



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



⑪ Número de publicación: **2 380 178**

⑯ Int. Cl.:
C12N 15/82 (2006.01)

⑫

TRADUCCIÓN DE PATENTE EUROPEA

T3

- ⑯ Número de solicitud europea: **02723399 .8**
⑯ Fecha de presentación: **08.03.2002**
⑯ Número de publicación de la solicitud: **1390381**
⑯ Fecha de publicación de la solicitud: **25.02.2004**

⑮

Título: **Reguladores del metabolismo de lípidos en plantas**

⑯ Prioridad:
08.03.2001 US 274170 P

⑯ Titular/es:
**MICHIGAN STATE UNIVERSITY
238 ADMINISTRATION BUILDING
EAST LANSING, MICHIGAN 48824, US**

⑯ Fecha de publicación de la mención BOPI:
09.05.2012

⑯ Inventor/es:
**BENNING, Christopher y
CERNAC, Alex**

⑯ Fecha de la publicación del folleto de la patente:
09.05.2012

⑯ Agente/Representante:
Carvajal y Urquijo, Isabel

ES 2 380 178 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 del metabolismo de lípidos en plantas.

Campo de la invención

5 La presente invención se refiere en general al metabolismo de los compuestos que se almacenan en las semillas en las plantas. Más específicamente, la presente invención se refiere a secuencias de ácidos nucleicos que codifican a las proteínas reguladoras del metabolismo de lípidos y el uso de estas secuencias en la producción de plantas transgénicas.

Antecedentes de la invención

10 El estudio y la manipulación genética de las plantas tiene una larga historia que comenzó incluso antes de que los estudios famosos de Gregor Mendel en el siglo 19. En el perfeccionamiento de esta ciencia, los científicos han logrado modificaciones de características particulares de las plantas que van desde los tubérculos de la patata que tienen mayor contenido de almidón hasta plantas oleaginosas como la canola y el girasol que tienen un mayor contenido o un contenido alterado de ácidos grasos. Con el aumento en el consumo y el uso de aceites vegetales, la modificación del contenido de aceite en las semillas y del nivel de aceite en las semillas se ha vuelto cada vez más generalizada. La producción o composición de aceite de semillas ha sido modificada en numerosas plantas tradicionales de semillas oleaginosas como la soja (Patente de EE.UU. N ° 5.955.650), canola (Patente de EE.UU. No. 5.955.650), girasol (Patente de EE.UU. No. 6.084.164) y colza (Toepfer et al. 1995, *Science* 268: 681 - 686), así como plantas de semillas oleaginosas no tradicionales, tales como tabaco (Cahoon et al. 1992, *Proc. Natl. Acad. EE.UU.* 89: 11184 - 11188).

20 20 Los aceites de las semillas de las plantas contienen tanto lípidos neutros como polares (véase la Tabla 1). Los lípidos neutros de la semilla contienen principalmente triacilglicerol, que es el lípido almacenado principal que se acumula en los cuerpos oleosos en las semillas. Los lípidos polares se encuentran principalmente en las diversas membranas de las células de las semillas, por ejemplo, el retículo endoplasmático, las membranas microsómicas y la membrana celular. Los lípidos neutros y polares contienen varios ácidos grasos comunes (véase la Tabla 2) y una variedad de ácidos grasos menos comunes. La composición de ácidos grasos de los lípidos de la membrana está muy regulada y sólo un número selecto de los ácidos grasos se encuentran en los lípidos de la membrana. Por otro lado, un gran número de ácidos grasos inusuales pueden ser incorporados en los lípidos neutros que se almacenan en las semillas de muchas especies de plantas (Van de Loo, F. J. et al. 1993, *Unusual Fatty Acids in Lipid Metabolism in Plants*, páginas 91 - 126, editor TS Moore Jr. CRC Press).

30 30 Los lípidos se sintetizan a partir de ácidos grasos y su síntesis se puede dividir en dos partes: la ruta procariota y la ruta eucariota (J Ohlrogge Browne & J 1995, *Lipid Biosynthesis Plant Cell* 7: 957 - 970). La ruta procariota está localizada en los plástidos que son el sitio primario de la biosíntesis de ácidos grasos. La síntesis de ácidos grasos comienza con la conversión de acetil-CoA a malonil-CoA por medio de la acetil-CoA carboxilasa (ACCasa). La malonil-CoA se convierte en malonil-ACP por medio de la malonil-CoA:ACP transacilasa. La enzima beta-ceto-acil-ACP-sintetasa III (KAS III) cataliza una reacción de condensación en la cual se transfiere el grupo acilo de la acetil-CoA a la malonil-ACP para formar 3-cetobutiril-ACP. En una serie posterior de reacciones de condensación, reducción y deshidratación, la cadena naciente de ácido graso sobre el cofactor ACP se alarga por medio de la adición paso a paso (condensación) de dos átomos de carbono donados por la malonil-ACP hasta que se forma una cadena saturada de ácido graso de 16 a 18 carbonos. La delta-9-acil-ACP desaturasa plastidial introduce el primer doble enlace insaturado en el ácido graso. Las tioesterasas escinden los ácidos grasos a partir del cofactor ACP y se exportan los ácidos grasos libres al citoplasma, donde participan como ésteres grasos de acil-CoA en la ruta eucariota. En esta ruta se esterifican los ácidos grasos por medio de la glicerol-3-fosfato aciltransferasa y la aciltransferasa del ácido lisofosfatídico en las posiciones sn-1 y sn-2 del glicerol-3-fosfato, respectivamente, para producir ácido fosfatídico (PA). El PA es el precursor de otros lípidos polares y neutros, siendo este último formado en la ruta de Kennedy (Shanklin y Cahoon 1998, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49: 611 - 641; Voelker 1996, *Genetic Engineering* ed.: J. K. Setlow 18:111 - 13; Frentzen 1998, *Lipid* 100 (4 - 5): 161 - 166).

40 40 La acetil-CoA en los plástidos es la precursora central para la biosíntesis de lípidos. La acetil-CoA se puede formar en los plástidos por medio de reacciones diferentes y la contribución exacta de cada reacción se sigue debatiendo (J & J Ohlrogge Navega 1995, *Lipid Biosynthesis Plant Cell* 7: 957 - 970). Sin embargo, se acepta que una gran parte de la acetil-CoA se deriva de la glucosa-6-fosfato y piruvato (es decir fosfoenolpiruvato) que se importan desde el citoplasma dentro de los plástidos. La sacarosa se produce en los órganos fuente (hojas, donde se realiza la fotosíntesis), y es transportada a las semillas en desarrollo que también son denominados órganos que se hunden. En las semillas en desarrollo, la sacarosa es el precursor de todos los compuestos de almacenamiento, es decir, almidón; lípidos y parcialmente las proteínas que se almacenan en las semillas. Por lo tanto, es claro que el metabolismo de carbohidratos en el cual la sacarosa juega un papel central es muy importante para la acumulación de compuestos que se almacenan en la semilla.

50 50 Aunque el contenido de lípidos y de ácido graso del aceite de la semilla puede ser modificado por medio de métodos tradicionales de fitomejoramiento de plantas, la llegada de la tecnología del ADN recombinante ha permitido una

manipulación más fácil del contenido de aceite de la semilla de una planta, y en algunos casos, ha permitido la alteración de los aceites de la semilla en formas que no podrían llevarse a cabo solamente mediante fitomejoramiento. Por ejemplo, la introducción de una secuencia de ácido nucleico para Δ^{12} -hidroxilasa en tabaco transgénico dio como resultado la introducción de un ácido graso novedoso, ácido ricinoleico, en el aceite de la semilla de tabaco (Van de Loo, et al. 1995, Proc. Natl. Acad. Sci. EE.UU. 92: 6743 - 6747). Las plantas de tabaco también han sido modificadas por ingeniería genética para producir bajos niveles de ácido petroselínico por medio de la introducción y expresión de una acil-ACP desaturasa del cilantro (Cahoon et al. 1992, Proc. Natl. Acad. Sci. EE.UU. 89: 11184 - 11188).

La modificación del contenido de aceite de la semilla en las plantas tiene importantes ramificaciones médicas, nutricionales y económicas. Con respecto a las ramificaciones médicas, los ácidos grasos de cadena larga (C18 y mayores) que se encuentran en muchos aceites de semilla se han vinculados a la reducción de la hipercolesterolemia y otros trastornos clínicos relacionados con la enfermedad cardíaca coronaria (Brenner R. R. 1976, Adv. Exp. Biol. 83: 85 - 101). Por lo tanto, el consumo de una planta que tiene niveles elevados de estos tipos de ácidos grasos puede reducir el riesgo de una enfermedad cardíaca. Niveles más altos del contenido de aceite en la semilla también aumentan la producción a gran escala y por lo tanto reducen el costo de estos aceites.

Con el fin de aumentar o alterar los niveles de compuestos tales como el aceite en la semilla en las plantas, se deben identificar las secuencias de ácido nucleico y las proteínas relacionadas con el metabolismo de ácido graso y de lípidos. Como se mencionó anteriormente, se han clonado varios ácidos nucleicos para desaturasa tales como el ácido nucleico para Δ^6 -desaturasa, el ácido nucleico para Δ^{12} -desaturasa y el ácido nucleico para acil-ACP desaturasa y se demostró que codifican las enzimas necesarias para la síntesis de ácidos grasos en varias especies de plantas. Se han clonado también secuencias de ácido nucleico para oleosina a partir de especies tan diferentes como *Brassica*, soja, zanahoria, pino y *Arabidopsis* también y se determinó que codifican proteínas asociadas con la membrana monocapa de fosfolípidos de los cuerpos oleosos en esas plantas.

Aunque se han clonado e identificado varios ácidos nucleicos que están involucrados en etapas enzimáticas del metabolismo de lípidos, ácidos grasos y almidones, probablemente existen una multitud de tales ácidos nucleicos de las plantas que todavía deben ser identificados. El análisis fenotípico de diferentes plantas oleaginosas y de otras plantas mutadas ha revelado otras proteínas putativas implicadas en el metabolismo de lípidos en las plantas, pero el estado de la técnica aún tiene que describir la localización genómica de estas proteínas, o de los ácidos nucleicos que las codifican.

Un ejemplo de un estudio es el del planta oleaginosa *Arabidopsis thaliana*. En 1998, Focks y Benning aislaron y caracterizaron un mutante arrugado de *Arabidopsis thaliana* denominado *wri1* (Plant, Physiology 1998, 118: 91 - 101). Este *wri1* tiene un menor contenido de aceite en la semilla que se especuló que es debido a un defecto en la regulación específica de la semilla del metabolismo de carbohidratos. En el mutante *wri1*, se redujeron las actividades de diferentes enzimas glicolíticas y se alteró en las semillas mutantes la incorporación de sacarosa y de glucosa en los lípidos triacilglicerol, mientras que las moléculas precursoras importantes para la biosíntesis de lípido plastidial, como piruvato y acetato, fueron incorporadas en tasas mayores. Esta evidencia bioquímica fue interpretada por Focks y Benning como una indicación de que la proteína WRI1 podría ser una proteína reguladora que gobierna el metabolismo de carbohidratos durante el desarrollo de la semilla o una hexoquinasa que puede actuar como un sensor de azúcar en las semillas en desarrollo, y por lo tanto que controla las actividades de diferentes enzimas glicolíticas (Fisiología Vegetal 1998, 118: 91 - 101). El fenotipo de *wri1* (semillas arrugadas) ha sido encontrado en dos diferentes mutantes alélicos de *Arabidopsis thaliana*, a saber, *wri1-1* y *wri1-2*.

Desde el descubrimiento del fenotipo de *wri1* por Focks y Benning, el genoma de *Arabidopsis thaliana* fue secuenciado en su totalidad (The *Arabidopsis* Genome initiative 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796 - 844). Estas secuencias genómicas han sido anotadas y se han asignado los marcos de lectura abiertos que codifican las proteínas putativas en un proceso automático. Es importante destacar sin embargo, que esta anotación se basa exclusivamente en homologías con otras secuencias con funciones conocidas, y por lo tanto, en ningún modo identifica la ubicación verdadera, la secuencia o la funcionalidad de una secuencia de ácido nucleico de *Arabidopsis thaliana*. La anotación y asignación de los marcos de lectura abiertos tampoco describen de ninguna manera la ubicación, la función o la secuencia del gen *wri1*.

Por lo tanto, lo que se necesita en la técnica es la elucidación de la ubicación y la identidad de uno o más de los ácidos nucleicos que están asociados con la mutación *wri1* en *Arabidopsis thaliana* junto con una comprensión de la funcionalidad de las proteínas y fragmentos de proteínas codificadas por aquellos ácidos nucleicos.

Breve descripción de los dibujos

La Figura 1 es una representación esquemática de la posición de los clones BAC, F24B22, T12E18, T14E10 y T5N23, los marcadores en ellos y el locus genético de LMR (denominado como "posición de WRI1").

La Figura 2 es una representación esquemática de la estrategia de complementación de *wri1* mediante la utilización de clones de cósidos de tipo silvestre. Los cósidos denominados como AG1-3 y 5 rescataron el fenotipo de *wri1* y restablecieron el contenido de aceite de la semilla hasta niveles cercanos al tipo silvestre (el cósido AG4 nunca

fue transformado con éxito). La región de solapamiento entre los cósmidos de rescate es de 8,5 kb, lo que indica que el gen *ARRUGADO1* está contenido dentro de esa región de 8,5 kb.

5 La Figura 3 es una alineación de la secuencia de aminoácidos de un ORF codificado por un ADN para *WRI1* (SEQ ID NO: 3) y las secuencias de aminoácidos de dos ORF separados (SEQ ID NO: 8 y SEQ ID NO: 9) que fueron predichas por medio de los programas de búsqueda automática de genes. La alineación muestra que contiene *WRI1* de ambos ORF predichos y además varios aminoácidos, que son el resultado de las diferencias en los sitios de empalme predichos.

10 Las Figuras 4 (A - C) muestran la secuencia de polinucleótidos del ADNc (SEQ ID NO: 1), la secuencia de nucleótidos del marco de lectura abierto (SEQ ID NO: 2) y la secuencia de aminoácidos (SEQ ID NO: 3) de un LMR de *Arabidopsis thaliana*.

Las Figuras 5 (A - C) muestran la secuencia de polinucleótido del ADNc (SEQ ID NO: 4), la secuencia de nucleótidos del marco de lectura abierto (SEQC ID NO: 5) y la secuencia de aminoácidos de un LMR de *Arabidopsis thaliana*.

15 La Figura 6 muestra la secuencia truncada de polinucleótidos del clon BAC de T12E18 (Genbank, número de acceso AL132971).

15 Resumen de la invención

La presente invención está dirigida a nuevo ácido nucleico aislado y a las secuencias de aminoácidos asociadas con el metabolismo de los compuestos almacenados en la semilla de las plantas. Un nuevo descubrimiento descrito en este documento se relaciona con la identificación de las secuencias de ácido nucleico que codifican el locus genético *WRINKLED1* (*WRI1*), y que codifica por lo tanto un locus genético del Regulador del Metabolismo de los Lípidos (LMR). Se mapeó primero la secuencia de ácido nucleico de *wri1* mutante para un intervalo de ciento cincuenta kilobases que se extiende entre los marcadores CDC2BG y TSA1 sobre el cromosoma III de *Arabidopsis thaliana*. La transformación del mutante de *Arabidopsis wri1* con vectores binarios cósmidos que contienen las secuencias genómicas que abarcan esta región conducen al descubrimiento de cuatro cósmidos superpuestos que complementaban al mutante *wri1* y restauró el fenotipo de las semillas de tipo silvestre y el contenido de aceite de las semillas (Figura 2 y el Ejemplo 13). El fragmento genómico común más pequeño sobre los cinco vectores binarios cósmidos que se complementan es de 8,5 kb, lo que indica que el gen *WRI1* está contenido dentro de esa región de 8,5 kb.

30 De acuerdo con la anotación generada automáticamente de la secuencia genómica en el GenBank, este fragmento de 8,5 kb contiene completamente dos marcos de lectura abiertos (los ORF, SEQ ID NOs: 8 y 9) y los fragmentos parciales de otros dos ORF. Sin embargo, mediante el uso de RT-PCR y RACE-PCR se determinó que los dos ORF (SEQ ID NOs: 8 y 9) son en realidad parte del mismo gen que contiene un intrón inusualmente largo de 1641 nucleótidos. Con base en el mutante *wri1* y los datos del mapeo, este gen fue denominado "*WRI1*". Se aisló el ADNc completo de *WRI1* por RT-PCR utilizando una polimerasa para corrección de errores y se determinó su secuencia nucleotídica (SEQ ID NO: 1). El ADNc codificado por este gen tiene una longitud de 1539 pb, una UTR 3' de 166 pb y contiene un ORF de 1293 pb (SEQ ID NO: 2) que codifica una proteína de 430 aminoácidos (SEQ ID NO: 3). El ADN para *WRI1* muestra pequeñas homologías de secuencia, con un factor de transcripción conocido, Aintegumenta, y contiene un dominio de enlazamiento de ADN tipo apetala-2. Los dominios tipo AP2 abarcan 64 - 134 y 166 - 225 aminoácidos de la SEQ ID NO: 3. Otras repeticiones interesantes en la SEQ ID NO: 3 abarcan 11 - 21, 266 - 272 y 401 - 423 aminoácidos. Las secuencia alternas para el ADNc de *WRI1* de longitud completa, el ORF y la proteína codificada se muestran en la SEQ ID NO: 4, SEQ ID NO: 5 y SEQ ID NO: 6, respectivamente. Por consiguiente, se propone que *WRI1* actúa como un factor de transcripción que regula el metabolismo lipídico y del compuesto que almacena la semilla durante el desarrollo de la semilla. Este ácido nucleico y la secuencia de aminoácidos recientemente descubiertos han sido etiquetados aquí como las secuencias del aminoácido y del ácido nucleico del Regulador del Metabolismo de Lípidos (o LMR).

45 En consecuencia, la presente invención se relaciona con el uso de ácidos nucleicos del LMR en la producción de plantas transgénicas que tengan un nivel modificado de compuesto que se almacena en la semilla. Un método para producir una planta transgénica con un nivel modificado de un compuesto que se almacena en las semillas incluye las etapas de transformación de una célula de la planta con un vector de expresión que comprende un ácido nucleico del LMR, y la generación de una planta con un nivel modificado del compuesto que se almacena en la semilla de la célula de la planta. En una realización preferida, la planta es una especie productora de aceite seleccionada de entre el grupo que consiste de colza, canola, linaza, soja, girasol, maíz, avena, centeno, cebada, trigo, pimienta, Tagetes, algodón, palma oleaginosa, palma cocotera, lino, ricino y maní, por ejemplo.

50 También se incluyen en este documento una semilla producida por una planta transgénica transformada por una secuencia de polinucleótidos del LMR, en donde la semilla contiene la secuencia de polinucleótidos del LMR y en donde la planta es una línea genéticamente pura para un nivel modificado de un compuesto que se almacena en la semilla. La presente invención incluye adicionalmente un aceite de semilla producido por la semilla antes mencionada.

De acuerdo con la presente invención, la composición y los métodos aquí descritos pueden ser utilizados para aumentar o disminuir el nivel de un lípido en un aceite de semillas, o para aumentar o disminuir el nivel de un ácido graso en un aceite de semillas, o para aumentar o disminuir el nivel de un almidón en una semilla o una planta. Un método para producir un nivel más alto de lo normal de compuesto de almacenamiento en una planta transgénica, que comprende expresar un ácido nucleico del LMR de *Arabidopsis thaliana* en la planta transgénica, en donde, la planta transgénica es una especie distinta de *Arabidopsis thaliana*. También se incluyen en este documento composiciones y métodos para la modificación de la eficiencia de producción de un compuesto que se almacena en la semilla.

5 En consecuencia, un objeto de la presente invención es proporcionar nuevos ácidos nucleicos aislados del LMR y aisladas secuencias de aminoácidos del LMR de *Arabidopsis thaliana*, así como fragmentos activos, análogos y ortólogos de los mismos.

10 Otro objetivo de la presente invención es proporcionar plantas transgénicas que tengan niveles modificados de los compuestos que se almacenan en la semilla y, en particular, niveles modificados de un lípido, un ácido graso o un almidón.

15 Un objetivo adicional de la presente invención es proporcionar métodos para producir tales plantas transgénicas antes mencionadas.

Otro objetivo de la presente invención es proporcionar semillas y aceites de semillas de plantas transgénicas tales como las anteriormente mencionadas.

20 Estos y otros objetivos, características y ventajas de la presente invención serán evidentes después de una revisión de la siguiente descripción detallada de las realizaciones descritas y las reivindicaciones adjuntas.

Descripción detallada

25 La presente invención está dirigida a una planta transgénica con un nivel modificado de un compuesto que se almacena en la semilla y un método para producir el mismo. La presente invención también abarca una semilla producida por una planta transgénica con un nivel de modificado de un compuesto que se almacena en la semilla y un método para la producción del mismo. Preferiblemente, la planta es de una especie productora de aceite y el compuesto que se almacena en la semilla es un ácido graso, un lípido o almidón.

30 La presente invención incluye adicionalmente secuencias de ácido nucleico que codifican y dirigen la expresión de proteínas que regulan el metabolismo de un compuesto que se almacena en la semilla. Estas secuencias de ácido nucleico son llamadas secuencias de ácido nucleico "reguladoras del metabolismo de lípidos" y se denominan en lo sucesivo como "ácidos nucleicos para el LMR" o "polinucleótidos LMR". Debe entenderse sin embargo que la nomenclatura utilizada en este documento de ninguna manera limita la funcionalidad de las secuencias de aminoácidos codificadas por la secuencia de ácido nucleico descrita en este documento. En una realización, un ácido nucleico LMR descrito aquí se introduce en una planta, preferiblemente de especies productoras de aceite, lo que resulta en la modificación de los niveles de un compuesto que se almacena en la semilla de las mismas.

35 "Plantas" se refieren a organismos fotosintéticos, tanto eucariotas como procariotas, mientras que el término "plantas superiores" se refiere a plantas eucarióticas. "Especie productora de aceite" como se usa aquí se refiere a especies de plantas que producen y almacenan triacilglicerol en órganos específicos y principalmente en las semillas. Las plantas de las especies productoras de aceite incluyen, pero no se limitan a, colza, canola, linaza, soja, girasol, maíz, avena, centeno, cebada, trigo, pimienta, tagetes, algodón, palma oleaginosa, palma de coco, lino, ricino y maní. El grupo de plantas de especies productoras de aceite también incluye especies no agronómicas que son útiles en el desarrollo de vectores de expresión apropiados, tales como tabaco, especie *Brassica* de ciclo rápido, *Arabidopsis thaliana* y especies silvestres que pueden ser una fuente de ácidos grasos únicos. Además de las plantas de especies productoras de aceite, también se incluyen plantas tuberosas en la presente invención.

45 Tal como se usa aquí, el término "compuesto que se almacena en la semilla" se refiere a un lípido, un ácido graso, un almidón o una proteína que se almacena en la semilla. El término "lípido" incluye, pero no está limitado a, un lípido de triacilglicerol, un lípido plastidial, o un lípido de la membrana. Las clases de lípidos incluidas en la presente invención se muestran en la Tabla 1 a continuación. En una realización de la presente invención, un ácido nucleico LMR de la presente invención regula el metabolismo de un lípido de triacilglicerol, un lípido de la membrana, o ambos. El término "ácido graso" es bien conocido por aquellos ordinariamente capacitados en el arte e incluye tanto ácidos grasos saturados como ácidos grasos poliinsaturados, incluyendo, pero sin limitarse a, aquellos ácidos grasos enlistados en la Tabla 2 más abajo. El término "almidón" es bien conocido por aquellos ordinariamente capacitados en el arte y como se usa aquí incluye amilosa y amilopectina. El término "proteína que se almacena en la semilla" es bien conocido por aquellos ordinariamente capacitados en el arte e incluye a la proteína de albúmina 2S, a la proteína cruciferina, a la proteína de albúmina 12S y oleosina.

Tabla 1

Clases de lípidos de las plantas	
Lípidos neutros	Triacilglicerol (TAG)
	Diacilglicerol (DAG)
	Monoacilglicerol (MAG)
Lípidos polares	Monogalactosildiacilglicerol (MGDG)
	Digalactosildiacilglicerol (DGDG)
	Fosfatidilglicerol (PG)
	Fosfatidilcolina (PC)
	Fosfatidiletanolamina (PE)
	Fosfatidilinositol (PI)
	Fosfatidilserina (PS)
	Sulfoquinovosildiacilglicerol

Tabla 2

Ácidos grasos comunes en las plantas	
16:0	Ácido palmítico
16:1	Ácido palmitoléico
16:3	Ácido palmitolénico
18:0	Ácido esteárico
18:1	Ácido oleico
18:2	Ácido linoleico
18:3	Ácido linolénico

(continuación)

Ácidos grasos comunes en las plantas	
g-18:3	Ácido gamma-linolénico *
20:0	Ácido araquidónico
22:6	Ácido docosahexanoico (DHA) *
20:2	Ácido eicosadienoico
20:4	Ácido araquidónico (AA)*
20:5	Ácido eicosapentaenoico (EPA) *
22:1	Ácido erúcico
* Estos ácidos grasos no se presentan normalmente en los aceites de las semillas de las plantas, pero su producción en los aceites de las semillas de plantas transgénicas es de importancia en la biotecnología de las plantas.	

La capacidad de producir una planta transgénica que tenga un nivel modificado de un compuesto que se almacena en la semilla se encuentra en un nuevo descubrimiento descrito en este documento que comprende la identificación de las secuencias de ácido nucleico que codifican el locus genético *WRI1*, y que por lo tanto codifican un locus genético regulador de metabolismo de los lípidos (LMR). Más particularmente, la presente invención proporciona secuencias de polinucleótidos LMR y de polipéptidos localizadas dentro del locus genético LMR. Con el fin de identificar las secuencias de polinucleótidos LMR y de polipéptidos, se mapeó primero la secuencia de ácido nucleico de *wri1* mutante para un intervalo de ciento cincuenta y cinco kilobases que se extiende entre los marcadores CDC2BG y TSA1 sobre el cromosoma III (véase la Tabla 3 y la Figura 1). Por lo tanto, el término "locus genético LMR" se refiere a una secuencia de ácido nucleico entre los marcadores CDC2BG y TSA1 sobre el cromosoma III de *Arabidopsis thaliana*. Como se muestra en la Figura 1, el marcador CDC2BG está situado sobre el clon BAC F24B22 (GenBank, número de acceso AL132957) mientras que el marcador TSA1 se encuentra tanto en el extremo 5' del clon BAC T14E10 (GenBank, número de acceso AL138656) y puede que sobre el extremo 3' del clon BAC T5N23 (GenBank, número de acceso AL132970). Porciones del locus genético LMR, se encuentra por lo tanto sobre los dos clones BAC, F24B22 y T14E10 así como un clon BAC interviniente, T12E18 (GenBank, número de acceso 132971).

Cualquiera persona ordinariamente capacitado en la técnica se dará cuenta que el descubrimiento de la ubicación exacta de una lesión genética como *wri1* sobre un cromosoma es un avance significativo y útil en la identificación de la secuencia de ácido nucleico y su proteína codificada que es responsable por el fenotipo de *WRI1*. De manera similar, cualquier persona ordinariamente capacitada en la técnica se dará cuenta que el mapeo de una mutación para ciertas secuencias de ácido nucleico con una función hasta ahora desconocida en ese lugar sobre el cromosoma, proporciona evidencia inequívoca de la función de los ácidos nucleicos en ese locus y sus proteínas codificadas.

Por lo tanto, en una realización preferida de la presente invención, una secuencia aislada de aminoácidos del LMR participa en el metabolismo de los compuestos necesarios para la construcción de membranas celulares, lípidos y ácidos grasos que se almacenan en la semilla, o almidón o proteína de la semilla en una célula de una planta, preferiblemente localizados en las semillas, o en el transporte de moléculas a través de sus membranas. En una realización preferida adicional, el ácido nucleico LMR que codifica la secuencia de aminoácidos del LMR está localizado en el cromosoma III de *Arabidopsis thaliana*. En una realización aún más preferida, el ácido nucleico LMR está situado sobre el locus genético LMR como se describe aquí. En particular, el ácido nucleico LMR está localizado entre los marcadores CDC2BG y TSA1 sobre el cromosoma III de *Arabidopsis thaliana*. En una realización de la presente invención, el ácido nucleico LMR comprende una secuencia de polinucleótidos seleccionada de entre el grupo que consta de un polinucleótido de la SEQ ID NO: 1; un polinucleótido de la SEQ ID NO: 2; un polinucleótido de la SEQ ID NO: 4; un polinucleótido de la SEQ ID NO: 5; una secuencia de polinucleótidos que codifica un polipéptido de la SEQ ID NO: 3; una secuencia de polinucleótidos que codifica un polipéptido de la SEQ ID NO: 6, un polinucleótido que comprende al menos 60 nucleótidos consecutivos de cualquiera de los polinucleótidos anteriormente mencionados; un polinucleótido que tiene al menos 70% de identidad de secuencia con cualquiera de los polinucleótidos antes mencionados; un polinucleótido que codifica un polipéptido que tiene al menos el 70% de identidad de secuencia con cualquiera de los polipéptidos antes mencionados; y un

polinucleótido complementario a cualquiera de los polinucleótidos antes mencionados. En una realización preferida adicional, el ácido nucleico LMR comprende una secuencia de polinucleótidos mostrada en la SEQ ID NO: 1, la SEQ ID NO: 2, la SEQ ID NO: 4 o la SEQ ID NO: 5.

5 Tal como se usan aquí, los términos "ácido nucleico" y "polinucleótido" se refieren al ARN o al ADN que es lineal o ramificado, monocatenario o bicatenario, o un híbrido de los mismos. El término también abarca híbridos de ARN / ADN. Estos términos también abarcan una secuencia no traducida localizada tanto en el extremo 3' como en el extremo 5' de la región de codificación del gen: al menos aproximadamente 1000 nucleótidos de la secuencia hacia arriba del extremo 5' de la región de codificación y al menos aproximadamente 200 nucleótidos de la secuencia hacia abajo del extremo 3' de la región codificadora del gen. La invención proporciona además un ácido nucleico aislado que codifica al LMR. En realizaciones preferidas, el ácido nucleico que codifica al LMR se selecciona entre un polinucleótido mostrado en la SEQ ID NO: 2 y un polinucleótido mostrado en la SEQ ID NO: 5.

10 Las bases menos comunes, tales como inosina, 5-metilcitosina, 6-metiladenina, hipoxantina y otras también pueden utilizarse para el apareamiento antisentido de ARNbc y ribozima. Por ejemplo, se ha demostrado que los polinucleótidos que contienen análogos de propino C-5 de uridina y citidina se unen con ARN con una alta afinidad y que son inhibidores potentes antisentido de la expresión génica. También pueden hacerse otras modificaciones, tales como la modificación de la columna vertebral de fosfodiéster, o de la 2'-hidroxi en el grupo azúcar de la ribosa del ARN. Los polinucleótidos antisentido y los ribozimas pueden consistir enteramente en ribonucleótidos, o puede contener ribonucleótidos y desoxirribonucleótidos mezclados. Los polinucleótidos de la invención pueden ser producidos por cualquier medio, incluyendo preparaciones genómicas, preparaciones de ADNc, síntesis *in vitro*, RT-PCR y transcripción *in vitro* o *in vivo*.

15 Una molécula "aislada" de ácido nucleico es una que está sustancialmente separada de otras moléculas de ácido nucleico, que están presentes en la fuente natural del ácido nucleico (es decir, secuencias que codifican otros polipéptidos). Preferiblemente, un ácido nucleico "aislado" está libre de algunas de las secuencias, que flanquean naturalmente al ácido nucleico (es decir, las secuencias localizadas en los extremos 5' y 3' del ácido nucleico) en su replicón de origen natural. Por ejemplo, un ácido nucleico clonado se considera aislado. En diversas realizaciones, la molécula de ácido nucleico aislada del LMR puede contener aproximadamente menos de 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de las secuencias de nucleótidos que flanquean naturalmente la molécula de ácido nucleico en el ADN genómico de la célula a partir de la cual se deriva el ácido nucleico (por ejemplo, una célula de *Arabidopsis thaliana*). Un ácido nucleico es también considerado aislado si ha sido alterado por intervención humana, o colocado en un locus o ubicación que no es su sitio natural, o si se introduce en una célula por medio de agroinfección. Además, una molécula "aislada" de ácido nucleico, tal como una molécula de ADNc, puede estar libre de algo de otro material celular con el que está asociado de forma natural, o medio de cultivo cuando es producida por técnicas recombinantes, o precursores químicos u otros productos químicos cuando se la sintetiza químicamente. En una realización, la expresión "sustancialmente libre de otro material celular" incluye preparaciones de secuencias de aminoácidos LMR que tienen aproximadamente menos de 30% (en peso seco) de la secuencia de aminoácidos sin LMR (denominada también aquí como "proteína contaminante"), más preferiblemente aproximadamente menos de 20% de proteína contaminante, aún más preferiblemente aproximadamente menos de 10% de proteína contaminante, y lo más preferiblemente aproximadamente menos de 5% de proteína contaminante.

20 25 30 35 40 45 50 55 60 Específicamente excluidos de la definición de "ácidos nucleicos aislados" están: los cromosomas de origen natural (tal como cromosomas esparcidos), bibliotecas cromosómicas artificiales, bibliotecas genómicas, y bibliotecas de ADNc que existen ya sea como una preparación de ácido nucleico *in vitro* o como una preparación de células huésped transfectadas/transformadas, en donde las células huésped son ya sea una preparación heterogénea *in vitro* o sembradas en placa como una población heterogénea de colonias individuales. También se excluyen específicamente las bibliotecas anteriores en donde un ácido nucleico específico representa menos del 5% del número de insertos de ácido nucleico en las moléculas de vector. Se excluyen más específicamente además ADN genómico de la célula entera o preparaciones de ARN de células enteras (incluidas las preparaciones de células enteras que son cortadas mecánicamente o digeridas enzimáticamente). Incluso se excluyen específicamente además las preparaciones de células enteras que se encuentran ya sea como una preparación *in vitro* o como una mezcla heterogénea separada por electroforesis en donde el ácido nucleico de la invención no ha sido separado adicionalmente de los ácidos nucleicos heterólogos en el medio de electroforesis (por ejemplo, separando aún más, por escisión de una sola banda a partir de una población de banda heterogénea en un gel de agarosa o transferencia a una membrana de nailon).

Los ácidos nucleicos del LMR de la presente invención se producen preferiblemente mediante técnicas de ADN recombinante. Por ejemplo, una molécula de ácido nucleico del LMR que codifica un polipéptido LMR se clona en un vector de expresión, se introduce el vector de expresión en una célula huésped, y el ácido nucleico del LMR se expresa en la célula huésped. Se puede aislar entonces la proteína LMR de las células mediante un esquema de purificación apropiado utilizando técnicas estándar de purificación de polipéptidos. Para los fines de la invención, el término "polinucleótido recombinante" se refiere a un polinucleótido que ha sido alterado, reordenado o modificado por medio de ingeniería genética. Los ejemplos incluyen cualquier polinucleótido clonado, y polinucleótidos que están enlazados o unidos a secuencias heterólogas. El término "recombinante" no se refiere a las alteraciones con polinucleótidos que resultan de eventos que ocurren naturalmente, tales como mutaciones espontáneas. Como

alternativa a la expresión recombinante, una proteína o péptido LMR pueden ser sintetizados químicamente usando técnicas estándar de síntesis de péptidos. Además, se puede aislar de las células una proteína nativa LMR (por ejemplo, *Arabidopsis thaliana*), por ejemplo utilizando un anticuerpo de proteína anti-LMR, que puede ser producido por medio de técnicas estándar que utilizan una proteína LMR o un fragmento de la misma.

5 Tal como se usa aquí, los términos "polipéptido" y "proteína" se refieren a una cadena de al menos cuatro aminoácidos unidos por enlaces peptídicos. La cadena puede ser lineal, ramificada, circular o combinaciones de las mismas. En una realización preferida, la proteína LMR es una proteína de *Arabidopsis thaliana* que regula el metabolismo de un compuesto que se almacena en la semilla en una planta. En una realización preferida adicional, la proteína LMR comprende una secuencia de polipéptidos como se muestra en la SEQ ID NO: 3 o en la SEQ ID NO: 6. También se debe entender que los polipéptidos LMR de la presente invención incluyen aquellos codificados por una secuencia de polinucleótidos muestrada en la SEQ ID NO: 2 o en la SEQ ID NO: 5, con o sin modificaciones post-traduccionales. Las modificaciones post-traduccionales incluyen la formación química de derivados *in vivo* y *in vitro* de polipéptidos, por ejemplo, acetilación, carboxilación, fosforilación o glicosilación, y tales modificaciones pueden presentarse durante la síntesis o el procesamiento de polipéptidos o después del tratamiento con enzimas modificadoras aisladas. En una realización de la presente invención, los polinucleótidos y polipéptidos LMR se derivan de una especie de *Arabidopsis*, y más preferiblemente, de *Arabidopsis thaliana*.

Además de las secuencias de polipéptidos y polinucleótidos LMR descritas anteriormente, la presente invención incluye fragmentos de polipéptidos y polinucleótidos LMR y plantas y células de plantas que contienen estos fragmentos. Un fragmento de polinucleótido LMR puede comprender una porción de la secuencia en la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 5, por ejemplo, un fragmento que puede ser utilizado como una sonda o un iniciador. Tal como se usa aquí, el término "fragmento" significa al menos 20 aminoácidos contiguos, preferentemente al menos 30 aminoácidos contiguos, más preferiblemente al menos 50 aminoácidos contiguos, y más preferible al menos 60 a 80 o más aminoácidos contiguos. Se pueden generar fragmentos de proteínas LMR por medio de métodos conocidos por aquellos capacitados en el arte o pueden resultar del procesamiento normal de proteína (por ejemplo la remoción de aminoácidos del polipéptido naciente que no se requieren para actividad biológica o remoción de aminoácidos por medio de empalme alternativo de ARNm o eventos que profesan proteína alternativa). En una realización preferida, un fragmento de proteína LMR retiene la actividad biológica de la proteína LMR de longitud completa. Como se usa aquí, el término "porción biológicamente activa de" un polipéptido LMR se entiende que incluye una porción, por ejemplo, un dominio / motivo, de un polipéptido LMR que participa en la modulación de los niveles de compuesto que se almacena en la semilla en una planta.

Además de los fragmentos de LMR, la presente invención incluye un homólogo, una variante alélica, un análogo, un ortólogo, o un parálogo de una proteína LMR de origen natural en una planta de una especie productora de aceite. Mientras que las diferencias en las secuencias de aminoácidos entre una proteína LMR de origen natural y su variante alélica son naturales, las diferencias en las secuencias de aminoácidos entre una proteína LMR de origen natural y su análogo, ortólogo o parálogo pueden ser naturales o inducidas. Tal como se utiliza aquí una "proteína de origen natural" se refiere a una secuencia de aminoácidos que se produce en la naturaleza. Preferiblemente, una proteína LMR de origen natural comprende una secuencia de aminoácidos correspondiente a la SEQ ID NO: 3 o a la SEQ ID NO: 6.

40 Un caso de homólogos de LMR es el de las variantes alélicas. Tal como se usa aquí, el término "variante alélica" se refiere a una secuencia de ácido nucleico que se produce en un locus genético del LMR en una planta de *Arabidopsis thaliana*. Tales variaciones alélicas naturales típicamente pueden dar lugar a una varianza de 1 - 5% en un ácido nucleico LMR. Al menos dos variantes alélicas de un ácido nucleico LMR han sido identificadas fenotípicamente y etiquetadas *wri1-1* y *wri1-2*. Se pueden identificar alelos alternativos por medio de la secuenciación de la secuencia de ácido nucleico de interés en una cantidad de plantas diferentes, que puede ser fácilmente llevada a cabo mediante el uso de sondas de hibridación para la identificación del mismo locus genético del LMR en esas plantas. Todas y cada una de las variantes de ácido nucleico y los polimorfismos de los aminoácidos resultantes o las variaciones en una proteína LMR que sean el resultado de una variación alélica natural y que no alteran la actividad funcional de una proteína LMR se pretende que queden incluidas dentro del alcance de la invención.

50 Además, las moléculas de ácido nucleico que codifican análogos, ortólogos y paralogs de LMR, que tienen una secuencia de polinucleótidos que difiere de aquella de un gen para LMR de *Arabidopsis thaliana*, se pretende que estén dentro del alcance de la presente invención. Como se usa aquí, el término "análogos" se refiere a dos ácidos nucleicos que tienen la misma o una función similar, pero que han evolucionado por separado en organismos no relacionados. Tal como se usa aquí, el término "ortólogos" se refiere a dos ácidos nucleicos de diferentes especies, pero que han evolucionado a partir de un gen ancestral común por especiación. Normalmente, los ortólogos codifican proteínas que tienen las mismas funciones o similares. Como también se usa aquí, el término "parálogos" se refiere a dos ácidos nucleicos que están relacionados por la duplicación dentro de un genoma. Normalmente, los parálogos tienen diferentes funciones, pero estas funciones pueden estar relacionadas (Tatusov, R. L. et al. 1997, *Science* 278 (5338): 631 - 637). Los ortólogos de la invención generalmente exhibirán por lo menos 80 - 85%, más preferiblemente 90%, y lo más preferible 95%, 96%, 97%, 98% o incluso 99% de identidad u homología con toda o parte de una secuencia de aminoácidos de LMR de origen natural. La longitud de comparación de la secuencia es al

menos de 15 residuos de aminoácidos, preferiblemente de al menos 25 residuos de aminoácidos, y más preferiblemente de al menos 35 residuos de aminoácidos.

Para determinar el porcentaje de identidad de secuencia de dos secuencias de aminoácidos (por ejemplo, las secuencias de la SEQ ID NO: 3 o de la SEQ ID NO: 6, y un mutante o homólogo de las mismas), se alinean las secuencias con el propósito de una comparación óptima (por ejemplo, se pueden introducir espacios en la secuencia de un polipéptido para una alineación óptima con el otro polipéptido). Se comparan los residuos de aminoácidos en las correspondientes posiciones de los aminoácidos. Cuando una posición en una secuencia (por ejemplo, la SEQ ID NO: 3 o la SEQ ID NO: 6) está ocupada por el mismo residuo aminoácido que la correspondiente posición en la otra secuencia (por ejemplo, un mutante u homólogo de la secuencia mostrada en la SEQ ID NO: 3 o en la SEQ ID NO: 6), entonces las moléculas son idénticas en esa posición. El mismo tipo de comparación se puede hacer entre dos secuencias de ácido nucleico.

En consecuencia, el porcentaje de identidad de secuencia entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (es decir, el porcentaje de identidad de secuencia = número de posiciones idénticas / el número total de posiciones x 100). Para los fines de la invención, el porcentaje de identidad de secuencia entre dos secuencias de polinucleótidos o de polipéptido se determina utilizando el paquete de software Vector NTI 6.0 (PC) (Informax, Wisconsin Ave. 7600., Bethesda, MD 20814). Se utilizan una penalización por apertura de espacio de 15 y una penalización por extensión de la apertura de 6,66 para determinar el porcentaje de identidad de dos polinucleótidos. Se utilizan una penalización por apertura de espacio de 10 y una penalización por extensión de la apertura de 0,1 para determinar el porcentaje de identidad de dos polipéptidos. Todos los demás parámetros se fijan en los ajustes predeterminados. Debe entenderse que para los propósitos de determinación de identidad de la secuencia cuando se compara una secuencia de ADN con una secuencia de ARN, un nucleótido de timidina es equivalente a un nucleótido de uracilo.

Preferiblemente, los homólogos aislados de aminoácidos incluidos en la presente invención son aproximadamente al menos 70 - 75%, 75 - 80%, 80 - 85%, 85 - 90% o 90 - 95%, y lo más preferible al menos aproximadamente 96%, 97%, 98%, 99% o más, idénticos a una secuencia entera de aminoácidos mostrada en la SEQ ID NO: 3 o en la SEQ ID NO: 6. En aún otra realización, los homólogos aislados de aminoácidos incluidos en la presente invención son al menos aproximadamente 70 - 75%, 75 - 80%, 80 - 85%, 85 - 90% o 90 - 95%, y lo más preferible al menos aproximadamente 96%, 97%, 98%, 99% o más, idénticos a una secuencia entera de aminoácidos codificada por una secuencia de ácido nucleico mostrada en la SEQ ID NO: 2 o en la SEQ ID NO: 5.

30 Se prefiere además que el homólogo del aminoácido del LMR actue para modular un compuesto que se almacena en la semilla en una planta.

En otra realización preferida, un homólogo del polinucleótido LMR de la invención incluye una secuencia de nucleótidos que es al menos aproximadamente 70 - 75%, 75 - 80%, 80 - 85%, 85 - 90% o 90 - 95%, y incluso más preferiblemente al menos aproximadamente 95%, 96%, 97%, 98%, 99% o más, idéntica a una secuencia de nucleótidos mostrada en la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 5. La longitud preferida de comparación de secuencias para ácidos nucleicos es la longitud total de la región de codificación. Se prefiere adicionalmente que el homólogo del polinucleótido LMR module un compuesto para almacenamiento en la semilla en una planta, por ejemplo, el nivel de un compuesto que se almacena en la semilla cuando se expresa o sobreexpresa allí.

40 Las moléculas de ácido nucleico correspondientes a variantes alélicas naturales y análogos, ortólogos y parálogos de un ADNc del LMR se pueden aislar con base a su identidad con los ácidos nucleicos del LMR de *Arabidopsis thaliana* aquí descritos utilizando los ADNc del LMR, o una porción de los mismos, como una sonda de hibridación de acuerdo con técnicas estándar de hibridación bajo condiciones estrictas de hibridación. Por lo tanto, la presente invención incluye una secuencia de ácido nucleico que hibrida (bajo condiciones estrictas o altamente estrictas) con una secuencia de ácidos nucleicos mostrada en la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 5.

Como se usa aquí con respecto a la hibridación para ADN con una transferencia de ADN, el término "condiciones estrictas" se refiere a la hibridación durante la noche a 60 ° C en solución de Denhart 10X, 6X SSC, SDS al 0,5% y 100 µg/ml de ADN de esperma de salmón desnaturalizado. Se lavan secuencialmente las transferencias a 62° C durante 30 minutos cada vez en 3X SSC/SDS al 0,1%, seguido por 1X SSC/SDS al 0,1% y finalmente 0,1 X SSC / SDS al 0,1%. Como también se usa aquí, "condiciones altamente estrictas" se refiere a la hibridación durante la noche a 65° C en solución de Denhart 10X, 6X SSC, SDS al 0,5% y 100 µg/ml de ADN de esperma de salmón desnaturalizado. Las transferencias se lavan secuencialmente a 65° C durante 30 minutos cada vez en 3X SSC/ SDS al 0,1%, seguido por 1X SSC/SDS al 0,1% y, finalmente 0,1X SSC/SDS al 0,1%. Los métodos para hibridaciones de ácido nucleico se describen en Meinkoth y Wahl, 1984, Anal. Biochem. 138: 267 - 284; Current Protocols in Molecular Biology, Capítulo 2, Ausubel et al. Eds, Greene Publishing and Wiley Interscience, Nueva York, 1995; y Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization with Nucleic Acid Probes, Parte I, Capítulo 2, Elsevier, Nueva York, 1993. Preferiblemente, una molécula aislada de ácido nucleico de la invención que se hibrida bajo condiciones estrictas o altamente estrictas con una secuencia de SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 5 corresponde a una molécula de ácido nucleico de origen natural.

- 5 El experto en la técnica se dará cuenta que pueden introducirse cambios por medio de mutación en la secuencia de aminoácidos de una proteína LMR, sin afectar la actividad biológica de la proteína LMR. Las mutaciones pueden ser introducidas por medio de técnicas estándar, tales como mutagénesis dirigida al sitio y mutagénesis mediada por PCR. Los residuos de aminoácidos que no se han conservado o sólo semi-conservado entre la proteína LMR de diferentes especies pueden ser no esenciales para la actividad y por lo tanto probablemente serían un objetivo para alteración. Un aminoácido "no esencial" es un residuo que puede ser alterado a partir de la secuencia de tipo silvestre de una proteína LMR (es decir, la SEQ ID NO: 3 o la SEQ ID NO: 6) sin alterar la actividad biológica, mientras que se requiere de un residuo de aminoácido "esencial" para la actividad biológica. Un aminoácido "esencial" incluye un residuo que puede ser alterado a partir de la secuencia de tipo silvestre de una proteína LMR (es decir, la SEQ ID NO: 3 o la SEQ ID NO: 6) para modificar o eliminar su actividad biológica, sin interferir en su enlazamiento con otras proteínas, ligandos o sitios de enlazamiento.
- 10 En una realización preferida, una proteína aislada LMR o una porción de la misma, aunque mutada a partir de la SEQ ID NO: 3 o la SEQ ID NO: 6, todavía es suficientemente similar a una secuencia de aminoácidos correspondiente a la SEQ ID NO: 3 o la SEQ ID NO: 6 de tal manera que la proteína o una porción de la misma mantiene la capacidad para participar en el metabolismo de los compuestos necesarios para los lípidos que se almacenan en la semilla, ácidos grasos y almidón en las semillas de la planta. Los aminoácidos que se conservan junto con las proteínas LMR de diferentes especies pueden ser esenciales para la actividad y, por tanto, probablemente no serían objetivos para alteración, a menos que uno desee aumentar, reducir o alterar la actividad de una proteína LMR. En consecuencia, un aspecto de la invención se refiere a moléculas de ácido nucleico que codifican proteínas LMR que contienen cambios en los residuos de aminoácidos que no son esenciales para la actividad. En una realización, la molécula aislada de ácido nucleico codifica una proteína que incluye una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 85%, 95% o 98% de identidad con la secuencia de aminoácidos de la SEQ ID NO: 3 o la SEQ ID NO: 6, y que puede participar en el metabolismo de un compuesto que se almacena en la semilla.
- 15 20 25 30 35 40 45 50 55 60
- En una realización preferida de la presente invención, los cambios realizados en uno o más residuos de aminoácidos predichos no esenciales son sustituciones conservadoras de aminoácidos. Una "sustitución conservadora de aminoácidos" es una en la cual se reemplaza el residuo de aminoácido con un residuo de aminoácido que tiene una cadena lateral similar. Las familias de los residuos de aminoácidos que tienen cadenas laterales similares han sido definidas en el arte. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales ramificadas en beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Por lo tanto, un residuo de aminoácido no esencial predicho en una proteína LMR es preferiblemente reemplazado con otro residuo de aminoácido de la misma familia de cadena lateral.
- La persona capacitada en la técnica se dará cuenta también que se pueden introducir cambios por la mutación en la secuencia de aminoácidos de una proteína LMR que cambia o elimina la actividad biológica de la proteína LMR sin afectar su capacidad de interacción con otras proteínas en la célula. Esta estrategia se conoce también como "mutación negativa dominante" en el arte y proporciona un método efectivo para la modulación de la actividad y la función de las proteínas o a las proteína reguladoras y enzimáticas, mediante lo cual se valora la actividad de la proteína endógena funcional por la forma no funcional de la proteína que compiten por interacciones con otras proteínas, ligandos o sitios de enlazamiento necesarios para su función. Los residuos de aminoácidos que están muy conservados o sólo semiconservados entre la proteína LMR de diferentes especies puede ser "esencial" para la actividad y por lo tanto probablemente serían los objetivos para tales alteraciones. En una realización preferida, una proteína LMR aislada o una porción de la misma es suficientemente similar a una secuencia de aminoácidos correspondiente a la SEQ ID NO: 3 o a la SEQ ID NO: 6 de tal manera que la proteína o una porción de la misma mantiene la capacidad de participar en el metabolismo de compuestos necesarios para almacenamiento en la semilla de lípidos, ácidos grasos y almidón en las semillas de la planta por competición con la secuencia de aminoácidos correspondiente a la SEQ ID NO: 3 o a la SEQ ID NO: 6 por interacciones con otras proteínas, ligandos o sitios de enlazamiento.
- Dos mutaciones que puede o no ser mutaciones negativas dominantes, pero que están incluidas dentro de la presente invención, son aquellas mutaciones encontradas en las plantas *wri1-1* y *wri1-2*. La planta *wri1-1* contiene una sola mutación en la posición del nucleótido 3197 sobre la secuencia BAC T12E18 suministrada en la SEQ ID NO: 7 y mostrada en la Figura 6. Esta mutación es una conversión de G por A en el sitio de empalme del intrón que provoca un empalme incompleto o alterado en comparación con la secuencia de ácido nucleico *wri1* de tipo silvestre. Una forma truncada y no activa de la proteína *wri1* es el resultado de la mutación de *wri1-1* ya que el empalme alterado provoca que el marco de lectura abierto se extienda dentro del intrón de tipo silvestre hasta un codón de detención. Por lo tanto, la presente invención abarca polinucleótidos LMR que contienen una mutación en un sitio correspondiente a la posición del nucleótido 3197 sobre la secuencia BAC T12E18 suministrada en la SEQ ID NO: 7.
- Existen una cantidad de mecanismos por medio de los cuales la mutación de un ácido nucleico o proteína LMR de la invención puede afectar directamente el rendimiento, la producción y / o la eficiencia de la producción de un

compuesto que se almacena en la semilla de una especie de planta productora de aceite. Por ejemplo, se puede incrementar una proteína LMR involucrada en el transporte de un compuesto que se almacena en la semilla en la célula en cantidad o en actividad de tal manera que se importen mayores cantidades de los precursores del compuesto que se almacena en la semilla desde el espacio exterior de la célula y se reparte dentro del flujo biosintético o se deposita en los compuestos que se almacenan en la semilla. El rendimiento, producción y/o la eficiencia de la producción de un compuesto que se almacena en la semilla pueden ser incrementados también cuando un incremento en una proteína LMR involucrada en el transporte de ese compuesto trae como resultado la distribución del compuesto a un compartimiento celular diferente de la planta. De manera similar, se puede incrementar una proteína LMR involucrada en la importación de nutrientes dentro de las células o en subcompartimientos celulares necesarios para la biosíntesis de uno o más compuestos que se almacenan en la semilla en cantidad o en actividad de tal manera que estos precursores, cofactores, o compuestos intermedios incrementan su concentración dentro de la célula o los plástidos o dentro de los compartimentos de almacenamiento de la semilla. Los compartimentos de almacenamiento de la semilla incluyen, pero no se limitan a, los cuerpos oleosos, microsomas, endospermo de la semilla, amiloplastos y gránulos de almacenamiento de proteína. Además, un aumento en el rendimiento, la producción y / o la eficiencia de la producción de un compuesto que se almacena en la semilla puede lograrse afectando la actividad de una o más proteínas LMR directa o indirectamente implicadas en la degradación de los compuestos que se almacenan en la semilla. Las proporciones relativas de los diferentes compuestos que se almacenan en la semilla (lípidos, almidón y proteínas) se pueden alterar para cambiar o mejorar las extracciones y el procesamiento de estos compuestos de las semillas o para hacerlos más adecuados para otros usos (por ejemplo en piensos para animales, alimento para humanos o para usos industriales).

En algunos casos especiales, la expresión alterada de una proteína de LMR, o de un análogo, ortólogo o parálogo de la misma, o la expresión de una versión alterada de una proteína de LMR, o de un análogo, ortólogo o parálogo de la misma, puede conducir al aumento de la producción de los compuestos que se almacenan en las plantas en las semillas de las plantas transgénicas sin que se manifieste por sí misma en un mayor rendimiento (rendimiento por hectárea) del compuesto deseado que se almacena en la semilla. Esto puede ser debido a la mayor degradación del compuesto que se almacena en la semilla que pueden ser tanto los lípidos, proteínas como almidón. Es posible utilizar el mayor flujo de productos intermedios a través de las rutas de degradación para la producción de otros compuestos de interés. Por ejemplo, el mayor flujo de ácidos grasos para la oxidación en beta del ácido graso se puede usar para aumentar la producción de polihidroxialcanoatos en peroxisomas de plantas transgénicas.

La modificación del balance de lípidos y / o de ácidos grasos en una semilla puede tener un profundo efecto sobre la composición lipídica de la membrana de las células de la semilla. Ya que cada tipo de lípido tiene diferentes propiedades físicas, una alteración en la composición lipídica de una membrana puede alterar significativamente la fluidez de la membrana. Los cambios en la fluidez de la membrana pueden afectar el transporte de moléculas a través de la membrana, así como la integridad de la célula, los cuales tienen un efecto profundo en la producción de "productos químicos finos" procedentes de las plantas. Estos cambios también pueden influir en otras características como la tolerancia hacia las condiciones de estrés abiótico y biótico.

El término "productos químicos finos" está reconocido por ley e incluye a las moléculas producidas por un organismo que tiene aplicaciones en diferentes industrias, tales como, pero sin limitarse a, la industria farmacéutica, agricultura y de cosméticos. Tales compuestos incluyen lípidos, ácidos grasos, cofactores y enzimas, aminoácidos, tanto proteinogénicos como no proteinogénicos, bases de purina y de pirimidina, nucleósidos, y ácidos nucleicos (por ejemplo como se describe en Kuninaka, A. (1996) Nucleic acids and related compounds, páginas 561 - 612, en Biotechnology Vol. 6, Rehm et al., eds VCH: Weinheim, y las referencias contenidas allí), lípidos, ácidos grasos tanto saturados como poliinsaturados (por ejemplo, ácido araquidónico), dioles (por ejemplo, de propano diol, y butano diol), carbohidratos (por ejemplo, ácido hialurónico y trehalosa), compuestos aromáticos (por ejemplo, aminas aromáticas, vanilina, e índigo), vitaminas y cofactores (como se describe en la Enciclopedia de Química Industrial de Ullmann, vol. A27, Vitaminas, páginas 443 - 613 (1996), VCH: Weinheim y las referencias citadas allí, y Ong, AS, Niki, E. & Packer, L. (1995) Nutrition, Lipids, Health and Disease, Proceedings of the UNESCO / Confederation of Scientific and Technological Associations in Malaysia, y la Society for Free Radical Research, Asia, celebrada del 1 al 3 de septiembre de 1994 en Penang, Malasia, AOCS Press (1995), enzimas, y todos los otros compuestos químicos descritos en Gutcho (1983) Chemicals by Fermentation. Noyes Data Corporation, y las referencias citadas allí).

Además de los ácidos nucleicos para el LMR descritos anteriormente, otro aspecto de la invención se refiere a secuencias aisladas de ácido nucleico que sean antisentido a las mismas. Una secuencia de ácido nucleico "antisentido" comprende una secuencia que es complementaria a una secuencia de ácido nucleico "sentido" que codifica una proteína, por ejemplo, complementaria a la cadena de codificación de una molécula de ADNc bicatenaria o complementaria a una secuencia de ARNm. El ácido nucleico antisentido puede ser complementario a un ácido nucleico LMR completo o únicamente a una porción del mismo. En una realización, una secuencia de ácido nucleico antisentido es antisentido para una "región de codificación" de la cadena de codificación de una secuencia de ácido nucleico que codifica una proteína LMR. Los términos "región de codificación" y "región no codificante" se definen más arriba. Una secuencia de ácido nucleico antisentido puede construirse utilizando síntesis química y reacciones enzimáticas de enlazamiento utilizando procedimientos conocidos por aquellos capacitados en el arte. Los ácidos nucleicos antisentido pueden utilizarse en un método para reducir la expresión de genes endógenos para

5 LMR con el fin de manipular el metabolismo del compuesto que se almacena en la semilla. Los constructos de ácido nucleico antisentido en general conducen a la formación de estructuras secundarias de ARN que son reconocidas y degradadas por las ARNasas celulares y por lo tanto conducen a una disminución de la abundancia de ARNm del gen objetivo, es decir, un gen endógeno. En consecuencia, los ácidos nucleicos antisentido que son complementarios a la totalidad o sólo a una porción de la secuencia de ácido nucleico LMR se muestran como la SEQ NO: 1, la SEQ NO: 2, la SEQ NO: 4 o la SEQ ID NO 5 se puede utilizar para manipular el metabolismo del compuesto que se almacena en la semilla.

10 Otro método para disminuir los niveles de transcripción de ARNm de los genes objetivo se llama "interferencia del ARN" o "interferencia de ARN bicatenario (iARNbc). Los expertos en la técnica están familiarizados con iARNbc como una alternativa efectiva a las técnicas antisentido. A fin de disminuir los niveles de transcripción del ARNm del gen objetivo, es decir, un ácido nucleico LMR, se hace una construcción que contenga una porción de la secuencia que codifica LMR en orientación sentido y antisentido, donde las secuencias sentido y antisentido se separan por una región enlazadora. La expresión de tal constructo de iARNbc en las células conduce a la formación de estructuras secundarias estables de ARN bicatenario, lo que a su vez conduce al reconocimiento y la destrucción de las estructuras y las transcripciones de LMR a través de los mecanismos celulares. Por consiguiente, las secuencias de ácidos nucleicos que son complementarias a la totalidad o sólo a una porción de una secuencia de ácido nucleico para LMR suministrada bajo la SEQ ID NO: 1, la SEQ ID NO: 2, la SEQ ID NO: 4 o la SEQ ID NO: 5, se puede utilizar para manipular el metabolismo del compuesto que se almacena en la semilla por medio del uso de interferencia de ARN.

15 20 Además de ácidos nucleicos antisentido para LMR, también se describen en este documento vectores de expresión recombinante y cassetes de expresión que comprenden los ácidos nucleicos para LMR descritos aquí y las células huésped dentro de las cuales se han introducido tales vectores y cassetes. Como se usa aquí, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico con el cual se ha enlazado. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN bicatenario circular dentro del cual se pueden ligar 25 segmentos adicionales de ADN. Otro tipo de vector es un vector viral, en donde los segmentos adicionales de ADN se pueden ligar en el genoma viral. Algunos vectores son capaces de replicación autónoma en una célula huésped en la cual se introducen (por ejemplo, vectores bacterianos que tiene un origen bacteriano de replicación y vectores de mamífero episomales). Otros vectores (por ejemplo, vectores de mamífero no episomales) se integran en el 30 genoma de una célula huésped después de la introducción en la célula huésped, y con ello se replican junto con el genoma del huésped. Además, ciertos vectores son capaces de dirigir la expresión de genes a los cuales están operativamente enlazados. Tales vectores se denominan aquí como "vectores de expresión". En general, los vectores de expresión de utilidad en las técnicas de ADN recombinante son a menudo en forma de plásmidos. En la 35 presente especificación, "plásmido" y "vector" se puede utilizar indistintamente yaque el plásmido es la forma más comúnmente utilizada de vector. Sin embargo, la invención pretende incluir tales otras formas de vectores de expresión, tales como vectores virales (por ejemplo, retrovirus de replicación defectuosa, adenovirus y virus adenovirus- 40 asociados), que sirven para funciones equivalentes.

45 El vector de expresión recombinante de la invención comprende un ácido nucleico de la invención en una forma adecuada para expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinantes incluyen una o más secuencias reguladoras, seleccionadas con base en las células huésped que se utilizan para la expresión, que están operativamente enlazadas a la secuencia de ácido nucleico que es expresada. Dentro de un vector de expresión recombinante, "operativamente enlazado" pretende significar que la secuencia de nucleótidos de interés está enlazada a la(s) secuencia(s) reguladora(s) en una forma que permite la expresión de la secuencia de nucleótidos (por ejemplo, en un sistema de transcripción / traducción *in vitro* o en una célula huésped cuando se introduce el vector en la célula huésped). El término "secuencia reguladora" pretende incluir promotores, reforzadores y otros elementos de control de la expresión (por ejemplo, señales de poliadenilación). Tales 50 secuencias reguladoras son descritas, por ejemplo, en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990) o visite: Gruber y Crosby, en: Methods in Plant Molecular Biology and Biotechnology, eds. Glick y Thompson, capítulo 7, 89 - 108, CRC Press: Boca Raton, Florida, incluidas las referencias citadas allí. Las secuencias reguladoras incluyen a aquellas que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huésped y aquellas que dirigen la expresión de la secuencia de nucleótidos sólo en ciertas células huésped o bajo ciertas condiciones. Se apreciará por parte de aquellos capacitados en el arte que el diseño del vector de expresión puede depender de factores tales como la escogencia de la célula huésped que va a ser transformada, el nivel de expresión del polipéptido deseado, etc. Los vectores de expresión de la invención pueden ser introducidos en células huésped para producir con ello 55 polipéptidos o péptidos, incluyendo polipéptidos o péptidos de fusión, codificados por los ácidos nucleicos como se describe aquí (por ejemplo, las proteínas LMR, las formas mutantes de las proteínas LMR, polipéptidos de fusión, etc.).

60 Tal como se usa aquí, el término "proteína recombinante LMR" se refiere a una proteína LMR expresada por un vector de expresión recombinante que comprende un ácido nucleico para el LMR. La célula huésped se puede derivar de bacterias, levaduras o material vegetal. En una realización, la célula huésped es capaz de almacenar compuestos que se almacenan en la semilla. En una realización adicional, la célula huésped es capaz de almacenar un compuesto que se almacena en la semilla de una manera que facilita la recolección del compuesto deseado de la

célula. Preferiblemente, la célula huésped es una célula del tejido de almacenamiento tal como una célula epidérmica o una célula de la semilla. La célula huésped como se describió anteriormente se puede encontrar o localizar en un tejido de la planta, en un órgano de la planta o en una planta entera. Debe entenderse que el término "célula huésped" se refiere no solamente a la célula objetivo particular, sino también a la progenie o a la progenie potencial de dicha célula. Debido a que puede presentarse cierta modificación en generaciones sucesivas debido ya sea a una mutación o a influencias ambientales, tal progenie puede no ser idéntica a la célula progenitora, pero están incluidas aún dentro del alcance del término utilizado aquí.

Los vectores de expresión adecuados para uso en plantas, y por lo tanto adecuados para su uso en la presente invención, son bien conocidos por aquellos capacitados en la técnica. Varios ejemplos de vectores de expresión adecuados pueden encontrarse en las siguientes referencias: *Plant Molecular Biology and Biotechnology*, Capítulos 6/7, S. 71 - 119. CRC Press, Boca Raton, Florida, 1993; F. F. White, *Vectors for Nucleic acid sequence Transfer in Higher Plants in transgenic Plants*, Vol. 1, engineering and Utilization, p. 44 - 38, eds.: Kung und R. Wu, Academic Press, 1993; B. Jenes et al., *Techniques for Nucleic acid sequence Transfer in Transgenic Plants*, Vol. 1, Engineering and Utilization, p. 128 - 143, eds.: Kung und R. Wu, Academic Press, 1993; Potrykus, 1991. Annu. Rev. Plant Physiol. Plant Molec. Biol. 42: 205 - 225.

La presente invención incluye adicionalmente métodos de expresión de las proteínas LMR descritas aquí. Como se describe adicionalmente en detalle más adelante, se introduce un ácido nucleico para LMR de tipo silvestre o mutante en una célula huésped y o bien se mantiene en un plásmido separado o se integra en el genoma de una célula huésped. La proteína expresada LMR se puede aislar a partir del medio en el que se cultiva la célula huésped o a partir de la célula huésped en sí misma. Se describen diferentes métodos no limitantes de expresión y aislamiento de proteínas LMR en los Ejemplos más adelante.

Debe entenderse que las secuencias de ácido nucleico del LMR descritas aquí pueden ser regiones reguladoras, tales como promotores, "reforzadores" o intrones que contienen un elemento regulador. Se proporcina aquí un método para utilizar tales regiones reguladoras para modular el metabolismo del compuesto que se almacena en la semilla. Cuando los ácidos nucleicos del LMR son regiones del regulador, se pueden utilizar los ácidos nucleicos del LMR en cassetes de expresión. Un casete de expresión que contiene un ácido nucleico del LMR está operativamente enlazado a la secuencia de codificación de una secuencia de ácido nucleico de interés de tal manera que la región reguladora del LMR sea capaz de controlar la expresión del producto codificado por la secuencia de ácido nucleico de interés. La naturaleza específica de la semilla del fenotipo mutante *wri1* indica que el promotor del LMR es específico de la semilla. Por lo tanto, se puede utilizar un promotor del LMR para la expresión específica de la semilla para otras proteínas con funciones no directamente relacionadas con la proteína LMR. Por ejemplo, se puede utilizar un promotor del LMR específico de la semilla para expresar proteínas que puedan controlar eventos en el desarrollo de la semilla.

Las "reforzadoras" son secuencias reguladoras que actúan sobre cis que provocan una mayor expresión de los genes situados a una distancia, preferentemente dentro de 5 kb del elemento reforzador. Los reforzadores de este tipo pueden ser utilizados para sobreexpresar otros genes de interés con el fin de modular el metabolismo del compuesto que se almacena en la semilla. Los intrones pueden contener elementos reguladores que afectan el nivel de expresión, de empalme y de estabilidad de la transcripción de ARN. Se descubrió que el gen *WRI1* contiene un intrón inusualmente grande de 1650 pb y que algunos productos diferenciales de empalme de ARN fueron observados durante análisis por RT-PCR. La manipulación del tamaño de intrón y de las secuencias se puede utilizar para modular el nivel de expresión y la abundancia relativa de los productos de empalme particulares que tienen la actividad biológica deseada y conducen a cambios deseados en el metabolismo del compuesto que se almacena en la semilla. Las secuencias reguladoras, tales como promotores, reforzadores y elementos reguladores contenidos en los intrones se puede encontrar en la SEQ ID NO: 7 mostrada en la Figura 6.

Los ácidos nucleicos para el LMR, las secuencias de aminoácidos, y análogos y fragmentos de los mismos, de la presente invención tienen una variedad de usos. Las moléculas de ácido nucleico que codifican la proteína LMR de la invención pueden ser utilizadas en la modificación por ingeniería genética de una amplia variedad de plantas para hacerlas productoras mejores o más eficientes de uno o más compuestos que se almacenan en la semilla. Las proteínas LMR descritas aquí son capaces, por ejemplo, de llevar a cabo una función implicada en el metabolismo de compuestos necesarios para la biosíntesis de lípidos o de ácidos grasos, y puede estar directamente implicadas en la repartición de los recursos vegetales a los diferentes compuestos que se almacenan en la semilla, tales como ácidos grasos, lípidos, almidón o proteínas. Por lo tanto las proteínas LMR se pueden utilizar para manipular las proporciones relativas de lípidos, almidón y proteínas en las semillas. Estas proteínas LMR también puede ser utilizadas para aumentar la proporción relativa de cualquier compuesto que se almacena en la semilla, tal como el almacenamiento de lípido, sin disminuir la abundancia de otro compuesto que se almacena en la semilla.

De conformidad con la presente invención, se produce una planta transgénica que contiene un ácido nucleico para el LMR, que consta de los elementos descritos anteriormente, lo que resulta en la modificación del nivel de un compuesto que se almacena en la semilla en la planta transgénica. Aquí se describe un método para producir una planta transgénica con un nivel modificado de un compuesto que se almacena en la semilla que comprende, la transformación de una célula de una planta con un vector de expresión que contiene un ácido nucleico para el LMR

- 5 y la regeneración de una planta con un nivel modificado del compuesto que se almacena en la semilla de la célula de la planta. Los métodos de la presente invención implican la introducción de un ácido nucleico mutante o de tipo silvestre para el LMR en una célula de la planta, ya sea mantenido en un plásmido separado, codificado por un virus genéticamente modificado que replica en una planta o integrado en el genoma de la célula huésped. Si se integra dentro del genoma, tal integración puede ser aleatoria, o puede tener lugar por recombinación de tal manera que la secuencia nativa de ácido nucleico es reemplazada por la copia introducida, o por medio del uso de un ácido nucleico para el LMR en posición trans de tal manera que el ácido nucleico está funcionalmente enlazado a una unidad de expresión funcional que contiene al menos una secuencia que facilita la expresión de un ácido nucleico y una secuencia que facilita la poliadenilación de un ácido nucleico funcionalmente transcrita.
- 10 10 La presente invención enseña una composición y métodos para la modificación del nivel de un compuesto que se almacena en la semilla en una planta o semilla, y preferiblemente un aceite vegetal o de semilla, en comparación con una variedad de tipo silvestre de esa planta. Como se usa aquí, "modificación" incluye un aumento en el nivel de un compuesto que se almacena en la semilla, una disminución en el nivel de un compuesto que se almacena en la semilla, un aumento en el transporte de un compuesto que se almacena en la semilla a microcuerpos lipídicos y gránulos donde se almacena proteína. Por ejemplo, las composiciones y métodos de la presente invención se puede utilizar para aumentar o disminuir el nivel de un lípido en un aceite de la semilla, o para aumentar o disminuir el nivel de un ácido graso en un aceite de la semilla, para aumentar o disminuir el nivel de un almidón en una semilla o en una planta, o para aumentar o disminuir el nivel de una proteína que se almacena en la semilla en una semilla o en una planta.
- 15 20 En una realización, un nivel más alto de lo normal de un compuesto que se almacena en la semilla se produce en una planta transgénica. El término "un nivel más alto de lo normal" se refiere a un nivel mayor que el encontrado en una variedad de tipo silvestre de la planta transgénica. En un caso especial, la mayor síntesis y acumulación de lípidos que se almacenan en la semilla provocada por la expresión alterada de un LMR homólogo o la expresión de una versión alterada de un LMR pueden ser utilizados para aumentar el rendimiento de ácidos grasos poliinsaturados como el ácido araquidónico (ARA) y / o el ácido eicosapentaenoico (EPA) y / o el ácido docosahexaenoico (DHA) en plantas transgénicas que expresan genes de musgos y de algas para la producción de estos neutraceuticalos. Finalmente, el término "modificación" incluye la introducción de un nuevo compuesto que se almacena en la semilla en una planta o semilla.
- 25 30 Los ácidos nucleicos para el LMR pueden utilizarse para modificar la proporción relativa de los compuestos que se almacenan en la planta en las semillas de plantas transgénicas. Especialmente en el caso de las plantas oleaginosas la relación de los lípidos que se almacenan puede ser modificada en relación con el contenido de proteína de la semilla y de almidón de la semilla. Los ácidos nucleicos para el LMR pueden ser utilizados para modificar las proporciones relativas de los compuestos que se almacenan en la semilla y también de otros compuestos que se acumulan en las semillas, influyendo en la partición del fotosíntesis en las respectivas rutas bioquímicas. Las rutas implicadas son responsables de la producción de ácidos grasos de los lípidos de la membrana y para almacenamiento, de aminoácidos para la síntesis de proteínas, de carbohidrato para la síntesis de almidón, o de precursores para las rutas que conducen a la síntesis de una gran variedad de compuestos como isoprenoides, lignina, celulosa, hemicelulosa, glucanos y pectina.
- 35 40 La modificación del nivel de un compuesto que se almacena en la semilla se puede lograr mediante una modificación del metabolismo de ese compuesto. El término "metabolismo" es reconocido en la técnica e incluye la totalidad de las reacciones bioquímicas que tienen lugar en un organismo. El "metabolismo" de un compuesto particular, por lo tanto, (por ejemplo, el metabolismo de un ácido graso) comprende la ruta biosintética completa, la modificación y degradación en la célula relacionada con este compuesto. Los términos "biosíntesis" y "ruta biosintética" son reconocidos en la técnica e incluyen la síntesis de un compuesto, preferiblemente de un compuesto orgánico, por una célula a partir de un compuesto intermedio en lo que puede ser un proceso de múltiples etapas y altamente regulado. Los términos "degradación" y "ruta de degradación" son reconocidos en la técnica e incluyen la degradación de un compuesto, preferiblemente de un compuesto orgánico, por una célula hasta los productos de degradación (moléculas más o menos complejas) en lo que puede ser un proceso de múltiples etapas y altamente regulado.
- 45 50 También se incluyen en este documento composiciones y métodos para la modificación de la eficiencia de producción o de un compuesto que se almacena en la semilla. El término "eficiencia de producción" incluye el tiempo necesario para que se logre un nivel particular de producción (por ejemplo, el tiempo que tarda la célula para alcanzar una tasa particular de producción de un compuesto que se almacena en la semilla). El término "rendimiento" es reconocido en la técnica, e incluye la cantidad de producto (es decir, del compuesto que se almacena en la semilla o las semillas) obtenido por unidad de superficie (es decir, hectárea o acre). Esto generalmente se describe, por ejemplo, como toneladas de producto por hectárea. Al aumentar el rendimiento o la producción del compuesto, la cantidad de moléculas recuperadas, o de moléculas útiles recuperadas, de ese compuesto a partir de una superficie dada de terreno cultivada se incrementa. Una mejora en la eficiencia de producción, o un aumento o disminución en la producción de un compuesto que se almacena en la semilla, puede ser debido a un efecto directo de la manipulación de un ácido nucleico de la invención, o puede ser debido a un efecto indirecto de tal manipulación.

La presente invención permite la producción de una variedad genéticamente pura de plantas que portan semillas que tienen niveles modificados de un compuesto que se almacena en la semilla en comparación con una variedad de tipo silvestre. El término "variedad" se refiere a un grupo de plantas dentro de una especie que comparten caracteres constantes que los separan de la forma típica y de otras variedades posibles dentro de esa especie. Si bien poseen al menos un rasgo distintivo, una variedad se caracteriza también por una cierta variación entre individuos dentro de la variedad, basadas principalmente en la segregación mendeliana de caracteres entre la progenie de generaciones sucesivas. Una variedad se considera una "línea genéticamente pura" para un rasgo particular si es genéticamente homocigota para ese rasgo, en la medida en que, cuando la variedad de la línea genéticamente pura es autogama, una cantidad significativa de segregación independiente del rasgo entre la progenie no es observado. En la presente invención, el rasgo surge de la expresión transgénica de una sola secuencia de ADN introducida en una variedad de planta.

En la producción de una planta transgénica de la presente invención, un vector expresión que contiene la secuencia introducida ácido nucleico se inserta en protoplastos, en tejidos intactos, tales como los embriones inmaduros y meristemos, en cultivos de callo o en células aisladas. Preferiblemente, se insertan los vectores de expresión en tejidos intactos. Los métodos generales de cultivo de tejidos vegetales son proporcionados, por ejemplo, por Miki et al., *Procedures for Introducing Foreign DNA into Plants* in *Methods in Plant Molecular Biology and Biotechnology*, p. 67 - 88, CRC Press, 1993.

Los métodos para la introducción de un vector de expresión en el tejido de una planta incluyen la infección directa o el cultivo conjunto del tejido de la planta con *Agrobacterium tumefaciens* (Horsch et al. 1985, *Science* 227: 1229). Preferiblemente, se utiliza un plásmido Ti desarmado como vector para una secuencia de ADN foraneo. El más preferido es el uso de *Agrobacterium* con explantes embrionarios divididos como lo describen Malone-Schoneberg et al. (1994, *Plant Science* 103: 199 - 207). Aunque *Agrobacterium* es un vector preferido, se pueden utilizar otros tipos de vectores para la transformación por medio de procedimientos tales como la transferencia directa de genes, transformación de protoplastos *in vitro*, transformación de plantas mediada por virus y la transformación mediada por liposomas. Transformación de plantas también se puede realizar utilizando bombardeo de partículas, absorción de ADN mediada por polietilén glicol o la técnica de la Fibra de Carburo de Silicio (Freeling y Walbot, *The Maize Handbook*, Springer Verlag, Nueva York 1993).

La naturaleza específica de semillas del fenotipo *wri1* indica que una proteína LMR es activa en las semillas en desarrollo, y que la expresión ectópica de una proteína LMR en otras partes de las plantas transgénicas, por ejemplo hojas, raíces o tubérculos, se puede utilizar en métodos para modular el metabolismo de carbohidratos primarios en estos órganos de la planta. Por lo tanto, la transformación de una planta con un ácido nucleico para el LMR como se describe aquí puede facilitar la producción de compuestos de almacenamiento como almidón, proteínas y lípidos, en órganos de la planta que habitualmente no se acumulan estos compuestos. Un ejemplo preferido es la acumulación de lípidos de almacenamiento en hojas, raíces o tubérculos. En otro ejemplo, la expresión ectópica de una proteína LMR conduce a una mayor biosíntesis de ácidos grasos que a su vez conduce a mayores tasas de degradación de los ácidos grasos. Tanto la mayor tasa de biosíntesis de ácidos grasos como la tasa de aumento de degradación de ácidos grasos en las hojas o en las raíces, por ejemplo, pueden ser utilizadas para aumentar la producción y acumulación de polihidroxialcanoatos (por ejemplo, PHB y PHA) en células transgénicas o en plantas. Los polihidroxialcanoatos son compuestos valiosos que pueden ser utilizados como materiales plásticos biodegradables a partir de recursos renovables.

Otro aspecto de la invención se refiere a un método para producir un compuesto que se almacena en la semilla. Este método implica el cultivo de células vegetales, tejidos, órganos o plantas enteras que contienen un vector que dirige la expresión de una molécula de ácido nucleico para LMR de la invención, de tal manera que se produce un compuesto que se almacena en la semilla. En una realización preferida, este método incluye además la etapa de obtención de una célula huésped que contiene dicho vector, en la cual se transforma la célula con el vector que dirige la expresión de un ácido nucleico para el LMR. En una realización particularmente preferida, la célula huésped es de una planta de especies productoras de aceite.

Los ácidos nucleicos para el LMR y las secuencias de aminoácidos también pueden ser utilizados como una herramienta de investigación en métodos que apuntan a la identificación de otras proteínas reguladoras que interactúan con las proteínas LMR. La identificación de los genes para LMR por medio del mapeo y clonación del gen *WRI1* aquí descrito hace posible la utilización de un gen para LMR en experimentos que identificarán proteínas que participan en las mismas rutas de regulación o similares. Por ejemplo, un gen para LMR puede ser utilizado en las selecciones de doble híbrido de levadura para identificar otras proteínas que interactúan con la proteína LMR. En otro ejemplo, una proteína LMR puede ser utilizada en experimentos de inmunoprecipitación que hacen uso de anticuerpos para precipitar conjuntamente una proteína LMR y las proteínas que interactúan con ella, conduciendo así a la identificación de las parejas de proteína que interactúan. Los anticuerpos pueden ser ya sea específicos para la proteína LMR o pueden ser específicos para epitopos que se han unido a la proteína LMR por medio de técnicas de clonación estándar. Estos experimentos específicos sólo se pueden realizar debido al descubrimiento de la identidad y la función de un gen para LMR como se describe aquí.

El promotor de un gen para LMR también puede ser utilizado como una herramienta de investigación en métodos que apuntan a la identificación de otras proteínas reguladoras que interactúan con un promotor del LMR. La identificación de los genes para LMR por medio del mapeo y clonación del gen *wri1* divulgado aquí hace posible la utilización de un gen para LMR y su promotor en los experimentos que identificarán proteínas adicionales que participan en las mismas rutas reguladoras o similares. La longitud de la secuencia del promotor de la planta puede variar considerablemente, pero es apropiado decir que la mayoría de los elementos activos en promotores de plantas se encuentran dentro de 3000 bases de nucleótidos secuencia arriba del gen. Por lo tanto se define aquí un "promotor del LMR" como las 3000 bases de nucleótidos de secuencia de la secuencia de ADN genómico 5' secuencia arriba de una región que codifica LMR. Se puede utilizar, por ejemplo, un promotor del LMR en selecciones de un solo híbrido de levadura para identificar otras proteínas que interactúan con elementos en el promotor.

En otro ejemplo, se puede empalmar un promotor del LMR a un gen reportero, por ejemplo GUS o GFP. Tal constructo de fusión promotor del LMR - reportero puede ser transformado en células o en plantas transgénicas. Estas células o plantas transgénicas pueden ser utilizadas en experimentos de selección de mutantes para encontrar proteínas que se enlacen a un promotor del LMR. En tales experimentos, la expresión alterada del gen reportero sirve como una indicación de que el gen mutado codifica una proteína que puede interactuar con un promotor del LMR. La identidad del gen mutado se puede establecer mediante técnicas estándares, por ejemplo, mapeo molecular o secuenciación del ADN genómico que flanquea el ADN-T o al transposición que ha sido utilizado para crear la mutación. Las proteínas que interactúan con un promotor del LMR se puede utilizar para controlar la expresión y la actividad de la proteína LMR y por lo tanto también pueden ser utilizadas para controlar la acumulación, las proporciones relativas y el rendimiento total de los compuestos que se acumulan en las semillas de plantas transgénicas. Estos experimentos específicos sólo se pueden realizar con un promotor del LMR debido al descubrimiento de la identidad y la función de un gen para LMR como se describe aquí.

Debe entenderse que lo anterior se relaciona con realizaciones preferidas de la presente invención y que pueden hacerse numerosos cambios en ellas sin apartarse del alcance de la invención. La invención se ilustra adicionalmente mediante los siguientes ejemplos, que no deben interpretarse en modo alguno como la imposición de limitaciones sobre el alcance de la misma.

EJEMPLOS

Ejemplo 1

30 Clonación con base en el mapa del gen mutante *wri1*

La mutación *wri1-1* fue generada en un ambiente de Columbia-2 (Col-2) de *Arabidopsis*, y para asegurar que las plantas sigan siendo viables y tengan un fenotipo consistente, las plantas con *wri1-1* fueron cruzadas nuevamente con plantas col-2 de *Arabidopsis* de tipo silvestre por tres generaciones. Se mantuvo el mutante *wri1-1* homocigoto por medio de autofecundación sucesiva de los mutantes cruzados nuevamente 3 veces. Para el mapeo, se utilizaron las numerosas diferencias en la secuencia de ADN entre los ecotipos Landsberg y Columbia de *Arabidopsis* y se generó un mapeo inicial de la población por medio del cruce de *wri1-1* en un ambiente de col-2 de *Arabidopsis* con Landsberg crecta de tipo silvestre. Se autofecundó la generación F1 y se cosechó la semilla de la generación F2.

40 Para el mapeo inicial, se utilizaron cincuenta plantas F2 del cruce *wri1-1* X Landsberg (ler) que eran homocigotas para la mutación *wri1-1*. Se permitió el mapeo de esas cincuenta plantas con *wri1-1* entre los marcadores CAPS BGL1 y AFC1. Se generó una segunda población de mapeo que consta de 1000 plantas F2 en la misma forma que la primera excepto porque en lugar de seleccionar plantas *wri1-1* homocigotas, se incluyó la población total, incluidas aquellas plantas heterocigóticas para *wri1-1* y homocigotas de tipo silvestre. Esta población más amplia permitió un mapeo más fino de *wri1-1* sin tener que clasificar a través de la gran cantidad de plantas que se hubiera requerido utilizar que tuvieran una población de mapeo homocigoto puramente *wri1-1*. La mutación *wri1-1* provoca la germinación reducida y por lo tanto *wri1-1* plantas homocigotas están sub-representadas en la población F2. Como ejemplo, la obtención de los originales de cincuenta plantas homocigóticas se requiere una puntuación superior a cuatrocientas plantas F2.

50 Se anotó la población F2 para *wri1-1* / *wri1-1*, *wri1-1* / wt, wt / wt y por la presencia de ADN de Landsberg o de Columbia en los marcadores CAPS, CDC2BG y TSA1. Tres recombinantes resultaron de la puntuación, una entre *wri1-1* y CDC2BG y dos entre *wri1-1* y TSA1, permitiendo colocar el gen *wri1* entre CDC2BG y TSA1. Véase la Tabla 3 a continuación. En consecuencia, anotando tanto el genotipo de las plantas F2, así como si son Columbia o Landsberg para un marcador de PCR dado, se identificó una región que contenía la mutación *wri1-1*, y por lo tanto un locus genético para LMR.

Tabla 3

Marcador	Recombinantes/Puntaje total	Clon BAC	GenBank, Número de acceso
AFC1	14/1838	F4P12	AL132966
C DC2BG	1/1839	F24B22	AL132957
TSA1	2/1934	T5N23	AL132970
CF1	8/1934	T44C9	AL138650
BGL1	72/1934	F28O9	AL137080

Ejemplo 2Secuenciación de dos alelos *wri1* diferentes de *Arabidopsis wri1-1* y *wri1-2*

5 Cincuenta y cuatro genes putativos para LMR fueron identificados en el locus genético para LMR. Se secuenció el ADN genómico de tres de los genes putativos para LMR en plantas *Arabidopsis* Col con *wri1-1*, *wri1-2* y de tipo silvestre. Se diseñó la superposición de iniciadores para PCR como se describe más adelante y se utilizaron los mismos iniciadores para las reacciones de secuenciación. El objetivo era identificar cambios en la secuencia en los dos mutantes *wri1* comparado con la secuencia del gen de tipo silvestre ya que la falta de correspondencia encontrada por medio de la secuenciación del ADN genómico de los mutantes pueden conducir a la identificación del gen *wri1* verdadero.

10 Las alineaciones de secuencia indicaron que el locus de *wri1* contenía dos faltas de correspondencia en los mutantes *wri1-1* y *wri1-2*. Análisis adicionales indicaron que estas faltas de correspondencia estaban situadas en intrones del gen putativo para LMR.

15 Diseño del iniciador: se diseñaron 112 iniciadores como 56 pares de iniciadores para PCR y para la secuenciación de los productos de PCR. Se suministraron tres muestras de ADN genómico de *Arabidopsis*, es decir, de tipo silvestre, *wri1-1* y *wri1-2*. Se obtuvieron productos de PCR por BPS en las instalaciones de DNA Landmarks y se los secuenció allí mismo. Para cada una de las tres muestras de ADN, se obtuvieron 56 productos de PCR y se secuenciaron un número total de 168 productos de PCR de ambos extremos para producir 336 secuencias. Se utilizó un ADN polimerasa de alta fidelidad / corrección de errores para la reacción de PCR, porque se deseaba la búsqueda de mutaciones puntuales (mutantes EMS) en los mutantes. En total, se obtuvieron nueve secuencias ensambladas (tres genes para cada una de las tres muestras de ADN). Los iniciadores fueron los siguientes.

Iniciadores para proteína tipo Aintegumenta (gen = "T12E18.10")

Secuencia del oligo	Ubicación	SEQ ID NO:
ACT TGG CCT TTC TCGTTT C	2873	10
TGT TTG CCT TTC TTG TTC TG	3364	11
TGG ATT TGG GTT TTG TCT TA	3238	12
GCG TGA AAG CAG TTG AG	3644	13
CGA GCC TCA TCT TTG GGA	3324	14
TCT TCC TTT GTC ACT CTC TG	3858	15
CCG ACA CCA TCT TGA AT	3708	16
ATC CGA GCC TCC CAT CTT CC	4248	17
ATT TCG GAA TTT ACT ACC AA	4001	18
CAA AAC TCT CGC CTC AC	4629	19

Ejemplo 3Complementación del mutante *wri1* con cósmidos que contienen ADN genómico de tipo silvestre

El mapeo fino colocó al gen *WRI1* en una región de ciento cincuenta y cinco kilobases entre los marcadores de PCR TSA1 y CDC2BG sobre el cromosoma tres de *Arabidopsis thaliana*. Después del mapeo fino, se ensambló un conjunto de superposición de constructos de vector cósmido con el propósito de llevar a cabo experimentos de rescate de transformación del fenotipo *wrinkled 1* (semillas arrugadas que contenían niveles reducidos de los aceites que se almacenan en la semilla). Los cósmidos fueron generadas por Knut Meyer, utilizando el vector binario pBIC20 y el ADN genómico de *Arabidopsis thaliana* digerido parcialmente con Hind III (Meyer, K., Leube, M. P. y Grill, E. 1994. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* 264: 1452 - 1455).

Con el fin de lograr un rescate del cósmido del fenotipo de la semilla con *wri1-1*, se seleccionó una biblioteca de ADN genómico de *Arabidopsis*, contenida en el vector binario pBIC 20 (Meyer, K. et al. 1994 *Science* 264: 1452 - 1455), utilizando productos de PCR marcados en forma radioactiva. Los productos de PCR se utilizan como plantillas para las sondas se obtuvieron a través de la amplificación de ADN genómico en la región del cromosoma tres entre los marcadores de PCR Tsal y CDC2BG. Se obtuvieron los productos de PCR como moldes para las sondas a través de la amplificación de ADN genómico en la región del cromosoma tres entre los marcadores de PCR TSA1 y CDC2BG. Los productos PCR estaban en el rango de 0,5 kb a 1,5 kb. Los iniciadores para amplificación fueron escogidos con base en la posición del producto, con el objetivo de diseñar una sonda para cada segmento de veinte kilobases de ADN en la región definida por TSA1 y CDC2BG. La biblioteca fue sembrada en placa con una densidad de 3.000 - 4.000 colonias de bacterias por caja de Petri de 150 mm y se seleccionaron las colonia de nueve cajas con cada sonda de acuerdo con los procedimientos esbozados en Sambrook et al. (1989 *Molecular Cloning: A Laboratory Manual* 2nd Edición (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY)).

Una vez identificados los positivos putativos, se tomaron tapones de agar que contenía bacterias en la región de la señal sobre la autorradiografía y se eluyeron las bacterias en LB, que contenía tetraciclina (10 mg por litro). Se realizaron diluciones en serie y se sembraron las bacterias en placa con una densidad de una a 300 colonias por caja de Petri de 100 mm. Se realizaron levantamientos de colonias y se seleccionaron nuevamente las positivas putativas. Se recogieron las colonias correspondientes a las señales positivas y se las seleccionó nuevamente. Una vez que se confirmaron las colonias, se aisló el plásmido que contenía el ADN genómico y cortó con la enzima de restricción Hind III. Se compararon los patrones de restricción de diferentes colonias y utilizó para designar su posición dentro del contigo cósmido. La superposición entre cónigos cósmidos se determinó a través de transferencia tipo Southern e hibridación entre cósmidos en cónigos vecinos. Se utilizaron productos minipreparados de plásmido para transformar *Agrobacterium tumefaciens*, que fueron luego utilizados para transformar plantas *wrinkled 1-1* mediante inmersión. Las plantas transformadas fueron obtenidas por selección sobre placas de kanamicina que contenían Clavamox para suprimir el crecimiento del *Agrobacterium* transformado. Las plantas transformadas fueron luego transferidas al suelo y se tomó la puntuación de la semilla de las plantas adultas para el rescate del fenotipo *wrinkled 1*. Véase la Figura 2 para un diagrama de la estrategia de complementación del cósmido y superposición del clon cósmido.

Estos cósmidos de superposición fueron transformados individualmente dentro del mutante *wri1-1* de *Arabidopsis* descrito en el Ejemplo 1 y se anotaron los transformantes para la restauración del fenotipo de la semilla de tipo silvestre (semillas normales con los niveles de aceites que se almacenan en la semilla, correspondientes al tipo silvestre). Las plántulas transgénicas que eran resistentes al marcador de selección de antibiótico fueron cultivadas en el suelo hasta la madurez. Las semillas de silicuas individuales fueron cosechadas de plantas transformadas y se les dio puntaje visual para el rescate del fenotipo encogido y arrugado de *wri1-1*. Todos los cósmidos correspondientes a los clones BAC CDC2B y TSA1 fueron utilizados para transformar *wri1-1* y ninguno de ellos recuperó el fenotipo de semilla arrugada de la mutación *wri1-1*. Los cósmidos designados como AGL, AG2, AG3 y GA5 rescatados del fenotipo *wri1* y el contenido restaurado de aceite de semilla hasta niveles cercanos a los de tipo silvestre. (El clon cósmido GA4 nunca fue transformado, ya que se dispuso de los resultados de los otros clones). Para estos cósmidos las semillas transgénicas T2 segregadas en una relación de 3:1 de semillas de tamaño normal y lisas con respecto a las semillas engogidas y arrugadas, indicó que las semillas eran heterocigotas para el inserto de tipo silvestre del fragmento de rescate de ADN. Las semillas de plantas *wri1-1* rescatadas también fueron analizadas por medio de CG y se comparó con el tipo silvestre y *wri1-1* (véanse los datos más abajo). En todos los casos, las semillas que puntuaron visualmente como rescatadas tenían un nivel elevado de TAG. Las plantas de control que contenían ya sea un vector vacío, o insertos que contenían cósmidos para otras regiones del área cercana al locus de *WRI1* no mostraron ningún aumento en el contenido de TAG y eran indistinguiblea simple vista de *wri1-1*. La restauración del fenotipo de las semillas de tipo silvestre implicó que los cósmidos que se complementan codifiquen el LMR de tipo silvestre y por lo tanto conduzcan a la identificación del gen para LMR.

Determinación del contenido de lípidos de la semillas T2 de cósmidos que se complementan

Genotipo	µg de ácidos grasos totales por semilla
semillas de tipo silvestre Col-2	10,65
semillas mutantes <i>wri1-1</i>	2,1
semillas transgénicas AG1	10,7
semillas transgénicas AG2	10,63
semillas transgénicas AG5	9,44
semillas transgénicas 3P3 *	1,5
semillas transgénicas cdc 16-2 *	1,7
semillas transgénicas del vector vacío BPS *	3,1

* 3P3 y cdc 16-2 representan cósmidos de la región CDC2B sobre BAC F24B22 (véase la Figura 2) que se transformaron en las plantas mutantes *wri1-1* y BPS es el control transgénico del vector vacío.

5 Los tamaños de los insertos de los cósmidos de rescate son: AG1 16,05 kb, AG2 19,35 kb, AG3 15,1 kb, AG4 18 kb, AG5 14,6 kb. La Tabla 4 indica la ubicación exacta de los clones de cósmidos sobre la secuencia genómica BAC. Aunque no está exactamente representada en la Figura 2, la región de solapamiento entre los cósmidos rescate es de 8,5 kb, indicando que el gen *WR1* está contenida dentro de esa región de 8,5 kb.

Tabla 4

Nombre del cósmido	Ubicación sobre BAC F24B22 (pb)	Ubicación sobre BAC T12E18 (pb)*	Tamaño del cósmido (kb)
AG1		0 - 16.146	16,1
AG2		0 - 19.450	19,4
AG3	90.033 - 96.928	0 - 8.522	15,1
AG4	87.481 - 96.928	0 - 8.522	18
AG5	89.877 - 96.928	0 - 8.522	14,6

* La SEQ ID NO: 6 contiene la secuencia genómica parcial sobre BAC T12E18.

Los procedimientos observados en el experimento anterior son los siguientes:

10 PCR para la producción de sondas marcadas en forma radioactiva para identificar clones cósmidos La PCR para la producción de sondas marcadas en forma radioactiva se realizó con los iniciadores enlistados bajo las siguientes condiciones:

Desnaturalización	95 ° C durante 4 minutos
Fusión	95°C durante 0,5 minutos
Hibridación	55°C durante 0,5 minutos
Alargamiento	72°C durante 1,5 minutos
Repetir las últimas tres etapas nueve veces	
Alargamiento	72°C durante 10 minutos

Sostenimiento 4°C

Las temperaturas de hibridación de los iniciadores específicos se enumeran a continuación con la secuencia del iniciador, donde se diferencian del anterior protocolo.

Iniciadores de la PCR para la producción de sondas marcadas en forma radioactiva:

5	CDC2BG CGT CTG AAG GTC TGC ACC TAG TC SEQ ID NO: 20										
	CGC TAA GAT ACT TCC ACG TCA C SEQ ID NO: 21										
	Tipo proteína de enlazamiento de ARN, hibrida a 54 ° C TGA TGG CTG CGA TGA CT SEQ ID NO: 22										
10	(posición 59701 sobre el clon BAC F24B22) TTC CAC CAT AAC TGC GTC TA SEQ ID NO: 23										
	(posición 60862 sobre el clon BAC F24B22)										
	Proteína tipo Aintegumenta (gen = "T12E18.10"), hibrida a 54,3°C										
	ACT TGC CCT TTC TCG TTT C SEQ ID NO: 24										
15	(posición 2873 sobre el clon BAC T12E18) TGT TTG CCT TTC TTG TTC TG SEQ ID NO: 25										
	(posición 3363 sobre el clon BAC T12 E18)										
	Vecindad del nucleoide que contiene al ADN tipo enlazamiento, hibrida a 52,8 ° C TGA TTA CCT GGG CAC ATA SEQ ID NO: 26										
20	(Posición 24683 sobre el clon BAC T12E18) CAT AAG AAT GCO AGA CAA SEC ID NO: 27										
	(posición 25751 sobre el clon BAC T12E10)										
	Oligopeptido, tipo transportador, hibrida a 53,2 ° C CGC TGC CCT TTC AT SEQ ID NO: 28										
25	(Posición 27312 sobre el clon BAC T14E10) CGC AAT CCT CTC TCT ATG GT SEQ ID NO: 29										
	(posición 28856 sobre el clon BAC T14E10)										
	T14E10 de 52K, hibrida a 52° C TTA TAC GCC TGG GAG ATA SEQ ID NO: 30										
30	(Posición 52556 sobre el clon BAC T14E10) GGG CAY TCA AAC AGO TA SEQ ID NO: 31										
	(posición 53602 sobre el clon BAC clon T14E 10)										
	T14E10 de 60K, hibrida a 52 ° C TAT CTG GAT TCT TCT CT GGG SEQ ID NO: 32										
35	(posición 45037 sobre el clon BAC T14E10) ATC AGG GAG CTT GAT el TTT clon GG SEQ BAC ID NO: T14E10 33 (posición 47495 sobre el clon BAC T14E10)										
40	TSA1 TCT TGG TAG del CAT GAT TCT clon GAT CAG TC SEQ BAC ID NO: T14E10 34 (final CCT TTC CCG CTT ACA GAT GAT C) SEQ BAC ID NO: T14E10 35 (final del clon BAC T14E10)										

Medios

El medio para el cultivo de bacterias consistía de Medio Luria-Bertani (LB) (Sambrook et al. 1989 Molecular Cloning: A Laboratory Manual, 2^a Edición, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), con o sin agar (1,2% si

con), y el antibiótico apropiado (10 mg por litro de tetraciclina para *Escherichia coli* que contenía pBIC20 y 50 mg por litro de kanamicina para *Agrobacterium tumefaciens* que contenía pBIC20). Los medios para las plantas consistían de 0,5X MS (Sigma M5524) que consistía de sacarosa al 2% y agar al 1,2% a un pH de 5,8 - 6,0 con KOH 5 M. Despues de que los medios se había enfriado a aproximadamente a 45° C, se añadió kanamicina a razón de una concentración de 50 mg por litro.

5 Siembra en placa de la biblioteca de cósmidos y selección de cósmidos por transferencia de colonias

Se sembró la biblioteca en una placa por raspado de una pieza del cultivo congelado fuera del tubo del patrón y diluyendo luego la pieza congelada en 1 ml de LB líquido con tetraciclina. Se llevaron a cabo diluciones seriales y se determinó la cantidad de bacterias por 0,001 ml de volumen. El volumen que contenía 3.000 - 4.000 bacterias fue colocado luego sobre una caja de Petri de 120 mm de diámetro y se añadieron 0,6 ml de LB líquido con tetraciclina para ayudar en la difusión de las bacterias.

10 Se tomaron pequeños tacos de la biblioteca utilizando el extremo de una pipeta azul para remover un círculo de medio correspondiente a una señal positiva en una autoradiografía de un levantamiento de colonias de una placa que contiene las bacterias que portan el ADN que hibrida con la sonda utilizada. El taco fue luego colocado en un tubo de centrífuga de 1,5 ml que contenía 0,6 ml de LB líquido. El tubo fue sometido a agitación tipo vórtice y se hicieron diluciones seriales hasta que la solución original con el taco se había diluido diez mil veces. A partir de la dilución 10^{-4} , se tomaron 0,005 y 0,010 ml y se sembraron en placa utilizando 0,050 ml de LB líquido para ayudar en la difusión de las bacterias sobre las cajas de Petri de 100 mm que contenían LB con agar y tetraciclina.

15 Se utilizaron luego las placas que contenían alrededor de 300 colonias para aislar aún más las bacterias que contienen el cósmido de interés haciendo levantamiento de colonias, sondeando los filtros y luego recogiendo las bacterias de la región que contenía una colonia de hibridación. Se cultivó luego la colonia individual, se aisló el cósmido, se lo cortó con Hind III y se comparó el patrón de restricción con otros cósmidos.

Transformación de *Agrobacterium*

20 25 *Agrobacterium tumefaciens* fue transformada con los vectores binarios seleccionados del cósmido por medio de electroporación en una cubeta de 0,2 cm. La electroporación se programó con 25 μ F, 200 ohm; y 2,5 V. Despues de la electroporación, se cultivaron las bacterias durante una hora en LB líquido sin antibióticos y luego se transfirió el cultivo completo a un matraz de cultivo que contenía LB líquido y los antibióticos apropiados.

Transformación de la planta

30 35 Se desarrollaron durante la noche cultivos de *Agrobacterium tumefaciens*, se centrifugaron a 4000 rpm durante diez minutos y se resuspendieron en 30 ml de sacarosa al 5%. Se transfirieron luego las bacterias a 300 ml de sacarosa al 5% que contenía 0,3 ml de surfactante Silwet (Semillas Lehle, catálogo # vis-01). Se sumergieron e hicieron girar las macetas que contenían 20 plantas con *wri1-1*, que estaban empezando a brotar, en la solución durante al menos treinta segundos y luego se las colocó de costado en una superficie plana cubierta. Tres macetas, cada una de las plantas, fueron sumergidas con cada construcción. Las plantas sumergidas fueron cultivadas en cámaras para crecimiento de plantas y se cosecharon las semillas.

Selección de plantas transgénicas que contienen constructos del vector binario cósmido

40 45 Se esterilizaron en la superficie las semillas de las plantas sumergidas utilizando una solución al 20% de blanqueador (Clorox) y Triton X100 al 0,2%. Se incubaron aproximadamente 2.000 semillas en un tubo de centrífuga de 1,5 ml, que contiene 1,2 ml de la solución, con agitación suave durante veinte minutos y luego se enjuagaron en un ambiente estéril con agua estéril. Las semillas fueron lavadas mediante la adición de un ml de agua a las semillas. Luego se permitió que las semillas se asentaran, se eliminó el agua y se añadió más agua. Este procedimiento se llevó a cabo un total de seis veces. Las semillas de *Arabidopsis* se cultivaron sobre placas MS que contenían kanamicina y Clavamox para suprimir el crecimiento de *Agrobacterium* transformado. Se seleccionaron las plantas transgénicas resistentes a la kanamicina (transformantes) y se las transfirió al suelo. Las plantas se cultivaron en cámaras para crecimiento de plantas.

Calificación del fenotipo de las semillas de plantas transgénicas para hallar los cósmidos complementarios y la confirmación de la identidad de un gen para LMR

50 Se calificaron visual y microscópicamente silicuas maduras individuales de cada transformante para el rescate del fenotipo de *wri1-1*. Las semillas de plantas transgénicas con *wri1-1* que contienen al cósmido complementario exhiben el fenotipo de semilla de tipo silvestre y pueden ser diferenciadas de las semillas con *wri1-1*. Dependiendo del cósmido complementario específico, la elección de los posibles genes candidatos para LMR se reducirá drásticamente y se puede identificar el LMR más importante. Este experimento de rescate de cósmido puede ser seguido por complementación con vectores binarios que contienen los ADNc individuales de los candidatos del LMR para identificar inequívocamente al LMR que causa el fenotipo de *wri1* en el mutante *wri1-1*. Además, el ADN

genómico de *wri1-1* y *wri1-2* que codifica este LMR puede ser secuenciado para determinar las lesiones genéticas o el gen para LMR en estos dos mutantes.

Ejemplo 4

Clonación de *WR1* por medio de RT-PCR

5 La transformación del mutante de *Arabidopsis wri1-1* con vectores binarios de cósmidos (Ejemplo 3) que contienen las secuencias genómicas que abarcan la región identificada por el mapeo molecular (Ejemplo 1) conducen al descubrimiento de 4 cósmidos que se superponen que complementaron al mutante *wri1-1* y restauraron el fenotipo de semillas de tipo silvestre y el contenido de aceite de las semillas (Figura 2 y Ejemplo 3). El fragmento genómico común más pequeño en los 4 vectores binarios cósmidos que se complementan es de 8,5 kb, lo que indica que el gen *WR1* está contenido dentro de esa región de 8,5 kb. De acuerdo con la anotación generada automáticamente de la secuencia genómica en el GenBank, este fragmento de 8,5 kb contiene completamente dos marcos de lectura abiertos (ORF, SEQ ID NOs: 8 y 9) y los fragmentos parciales de otros dos ORF. Sin embargo, mediante el uso de RT-PCR, se descubrió que los dos ORF (SEQ ID NOs: 8 y 9) son en realidad parte del mismo gen que contiene un intrón inusualmente largo de 1641 nucleótidos. Con base en el mutante *wri1* y en nuestros datos de mapeo, denominamos a este gen "*WR1*". El producto RT-PCR *WR1* tenía una longitud de 1,6 kb. La amputación de un producto PCR genómico de 3,8 kb en las reacciones de RT-PCR que no fueron tratados con ADNasa indicó que el producto RT-PCR de 1,6 kb se deriva realmente de un transcripto de ARNm de 1,6 kb. Los productos de RT-PCR derivados del mutante *wri1-1* mostraron un pequeño aumento en tamaño en comparación con el mutante *wri1-2* y de tipo silvestre. Es posible que el empalme diferencial del transcripto de ARNm pueda conducir a tales diferencias de tamaño del transcripto.

20 Se aisló un ADNc de *WR1* de longitud completa por medio de RT-PCR utilizando una polimerasa de corrección de errores (Método descrito en el Ejemplo 5) y se determinó su secuencia de nucleótidos (SEQ ID NO: 1). Los iniciadores utilizados en esta reacción eran los siguientes: GGT ACC AAA TCT AAA CTT TCT CAG AG (SEQ ID NO: 36), ACT AGT AAA TCT AAA CTT tut CAG AG (SEQ ID NO: 37), TCT AGA AAA TCT AAA CTT TCT CAG AG (SEQ ID NO: 38) y TCT AGA GGC AAA GAC ATT GAT TAT TC (SEQ ID NO: 39). El ADNc codificado por este gen tiene una longitud de 1539 pb, tiene un UTR 3' de 166 pb y contiene un ORF de 1293 pb (SEQ ID NO: 2) que codifica una proteína de 430 aminoácidos (SEQ ID NO: 3). La secuencia de proteína codificada por el ADNc de *WR1* muestra una ligera similitud de secuencia con un factor de transcripción conocido, Aintegumenta, y contiene una dominio de enlazamiento de ADN tipo Apetala-2. Por lo tanto se propone que *WR1* actúa como un factor de transcripción que regula el metabolismo de lípidos y del compuesto que se almacena en la semilla durante el desarrollo de la semilla.

Ejemplo 5

Clonación del los ADNc de *Arabidopsis* que codifican proteínas LMR

35 El ADNc de longitud completa o parcial para LMR de *Arabidopsis* puede obtenerse por medio de RT-PCR utilizando ARN total de silicuas en desarrollo de *Arabidopsis*. Los iniciadores esenciales de la PCR específicos para LMR localizados secuencia arriba del codón de inicio y secuencia abajo del codón de detención pueden ser diseñados con base en la disponibilidad de la secuencia de ADN genómico. El mapeo del locus genético del LMR *wri1* como se describe en el Ejemplo 1 es de crucial importancia para este, ya que indica cual secuencia de ADN genómico tiene que ser utilizada para este propósito.

40 El ARN total se puede aislar de las semillas o las silicuas que se desarrollan de *Arabidopsis thaliana* por medio del método de Van Slooteren (1983, Plant Mol. Biol. 2: 321 - 333) con ligeras modificaciones. Para este método, se congela el tejido silicua en desarrollo (200 mg) con nitrógeno líquido y se muele hasta un polvo fino con un mortero con pistilo. El polvo se coloca en un tubo de microcentrifuga y se extrae el ARN con 500 µl de amortiguador de extracción (fenol: LiCl 0,1 M, Tris-HCl 100 mM [pH 8,0], EDTA 10 mM, SDS al 1% (p / v) [1:1]) precalentado a 90° C. 45 La mezcla se calienta adicionalmente durante 1 minuto a 90° C y luego se somete a agitación tipo vórtice durante 5 minutos. Las proteínas se extraen mediante la adición de 250 µl de cloroformo: alcohol isoamílico (24:1) y la mezcla se somete a agitación tipo vórtice durante 5 minutos y se centrifuga durante 44 minutos a 13.000 rpm en una centrífuga Eppendorf 5414 a 4° C. La capa acuosa se separa y se repite dos veces más la extracción de proteínas.

50 Se añade un volumen de LiCl 4 mM y se permite que precipite el ARN durante la noche a 4° C. Para recoger el ARN, se centrifuga la mezcla durante 44 minutos a 4° C a 13.000 rpm en una centrífuga Eppendorf 5414. Se resuspende el precipitado en 250 µl de agua estéril, desionizada. Para precipitar el ARN, se añaden 0,1 volúmenes de acetato sódico 3 M (pH 5,2) y 2 volúmenes de etanol al 100%. Se toma una alícuota y se centrifuga durante 20 minutos a 4° C a 13.000 rpm en una centrífuga Eppendorf 5414. Se lava el precipitado con etanol al 70% para eliminar las sales del precipitado y se seca utilizando un Speed Vac. Se resuspende el precipitado en 25 µl de H₂O tratada con DEPC y se analiza la integridad a través de electroforesis. El ARN se almacena a -70 ° C.

55 Para la RT-PCR y la clonación de genes para LMR de *Arabidopsis*, se logra la síntesis de la primera hebra de ADNc utilizando la transcriptasa inversa AMV (Roche, Mannheim, Alemania). Se amplifica el ADNc monocatenario

resultante a través de la Reacción en Cadena de Polimerasa (PCR) utilizando los dos iniciadores específicos del gene. Las condiciones para la reacción son las condiciones estándar descritas en el sistema PCR Expand High Fidelity (Roche). Los parámetros para la reacción son: cinco minutos a 94° C seguido por cinco ciclos de 40 segundos en 94° C, 40 segundos a 50° C y 1,5 minutos a 72° C. Esto es seguido por treinta ciclos de 40 segundos a 94° C, 40 segundos a 65° C y 1,5 minutos a 72° C.

5 El fragmento se extrae a partir de gel de agarosa con un kit de extracción de gel QIAquick (Qiagen) y se liga en un vector de clonación, por ejemplo el vector TOPO pCR2.1 (Invitrogen), siguiendo las instrucciones del fabricante. Los vectores recombinantes se transforman en células Top10 (Invitrogen) utilizando condiciones estándar. Las células transformadas se seleccionan sobre agar LB que contiene 100 µg / ml de carbenicilina, 0,8 mg de X-gal (5-bromo-4-cloro-3-indolil-β-D-galactósido) y 0,8 mg de IPTG (isopropiltiо-β-D-galactósido) cultivadas durante la noche a 37° C. Se seleccionan las colonias blancas y se utilizan para inocular 3 ml de LB líquido que contiene 100 µg / ml de ampicilina y cultivadas durante la noche a 37 ° C. Se extrae el ADN del plásmido utilizando el kit QIAprep Spin Miniprep (Qiagen) siguiendo las instrucciones del fabricante.

10 15 Se secuencia el gen para LMR de *Arabidopsis* clonado por RT-PCR para verificar que la secuencia completa de ADNc corresponda con la secuencia genómica (SEQ ID NO: 6). Se escinde el fragmento que contiene el ADNc para LMR de *Arabidopsis* a partir del vector recombinante TOPO PCR2.1 por medio de digestión con enzimas de restricción de acuerdo con las instrucciones del fabricante. Los fragmentos posteriores se cortan de un gel de agarosa con el kit de extracción de gel QIAquick (Qiagen) de acuerdo con las instrucciones del fabricante y se ligan en el vector binario vegetal utilizado para la sobreexpresión del gen en *Arabidopsis*.

20 25 El vector binario de la planta se construye por medio de digestión con enzimas de restricción. Un fragmento de ADN que codifica el gen para LMR se clona en el vector binario entre un promotor y un terminador de modo que una proteína LMR funcional puede ser expresada por este constructo. El constructo también contiene genes marcadores de resistencia a antibióticos que permiten la selección del vector binario en *E. coli*, *Agrobacterium* y en plantas transgénicas, respectivamente.

25 **Ejemplo 6**

Clonación y secuenciación de los ADNc alélicos para *wri1-1* y *wri1-2*

30 35 El aislamiento y secuenciación de los ADNc alélicos *wri1* de los mutantes *wri1-1* y *wri1-2* indicarán los cambios en la secuencia del gen *WR1* que conducen al fenotipo en el mutantes. Ambos mutantes son capaces de expresar el gen mutado y muestran una banda de la RT-PCR de aproximadamente el tamaño correcto (Ejemplo 4). Los productos de la RT-PCR obtenidos a partir del mutante *wri1-1* fueron secuenciados y se determinó que la secuencia contiene una mutación puntual en una posición correspondiente al nucleótido 3197 de la secuencia BAC T12E18 mostrada en la SEQ ID NO: 7, y en la Figura 6. Esta mutación puntual da como resultado una conversión de G por A en el sitio de empalme del intrón del gen *wri1-1* de tal manera que el ARNm del mutante *wri1-1* no se empalme más correctamente y el ORF continue hasta que se alcance un codón de detención en el intrón. Este empalme alterado da como resultado una forma truncada y no activa de la proteína de *WR1*. Estos mismos procedimientos son seguidos para identificar la(s) mutación(es) en el gen *wri1-2*.

35 **Ejemplo 7**

Plásmidos para la transformación de la planta

40 45 Para la transformación de la planta, se pueden utilizar vectores binarios tales como pBinAR (Höfgen y Willmitzer, 1990, Plant Science 66: 221 - 230). La construcción de los vectores binarios se puede llevar a cabo mediante ligación del ADNc en orientación sentido o antisentido o en el ADN-T. 5 prima al ADNc, un promotor de una planta activa la transcripción del ADNc. Una secuencia de poliadenilación se localiza 3 prima al ADNc.

La expresión específica del tejido se puede lograr mediante el uso de un promotor específico del tejido. Por ejemplo, la expresión específica de la semilla puede conseguirse por clonación de la napina o LeR4 o un promotor desconocido de proteína de semilla (USP) 5' al ADNc. También, se puede utilizar cualquier otro promotor específico de la semilla. Para expresión constitutiva dentro de la planta completa, se puede utilizar el promotor 35S del CaMV.

50 La proteína expresada puede ser dirigida a un compartimento celular utilizando un péptido señal, por ejemplo para plástidos, mitocondria o retículo endoplásmico (Kermode 1996, Crit. Rev. Plant Sci. 44, 4: 285 - 423). El péptido señal se clona 5 prima en el marco al ADNc para archivar la localización subcelular de la proteína de fusión.

50 **Ejemplo 8**

Transformación de la planta mediada por *Agrobacterium*

La transformación de la planta mediada por *Agrobacterium* con los ácidos nucleicos para el LMR descrita aquí se puede realizar utilizando técnicas estándar de transformación y regeneración (Gelvin, Stanton B. y Schilperoort, Robert A, Plant Molecular Biology Manual, 2^{da} edición. Kluwer Academic Publ., Dordrecht 1995 en Sect., Ringbuc

Zentrale Signatur: BT11-P; Glick, Bernard R. y Thompson, John E. *Methods in Plant Molecular Biology and Biotechnology*, S. 360, CRC Press, Boca Raton, 1993). Por ejemplo, la transformación mediada por *Agrobacterium* se puede realizar utilizando a cepa GV3 (pMP90) (Konez y Schell 1986, Mol. Gen. Genet. 204: 383 - 396) o la cepa LBA4404 (Clontech) de *Agrobacterium tumefaciens*.

- 5 *Arabidopsis thaliana* puede ser cultivada y transformada de acuerdo a condiciones estándar (Bechtold 1993, Acad. Sci. Paris 316: 1194 - 1199; Bent et al, 1994, Science 265: 1856 - 1860). Además, la colza puede ser transformada con los ácidos nucleicos para el LMR de la presente invención a través de la transformación del cotiledón o del hipocotilo (Moloney et al., 1989, *Planta Cell Report* 8: 238 - 242; De Block et al., 1989, *Plant Physiol* 91: 694 - 701). El uso de antibióticos para la selección de plantas y de *Agrobacterium* depende del vector binario y de la cepa de *Agrobacterium* utilizada para la transformación. La selección de la colza se realiza normalmente mediante el uso de kanamicina como marcador seleccionable de la planta. Además, la transferencia génica mediada por *Agrobacterium* al lino se puede realizar utilizando, por ejemplo, una técnica descrita por Mlynarova et al. (1994, *Plant Cell Report* 13: 282 - 285).
- 10 La transformación de la soja puede realizarse utilizando por ejemplo una técnica descrita en EP 0 424 047, Patente de EE.UU. No. 5.322.783 (Pioneer Hi-Bred International) o en el documento EP 0 397 687, Patente de EE.UU. No. 5.376.543 o en la Patente de EE.UU. No. 5.169.770 (Universidad Toledo). Las semillas de soja se esterilizan en la superficie con etanol al 70% durante 4 minutos a temperatura ambiente con agitación continua, seguido por Clorox al 20% (v/v) suplementado con Tween al 0,05% (v/v) durante 20 minutos con agitación continua. A continuación, se enjuagan las semillas 4 veces con agua destilada y se colocan sobre papel de filtro humedecido estéril en una caja de Petri a temperatura ambiente durante 6 a 39 horas. Se pelan las cubiertas de las semillas, y se separan los cotiledones del eje embrionario. El eje embrionario se examina para asegurarse de que la región meristemática no está dañada. Se recogen los ejes embrionarios cortados en una cala de Petri estéril medio abierto y se secan al aire hasta un contenido de humedad menor al 20% (peso fresco) en una caja de Petri sellada hasta su uso posterior.
- 15 El método de transformación de la planta es también aplicable a cultivos de *Brassica* y a otros cultivos. En particular, se esteriliza la superficie de semillas de canola con etanol al 70% durante 4 minutos a temperatura ambiente con agitación continua, seguido por Clorox al 20% (v/v) suplementado con Tween al 0,05% (v/v) durante 20 minutos, a temperatura ambiente con agitación continua. A continuación, se enjuagan las semillas 4 veces con agua destilada y se colocan sobre papel de filtro humedecido estéril en una caja de Petri a temperatura ambiente durante 18 horas. Se separan las cubiertas de las semillas y las semillas se secan al aire durante una noche en una caja de Petri estéril medio abierto. Durante este período, las semillas pierden aproximadamente 85% de su contenido de agua. Las semillas se almacenan luego a temperatura ambiente en una caja de Petri sellada hasta su uso posterior.
- 20 Se prepara un cultivo de *Agrobacterium tumefaciens* a partir de una sola colonia en medio LB sólido más antibióticos apropiados (por ejemplo, 100 mg /l de estreptomicina, 50 mg /l de kanamicina) seguido por el crecimiento de la única colonia en medio LB líquido hasta una densidad óptica a 600 nm de 0,8. A continuación, se precipita el cultivo de bacterias a 7000 rpm durante 7 minutos a temperatura ambiente, y resuspende en medio MS (Murashige y Skoog, 1962) suplementado con acetosiringona 100 µM. Se incuban cultivos de bacterias en este medio de preinducción durante 2 horas a temperatura ambiente antes de usarlos. Los ejes de embriones de semillas de soja cigóticos con un contenido de humedad aproximadamente del 44% se empapan durante 2 horas a temperatura ambiente con el cultivo en suspensión preinducido de *Agrobacterium*. (La imbibición de embriones secos con un cultivo de *Agrobacterium* es también aplicable a los ejes de embrión de maíz).
- 25 Los embriones se retiran del cultivo de imbibición y se transfieren a cajas de Petri que contienen medio MS sólido suplementado con sacarosa al 2% y se incuban durante 2 días, en la oscuridad a temperatura ambiente. Alternativamente, los embriones se colocan en la parte superior de papel filtro estéril humedecido (medio MS líquido) en una caja de Petri y se incuban en las mismas condiciones descritas anteriormente. Después de este período, los embriones se transfieren a medio MS ya sea sólido o líquido suplementado con 500 mg/L de carbenicilina o 300 mg/L de cefotaxima para matar la agrobacteria. El medio líquido se usa para humedecer el papel de filtro estéril. Los embriones se incuban durante 4 semanas a 25° C, bajo 440 µmol m⁻²s⁻¹ y un período de luz de 12 horas. Una vez que las plántulas han producido raíces, se transfieren a suelo Metromix estéril. El medio de las plantas *in vitro* se lava antes de transferir las plantas al suelo. Las plantas se mantienen bajo una cubierta de plástico durante 1 semana, para favorecer el proceso de aclimatación. Luego, se transfieren las plantas a una sala de crecimiento donde se incuban a 25° C, bajo 440 µmol m⁻²s⁻¹ de intensidad de luz y un período de luz de 12 horas durante unos 80 días.
- 30 Las muestras de las plantas transgénicas primarias (T0) se analizan por medio de PCR para confirmar la presencia de ADN-T. Estos resultados se confirman por medio de hibridación tipo Southern en donde se somete a electroforesis el ADN sobre un gel de agarosa al 1% y se hace transferencia a una membrana de nylon cargada positivamente (Roche Diagnostics). El kit PCR DIG Probe Synthesis (Roche Diagnostics) se utiliza para preparar una sonda marcada con digoxigenina por medio de PCR, y se utiliza como lo recomienda el fabricante.
- 35
- 40
- 45
- 50
- 55

Ejemplo 9

Transferencias tipo Northern de las semillas en desarrollo de mutantes de *Arabidopsis* y de *Arabidopsis* tipo silvestre

El análisis de transferencias tipo Northern de los niveles de expresión de transcritos de genes para LMR en semillas en desarrollo de tipo silvestre de *Arabidopsis* y los mutantes *wri1-1* y *wri1-2* se realiza de acuerdo con procedimientos estándar (Sambrook et al. 1989). No se esperan diferencias significativas en los niveles de expresión de un gen para LMR en el caso de mutaciones puntuales inducidas por EMS en un gen para LMR que conducen a un desplazamiento de marco o que introduce un codón de detención prematuro, excepto en los siguientes casos. En primer lugar, la ausencia de transcritos de ARNm para LMR en los mutantes *wri1* en comparación con el tipo silvestre puede indicar que la mutación inducida por EMS se produjo en partes críticas del promotor del LMR o en el sitio de iniciación de la transcripción. En segundo lugar, diferencias significativas observables en los niveles de expresión de un gen para LMR entre el tipo silvestre y los mutantes *wri1* pueden indicar que la transcripción del gen para LMR está sujeta a un favorecimiento o reducción de la expresión como parte del mecanismo de regulación en la que también participa el LMR. Por ejemplo, un mecanismo que percibe la falta de acumulación de compuestos de almacenamiento en las semillas puede provocar el favorecimiento de la expresión de genes positivos para LMR secuencia arriba en la cascada reguladora, o la reducción de la expresión de un gen para LMR que actúa negativamente secuencia arriba en la cascada reguladora. Esto será una prueba más de la función esencial de una proteína LMR en el metabolismo de la semilla y, en particular, la acumulación de compuestos que se almacenan en la semilla. Finalmente, el tamaño del transcríto de ARNm del gen mutante para LMR puede ser cambiado debido a la mutación, indicando que la mutación inducida por EMS en el gen *wri1* provoca un empalme incorrecto del transcríto de ARNm, lo que conduce a la vez al mal funcionamiento de la proteína WRI1.

Ejemplo 10

Evaluación de la expresión y la actividad de una proteína recombinante LMR

La expresión de una proteína recombinante LMR puede ser evaluada a nivel transcripcional utilizando transferencias tipo Northern que son bien conocidas por aquellos capacitados en la técnica y como se describe en el Ejemplo 9 anterior. Para evaluar la presencia o la cantidad relativa de proteína traducida a partir de este ARNm, se pueden emplear técnicas estándar, tales como una transferencia tipo Western. Además, las actividades y los parámetros cinéticos de las enzimas se pueden determinar por métodos bien establecidos en la técnica. Los experimentos para determinar la actividad de cualquier enzima alterada dada deben ser adaptados a la actividad específica de la enzima de tipo silvestre, que está dentro de la capacidad de aquellos capacitados en la técnica.

La actividad de las proteínas LMR que se unen al ADN puede ser medida por diferentes métodos bien establecidos, tal como el ensayo de desplazamiento de la banda de ADN (también llamados ensayos de retardo en gel). El efecto de tales proteínas LMR sobre la expresión de otras moléculas se puede medir utilizando ensayos del gen reportero (tal como se describe en Kolmar, H. et al 1995, EMBO J. 14: 3895 - 3904 y las referencias citadas allí). Los sistemas de prueba del gene reportero son bien conocidos y están bien establecidos para las aplicaciones tanto en células procariotas como eucariotas, utilizando enzimas tales como beta-galactosidasa, proteína fluorescente verde, y varias otras.

La determinación de la actividad de proteínas de transporte de membrana LMR puede realizarse de acuerdo con técnicas tal como las descritas en Gennis RB, 1989, Pores, Channels and Transporters, en Biomembranes, Molecular Structure and Function, Springer: Heidelberg, páginas 85 - 137, 199 - 234 y 270 - 322.

Ejemplo 11

Expresión específica en la semilla de un promotor del LMR

La región promotora de un gen para LMR se clona en un vector binario, dirigiendo al gen reportero GUS (Jefferson et al, 1987). El constructo resultante se transforma en plantas de *Arabidopsis* como se describe en el Ejemplo 8. La actividad del promotor en las semillas en desarrollo se prueba por coloración de GUS.

Ejemplo 12

Análisis del impacto de la transformación de una planta con ácidos nucleicos recombinante para el LMR sobre la producción de un compuesto deseado para almacenamiento en la semilla de la planta

El efecto de la modificación genética en plantas o sobre la producción de un compuesto deseado para almacenamiento en la semilla (tal como un ácido graso) se puede evaluar en ellos por medio del crecimiento de la planta que sirve como modelo bajo condiciones adecuadas y analizar las semillas o cualquier otro órgano de la planta para aumentar la producción del producto deseado (es decir, un lípido o un ácido graso). Tales técnicas de análisis son bien conocidas por aquellos capacitados en la técnica, e incluyen espectroscopia, cromatografía en capa fina, métodos de coloración de diversos tipos, métodos enzimáticos y microbiológicos, y cromatografía analítica tal como cromatografía líquida de alto rendimiento (véase, por ejemplo, Ullman 1985, Encyclopedia of Industrial Chemistry, Vol. A2, páginas 89 - 90 y 443 - 613, VCH: Weinheim; Fajlon, A. et al, 1987 Applications of HPLC in Biochemistry en: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehurt et al., 1993, Product recovery and purification, Biotechnology, vol. 3, Capítulo III, páginas 469 - 714, VCH: Weinheim; Better, P. A. et al. 1988, Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J. F. y Cabral, J.

- M. S. 1992, Recovery processes for biological materials, John Wiley and Sons: Shaeiwitz, J. A. y Henry, J. D., 1988, Biochemical separations en: Ullmann's Encyclopedia of Industrial Chemistry, Separation and purification techniques in biotechnology, vol. B3, Capítulo 11, páginas 1 - 27, VCH: Weinheim y Dechow, F. J. 1989).
- 5 Además de los métodos mencionados anteriormente, los lípidos de las plantas se extraen a partir de material vegetal como se describe por parte de Cahoon et al. (1999, Proc. Natl. Acad. Sci. 96, 22: 12935 - 12940 EE.UU.) y de Browse et al. (1986, Analytic Biochemistry 442: 141 - 145). El análisis cualitativo y cuantitativo de lípidos o de ácidos grasos es descrito en Christie, William W., Advances in Lipid Methodology. Ayr/Scotland: Oily Press. - (Oily Press Lipid Library; 2): Christie, William W., Gas Chromatography and Lipids, A Practical Guide - Ayr, Scotland: Oily Press. 1989, Repr. 1992. - IX, 307 S. - (Oily Press Lipid Library 1); "Progress in Lipid Research. Oxford: Pergamon Press, 1 10 (1952) - 16 (1977) u. d. T.: Progress in the Chemistry of Fats and Other Lipids CODEN.
- 10 Un método estándar común para el análisis de azúcares, sobre todo de almidón, es publicada por Stitt, R. Mc. C Lilley, R. Gerhard y Heldt MW (1989, "Determinación de los niveles de metabolitos en células específicas y en los compartimentos subcelulares de las hojas de la planta". Methods Enzymol. 174: 518 - 552; para otros métodos véase también Härtel et al. 1998, Plant Physiol Biochem 36: 407 - 417 y Focks & Benning y 1998, Plant Physiol 118: 91-15 101).
- 15 Para la extracción de azúcares solubles y almidón, se homogenizan 50 semillas en 500 μ l de etanol al 80% (v/v) de etanol en un tubo de ensayo de polipropileno de 1,5 ml e incubaron a 70° C durante 90 min. Tras la centrifugación a 16.000 g durante 5 min, se transfiere el sobrenadante a un nuevo tubo de ensayo. Se extrae dos veces el precipitado con 500 μ l de etanol al 80%. El disolvente de los sobrenadantes combinados se evapora a temperatura ambiente al vacío. El residuo se disuelve en 50 μ l de agua, que representa la fracción de carbohidratos solubles. El precipitado que queda de la extracción con etanol, que contiene los carbohidratos insolubles incluido el almidón, se homogeneiza en 200 μ l de KOH 0,2 N, y se incuba la suspensión a 95° C durante 1 hora para disolver el almidón. Despues de la adición de 35 μ l de ácido acético 1 N y centrifugación durante 5 min a 16.000 g, se utiliza el sobrenadante para cuantificación del almidón.
- 20 25 Para cuantificar los azúcares solubles, se añaden 10 μ l del extracto de azúcar a 990 μ l del amortiguador de reacción que contiene imidazol 100 mM, pH 6,9, MgCl₂ 5 mM, NADP 2 mM, ATP 1 mM, y 2 unidades de 2 ml^{-1} de la glucosa-6-P-deshidrogenasa. Para la determinación enzimática de glucosa, fructosa y sacarosa, se añaden 4,5 unidades de hexoquinasa, 1 unidad de fosfoglucoisomerasa, y 2 μ l de una solución saturada de fructosidasa sucesivamente. La producción de NADPH se controla fotométricamente a una longitud de onda de 340 nm. Del mismo modo, se analiza el almidón en 30 μ l de la fracción insoluble de carbohidrato con un kit de Boehringer Mannheim.
- 30 35 40 Un ejemplo para analizar el contenido de proteína en las hojas y las semillas se pueden encontrar en Bradford M. M. (1976, "Un método rápido y sensible para la cuantificación de cantidades en microgramos de proteína utilizando el principio de enlazamiento de colorante por la proteína". Anal. Biochem 72: 248 - 254). Para la cuantificación de la proteína total de la semilla, se homogenizan 15 - 20 semillas en 250 μ l de acetona en un tubo de ensayo de polipropileno de 1,5 ml. Despues de centrifugación a 16.000 g, se desechará el sobrenadante y se resuspendió el precipitado secado al vacío en 250 μ l de amortiguador de extracción que contiene Tris-HCl 50 mM, pH 8,0, NaCl 250 mM, EDTA 1 mM, y SDS al 1% (p/v). Despues de incubación durante 2 horas a 25° C, se centrifuga el homogenato a 16.000 g durante 5 min y se utilizarán 200 ml del sobrenadante para las mediciones de proteína. En el ensayo se utiliza y-globulina para la calibración. Para las mediciones de proteína se utilizan el ensayo de proteína DC de Lowry (Bio-Rad) o el ensayo de Bradford (Bio-Rad).
- 45 Los ensayos enzimáticos de la hexoquinasa y la fructoquinasa se realizan espectrofotométricamente de acuerdo con Renz et al. (1993, Planta 190: 156 - 165), de fosfoglucoisomerasa, 6-fosfofructoquinasa que depende de ATP, 6-fosfofructoquinasa que depende de pirofosfato, Fructosa-1,6-bifosfato aldolasa, triosa fosfato isomerasa, gliceral-3-P deshidrogenasa, fosfoglicerato quinasa, fosfoglicerato mutasa, enolasa y piruvato quinasa se realizaron de acuerdo con Burrell et al. (1994, Planta 194: 95 - 101) y de la UDP-glucosa pirofosforilasa de acuerdo a Zrenner et al. (1995, Planta 17: 97 - 107).
- 50 55 Los intermediarios del metabolismo de carbohidratos, como la glucosa-1-fosfato, la glucosa-6-fosfato, la fructosa-6-fosfato, el fosfoenolpiruvato, el piruvato y el ATP se miden como se describe en Härtel et al. (1998, Plant Physiol Biochem 36: 407 - 417) y los metabolitos se miden como se describe en Jelitto et al. (1992, Planta 188: 238 - 244). Además de la medición del compuesto que se almacena en la semilla (es decir, lípidos, almidón o proteína que se almacena), también es posible analizar otros componentes de las rutas metabólicas utilizadas para la producción de un compuesto deseado que se almacena en la semilla, tal como los productos intermedios y los productos secundarios, para determinar la eficiencia total de la producción del compuesto (O'Fiehn et al. 2000, Nature-Biotechnology 18: 1447 - 1161).
- Una prueba inequívoca de la presencia de productos de ácidos grasos puede ser obtenida por medio del análisis de plantas transgénicas siguiendo procedimientos analíticos estándar: CG, CG-MS o TLC en sus distintas versiones descritas por Christie y en las referencias citadas allí (1997, en: Advances on Lipid Methodology, 4^a edición: Christie, Oily Press, Dundee, páginas 119 - 169, 1998).

El material que va a ser analizado puede ser desintegrado por medio de sonicación, molienda con vidrio, nitrógeno líquido y molienda o a través de otros métodos aplicables. El material tiene que ser centrifugado después de la desintegración. El sedimento se resuspende en agua destilada, se calienta durante 10 minutos a 100° C, se enfriá sobre hielo y se centrifuga de nuevo seguido por extracción en ácido sulfúrico 0,5 M en metanol que contiene dimetoxipropano al 2% durante 1 hora a 90° C que conduce a aceite hidrolizado y compuestos lipídicos que resultan en lípidos transmetilados. Estos ésteres metílicos de ácidos grasos se extraen en éter de petróleo y finalmente se someten a análisis por GC utilizando una columna capilar (Chrompack, sílice fundida WCOT, CP-Wax-52 CB, 25 m, 0,32 mm) con un gradiente de temperatura entre 170° C y 240° C durante 20 minutos y 5 minutos a 240° C. La identidad de los ésteres metílicos de ácidos grasos resultantes se define por medio de la utilización de estándares disponibles en el comercio (es decir, Sigma).

En el caso de los ácidos grasos donde no se encuentran disponibles los estándares, la identidad de las moléculas se muestra a través de la formación de derivados y posterior análisis por GC-MS. Por ejemplo, la localización de ácidos grasos con enlace triple es mostrada a través de GC-MS después de la formación de derivados a través de 4,4-dimetioxiazolina (Christie, 1998).

15 Ejemplo 13

Purificación de una proteína LMR de una planta transgénica

Una proteína LMR puede ser recuperada de una célula de una planta (o de cualquier otra célula huésped) o sobrenadante celular utilizando diferentes métodos bien conocidos en el arte. Si la proteína LMR es secretada a partir de las células deseadas, entonces se remueven las células del cultivo por medio de centrifugación a baja velocidad, y se retiene la fracción del sobrenadante para purificación adicional. Si la proteína LMR no es secretada a partir de las células, puede ser recogida del cultivo por medio de centrifugación a baja velocidad y luego lisada por medio de técnicas estándar, tales como fuerza mecánica o sonicación. Una proteína LMR puede ser aislada también a partir de órganos y tejidos de la planta. Los órganos de las plantas se pueden separar mecánicamente de otros órganos o tejidos antes del aislamiento del compuesto que se almacena en la semilla del órgano de la planta. Despues de la homogenización del material de la planta, se remueven los residuos celulares por medio de centrifugación, y se retiene la fracción sobrenadante que contiene las proteínas solubles para una purificación adicional de la proteína LMR deseada.

La fracción sobrenadante de cualquier método de purificación es sometida a cromatografía con una resina adecuada, en la cual la molécula deseada es o bien retenida sobre una resina de cromatografía mientras muchas de las impurezas en la muestra no lo son, o donde las impurezas son retenidas por la resina mientras que la muestra no lo es. Tales etapas de cromatografía pueden ser repetidas cuantas veces sea necesario, utilizando las mismas o diferentes resinas de cromatografía. Alguien ordinariamente capacitado en el arte está bien versada en la selección de las resinas apropiadas de cromatografía y en su aplicación más eficaz para la purificación de una molécula particular. El producto purificado puede ser concentrado por medio de filtración o de ultrafiltración, y almacenado a una temperatura a la cual se maximiza la estabilidad del producto.

Existen una amplia gama de métodos de purificación que es conocida por aquellos capacitados en el arte y el método anterior de purificación no se constituye como una limitante. Tales técnicas de purificación están escritas, por ejemplo, en Bailey, J. E. & Ollis, D. F. Biochemical Engineering Fundamentals, McGraw-Hill, New York 1986.

40 La identidad y la pureza de las proteínas LMR aisladas pueden ser evaluadas por medio de técnicas estándar en el arte. Estas incluyen cromatografía líquida de alto rendimiento (HPLC), métodos espectroscópicos, métodos de coloración, cromatografía en capa fina, MRS, ensayos enzimáticos y métodos microbiológicos. Tales métodos de análisis son revisados por Patek et al. (1994 Appl. Environ. Microbiol. 60: 133 - 140), Malakhova et al. (1996 Biotekhnologiya 11: 27 - 32), Schmidt et al. (1998 Bioprocess Engineer. 19: 67 - 70), Ulmann's Encyclopedia of Industrial Chemistry (1996 vol. A27, VCH: Weinheim, páginas 89 - 90, 521 - 540, 540 - 547, 559 - 566, 575 - 581 y 581 - 587), Michal, G. (1999 Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons) y Fallon, A. et al. (1987 Applications of HPLC in Biochemistry en: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17).

45 Ejemplo 14

Identificación de ortólogos de proteína LMR en soja, Brassica, girasol y maíz

50 Las secuencias de ácido nucleico de las proteínas LMR de *Arabidopsis* pueden ser utilizadas para encontrar ortólogos en otras especies de plantas. El objetivo es encontrar proteínas LMR que se expresan en semillas o en órganos de almacenamiento como tubérculos. Ya que las proteínas LMR de *Arabidopsis* juegan un papel central en la regulación de la acumulación de lípidos que se almacenan en la semilla, se espera que los ortólogos de la proteína LMR en otras especies de plantas jueguen un papel igualmente importante y similar. Se pueden utilizar homologías de secuencia (identidades y similitudes) para hallar estos ortólogos por medio del uso de los métodos descritos anteriormente, y en particular, por medio de la búsqueda en bases de datos de secuencia con algoritmos de búsqueda como FASTA, TBLASTN o aquellos divulgados en la descripción anterior. Las búsquedas de similitud

de secuencia pueden ser llevadas a cabo sobre bases de datos de secuencia públicas o privadas y la identificación de los ortólogos de LMR de *Arabidopsis* en otras especies de plantas pueden ser seguidas con la clonación de los ADNc de longitud completa o parcial que codifican estos ortólogos de LMR.

5 Alternativamente, el complemento de todos los transcritos presentes en las semillas en desarrollo u órganos de almacenamiento, representados por ejemplo por bibliotecas de ADNc, puede ser seleccionado por hibridación para encontrar genes similares a los ácidos nucleicos para LMR de *Arabidopsis*. Los ácidos nucleicos para LMR de *Arabidopsis* o fragmentos cortos de los mismos, o sondas de oligonucleótidos con base en los ácidos nucleicos para LMR de *Arabidopsis* pueden ser utilizados para este propósito. Se puede llevar a cabo una selección por hibridación por medio del uso de protocolos estándar (Sambrook et al., 1989). Estos procedimientos permitirán la identificación 10 de clones de ADNc de longitud completa o parcial de genes para LMR de otras especies vegetales.

Ejemplo 15

Mutagénesis *in vivo*

15 La mutagénesis *in vivo* de ácidos nucleicos para LMR puede ser llevada a cabo por medio de la incorporación de un ácido nucleico para LMR en un plásmido (u otro vector) y el paso del plásmido a través de *E. coli* u otro microorganismo (por ejemplo *Bacillus* spp. o levaduras tales como *Saccharomyces cerevisiae*) que se ve afectada en su capacidad para mantener la integridad de su información genética. Las cepas típicas que mutan tienen 20 mutaciones en los genes para el sistema de reparación de ADN (por ejemplo, mutHLS, mutD, mutT, etc.; para referencia, véase Rupp, W. D., 1996, mecanismos de reparación de ADN, en: *Escherichia coli and Salmonella*, p. 2277 - 2294, ASM Washington). Tales cepas son bien conocidas por aquellos capacitados en el arte. El uso de tales cepas es ilustrada, por ejemplo, en Greener, A. y Callahan, M., 1994, *Strategies* 7: 32 - 34. La transferencia de ácidos nucleicos mutados para LMR en plantas es preferiblemente hecha después de la selección y prueba en microorganismos. Las plantas transgénicas se generan de acuerdo con diferentes ejemplos dentro de los ejemplos de este documento.

REIVINDICACIONES

1. Un ácido nucleico aislado que comprende una secuencia de polinucleótidos que codifica un polipéptido que actúa como un modulador de un compuesto que se almacena en la semilla en una planta, en donde el nivel del compuesto que se almacena en la semilla se incrementa, seleccionado de entre el grupo que consiste de:
- 5 (a) un polinucleótido como el descrito por la SEQ ID NO: 1;
 (b) un polinucleótido como el descrito por la SEQ ID NO: 2;
 (c) un polinucleótido como el descrito por la SEQ ID NO: 4;
 (d) un polinucleótido como el descrito por la SEQ ID NO: 5;
 (e) un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 3;
 10 (f) un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 6;
 (g) un polinucleótido que codifica un polipéptido que tiene al menos 70% de identidad de secuencia con el polinucleótido de cualquiera de (e) hasta (f), en donde la longitud de la comparación de secuencia es la longitud completa del polipéptido de cualquiera de (e) hasta (f); y
 (h) un polinucleótido complementario a un polinucleótido de cualquiera de (a) hasta (g).
- 15 2. El ácido nucleico aislado de la reivindicación 1, en donde el ácido nucleico comprende al polinucleótido como el descrito por la SEQ ID NO: 1.
3. El ácido nucleico aislado de la reivindicación 1, en donde el ácido nucleico comprende al polinucleótido como el descrito por la SEQ ID NO: 2.
- 20 4. El ácido nucleico aislado de la reivindicación 1, en donde el ácido nucleico comprende al polinucleótido como el descrito por la SEQ ID NO: 4.
5. El ácido nucleico aislado de la reivindicación 1, en donde el ácido nucleico comprende el polinucleótido como el descrito por la SEQ ID NO: 5.
6. El ácido nucleico aislado de la reivindicación 1, en donde el ácido nucleico comprende un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 3.
- 25 7. El ácido nucleico aislado de la reivindicación 1, en donde el ácido nucleico comprende un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 6.
8. El ácido nucleico aislado de la reivindicación 1, en donde el ácido nucleico comprende un polinucleótido que tiene al menos 90% de identidad de secuencia con la secuencia de nucleótidos de la SEQ ID NO: 1, la SEQ ID NO: 2, la SEQ ID NO: 4 o la SEQ ID NO: 5.
- 30 9. Un primer ácido nucleico aislado que hibrida bajo condiciones rigurosas hasta un segundo ácido nucleico seleccionado de entre el grupo que consiste de:
- (a) un segundo ácido nucleico que comprende al polinucleótido como el descrito por la SEQ ID NO: 1, la SEQ ID NO: 2, la SEQ ID NO: 4 o la SEQ ID NO: 5, y
 35 (b) un segundo ácido nucleico que codifica un polipéptido como es descrito por la SEQ ID NO: 3 o la SEQ ID NO: 6 donde el primer ácido nucleico codifica un polipéptido que actúa como un modulador de un compuesto que se almacena en la semilla en una planta,
 en donde las condiciones rigurosas comprenden el lavado a 62°C durante 30 minutos cada vez en SSC 3 veces/SDS al 0,1%, SSG 1 vez/SDS al 0,1% y SSC 1 vez, SDS al 1%.
- 40 10. Un vector de expresión recombinante que comprende al ácido nucleico de la reivindicación 1, en donde la expresión del vector en una célula huésped modifica un nivel de un compuesto que se almacena en la semilla en la célula huésped, en donde el compuesto que se almacena en la semilla se selecciona de entre el grupo que consiste de un lípido, un ácido graso, un almidón y una proteína que se almacena en la semilla.
11. La proteína del vector de expresión recombinante de la reivindicación 10, en el que la célula huésped es una célula de una planta.
- 45 12. Una célula de una planta transgénica que comprende un transgén que consiste de un polinucleótido como se duvulga en la reivindicación 1.

13. La célula de la planta transgénica de la reivindicación 12, en donde la expresión del ácido nucleico en la célula de la planta da como resultado un nivel modificado de un compuesto que se almacena en la semilla en la célula de la planta comparado con una variedad de tipo silvestre de la célula de la planta.
- 5 14. Una planta transgénica que comprende un transgén que consiste de un polinucleótido como el divulgado en la reivindicación 1.
15. La planta transgénica de la reivindicación 14, en donde la planta es una planta dicotiledónea.
16. La planta transgénica de la reivindicación 14, en donde la planta es una planta monocotiledónea.
17. La planta transgénica de la reivindicación 14, donde la planta es de una especie productora de aceite.
- 10 18. La planta transgénica de la reivindicación 14, en donde la planta se selecciona de entre el grupo que consiste de colza, canola, linaza, soja, girasol, maíz, avena, centeno, cebada, trigo, pimienta, Tagetes, algodón, palma oleaginosa, palma cocotera, lino, ricino y maní.
19. La planta transgénica de la reivindicación 14, en donde la expresión del ácido nucleico en la planta da como resultado un nivel modificado de un compuesto que se almacena en la semilla en la planta en comparación con una variedad de tipo silvestre de la planta.
- 15 20. La planta transgénica de la reivindicación 19, en donde el compuesto que se almacena en la semilla se selecciona de entre el grupo que consiste de un lípido, un ácido graso, un almidón y una proteína que se almacena en la semilla.
- 20 21. Una semilla producida por la planta transgénica de la reivindicación 19, en donde la semilla contiene la secuencia de polinucleótidos la planta es de una línea genéticamente pura para un nivel modificado del compuesto que se almacena en la semilla en comparación con una variedad de tipo silvestre de la planta.
22. Un método para producir una planta transgénica que tiene un nivel modificado de un compuesto que se almacena en la semilla que comprende la transformación de una célula de una planta con un vector de expresión que comprende un ácido nucleico y la generación a partir de la célula de la planta de la planta transgénica, en donde el ácido nucleico comprende un polinucleótido seleccionado de entre el grupo consistente de:
- 25 (a) el polinucleótido como el descrito por la SEQ ID NO: 1;
- (b) el polinucleótido como el descrito por la SEQ ID NO: 2;
- (c) el polinucleótido como el descrito por la SEQ ID NO: 4;
- (d) el polinucleótido como el descrito por la SEQ ID NO: 5;
- (e) un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 3;
- 30 (f) un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 6;
- (g) un polinucleótido que tiene al menos 70% de identidad de secuencia con el polinucleótido de cualquiera de a) hasta d), en donde la longitud de comparación de la secuencia es la longitud completa de la región de codificación;
- 35 (h) un polinucleótido que codifica un polipéptido que tiene al menos 70% de identidad de secuencia con el polipéptido de cualquiera de e) hasta f), en donde la longitud de comparación de la secuencia es la longitud completa del polipéptido de cualquiera de (e) hasta (f); e
- (i) un polinucleótido complementario a un polinucleótido de cualquiera de a) hasta h).
23. El método de la reivindicación 22, en donde el ácido nucleico comprende al polinucleótido como el descrito por la SEQ ID NO: 2.
- 40 24. El método de la reivindicación 22, en donde el ácido nucleico comprende al polinucleótido como el descrito por la SEQ ID NO: 4.
25. El método de la reivindicación 22, en donde el ácido nucleico comprende un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 3.
26. El método de la reivindicación 22, en donde el ácido nucleico comprende un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 6.
- 45 27. El método de la reivindicación 22, en donde la planta es una planta dicotiledónea.
28. El método de la reivindicación 22, en donde la planta es una planta monocotiledónea.

29. El método de la reivindicación 23, en donde la planta es de una especie productora de aceite.
30. El método de la reivindicación 22, en donde la planta se selecciona de entre el grupo que consiste de colza, canola, linaza, soja, girasol, maíz, avena, centeno, cebada, trigo, pimienta, Tagetes, algodón, palma oleaginosa, palma de coco, lino, ricino y maní.
- 5 31. El método de la reivindicación 22, en donde el nivel del compuesto que se almacena en la semilla se incrementa.
32. El método de la reivindicación 22, donde el compuesto que se almacena en la semilla se selecciona de entre el grupo que consiste en un lípido, un ácido graso, un almidón y una proteína que se almacena en la semilla.
- 10 33. Un método de modulación del nivel de un compuesto que se almacena en la semilla en una planta que comprende, la modificación de la expresión de un ácido nucleico en la planta, en donde el ácido nucleico se selecciona de entre el grupo que consiste de:
- (a) el polinucleótido como el descrito por la SEQ ID NO: 1;
 - (b) el polinucleótido como el descrito por la SEQ ID NO: 2;
 - (c) el polinucleótido como el descrito por la SEQ ID NO: 4;
 - (d) el polinucleótido como el descrito por la SEQ ID NO: 5;
- 15 (e) un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 3;
- (f) un polinucleótido que codifica al polipéptido como el descrito por la SEQ ID NO: 6;
- 20 (g) un polinucleótido que tiene al menos 70% de identidad de secuencia con el polinucleótido de cualquiera de a) hasta d), en donde la longitud de comparación de la secuencia es la longitud completa de la región de codificación;
- (h) un polinucleótido que codifica un polipéptido que tiene al menos 70% de identidad de secuencia con el polipéptido de cualquiera de e) hasta f); en donde la longitud de comparación de la secuencia es la longitud completa del polipéptido de cualquiera de (e) hasta (f); e
- (i) un polinucleótido complementario a un polinucleótido de cualquiera de a) hasta h).
34. El método de la reivindicación 33, en donde el ácido nucleico comprende al polinucleótido como el descrito por la SEQ ID NO: 2.
- 25 35. El método de la reivindicación 33, en donde el ácido nucleico comprende al polinucleótido como el descrito por la SEQ ID NO: 4.
36. El método de la reivindicación 33, en donde el ácido nucleico comprende un polinucleótido que codifica al polipéptido como se describe en la SEQ ID NO: 3.
- 30 37. El método de la reivindicación 33, en donde el ácido nucleico comprende un polinucleótido que codifica al polipéptido como se describe en la SEQ ID NO: 6.
38. El método de la reivindicación 33, en donde la planta es una planta dicotiledónea.
39. El método de la reivindicación 33, en donde la planta es una planta monocotiledónea.
- 40 40. El método de la reivindicación 33, en donde la planta es de una especie productora de aceite.
41. El método de la reivindicación 33, en donde la planta se selecciona de entre el grupo que consiste de colza, canola, linaza, soja, girasol, maíz, avena, centeno, cebada, trigo, pimienta, Tagetes, algodón, palma oleaginosa, palma de coco, lino, ricino y maní.
42. El método de la reivindicación 33, en donde el nivel del compuesto que se almacena en la semilla se incrementa.
43. El método de la reivindicación 33, donde el compuesto que se almacena en la semilla se selecciona de entre el grupo que consiste en un lípido, un ácido graso, un almidón y una proteína que se almacena en la semilla.
- 40 44. El método de la reivindicación 33, en el que la planta es transgénica.
45. El método de la reivindicación 33, en donde la planta no es transgénica.

FIGURA 1



FIGURA 2

Diagrama que ilustra la complementación del mutante de *Arabidopsis wri1* con clones cósmidos de tipo silvestre

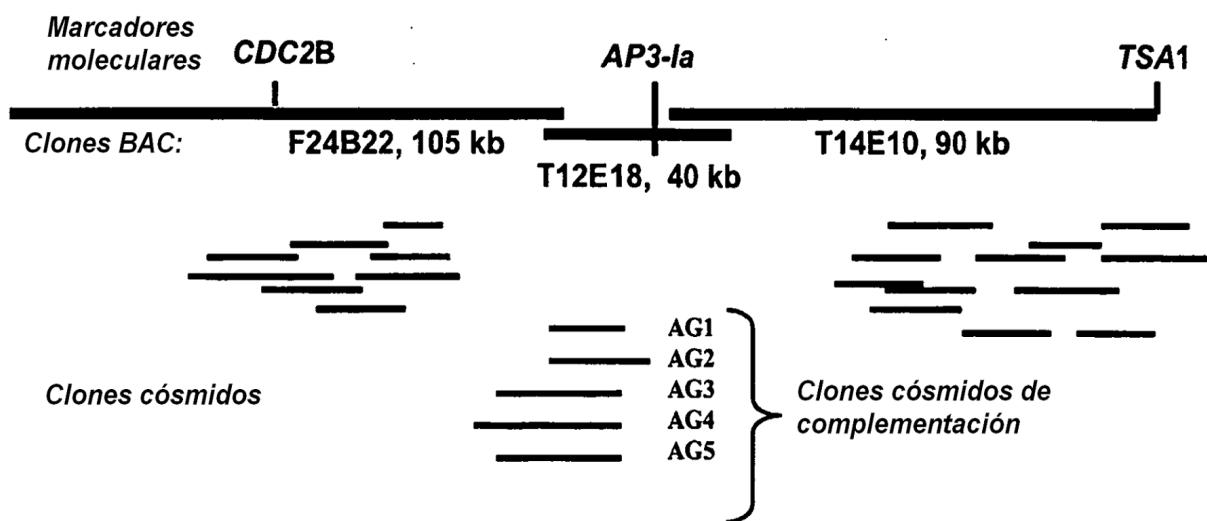


FIGURA 3

Alineamiento de secuencias de una secuencia de proteína LMR (WRI1, SEQ ID NO: 3) y secuencias candidatas de LMR SEQ ID NO: 8 y SEQ ID NO: 9

Nombre: WRI1 SEQ ID NO: 3 Longitud: 430

Nombre: WRI1 SEQ ID NO: 8 candidata Longitud: 205

Nombre: WRI1 SEQ ID NO: 9 candidata Longitud: 221

1	50
WRI1	MKKRLTTSTC SSSPSSSVSS STTTSSPIQS EAPRPKRAKR AKKSSPSGDK
Seq_8	MKKRLTTSTC SSSPSSSVSS STTTSSPIQS EAPRPKRAKR AKKSSPSGDK
Seq_9
51	100
WRI1	SDNPTSPAST RRSSIYRGVT RHRWTGRFEA HLWDKSSWNS IQNKKGKQVY
Seq_8	SHNPTSPAST RRSSIYRGVT RHRWTGRFEA HLWDKSSWNS IQNKKGK...
Seq_9
101	150
WRI1	LGAYDSEAAA AHTYELAALK YWGPDTILNF PAETYTKELE EMQRVTKEEY
Seq_8	QGAYDSEAAA AHTYDLAALK YWGPDTILNF PAETYTKELE EMQRVTKEEY
Seq_9
151	200
WRI1	LASLRRQSSG FSRGVSKYRG VARHHHNGRW EARIGRVFGN KYLYLGTYNT
Seq_8	LASLRRQSSG FSRGVSKYRG VARHHHNGRW EARIGRVFGN KYLYLGTYST
Seq_9
201	250
WRI1	QEEAAAAYDM AAIEYRGANA VTNFDISNYI DRLKKKGVVP FPVNQAKDQE
Seq_8	LSPFPFFF.....
Seq_9M AAIEYRGANA VTNFDISNYI DRLKKKGVFP FPVNQANHQE

FIGURA 3 (CONTINUACIÓN)

	251	300
WRI1	GILVEAKQEV ETREAKEEPR EEVKQQYVEE PPQEEEEKEE EKAEQQEAEI	
Seq_8	
Seq_9	GILVEAKQEV ETREAKEEPR EEVKQQYVEE PPQEEEEKEE EKAEQQEAEI	
	301	350
WRI1	VGYSEAAVV  ICCIDSSTIM EMDRCGDNNE LAWNFCMMDT GFSPFLTDQN	
Seq_8	
Seq_9	VGYSEAAVV  NCCIDSSTIM EMDRCGDNNE LAWNFCMMDT GFSPFLTDQN	
	351	400
WRI1	LANENPIEYP ELFLELAFED NIDFMFDDGK HECLNLELLD CCVVGRESPP	
Seq_8	
Seq_9	LANENPIEYP ELFNELAFED NIDFMFDDGK HECLNLENLD CCVVGRESPP	
	401	430
WRI1	SSSSPLSCLS TDSASSTTTT TTSVSCNYLV	
Seq_8	
Seq_9	SSSSPLSCLS TDSASSTTTT TTSVSCNYLV	

FIGURA 4A

Secuencia de nucleótidos de un ADNc para LMR de *Arabidopsis thaliana* (SEQ ID NO: 1)

AAACCACCTCTGCTTCCTCTGAGAAATCAAATCACTCACACTCCAAAAAAA
 ATCTAAACCTTCTCAGAGTTAATGAAGAAGCGCTTAACCACCTCCACTGTTCTCT
 TCTCCATCTTCCTGTTCTTCTACTACTACTTCCTCTCTATTAGTCGGAGGC
 TCCAAGGCCTAACGAGCCAAAGGGCTAAGAAATCTTCTCCTCTGGTATAAATC
 TCATAACCCGACAAGCCCTGCTTCTACCCGACGCAGCTCTATCTACAGAGGAGTCAC
 TAGACATAGATGGACTGGGAGATTCGAGGCTATCTTGGGACAAAGCTCTGGA
 ATTGATTAGAACAAGAAAGGCAAACAAGTTATCTGGGAGCATATGACAGTGAA
 GAAGCAGCAGCACATACGTACGATCTGGCTGCTCAAGTACTGGGGACCCGACAC
 CATCTGAATTTCGGCAGAGACGTACACAAAGGAATTGGAAGAAATGCAGAGAG
 TGACAAAGGAAGAATATTGGCTCTCCGCCAGAGCAGTGGTTCTCCAGAG
 GCGTCTCTAAATATCGCGCGTCGCTAGGCATCACCACAAAGGAAGATGGGAGGCT
 CGGATCGGAAGAGTGTGGAAACAAGTACTTGTACCTCGGCACCTATAATACGCA
 GGAGGAAGCTGCTGCAGCATATGACATGGCTCGATTGAGTATCGAGGCGAAACG
 CGGTTACTAATTGACATTAGTAATTACATTGACCGGTTAAAGAAGAAAGGTGTTG
 TCCCCTCTGTGAACCAAGCTAAGGATCAAGAGGGTATTCTTGTGAAGCCAAAC
 AAGAAGTTGAAACGAGAGAAGCGAAGGAAGAGCCTAGAGAAGAAGTGAACAAACA
 GTACGTGAAGAACCAACCGCAAGAAGAAGAAGAGAAGAAGAAGAGAAAGCAGA
 GCAACAAGAAGCAGAGATTGTAGGATATTCAAAGAAGAAGCAGCAGTGGTCAATTGCT
 GCATAGACTCTCAACCATAATGAAATGGATCGTTGTGGGACAACAATGAGCTG
 GCTTGGAACTCTGTATGGATACTGGTTCTCCGTTTGTACTGATCAGAATC
 TCGGAATGAGAATCCATAGAGTATCCGGAGCTATTCAAATGAGTTAGCATTGAGG
 ACAACATCGACTTCATGGATGGAAAGAGAGAGGCCACCCCTCTTCTTCAACCATTGCT
 TGGATTGTTGGTGGTGGAAAGAGAGAGGCCACCCCTCTTCTTCAACCATTGCT
 CTTATCTACTGACTCTGCTTCATCAAACAACAACAACACCTCGGTTCTGTAAAC
 TATTGGTCTGAGAGAGAAGAGCTTGCCTCTAGTTGAATTCTATTCTTCCGCT
 TCTTCTTCTTTCTTGTGGTTCTGCTTAGGGTTGTATTCAAGTTCAAGTCAGGG
 CTTGTTCGTTGGTCTGAATAATCAATGTCTTGCCTTCTAATGCTCCAAGTTC
 AGAT

FIGURA 4B

Secuencia de nucleótidos del marco de lectura abierto de un ADNc para LMR de *Arabidopsis thaliana* (SEQ ID NO: 2)

ATGAAGAAGCGCTAACCACTTCCACTGTTCTTCTCCATCTCCTCTGTTCTTC
TTCTACTACTACTTCCTCTCCTATTCACTGGAGGCTCCAAGGCCTAACAGAGCCAA
AAGGGCTAAGAAATCTTCTCCTCTGGTGATAAAATCTCATACCCGACAAGCCCTGC
TTCTACCCGACGCAGCTCTATCTACAGAGGAGTCACTAGACATAGATGGACTGGGA
GATTGAGGCTCATCTTGGACAAAAGCTCTTGGATTGATTCAGAACAAGAAA
GGCAAACAAGTTATCTGGGAGCATATGACAGTGAAGAACAGCAGCACATACGTA
CGATCTGGCTGCTCTCAAGTACTGGGGACCCGACACCCTTGAAATTTCGGCAGA
GACGTACACAAAGGAATTGGAAGAAATGCAGAGAGTGACAAAGGAAGAATATTG
GCTTCTCTCCGCCAGAGCAGTGGTTCTCCAGAGGCCTCTAAATATCGCCGC
GTCGCTAGGCATCACCACAAACGGAAGATGGGAGGCCTGGATCGGAAGAGTGTGG
GAACAAGTACTTGTACCTCGGCACCTATAATACGCAGGAGGAAGCTGCTGCAGCAT
ATGACATGGCTGCGATTGAGTATCGAGGCACAAACGCGTTACTAATTGACACATTA
GTAATTACATTGACCGGTTAAAGAAGAAAGGTGTTGTCCCGTTCCCTGTGAACCAAG
CTAAGGATCAAGAGGGTATTCTTGTGAAGCCAAACAAGAAGTTGAAACGAGAGAA
GCGAAGGAAGAGCCTAGAGAAGAAGTGAACACAGTACGTGGAAGAACCCACCGC
AAGAAGAAGAAGAGAAGGAAGAAGAGAAAGCAGAGCAACAAAGAAGCAGAGATTG
TAGGATATTAGAAGAACAGCAGTGGTCAATTGCTGCATAGACTCTCAACCATAA
TGGAAATGGATCGTTGTTGGGACAACAATGAGCTGGCTGGAACCTCTGTATGATGG
ATACTGGGTTTCTCCGTTTGACTGATCAGAATCTCGCGAATGAGAATCCCATAG
AGTATCCGGAGCTATTCAATGAGTTAGCATTGAGGACAACATCGACTCATGTCG
ATGATGGGAAGCACGAGTGCTGAACCTGGAAAATCTGGATTGTTGCGTGGTGGGA
AGAGAGAGGCCACCCTCTTCTCACCATTGCTCTTATCTACTGACTCTGCTT
CATCAACAAACAACAACACACCTCGTTCTGTAACATTGGTCTGA

FIGURA 4C

Secuencia de aminoácidos de una proteína LMR de *Arabidopsis thaliana* (SEQ ID NO: 3)

MKKRLTTSTCSSPSSSVSSSTTSSPIQSEAPRPKRAKKSSPSGDKSHNPTSPASTRR
SSIYRGVTRHRWTGRFEAHLWDKSSWNSIQNKKQKVYLGAYDSEEAAAHTYDLAAL
KYWGPDTILNFPAPAETYTKELEEMQRVTKEEYLASLRRQSSGFSRGVSKYRGVARHHN
GRWEARIGRVFGNKYLYLGTNTQEEAAAAYDMAAIEYRGANAVTNFDISNYIDRLKK
KGVVPFPVNQAKDQEGILVEAKQEVTREAKEPREEVKQQYVEEPPQEEEKEEKA
QQEAEIVGYSEAAVVNCCIDSSTIMEDRCGDNNEAWNFCMMDTGFSPFLTDQNL
NENPIEYPELFNELAFEDNIDMFDDGKHECLNLENLDCCVVGRESPPSSSPLSCLSTD
ASSTTTTSVSCNYLV

FIGURA 5A

Secuencia de nucleótidos de un ADNc para LMR de *Arabidopsis thaliana* (SEQ ID NO: 4)

TTGGTACCAAATCTAAACTTCTCAGAGTTAATGAAGAAGCGCTTAACCACCTCCA
 CTTGTTCTCTTCTCCATCTCCTCTGTTCTTCTACTACTACTCTCCTCTCCTATTC
 AGTCGGAGGCTCCAAGGCCTAAACGAGCCAAAGGGCTAAGAAATCTTCTCCTCT
 GGTGATAAAATCTCATAACCCGACAAGCCCTGTTCTACCCGACGCAGCTCTATCTAC
 AGAGGAGTCACTAGACATAGATGGACTGGAGATTGAGGCTCATCTTGGGACAA
 AAGCTCTTGGATTGATTGAGGAAACAAGAAAGGCAAACAAGTTATCTGGGAGCAT
 ATGACAGTGAAGAAGCAGCAGCACATACGTACGATCTGGCTGCTCTCAAGTACTGG
 GGACCCGACACCATCTGAATTTCGGCAGAGACGTACACAAAGGAATTGGAAGA
 AATGCAGAGAGTGACAAAGGAAGAATATTGGCTCTCCGCCAGAGCAGTG
 GTTCTCCAGAGGCCTCTAAATATCGCGCGTCGCTAGGCATCACCACAAACGGAA
 GATGGGAGGCTCGGATCGGAAGAGTGTGTTGGGACAAAGTACTTGTACCTCGGCACC
 TATAATACGAGGAGGAAGCTGCTGCAGCATATGACATGGCTGCGATTGAGTATCG
 AGGCGCAAACCGGTTACTAATTGACATTAGTAATTACATTGACCGGTTAAAGAA
 GAAAGGTGTTCCCGTCCCTGTGAACCAAGCTAACCATCAAGAGGGTATTCTGT
 TGAAGCCAACAAGAAGTTGAAACGAGAGAAGCGAAGGAAGAGCCTAGAGAAGAA
 GTGAAACAACAGTACGTGGAAGAACCAACCGCAAGAAGAAGAAGAGAAGGAAGAAG
 AGAAAGCAGAGCAACAAGAACGAGAGATTGTAGGATATTCAAAGAAGCAGCAGT
 GGTCAATTGCTGCATAGACTCTCAACCATAATGGAATGGATCGTTGGGGACAA
 CAATGAGCTGGCTTGGAACTTCTGTATGGATACAGGGTTCTCCGTTTGACT
 GATCAGAATCTCGCGAATGAGAACATCCCATAGAGTATCCGGAGCTATTCAATGAGTTA
 GCATTGAGGACAACATCGACTCATGTTGATGATGGGAAGCAGCAGTGCTGAAC
 TTGGAAAATCTGGATTGTTGCGTGGTGGGAAAGAGAGAGAGGCCACCTCTTCTCA
 CCATTGTCTGCTTACTGACTCTGCTTCAACAAACAACAACAACCTCG
 GTTCTTGTAACTATTGGTCTGAGAGAGAGAGCTTGCCTCTAGTTGAATTCTA
 TTTCTCCGCTTCTCTTCTTTCTTTGTTGGGTTCTGCTTAGGGTTGTATTTC
 AGTTCAAGGGCTTGTGCGTGGTCTGAATAATCAATGTCTTGCC

FIGURA 5B

Secuencia de nucleótidos del marco de lectura abierto de un ADNc para LMR de *Arabidopsis thaliana* (SEQ ID NO: 5)

ATGAAGAAGCGCTAACCACTTCCACTTGTCTTCTCCATCTCCTCTGTTCTTC
TTCTACTACTACTTCCCTCCTATTCACTCGGGAGGCTCCAAGGCTAAACGAGCCAA
AAGGGCTAAGAAATCTTCTCCTCTGGTATAAAATCTCATAACCCGACAAGGCCCTGC
TTCTACCCGACGCAGCTCTATCTACAGAGGAGTCACTAGACATAGATGGACTGGGA
GATTGAGGCTCATCTTGGACAAAAGCTTGGATTGATTGAGAACAGAACAGAACAGA
GGCAAACAAGTTATCTGGGAGCATATGACAGTGAAGAACAGCAGCACATACGTA
CGATCTGGCTGCTCTCAAGTACTGGGGACCCGACACCCTTGAATTTCCGGCAGA
GACGTACACAAAGGAATTGGAAGAAATGCAGAGAGTGACAAAGGAAGAATATTG
GCTTCTCTCCGCCAGAGCAGTGGTTCTCCAGAGGCGTCTAAATATCGCCGC
GTCGCTAGGCATCACCACAAACGGAAGATGGGAGGCTGGATCGGAAGAGTGTTGG
GAACAAGTACTTGTACCTCGGCACCTATAATACGCAGGAGGAAGCTGCTGCAGCAT
ATGACATGGCTGCGATTGAGTATCGAGGGCGAAACGCGGTTACTAATTGACATTA
GTAATTACATTGACCGGTTAAAGAAGAAAGGTGTTTCCGTTCCCTGTGAACCAAG
CTAACCATCAAGAGGGTATTCTTGTGAAGCCAAACAAGAAGTTGAAACGAGAGAA
GCGAAGGAAGAGCCTAGAGAAGAAGTGAACAAACAGTACGTGGAAGAACACCACGC
AAGAAGAAGAAGAGAAGGAAGAAGAGAAAGCAGAGCAACAAGAAGCAGAGATTG
TAGGATATTCAAAGAAGCAGCAGTGGTCAATTGCTGCATAGACTCTCAACCATAA
TGGAAATGGATCGTTGGGACAACAATGAGCTGGCTGGAACTTCTGTATGATGG
ATACAGGGTTTCTCCGTTTGACTGATCAGAATCTCGCGAATGAGAAATCCCATAG
AGTATCCGGAGCTATTCAAATGAGTTAGCATTGAGGACAACATCGACTTCATGTCG
ATGATGGGAAGCACGAGTGCTGAACCTGGAAATCTGGATTGTTGCGTGGTGGGA
AGAGAGAGGCCACCCCTCTTCTCACCATGTCTTCTTACTGACTCTGCTT
CATCAACAACAACAACACCGGTTCTGTAACTATTGGTCTGA

FIGURA 5C

Secuencia de aminoácidos de una proteína LMR de *Arabidopsis thaliana* (SEQ ID NO: 6)

MKKRLTTSTCSSPSSSVSSSTTSSPIQSEAPRKRAKKSSPSGDKSHNPTSPASTRR
SSIYRGVTRHRWTGRFEAHLWDKSSWNSIQNKKKGKQVYLGAYDSEEEAAHTYDIAAL
KYWGPDTILNFPAETYTKELEEMQRVTKEEYLASLRRQSSGFSRGVSKYRGVARHHHN
GRWEARIGRVFGNKYLYLGTNTQEEAAA YDMAAIEYRGANAVTNFDISNYIDRLKK
KGVFPFPVNQANHQEGLVEAKQEVEETREAKEPREEVKQQYVEEPPQEEEEKEEEKAE
QQEAEIVGYSEAAVNCCIDSSTIMEDRCGDNNELAWNFCMMDTGFSPFLTDQNL
NENPIEYPELFNELAFEDNIDMFDDGKHECLNLENLDCCVVGRESPPSSSPLSCLSTD
ASSTTTTSVSCNYLV

FIGURA 6

Secuencia de ácidos nucleicos truncada del clon BAC T12E18 (GenBank, número de acceso AL132971) (SEQ ID NO: 7)

AAGCTTCAATATGATTCACTAAAATGTGGTAAGTAGGGGAGATGAAAAGGACTTA
 TCTTATAGAAATTGAGTCGTTATTTGGTCCACCATACATTACATGAAATGGT
 TACATCAGGACGAGTGTGAGACCTAGAAAATCAGAAACCAAAACCACAAAA
 GAATTAAAATGAAATTACAGACACAATCATGAATCTCTAACATAAGAAATAACAAA
 ACCACCTCACGTGATACAACACTCTTGGTTCTCAACACAAAATCCCATTGTGTA
 CAGCTCTATACCAACATTGACATAAGACTTGGGACATTGTACGAAACACATT
 TCTTCAATTCTCCAAAGTGCTTTAGTTAACGGACCATCTCTAGAGTTCTGATGA
 ACTAACACACTGCATCAATTGATAAGATAACAATCGTAAACTGTATTCAACATA
 AAAAAGACTTGATTAGCGAGAAATGCGCAATCCAACATAGTTCTAAGTTGTCTAG
 ATAGTAGATAACCGACTCAACTATAGTGTCCAAGTTGTCAACACAGTTCAATCCAC
 TACAAACACCTACGGTAATTACAATGCTCTTAAACCAACACTAGGCTAATCACTAA
 CCCTGCCCGAGGTAAAACAGAGTTGTGTGTCCTAAAACGTGAAATTGAGGA
 GAAGTCAAAAACCTTACTCATGGAAAACAACATTACCAACCGAGCGATCAGTCTG
 AGTTGACATCCTCTGATCCACCATACCTAGCAACGTCTTGTGTCCTATCAAATC
 CCATTATAATGTCATCCTCACTTCTGTCACCCATAACCATGGATTTCAGTT
 TCCCAAAATCACTCTCGCTAAGATTACCAAGATTTTCTCTGCGTCTTCCTCCAC
 TCTGAAGAACCGGGCACAAGTTGAAAGTTGCTATCCTCTGGATCCGAACCAACGAG
 AAACCATCTAGAAATGTAACCGTTGCGTAGATTATCTTGTATTAGTACTCCGAC
 GTTTTATTGTGTTTTAAGAAGAGAACGGTGGAACAGAGGATGCTCTCAATGCAAC
 ATTTCAAGTCGTTATGTGGTGGAACGTAAGACATCTGAAACGAATATGATA
 CTAGGAAAATTAAAGCTGATTCAAAAAGATAACTCAAGTTCCGAGCAACGAAG
 TATGAGAAGAACAAAAGAGTACCTTTAATTATATCAGCTTAGGTCAATTCAA
 AGAAGAACGACGATCATCGAGAGGAAACTGAAAAGGCGCAGAGCAGTGA
 AGTGGAAAGAGACGAGACGGTTGGCTTTTATCACTAGAAGTTAGTCTGGAAAT
 GTGCATAATATTAAATTCAATTCAATTCAAATCAAAAATAAAACCGATAACGA
 AACTCTGTTATCATATTATGAGCGTTGATAATTAGAGAATCTCTGTTCTTAGAA
 AATTGTATGAACATTGTTAAATAACTATTAAATGCTAAACCTCATTGTT
 TAGTCCTAAATGTTAGGAATCTTAGTATTCACTTTCTCTCTTATGATACACTT
 CCGGTTAGATTCTAGGACGTTAAATTAAATTACTGCTAATAAAATAGTTA
 TTTATTAAACAAATTAAATAAAACATTGATCTCTAGAGTAAAAAGTTAA
 AAATTGGTGAATGTTTATTGGCAACCCAACTTAATTCTGAAACGAATCACG
 AGAATTATTAAACGTACTACAAAAGGTACATCTAATTCTAAACCGAATTAGACT
 ATCGTCTGGTTGTTGGATGTTGCACTGCTATATTATGATAATTCTGTAAGC
 ATTACAATTAAATCGTTCTGTTTACTAAAGGAGTTGACTGTTAAACGACTG
 TAAACACTCGTGTGCTGCTTAAATCATGTTAAAGTTAAATGATAACAGTTAGAAC
 ATCCAACCTTACATATGCTTGTGATTAGACATTCTACGATATGTTGGATTGACAC
 GTGCTAATCATCTGATCATTCTGTTTACATATGTTCTAATTAGGCATTTCG
 ACGATACATTAGATCGATAATGTTAAATCATCTGACATTGTTACTGTTAT
 GTTCTAATTATGTTATTACGATAAACATTAGATCGACACATGCAAAACATTGAC
 AATTGTTGAACACTGAAGACCTAATTGGTATAATACAGCTCAGCGAAGATACTT
 ACAACTAATTGGAGCGTTATAGAAAATAGGCAAATAGCCAAATATGCTAATA
 AAAAGATTAGAACTTAATAGGTAACTTAAACACATTTATAACACACAAGCATATA

FIGURA 6 (CONTINUACIÓN)

TATTAAAAGGGAAAAAACTAGTCACCTTCATAGACCATCATTACCAATAAACTC
 AATTCAATCAAATTAAATTGGCTACTATTTCTTCTAAAACCAAAACATCA
 TCATAGTAGATTCTCAATTAAAATAGAATTCAAATTGTTGACCAGGGAAATAA
 ATAGAATTGCAAAAAACAAATTAAAGATATTACATTCAACTTTCCCCAAAAA
 ATTAGAGAAAAAAATCAAATTAAATTCAATAATTCCAAAACATAAAACGTACAAAA
 AAGTTAATGATAAAAAACGTAAAAAAATACAAAATACAAAGAAAAAGAAA
 GACAGCGTGGAGAGTAAAGCTATTGGTAGCATGGCTGCAACAGAAGGCCAAGTGT
 TGGTCTTACAATTCAATTAACTCTTGTGTTAATTAAACCAATTAGCCCTTCCTC
 CTTCAGGGTTATTAACTTGCCCTTCTCGTTCTCCTTTCTTAAACCACTCT
 GCTTCCTCTCCTCTGAGAAATCAAATCACTCACACTCCAAAAAAATCTAAACCT
 TCTCAGAGTTAATGAAGAAGCGCTAACCAACTCCACTTGTCTTCTCCATCTT
 CCTCTGTTCTCTACTACTACTTCCTCTCCTATTCACTCGGAGGCTCAAGGCCT
 AAACGAGCCAAAAGGGCTAAGAAATCTCTCCTCTGGTGATAAAATCTCATAACCCG
 ACAAGCCCTGCTTCAACCGACGCAGCTATCTACAGAGGAGTCAGAGTTTTA
 TTTTTGGAAATTAAATGATTGGTGTGAGATTGGATTGGGTTTGTCTTAAAC
 TGCATTGTAAGATTGCATGTTGTTGGATTTCAGACATAGATGGACTGG
 AGATTGAGGCTCATCTTGGGACAAAAGCTCTGGAATTGAGAACAGAA
 AGGCAAACAAGGTTCGTCTCTTTCTCGTAACTCATGATTGTTGTT
 TAATAAAGATCTGGACTTAACTGATAAAATTGGTTCTTGATCTGTTGTTGATCT
 CAACTCGTCACAACCTCACCAAGTTATCTGGTAAGCTCTAATTCTCTGAAACAAA
 AGATCAATTGTTTACTAAATTGAAAGAGAAAAAAAGAGAGAGTGGATTGG
 TCAGAAGAATCTCAACTGCTTCACGCGTAATTGCAGGAGCATATGACAGTGAAGA
 AGCAGCAGCACATACGTACGATCTGGCTCTCAAGTACTGGGGACCCGACACCA
 TCTGAATTTCGGTAAACCAAAAAACAAAAATCAGATTGTTGATATGCATGTT
 TGTGATTGGAATCTGGATTATAATTAAAAAAATGGGAAATCAGGCAGA
 GACGTACACAAAGGAATTGGAAGAAATGCAGAGAGTGACAAAGGAAGAATTG
 GCTTCTCTCCGCCAGAGCAGTGGTTCTCCAGAGGCGTCTAAATATCGCCGC
 GTCGCTAGGTTCTTTCTCTCTTTTAATTCTTAGATTATTAA
 ATTTCGGAATTACTACCAAAATTGAGAAATGATTTCCTATTTCGGATATCTGAAT
 AACAGAATTATTAGGAAAAAAATCTGATCATGAAAATTGCTTTAGAATATT
 CTCTTTCTTAAACCAATTATGTTTCTCAGCACTGCTAAAGTTATG
 GATTCAATAGTTGGCATTATTCTTAAACCAATTAGGATTATTGTTCTAAAC
 AGGCATCACCACACCGAAGATGGGAGGCTGGATCGGAAGAGTGTTGGAAACAA
 GTACTTGTACCTCGGCACCTATAGTACGTTATCTCCTTCCCTTTCTTAGTAAT
 TTTAGAAAAAAATAGATATGTAATTGTTCTGGTAATTAAAATAATTAGCGTAATTAT
 TGACTTTTATAACTACCGGGCATAACGGATCCTTACCTGTTATGCTTATAA
 TATATAATTGTAAGTATAAAATAGAGTGTGATAATTGTTAGACTGTTTTGTGTT
 TGTTTATAAGAGTGATTAAAGAATATTAAATTGTTAGTGAGACAATTAGAATAATA
 TAATGGGAAGCAGTGGCAGTGGGTTGAATTTCACACACACTGACTCACGTGAG
 GCGAGAGTTGACATCATGCTTAAATTGATTCTTAAATCAAATCAACT
 TTTTTCTCTTCTTTAATTATCTGATCCCTCTGCATAATTACCTTAAATTCTGC
 ATTTTTGTGGATCCGATACTCTGAATACAAAATTGAGAACTCTGCAGAAGGGAAAT
 ATTAACAACAACCTTTACTGAAAAGTAATTCCATTTTTATTGTTTGCTGA
 CTCTTACCGGGCCTTAGATTATTAAAGGACCTCTAACTTCAACAAATCTCAAAT
 TTTGAAAATTAGATTTTAAAAAGTAATACAAATTGAGATTGCAAAATATA
 GGCATTGTTCTATAACAAAGATACTTATTACCAAAAAAAAGAAAAGAGTT
 GTCAAGAGCATAATTACAAAAAAATTGAAATAAGTAGTAGTTGTGAAATT

FIGURA 6 (CONTINUACIÓN)

GTATAGAAAAATAATGTAGGTTACAAGTGTAAAGGCGCGTAGCGCGTAGGTC
 ACGTGATAACACTCTCACTCATAAAAGGACAAAATAGTCAGAGAGGCTTAGGA
 CCAAACCCGAGGTCGATCTGGTTGCTCTGTTGGTTATTAAATTGGATAA
 TTAGTTAGAGTTGAGACTTGCTCTGAGTAGGAGTCACGAGCCCTACGTGCACTGCT
 CATCTCTCTCATCTCTACCATATCTTCATCTTGTCTCCGAACAAAATCTGGTC
 TAACTTTATCTTCTTTCTTTAATAATTGTCTCTACTTACCAATTCTCTA
 GATTATTCTGGTACCAAACTTATAACTTAATAGTTGATTAGTGTAGAGTTGACACT
 AGGTTGGTGTAAATTATTGTTAATTAACTCAGTAATCGTACGTTCGTGTATTAATT
 ATATACACAAATTGTTGCGATGACTAAATTAAAGTCACAAGTTGGACTCTTAGTG
 TTTAGAGCGGCGCAGTGGAGGAGAAATGGTCTTGTACACGCCTCACATCTCCACAC
 AAATCGTGTAAACCCTAGTTGTCCTACAAAAACACGTACCAAATTCTATTGATTCT
 TTGTCCTTATTAGGTATCATAAATTCTCTAATTAAATATGAAACGACAAAGAAGAA
 ACTCTCTTTAAACACTTGTGGTCTTATTGGTTAGCCACTTACCAAGGTAATGT
 GAAAGTTAACAAACAAGTGTGTTAACACTTCAAAACACTGACTTGTGATTGCTT
 GATTAGAAGCTTTAGCTAACACAGTTGTCTATTATTGGTTACAGATACGCAG
 GAGGAAGCTGCTGCAGCATATGACATGGCTGCGATTGAGTATCGAGGCGCAAACGC
 GTTACTAATTGACATTAGTAATTACATTGACCGGTTAAAGAAGAAAGGTGTTT
 CCCGTTCCCTGTGAACCAAGCTAACCATCAAGAGGGTATTCTTGTGAAGCCAACA
 AGAAGTTGAAACGAGAGAAGCGAAGGAAGAGCCTAGAGAAGAAGTGAACAAACAG
 TACGTGGAAGAACCGACCGCAAGAAGAAGAAGAGAAGGAAGAGAAAGCAGAG
 CAACAAGAAGCAGAGATTGAGGATATTCTAGAAGAAGCAGCAGTGGTCAATTGCTG
 CATAGACTCTCAACCATAATGAAATGGATCGTTGTGGGGACAACAATGAGCTGG
 CTTGGAACCTCTGTATGATGGATACAGGGTTCTCCGTTTGTACTGATCAGAATCT
 CGCGAATGAGAATCCCAGAGATCCGGAGCTATTCAATGAGTTAGCATTGAGG
 ACAACATCGACTTCAATTGTCGATGATGGGAAAGCACGAGTGTGACTTGGAAAATC
 TGGATTGTTGCGTGGTGGGAAAGAGAGAGAGGCCACCCCTCTCTTCTCACCATTGCTT
 CTTATCTACTGACTCTGCTTCATCAACAACAACAACAACCTCGGTTCTGTAAAC
 TATTGGTCTGAGAGAGAGAGCTTGCCTCTAGTTGAATTCTATTCTTCCGCTT
 CTTCTCTTTTGTGGGTTCTGCTTAGGGTTGTATTCACTGAGTTCAAGG
 TTGTTGCTGGTCTGAATAATCAATGTTGCCCCCTTCTAATGCTCCAAGTCA
 GATAAGAAATAAAACTAAATGAACCGTAAACACAAAGGAGGCTAAGATGTCTAT
 GAATTCTCTGGCATATGTGCAATAAGGAAGAGCTCAAGGAATCTGAACAAAGA
 GAAGGCTCAAAAGGCTAGACCTCTCAAGTTCCCTGTCTTATTATTGACCGGACTA
 TGAAGTTCTAGGAAGGTTATAGGCAAAAGTTATTCTTAACCTTAATTCTAA
 TTCTGATTGCGATGATTGGTTGCTAGATAGTTAGTAAAGAGCCTTACCATCC
 AAACCTGTGATCAATATTGTAACCGTGAACGTCTCATTAGGGTAATATTGAAT
 ATCTTGTGTTCCGGTTACACATAACAACCCGAAAAGTTGTCATTGGCTCGGAA
 TACATCTGTATAAAATTGCTTTACTTTAGTTGTAATTCATCAGACAATATATA
 GGTAGTCTATTGTTGAAACAACCTCAATTGTAATTCAATAGTATGATT
 TGCATTCAAGTAGAAAATCAAATCGAGTAAATTGAGCATAATTAAAGTAGAGTGGT
 CAATATAATAATACGATACATGATAAAAGTCTGAAAACATTCCACACACAATAAG
 CCTAAGCACTACTGATCACACAATCCATATTCTTCTAGGCTTAAACACATAACTA
 AGAAACATGACAGTTAAGGTTAGCAACACCATGCCCTATGTTTAAGTAGTTAC
 TAAATAATTAGATAGACAATGATGGCACCAGCAAACCTTTAGCCTTAATTATTCA
 AGAAGATGGAAGGTAATGATGTCAGAGGCAGAGGGTGCATGAAGGCCATGGTGG
 GGTAAATAGTGGTGTGGTCTGGTGGAAACGAAGAGCGTAAGCACGTGACCCCTCG
 ATTGGTATCCAAGAACTGAGTCGAATCTCCTCATTGTCTACTAGTCATAGTGA

FIGURA 6 (CONTINUACIÓN)

GGATCTCAGCTCTAGTCCTAATTACCAAAACATGTTTAAGCCATATTCCACAT
 AACTAAAACCAATATTATTAAATCGAAATGACATGACATTGATATATGATTAAC
 AGGAGATATTCGATAAGAGAATCACAAGAAAACATAAGAGTGAUTGATCAATCTA
 CACTACATTCTAGATATCTTTAACCTGGCTCCGTATCCATGAATGTATTATGCCT
 AAACCAACATTGAGAAAACAGAAAGAGATATTACCAAGCTCATGTATGAGATTCT
 TTTGTATGTCTTGTGACTTTGTTCTGCAATTGTGAAACCACAAGAGAAAATGATGA
 AGAAATATTCCCTCCCAAAAAATTATGAAGATATACCCAAGTTATTATGACATC
 ACTATATAGTATACTTGGTATTTCTTATCCATTAGCTAAACACACACACACAC
 ACAAAATATAATTGCCAAGAAAAGTGTATCAACAGTTAGTAAGTTGATAAATT
 AATTGACACAAAATTGTTGAGACAAATAATGATAAAAATGGAAAAGTGAUTGG
 CATAGAAGATGTGACCTTTCTTGGTGGTCTCGATCTGATTCCAAGAGATTTGAA
 CTGTACAGTACAATACAATTACATATTATCCAACAGTAAACAAAATGTTCTT
 TTCTAAAACATATACACAATAAAATTCAATTAAATTCAATTCAACTG
 AAGCTTGGTATAGGATTAATTGAAACCCCTAGGTAAAGGGAAAAGTATGAAGAGA
 ACCTTGCCTCGCGAACGAGTTGAAAGTGTCCATTCACTCCTCAAGACGACGC
 AGCTCCTGAATGTCAAGCTCGCCAAACACTCACCTAGCCTCTGCCTACAATT
 AAATATGTATATTCGTAATTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 AGAGAAAACAAATACTTGATCTGAGTCCGGAGATTCTATTGTCTCCACAGTTCC
 TCTTGGTTCTGCATTGCTGCAGCAAAACGCATCAAGAATTAAACCAACCAGCGA
 AACACATATCAAGAAGAAGAAAGATCTAAGAAGGAAAAACCTCATATTGA
 GTGGCCCAAACATCGACATCAGAAATAGTTGGTACAGATCTACGATCTCCTCGTT
 CTATAACCAACAAAATACCAAAAAAGTAGTGGTTAATTATAAGAAAACAAAGCAA
 CTCATTAGCTTAATTAAAGAGTGGTGTAGAGAGATGGTGTACGTGGTGTAGGG
 CTGATATACTCATGAAGCTTGTGGAGCTAGAGAACATGATAATCGAAACCCCTAGCA
 TCACACAAAACCGTGAGCTCATGTGCTTCTGAATAAACCTTCTCTTCTTGAAT
 ACGTCACTTGTCTGTTCTGGTTCTCTATCCTCTTGATCTGGATCTCCCTCGCC
 ATATTCTCTCTCTTGTAAATCTTTGTTGAAGAGATTGGTGGAGAGGACAAGA
 GATATAATGAGAAGAAGAAAGTGAATAGAAAGAGAGAGATGGAGTTGAAGAAGT
 AAAGGGTCCACTGAGTTACTAAAAATGGAAAGTATTGCTTAATCCATGAAAGGTA
 AGTTCAGAAGTTGATGAAAACTAAGTGAATAATTCCATTATGTAAACTGATT
 TTTAATTACAAGACACTGTTAAATCAACGGACATCGATGGTTGTATAATTACAAGA
 CACTGCTAAATCAATTGCACTTGGTAAAAGAGGGGGACCAAAGCTAAAACCTG
 GTTGCTTTCGTCAATTATTAAACGGAGCTCCGTTAGCTCTACTTGTGTTGTTA
 GACGGTGACAAACTTAACGGCGTTGAAACGTCAAATCAGTGCAGGTGGATTGGT
 TGGCTTTACACTTACCTAAAGTAAACAAAAGTGTGTTAGGTGTGAACAAGTT
 ATATTGGTGTAAAAGCCAATTCTACTTTGTCTGTTAATTGGGAAATAAGAAG
 AAAGAGATATATGTTGGATTAATCGTCACTCCATCTAACTCAAATCTTTATAT
 GTTCCCCAAATGATTACAATAATTCCGGATTGCTTAACTAAAGATCCTGG
 AAATAGAATATTACCCCTAGGGTTAAGAAGGAATCACAAGTTAGGAGAAAAGTCT
 TTGTTGTAGATAATGTTGGTCTAACCTCAAACAGTCCATCAAATCAGATATATG
 ATGAAACTCTTTCTTGTGCTCGACCAGTTAGTGTGTTCAAAACTAAGTTGG
 CCACAGTTACTGTTGTTGAAA