis thaliana

<220>
 <223> la secuencia de polipéptido predicha G2133 es paróloga de G47
 <400> 11

```

atctcatctt catccaccca aaaacatgga ttcaagagac accggagaaa ctgaccagag    60
caagtacaaa ggtatccgtc gtcggaaatg gggaaaatgg gtatcagaga ttcgtgtccc    120
gggaactcgt caacgtctct ggttaggctc tttctccacc gcagaaggcg ctgccgtagc    180
ccacgacgtc gcttttact gcttgacccg accatcttcc ctcgacgacg aatcttttaa    240
cttccctcac ttacttaca cctccctcgc ctccaatata tctcctaagt ccatccaaaa    300
agctgcttcc gacgccggca tggccgtgga cgccggattc catggtgctg tgtctgggag    360
tggtggttqt gaagagagat ctccatggc gaatatggag gaggaggaca aacttagtat    420
ctcctgtgat gattatcttg aagacgatct cgtttgatct atacgagtac gtttttagca    480
gttaa                                                                    485
    
```

10 <210> 12
 <211> 143
 <212> PRT
 <213> Arabidopsis thaliana
 <220>

15 <223> polipéptido G2133 parólogo de G47
 <400> 12

Met Asp Ser Arg Asp Thr Gly Glu Thr Asp Gln Ser Lys Tyr Lys Gly
 1 5 10 15

Ile Arg Arg Arg Lys Trp Gly Lys Trp Val Ser Glu Ile Arg Val Pro
 20 25 30

Gly Thr Arg Gln Arg Leu Trp Leu Gly Ser Phe Ser Thr Ala Glu Gly
 35 40 45

Ala Ala Val Ala His Asp Val Ala Phe Tyr Cys Leu His Arg Pro Ser
 50 55 60

Ser Leu Asp Asp Glu Ser Phe Asn Phe Pro His Leu Leu Thr Thr Ser
 65 70 75 80

Leu Ala Ser Asn Ile Ser Pro Lys Ser Ile Gln Lys Ala Ala Ser Asp
 85 90 95

Ala Gly Met Ala Val Asp Ala Gly Phe His Gly Ala Val Ser Gly Ser
 100 105 110

Gly Gly Cys Glu Glu Arg Ser Ser Met Ala Asn Met Glu Glu Glu Asp
 115 120 125

Lys Leu Ser Ile Ser Val Tyr Asp Tyr Leu Glu Asp Asp Leu Val
 130 135 140

<210> 13
 <211> 891
 <212> ADN

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia G2999: identificador de clado

<400> 13

tattagggtt ttctttatcc agaagcttca agatcagaaa agatggaagt tagagagaag 60

aaagatgaga aaatggagat gacaagaaga aaatcatcag ctctagatca tcatcgcttt 120

cctccgtaca ctactccca aaccgccaac aaagagaaac ccaccaccaa aagaaacggg 180

tcggaccggg atccggatcc ggatcttgat actaacccaa tttctatata tcatgctcca 240

agatcatatc ctaggcctca aactacttct ccggggaagg caaggtatag agaatgcca 300

```

aagaatcacg cggcgagctc cggtaggacac gttgtcgacg gttgcggcga gtttatgtct 360
tccggggaag aagcacggt ggagctctctt ctttgtgcag cttgtgattg ccaccggagc 420
ttccatagaa aagagatcga tggtttgttt gttgtcaatt tcaatagctt tggccattca 480
cagagaccac tcggaagccg ccacgtgtct ccgataatga tgagttttgg tgggtggtgt 540
ggttgtgcgg cggagtcctc gacggaggat ctaaacaagt ttcacatc ttttagtggt 600
tatggagtgg atcagtttca tcattatcaa cccaagaaac ggtttaggac aaagttaaat 660
gaagaacaga aggagaagat gatggagttt gctgagaaga ttggttgag aatgacaaaa 720
ctagaagatg atgaagtcaa tcggttttgt cgtgagatta aagttaaaag acaagtgttt 780
aaagtttggg tgcacaacaa taaacaagca gcaaagaaga aggatttga agagatttct 840
atatatatat atatactatt tttactctc tttgttcaac attgctttaa c 891

```

<210> 14

<211> 262

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia de polipéptido G2999; identificador de clado

<400> 14

```

Met Glu Val Arg Glu Lys Lys Asp Glu Lys Met Glu Met Thr Arg Arg
 1                               5 10 15

Lys Ser Ser Ala Leu Asp His His Arg Leu Pro Pro Tyr Thr Tyr Ser
 20 25 30

Gln Thr Ala Asn Lys Glu Lys Pro Thr Thr Lys Arg Asn Gly Ser Asp
 35 40 45

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

Ala Pro Arg Ser Tyr Ala Arg Pro Gln Thr Thr Ser Pro Gly Lys Ala
 65 70 75 80

Arg Tyr Arg Glu Cys Gln Lys Asn His Ala Ala Ser Ser Gly Gly His
 85 90 95

Val Val Asp Gly Cys Gly Glu Phe Met Ser Ser Gly Glu Glu Gly Thr
 100 105 110

Val Glu Ser Leu Leu Cys Ala Ala Cys Asp Cys His Arg Ser Phe His
 115 120 125

```

Arg Lys Glu Ile Asp Gly Leu Phe Val Val Asn Phe Asn Ser Phe Gly
 130 135 140

His Ser Gln Arg Pro Leu Gly Ser Arg His Val Ser Pro Ile Met Met
 145 150 155 160

Ser Phe Gly Gly Gly Gly Gly Cys Ala Ala Glu Ser Ser Thr Glu Asp
 165 170 175

Leu Asn Lys Phe His Gln Ser Phe Ser Gly Tyr Gly Val Asp Gln Phe
 180 185 190

His His Tyr Gln Pro Lys Lys Arg Phe Arg Thr Lys Phe Asn Glu Glu
 195 200 205

Gln Lys Glu Lys Met Met Glu Phe Ala Glu Lys Ile Gly Trp Arg Met
 210 215 220

Thr Lys Leu Glu Asp Asp Glu Val Asn Arg Phe Cys Arg Glu Ile Lys
 225 230 235 240

Val Lys Arg Gln Val Phe Lys Val Trp Met His Asn Asn Lys Gln Ala
 245 250 255

Ala Lys Lys Lys Asp Leu
 260

- <210> 15
- <211> 1621
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> secuencia de referencia G3086; identificador de clado
- <400> 15

```

aagtttctct cacgttctct tttttaattt taatttctcg ccggaacaa tctcatctcc    60
cggcgaacga aacttccggt gtggtactgc aaacggagaa aaaaataacc aaagaagaga    120
gaaactcaaa agctactaag atggaatcag aattccagca acatcacttc cttctccacg    180
atcatcaaca ccagagacca agaaactcag gattgattcg ttaccaatca gcaccaagtt    240
cgtacttttc gagtttcggt gaatcaatcg aagagttttt agatcgaccc acaagtctctg    300
aaactgagcg aatcttatct ggctttttac aaaccaccga cacaagcgac aacgttgata    360
gtttccttca ccatactttt aacagtgatg gaactgagaa gaaacctccg gaagttaaaa    420
cagaggacga agatgctgaa attccggtga ctgacgacggc gacggcgatg gaggttgttg    480
tttccggtga tggtgaaatc tcagtgaaatc ctgaagtatc gattgggtat gtggcttcgg    540
    
```

```

tttcgaggaa taagaqacca agagagaaag atgatcggac tccggatgaat aatctagctc 600
gtcataatag ttcaccggcc ggattatfff catccattga tgttgaaaca gcttatgcag 660
ctgtaatgaa aagtatggga ggttttggag gaagtaatgt gatgagtaca agcaatactg 720
aagettcgtc tcttactcct agaagcaagt tacttcctcc tacttctaga gcgatgagtc 780
cgatctctga ggttgatggt aaaccgggtt tctcgtctag attgcctcct cggacgcttt 840
ccggtggggt taatcgttct ttggggaatg aagggtctgc ttcttccaag cttacagctc 900
ttgctagac ccaatctgga ggtctagatc aatacaaaac caaggatgag gattcagcaa 960
gtagacgtcc tcctttggca catcacatga gtttgcccaa gtctttatca gatattgaac 1020
agttactgtc agattctatc ccatgtaaga tcagagccaa gcgggggtgt gcaactcctc 1080
ctcgaagcat agccgagagg gtgagaagaa ccaagatcag tgaaagaatg aggaagctgc 1140
aagaccttgt tccaacatg gacacgcaa caaacacagc agacatggtg gatcttgcgg 1200
ttcaatacat caaggacctg caagaacaag tgaaggcgtc cgaagagagt cgggcaagat 1260
gtagatgctc tagtgcgtga aactagaagt gggagtatgc gcgagtgcta gccagggagg 1320
gagttgtgca tagaagtatc ggttcggcct ttggagaaaa gtcgaagata gcaaagtaga 1380
gaagagatca tgaacaaagc taaatttggg ggtggtggtg aaaagggttt ttgtaaagtt 1440
ggaaaccttt tttggtaggg aagaaaagtag caagggtgtg taatggtccg aactccaatg 1500
ctattgtatg ttcttcatc caaaaaaaaaa aaaagcagaa gagaagtgat gtacaatgag 1560
aagcttattt tattgtaact caagaataat tcaggtaatt agtttaaaaa aaaaaaaaaa 1620
a 1621

```

- <210> 16
- <211> 379
- <212> PRT
- 5 <213> Arabidopsis thaliana
- <220>
- <223> secuencia de referencia del polipéptido G3086; identificador de clado
- <400> 16

```

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

His Gln Arg Pro Arg Asn Ser Gly Leu Ile Arg Tyr Gln Ser Ala Pro
20           25           30

Ser Ser Tyr Phe Ser Ser Phe Gly Glu Ser Ile Glu Glu Phe Leu Asp
35           40           45

Arg Pro Thr Ser Pro Glu Thr Glu Arg Ile Leu Ser Gly Phe Leu Gln

```



```

Arg Ala Lys Arg Gly Cys Ala Thr His Pro Arg Ser Ile Ala Glu Arg
305                               310                               315                               320

Val Arg Arg Thr Lys Ile Ser Glu Arg Met Arg Lys Leu Gln Asp Leu
                               325                               330                               335

Val Pro Asn Met Asp Thr Gln Thr Asn Thr Ala Asp Met Leu Asp Leu
                               340                               345                               350

Ala Val Gln Tyr Ile Lys Asp Leu Gln Glu Gln Val Lys Ala Leu Glu
                               355                               360                               365

Glu Ser Arg Ala Arg Cys Arg Cys Ser Ser Ala
370                               375

```

- <210> 49
- <211> 729
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> la secuencia de polipéptido predicha G2992 es paróloga de G2999
- <400> 49

```

atggatttgt cttccaaacc acaacaacaa cttctaaact ctcttcccat cgcgagagaa      60
ctaactgtca ccggagaagt ggggtgttgt tacaaagagt gtttgaaaaa ccacgcggct      120
aatctcggcg gccatgtctt cgacggttgc ggcgagtta tgccgagtcc cacggctact      180
tccaccgatc cttcttctct ccgttgcgct gcttgtgct gccaccgtaa cttccaccgc      240
cgtgaccctt ccgagaatct caacttcttc accgcgcttc cgatttcttc tccctccggc      300
actgagtcgc cgcggtctcg tcacgtgtct tctctcttct cttgctctta ctacacctca      360
gctctccac atcacgtgat tctctctctc agctccggtt tccctggacc gtcagatcaa      420
gatccaacgg tggttaggtc agagaacagt tcaagaggag caatgaggaa acgaacaaga      480
accaaattca cgcgagaca gaagatcaag atgagagcct tcgcagagaa agctggttgg      540
aaaatcaacg gatgtgatga aaagtcggtg agagagtttt gtaacgaagt tgggattgag      600
agaggagttc ttaaagtgtg gatgcataac aacaagtact cacttctcaa tggcaagatc      660
agagagatcg aacatggcct gtgtctgaac actcacagca atgatggtga tgggtcttca      720
tcttcttga                                     729

```

- 10 <210> 50
- <211> 242
- <212> PRT
- <213> Arabidopsis thaliana
- <220>
- 15 <223> polipéptido G2992 parólogo de G2999
- <400> 50

Met Asp Leu Ser Ser Lys Pro Gln Gln Gln Leu Leu Asn Ser Leu Pro
1 5 10 15

Ile Ala Gly Glu Leu Thr Val Thr Gly Glu Met Gly Val Cys Tyr Lys
20 25 30

Glu Cys Leu Lys Asn His Ala Ala Asn Leu Gly Gly His Ala Leu Asp
35 40 45

Gly Cys Gly Glu Phe Met Pro Ser Pro Thr Ala Thr Ser Thr Asp Pro
50 55 60

Ser Ser Leu Arg Cys Ala Ala Cys Gly Cys His Arg Asn Phe His Arg
65 70 75 80

Arg Asp Pro Ser Glu Asn Leu Asn Phe Leu Thr Ala Pro Pro Ile Ser
85 90 95

Ser Pro Ser Gly Thr Glu Ser Pro Pro Ser Arg His Val Ser Ser Pro
100 105 110

Val Pro Cys Ser Tyr Tyr Thr Ser Ala Pro Pro His His Val Ile Leu
115 120 125

Ser Leu Ser Ser Gly Phe Pro Gly Pro Ser Asp Gln Asp Pro Thr Val
130 135 140

Val Arg Ser Glu Asn Ser Ser Arg Gly Ala Met Arg Lys Arg Thr Arg
145 150 155 160

Thr Lys Phe Thr Pro Glu Gln Lys Ile Lys Met Arg Ala Phe Ala Glu
165 170 175

Lys Ala Gly Trp Lys Ile Asn Gly Cys Asp Glu Lys Ser Val Arg Glu
180 185 190

Phe Cys Asn Glu Val Gly Ile Glu Arg Gly Val Leu Lys Val Trp Met
195 200 205

His Asn Asn Lys Tyr Ser Leu Leu Asn Gly Lys Ile Arg Glu Ile Glu
210 215 220

His Gly Leu Cys Leu Asn Thr His Ser Asn Asp Gly Asp Gly Ser Ser
225 230 235 240

<210> 87

<211> 1082

<212> ADN

5 <213> Glycine max

<220>

<223> la secuencia de polipéptido predicha G3643 es ortóloga de G47

<400> 87

```

ccacgcgtcc gcttcttctc tcagaataca caacacaaag tcaatataat tatagtatat    60
ccctatgagt aggagttcgg cgatgcatgg aattacaagc acaacaaca agttgaaggg    120
agttcggcgt cgaaaatggg gcaaatgggt gtcggagatt cgtgttccgg gcacgcaaga    180
gcgttctgag ttgggaacct acgccacgcc ggaggctgcc gcggtggctc acgacgttgc    240
cgtctactgt ctaagtaggc cttcttcggt ggacaaactt aacttccccg aaaccttgtc    300
ttcgtacagt gttcagctca gggacatgtc tccgaggtct gtgcagaagg tggcttccga    360
tgttggcatg gatgttgatg caagaaacat tgttgcgggc aaaacttcaa cggtgggggc    420
agaaactaat tgcgagagtg atgagaggac tagtactgcg tctgtgtgta atgttgttgg    480
agaaggtggt gctgatcatt cggatgtggt ttggtgggat gatgatggtg ggtcttgcca    540
tggaaagtgt ggagattcta cggaaagga tgccttgagc atttccattg aagattatct    600
ttagctgttc taggtttcaa ctttagttat cttttttttt ttttttttga gttatgatcg    660
atgcgcatag ttagtagtta cagataatta ctgctagtgt tgggtgttta aacggtcaga    720
gatgatgata tataaatttg atgtgcgcta gctgcctttt tgaagaaact aaaaaaagg    780
aaaaaagaa agattgtacc cccaaaatga acttgggtcaa tttcacactt tcacgccatt    840
tgcattttgt gcacgttgtt agttttcaaa ttgtaattc ccttgctaaa cccactcaa    900
cttgagtgcg ttcacctttt attaccacgt agtaactgag agttaacgaa aatatttctt    960
tataaaaatt atttaccatt tagtgtcttt cgtttacttg taattattca acttgtcaat   1020
aagatgaaag ttctaatttt agatataata aaattcagct agtattttgtg aattaccac   1080
tt

```

<210> 88
 <211> 179
 <212> PRT

5 <213> Glycine max

<220>

<223> polipéptido G3643 ortólogo de G47

<400> 88

Met Ser Arg Ser Ser Ala Met His Gly Ile Thr Ser Thr Asn Asn Lys


```

tggctggggt cctacgccac gccggaggcc gccgccgtcg cgcacgacac ggccgtctac   300
ttcctccgcg gaggcgcggg cgacggcggt ggcggcgccg cgacgctcaa cttcccggag   360
cgcgcggcgg ccacgtacgg cggcggcgcc gccgtggcgc gcctgtcgcc gcggtcctg   420
cagcgcgtgg cgtccgacgc cggcatggcc gccgacgcgc agctcgtggc ggcgcgggac   480
ggcgcgcccg cgcgccgcc ggcgacggcg tacgcgcgcc cggatcactg cgcgcggcgc   540
acgacggcgc ggcacgacga gctggcgcgc cgcgggatgt acggcgctca cgcgcatgcc   600
gccggcgcga acgccaggac gagcggcgag cggcagctcg tctgtgccga ggagattagc   660
gtggatgaca tggagatcct gatgtaa                                     687
    
```

<210> 90

<211> 228

<212> PRT

5 <213> Oryza sativa (grupo de variedad cultivada japónica)

<220>

<223> polipéptido G3644 ortólogo de G47

<400> 90

```

Met Pro Phe Val Arg Thr Leu Ala Phe Pro Thr Leu Ala Asn Gln Ile
 1                               5                               10                               15

Glu Gly Glu Ile Lys Ser Arg Thr Glu Arg Asp His Gly Glu Val Asp
 20                               25                               30

Leu Ser Ser Pro Glu Glu Ala Met Ser Arg Ala Glu Cys Gly Gly Gly
 35                               40                               45

Glu Glu Glu Glu Arg Cys Arg Tyr Arg Gly Val Arg Arg Arg Arg Trp
 50                               55                               60

Gly Lys Trp Val Ser Glu Ile Arg Val Pro Gly Thr Arg Glu Arg Leu
 65                               70                               75                               80

Trp Leu Gly Ser Tyr Ala Thr Pro Glu Ala Ala Ala Val Ala His Asp
 85                               90                               95

Thr Ala Val Tyr Phe Leu Arg Gly Gly Ala Gly Asp Gly Gly Gly Gly
 100                              105                              110

Gly Ala Thr Leu Asn Phe Pro Glu Arg Ala Ala Ala Thr Tyr Gly Gly
 115                              120                              125

Gly Ala Ala Val Ala Arg Leu Ser Pro Arg Ser Val Gln Arg Val Ala
 130                              135                              140
    
```

Ser Asp Ala Gly Met Ala Ala Asp Ala Gln Leu Val Ala Ala Arg Asp
145 150 155 160

Ala Ala Pro Ala Pro Ala Pro Ala Thr Ala Tyr Ala Arg Pro Asp His
165 170 175

Cys Ala Gly Ala Thr Thr Ala Arg His Asp Glu Leu Ala Arg Arg Gly
180 185 190

Met Tyr Gly Ala His Ala His Ala Ala Gly Ala Asn Ala Arg Thr Ser
195 200 205

Gly Glu Arg Gln Leu Val Cys Ala Glu Glu Ile Ser Val Asp Asp Met
210 215 220

Glu Ile Leu Met
225

<210> 91

<211> 656

<212> ADN

5 <213> Brassica rapa subsp. Pekinensis

<220>

<223> la secuencia de polipéptido predicha G3645 es ortóloga de G47

<400> 91

```

cccacgcgctc cgatggatta catcgacaac accgctcgaaa ctcaatcaaa gtacaaaggc      60
atccgctcgcc ggaatgggg gaaatgggta tcggagattc gagttccggg aactcggcag      120
cgtctctggt taggetcatt ctccacggcg gaaggcgag ccgtggcgca cgacgtggct      180
ttctactggt tacaccaacc aaactcgctc gaatctctca acttccctca cttgcttctc      240
ccttccattg tttccaagac ttcgcccagg tctatccagc aagctgcttc taatgccgga      300
atggccqttg acgccggaat cgttaacagc tgtgatcagc cgtcagggaa ctctgggaat      360
ggagatacaa cgacggcgta ttgtgagaat ggaggtgctt tgaatatatc agtgtatgat      420
tatttgagc gtcacgatca cgtttgaact tcatctctct gtttttctgt ttaaagatac      480
agctactcaa aagaagcagt gatggagcct ggtgtgtaag caagcaaac gttgtgaata      540
tatataccgg tatgtttcgc tgttgccca atgcaagaaa cttttgtagt acgaatatat      600
atattttatt ttgaatgcct tcaagaggat tataatgcga gtgaagcttt gtttca      656
    
```

10 <210> 92

<211> 144

<212> PRT

<213> Brassica rapa subsp. Pekinensis

<220>

15 <223> polipéptido G3645 ortólogo de G47

<400> 92

Met Asp Tyr Ile Asp Asn Thr Val Glu Thr Gln Ser Lys Tyr Lys Gly
 1 5 10 15

Ile Arg Arg Arg Lys Trp Gly Lys Trp Val Ser Glu Ile Arg Val Pro
 20 25 30

Gly Thr Arg Asp Arg Leu Trp Leu Gly Ser Phe Ser Thr Ala Glu Gly
 35 40 45

Ala Ala Val Ala His Asp Val Ala Phe Tyr Cys Leu His Gln Pro Asn
 50 55 60

Ser Leu Glu Ser Leu Asn Phe Pro His Leu Leu Pro Pro Ser Ile Val
 65 70 75 80

Ser Lys Thr Ser Pro Arg Ser Ile Gln Gln Ala Ala Ser Asn Ala Gly
 85 90 95

Met Ala Val Asp Ala Gly Ile Val Asn Ser Cys Asp His Ala Ser Gly
 100 105 110

Asn Ser Gly Asn Gly Asp Thr Thr Thr Ala Tyr Cys Glu Asn Gly Gly
 115 120 125

Ala Leu Asn Ile Ser Val Tyr Asp Tyr Leu Asp Gly His Asp His Val
 130 135 140

<210> 93
 <211> 649
 <212> ADN

5 <213> Brassica oleracea

<220>

<223> la secuencia de polipéptido predicha G3646 es ortóloga de G47

<400> 93

```

cagccgcata acatattact attccccctg gtcatatctt tgacttttat ttttatttgc      60
tctaacctaa aagtattata taagtattgc acaattcata caatcggaaat tgactttctt      120
ctcctccaat cgtattttta ttcaacagtt cctctcaaga tcatcaactc aaaaaatgga      180
tcttagagac ggcggagaaa cccatcaggc caagtacaaa ggcacccgtc gccggaatg      240
gggaaaatgg gtatcggaga ttagggttcc agcaactcgt gaacgactct ggtaggctc      300
tttctccacc gccgaaggag ctgcggtagc ccacgacgtc gctttttact gcttgcaccg      360
accatcttct ctgcacaacg aagcttttaa cttccctcac ttgctgcaac cttcccttgc      420

ctccaacaca tctcetaagt ccatacaaaa agctgcttcg gacgcaggca tgggctaga      480
cgcaggattc gccctaaaca acgacagcgc gagtgggtggc gtggaggaag gcaccgaacg      540
ggaaacgttg aacatctcgg tgtacgatta tctagacgac ggtcgcattt gatatatggg      600
tttatatcta cgagcacctt atattagtaa ttaatatagg atgtgaata      649
    
```

- <210> 94
- <211> 138
- <212> PRT
- <213> Brassica oleracea
- 5 <220>
- <223> polipéptido G3646 ortólogo de G47
- <400> 94

```

Met Asp Pro Arg Asp Gly Gly Glu Thr His Gln Ala Lys Tyr Lys Gly
 1          5          10

Ile Arg Arg Arg Lys Trp Gly Lys Trp Val Ser Glu Ile Arg Val Pro
          20          25          30

Ala Thr Arg Glu Arg Leu Trp Leu Gly Ser Phe Ser Thr Ala Glu Gly
          35          40          45

Ala Ala Val Ala His Asp Val Ala Phe Tyr Cys Leu His Arg Pro Ser
 50          55          60

Ser Leu Asp Asn Glu Ala Phe Asn Phe Pro His Leu Leu Gln Pro Ser
 65          70          75          80

Leu Ala Ser Asn Thr Ser Pro Lys Ser Ile Gln Lys Ala Ala Ser Asp
          85          90          95

Ala Gly Met Gly Val Asp Ala Gly Phe Ala Leu Asn Asn Asp Ser Ala
          100          105          110

Ser Gly Gly Val Glu Glu Gly Thr Glu Arg Glu Thr Leu Asn Ile Ser
          115          120          125

Val Tyr Asp Tyr Leu Asp Asp Gly Arg Ile
 130          135
    
```

- <210> 95
- 10 <211> 495
- <212> ADN
- <213> Zinnia elegans
- <220>
- <223> la secuencia de polipéptido predicha G3647 es ortóloga de G47
- 15 <400> 95

```

attcgcggcc gcgaatatga gtaccagctc agatgaaggt aacaactggt taagccaaaa    60
gacttacaaa ggcgtaggt gccgacgatg gggcaaatgg gtgtcagaga ttcgagttcc    120
aggaagtcca gaacggctct ggctaggcac gtactctacg cctgagggtg cagctgtggc    180
tcatgatgta gcctcgtaact gtttaaaagg gaatacgtct tttcataaac ttaatatcc    240
gtctatgtta cctccgacag cacggacaga cctatctcct aggtccatcc aasaggctgc    300
gtctgatgct ggtatggcca tagacgcacg gtttatcgcg tctagagata ccacaccgac    360
taatgaggcg ttgaacattt ctgtagatga ttatctttaa attttgaqaa ctaatatgtt    420
gtcaccaata ttgtaagtcg atctacattg gcaaacacaa tgtacgtggt tggtggcact    480
tccagattat tgttt                                                    495

```

- <210> 96
- <211> 127
- <212> PRT
- 5 <213> Zinnia elegans
- <220>
- <223> polipéptido G3647 ortólogo de G47
- <400> 96

```

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

```

- 10 <210> 97
- <211> 594
- <212> ADN
- <213> Oryza sativa (grupo de variedad cultivada japónica)
- <220>
- 15 <223> la secuencia de polipéptido predicha G3649 es ortóloga de G47
- <400> 97

```

atgggccggg tggcggcgag cggcggcggc gccggcggag gggagatgat gaggtaacagg      60
ggcgtgcggc ggcggcggtg ggggaagtgg gtgtcggaga tccgggtgcc cgggacgcgg      120
gagcgcctgt ggetcggctc ctacgccacc gccgaggccg ccgccgtcgc gcacgacgcc      180
gccgtctgcc tctccggct cggcggcggc cgcgcgccg ccgcaggcgg aggcggcggg      240
ctcaacttcc ccgcccgccg gctcgcggc ggcggcggcc cctcctcta cggcggcggc      300
ggcgggtccc tgtcccggc ctcctgcag cgcgtggcgt ccgacgccgg catggcccgc      360
gacgcgtcgc tcgtggacct gcgccggac caccgcccg ccgcccgccg cgcctcatcc      420
tccggcagcg gcgtggcggg agacgggtgca agaaagcaag ggacacgtgg cgaggttagc      480
gacacgtatt ggtgtaggaa tggagaggat gggagcagaa gccggagctc cgggagtgg      540
gagctcattg tttacgaggg cttaatgtga gatgacatgg aaattttgat gtaa      594
    
```

<210> 98

<211> 197

<212> PRT

5 <213> Oryza sativa (grupo de variedad cultivada japónica)

<220>

<223> polipéptido G3649 ortólogo de G47

<400> 98

```

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

Tyr Gly Gly Ala Gly Gly Leu Leu Ser Pro Arg Ser Val Gln Arg Val
 100 105 110
 Ala Ser Asp Ala Gly Met Ala Ala Asp Ala Gln Leu Val Asp Leu Arg
 115 120 125
 Arg Asp His Pro Pro Ala Ala Ala Ala Ala Ser Ser Ser Gly Ser Gly
 130 135 140
 Val Ala Gly Asp Gly Ala Arg Lys Gln Gly Thr Arg Gly Glu Val Ser
 145 150 155 160
 Asp Thr Tyr Trp Cys Arg Asn Gly Glu Asp Gly Ser Arg Ser Arg Ser
 165 170 175
 Ser Gly Ser Glu Glu Leu Ile Val Tyr Glu Gly Leu Ser Val Asp Asp
 180 185 190
 Met Glu Ile Leu Met
 195

- <210> 99
- <211> 702
- <212> ADN
- 5 <213> Oryza sativa (grupo de variedad cultivada japónica)
- <220>
- <223> la secuencia de polipéptido predicha G3651 es ortóloga de G47
- <400> 99

```

atggetgacc tcccatgcat atatatacgc gtagtacgta caettgcitt cccaacgctc    60
gcgaatcaaa tccgaggtga aattaagtca agaacggaga gagatcacgg tgaggttgar    120
ctcagctcgc cggaggaggc aatgagccgg gcgagtgccg gcggcggcga ggaggaggag    180
cgggtgcagg acagggggct gcggcggcgg cgggtgggga agtgggtgtc ggagatccgg    240
gtgcccggca cgcgggagcg gctgtggctg gggtcctacg ccacgcccga ggccgcccgc    300
gtcgcgcacg acacggccgt ctacttctc cgcggaggcg cgggcgacgg cggtgccggc    360
ggcgcgaccg ctcaacttc cggagcgcgc ggcggccacc gtacggcggc cgcctggcg    420
cgctgtcgc cgcggtccgt gcagcgcgtg gcgtccgacg cggcatggcc gccgacgcgc    480
agctcggcgt gccgggaccc ggcccggccc gcgcccggca cggcgtacgc gcgcccggat    540
cactgcgccg gcgcgacgac ggcgcggcac gacgagctgg cgcgccggcg gatgtacggc    600
gctcacgcgc atgcccggcg cgcgaacgcc aggacgagcg gcgagcggca gctcgtctgt    660
gccgaggaga ttgctgtgga tgacatggag atcctgatgt aa                            702
    
```

- 10 <210> 100
- <211> 233
- <212> PRT
- <213> Oryza sativa (grupo de variedad cultivada japónica)
- <220>
- 15 <223> polipéptido G3651 ortólogo de G47
- <400> 100

Met Ala Asp Leu Pro Cys Ile Tyr Ile Arg Val Val Arg Thr Leu Ala
 1 5 10 15

Phe Pro Thr Leu Ala Asn Gln Ile Glu Gly Glu Ile Lys Ser Arg Thr
 20 25 30

Glu Arg Asp His Gly Glu Val Asp Leu Ser Ser Pro Glu Glu Ala Met
 35 40 45

Ser Arg Ala Glu Cys Gly Gly Gly Glu Glu Glu Glu Arg Cys Arg Tyr
 50 55 60

Arg Gly Val Arg Arg Arg Arg Trp Gly Lys Trp Val Ser Glu Ile Arg
 65 70 75 80

Val Pro Gly Thr Arg Glu Arg Leu Trp Leu Gly Ser Tyr Ala Thr Pro
 85 90 95

Glu Ala Ala Ala Val Ala His Asp Thr Ala Val Tyr Phe Leu Arg Gly
 100 105 110

Gly Ala Gly Asp Gly Gly Gly Gly Gly Ala Thr Ala Gln Leu Pro Gly
 115 120 125

Ala Arg Gly Gly His Arg Thr Ala Ala Ala Val Ala Arg Leu Ser Pro
 130 135 140

Arg Ser Val Gln Arg Val Ala Ser Asp Ala Ala Trp Pro Pro Thr Arg
 145 150 155 160

Ser Ser Ala Cys Arg Asp Pro Ala Arg Pro Ala Pro Ala Thr Ala Tyr
 165 170 175

Ala Arg Pro Asp His Cys Ala Gly Ala Thr Thr Ala Arg His Asp Glu
 180 185 190

Leu Ala Arg Arg Gly Met Tyr Gly Ala His Ala His Ala Ala Gly Ala
 195 200 205

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

Ser Val Asp Asp Met Glu Ile Leu Met
 225 230

- <210> 223
- <211> 2217
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>

<221> misc_feature
 <222> (2192)..(2192)
 <223> n es a, c, g, o t
 <220>

5 <223> secuencia de referencia G175; la secuencia de polipéptido predicha es paróloga de G1274
 <400> 223

```

atttctaaaa atagttaaac tgttggatg ttttggtaa ctgattaatt aacacaattg    60
gagaaaacaa attgaaactt ttttgttctg tagagagtta aggagaaatt ttatctaacc    120
aagctgtgta aatctcttct agtttttctg ccgatataca ttttcattgt gttgagggta    180
aacgataaat aagaacgaga gagagagaga gcaagagcaa gagatttcta ctacagaaga    240
tttattatat tgatcatttt gtgtgatcaa cccataaaaa cagagagaca tagacaagtc    300
catgtttcga tgtttcgatc tctcttactg tctaaacggc gaaataaaaa gtctgatggg    360
tgtcacttat tgcatgtata ttagtaaact agcttgagcc caagttaaag ctgaaacttg    420
ggtttgcaat ggctggattt gataataaag ctgctgtaat gggagaaatg ttcgactgta    480
gtactactaa ccacaggaag agatcgaaag cggaaacttg tagagagttt tctttaaatt    540
acatcaagaa tgaggattct ttgcaaacca cctttcaaga aagttcacga ggagctcttc    600
gtgaaaggat tgctcgcaga tccgggttta atgcaccgtg gttaaacact gaggatattc    660
ttcagtcgaa atctttaacc atctcttctc ctggctctag tctcgaact ctgttagagt    720
ctcctgtttt cctctcaaac cctttgctat ctccaacaac cgggaagctc tcatcagtac    780
cttctgataa ggctaaagct gagttatttg acgacattac cacatcetta gccttccaaa    840
ccatttcagg aagtggcctt gatcctacta acatcgcttt agaaccgat gattcccaag    900
actatgaaga aagacagctc ggcggtttag gagactcgat ggcttgttgt gcacctgcag    960
atgatggata caactggaga aatatggac aaaagctagt taaaggaagt ggtatccgc    1020
ggagctatta caagtgcacg caccgaatt gtgaggccaa gaagaagggt gaacggctctc    1080
gggaaggtca tattatagag atcatataca caggagatca tatacacagc aaacctccac    1140
ctaaccgccg gtcagggatt ggatcatccg gtactggcca agacatgcaa atagatgcaa    1200
ccgaatcaga aggttttgct ggaaccaatg agaacataga atggacatca cctgtatctg    1260
    
```

```

cagagctcga atacgggagc cattcaggat caatgcaggt tcaaacggg actcatcagt 1320
tcgggtatgg tgatgcagca gctgatgcct tatatagaga tgaaacgaa gatgatcgca 1380
cgtcccacat gagtgtttcc ctgacttacg atggagaggt agaagagtcc gaatcaaga 1440
gaaggaaact agaagcttat gcaacagaaa cgagtggatc aaccagagcc agccgtgagc 1500
caagagttgt ggtgcagacc acaagtgaca ttgacatcct cgatgatggt tatcgctggc 1560
gcaagtatgg gcaaaaagtc gttaaaggaa acccgaatcc aaggagctac tataaatgca 1620
cagctaattg atgtaccgta acgaagcatg tagagagagc ctctgatgac ttcaagagcg 1680
tactaacaac ttatataggc aagcacacc accgtgtacc agcagcacgc aacagcagcc 1740
acgtcgggagc aggcagttca gggactctcc aaggcagttt agcgaactag acccacaacc 1800
acaatgtgca ctatccaatg ccacacagta gatctgaggg actggccaca gccaaactcat 1860
ctctatttga ctccagtc caccctgagc atcctacagg tttctccggt tacataggcc 1920
aatctgagct ttctgatctt tcaatgcctg gtctaactat tgggcaagag aagcttacca 1980
gcctgcaggg gcctgacatt ggggatccaa ctggccta atgtgcagta gcagcacagc 2040
cgaaggtgga accagtgta ccacaacag gacttgattt gtcagcagc tcattgatat 2100
gcagagagat gttgagtaga ttacgacaga tatgaaacaa atctctttgt tcaactgattg 2160
ctcaaatttt tgataaatg aaaaattqag anaaaaaaaa aaaaaaaaaa aaaaaaa 2217

```

<210> 224

<211> 568

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia de polipéptido G175; paróloga de G1274

<400> 224

```

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

```

Lys Ser Leu Thr Ile Ser Ser Pro Gly Leu Ser Pro Ala Thr Leu Leu
 85 90 95
 Glu Ser Pro Val Phe Leu Ser Asn Pro Leu Leu Ser Pro Thr Thr Gly
 100 105 110
 Lys Leu Ser Ser Val Pro Ser Asp Lys Ala Lys Ala Glu Leu Phe Asp
 115 120 125
 Asp Ile Thr Thr Ser Leu Ala Phe Gln Thr Ile Ser Gly Ser Gly Leu
 130 135 140
 Asp Pro Thr Asn Ile Ala Leu Glu Pro Asp Asp Ser Gln Asp Tyr Glu
 145 150 155 160
 Glu Arg Gln Leu Gly Gly Leu Gly Asp Ser Met Ala Cys Cys Ala Pro
 165 170 175
 Ala Asp Asp Gly Tyr Asn Trp Arg Lys Tyr Gly Gln Lys Leu Val Lys
 180 185 190
 Gly Ser Glu Tyr Pro Arg Ser Tyr Tyr Lys Cys Thr His Pro Asn Cys
 195 200 205
 Glu Ala Lys Lys Lys Val Glu Arg Ser Arg Glu Gly His Ile Ile Glu
 210 215 220
 Ile Ile Tyr Thr Gly Asp His Ile His Ser Lys Pro Pro Pro Asn Arg
 225 230 235 240
 Arg Ser Gly Ile Gly Ser Ser Gly Thr Gly Gln Asp Met Gln Ile Asp
 245 250 255
 Ala Thr Glu Tyr Glu Gly Phe Ala Gly Thr Asn Glu Asn Ile Glu Trp
 260 265 270
 Thr Ser Pro Val Ser Ala Glu Leu Glu Tyr Gly Ser His Ser Gly Ser
 275 280 285
 Met Gln Val Gln Asn Gly Thr His Gln Phe Gly Tyr Gly Asp Ala Ala
 290 295 300
 Ala Asp Ala Leu Tyr Arg Asp Glu Asn Glu Asp Asp Arg Thr Ser His
 305 310 315 320
 Met Ser Val Ser Leu Thr Tyr Asp Gly Glu Val Glu Glu Ser Glu Ser
 325 330 335

Lys Arg Arg Lys Leu Glu Ala Tyr Ala Thr Glu Thr Ser Gly Ser Thr
 340 345 350

Arg Ala Ser Arg Glu Pro Arg Val Val Val Gln Thr Thr Ser Asp Ile
 355 360 365

Asp Ile Leu Asp Asp Gly Tyr Arg Trp Arg Lys Tyr Gly Gln Lys Val
 370 375 380

Val Lys Gly Asn Pro Asn Pro Arg Ser Tyr Tyr Lys Cys Thr Ala Asn
 385 390 395 400

Gly Cys Thr Val Thr Lys His Val Glu Arg Ala Ser Asp Asp Phe Lys
 405 410 415

Ser Val Leu Thr Thr Tyr Ile Gly Lys His Thr His Val Val Pro Ala
 420 425 430

Ala Arg Asn Ser Ser His Val Gly Ala Gly Ser Ser Gly Thr Leu Gln
 435 440 445

Gly Ser Leu Ala Thr Gln Thr His Asn His Asn Val His Tyr Pro Met
 450 455 460

Pro His Ser Arg Ser Glu Gly Leu Ala Thr Ala Asn Ser Ser Leu Phe
 465 470 475 480

Asp Phe Gln Ser His Leu Arg His Pro Thr Gly Phe Ser Val Tyr Ile
 485 490 495

Gly Gln Ser Glu Leu Ser Asp Leu Ser Met Pro Gly Leu Thr Ile Gly
 500 505 510

Gln Glu Lys Leu Thr Ser Leu Gln Ala Pro Asp Ile Gly Asp Pro Thr
 515 520 525

Gly Leu Met Leu Gln Leu Ala Ala Gln Pro Lys Val Glu Pro Val Ser
 530 535 540

Pro Gln Gln Gly Leu Asp Leu Ser Ala Ser Ser Leu Ile Cys Arg Glu
 545 550 555 560

Met Leu Ser Arg Leu Arg Gln Ile
 565

- <210> 225
- <211> 1072
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> secuencia de referencia G303
- <400> 225

```

aaaagataac gtagcatta acttttatgc attaactcct tcatttcttt ttgtgctcgt      60
ttggttgaga gagagagagg ctatatatga tgttccaaca agattaccct catggctttt      120
cactcgtgga aacatcctta agttacgaaa tgttggatta ctttcaaac atcgtcgttt      180
cgaactctga agacgtggcg tcacagcaaa attccatttc gtctcttctt tattcatcag      240
cgacactctc ctgtccata acagagcaaa aatctcactt aactgaaaag ttatctcttc      300
tacgagaaag atatggttgc ggtgactttc tgtcgcggaa gaggagaag agaaagtgaaa      360
aaacgattgt agataaagag .aatcaagga tgaatcacat tgccgtcgag cgtaacggga      420
gaaaacagat gaatcatttt ctgtctatcc tcaagtctat gatgcctctc tcttattctc      480
aacctaataa ccaagcatca atcatagaag ggaccattag ctatctgaag aagctagaac      540
aacgtctcca atctctcgaa gcccaattaa aagctactaa actcaatcaa tcaccaataa      600
tattttccga cttcttcatg ttccctcaat actccaccgc cactgccact gccaccgcca      660
ctgcctcttc atcctctctg agccaccacc atcacaagcg actagagggt gttgctgacg      720
tggaggttac aatggtagaa agacatgcca acattaaagt gttaacgaag acacagccaa      780
gattgctctt caagattatc aatgagtta actctttagg ttttaagtact cttcatctca      840
aectcacaac ttccaaagac atgtctctct tcacttttag cgtcaaggta gaggcagatt      900
gtcaattgac gccttctggt aatgaggtcg caaatcggg gcatgaagtc gttagaagag      960
ttcacaagga acgttgaatt ttgtttacat actagctaac ttgaaattc tattttattg     1020
tataaacaat ctcttatgtg tgaatttac atatatacaa ttaattaaca tt             1072

```

<210> 226

<211> 296

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G303

<400> 226

```

Met Met Phe Gln Gln Asp Tyr Pro His Gly Phe Ser Leu Val Glu Thr
1          5          10          15
Ser Leu Ser Tyr Glu Met Leu Asp Tyr Phe Gln Asn Ile Val Val Ser
20          25          30

```

Asn Ser Glu Asp Val Ala Ser Gln Gln Asn Ser Ile Ser Ser Ser Ser
 35 40 45
 Tyr Ser Ser Ala Thr Leu Ser Cys Ser Ile Thr Glu Gln Lys Ser His
 50 55 60
 Leu Thr Glu Lys Leu Ser Pro Leu Arg Glu Arg Tyr Gly Cys Gly Asp
 65 70 75 80
 Phe Leu Ser Arg Lys Arg Arg Arg Arg Ser Glu Lys Thr Ile Val Asp
 85 90 95
 Lys Glu Asn Gln Arg Met Asn His Ile Ala Val Glu Arg Asn Arg Arg
 100 105 110
 Lys Gln Met Asn His Phe Leu Ser Ile Leu Lys Ser Met Met Pro Leu
 115 120 125
 Ser Tyr Ser Gln Pro Asn Asp Gln Ala Ser Ile Ile Glu Gly Thr Ile
 130 135 140
 Ser Tyr Leu Lys Lys Leu Glu Gln Arg Leu Gln Ser Leu Glu Ala Gln
 145 150 155 160
 Leu Lys Ala Thr Lys Leu Asn Gln Ser Pro Asn Ile Phe Ser Asp Phe
 165 170 175
 Phe Met Phe Pro Gln Tyr Ser Thr Ala Thr Ala Thr Ala Thr Ala Thr
 180 185 190
 Ala Ser Ser Ser Ser Ser Ser His His His His Lys Arg Leu Glu Val
 195 200 205
 Val Ala Asp Val Glu Val Thr Met Val Glu Arg His Ala Asn Ile Lys
 210 215 220
 Val Leu Thr Lys Thr Gln Pro Arg Leu Leu Phe Lys Ile Ile Asn Glu
 225 230 235 240
 Phe Asn Ser Leu Gly Leu Ser Thr Leu His Leu Asn Leu Thr Thr Ser
 245 250 255
 Lys Asp Met Ser Leu Phe Thr Phe Ser Val Lys Val Glu Ala Asp Cys
 260 265 270
 Gln Leu Thr Pro Ser Gly Asn Glu Val Ala Asn Thr Val His Glu Val
 275 280 285
 Val Arg Arg Val His Lys Glu Arg
 290 295

- <210> 227
- <211> 628
- <212> ADN
- <213> Arabidopsis thaliana
- 5 <220>
- <223> secuencia de referencia G354
- <400> 227

```

cctagaagtc actaagtcga ttcaaaatgg ttgcgagaag tgaggaaatt gtgatagtgg      60
aagaagatac gactgcgaaa tgtttgatgt tgttatcaag agtcggagaa tgcggcggcg      120
gctgcggggg agatgaacgt gttttccgat gcaagacttg tcttaaagag ttctcatcgt      180
ttcaagcttt gggagggtcat cgtgcaagcc acaagaaact tatcaacagt gacaatccat      240
cacttcttgg atccttgtec aacaagaaaa ctaaaacgtc tcatccttgt ccgatatgtg      300
gagtgaagtt tccgatggga caagctcttg gtggtcacat gaggagacat aggaacgaga      360
aagtctcagg ctcggtggtt acacgttctt ttctaccgga gacgacgacg gtgacggctt      420
tgaagaaatt tagtagtggg aagagagtgg cttgtttggg tttggactta gattcgatgg      480
agagtttggg caattggaag ttggagttgg gaagaacgat ttcttggagt taagtttttg      540
ggttgtatac agtttcacat gattttgtaa tctttgttga tccaattatc gtaccgatcg      600
atgtgaatat tattttgata caataaaa                                         628
    
```

- <210> 228
- 10 <211> 168
- <212> PRT
- <213> Arabidopsis thaliana
- <220>
- <223> secuencia de referencia del polipéptido G354
- 15 <400> 228

```

Met Val Ala Arg Ser Glu Glu Ile Val Ile Val Glu Glu Asp Thr Thr
 1          5          10
Ala Lys Cys Leu Met Leu Leu Ser Arg Val Gly Glu Cys Gly Gly Gly
 20          25          30
Cys Gly Gly Asp Glu Arg Val Phe Arg Cys Lys Thr Cys Leu Lys Glu
 35          40          45
Phe Ser Ser Phe Gln Ala Leu Gly Gly His Arg Ala Ser His Lys Lys
 50          55          60
    
```

Leu Ile Asn Ser Asp Asn Pro Ser Leu Leu Gly Ser Leu Ser Asn Lys
 65 70 75 80

Lys Thr Lys Thr Ser His Pro Cys Pro Ile Cys Gly Val Lys Phe Pro
 85 90

Met Gly Gln Ala Leu Gly Gly His Met Arg Arg His Arg Asn Glu Lys
 100 105 110

Val Ser Gly Ser Leu Val Thr Arg Ser Phe Leu Pro Glu Thr Thr Thr
 115 120 125

Val Thr Ala Leu Lys Lys Phe Ser Ser Gly Lys Arg Val Ala Cys Leu
 130 135 140

Asp Leu Asp Leu Asp Ser Met Glu Ser Leu Val Asn Trp Lys Leu Glu
 145 150 155 160

Leu Gly Arg Thr Ile Ser Trp Ser
 165

<210> 229
 <211> 869
 <212> ADN

5 <213> Arabidopsis thaliana

<220>
 <223> secuencia de referencia G489
 <400> 229

tggtagtgaa tggatcaaca agaccatgga cagtctggag ctatgaacta tggcacaac 60

ccataccaaa ccaaccgat gagcaccact gctgctactg tagcaggagg tgcggcaca 120

ccaggccagc tggcgttcca ccagatccat cagcagcagc agcagcaaca gctggccag 180

cagcttcaag cattttggga gaaccaatc aaagagattg agaagactac cgatttcaag 240

aaccacagcc ttccccttgc gagaatcaag aaaatcatga aagcggatga agatgtccgt 300

atgatctcgg ctgaggcgcc ggtcgtgttt gcaagggcct gtgagatggt catcctggag 360

ctgacactca ggtcgtgga ccacactgag gagaataaga ggcggacggt gcagaagaac 420

gatattgctg ctgctgtgac tagaaccgat atttttgatt tccttgtgga tattgttccc 480

cgggaggatc tccgagatga agtcttggga agtattccga ggggcactgt cccggaagct 540

gctgctgctg gttaccgta tggatacttg cctgcaggaa ctgctccaat aggaaatccg 600

ggaatggtta tgggtaatcc cggtggtgcg tatccaccta atccttatat gggtaacca 660

atgtggcaac aacaggcacc tgaccaacct gaccaggaaa attagcaaga aactgtgagt 720

cttccagctt cttttaggcc tacctttagt tcttggggtt ttgtttctgt tttogaataa 780

tggtaacctt tgtataactt atttcagtat cgtctcagtt tggtaactatg tcagttttgg 840

tattcttata ttattgaagt aaaatgtg 869

10 <210> 230
 <211> 220
 <212> PRT

<213> Arabidopsis thaliana
 <220>
 <223> secuencia de referencia del polipéptido G489
 <400> 230

```

Met Asn Tyr Gly Thr Asn Pro Tyr Gln Thr Asn Pro Met Ser Thr Thr
 1          5          10          15

Ala Ala Thr Val Ala Gly Gly Ala Ala Gln Pro Gly Gln Leu Ala Phe
 20          25          30

His Gln Ile His Gln Gln Gln Gln Gln Gln Leu Ala Gln Gln Leu
 35          40          45

Gln Ala Phe Trp Glu Asn Gln Phe Lys Glu Ile Glu Lys Thr Thr Asp
 50          55          60

Phe Lys Asn His Ser Leu Pro Leu Ala Arg Ile Lys Lys Ile Met Lys
 65          70          75          80

Ala Asp Glu Asp Val Arg Met Ile Ser Ala Glu Ala Pro Val Val Phe
 85          90          95

Ala Arg Ala Cys Glu Met Phe Ile Leu Glu Leu Thr Leu Arg Ser Trp
100          105          110

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

Ala Ala Ala Val Thr Arg Thr Asp Ile Phe Asp Phe Leu Val Asp Ile
130          135          140

Val Pro Arg Glu Asp Leu Arg Asp Glu Val Leu Gly Ser Ile Pro Arg
145          150          155          160

Gly Thr Val Pro Glu Ala Ala Ala Ala Gly Tyr Pro Tyr Gly Tyr Leu
165          170          175

Pro Ala Gly Thr Ala Pro Ile Gly Asn Pro Gly Met Val Met Gly Asn
180          185          190

Pro Gly Gly Ala Tyr Pro Pro Asn Pro Tyr Met Gly Gln Pro Met Trp
195          200          205

Gln Gln Gln Ala Pro Asp Gln Pro Asp Gln Glu Asn
210          215          220
    
```

5

<210> 231
 <211> 798
 <212> ADN
 10 <213> Arabidopsis thaliana
 <220>
 <223> secuencia de referencia G634
 <400> 231

```

atggagcaag gaggagggtg tggtggtaat gaagtgtgag aggaagcttc acctattagt    60
tcaagacctc ctgctaaca cttagaagag cttatgagat tctcagccgc cgcggatgac    120
ggtggattag gaggtggagg tggaggagga ggaggaggaa gtgcttcttc ttcacatcggg    180
aatcgatggc cgagagaaga aacttttagct cttcttcgga tccgatccga tatggattct    240
acttttcgtg atgctactct caaagctcct ctttgggaac atgtttccag gaagctattg    300
gagttagggt acaaacgaag ttcaaagaaa tgcaaagaga aattcgaaaa cgttcagaaa    360
tattacaaac gtactaaaga aactcgcggt ggtcgcatg atggtaaagc ttacaagttc    420
ttctctcagc ttgaagctct caacactact cctcctcctc ctccttctca tcttcacgct    480
catcaaccag aacagaaaca acaacaaca ccacaacaag agatggtcac gagctcggaa    540
caatcatcat taccatcatc atcaagatgg ccaaaggcag agattctagc gcttataaac    600
ctgagaagtg gaatggaacc aaggtacca gataatgtac ctaaaggact tctatgggaa    660
gagatctcaa cttcaatgaa gagaatggga tacaacagaa acgctaagag atgtaaagag    720
aatgggaaa acataaaca atactacaag aaagttaaag aaagcaaca cagcaactac    780
aacaacaaga atcaatga                                         798
    
```

<210> 232

<211> 265

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G634

<400> 232

```

Met Glu Gln Gly Gly Gly Gly Gly Gly Asn Glu Val Val Glu Glu Ala
 1                               5 10 15
Ser Pro Ile Ser Ser Arg Pro Pro Ala Asn Asn Leu Glu Glu Leu Met
 20 25 30
    
```

Arg Phe Ser Ala Ala Ala Asp Asp Gly Gly Leu Gly Gly Gly Gly Gly
 35 40 45

Gly Gly Gly Gly Gly Ser Ala Ser Ser Ser Ser Gly Asn Arg Trp Pro
 50 55 60

Arg Glu Glu Thr Leu Ala Leu Leu Arg Ile Arg Ser Asp Met Asp Ser
 65 70 75 80

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

Arg Lys Leu Leu Glu Leu Gly Tyr Lys Arg Ser Ser Lys Lys Cys Lys
 100 105 110

Glu Lys Phe Glu Asn Val Gln Lys Tyr Tyr Lys Arg Thr Lys Glu Thr
 115 120 125

Arg Gly Gly Arg His Asp Gly Lys Ala Tyr Lys Phe Phe Ser Gln Leu
 130 135 140

Glu Ala Leu Asn Thr Thr Pro Pro Pro Pro Pro Ser His Pro His Ala
 145 150 155 160

His Gln Pro Glu Gln Lys Gln Gln Gln Gln Pro Gln Gln Glu Met Val
 165 170 175

Met Ser Ser Glu Gln Ser Ser Leu Pro Ser Ser Ser Arg Trp Pro Lys
 180 185 190

Ala Glu Ile Leu Ala Leu Ile Asn Leu Arg Ser Gly Met Glu Pro Arg
 195 200 205

Tyr Gln Asp Asn Val Pro Lys Gly Leu Leu Trp Glu Glu Ile Ser Thr
 210 215 220

Ser Met Lys Arg Met Gly Tyr Asn Arg Asn Ala Lys Arg Cys Lys Glu
 225 230 235 240

Lys Trp Glu Asn Ile Asn Lys Tyr Tyr Lys Lys Val Lys Glu Ser Asn
 245 250 255

Asn Ser Asn Tyr Asn Asn Lys Asn Gln
 260 265

- <210> 233
- <211> 341
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> secuencia de referencia G682
- <400> 233

```

acatggataa ccctcgcagg actaagcaac ccaagaccaa ctccatcgtt acttcttctt      60
ctgaagggaac agaagtgagt agtcttgagt gggaaqttgt gaacatgagt caagaagaag      120
aagatttggg ctctcgaatg cataagcttg tcggtgacag gtgggaactg atagctggga      180
ggatcccagg aagaaccgct ggagaaattg agaggttttg ggtcatgaaa aattgaaaag      240
tgatcaatat atctctatag tttcactgat tgtttcgatt atattccggt gtattttaaa      300
tttaccaact ctgttataag aagaaaactt tcttaaatcc c                               341

```

<210> 234

<211> 77

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G682

<400> 234

```

Met Asp Asn His Arg Arg Thr Lys Gln Pro Lys Thr Asn Ser Ile Val
 1                               5                               10                               15

Thr Ser Ser Ser Glu Gly Thr Glu Val Ser Ser Leu Glu Trp Glu Val
                20                               25                               30

Val Asn Met Ser Gln Glu Glu Glu Asp Leu Val Ser Arg Met His Lys
                35                               40                               45

Leu Val Gly Asp Arg Trp Glu Leu Ile Ala Gly Arg Ile Pro Gly Arg
 50                               55                               60

Thr Ala Gly Glu Ile Glu Arg Phe Trp Val Met Lys Asn
65                               70                               75

```

10 <210> 235

<211> 1797

<212> ADN

<213> Arabidopsis thaliana

<220>

15 <223> secuencia de referencia G916

<400> 235

```

aaaaaaaaa acaatctaag atatctatca agatcttaag atatctacat ttatagagat      60
gtttcgtttt ccggtaatgc ttggaggagg tccacgtgag aatctgaagc catcagatga      120

```

```

gcagcatcaa cgtgcggtg tgaatgaggt tgactttttc cgatcagcgg agaagagaga 180
tagggtttca cgtgaagaac aaaacattat cgccgatgag actcataggg ttcattgtcaa 240
aagggagaat tcacgtgttg atgatcatga cgatcgttct actgatcaca tcaatattgg 300
acttaatctt ctcaactgca atacgggaag cgacgagtca atggtggatg atggattgtc 360
tgtggatatg gaagagaaac gtacaaagtg tgagaatgca caacttcgcg aagagctaaa 420
gaaggcgagt gaagataatc aaagactaaa gcaaatgcta agtcaaaca ccaacaactt 480
caattctctg cagatgcaac ttgttgctgt catgaggcaa caagaagatc atcatcacct 540
agctacgacc gagaacaatg acaatgtaaa gaaccgacat gaaatgcctg aatggttcc 600
aagacagttc atcgatttg gaccgattc tgacgaagtg tcgtccgagg agaggacgac 660
ggttcggtcg ggatctctc cctcgttct agagaaatct agctcagtc aaaacggaaa 720
gagagtgtt gtaagagaag aaagcccga aaccgaatcc aacggctgga gaaaccctaa 780
caaagttcct aaacaccatg catcatccag catttgcggt ggcaatggca gtgaaaatgc 840
aagtacgaag gtcattgagc aagcggccgc cgaagccacc atgcgtaaag cccgtgtctc 900
ggttcgtgct cgatccgaag ctcccatggt aagcgatgga tgtcaatgga gaaaatacgg 960
acaaaaaatg gcgaaaggaa acccgtgccc tcgagcttat taccgttgca caatggctgt 1020
tggatgtcct gttcgcaagc aagtgcacg ttgcgcggaa gatagaacca ttctcataac 1080
aacctacgaa ggaaaccata accatccatt acctcctgcg gctatgaaca tggcttcaac 1140
tacaacagca gccgcaagca tgcttctctc aggtccacc atgtcgaacc aagatggtt 1200
aatgaaccca acaaatctct tggctcgaac catattaccg tgttcctcaa gcatggctac 1260
tatctcagcc tctgcacat tcccaacct tacattagac ctacagagt cacccaacgg 1320
gaacaatcca accaataacc cgtgatgca attctctcaa cggctctggt tgggtggagt 1380
gaaccaatcg gttttgcctc atatgatggg tcaggctttg tactacaacc aacagtctaa 1440
gttttcgggt ttacatatgc cgtctcagcc gctaaacgct ggtgagagtg tttagcggc 1500
tactgccga atcgctcca atcccaact tgccgcggt ctactgag ccataacttc 1560
gattatcaac ggttcgaaca atcagcagaa tgggaacaac aataacagta atgttacaac 1620
gagcaacggt gacaataggc aataacattt ttttataagt tttagttagg gactttttat 1680
cggctgattt tgtttgttt tctttttatt acatttttt ttagttacgg cttttttttg 1740
ttttttttt tctttcccaa acaacaagta ttgagagcaa ttcccccccc cccccct 1797

```

<210> 236

<211> 528

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G916

<400> 236

Met Phe Arg Phe Pro Val Ser Leu Gly Gly Gly Pro Arg Glu Asn Leu
1 5 10 15

Lys Pro Ser Asp Glu Gln His Gln Arg Ala Val Val Asn Glu Val Asp
20 25 30

Phe Phe Arg Ser Ala Glu Lys Arg Asp Arg Val Ser Arg Glu Glu Gln
35 40 45

Asn Ile Ile Ala Asp Glu Thr His Arg Val His Val Lys Arg Glu Asn
50 55 60

Ser Arg Val Asp Asp His Asp Asp Arg Ser Thr Asp His Ile Asn Ile
65 70 75 80

Gly Leu Asn Leu Leu Thr Ala Asn Thr Gly Ser Asp Glu Ser Met Val
85 90 95

Asp Asp Gly Leu Ser Val Asp Met Glu Glu Lys Arg Thr Lys Cys Glu
100 105 110

Asn Ala Gln Leu Arg Glu Glu Leu Lys Lys Ala Ser Glu Asp Asn Gln
115 120 125

Arg Leu Lys Gln Met Leu Ser Gln Thr Thr Asn Asn Phe Asn Ser Leu
130 135 140

Gln Met Gln Leu Val Ala Val Met Arg Gln Gln Glu Asp His His His
145 150 155 160

Leu Ala Thr Thr Glu Asn Asn Asp Asn Val Lys Asn Arg His Glu Val
165 170 175

Pro Glu Met Val Pro Arg Gln Phe Ile Asp Leu Gly Pro His Ser Asp
180 185 190

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

Ser Leu Leu Glu Lys Ser Ser Ser Arg Gln Asn Gly Lys Arg Val Leu
210 215 220

Val Arg Glu Glu Ser Pro Glu Thr Glu Ser Asn Gly Trp Arg Asn Pro
225 230 235 240

Asn Lys Val Pro Lys His His Ala Ser Ser Ser Ile Cys Gly Gly Asn
 245 250 255
 Gly Ser Glu Asn Ala Ser Ser Lys Val Ile Glu Gln Ala Ala Ala Glu
 260 265 270
 Ala Thr Met Arg Lys Ala Arg Val Ser Val Arg Ala Arg Ser Glu Ala
 275 280 285
 Pro Met Leu Ser Asp Gly Cys Gln Trp Arg Lys Tyr Gly Gln Lys Met
 290 295 300
 Ala Lys Gly Asn Pro Cys Pro Arg Ala Tyr Tyr Arg Cys Thr Met Ala
 305 310 315 320
 Val Gly Cys Pro Val Arg Lys Gln Val Gln Arg Cys Ala Glu Asp Arg
 325 330 335
 Thr Ile Leu Ile Thr Thr Tyr Glu Gly Asn His Asn His Pro Leu Pro
 340 345 350
 Pro Ala Ala Met Asn Met Ala Ser Thr Thr Thr Ala Ala Ala Ser Met
 355 360 365
 Leu Leu Ser Gly Ser Thr Met Ser Asn Gln Asp Gly Leu Met Asn Pro
 370 375 380
 Thr Asn Leu Leu Ala Arg Thr Ile Leu Pro Cys Ser Ser Ser Met Ala
 385 390 395 400
 Thr Ile Ser Ala Ser Ala Pro Phe Pro Thr Ile Thr Leu Asp Leu Thr
 405 410 415
 Glu Ser Pro Asn Gly Asn Asn Pro Thr Asn Asn Pro Leu Met Gln Phe
 420 425 430
 Ser Gln Arg Ser Gly Leu Val Glu Leu Asn Gln Ser Val Leu Pro His
 435 440 445
 Met Met Gly Gln Ala Leu Tyr Tyr Asn Gln Gln Ser Lys Phe Ser Gly
 450 455 460
 Leu His Met Pro Ser Gln Pro Leu Asn Ala Gly Glu Ser Val Ser Ala
 465 470 475 480
 Ala Thr Ala Ala Ile Ala Ser Asn Pro Asn Phe Ala Ala Ala Leu Ala
 485 490 495

Ala Ala Ile Thr Ser Ile Ile Asn Gly Ser Asn Asn Gln Gln Asn Gly
 500 505 510

Asn Asn Asn Asn Ser Asn Val Thr Thr Ser Asn Val Asp Asn Arg Gln
 515 520 525

- <210> 237
- <211> 768
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> secuencia de referencia G975
- <400> 237

```

attactcacc atcaagttcc tactttctct ctgacaaaca tcacagagta agtaagaatg      60
gtacagacga agaagttcag aggtgtcagg caacgccatt ggggttcttg ggtcgctgag      120
attcgtcacc ctctcttgaa acggaggatt tggctagggg cgttcgagac cgcagaggag      180
gcagcaagag catacgacga ggccgccgtt ttaatgagcg gccgcaacgc caaaaccaac      240
tttccctca acaacaacaa caccggagaa acttccgagg gcaaaaaccga tatttcagct      300
tcgtccacaa tgtcatcctc aacatcatct tcatcgetct cttecatcct cagcgccaaa      360
ctgaggaat gctgcaagtc tccttcccca tccctcacct gcctccgtct tgacacagcc      420
agctcccata tcggcgtctg gcagaaacgg gccggttcaa agtctgactc cagctgggtc      480
atgacggtgg agctaggtcc cgcaagctcc tccaagaga ctactagtaa agcttcacaa      540
gacgctattc ttgctccgac cactgaagtt gaaattggtg gcagcagaga agaagtattg      600
gatgaggaag aaaaggttgc tttgcaaatg atagaggagc ttctcaatac aaactaaatc      660
ttatttgctt atatatatgt acctatttcc attgctgatt tacagccaaa ataatcaatt      720
ataccgtgta ttttatagat gttttatatt aaaaggttgc tagatata      768
    
```

- 10 <210> 238
- <211> 199
- <212> PRT
- <213> Arabidopsis thaliana
- <220>
- 15 <223> secuencia de referencia del polipéptido G975
- <400> 238

Met Val Gln Thr Lys Lys Phe Arg Gly Val Arg Gln Arg His Trp Gly
 1 5 10 15

Ser Trp Val Ala Glu Ile Arg His Pro Leu Leu Lys Arg Arg Ile Trp
 20 25 30

Leu Gly Thr Phe Glu Thr Ala Glu Glu Ala Ala Arg Ala Tyr Asp Glu
 35 40 45

Ala Ala Val Leu Met Ser Gly Arg Asn Ala Lys Thr Asn Phe Pro Leu
 50 55 60

Asn Asn Asn Asn Thr Gly Glu Thr Ser Glu Gly Lys Thr Asp Ile Ser
 65 70 75 80

Ala Ser Ser Thr Met Ser Ser Ser Thr Ser Ser Ser Ser Leu Ser Ser
 85 90 95

Ile Leu Ser Ala Lys Leu Arg Lys Cys Cys Lys Ser Pro Ser Pro Ser
 100 105 110

Leu Thr Cys Leu Arg Leu Asp Thr Ala Ser Ser His Ile Gly Val Trp
 115 120 125

Gln Lys Arg Ala Gly Ser Lys Ser Asp Ser Ser Trp Val Met Thr Val
 130 135 140

Glu Leu Gly Pro Ala Ser Ser Ser Gln Glu Thr Thr Ser Lys Ala Ser
 145 150 155 160

Gln Asp Ala Ile Leu Ala Pro Thr Thr Glu Val Glu Ile Gly Gly Ser
 165 170 175

Arg Glu Glu Val Leu Asp Glu Glu Glu Lys Val Ala Leu Gln Met Ile
 180 185 190

Glu Glu Leu Leu Asn Thr Asn
 195

- <210> 239
- <211> 1116
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> Secuencia de referencia G1069; funcionalmente relacionada, homóloga de G1073
- <400> 239

```

ttggaaccct agaggccttt caagcaaatc atcagggtaa caatttcttg atctttcttt      60
ttagcgaatt tccagttttt ggtcaatcat ggcaaaccct tgggtggacga accagagtgg      120
tttagcgggc atggtggacc attcggctct ctcaggccat caccaaaacc atcaccacca      180
aagtcttctt accaaaggag atcttggaat agccatgaat cagagccaag acaacgacca      240
agacgaagaa gatgatccta gagaaggagc cgttgagggtg gtcaaccgta gaccaagagg      300
    
```

```

tagaccacca ggatccaaaa acaaacccaa agctccaatc tttgtgacaa gagacagccc 360
caacgcactc cgtagccatg tcttggagat ctccgacggc 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 ccggggtcca gggtcaggtc gttggaggtg gcggtgtagg 660
ccactctta gccataggtt cggatcaggt gattgctgct actttctcta acgctactta 720
tgagagattg cccatggaag aagaggaaga cggtgccggc tcaagacaga ttcacggagg 780
cggtgactca ccgccagaa tcggtagtaa cctgcctgat ctatcagggg tggccgggcc 840
aggctacaat atgccccgc atctgattcc aaatggggct ggtcagctag ggcacgaacc 900
atatacatgg gtccacgcaa gaccacctta ctgactcagt gagccatttc tatatataat 960
ggtctatata aataaatata tagatgaata taagcaagca atttgaggtg gtctattaca 1020
aagcttttgc tctggttggg aaaataaata agtatcaaag ctttgtttgt tcttaatgga 1080
aatatagagc ttgggaaggt agaaagagac gacatt 1116

```

<210> 240

<211> 281

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> Secuencia de referencia G1069; funcionalmente relacionada, homóloga de G1073

<400> 240

```

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> 241

<211> 1371

<212> ADN

5 <213> Arabidopsis thaliana

<220>

<223> Secuencia de referencia G1452; funcionalmente relacionada, homóloga de G512

<400> 241

atttattaag catcaatgag agaacttcag agctggggtt gagttctgtc caataataca 60

taaccacggt atcatttttg tcctttacta tctcattaca ctcttctggt attogcccaa 120

ttctttacagt cactactctc tatagggctc gagcggccgc ccgggcaggt ttctatgcag 180

```

atggttcaca cttcccqctc cattgcccag attgggttcg gtgtaagtc gcaattagta 240
ctcactatag ggctcgagcg gccgcccggg caggtaaaag atcaaacaat gtctaaagaa 300
gctgagatgt cgatcgcggt gtcggctttg ttccttggtt ttagattctc tcctactgat 360
gttgaactta tctcgtacta tcttcgctcg aaaatcgatg gtgatgagaa ctctgttgc 420
gtgattgctg aggtcgagat ttacaagttc gagccgtggg acttgccaga ggaatcgaaa 480
ctgaaatcgg agaacgagtg gttttacttc tgcgcgaggg ggaggaagta cccgcacggg 540
tcacaagacc ggcgagccac acagctagga tattggaaag cgaccggtaa agagcggagt 600
gttaaatccg ggaaccaagt tgttgaacc aagagaacgc ttgtatttca taccggtcgg 660
gctcctcgtg gcgagagAAC ggagtggatt atgcatgaat actgcatcca tggagcccca 720
caggatgcat tagtgggtg cgggttaaga aaaaatgctg attttcgggc tagttcgacc 780
caaaaaattg aggatggtgt tgtgcaagac gatggctacg ttggccaaag aggtggtttg 840
gacaaggagg acaaatccta ctatgaatct gagcatcaga taccaaatgg tgacatcgca 900
gaatcatcaa atgttgttga ggatcaggcc gataccgatg atgattgta cggcagatt 960
ctgaaatgatg atataataaa gctcgacgaa gaagcgttga aagctagcca agcgtttcga 1020
ccaactaatc caactcatca agaaacaata tcaagcgagt catcgagtaa gaggtcaaaa 1080
tgtgttataa aaaaagaatc aacggaaaca atgaattggt acgctttggt caggatcaag 1140
aacgttgccg gaaccgactc cagctggaga ttcccgaacc cgttcaaaat caagaaagat 1200
gatagccaga gattgatgaa gaatgttctg gccactactg ttttcttggc tatcttattt 1260
tctttctttt ggactgtatt aatagctagg aactaaagct agttacgaca tacatattat 1320
ttatacataa ataaatafag tattttgtct atggcaaaaa aaaaaaaaaa a 1371

```

<210> 242

<211> 373

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G1452; funcionalmente relacionada, homóloga de G512

<400> 242

```

Met Gln Met Val His Thr Ser Arg Ser Ile Ala Gln Ile Gly Phe Gly
1                               5                10                15

Val Lys Ser Gln Leu Val Leu Thr Ile Gly Leu Glu Arg Pro Pro Gly
                20                25                30

Gln Val Lys Asp Gln Thr Met Ser Lys Glu Ala Glu Met Ser Ile Ala
                35                40                45

```

Val Ser Ala Leu Phe Pro Gly Phe Arg Phe Ser Pro Thr Asp Val Glu
50 55 60

Leu Ile Ser Tyr Tyr Leu Arg Arg Lys Ile Asp Gly Asp Glu Asn Ser
65 70 75 80

Val Ala Val Ile Ala Glu Val Glu Ile Tyr Lys Phe Glu Pro Trp Asp
85 90 95

Leu Pro Glu Glu Ser Lys Leu Lys Ser Glu Asn Glu Trp Phe Tyr Phe
100 105 110

Cys Ala Arg Gly Arg Lys Tyr Pro His Gly Ser Gln Ser Arg Arg Ala
115 120 125

Thr Gln Leu Gly Tyr Trp Lys Ala Thr Gly Lys Glu Arg Ser Val Lys
130 135 140

Ser Gly Asn Gln Val Val Gly Thr Lys Arg Thr Leu Val Phe His Ile
145 150 155 160

Gly Arg Ala Pro Arg Gly Glu Arg Thr Glu Trp Ile Met His Glu Tyr
165 170 175

Cys Ile His Gly Ala Pro Gln Asp Ala Leu Val Val Cys Arg Leu Arg
180 185 190

Lys Asn Ala Asp Phe Arg Ala Ser Ser Thr Gln Lys Ile Glu Asp Gly
195 200 205

Val Val Gln Asp Asp Gly Tyr Val Gly Gln Arg Gly Gly Leu Asp Lys
210 215 220

Glu Asp Lys Ser Tyr Tyr Glu Ser Glu His Gln Ile Pro Asn Gly Asp
225 230 235 240

Ile Ala Glu Ser Ser Asn Val Val Glu Asp Gln Ala Asp Thr Asp Asp
245 250 255

Asp Cys Tyr Ala Glu Ile Leu Asn Asp Asp Ile Ile Lys Leu Asp Glu
260 265 270

Glu Ala Leu Lys Ala Ser Gln Ala Phe Arg Pro Thr Asn Pro Thr His
275 280 285

Gln Glu Thr Ile Ser Ser Glu Ser Ser Ser Lys Arg Ser Lys Cys Gly

```

                290                295                300
Ile Lys Lys Glu Ser Thr Glu Thr Met Asn Cys Tyr Ala Leu Phe Arg
305                310                315                320

Ile Lys Asn Val Ala Gly Thr Asp Ser Ser Trp Arg Phe Pro Asn Pro
                325                330                335

Phe Lys Ile Lys Lys Asp Asp Ser Gln Arg Leu Met Lys Asn Val Leu
                340                345                350

Ala Thr Thr Val Phe Leu Ala Ile Leu Phe Ser Phe Phe Trp Thr Val
                355                360                365

Leu Ile Ala Arg Asn
                370

```

- <210> 243
- <211> 732
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> Secuencia de referencia G1820
- <400> 243

```

ctcacttcca acatccaaat ccctagaat tgtaaatggc tgagaacaac aacaacaacg      60
gcgacaacat gaacaacgac aaccaccagc aaccaccgtc gtactcgcag ctgccgccga     120
tggcatcadc caaccctcag ttacgtaatt actggattga gcagatgga accgtctcgg     180
atttcaaaaa ccgtcagctt ccattggctc gaattaagaa gatcatgaag gctgatccag     240
atgtgcacat ggtctccgca gaggctccga tcatcttcgc aaaggettgc gaaatgttca     300
tcgttgatct cacgatcggg tcgtggctca aagccgagga gaacaaacgc cacacgcttc     360
agaaatcgga tatctccaac gcagtggcta gctctttcac ctacgatttc cttcttgatg     420
ttgtccctaa ggacgagtct atcgccaccg ctgatcctgg ctttgtggct atgccacatc     480
ctgacggtgg aggagtaccg caatattatt atccaccggg agtgggtgat ggaactccta     540
tggttgtag tggaatgtac gcgccatcgc aggcctggcc agcagcggct ggtgacgggg     600
aggatgatgc tgaggataat ggaggaaacg gcggcggaaa ttgaagtga gatttagggg     660
ttgtaaccgc ctatgtggga aatttgaaat ttggtggtgt ttattagggg tottcaattc     720
gtcggatttg ct                                                                732

```

- 10 <210> 244
- <211> 202
- <212> PRT
- <213> Arabidopsis thaliana
- <220>
- 15 <223> secuencia de referencia del polipéptido G1820
- <400> 244

```

Met Ala Glu Asn Asn Asn Asn Asn Gly Asp Asn Met Asn Asn Asp Asn
1          5          10          15

His Gln Gln Pro Pro Ser Tyr Ser Gln Leu Pro Pro Met Ala Ser Ser
          20          25          30

Asn Pro Gln Leu Arg Asn Tyr Trp Ile Glu Gln Met Glu Thr Val Ser
          35          40          45

Asp Phe Lys Asn Arg Gln Leu Pro Leu Ala Arg Ile Lys Lys Ile Met
50          55          60

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

Phe Ala Lys Ala Cys Glu Met Phe Ile Val Asp Leu Thr Met Arg Ser
          85          90          95

Trp Leu Lys Ala Glu Glu Asn Lys Arg His Thr Leu Gln Lys Ser Asp
          100          105          110

Ile Ser Asn Ala Val Ala Ser Ser Phe Thr Tyr Asp Phe Leu Leu Asp
          115          120          125

Val Val Pro Lys Asp Glu Ser Ile Ala Thr Ala Asp Pro Gly Phe Val
          130          135          140

Ala Met Pro His Pro Asp Gly Gly Gly Val Pro Gln Tyr Tyr Tyr Pro
145          150          155          160

Pro Gly Val Val Met Gly Thr Pro Met Val Gly Ser Gly Met Tyr Ala
          165          170          175

Pro Ser Gln Ala Trp Pro Ala Ala Ala Gly Asp Gly Glu Asp Asp Ala
          180          185          190

Glu Asp Asn Gly Gly Asn Gly Gly Gly Asn
          195          200

```

<210> 245

<211> 866

<212> ADN

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia G270; la secuencia de polipéptido predicha es paróloga de G1634

<400> 245

```

gtgttttag ttgaaactta ttcttcctt tttttgttt taggtatgga gactctgcat    60
ccattctctc acctacctat ctctgaccac cggttcgttg ttcaagagat ggtgagctta    120
cacagctcga gtagcggtag ctggactaaa gaagagaaca agatgttcga acgagctctt    180
gcgatatacg ctgaaagactc gcctgatcgc tggtttaaag ttgcttccat gatccctgga    240
aagactgttt ttgatgttat gaagcaatat agtaagcttg aagaagacgt ttcgatatt    300
gaagcaggac gtgttcccat tcctggttat cctgcagctt cttctccctt ggggtttgac    360
acggacatgt gtcgtaaacy gcctagtggg gctagaggat ctgatcaaga tcgaaagaaa    420
ggagtccctt ggacagagga agaacacagg agattcttgt taggccttct caagtacggt    480
aaagagatt ggagaaacat atcgagaaac ttcgtggtgt caaagacycc aacgcaagtg    540
gctgagccacg cccaaaagta ttaccagaga cagctctccg gagccaagga caaacgcagg    600
ccaagtatcc atgacatcac aaccggcaat cttctcaatg ccaatctcaa ccgttcttt    660
tccgatcata gagatattct ccctgattta gggtttatcg ataaggatga tacggaggag    720
ggagtaatat ttatgggtca gaatctctct tcagaaaatc tgtttctctc atcaccaact    780
tcattcgaag ctgccattaa cttcgcgga gaaaatgtct tcagtgcccg agcttaaggc    840
aacatagaat ccccaaactc agcggc                                     866
    
```

<210> 246

<211> 263

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G2701; paróloga de G1634

<400> 246

```

Met Glu Thr Leu His Pro Phe Ser His Leu Pro Ile Ser Asp His Arg
 1                               5                               10                               15

Phe Val Val Gln Glu Met Val Ser Leu His Ser Ser Ser Ser Gly Ser
                20                               25                               30

Trp Thr Lys Glu Glu Asn Lys Met Phe Glu Arg Ala Leu Ala Ile Tyr
                35                               40                               45

Ala Glu Asp Ser Pro Asp Arg Trp Phe Lys Val Ala Ser Met Ile Pro
 50                               55                               60
    
```

Gly Lys Thr Val Phe Asp Val Met Lys Gln Tyr Ser Lys Leu Glu Glu
65 70 75 80

Asp Val Phe Asp Ile Glu Ala Gly Arg Val Pro Ile Pro Gly Tyr Pro
85 90 95

Ala Ala Ser Ser Pro Leu Gly Phe Asp Thr Asp Met Cys Arg Lys Arg
100 105 110

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

Trp Thr Glu Glu Glu His Arg Arg Phe Leu Leu Gly Leu Leu Lys Tyr
130 135 140

Gly Lys Gly Asp Trp Arg Asn Ile Ser Arg Asn Phe Val Val Ser Lys
145 150 155 160

Thr Pro Thr Gln Val Ala Ser His Ala Gln Lys Tyr Tyr Gln Arg Gln
165 170 175

Leu Ser Gly Ala Lys Asp Lys Arg Arg Pro Ser Ile His Asp Ile Thr
180 185 190

Thr Gly Asn Leu Leu Asn Ala Asn Leu Asn Arg Ser Phe Ser Asp His
195 200 205

Arg Asp Ile Leu Pro Asp Leu Gly Phe Ile Asp Lys Asp Asp Thr Glu
210 215 220

Glu Gly Val Ile Phe Met Gly Gln Asn Leu Ser Ser Glu Asn Leu Phe
225 230 235 240

Ser Pro Ser Pro Thr Ser Phe Glu Ala Ala Ile Asn Phe Ala Gly Glu
245 250 255

Asn Val Phe Ser Ala Gly Ala
260

<210> 247
<211> 1040
<212> ADN

5 <213> Arabidopsis thaliana

<220>

<223> Secuencia de referencia G2789; la secuencia de polipéptido predicha es paróloga de G596

<400> 247

ctttagggac accaaatcta ttcaacctaa aagccttctt ttcccctata ttgaccaact 60

```

ttttagcgaa tcagaagagg aatggatgag gtatctcgtt ctcatacacc gcaatttcta 120
tcaagtgatc atcagcacta tcaccatcaa aacgctggac gacaaaaacg cggcagagaa 180
gaagaagagag ttgaacccaa caatataggg gaagacctag ccacctttcc ttccggagaa 240
gagaatatca agaagagaag gccacgtggc agacctgctg gttccaagaa caaacccaaa 300
gcaccaatca tagtcaactc gactccgcg aacgccttca gatgtcacgt catggagata 360
accaacgcct gcgatgtaat ggaagccta gccgtcttcg cttagacgcc tcagcgtggc 420
gtttgctctc tgaccgaaa cggggccgtt acaaacgtca ccgttagaca acctggcggg 480
ggcgtcgtca gtttacacgg acggtttgag attctttctc tctcgggttc gtttcttctt 540
ccaccggcac caccagctgc gtctggtta aaggtttact tagccggtgg tcaaggtaa 600
gtgatcggag gcagtgtggt gggaccgctt acggcatcaa gtcgggtggt cgttatggca 660
gcttcatttg gaaacgcac ttacgagagg ctgccactag aggaggagga gaaactgaa 720
agagaaatag atggaacgc ggctagggcg attggaacgc aaacgcagaa acagttaatg 780
caagatgcga catcgtttat tgggtcggcg tcgaatttaa ttaactctgt ttcggtgcca 840
ggtgaagctt attggggaac gcaacgaccg tctttctaag ataatatcat tgataatata 900
agtttctctc tcttattctt tttcactttt taccttttcc actttcttag gttttgtttt 960
aacgtttgat taatacctga aggtttttgg aaaattttcg atcggataaa aggatttatg 1020
ttgcgagccg aaacgcggcc 1040

```

<210> 248

<211> 265

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G2789; paróloga de G596

<400> 248

```

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
Cys Asp Val Met Glu Ser Leu Ala Val Phe Ala Arg Arg Arg Gln Arg
      100             105
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
Ser Gly Leu Lys Val Tyr Leu Ala Gly Gly Gln Gly Gln Val Ile Gly
      165             170
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> 249

<211> 495

<212> ADN

5 <213> Arabidopsis thaliana

<220>

<223> Secuencia de referencia G2839; la secuencia de polipéptido predicha es paróloga de G354

<400> 249

atggttgca gaagtgagga agttgagata gtggaagata cggcggcga atgtttgatg 60

```

ttgttatcaa gagttggaga atgCGGCGga ggaggagaga aacgagtttt cCGatgcaag 120
acttgtctta aagagttttc gtcGtttcaa gctttgggag gtcacgcgtgc aagccacaag 180
aaactcatta acagtagcga tccatcactt cttggatcct tgtctaacaa gaaaactaaa 240
acggcgacgt ctcaccttg tccgatatgt ggcgtggagt tccgatggg gcaagctctt 300
ggtggtcaca tgaggagaca taggagtgag aaagcctcac caggcacgtt ggttacacgt 360
tcttttttac cggagacgac gacggtgacg actttgaaaa aatcgagtag tgggaagaga 420
gtggcttgtt tggacttaga ttcgatggag agtttaqtca attggaagtt ggagttggga 480
agaacgattt cttga 495
    
```

<210> 250

<211> 164

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> secuencia de referencia del polipéptido G2839; paróloga de G354

<400> 250

```

Met Val Ala Arg Ser Glu Glu Val Glu Ile Val Glu Asp Thr Ala Ala
 1           5           10           15

Lys Cys Leu Met Leu Leu Ser Arg Val Gly Glu Cys Gly Gly Gly Gly
 20           25           30

Glu Lys Arg Val Phe Arg Cys Lys Thr Cys Leu Lys Glu Phe Ser Ser
 35           40           45

Phe Gln Ala Leu Gly Gly His Arg Ala Ser His Lys Lys Leu Ile Asn
 50           55           60

Ser Ser Asp Pro Ser Leu Leu Gly Ser Leu Ser Asn Lys Lys Thr Lys
 65           70           75           80

Thr Ala Thr Ser His Pro Cys Pro Ile Cys Gly Val Glu Phe Pro Met
 85           90           95

Gly Gln Ala Leu Gly Gly His Met Arg Arg His Arg Ser Glu Lys Ala
100           105           110

Ser Pro Gly Thr Leu Val Thr Arg Ser Phe Leu Pro Glu Thr Thr Thr
115           120           125

Val Thr Thr Leu Lys Lys Ser Ser Ser Gly Lys Arg Val Ala Cys Leu
130           135           140

Asp Leu Asp Ser Met Glu Ser Leu Val Asn Trp Lys Leu Glu Leu Gly
145           150           155           160

Arg Thr Ile Ser
    
```

10

<210> 251

<211> 1569
 <212> ADN
 <213> Arabidopsis thaliana
 <220>

5 <223> Secuencia de referencia G2854; la secuencia de polipéptido predicha es paróloga de G1940
 <400> 251

```

tctctctttt gcccttacgt cctcgagctc aaccctttaa ttgcatctac gacagtcgta      60
ccttttgcta ggttatcgtc gtcgatgatgc agcagccacc tccagcttcc aacggtgctg      120
caacagggcc agggcagatt ccttccgacc aacaagctta cctccagcag cagcagtcgt      180
ggatgatgca gcaccagcag caacaacaag gtcagccgcc tgcaggatgg aatcagcagt      240
ctgcacgcgc ttctggtaaa ccacagcagc agcagtatgg tggtggtgga tctcagaatc      300
caggatcagc tggtgagatc cggtcctcgt ggatcgggta cttgcagcca tggatggatg      360
agaactatct catgaacgtc tttggtctta ctggcgaggc tacagcagct aaagttatc      420
gcaataaaca gaacggatat tcagaagggt atggctttat tgagtttggt aacctgcta      480
cagctgagag gaatttacag acttacaatg gtgctccgat gccgagcagt ggcagggcct      540
tcaggttgaa ctgggctcag cttggagctg gagagagacg ccaggctgaa gggcctgagc      600
acacagtttt tgttgagagc ttggcacctg atgttaccga ccacatgctt actgaaacgt      660
ttaaagctgt gtattcctct gtcgaaggag ctaaagttgt gaatgatagg actactggac      720
ggtccaaggg ttatggattt gtcaggtttg cggatgaaag tgagcagatt cgtgccatga      780
ctgaaatgaa tgggtcaatac tgctcatcaa ggcctatgag tactggctct gctgccaaca      840
agaagcctct tacaatgcaa ccagcttcat atcagaacac tcaaggaaat tcaggagaaa      900
gtgatccaac taacacaaca atttttggtg gagctgtgga tcaaagtgta acagaagatg      960
atthgaagtc agtttttggg caatttggtg aactagtcca tgtgaaaata cccgcaggaa     1020
aacgttgccg atttgttcaa tacgccaata gggcatgtgc tgagcaagca ctttctgtgt     1080
tgaacggaac acaacttggg ggacaaagca ttcgtctttc atggggctgc agtcctcca     1140
acaacacagac tcaacctgat caagcccagt atggtggtgg tggaggatac tatgggtatc     1200
ctcctcaagg atatgaagca tacggatatg cacctcctcc tcaggacctt aacgcctact     1260
acggtggtta tgctgggggc ggctatggaa actaccagca gcctgggtgga taccagcagc     1320
aacagcagtg aagcatgatg tgtaagtcgt gatcgagctg aagttgtaga aagctgtaac     1380

cgtttgacaa tggatgaatga attattgaag tggatgtgta ttagctcttc ttgtctcaga     1440
ttctctcttc aactcttttt ttttctttt ggattattht tctttcgcgt ggctattggt     1500
gagtgctcct gcaagtcata acaaaaagct taaaccaag tgttttttgt catgactaat     1560
atgaccatt                                         1569
    
```

<210> 252
 10 <211> 415
 <212> PRT
 <213> Arabidopsis thaliana
 <220>

<223> secuencia de referencia del polipéptido G2854; paróloga de G1940
 <400> 252

```

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

Gly Gln Ile Pro Ser Asp Gln Gln Ala Tyr Leu Gln Gln Gln Gln Ser
 20          25          30

Trp Met Met Gln His Gln Gln Gln Gln Gly Gln Pro Pro Ala Gly
 35          40          45

Trp Asn Gln Gln Ser Ala Pro Ser Ser Gly Gln Pro Gln Gln Gln Gln
 50          55          60

Tyr Gly Gly Gly Gly Ser Gln Asn Pro Gly Ser Ala Gly Glu Ile Arg
 65          70          75          80

Ser Leu Trp Ile Gly Asp Leu Gln Pro Trp Met Asp Glu Asn Tyr Leu
 85          90          95

Met Asn Val Phe Gly Leu Thr Gly Glu Ala Thr Ala Ala Lys Val Ile
100          105          110

Arg Asn Lys Gln Asn Gly Tyr Ser Glu Gly Tyr Gly Phe Ile Glu Phe
115          120          125

Val Asn His Ala Thr Ala Glu Arg Asn Leu Gln Thr Tyr Asn Gly Ala
130          135          140

Pro Met Pro Ser Ser Glu Gln Ala Phe Arg Leu Asn Trp Ala Gln Leu
145          150          155          160

Gly Ala Gly Glu Arg Arg Gln Ala Glu Gly Pro Glu His Thr Val Phe
165          170          175

```

Val Gly Asp Leu Ala Pro Asp Val Thr Asp His Met Leu Thr Glu Thr
 180 185 190

Phe Lys Ala Val Tyr Ser Ser Val Lys Gly Ala Lys Val Val Asn Asp
 195 200 205

Arg Thr Thr Gly Arg Ser Lys Gly Tyr Gly Phe Val Arg Phe Ala Asp
 210 215 220

Glu Ser Glu Gln Ile Arg Ala Met Thr Glu Met Asn Gly Gln Tyr Cys
 225 230 235 240

Ser Ser Arg Pro Met Arg Thr Gly Pro Ala Ala Asn Lys Lys Pro Leu
 245 250 255

Thr Met Gln Pro Ala Ser Tyr Gln Asn Thr Gln Gly Asn Ser Gly Glu
 260 265 270

Ser Asp Pro Thr Asn Thr Thr Ile Phe Val Gly Ala Val Asp Gln Ser
 275 280 285

Val Thr Glu Asp Asp Leu Lys Ser Val Phe Gly Gln Phe Gly Glu Leu
 290 295 300

Val His Val Lys Ile Pro Ala Gly Lys Arg Cys Gly Phe Val Gln Tyr
 305 310 315 320

Ala Asn Arg Ala Cys Ala Glu Gln Ala Leu Ser Val Leu Asn Gly Thr
 325 330 335

Gln Leu Gly Gly Gln Ser Ile Arg Leu Ser Trp Gly Arg Ser Pro Ser
 340 345 350

Asn Lys Gln Thr Gln Pro Asp Gln Ala Gln Tyr Gly Gly Gly Gly Gly
 355 360 365

Tyr Tyr Gly Tyr Pro Pro Gln Gly Tyr Glu Ala Tyr Gly Tyr Ala Pro
 370 375 380

Pro Pro Gln Asp Pro Asn Ala Tyr Tyr Gly Gly Tyr Ala Gly Gly Gly
 385 390 395 400

Tyr Gly Asn Tyr Gln Gln Pro Gly Gly Tyr Gln Gln Gln Gln Gln
 405 410 415

- <210> 253
- <211> 914
- <212> ADN
- 5 <213> Arabidopsis thaliana
- <220>
- <223> secuencia de referencia G3083
- <400> 253

```

caaaatatga gccaaagagac agcगतगca agcttcaaga aattccaaca gtcатgगac 60
gagcagctac gaaaccacct gaaccacctc cgctcgcgcc aaaaccatca ccggaactct 120
gctaccggtg atgaggaacg gttaaggag gctgtggaca gagtगतgga acactttaga 180
gaatatcaca gggccaagtg ggctgcgacg gataaagacg tcatcgaagt tatggcttct 240
ccttgggctt cagctcttga acggctcctt cagtgggtcg gtggtggcg accaaccacc 300
ttgttcctac tggtttacac tgagtcgagt attttatttg agtctcgat cgtगतata 360
cttcgtgctt ttcgcaccgg tgatctcagt gatctctccc cttctcagtt caggacgगta 420
agtगतctac aatgtgagac cgtgaaggaa gagaatcga taucggaaga gttगतcgag 480
tggcaagacg atgcगतgga ccttगतctg gggacatcgt cagatcccga ccagaggatc 540
cgacggctag cagaaatcgt ccaccgaact gatगतctgc gactgagaac gatcacacgt 600
gtggtggagg tcttgagtcc gctccaacaa gcggagtttc ttgctgctgc gगतगतctt 660
cgtacaggcg ttगतggttg ggggactagc cacgaccgtc gtcgaagttc cgaagtttaa 720
agctaactaa aaccggtcta aaggtaatga gaatcगगga agaaaaatag tttttgगata 780
ttaattatat gtttttacca tgtttttggt tatccgact gagtttggtt ttacatggtt 840
atgaaaacaa aaaaaactca tatgttatgt ctctacggtt gcaataaaga tcaaccacg 900
tgtggtgctt gtta 914

```

<210> 254

<211> 237

<212> PRT

5 <213> Arabidopsis thaliana

<220>

<223> polipéptido G3083: secuencia de referencia

<400> 254

```

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

```

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

REIVINDICACIONES

1. Un procedimiento para producir una planta transgénica que tiene mayor tolerancia a la falta de agua que una planta de control, comprendiendo el procedimiento las etapas de:
- 5 (a) producir un vector de expresión que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de al menos 70% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12; en el que el polipéptido tiene una propiedad de la SEQ ID NO: 12 de aumentar la tolerancia a la falta de agua en una planta con respecto a la planta de control; y
- 10 (b) introducir el vector de expresión en una planta objetivo para producir una planta transgénica; en el que el polipéptido es expresado en exceso en la planta transgénica y dicho exceso de expresión da como resultado que la planta transgénica tiene una mayor tolerancia a la falta de agua que la planta de control.
2. El procedimiento de la reivindicación 1, en el que el dominio conservado tiene una identidad de secuencia de al menos 89% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12.
- 15 3. El procedimiento de la reivindicación 1 o 2, en el que el vector de expresión comprende un promotor constitutivo, inducible o específico de tejido operativamente unido a la secuencia de nucleótidos.
4. Una semilla transgénica que comprende un vector de expresión que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de al menos 70% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12; producido por una planta transgénica producida por cualquiera de los procedimientos de las reivindicaciones 1 a 3, en el que una planta descendiente que ha crecido a partir de la semilla transgénica tiene una mayor tolerancia a la sequía con respecto a la planta de control, y en el que dicha secuencia está operativamente unida a un promotor que es inducible en respuesta al calor o sequía.
- 25 5. Una semilla, fruta, hoja, raíz o descendiente de una planta que comprende un vector de expresión que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de al menos 70% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12; en el que el polipéptido tiene una propiedad de la SEQ ID NO:12 de aumentar la tolerancia a la falta de agua en una planta con respecto a la planta de control, que se puede obtener por el procedimiento de una cualquiera de las reivindicaciones 1 a 3, en el que dicha planta, semilla, fruta, hoja, raíz, célula de planta o descendiente tiene una mayor tolerancia a la falta de agua que la planta de control, y en el que dicha secuencia está operativamente unida a un promotor que es inducible en respuesta al estrés por calor o sequía.
- 35 6. Uso de un vector de expresión para producir una planta transgénica que tiene una mayor tolerancia a la falta de agua con respecto a una planta de control, en el que el vector de expresión codifica un polipéptido que comprende un dominio conservado con una identidad de secuencia de aminoácidos de al menos 70% con las coordenadas de aminoácidos 10-77 de la SEQ ID NO: 12.
- 40

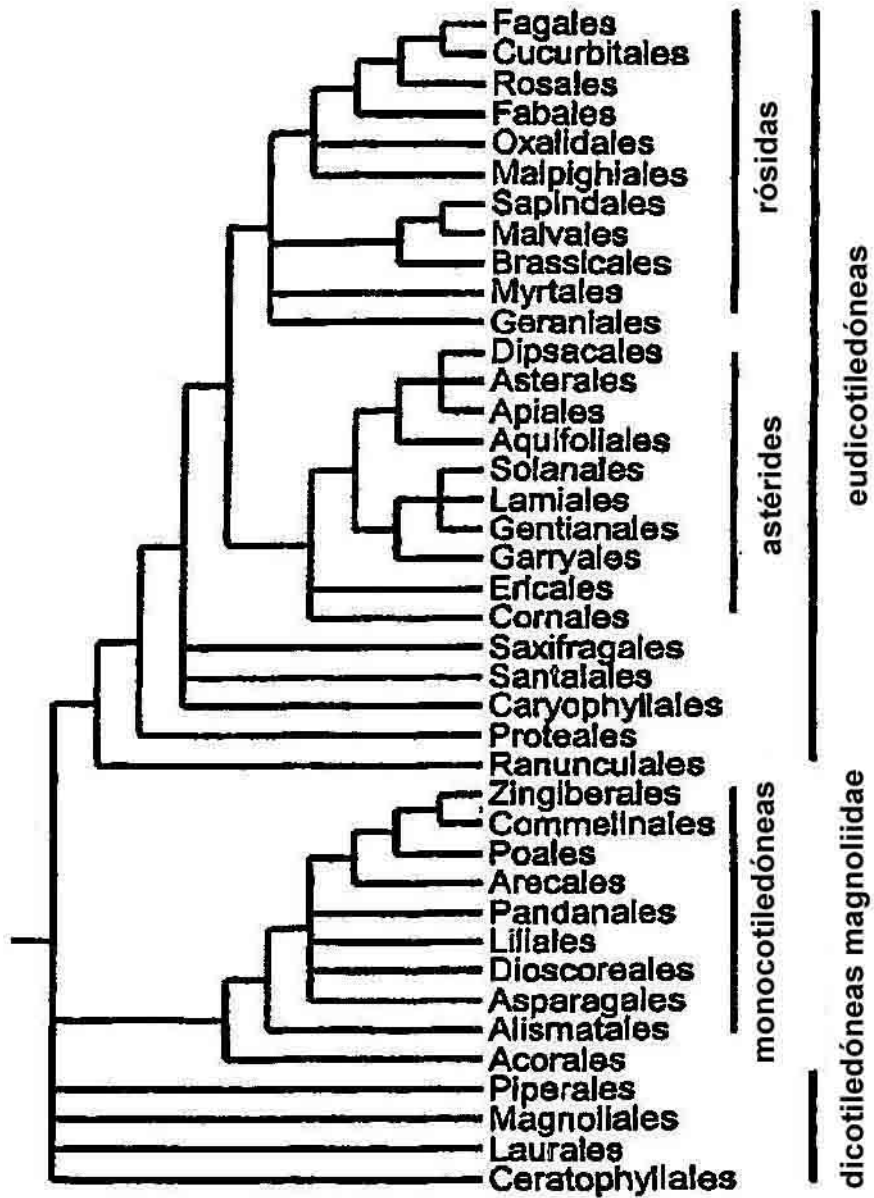


FIGURA 1

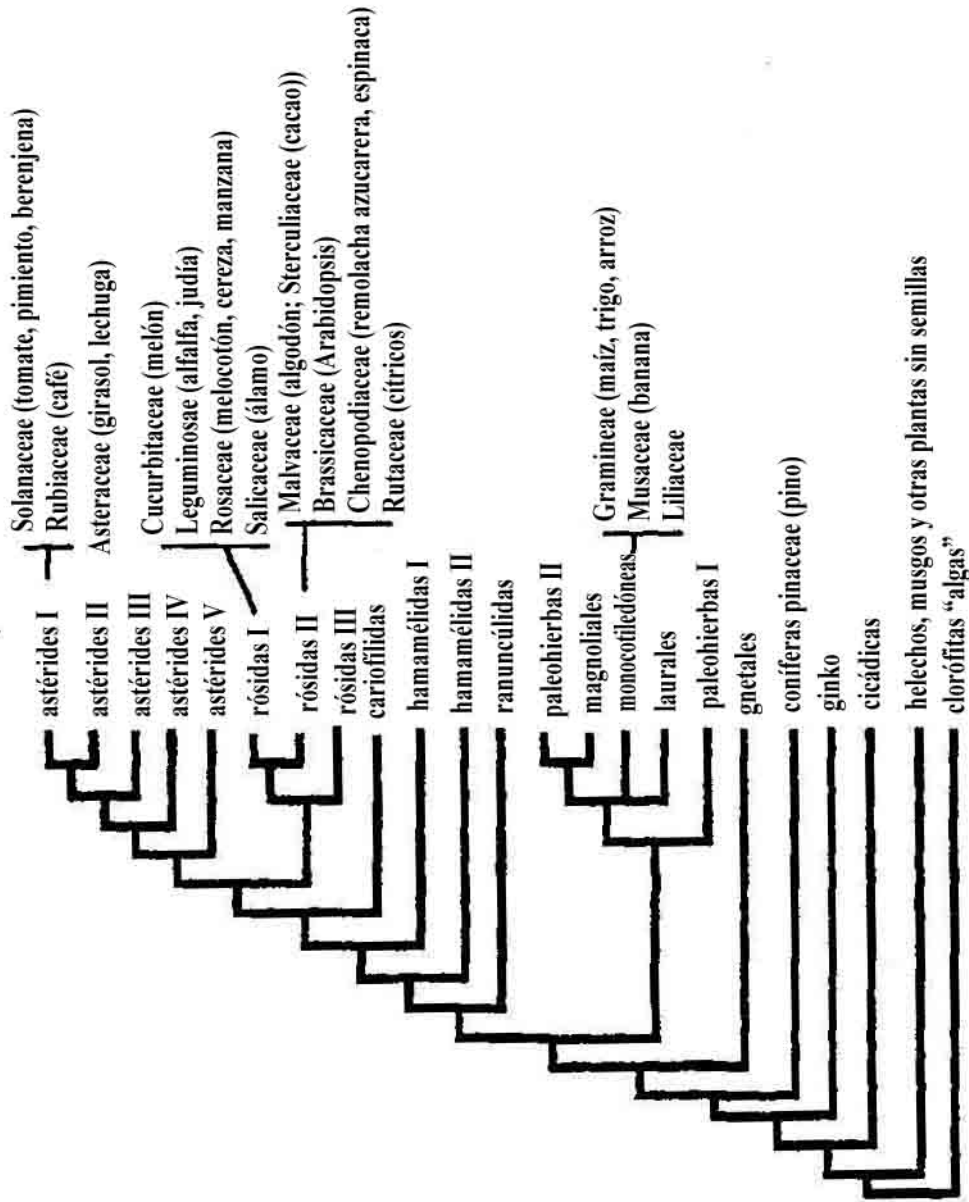


FIGURA 2

G2133	DQSKYKGI	RRRKWGWVSEIRV	PGTRQRLWLG	SFSTAEGAAVAHDVA	FYC	
G3646	HQAKYKGI	RRRKWGWVSEIRV	PATRERLWLG	SFSTAEGAAVAHDVA	FYC	
G3645	TQSKYKGI	RRRKWGWVSEIRV	PGTRDRRLWLG	SFSTAEGAAVAHDVA	FYC	
G47	SQSKYKGI	RRRKWGWVSEIRV	PGTRDRRLWLG	SFSTAEGAAVAHDVA	FFC	
G3643	TNNKLG	VRRRKWGWVSEIRV	PGTOERLWLG	TYATPEAAVAHDVA	NYC	
G3647	SQKTYK	GVRCRRWGWVSEIRV	PGSRERLWLG	TYSTPEGAAVAHDVA	SYC	
G3648	SNKFKG	VRRRKWGWVSEIRV	PGTOERLWLG	TYATPEAAVAHDVA	FYC	
G3644	ERCYRG	VRRRKWGWVSEIRV	PGTRERLWLG	SYATPEAAVAHDVA	VYF	
G3649	EMRYRG	VRRRKWGWVSEIRV	PGTRERLWLG	SYATAEA	AAVAHDVA	VCL
G3650	RRCRYG	VRRRKWGWVSEIRV	PGTRERLWLG	SYAAPEAAVAHDVA	ACL	
G3651	ERCYRG	VRRRKWGWVSEIRV	PGTRERLWLG	SYATPEAAVAHDVA	VYF	
G3654	TTTKYR	GVRLRKWGWVSEIRL	PNSRERLWLG	SYDTPEEAARA	FDAAEVFC	
G3652	KETRYK	GVRLRQWGWVSEIRL	PNSRKRILWLG	SYTYTPEKAARA	FDAAFIC	
G3653	VERKYR	GVRLRQWGWVSEIRL	PNSLKRILWLG	SYDSPEKAARA	FDAAFIC	
G3655	ERRKYK	GVRLRQWGWVSEIRL	PSSCERILWLG	SYDTPEKAARA	FDAAFIC	
G2576	MQSKYK	GVRRKRWGWVSEIRL	PNSRERILWLG	SYDTPEKAARA	FDAALYC	
G872	MQSKYK	GVRRKRWGWVSEIRL	PHSRERILWLG	SYDTPEKAARA	FDAAQFC	
G1277	REPFKG	IRMRKRWGWVSEIRE	PNKRSRLWLG	SYSTPEAAARA	YDTAVFY	
G12	RDKPYK	GIRMRKRWGWVSEIRE	PNKRSRIWLG	SYSTPEAAARA	YDTAVFY	
G24	DLKPYK	GIRMRKRWGWVSEIRE	PNKRSRIWLG	SYATPEAAARA	YDTAVFY	
G867	PSSKYK	GVVFPQNGRWGAQI	YEKHQRVWLG	TFNEE	DEAARAYDVAVHR --	
G40	RHPIYR	GVQRNSGKWVSEVRE	PNKTRIWLG	TFQTAEMA	ARAHDAALA	
G2294	RMKYYK	GVRRMSWGSWVSEIRA	PNQKTRIWLG	SYSTAEAAARA	YDAAALLC	
G2115	KIKKYYK	GVRRMSWGSWVSEIRA	PNQKTRIWLG	SYSTAEAAARA	YDVAALLC	

FIGURA 3

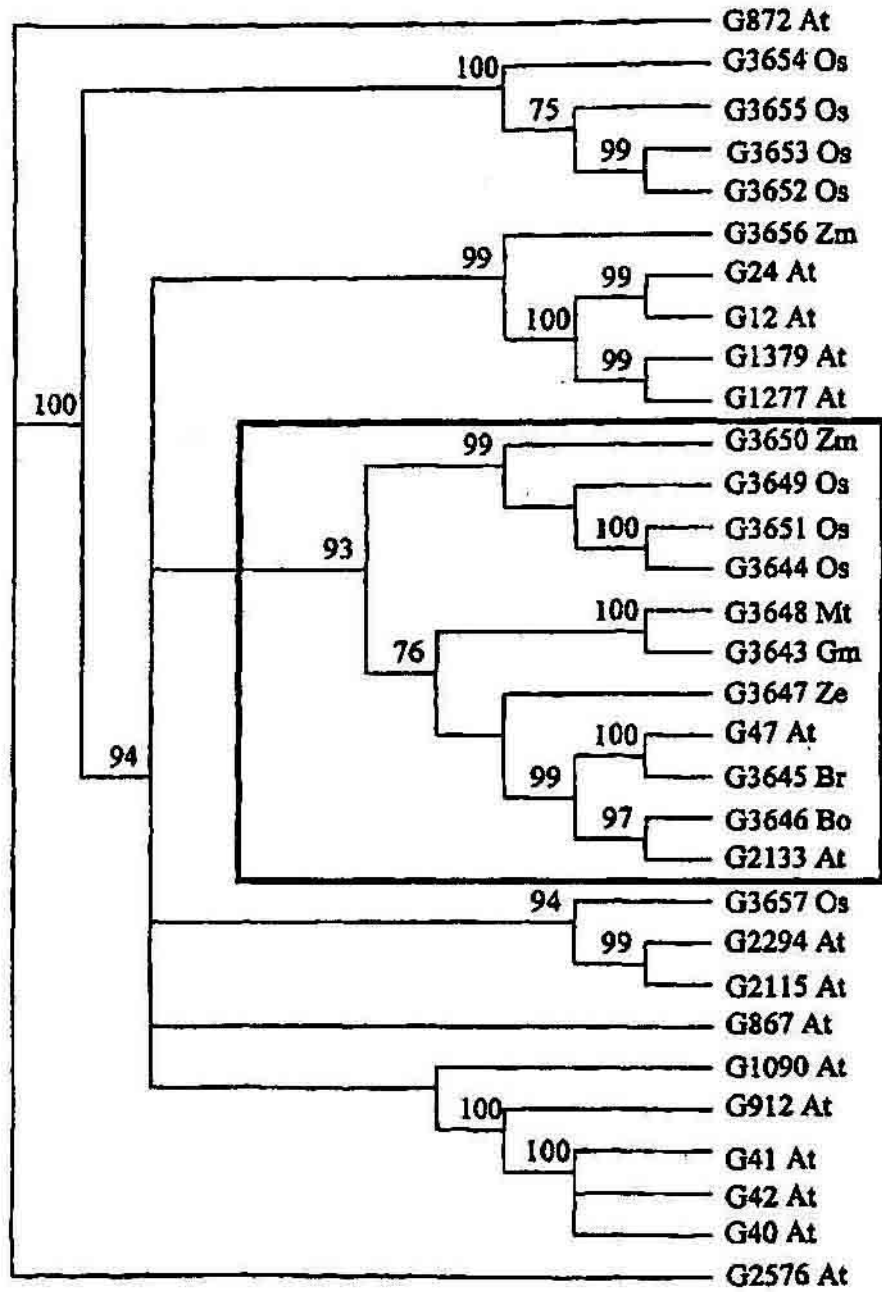


FIGURA 4



FIGURA 5A

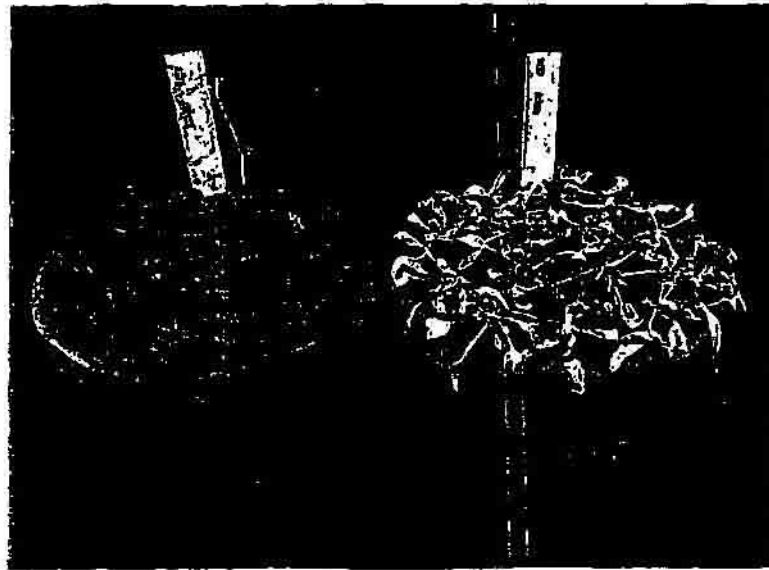


FIGURA 5B